Az információszámítás néhány fontos fogalma és eredménye.

1. Az entrópia és feltételes entrópia fogalma és tulajdonságai.

Annak érdekében, hogy megértsük az entrópia fogalmát és azt, hogy milyen problémák vezettek annak megalkotásához tekintsük a következő kérdést. Tudunk-e nyerni a totón, ha jól ismerjük a totóban szereplő csapatok erejét, és ezért nagy valószínűséggel meg tudjuk tippelni a mérkőzések eredményét? Mivel igazán nagy nyereményt csak telitalálatos szelvénnyel lehet nyerni, foglalkozzunk azzal a kérdéssel, hogy hány szelvényt kell kitölteni a telitalálat elérése érdekében annak, aki ért és annak aki nem ért a futballhoz. Azt várjuk, hogy egy futballhoz értőnek sokkal jobbak a nyerési esélyei. Ez így is van. De ahhoz, hogy ezt jobban megértsük és a problémát alaposabban vizsgálhassuk először meg kell fogalmazni a kérdést pontosabban.

Ha biztos telitalálatot szeretnénk elérni, akkor hiába tudjuk az eredményeket nagy, de nem 100 százalékos biztonsággal eltalálni, célunkat csak úgy érhetjük el, ha minden lehetséges kimenetre fogadunk. Ebben az esetben tehát nem tudunk mást tenni, mint egy olyan fogadó, aki semmit sem tud az egyes mérkőzések valószínű eredményéről. Más azonban a helyzet, ha megelégszünk azzal, hogy nagy, mondjuk 0.95 valószínűséggel nyerjünk. Ekkor a futballhoz értő határozott előnyben van a futballhoz nem értővel szemben. Neki sokkal kevesebb szelvényt kell kitölteni e cél elérése érdekében, mint a másiknak. Másrészt elképzelhető, hogy érdemes totózni, ha 10000 szelvény kitöltésével tudjuk biztosítani a majdnem biztos nyerést, de nem érdemes akkor, ha ehhez 100000 szelvényt kell kitöltenünk.

A fenti kérdés természetesen elvezet a következő problémához. Tegyük fel, hogy n mérkőzés van, ezek eredménye egymástól független, és meg tudjuk itélni, hogy a pályaválasztó p_1, a vendégcsapat p_2 valószínűséggel nyer, és p_3 valószínűséggel lesz az eredmény döntetlen. Tegyük fel, hogy ezek a p_1, p_2 és p_3 valószínűségek mindegyik találkozón ugyanazok a számok, és az egyes mérkőzések eredményei egymástól függetlenek. Jelöljük a pályaválasztó nyerésének bekövetkeztét 1-gyel, a vendégcsapatét 2-vel, a döntetlen eredményt pedig x-szel, úgy ahogy az a totóban szokás. Arra vagyunk kiváncsiak, hogy hány tippet, hány n hosszúságú 1, 2, x sorozatot kell megadnunk ahhoz, hogy p valószínűséggel ezen tippek valamelyike tartalmazza az összes mérkőzés helyes végeredményét. A p szám egy az 1 számnál kicsit kisebb rögzített (azaz a mérkőzések n számától nem függő) szám, és minket az érdekel, hogy körülbelül hány tippet kell megadnunk célunk elérése érdekében akkor, ha n, azaz a mérkőzések száma nagyon nagy. Világos, hogy ez a szám függ a p_1, p_2 és p_3 számoktól, azaz attól, hogy milyen biztonsággal tudjuk megtippelni az eredményeket. Az, hogy semmit nem tudunk a lehetséges végeredményről azt jelenti, hogy $p_1 = p_2 = p_3 = \frac{1}{3}$.

Először pongyolán fogalmazom meg az eredményeket, és azokra egy heurisztikus indoklást adok, majd megadom az állítások és felhasznált fogalmak pontos megfogalmazását az általános esetben, és ismertetem a precíz bizonyításokat.

Meg tudjuk mondani, hogyan kell kitölteni a szelvényeket, ha pontosan k tippet tehetünk, és az a célunk, hogy a lehető legnagyobb valószínűséggel legyen telitalálatunk. Ha $k = 1$, akkor a mérkőzéssorozat legvalószínűbb eredményére érdemes tippelni. Ha

$k = 2$, akkor a két legvalószínűbb eredményre tippeljünk, és általános k esetén a legjobb stratégia a k legvalószínűbb eredményre fogadni. Ezután ha meghatározzuk, hogy melyik az a legkisebb $k = k(n)$ szám, amelyre a k legvalószínűbb eredmény valamelyike legalább p valószínűséggel bekövetkezik, akkor megoldjuk a feladatot. Ennek a $k = k(n)$ számnak a pontos meghatározása azonban nehéz. Ennél sokkal egyszerűbb az alábbi érvelés, amely a nagy számok (gyenge) törvénye segítségével jó közelítő értéket ad a keresett $k = k(n)$ számra.

Be fogjuk látni, hogy akkor érhető el, hogy majdnem 1 valószínűséggel lesz telitalálatunk n mérkőzés tippelése során, ha körülbelül 2^{Hn} szelvényt töltünk ki alkalmas módon, ahol H egy a p_1, p_2 és p_3 valószínűségektől függő szám. Azaz, exponenciálisan sok szelvényt kell kitölteni, de az hogy milyen H együttható szerepel a kitevőben attól függ, hogy milyen biztonsággal tudjuk eltalálni a mérkőzések eredményét. Ennek a H számnak a kiszámítása vezet el az entrópia fogalmának bevezetéséhez.

A mérkőzéssorozat eredménye egy n hosszúságú véletlen 1, 2 és x jelekből álló sorozat, ahol mindegyik jel a többiektől függetlenül p_1 valószínűséggel vesz fel 1 p_2 valószínűséggel 2 és p_3 valószínűséggel x értéket. Célunk viszonylag kevés sorozat kiválasztása úgy, hogy annak valószínűsége, hogy a megjelenő véletlen sorozat ezen sorozatok valamelyike legyen majdnem 1. Vegyük észre, hogy a nagy számok (gyenge) törvénye szerint majdnem minden sorozat olyan, hogy körülbelül np_1 darab 1 értéket, np_2 darab 2 értéket és np_3 darab x értéket tartalmaz. Nevezzünk egy ezzel a tulajdonsággal rendelkező sorozatot tipikusnak. Nagy n számra elég a tipikus sorozatokra tippelni, mert annak a valószínűsége, hogy valamelyik nem tipikus sorozat jelenik meg majdnem nulla. Ezért azt kell meghatároznunk, hogy hány tipikus sorozat van.

A tipikus sorozatok számát a következő heurisztikus érvelés segítségével határozhatjuk meg. Egy rögzített tipikus sorozat megjelenésének a valószínűsége körülbelül $p_1^{np_1}p_2^{np_2}p_3^{np_3}$, a tipikus sorozatok összvalószínűsége majdnem 1, ezért a tipikus sorozatok száma körülbelül $\frac{1}{p_1^{np_1}p_2^{np_2}p_3^{np_3}} = 2^{nH}$, ahol $H = -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3$. Itt és a továbbiakban is a log jel 2-es alapú logaritmust fog jelenteni. A természetes logaritmust az ln kifejezés fogja jelölni.

Megfogalmazom és bebizonyítom a fenti heurisztikus érvelés segítségével kapott eredmény egy természetes általánosítását. Előtte ismertetem az eredményben megjelenő entrópia fogalmát.

Entrópia definiciója. *Legyen ξ egy értékeit egy véges vagy megszámlálhatóan végtelen $X = \{x_1, x_2, \dots\}$ halmazon felvevő valószínűségi változó, amelyre $P(\xi = x_j) = p(x_j)$, $j = 1, 2, \dots$, ahol $\sum_j p(x_j) = 1$. A ξ valószínűségi változó entrópiája a*

$$H(\xi) = -\sum_j p(x_j) \log p(x_j)$$

esetleg végtelen értéket felvevő mennyiség, ahol log *a 2-es alapú logaritmust jelöli.*

1. Megjegyzés. *Kényelmi okokból megengedjük, hogy az entrópia fenti definiciójában $P(\xi = x_j) = 0$ legyen bizonyos x_j értékekre. Annak érdekében, hogy a definició ebben az esetben is értelmes legyen bevezetjük a $0 \log 0 = 0$ konvenciót.*

2. Megjegyzés. *Ha több $\xi_1, \dots, \xi_k$ véges vagy megszámlálható sok értéket felvevő valószínűségi változónk van, akkor ezek együttes $H(\xi_1, \dots, \xi_k)$ entrópiáját úgy definiáljuk, hogy a $\xi_1, \dots, \xi_k$ valószínűségi változó sorozatot természetes módon azonosítjuk a $(\xi_1, \dots, \xi_k)$ véletlen vektorral, és annak entrópiáját definiáljuk, mint a $H(\xi_1, \dots, \xi_k)$ entrópiát. Természetesen a tekintett véletlen vektor eloszlását a*

$$P((\xi_1, \dots, \xi_k) = (x_{i_1}^{(1)}, \dots, x_{i_k}^{(k)})) = P(\xi_1 = x_{i_1}^{(1)}, \dots, \xi_k = x_{i_k}^{(k)})$$

képlet definiálja minden $x_{i_1}^{(1)}, \dots, x_{i_k}^{(k)}$ sorozatra.

Vegyük észre, hogy egy ξ valószínűségi változó entrópiája nem függ attól, hogy ξ milyen értékeket vesz fel. Az csak a valószínűségi mezőnek a ξ valószínűségi változó által meghatározott particiójától függ, azaz attól a particiótól, amelynek elemei a ξ valószínűségi változó nívóhalmazai, vagyis azok a halmazok, ahol ξ valamilyen rögzített értéket vesz fel. Hasonló megjegyzés érvényes a később bevezetendő feltételes entrópiára is.

A következő eredményt fogjuk bizonyítani.

Tétel független valószínűségi változókból álló tipikus sorozatok számáról. *Legyen ξ egy értékeit valamely véges $X = \{x_1, x_2, \dots, x_r\}$ halmazon fölvevő valószínűségi változó, $\xi_1, \dots, \xi_n$ pedig független, a ξ valószínűségi változóval azonos eloszlású valószínűségi változók sorozata. Ekkor minden $\varepsilon > 0$ és $\delta > 0$ számhoz létezik olyan $n_0 = n_0(\varepsilon, \delta)$ küszöbindex, hogy $n \ge n_0$ esetén minden a $P((\xi_1, \dots, \xi_n) \in A) \ge \delta$ feltételt teljesítő $A = \{(x_{j_1}^{(k)}, \dots, x_{j_n}^{(k)}),\ 1 \le k \le L\}$ X halmazbeli n-hosszúságú sorozatokból álló halmaz L elemszáma teljesíti az $L = |A| \ge 2^{(1-\varepsilon)nH(\xi)}$ egyenlőtlenséget.*

Megfordítva, létezik olyan $\bar{A} = \{(x_{j_1}^{(k)}, \dots, x_{j_n}^{(k)}),\ 1 \le k \le \bar{L}\}$ $\bar{L}$ darab X halmazbeli n-hosszúságú sorozatból álló halmaz, amelyre $P((\xi_1, \dots, \xi_n) \in \bar{A}) \ge 1 - \delta$, és az $\bar{A}$ halmaz $\bar{L}$ elemszáma teljesíti az $\bar{L} = |\bar{A}| \le 2^{(1+\varepsilon)nH(\xi)}$ egyenlőtlenséget. A fenti egyenlőtlenségekben $H(\xi)$ a ξ valószínűségi változó entrópiáját jelöli.

Megjegyzés. A fenti tétel akkor is érvényes, ha a ξ valószínűségi változó nem csak véges sok, hanem megszámlálhatóan végtelen sok értéket is felvehet. Egy kiegészítésben ismertetem ennek az általánosabb eredménynek a bizonyítását. Ez a bizonyítás néhány új gondolatot igényel. Egy később tárgyalandó híres eredménynek, az úgynevezett Shannon–McMillan–Breiman tételnek egy speciális és egyszerűen igazolható esetét fogjuk felhasználni.

Bizonyítás. Jelölje B az összes olyan X-beli elemekből álló n hosszúságú $x_{j_1}, \dots, x_{j_n}$ sorozatok halmazát, amely sorozatok legalább $np(x_k)(1 - \varepsilon/2)$ multiplicitással tartalmazzák az x_k jelet minden $1 \le k \le r$ indexre, ahol $p(x_k) = P(\xi = x_k)$. A nagy számok gyenge törvénye szerint $P((\xi_1, \dots, \xi_n) \in B) \ge 1 - \frac{\delta}{2}$, ha $n \ge n_0(\varepsilon, \delta)$ alkalmas $n_0(\varepsilon, \delta)$ küszöbindexszel. Ezért $P((\xi_1, \dots, \xi_n) \in A \cap B) \ge \frac{\delta}{2}$, ha $P((\xi_1, \dots, \xi_n) \in A) \ge \delta$. Elég belátni, hogy az $A \cap B$ halmaz számossága nagyobb, mint $2^{n(1-\varepsilon)H(\xi)}$. Ennek érdekében definiáljuk a következő mennyiségeket. Legyen $s(k, x^{(n)})$ az $x^{(n)} = (x_{j_1}, \dots, x_{j_n})$ sorozatban levő x_k jelek száma minden $1 \le k \le r$ indexre. Ekkor minden $x^{(n)} =$

$(x_{j_1},\dots,x_{j_n}) \in A\cap B$, illetve általánosabban minden $x^{(n)}=(x_{j_1},\dots,x_{j_n})\in B$ sorozatra

$$P((\xi_1,\dots,\xi_n)=(x_{j_1},\dots,x_{j_n}))$$
$$=\prod_{k=1}^{r} p(x_k)^{s(k,x^{(n)})} \le p(x_1)^{n(1-\varepsilon/2)p(x_1)}\cdots p(x_r)^{n(1-\varepsilon/2)p(x_r)} = 2^{-n(1-\varepsilon/2)H(\xi)},$$

és mivel $P((\xi_1,\dots,\xi_n)\in A\cap B)\ge \frac{\delta}{2}$, az $A\cap B$ halmaz számossága nagyobb, mint $\frac{\delta}{2}2^{n(1-\varepsilon/2)H(\xi)} \ge 2^{n(1-\varepsilon)H(\xi)}$, ha az $n_0(\varepsilon,\delta)$ küszöbindexet elég nagyra választjuk.

Olyan $\bar{A}$ halmazt, amelyre $P((\xi_1,\dots,\xi_n)\in\bar{A})>1-\delta$, és elemszáma teljesíti a kívánt felső becslést választhatuk úgy, mint az összes olyan n hosszúságú X-beli elemeket tartalmazó $(x_{j_1},\dots,x_{j_n})$ sorozatból álló halmazt, amely sorozatok legfeljebb $np(x_k)(1+\varepsilon)$ multiplicitással tartalmazzák az x_k jelet minden $1\le k\le r$ indexre. Ismét a nagy számok gyenge törvényére hivatkozva kapjuk, hogy $P((\xi_1,\dots,\xi_n)\in\bar{A})\ge 1-\delta$, ha $n\ge n_0(\varepsilon,\delta)$ alkalmas $n_0(\varepsilon,\delta)$ küszöbindexszel. Másrészt minden $x^{(n)}=(x_{j_1},\dots,x_{j_n})\in\bar{A}$ sorozatra

$$P((\xi_1,\dots,\xi_n)=(x_{j_1},\dots,x_{j_n}))$$
$$=\prod_{k=1}^{r} p(x_k)^{s(k,x^{(n)})} \ge p(x_1)^{n(1+\varepsilon)p(x_1)}\cdots p(x_r)^{n(1+\varepsilon)p(x_r)} = 2^{-n(1+\varepsilon)H(\xi)},$$

és mivel az ilyen sorozatok összvalószínűsége kisebb vagy egyenlő mint 1, innen következik, hogy az $\bar{A}$ halmaz számossága kisebb, mint $2^{n(1+\varepsilon)H(\xi)}$.

A fent bizonyított eredményt a következő módon is interpretálhatjuk. Tekintsük egy ξ valószínűségi változóval azonos eloszlású, független valószínűségi változók n-hosszúságú sorozatait, és válasszuk ki e véletlen sorozatok p-ed részét alkalmas módon, úgy hogy csak viszonylag kevés sorozatot kelljen kiválasztanunk. A p-ed rész kifejezés itt azt jelenti, hogy a kiválasztott sorozatok összvalószínűsége legalább p. Az ebben a feladatban szereplő p szám teljesíti a $0<p<1$ egyenlőtlenséget, egyébként tetszőlegesen választhatjuk. Azt láttuk be, hogy célunkat elérhetjük $2^{nH(\xi)+o(n)}$ alkalmasan választott sorozat megadásával, de kevesebbel már nem. Ennek az eredménynek megfogalmazhatjuk az alábbi következményét. Ha a tekintett véletlen sorozatok nagy részét meg akarjuk jelölni különböző, de azonos hosszúságú véletlen 0–1 sorozatokkal, ahol a 'nagy részét' kifejezés azt jelenti, hogy a sorozatok kis (alkalmasan választott) ε valószínűségi részét figyelmen kívül hagyhatjuk, akkor az egyes sorozatokat körülbelül $nH(\xi)$ hosszúságú 0–1 sorozatokkal kell megjelölnünk. Ezt szokás úgy interpretálni, hogy a tekintett véletlen sorozat egyes tagjainak a megnevezéséhez $H(\xi)$ bit szükséges, azaz ennyi információ kell annak megismeréséhez.

Olyan problémát tekintettünk, amelynek vizsgálatában természetes módon megjelent az entrópia fogalma. Tekintsük ennek a problémának egy olyan változatát, ahol az előbb tárgyalt feladathoz hasonlóan egy véletlen sorozatot akarunk nagy valószínűséggel eltalálálni, de rendelkezünk bizonyos plusz információval. Nevezetesen ismerjük egy

másik, a minket érdeklő véletlen sorozattal kapcsolatban levő véletlen sorozat értékeit, és ezt a plusz információt is fel kívánjuk használni. A probléma jobb megértése érdekében tekintsük a korában vizsgált totózásról szóló feladat egy olyan változatát, amelyben ilyen kérdés merül fel.

A következő feladatot vizsgáljuk. Megint egy mérkőzéssorozat eredményeire akarunk jól tippelni a totón. Viszont a mérkőzések előtti napon az egyes találkozókon résztvevő együttesek ifjúsági csapatai is játszanak egymás ellen, és annak eredményét megismerhetjük a totószelvény kitöltése előtt. Az, hogy az ifjúsági csapatok milyen eredményt érnek el, hogy vannak felkészülve, információt ad a nagy csapatok felkészültségéről is, és ez megváltoztatja megitélésünket a lehetséges végeredmények valószínűségéről. A totószelvények kitöltésénél érdemes ezt az információt is figyelembe venni. A kérdés az, hogy hogyan vegyük ezt figyelembe, és ezen plusz információk felhasználása esetén hány szelvényt kell kitöltenünk annak érdekében, hogy nagy valószínűséggel telitalálatot érjünk el.

Fogalmazzuk meg a feladatot pontosabban. Tekintsük n mérkőzéspár eredményeit, amelyeket jelöljünk a (ξ_l, η_l), $1 \le l \le n$, jelpárokkal. (Az l-ik mérkőzéspár eredménye a totó l-ik fordulójában szereplő felnőtt és a nekik megfelelő ifjúsági csapatok mérkőzésének az eredménye, amelyeket ξ_l-lel illetve η_l-lel jelölünk.) Tegyük fel, hogy ezek a (ξ_l, η_l) vektorok egymástól függetlenek, és azonos eloszlásúak, továbbá ezt az eloszlást ismerjük. Vezessük be az $r(i,j) = P(\xi_l = i, \eta_l = j)$ valószínűségeket, ahol az i és j változók az 1, 2 és x értékeket veheti fel. A kérdés az, hogy ismerve az $\eta_1, \dots, \eta_n$ valószínűségi változók értékeit, hány n hosszúságú 1, 2, x sorozatot kell (alkalmasan) a totószelvényen megadni, ha azt akarjuk elérni, hogy ezen tippsorozatok valamelyike majdnem 1 valószínűséggel megegyezzen a véletlen $\xi_1, \dots, \xi_n$ sorozattal. Most is feltesszük, hogy a mérkőzések n száma nagy.

Először a feladat heurisztikus megoldását ismertetem, majd megfogalmazok egy általánosabb eredményt, és megadom annak a bizonyítását.

Jelölje $p(i) = r(i,1) + r(i,2) + r(i,x)$ a ξ_l és $q(j) = r(1,j) + r(2,j) + r(x,j)$ az η_l valószínűségi változók eloszlását, ahol i és j az 1, 2 és x értékeket veszi fel, és jelölje $r(i|j) = \frac{r(i,j)}{q(j)} = P(\xi_l = i | \eta_l = j)$ a ξ_l valószínűségi változó feltételes eloszlását feltéve az η_l valószínűségi változó értékét. Tekintsük az olyan $y_{v_1}, \dots, y_{v_l}$ sorozatokat, amelyek körülbelül $nq(j)$ darab j jelet tartalmaznak, ahol $j = 1, 2$ vagy x. Ha az $\eta_1, \dots, \eta_n$ sorozat, azaz az ifjúsági mérkőzések eredményeinek a sorozata ilyen arányban veszi fel ezeket az értékeket, akkor tippeljünk úgy, hogy az összes olyan tippet megadjuk, amelyekben azon körülbelül $nq(j)$ mérkőzés közül, amelyeknek ifjúsági mérkőzés megfelelőjében az $\eta_l = j$ eredmény született körülbelül $nq(j)r(i|j) = np(i,j)$ mérkőzés eredményét tippeljük i-nek, $i = 1, 2$ vagy x. Ha az η_l, $1 \le l \le n$, mérkőzések eredményei nem teljesítik a kívánt feltételt, akkor a totószelvényeket tetszőlegesen kitölthetjük, csupán arra ügyelve, hogy ne töltsünk ki túl sok szelvényt.

A nagy számok törvényéből következik, hogy az adott módon kitöltve a szelvényeket majdnem egy valószínűséggel lesz telitalálatunk. Azt kell még megbecsülünk, hogy hány szelvényt töltöttünk ki. Ezt az előző feladatban alkalmazott heurisztikus érveléshez hasonlóan tehetjük meg. Elég csak azokat az eseteket nézni, amelyekben az $\eta_1, \dots, \eta_l$

valószínűségi változók által felvett $y_{v_1}, \dots, y_{v_n}$ eredmények körülbelül $nq(j)$ számú j eredményt tartalmaznak, $j = 1, 2$ vagy x. Ebben az esetben egy olyan $\xi_1 = x_{u_1}$, $\dots$, $\xi_n = x_{u_n}$ eredménynek, amelyre tippeltünk a feltételes valószínűsége a $P(\cdot|\eta_1 = y_{v_1}, \dots, \eta_n = y_{v_n})$ feltételes eloszlás szerint körülbelül $\prod_{i\in\{1,2,x\}} \prod_{j\in\{1,2,x\}} r(i|j)^{nr(i,j)}$, és mivel annak feltételes valószínűsége a tekintett feltételes valószínűség szerint, hogy lesz telitalálatunk majdnem 1, ezért a kitöltött szelvények száma körülbelül

$$\prod_{i\in\{1,2,x\}} \prod_{j\in\{1,2,x\}} r(i|j)^{-nr(i,j)} = 2^{n\bar{H}},$$

ahol $\bar{H} = -\sum_{i\in\{1,2,x\}} \sum_{j\in\{1,2,x\}} r(i,j) \log \frac{r(i,j)}{q(j)}$. Tehát körülbelül $2^{n\bar{H}}$ szelvény kitöltésével tudunk majdnem biztosan telitalálatot elérni.

Annak érdekében, hogy a fenti heurisztikus tárgyalásban kapott eredményt pontosabban megfogalmazhassuk vezessük be a következő fogalmat.

A feltételes entrópia definiciója. *Legyen adva két ξ és η valószínűségi változó, amelyek értékeiket egy véges vagy megszámlálhatóan végtelen $X = \{x_1, x_2, \dots\}$ illetve $Y = \{y_1, y_2, \dots\}$ halmazon veszik fel, és együttes eloszlásuk valamely $r(x_i, y_j) = P(\xi = x_i, \eta = y_j)$, $x_i \in X$, $y_j \in Y$, függvény. Vezessük be a $q(y_j) = P(\eta = y_j) = \sum_{x_i\in X} r(x_i, y_j)$ valószínűségeket is minden $y_j \in Y$ értékre. A ξ valószínűségi változó feltételes entrópiája az η valószínűségi változóra vonatkozólag a*

$$H(\xi|\eta) = -\sum_{x_i\in X} \sum_{y_j\in Y} r(x_i, y_j) \log \frac{r(x_i, y_j)}{q(y_j)}$$

esetleg végtelen értéket felvevő mennyiség, ahol log *a 2-es alapú logaritmust jelöli. A fenti definióban megengedjük az $r(x_i, y_j) = 0$ lehetőséget bizonyos (x_i, y_j) párokra. Annak érdekében, hogy a fenti összeget ekkor is értelmezhessük bevezetjük a $0 \log 0 = 0 \log \frac{0}{0} = 0$ konvenciót.*

Megjegyzés. *Abban az esetben, ha $H(\eta) < \infty$, érvényes a*

$$H(\xi|\eta) = -\sum_{x_i\in X} \sum_{y_j\in Y} r(x_i, y_j) \log r(x_i, y_j) + \sum_{y_j\in Y} q(y_j) \log q(y_j) = H(\xi, \eta) - H(\eta)$$

azonosság.

A következő tétel megfogalmazásának érdekében vezessünk be néhány jelölést. Legyen adva egy $X = \{x_1, \dots, x_r\}$ halmaz, és jelölje X^n az X halmazbeli elemekből álló n hosszúságú $(x_{i_1}, \dots, x_{i_n})$ sorozatok halmazát, ahol $x_{i_k} \in X$ minden $1 \le k \le n$ indexre. Hasonlóan, legyen adva egy $Y = \{y_1, \dots, y_s\}$ halmaz, és jelölje Y^n az $y_{j_k} \in Y$, $1 \le k \le n$, elemekből álló n hosszúságú $(y_{j_1}, \dots, y_{j_n})$ sorozatok halmazát. Továbbá, jelölje $X^n \times Y^n$ az $(x^{(n)}, y^{(n)}) = ((x_{i_1}, \dots, x_{i_n}), (y_{j_1}, \dots, y_{j_n}))$, $x^{(n)} \in X^n$, $y^{(n)} \in Y^n$

sorozatok halmazát. Adva egy $A \subset X^n \times Y^n$ halmaz és egy $(y_{j_1}, \dots, y_{j_n}) \in Y^n$ sorozat legyen $A(y_{j_1}, \dots, y_{j_n})$ az A halmaz metszete az $X^n \times \{(y_{j_1}, \dots, y_{j_n})\}$ halmazzal, azaz

$$A(y_{j_1}, \dots, y_{j_n}) = \{(x_{i_1}, \dots, x_{i_n})\colon\ ((x_{i_1}, \dots, x_{i_n}), (y_{j_1}, \dots, y_{j_n})) \in A\}.$$

Jelölje $|A|$ egy (véges) A halmaz elemszámát.

Az alább megfogalmazott tétel jobb megértése érdekében leírom előbb annak heurisztikus tartalmát. A következő problémával foglalkozunk. Ha adva van független, egyforma eloszlású (ξ_j, η_j), $1 \le j \le n$, véletlen vektorok egy sorozata, akkor az $\eta^{(n)} = (\eta_1, \dots, \eta_n)$ sorozat ismeretében meg akarunk adni egy olyan viszonylag kevés $x^{(n)} \in X^n$ sorozatból álló halmazt, amely nagy valószínűséggel tartalmazza az $\xi^{(n)} = (\xi_1, \dots, \xi_n)$ véletlen sorozatot. Ez azt jelenti, hogy olyan $A \subset X^n \times Y^n$ halmazt akarunk definiálni, amelyre a $P(\xi^{(n)} \in A(y^{(n)}) | \eta^{(n)} = y^{(n)})$ feltételes valószínűség viszonylag nagy, és az $A(y^{(n)})$ halmaz $|A(y^{(n)})|$ számossága viszonylag kicsi az $y^{(n)} \in Y^n$ sorozatok nagy részére, azaz az $y^{(n)}$ sorozatok egy olyan alkalmas $B \subset Y^n$ halmazára, amelyre $P(\eta^{(n)} \in B)$ majdnem 1-gyel egyenlő. A tétel azt állítja, hogy ahhoz, hogy a tekintett feltételes valószínűségek teljesítsék a kívánt feltételt az A halmazt úgy kell választani, hogy $|A(y^{(n)})| > 2^{(1-\varepsilon)nH(\xi|\eta)}$ legyen minden $y^{(n)} \in B$ sorozatra. Másrészt meg lehet adni a kívánt tulajdonsággal rendelkező A halmazt úgy, hogy az $|A(y^{(n)})| < 2^{(1+\varepsilon)nH(\xi|\eta)}$ egyenlőtlenség teljesüljön minden $y^{(n)} \in B$ sorozatra.

Tétel független valószínűségi változókból álló és egy másik független valószínűségi változókból álló véletlen sorozat szerint tipikus sorozatok számáról. *Legyen adva egy (ξ, η) véletlen vektor, amelynek koordinátái közül ξ értékeit egy $X = \{x_1, \dots, x_r\}$ η pedig egy $Y = \{y_1, \dots, y_s\}$ véges halmazon veszi fel. Legyen adva független és a (ξ, η) párral azonos eloszlású $((\xi_1, \eta_1), \dots, (\xi_n, \eta_n))$ véletlen vektorok egy sorozata. Ekkor minden $0 < \varepsilon, \delta < 1$ számpárhoz van egy olyan $n_0 = n_0(\varepsilon, \delta)$ küszöbindex, hogy minden $n \ge n_0$ számra igaz a következő állítás. Ha $A \subset X^n \times Y^n$ olyan halmaz, amelyre $P(((\xi_1, \dots, \xi_n), (\eta_1, \dots, \eta_n)) \in A | \eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) \ge \delta$ minden $(y_{j_1}, \dots, y_{j_n}) \in Y^n$ sorozatra, akkor van olyan $B_0 \subset Y^n$ halmaz, amelyre $P((\eta_1, \dots, \eta_n) \in B_0) > 1 - \delta$, és az $A(y_{j_1}, \dots, y_{j_n})$ halmaz számossága teljesíti az $|A(y_{j_1}, \dots, y_{j_n})| > 2^{n(1-\varepsilon)H(\xi|\eta)}$ egyenlőtlenséget minden $(y_{j_1}, \dots, y_{j_n}) \in B_0$ sorozatra.*

Igaz a következő fordított irányú egyenlőtlenség is. Léteznek olyan $A \subset X^n \times Y^n$ és $B_1 \subset Y^n$ halmazok, amelyekre $P((\eta_1, \dots, \eta_n) \in B_1) \ge 1 - \delta$,

$$P(((\xi_1, \dots, \xi_n), (\eta_1, \dots, \eta_n)) \in A | \eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) \ge 1 - \delta$$

minden $(y_{j_1}, \dots, y_{j_n}) \in B_1$ sorozatra, és az $A(y_{j_1}, \dots, y_{j_n})$ halmaz számossága teljesíti az $|A(y_{j_1}, \dots, y_{j_n})| \le 2^{n(1+\varepsilon)H(\xi|\eta)}$ egyenlőtlenséget minden $(y_{j_1}, \dots, y_{j_n}) \in B_1$ sorozatra. A fenti egyenlőtlenségekben $H(\xi|\eta)$ a ξ valószínűségi változó feltételes entrópiáját jelöli az η valószínűségi változóra vonatkozólag.

Feladat: *Bizonyítsuk be a fenti tétel olyan élesebb formáját, amelyben megengedjük azt is, hogy az X és Y halmazok megszámlálhatóan végtelen számosságúak legyenek.*

A tétel bizonyítása. Vezessük be az $r(x_i, y_j) = P(\xi = x_i, \eta = y_j)$, $q(y_j) = P(\eta = y_j) = \sum_{i=1}^{r} r(x_i, y_j)$ és $r(x_i|y_j) = P(\xi = x_i|\eta = y_j) = \frac{r(x_i,y_j)}{q(y_j)}$, $1 \le i \le r$, $1 \le j \le s$, mennyiségeket. Adva egy $y^{(n)} = (y_{j_1}, \dots, y_{j_n}) \in Y^n$ vektor, jelölje $s(y^{(n)}, j)$ az $y^{(n)}$ sorozatban szereplő y_j elemek számát, $1 \le j \le s$, és definiáljuk a $B_0 \in Y^n$ halmazt, mint

$$B_0 = \{y^{(n)} = (y_{j_1}, \dots, y_{j_n})\colon\ y^{(n)} \in Y^n,\ s(y^{(n)}, j) \ge (1 - \tfrac{\varepsilon}{4})nq(y_j) \\ \text{minden } 1 \le j \le s \text{ indexre}\}.$$

Ekkor a nagy számok törvénye szerint $P((\eta_1, \dots, \eta_n) \in B_0) \ge 1 - \delta$, ha $n \ge n_0$ egy elég nagy $n_0 = n_0(\varepsilon, \delta)$ küszöbindexszel.Definiáljuk az $\ell(x^{(n)}, y^{(n)}, i, j)$ mennyiséget minden $(x^{(n)}, y^{(n)}) = ((x_{i_1}, \dots, x_{i_n}), (y_{j_1}, \dots, y_{j_n})) \in X^n \times Y^n$ sorozatra és $1 \le i \le r$, $1 \le j \le s$ indexekre úgy, mint az $(x^{(n)}, y^{(n)})$ vektorban szereplő olyan (x_{i_k}, y_{j_k}), $1 \le k \le n$, párok számát, amelyek egyenlőek az (x_i, y_j) párral. Adva egy $y^{(n)} = (y_{j_1}, \dots, y_{j_n}) \in Y^{(n)}$ vektor definiáljuk a következő $C(y^{(n)}) \subset X^n$ halmazt.

$$C(y^{(n)}) = \{x^{(n)} = (x_{i_1}, \dots, x_{i_n})\colon\ (x^{(n)}, y^{(n)}) \in X^n \times Y^n, \\ \ell(x^n), y^{(n)}, i, j) \ge (1 - \tfrac{\varepsilon}{2})nr(x_i, y_j) \quad \text{minden } 1 \le i \le r,\ 1 \le j \le s \text{ párra}\}.$$

Megmutatom a nagy számok törvénye segítségével, hogy

$$P((\xi_1, \dots, \xi_n) \in C(y^{(n)})|\eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) \ge 1 - \frac{\delta}{2}, \text{ ha } y^{(n)} = (y_{j_1}, \dots, y_{j_n}) \in B_0.$$

Ennek érdekében először rögzítek egy i és j számot, és megmutatom, hogy az $\eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}$ feltétel mellett, ahol $y^{(n)} = (y_{j_1}, \dots, y_{j_n}) \in B_0$, annak feltételes valószínűsége, hogy a $(\xi_1, \dots, \xi_n)$ sorozat olyan $x^{(n)} = (x_{i_1}, \dots, x_{i_n})$ értéket vesz fel, amelyre az $(x^{(n)}, y^{(n)})$ vektornak legalább $n(1 - \frac{\varepsilon}{2})$ koordinátája egyenlő az (x_i, y_j) párral, nagyobb, mint $(1 - \frac{\delta}{2rs})$. Valóban, ha $y^{(n)} \in B_0$, akkor e feltételes valószínűség feltételében $s(y^{(n)}, j) \ge nq(y_j)(1 - \frac{\varepsilon}{4})$ olyan k index van, amelyre $\eta_k = y_j$. A ξ_k valószínűségi változók együttes eloszlása ezen k indexekre a tekintett feltételes eloszlás szerint megegyezik $s(y^{(n)}, j) \ge (1 - \frac{\varepsilon}{4})nq(y_j)$ független, $r(\cdot|y_j)$ eloszlású valószínűségi változó együttes eloszlásával. Ezért ez a sorozat a nagy számok törvénye szerint több, mint $(1 - \frac{\delta}{2rs})$ valószínűséggel tartalmaz legalább $s(y^{(n)}, j)(1 - \frac{\varepsilon}{4})r(x_i, y_j) \ge n(1 - \frac{\varepsilon}{4})^2 q(y_j)r(x_i|y_j) \ge n(1 - \frac{\varepsilon}{2})r(x_i, y_j)$ számú x_i elemet minden $1 \le i \le r$ indexre, ha $n \ge n_0(\varepsilon, \delta)$. Mivel ez az egyenlőtlenség minden (i, j), $1 \le i \le r$, $1 \le j \le s$ párra érvényes, innen következik a bizonyítani kívánt egyenlőtlenség is.

A most bizonyított egyenlőtlenségből és a tétel feltételeiből következik, hogy tetszőleges $y^{(n)} = (y_{j_1}, \dots, y_{j_n}) \in B_0$ vektorra

$$P(((\xi_1, \dots, \xi_n), (\eta_1, \dots, \eta_n)) \in A, (\xi_1, \dots, \xi_n) \in C(y^{(n)})|\eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) \ge \frac{\delta}{2},$$

ami úgy is írható, hogy

$$P((\xi_1, \dots, \xi_n) \in C(y^{(n)}) \cap A(y^{(n)})|\eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) \ge \frac{\delta}{2},$$

ahol $A(y^{(n)}) = A(y_{j_1}, \dots, y_{j_n})$. Felhasználva ezt az egyenlőtlenséget és a $C(y^{(n)})$ halmaz elemeinek a tulajdonságait belátjuk, hogy $|A(y^{(n)}) \cap C(y^{(n)})| \ge 2^{n(1-\varepsilon)H(\xi|\eta)}$. Ennek érdekében vegyük észre, hogy tetszőleges $y^{(n)} = (y_{j_1}, \dots, y_{j_n}) \in B_0$ és $(x_{i_1}, \dots, x_{i_n}) \in C(y^{(n)})$ vektorokra

$$P(\xi_1 = x_{i_1}, \dots, \xi_n = x_{i_n} | \eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) = \prod_{i=1}^{r} \prod_{j=1}^{s} r(x_i|y_j)^{\ell(x^{(n)}, y^{(n)}, i, j)}$$

$$\le \prod_{i=1}^{r} \prod_{j=1}^{s} r(x_i|y_j)^{(1-\varepsilon/2)nr(x_i, y_j)} = 2^{-n(1-\varepsilon/2)H(\xi|\eta)}.$$

Az utolsó két egyenlőtlenségből következik, hogy

$$|A(y_{j_1}, \dots, y_{j_n})| \ge |A(y_{j_1}, \dots, y_{j_n}) \cap C(y^{(n)})| \ge \frac{\delta}{2} 2^{n(1-\varepsilon/2)H(\xi|\eta)} \ge 2^{(1-\varepsilon)nH(\xi|\eta)},$$

ha $y^{n)} = (y_{j_1}, \dots, y_{j_n}) \in B_0$.

A másik irányú becslést az alkalmas $B_1 \subset Y^n$ és $A \subset X^n \times Y^n$ halmazok definiciójával hasonlóan bizonyíthatjuk. Legyen

$$B_1 = \{y^{(n)} = (y_{j_1}, \dots, y_{j_n})\colon y^{(n)} \in Y^n,\ s(y^{(n)}, j) \le (1 + \tfrac{\varepsilon}{2})nq(y_j) \text{ minden } 1 \le j \le s \text{ indexre}\},$$

és

$$A = \{(x^{(n)}, y^{(n)}) = ((x_{i_1}, \dots, x_{i_n}), (y_{j_1}, \dots, y_{j_n}))\colon (x^{(n)}, y^{(n)}) \in X^n \times Y^n,\ y^{(n)} \in B_1,$$
$$\ell(x^{(n)}, y^{(n)}, i, j) \le (1 + \varepsilon)nr(x_i, y_j) \text{ minden } 1 \le i \le r,\ 1 \le j \le s \text{ párra}\}.$$

Az előző eset érveléséhez hasonlóan bizonyíthatjuk a nagy számok törvénye segítségével, hogy $P((\eta_1, \dots, \eta_n) \in B_1) \ge 1 - \delta$, ha $n \ge n_0$, és

$$P(((\xi_1, \dots, \xi_n), (\eta_1, \dots, \eta_n)) \in A | \eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n})$$
$$= P((\xi_1, \dots, \xi_n), \in A(y_{j_1}, \dots, y_{j_n}) | \eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) \ge 1 - \delta$$

minden $(y_{j_1}, \dots, y_{j_n}) \in B_1$ sorozatra. Továbbá

$$P(\xi_1 = x_{i_1}, \dots, \xi_n = x_{i_n} | \eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) = \prod_{i=1}^{r} \prod_{j=1}^{s} r(x_i|y_j)^{\ell(x^{(n)}, y^{(n)}, i, j)}$$

$$\ge \prod_{i=1}^{r} \prod_{j=1}^{s} r(x_i|y_j)^{(1+\varepsilon)nr(x_i, y_j)} = 2^{-n(1+\varepsilon)H(\xi|\eta)},$$

ha $(y_{j_1}, \dots, y_{j_n}) \in B_1$ és $(x_{i_1}, \dots, x_{i_n}) \in A(y_{j_1}, \dots, y_{j_n})$. Felhasználva a (triviális) $P((\xi_1, \dots, \xi_n) \in A(y_{j_1}, \dots, y_{j_n}) | \eta_1 = y_{j_1}, \dots, \eta_n = y_{j_n}) \le 1$ egyenlőtlenséget innen kapjuk, hogy $|A(y_{j_1}, \dots, y_{j_n})| \le 2^{n(1+\varepsilon)H(\xi|\eta)}$ minden $(y_{j_1}, \dots, y_{j_n}) \in B_1$ sorozatra.

A most bizonyított eredményt az előző eredményhez hasonlóan a következőképp is interpretálhatjuk. Legyen adva egy véges sok értéket felvevő (ξ, η) véletlen vektor (valójában az alább megfogalmazott állítás akkor is igaz, ha ez a véletlen vektor végtelen sok értéket is felvehet, de ezt nem bizonyítottuk be), és független véletlen vektoroknak egy ezzel a véletlen vektorral azonos eloszlású, n-hosszúságú (ξ_1, η_1), ..., (ξ_n, η_n) sorozata. Ismerve az $\eta_1, \dots, \eta_n$ valószínűségi változók értékeit, ki akarunk választani viszonylag kevés sorozatot úgy, hogy ezek egyike majdnem biztos megegyezzék a $\xi_1, \dots, \xi_n$ véletlen sorozat értékével. A majdnem biztos kiválasztás itt azt jelenti, hogy rögzítünk egy kis $\varepsilon > 0$ számot, megadjuk az $(\eta_1, \dots, \eta_n)$ sorozatok lehetséges értékeinek egy legalább $1 - \varepsilon$ mértékű halmazát, és amennyiben az $(\eta_1, \dots, \eta_n)$ sorozat ezek valamelyikével egyenlő, akkor kijelöljük n hosszúságú sorozatok egy viszonylag kevés elemből álló halmazát úgy, hogy annak a valószínűsége, hogy a $(\xi_1, \dots, \xi_n)$ véletlen sorozat megegyezik ezek valamelyikével legalább $1 - \varepsilon$. A kiválasztott sorozatok halmaza függhet az $(\eta_1, \dots, \eta_n)$ vektor értékétől. Azt láttuk be, hogy ezt megtehetjük $2^{nH(\xi|\eta)+o(n)}$ alkalmasan választott sorozat segítségével, de kevesebbel már nem. Ennek az eredménynek megfogalmazhatjuk az alábbi következményét.

Meg akarjuk nevezni a $(\xi_1, \dots, \xi_n)$ sorozatokat az $\eta_1, \dots, \eta_n$ sorozat ismeretében $m = m(n)$ hosszúságú az $\eta_1, \dots, \eta_n$ sorozattól függő választással 0–1 sorozatokkal úgy, hogy $1-\varepsilon$ valószínűséggel egy $(\xi_1, \dots, \xi_n)$ sorozatot megnevezünk, és rögzített $\eta_1, \dots, \eta_n$ sorozat megjelenése esetén két különböző $\xi_1, \dots, \xi_n$ sorozat elnevezése különböző. Ez lehetséges $m = nH(\xi|\eta) + o(n)$ hosszúságú 0–1 sorozatokkal, de rövidebb sorozatokkal nem. Ezt szokás úgy interpretálni, hogy az $\eta_1, \dots, \eta_n$ sorozat ismeretében a $\xi_1, \dots, \xi_n$ sorozat egyes tagjainak a megnevezéséhez $H(\xi|\eta)$ bit szükséges, azaz az η_k változók értékének az ismeretében ennyi információ kell az egyes ξ_k véletlen változók megismeréséhez.

Megfogalmazom és bebizonyítom az entrópiával és feltételes entrópiával kapcsolatos legfontosabb egyenlőtlenségeket. Ezek mindegyike heurisztikus szinten 'nyilvánvaló' következménye az entrópia és feltételes entrópia szemléletes tartalmának.

Tétel az entrópiával és feltételes entrópiával kapcsolatos fontos egyenlőtlenségekről. *Legyenek ξ, η és ζ véges vagy megszámlálható sok értéket fölvevő valószínűségi változók. Ezek teljesítik a következő egyenlőtlenségeket:*

a1.) $H(\xi) \geq 0$, *és egyenlőség akkor és csak akkor érvényes, ha a ξ valószínűségi változó egy valószínűséggel egy konstanssal egyenlő.*

a2.) Ha a ξ valószínűségi változó n értéket vesz fel, akkor $H(\xi) \leq \log n$. *Egyenlőség akkor és csak akkor érvényes, ha a ξ valószínűségi változó által felvett $x_1, \dots, x_n$ értékekre* $P(\xi = x_k) = \frac{1}{n}$ *minden* $1 \leq k \leq n$ *indexre.*

b.) $H(\xi, \eta) \leq H(\xi) + H(\eta)$, *illetve kissé általánosabban* $H(\xi|\eta) \leq H(\xi)$. *Egyenlőség akkor és csak akkor teljesül a második, általánosabb egyenlőtlenségben, ha ξ és η függetlenek, vagy* $H(\xi|\eta) = \infty$.

c.) $H(\eta) \leq H(\xi, \eta)$, *illetve kissé általánosabban* $H(\xi|\eta) \geq 0$. *Egyenlőség akkor és csak akkor teljesül a második, általánosabb egyenlőtlenségben, ha* $\xi = f(\eta)$ *valamely* $f(\cdot)$

függvénnyel, azaz ξ az η függvénye.

d.) $H(\xi|\eta,\zeta) \le H(\xi|\eta)$. *Abban az esetben, ha létezik olyan a (ξ,η) véletlen vektortól független Z valószínűségi változó, amelyre $\zeta = h(\eta, Z)$ valamely alkalmas h függvénnyel, akkor egyenlőség áll fenn.*

Megjegyzés. A feltételes entrópiáról szóló b) és c) pontban megfogalmazott egyenlőtlenségek akkor is érvényesek, ha $H(\eta) = \infty$. Ebben az esetben ezek az állítások többet mondanak, mint a nekik megfelelő az entrópiáról megfogalmazott egyenlőtlenségek.

Következmény.

a.) $H(f(\xi)) \le H(\xi)$ *tetszőleges $f(\cdot)$ függvényre. Vegye fel egy ξ valószínűségi változó értékeit egy végtelen $X = \{x_1, x_2, \dots\}$ halmazban, legyen $H(\xi) < \infty$, és definiáljuk minden $K \ge 1$ számra egy olyan $\xi^{(K)}$, valószínűségi változót, amelyre $\xi = x_j$ esetén $\xi^{(K)} = x_j$, ha $j \le K$, és $\xi^{(K)} = x^*$ valamely $x^* \ne x_s$, $1 \le s \le K$, értékkel, ha $j > K$. Ezzel a választással $H(\xi^{(K)}) \le H(\xi)$, és minden $\varepsilon > 0$ számhoz létezik olyan $K_0 = K_0(\varepsilon)$ index, hogy $H(\xi^{(K)}) \ge H(\xi) - \varepsilon$, ha $K \ge K_0$.*

b.) *Érvényesek a $H(\xi,\eta|\zeta) = H(\eta|\xi,\zeta) + H(\xi|\zeta)$ és $H(\xi,\eta) = H(\eta|\xi) + H(\xi)$ azonosságok. Továbbá $H(\xi,\eta|\zeta) \ge H(\xi|\zeta)$, és egyenlőség akkor és csak akkor áll fenn, ha vagy $\eta = f(\xi,\zeta)$ alkalmas f függvénnyel vagy $H(\xi|\zeta) = \infty$. Továbbá $H(\xi,\eta|\zeta) \le H(\xi|\zeta) + H(\eta|\zeta)$.*

A következmény bizonyítása. Az a) rész bizonyítása: $H(\xi) = H(\xi, f(\xi)) \ge H(f(\xi))$ a tétel c) pontja szerint. Innen következik a $H(\xi^{(K)}) \le H(\xi)$ egyenlőtlenség, az X téren definiált $f(x_j) = x_j$, ha $j \le K$, és $f(x_j) = x^*$, ha $j \ge K$ választással. Legyen $p(x_j) = P(\xi = x_j)$, $j = 1, 2, \dots$. A $H(\xi) = -\sum_{j=1}^{\infty} p(x_j) \log p(x_j) < \infty$ feltételből következik, hogy létezik olyan $K_0 = K_0(\varepsilon)$ index, hogy $-\sum_{j=1}^{K_0} p(x_j) \log p(x_j) \ge H(\xi) - \varepsilon$. Ilyen $K_0 = K_0(\varepsilon)$ választással igaz az a) rész utolsó állítása is.

A következmény b) részében szereplő első azonosság következik a $H(\xi,\eta|\zeta) = \sum_{i,j,k} P(\xi = i, \eta = j, \zeta = k) \log \frac{P(\xi=i,\eta=j,\zeta=k)}{P(\zeta=k)}$ és $H(\eta|\xi,\zeta) + H(\xi|\zeta) = \sum_{i,j,k} P(\xi = i, \eta = j, \zeta = k)(\log \frac{P(\xi=i,\eta=j,\zeta=k)}{P(\xi=i,\zeta=k)} + \log \frac{P(\xi=i,\zeta=k)}{P(\zeta=k)})$ relációkból. A második azonosság hasonlóan indokolható. A b) rész további állítása következik a $H(\xi,\eta|\zeta) = H(\eta|\xi,\zeta) + H(\xi|\zeta)$ azonosságból és a tétel c) illetve d) részének az állításából.

Értsük meg, hogy az előbb megfogalmazott tétel egyenlőtlenségei az entrópia és feltételes entrópia szemléletes tartalmának megfelelő tulajdonságokat fejeznek ki. Az a1) tulajdonság azt mondja, hogy a ξ valószínűségi változó által felvett érték megismeréséhez pozitív információ szükséges, kivéve azt az elfajuló esetet, amikor ξ értéke (ismert) konstans, és ezért nulla információ is elegendő. Az a2) állítás szerint egy n értéket felvevő valószínűségi változó értékének a megismeréséhez akkor kell a legtöbb információ, ha minden értéket egyforma valószínűséggel vesz fel, amit úgy is interpretálhatunk, hogy

azonkívül, hogy tudjuk, hogy ξ n értéket vesz fel, semmilyen plusz információnk nincs annak viselkedéséről.

A b) tulajdonság azt fejezi ki, hogy egy η valószínűségi változó értékének az ismerete csökkentheti egy ξ valószínűségi változó értékének megismeréséhhez szükséges információt. Akkor nincs csökkenés, ha ξ és η függetlenek, és ezért η ismerete semmilyen értékes információt nem ad ξ viselkedéséről. A c) tulajdonság szerint egy ξ valószínűségi változó értékének a megismeréséhez pozitív információ szükséges egy η valószínűségi változó értékének az ismeretében is. Akkor elég nulla információ, ha ξ az η ismert függvénye.

A d) tulajdonság jelentése az, hogy egy ξ valószínűségi változó értékének a megismeréséhez kevesebb információ szükséges, ha egy η valószínűségi változó értékén kívül egy másik ζ valószínűségi változó értékét is ismerjük. Semmilyen nyereséget nem jelent viszont ζ ismerete, ha az a már ismert η és egy mind a ξ mind az η valószínűségi változótól független valószínűségi változó függvénye.

A tétel bizonyításában fontos szerepet játszik egy egyszerű állítás, amelyet a bizonyítás jobb áttekinthetősége érdekében külön lemmában fogalmazok meg.

Lemma az $x\log x$ függvény viselkedéséről. *A $g(x) = x\log x$, ha $x > 0$, $g(0) = 0$ függvény egy a $[0,\infty)$ félegyenesen folytonos, szigorúan konvex függvény, amelyre $g(0) = g(1) = 0$.*

A lemma bizonyítása. Könnyen látható, hogy $g(x)$ folytonos függvény a $[0,\infty)$ félegyenesen, és $g(0) = g(1) = 0$. Ezenkívül $g''(x) = \frac{\log e}{x} > 0$ minden $x > 0$ számra, ahonnan következik, hogy $g(x)$ szigorúan konvex függvény.

A tétel bizonyítása. Jelölje $X = \{x_1, x_2, \dots\}$ a ξ, $Y = \{y_1, y_2, \dots\}$ az η és $Z = \{z_1, z_2, \dots\}$ a ζ valószínűségi változó értékeit. Az a) b) és c) részben használjuk a $p(x_i) = P(\xi = x_i)$, $q(y_j) = P(\eta = y_j)$ és $r(x_i, y_j) = P(\xi = x_i, \eta = y_j)$ jelölést. Az a1) állítás nyilvánvaló, mert $-p(x_i)\log p(x_i) > 0$, ha $0 < p(x_i) < 1$, és $0\log 0 = 1\log 1 = 0$. Az a2) állítás következik a

$$-H(\xi) = n\sum_{i=1}^{n}\frac{1}{n}g(p(x_i)) \geq ng\left(\frac{1}{n}\sum_{i=1}^{n}p(x_i)\right) = ng\left(\frac{1}{n}\right) = -\log n$$

egyenlőtlenségből, ahol $g(x)$ a lemmában szereplő konvex függvény. Mivel a $g(\cdot)$ függvény szigorúan konvex egyenlőség csak a $p(x_i) = \frac{1}{n}$, $1 \leq i \leq n$, esetben lehetséges.

A b) állítás bizonyítása érdekében írjuk fel a

$$H(\xi|\eta) = -\sum_{i,j} q(y_j)\frac{r(x_i,y_j)}{q(y_j)}\log\frac{r(x_i,y_j)}{q(y_j)} = -\sum_i\left(\sum_j q(y_j)g\left(\frac{r(x_i,y_j)}{q(y_j)}\right)\right)$$
$$\leq -\sum_i\left(g\left(\sum_j q(y_j)\frac{r(x_i,y_j)}{q(y_j)}\right)\right) = -\sum_i g(p(x_i)) = H(\xi)$$

egyenlőtlenséget. E számolásban felhasználtuk a $g(x) = x \log x$ függvény konvexitását a $q(y_j)$, $q(y_j) > 0$, $\sum_j q(y_j) = 1$, súlyokkal, azaz azt, hogy $\sum_j q(y_j) g(u_j) \le g\left(\sum_j q(y_j) u_j\right)$ minden $u_1 \ge 0$, $u_2 \ge 0$, ... számsorozatra, és a $\sum_j r(x_i, y_j) = p(x_i)$ azonosságot. Felhasználva a $g(\cdot)$ függvény szigorú konvexitását kapjuk, hogy egyenlőség akkor és csak akkor lehetséges, ha vagy $H(\xi|\eta) = \infty$ vagy bármely rögzített i indexre $\frac{r(x_i,y_j)}{q(y_j)} = \alpha_i$ valamely α_i számmal minden j indexre. Ez azt jelenti. hogy $r(x_i, y_j) = \alpha_i q(j)$, és ezt az azonosságot összegezve a j változóra azt kapjuk, hogy $p(x_i) = \sum_j r(x_i, y_j) = \alpha_i$, azaz $r(x_i, y_j) = p(x_i)q(y_j)$ minden i és j indexre, tehát a ξ és η valószínűségi változók függetlenek.

A c) állítás bizonyítása érdekében vegyük észre, hogy a

$$H(\xi|\eta) = -\sum_{i,j} r(x_i, y_j) \log \frac{r(x_i, y_j)}{q(y_j)}$$

azonosság jobboldalán csak nem-pozitív tagok szerepelnek a $0 \le \frac{r(x_i,y_j)}{q(y_j)} \le 1$ reláció miatt. Innen következik, hogy $H(\xi|\eta) \ge 0$. (A fenti összeg tagjainak azonos előjelét implicite a b) rész bizonyításában is felhasználtuk. Ez feljogosított minket arra, hogy a bizonyításban vizsgált összeget a számunkra megfelelő módon átrendezzük.) Egyenlőség csak akkor lehetséges, ha mindegyik $\frac{r(x_i,y_j)}{q(y_j)}$ tag vagy nullával vagy eggyel egyenlő. Ez azt jelenti, hogy létezik egy olyan $i(j)$ index, hogy $\frac{r(x_{i(j)},y_j)}{q(y_j)} = 1$, azaz $P(\xi = x_{i(j)}|\eta = y_j) = 1$. Ezért egyenlőség akkor és csak akkor teljesül, ha $\xi = f(\eta)$ valamely alkalmas f függvénnyel.

A d) rész állításának vizsgálatában vezessük be a $p(x_i) = P(\xi = x_i)$ és $q(y_j) = P(\eta = y_j)$ mennyiségek mellett az $u(x_i, y_j) = P(\xi = x_i, \eta = y_j)$, $v(y_j, z_k) = P(\eta = y_j, \zeta = z_k)$ valamint a $t(x_i, y_j, z_k) = P(\xi = x_i, \eta = y_j, \zeta = z_k)$ mennyiségeket is. Ezekkel a jelölésekkel felírhatjuk, hogy

$$\begin{aligned}
H(\xi|\eta,\zeta) &= -\sum_{x_i,y_j,z_k} t(x_i, y_j, z_k) \log \frac{t(x_i, y_j, z_k)}{v(y_j, z_k)} \\
&= -\sum_{x_i,y_j} \left(\sum_{z_k} q(y_j) \frac{v(y_j, z_k)}{q(y_j)} \frac{t(x_i, y_j, z_k)}{v(y_j, z_k)} \log \frac{t(x_i, y_j, z_k)}{v(y_j, z_k)} \right) \\
&= -\sum_{x_i,y_j} q(y_j) \left(\sum_{z_k} \frac{v(y_j, z_k)}{q(y_j)} g\left(\frac{t(x_i, y_j, z_k)}{v(y_j, z_k)} \right) \right) \\
&\le -\sum_{x_i,y_j} q(y_j) g\left(\sum_{z_k} \frac{v(y_j, z_k)}{q(y_j)} \frac{t(x_i, y_j, z_k)}{v(y_j, z_k)} \right) \\
&= -\sum_{x_i,y_j} q(y_j) g\left(\sum_{z_k} \frac{t(x_i, y_j, z_k)}{q(y_j)} \right) = -\sum_{x_i,y_j} q(y_j) g\left(\frac{u(x_i, y_j)}{q(y_j)} \right)
\end{aligned}$$

$$= -\sum_{x_i,y_j} P(\xi = x_i, \eta = y_j) \log \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} = H(\xi|\eta).$$

E számolásokban felhasználtuk azt, hogy a $g(\cdot)$ függvény konvex, $\sum\limits_k \frac{v(y_j,z_k)}{q(y_j)} = 1$, és $\sum\limits_k \frac{t(x_i,y_j,z_k)}{q(y_j)} = \frac{u(x_i,y_j)}{q(y_j)}$. A kapott egyenlőtlenségben azonosságot írhatunk abban a speciális esetben, ha $\frac{t(x_i,y_j,z_k)}{v(y_j,z_k)} = \alpha(i,j)$, azaz, ha ez a tört csak az csak az x_i és y_j változótól függ. Ez az egyenlőség teljesül abban a speciális esetben, ha $\zeta = Z$, ahol Z a (ξ,η) vektortól független valószínűségi változó, mert ekkor $\frac{t(x_i,y_j,z_k)}{v(y_j,z_k)} = \frac{P(\xi=x_i,\eta=y_j,Z=z_k)}{P(\eta=y_j,\zeta=z_k)} = \frac{P(\xi=x_i,\eta=y_j)P(Z=z_k)}{P(\eta=y_j)P\zeta=z_k)} = \alpha(i,j)$. Ha $\zeta = h(\eta,Z)$ egy ilyen tulajdonságú Z valószínűségi változóval, akkor a már bizonyított állítások alapján felírhatjuk, hogy $H(\xi|\eta) = H(\xi|\eta,Z) = H(\xi|\eta,Z,\zeta) \le H(\xi|\eta,\zeta) \le H(\xi|\eta)$. Ezért az utolsó formulasorozatban mindenütt egyenlőség van, és $H(\xi|\eta,\zeta) = H(\xi|\eta)$.

Feladat:

Legyen $\xi_1,\dots,\xi_n$ egy Markov lánc. Bizonyítsuk be, hogy $H(\xi_n|\xi_{n-1},\dots,\xi_1) = H(\xi_n|\xi_{n-1})$.

Segítség: Lássuk be a d) rész bizonyításában szereplő $\frac{t(x_i,y_j,z_k)}{v(y_j,z_k)} = \alpha(i,j)$ azonosságot, ahol a $t(x_i,y_j,z_k)$ és $v(y_j,z_k)$ függvényeket a következő ξ, η és ζ valószínűségi változók segítségével definiáljuk: $\xi = \xi_n$, $\eta = \xi_{n-1}$, $\zeta = (\xi_{n-2},\dots,\xi_1)$. A kívánt azonosság következik a Markov tulajdonságból.

Kiegészítés: *Független, értékeiket esetleg végtelen halmazban felvevő valószínűségi változókból álló tipikus sorozatok számának a becslése.*

A *független valószínűségi változókból álló tipikus sorozatok számáról* szóló tételben becslést adtunk arra, hogy hány tipikus sorozatot tartalmaz egy ξ valószínűségi változóval azonos eloszlású független valószínűségi változók $\xi_1,\dots,\xi_n$ sorozata. Pontosabban fogalmazva azt becsültük meg, hogy rögzítve egy kis $\varepsilon > 0$ számot körülbelül hány sorozatot tartalmaz a $(\xi_1,\dots,\xi_n)$ véletlen sorozatok egy $1-\varepsilon$ mértékű alkalmasan választott részhalmaza. A válasz a ξ valószínűségi változó $H(\xi)$ entrópiájától függ. Ezt az eredményt csak abban az esetben láttuk be, ha a ξ valószínűségi változó csak véges sok értéket vehet fel. Ugyanakkor természetes azt várni, hogy a tétel állítása érvényes megszámlálhatóan végtelen sok értéket felvevő ξ valószínűségi változó esetén is. Belátjuk, hogy ez tényleg így van. Az egyszerű fogalmazás érdekében feltesszük, hogy $H(\xi) < \infty$, hiszen minket valójában csak ez az eset érdekel. A következő eredményt fogom bebizonyítani.

Tétel független, véges vagy megszámlálhatóan végtelen sok értéket felvevő valószínűségi változókból álló tipikus sorozatok számáról. *A független valószínűségi változókból álló tipikus sorozatok számáról megfogalmazott tétel állítása akkor is érvényes, ha az abban szereplő ξ valószínűségi változó értékeit valamely véges vagy*

végtelen $X = \{x_1, x_2, \dots, \}$ halmazon veszi fel, és mindössze annyit teszünk fel róla, hogy $H(\xi) < \infty$.

A tétel bizonyítása. Definiáljuk a $p(x_k) = P(\xi = x_k)$, $k = 1, 2, \dots$, függvényt, és adva független a ξ valószínűségi változóval azonos eloszlású, független valószínűségi változóknak egy $\xi_1, \dots, \xi_n$ sorozata vezessük be a $\zeta_j = -\log p(\xi_j)$, $1 \le j \le n$, valószínűségi változókat. Vegyük észre, hogy $E\zeta_j = H(\xi)$, és a nagy számok (gyenge) törvénye alapján

$$\frac{1}{n}\sum_{j=1}^{n} \zeta_j \Rightarrow H(\xi), \quad \text{ha } n \to \infty,$$

ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl. Jelölje $\eta_k(n)$ a $\xi_1, \dots, \xi_n$ sorozatban szereplő x_k elemek számát. Ekkor az előbb felírt formula ekvivalens a következő állítással:

$$\frac{1}{n}\sum_{k}(\eta_k(n) - np(x_k))\log p(x_k) \Rightarrow 0, \quad \text{ha } n \to \infty, \tag{A1}$$

Megmutatom, hogy a (A1) relációból következik a tétel állítása. E célból válasszunk olyan $\varepsilon_n \to 0$ és $\delta_n \to 0$ sorozatokat, amelyekre

$$P\left(\left|\sum_{k}(\eta_k(n) - np(x_k))\log p(x_k)\right| > n\varepsilon_n\right) \le \delta_n. \tag{A2}$$

Ez az (A1) reláció érvényessége miatt megtehető. Adva egy $x^{(n)} = (x_{j_1}, \dots, x_{j_n}) \in X^n$ sorozat jelölje $s(k, x^{(n)})$ azon r, $1 \le r \le n$, indexek számát, amelyekre $x_{j_r} = x_k$. Definiáljuk minden $n = 1, 2, \dots$ indexre az

$$A_1(n) = \left\{x^{(n)} = (x_{j_1}, \dots, x_{j_n})\colon\ x^{(n)} \in X^n,\ \sum_{k=1}^{\infty}(s(k, x^{(n)}) - np(x_k))\log p(x_k) \le n\varepsilon_n\right\}$$

és

$$A_2(n) = \left\{x^{(n)} = (x_{j_1}, \dots, x_{j_n})\colon\ x^{(n)} \in X^n,\ \sum_{k=1}^{\infty}(s(k, x^{(n)}) - np(x_k))\log p(x_k) \ge -n\varepsilon_n\right\}$$

halmazokat, ahol ε_n megegyezik az (A2) formulában szereplő ε_n számmal. Vegyük észre, hogy mivel $(\xi_1(\omega), \dots, \xi_n(\omega)) = (x_{j_1}, \dots, x_{j_n})$, $x^{(n)} = (x_{j_1}, \dots, x_{j_n})$ esetén $\eta_k(n)(\omega) = s(k, x^{(n)})$. Ezért az (A2) formula alapján $P((\xi_1, \dots, \xi_n) \in A_j(n)) \ge 1 - \delta_n \ge 1 - \delta/2$ mind $j = 1$ mind $j = 2$ indexszel, ha $n \ge n_0(\delta, \varepsilon)$ alkalmas n_0 küszöbindexszel. A fő részben ismertetett bizonyítást alkalmazhatjuk a most tárgyalt esetben is minimális változtatásokkal, ha megmutatjuk, hogy egy $x^{(n)} = (x_{j_1}, \dots, x_{j_n}) \in X^n$ vektorra

$$P((\xi_1, \dots, \xi_n) = (x_{j_1}, \dots, x_{j_n})) \le 2^{-n(1-\varepsilon)H(\xi)}, \quad \text{ha } x^{(n)} = (x_{j_1}, \dots, x_{j_n}) \in A_1(n).$$

és

$$P((\xi_1,\dots,\xi_n)=(x_{j_1},\dots,x_{j_n}))\ge 2^{-n(1+\varepsilon)H(\xi)},\quad \text{ha } x^{(n)}=(x_{j_1},\dots,x_{j_n})\in A_2(n),$$

feltéve, hogy $n\ge n_0(\varepsilon,\delta)$ egy alkalmas n_0 küszöbindexszel. Ezekből az egyenlőtlenségekből ugyanis a fő részben ismertetett bizonyítás módszerével következik a számunkra szükséges becslés az $A_1(n)$ illetve az $A_2(n)$ halmaz elemszámára.

A kívánt becslések bizonyításának az érdekében írjuk fel az alábbi azonosságot minden $x^{(n)}=(x_{j_1},\dots,x_{j_n})\in X^n$ vektorra.

$$\begin{aligned}
&P((\xi_1,\dots,\xi_n)=(x_{j_1},\dots,x_{j_n}))\\
&\quad=\prod_{k=1}^{\infty}p(x_k)^{s(k,x^{(n)})}=\prod_{k=1}^{\infty}p(x_k)^{np_k(x)}\prod_{k=1}^{\infty}p(x_k)^{s(k,x^{(n)})-np_k(x)}\\
&\quad=\exp\left\{\frac{1}{\log 2}\sum_{k=1}^{\infty}np(x_k)\log p(x_k)\right\}\exp\left\{\frac{1}{\log 2}\sum_{k=1}^{\infty}(s(k,x^{(n)})-np(x_k))\log p(x_k)\right\}.
\end{aligned}$$

Innen következik, hogy

$$P((\xi_1,\dots,\xi_n)=(x_{j_1},\dots,x_{j_n}))\le 2^{-nH(\xi)}\cdot 2^{n\varepsilon_n}\ge 2^{-n(1-\varepsilon/2)H(\xi)}$$

minden $x^{(n)}=(x_{j_1},\dots,x_{j_n})\in A_1(n)$ vektorra, és

$$P((\xi_1,\dots,\xi_n)=(x_{j_1},\dots,x_{j_n}))\ge 2^{-nH(\xi)}\cdot 2^{-n\varepsilon_n}\ge 2^{-n(1+\varepsilon)H(\xi)}$$

minden $x^{(n)}=(x_{j_1},\dots,x_{j_n})\in A_2(n)$ vektorra, ha $n\ge n_0(\varepsilon,\delta)$. Innen következik, hogy az $A_1(n)$ és $A_2(n)$ halmazok elemszáma teljesíti az $|A_1(n)|\ge 2^{n(1-\varepsilon)H(\xi)}$ és $|A_2(n)|\le 2^{n(1+\varepsilon)H(\xi)}$ egyenlőtlenségeket. Sőt az is igaz, hogy amennyiben $B\subset X^n$ olyan halmaz, amelyre $P((\xi_1,\dots,\xi_n)\in B)\ge\delta$, akkor $|A_1(n)\cap B|\ge 2^{n(1-\varepsilon)H(\xi)}$.

2. Forráskódolás és dekódolás.

Az információelmélet egyik legfontosabb problémája a következő kódolási problémának nevezett kérdés. Legyen adva valószínűségi változók valamely $\xi_1,\xi_2,\dots$ véges vagy végtelen sorozata, amelynek tagjai értéküket valamely X véges vagy megszámlálhatóan végtelen számosságú halmazban veszik fel. A $\xi_1,\xi_2\dots$ sorozatot forrásnak, a ξ_j valószínűségi változók által felvett értékek X halmazát pedig (forrás) ABC-nek szokás nevezni az irodalomban. Ennek a $\xi_1,\xi_2\dots$ sorozatnak az értékeit szeretnénk közölni valakivel, akit felhasználónak nevezünk. E cél elérése érdekében bizonyos jeleket leadunk a felhasználónak, aki ezeket a jeleket, esetleg némi hibával, megkapja. Azt az apparátust, amely ezeket a jeleket továbbítja csatornának nevezzük. A felhasználó, a csatornán keresztül megkapott jelek segítségével megpróbálja rekonstruálni az eredeti $\xi_1,\xi_2,\dots$ üzenetet. Tegyük fel, hogy a kiinduló $\xi_1,\xi_2,\dots$ sorozat egymás utáni jelei megérkeznek bizonyos sebességgel, és mi le tudjuk adni a csatornán a jeleinket a felhasználónak bizonyos sebességgel. A kérdés az, hogy mikor tudjuk egy a felhasználóval

korábban egyeztetett módszer segítségével elérni, hogy ő a kapott jelsorozat segítségével viszonylag kis hibával rekonstruálni tudja az eredeti $\xi_1, \xi_2, \dots$ forrást.

Az előbb megfogalmazott kódolási probléma egy tipikus esete az, ha egy szöveg egymás utáni betűi érkeznek (a szóközi szüneteket külön jelnek tekintjük), és ezt a folyamatosan érkező szöveget akarjuk közölni egy tőlünk távol levő ismerősünknek. Ennek érdekében le tudunk adni egymás után 0 és 1 jeleket egy távírón. Azt akarjuk elérni, hogy ismerősünk, aki e jeleket megkapja képes legyen viszonylag pontosan rekonstruálni az eredeti szöveget még akkor is, ha a jeltovábbításban időnként hibák lépnek fel. Azt vizsgáljuk, hogy ez mikor lehetséges. Célszerű az egyes betűknek bizonyos jelsorozatot (kódot) megfeleltetni. Úgy kívánjuk ezt tenni, hogy ismerősünk képes legyen a kapott jelsorozat segítségével viszonylag pontosan rekonstruálni (dekódolni) az eredeti sorozatot még akkor is, ha a leadott jelsorozat egyes jelei hibásan érkeznek meg hozzá.

Célunkat elérhetjük például úgy, hogy mindegyik betűnek ugyanolyan hosszú és egymástól különböző kódszót feleltetünk meg, és mindegyik kódszót egymás után százszor küldjük el a csatornán. Ekkor kicsi annak a valószínűsége, hogy egy ilyen 100-szor elküldött jel az esetek többségében helytelenül érkezik, így ismerősünk nagy valószínűséggel rekonstruálni tudja az eredeti üzenetet. A probléma az, hogy ilyen módszerrel az üzenet továbbítása túlságosan sok időt vesz igénybe, és ezért esetleg nem tudjuk követni az eredetileg érkező jelsorozat sebességét. Célunk tehát az, hogy egy olyan módszert dolgozzunk ki, amelynek segítségével az eredeti hírforrást viszonylag gyorsan és pontosan meg tudjuk ismertetni a felhasználóval. A módszer kidolgozásánál érdemes figyelembe venni azt, hogy a (véletlen) $\xi_1, \xi_2, \dots$ forrás milyen valószínűségi törvénynek tesz eleget, illetve, hogy milyen valószínűséggel következnek be bizonyos hibák, amikor a jeleket leadjuk a csatornán.

Az előbb megfogalmazott kódolási problémát érdemes két probléma vizsgálatának a segítségével megoldani. Az első probléma a következő. Adva egy rögzített n szám, és egy rögzített $V = \{v_1, v_2, \dots\}$ véges vagy megszámlálhatóan végtelen halmaz, tekintsük a

$$\xi_{ln+1}, \dots, \xi_{l(n+1)}, \qquad l = 1, 2, \dots,$$

blokkokat. Válaszuk ki ezenkívül a V halmazbeli elemeket tartalmazó n-hosszúságú sorozatoknak egy alkalmas $A = A(n) \subset V^n$ részhalmazát. (Itt, és a továbbiakban, V^n jelöli a V halmazbeli, és X^n az X halmazbeli elemeket tartalmazó n hosszúságú sorozatok halmazát.) Olyan $f\colon X^n \to A(n)$ és $g\colon A(n) \to X^n$ függvényeket keresünk, amelyekre

$$P(g(f(\xi_{ln+1}, \dots, \xi_{(l+1)n})) = (\xi_{ln+1}, \dots, \xi_{(l+1)n})) \geq 1 - \varepsilon \quad \text{minden } l = 0, 1, \dots \text{ indexre} \tag{2.1}$$

egy kis fix $\varepsilon > 0$ számmal. Ha találunk ilyen $f, g)$ függvénypárt akkor $f(\cdot)$ függvényt kódfüggvénynek, az $f(x_{p(1)}, \dots, x_{p(n)})$ sorozatot az $x_{p(1)}, \dots, x_{p(n)}$ sorozat kódjának nevezzük, a $g(\cdot)$ függvényt pedig dekódoló függvénynek hívjuk. Ebben az esetben azt mondjuk, hogy ε-nál kisebb hibával kódoltunk.Olyan $A(n) \subset V^n$ halmazt és a (2.1) formulát teljesítő $f(\cdot)$, $g(\cdot)$ függvénypárt szeretnénk találni, amelyekre az n blokkhossztól függő $A(n)$ halmaznak viszonylag kicsi az elemszáma. E feladat megoldását nevezzük forrás kódolásnak és forrás dekódolásnak.

A második feladat arról szól, hogy amikor a forrás kódolásaként kapott sorozat értékét alkalmas csatorna esetleg hibás közvetítése segítségével közöljük a felhasználóval, akkor ő hogyan tudja viszonylag kis hibával rekonstruálni a csatornán keresztül elküldött sorozatot. Ezt a feladatot, amelyet csatorna kódolásnak és dekódolásnak neveznek a következő fejezetben fogom tárgyalni, és magát a feladatot is csak ott fogom pontosan megfogalmazni.

E fejezet témája a forrás kódolás és dekódolás. Csak azzal az esettel fogok foglalkozni, amikor a $\xi_1, \xi_2, \dots$ forrás független és egyforma eloszlású véges vagy megszámlálhatóan végtelen sok értéket felvevő valószínűségi változók sorozata. Ebben az esetben egy jó forrás kódolás és dekódolás megtalálása viszonylag egyszerű feladat, az könnyen megtehető az előző fejezet eredményeinek a segítségével. Tárgyalni fogom továbbá ennek a feladatnak egy önmagában is érdekes változatát, amelyben egy véletlen sorozatot akarunk viszonylag rövid sorozattal úgy kódolni, hogy az hibátlanul dekódolható legyen. A kódolt sorozat hossza függhet a véletlentől, és e véletlen szóhossz várható értékét szeretnénk kicsivé tenni.

Megfogalmazom ezt a módosított feladatot pontosabban. Legyen adva egy ξ valószínűségi változó, amely értékeit egy véges $X = \{x_1, \dots, x_M\}$ halmazban veszi fel, és amelynek ismert a $P(\xi = x_i) = p(i)$, $1 \le i \le M$, eloszlása. (Az egyszerűség kedvéért feltettem, hogy a ξ valószínűségi változó X értékkészlete egy véges halmaz, bár több alább ismertetendő eredmény akkor is érvényes, ha az X halmaz számossága megszámlálhatóan végtelen is lehet.) Legyen adva egy véges d-elemű $Y = \{y_1, \dots, y_d\}$ halmaz, amelyet a továbbiakban ABC-nek fogunk nevezni, és a ξ valószínűségi változóval azonos eloszlású $\xi_1, \dots, \xi_l$ valószínűségi változóknak egy sorozata. Nevezzük az Y halmaz elemeiből álló véges $y_{j_1}, \dots, y_{j_n}$ sorozatokat szavaknak. Minden $x_i \in X$ elemnek meg akarunk feleltetni egy $u(x_i) = y_{j_1}^{(i)}, \dots, y_{j_{n(i)}}^{(i)}$ $n(i)$ hosszúságú Y halmaz elemeiből álló sorozatot, amelyet az x_i sorozat nevének fogunk nevezni. Úgy kívánjuk ezt a megfeleltetést csinálni, hogy ha egymás után felsorolják nekünk egy $x_{i_1}, \dots, x_{i_l}$ sorozat elemeinek $u(x_{i_1}), \dots, u(x_{i_l})$ neveit, akkor képesek legyünk ennek alapján az $x_{i_1}, \dots, x_{i_l}$ sorozatot egyértelműen rekonstruálni. Az ilyen $x_i \to u(x_i)$, $x_i \in X$, leképezéseket egyértelműen dekódolható kódolásoknak fogjuk nevezni. Ennek pontos definicióját alább ismertetem. Azzal a kérdéssel fogunk foglalkozni, hogy milyen kicsivé tudjuk tenni egy egyértelműen dekódolható $x_i \to u(x_i)$, $1 \le i \le M$, kódolás $n(i)$ hosszának a várható értékét, azaz a $\sum_{i=1}^{M} p(i)n(i)$ mennyiséget.

Az egyértelműen dekódolható kódolás definiciója. *Legyen $X = \{x_1, x_2, \dots\}$ egy véges vagy megszámlálhatóan végtelen számosságú halmaz és $Y = \{y_1, \dots, y_d\}$ egy másik d elemszámú halmaz (ABC). Egy az X halmaz elemeit az Y halmaz elemeiből álló véges elemszámú $u(x_i) = y_{j_1}^{(i)} \dots y_{j_{n(i)}}^{(i)}$ sorozatokba képező leképezését az X halmaznak az Y ABC szavaival végzett egyértelműen dekódolható kódolásának nevezünk, ha minden $x_{i_1}, \dots, x_{i_l}$, $l = 1, 2, \dots$, $x_{i_j} \in X$, $1 \le j \le l$, sorozatnak az $u(x_{i_1}), \dots, u(x_{i_l})$ sorozatot megfeleltetve teljesül az $u(x_{i_1}), \dots, u(x_{i_{l_1}}) \ne u(x'_{i_1}), \dots, u(x'_{i_{l_2}})$ reláció, ha $x_{i_1}, \dots, x_{i_{l_1}} \ne x'_{i_1}, \dots, x'_{i_{l_2}}$. (Megjegyzem, hogy az $u(x_i)$ sorozat $n(i)$ hossza függhet az*

$x_i \in X$ *elemtől.)*

Először az eredeti kódolási feladattal foglalkozom. Megfogalmazom azt az eredményt, amely megadja, hogy milyen $N(n)$ elemszámú $A(n) \subset V^n$ halmaz segítségével lehet megadni valamely független, egyforma eloszlású $\xi_1, \dots, \xi_n$ valószínűségi változókból álló forrásnak egy ε-nál kisebb hibájú, azaz a (2.1) relációt teljesítő $f(\cdot)$ kódolását és $g(\cdot)$ dekódolását. A most ismertetett tétel valójában az előző fejezet bizonyos eredményeinek egyszerű következménye.

Tétel független, egyforma eloszlású valószínűségi változókból álló forrás kis hibájú kódolásáról és dekódolásáról. *Legyen ξ egy értékeit valamely véges vagy megszámlálhatóan végtelen számosságú $X = \{x_1, x_2, \dots\}$ halmazon fölvevő valószínűségi változó, $\xi_1, \dots, \xi_n$ pedig független, a ξ valószínűségi változóval azonos eloszlású valószínűségi változók sorozata. Legyen ezenkívül adva egy $V = \{v_1, v_2, \dots\}$ halmaz. Jelölje X^n az X halmazból, V^n a V halmaz elemeiből álló n hosszúságú sorozatok halmazát. Válasszunk egy $N = N(n)$ elemszámú $A = A(n) \subset V^n$ halmazt. Minden $\varepsilon > 0$ és $\delta > 0$ számhoz létezik $n_0 = n_0(\varepsilon, \delta)$ küszöbindex úgy, hogy ha $n \ge n_0$ és $N(n) \ge 2^{(1+\delta)H(\xi)n}$, akkor léteznek olyan $f\colon X^n \to A(n)$ és $g\colon A(n) \to X^n$ függvények, amelyekre $P(g(f(\xi_1, \dots, \xi_n)) = (\xi_1, \dots, \xi_n)) \ge 1 - \varepsilon$. Megfordítva, ha $N(n) \le 2^{(1-\delta)H(\xi)n}$, és $n \ge n_0(\varepsilon, \delta)$ akkor minden $f\colon X^n \to A(n)$ és $g\colon A(n) \to X^n$ függvénypárra $P(g(f(\xi_1, \dots, \xi_n)) = (\xi_1, \dots, \xi_n)) \le \varepsilon$.*

Bizonyítás. Láttuk a független valószínűségi változókból álló tipikus sorozatok számáról szóló tétel bizonyításában (ha megszámlálhatóan végtelen számosságú halmazokkal akarunk dolgozni, akkor e tételnek a kiegészítésben tárgyalt általánosítását kell tekintenünk), hogy $n \ge n_0(\varepsilon, \delta)$ esetén létezik olyan $B \subset X^n$ halmaz, amelyre az B halmaz kevesebb, mint $2^{(1+\delta)H(\xi)n}$ elemet tartalmaz, és $P((\xi_1, \dots, \xi_n) \in B) \ge 1 - \varepsilon$. Válasszunk egy ilyen B halmazt, és soroljuk fel az elemeit, mint $B = \{x_1^{(n)}, x_2^{(n)}, \dots, x_{\bar{N}(n)}^{(n)}\}$ valamely $\bar{N}(n) \le 2^{(1+\delta)nH(\xi)}$ számmal. Soroljuk fel az $A(n)$ halmaz elemeit is, mint $A(n) = \{v_1^{(n)}, v^{(n)}, \dots, v_{N(n)}^{(n)}\}$. Ha $N(n) \ge 2^{(1+\delta)H(\xi)n}$, és $n \ge n_0(\varepsilon, \delta)$ akkor $\bar{N}(n) \le N(n)$. Definiáljuk ebben az esetben az $f(x)$ függvényt egy olyan $x = (x_1, \dots, x_n) \in X^n$ elemre, amelyre $x \in B$, és ezért felírható $x = x_k^{(n)}$, $1 \le k \le \bar{N}(n)$, alakban, úgy mint $f(x) = f(x_k^{(n)}) = v_k^{(n)}$. Ha $x \notin B$, legyen $f(x) = v_1^{(n)}$. Definiáljuk a $g(v_k^{(n))})$ függvényt, mint $g(v_k^{(n)}) = x_k^{(n)}$, ha $k \le \bar{N}(n)$. Ha $\bar{N}(n) < k \le N(n)$, akkor a $g(v_k^{(n)})$ függvényt tetszőleges módon definiálhatjuk. Ilyen választással $P(g(f(\xi_1, \dots, \xi_n)) = (\xi_1, \dots, \xi_n)) \ge P((\xi_1, \dots, \xi_n) \in B) \ge 1 - \varepsilon$.

A tétel második felének bizonyításában azt használjuk ki, hogy $n \ge n_0$ esetén létezik olyan $\bar{B} \subset X^n$ halmaz, amelyre $P((\xi_1, \dots, \xi_n) \in \bar{B}) \ge 1 - \frac{\varepsilon}{2}$, és a $\bar{B}$ halmaz elemei viszonylag nagy valószínűséggel jelennek meg. Pontosabban, $P((\xi_1, \dots, \xi_n) = (x_{j_1}, \dots, x_{j_n})) \le 2^{-(1-\frac{\delta}{2})H(\xi)n}$ az $(x_{j_1}, \dots, x_{j_n}) \in \bar{B}$ vektorokra. Ha $N(n) \le 2^{(1-\delta)H(\xi)n}$ esetén is lenne olyan $f(\cdot)$, $g(\cdot)$ függvénypár, amelyre $P(g(f(\xi_1, \dots, \xi_n)) = (\xi_1, \dots, \xi_n)) \ge \varepsilon$, akkor létezne olyan $B_0 \subset X^n$ halmaz, amelyre $P((\xi_1, \dots, \xi_n) \in B_0) \ge \varepsilon$, és az összes $x = (x_{j_1}, \dots, x_{j_n}) \in B_0$ vektor $f(x) \in A(n)$ képe különböző lenne. (A B_0 halmazt úgy választhatjuk, mint azon $x \in X^n$ vektorok halmazát, amelyekre $g(f(x)) = x$.)

Ekkor az is igaz lenne, hogy $P((\xi_1,\dots,\xi_n)\in B_0\cap\bar{B})\ge\frac{\varepsilon}{2}$, és minden $x\in B_0\cap\bar{B}$ vektorra $P((\xi_1,\dots,\xi_n)=x)\le 2^{-(1-\frac{\delta}{2})H(\xi)n}$. Innen következne, hogy az $B_0\cap\bar{B}$ halmaz elemszáma nagyobb, mint $\frac{\varepsilon}{2}2^{(1-\frac{\delta}{2})H(\xi)n}>2^{(1-\delta)H(\xi)n}$, ha $n\ge n_0(\varepsilon,\delta)$. Ez ellentmond annak, hogy minden $x\in B_0\cap\bar{B}$ vektor $f(x)\in A(n)$ képe különböző, mert az $A(n)$ halmaz elemszáma $N(n)\le 2^{(1-\delta)H(\xi)n}$.

Megjegyzés. Mivel a tekintett $\xi_1,\xi_2,\dots$ sorozat független és egyforma eloszlású valószínűségi változókból áll, ezért az előző tétel eredményét alkalmazhatjuk nemcsak a $(\xi_1,\dots,\xi_n)$ hanem a $(\xi_{ln+1},\dots,\xi_{(l+1)n})$ sorozatra is minden $l=1,2,\dots$ indexre. Ezért a (2.1) reláció is teljesül egy alkalmas $f(\cdot)$, $g(\cdot)$ függvénypárral, ha az $A(n)$ halmaz $N(n)$ elemszámára $N(n)\ge 2^{(1+\delta)H(\xi)n}$, és $n\ge n_0(\varepsilon,\delta)$. Továbbá $P(g(f(\xi_{ln+1},\dots,\xi_{(l+1)n}))=(\xi_{ln+1},\dots,\xi_{(l+1)n}))\le\varepsilon$ minden $l=0,1,\dots$ indexre és f,g függvénypárra, ha $N(n)\le 2^{(1-\delta)H(\xi)n}$, és $n\ge n_0(\varepsilon,\delta)$.

Rátérek az egyértelműen dekódolható kódok vizsgálatára. Azt vizsgáljuk, hogyan lehet egy M elemből álló $X=\{x_1,\dots,x_M\}$ halmaz elemeit, pontosabban ezen elemekből álló sorozatokat hibátlanul kódolni és dekódolni viszonylag rövid kódokkal egy d elemű $Y=\{y_1,\dots,y_d\}$ ABC segítségével. Az általánosság megszorítása nélkül feltehetjük, hogy $Y=\{1,\dots,d\}$. Érdemes bevezetni a következő fogalmat.

Prefix kódok definiciója. *Legyen adva egy $X=\{x_1,\dots,x_M\}$ halmaz, és egy $Y=\{1,\dots,d\}$ ABC. Feleltessünk meg minden $x_i\in X$ elemnek egy $u(x_i)=j_1^{(i)},\dots j_{n(i)}^{(i)}$, $1\le j_s^{(i)}\le d$, $1\le s\le n(i)$, sorozatot. Azt mondjuk, hogy ez a megfeleltetés az X halmaz elemeinek prefix kódja, ha nincs olyan $x_p\in X$, $x_q\in X$ pár, amelyre az $u(x_p)$ sorozat az $u(x_q)$ sorozat megszorítása annak elejére, azaz semmilyen x_p, x_q párra nem lehet megkapni az $u(x_p)$ sorozatot úgy, hogy alkalmas számú jelet letörlünk az $u(x_q)$ sorozat végéről.*

Fontos lesz számunkra az alábbi lemmában megfogalmazott egyszerű észrevétel.

Lemma a prefix kódok egy tulajdonságáról. *Minden prefix kód egyértelműen dekódolható kód.*

Bizonyítás. Ha egy $x_1,\dots,x_n$ sorozatnak egy $u(x_1),u(x_2),\dots,u(x_n)$ prefix kód felel meg, akkor a prefix tulajdonság alapján meg tudjuk állapítani, hol fejeződik be e sorozatban az $u(x_1)$, sorozat, és így azonosítható az x_1 jel. Ezután rekurzív módon meg tudjuk határozni egymás után az x_2, x_3, ... értéketet is.

Léteznek nem prefix, de egyértelműen dekódolható kódok. Például, ha X két elemű halmaz, és $D=\{1,2\}$ választhatjuk x_1 kódjának az 1, x_2 kódjának az 1,2 sorozatot. Vagy x_1 kódjának az 1, x_2 kódjának az $1,\dots,1,2$ sorozatot, ahol az 1 jeleknek egy n hosszúságú sorozatát vettük a 2 jel előtt egy tetszőleges (ismert) pozitív egész n számmal. Ezek egyértelműen dekódolható, de nem prefix kódok. Viszont, mint látni fogjuk, minden egyértelműen dekódolható kódhoz lehet találni egy legalább ugyanolyan jó prefix kódot, ezért figyelmünket koncentrálhatjuk a prefix kódokra. Ezeknek megvan az az előnyük is, hogy gyorsan dekódolhatóak. Ha egymás után megérkeznek

az x_{j_1}, $x_{j_2}, \ldots$, jelek prefix kódjai, akkor mihelyt megérkezett egy jel kódja, azonnal dekódolhatjuk azt. Nem prefix, de egyértelműen dekódolható kódok esetében a helyzet bonyolultabb. Például, ha a két elemű X halmaz x_1 elemének a kódszava 1, az x_2 elem kódszava $1, \ldots, 1, 2$, (n darab 1 jellel), akkor egy 1 jel megérkezése után lehet, hogy még n jel megérkezését is meg kell várni, és csak azután tudjuk eldönteni, hogy ez az 1-es jel az x_1 kódszava vagy az x_2 kódszavának az első eleme volt-e. Egyébként létezik egy módszer annak eldöntésére, hogy mikor dekódolható egy kód egyértelműen, (lásd (Robert Ash: Information theory, Theorem 2.2.1), de erre a meglehetősen bonyolultan bizonyítható eredményre nem lesz szükségünk. Ezért azt nem tárgyalom.

A következő eredményben, amelyet Kraft–egyenlőtlenségnek is hívnak az irodalomban, megadjuk, hogy milyen hosszúak lehetnek egy prefix kód kódszavai.

Tétel egy prefix kód szavainak lehetséges hosszáról. (Kraft egyenlőtlenség.) *Legyen $X = \{x_1, \ldots, x_M\}$ egy M elemű halmaz, és legyen $u(x_i)$, $1 \le i \le M$, e halmaz egy prefix kódja az $Y = \{1, \ldots, d\}$ d-elemű ABC-vel. Jelölje $n(i)$ az $u(x_i)$ szó kódhosszát. Ekkor teljesül a $\sum_{i=1}^{M} d^{-n(i)} \le 1$ egyenlőtlenség. Megfordítva, ha az $n(i)$, $1 \le i \le M$, pozitív egész számok sorozata teljesíti a $\sum_{i=1}^{M} d^{-n(i)} \le 1$ egyenlőtlenséget, akkor létezik az X halmaz elemeinek olyan $u(x_i)$, $1 \le i \le M$, prefix kódja az $Y = \{1, \ldots, d\}$ d-elemű ABC-vel, amelyre az $u(x_i)$ szó kódhossza $n(i)$.*

Bizonyítás. Feltehetjük, hogy az $x_i \in X$ elemek úgy vannak indexelve, hogy $n(1) \le n(2) \le \cdots \le n(M)$. Ha adva van egy $u(x_i)$, $1 \le i \le M$, prefix kód, akkor minden egyes $u(x_i)$ kódszónak feleltessük meg az összes olyan $n(M)$ hosszúságú, az $Y = \{1, \ldots, d\}$ halmaz elemeiből álló sorozatok halmazát, amely sorozatok az $u(x_i)$ kódszó folytatásai. Az $u(x_i)$ kódszónak $d^{n(M)-n(i)}$ ilyen folytatása van, és az $u(\cdot)$ kód prefix tulajdonsága miatt íly módon csupa különböző $n(M)$ hosszúságú az $\{1, \ldots, d\}$ halmaz elemeiből álló sorozatot kapunk. Mivel összesen $d^{n(M)}$ ilyen sorozat van, ezért $\sum_{i=1}^{M} d^{n(M)-n(i)} \le d^{n(M)}$. Innen $d^{n(M)}$-mel osztva megkapjuk a $\sum_{i=1}^{M} d^{-n(i)} \le 1$ egyenlőtlenséget.

Ha adva van egy $1 \le n(1) \le n(2) \le \cdots \le n(M)$ a $\sum_{i=1}^{M} d^{-n(i)} \le 1$ egyenlőtlenséget teljesítő sorozat, akkor a következő módon tudunk a kívánt hosszúságú kódszavakból álló prefix kódot konstruálni. Legyen $u(x_1)$ tetszőleges $n(1)$ hosszúságú az $Y = \{1, \ldots, d\}$ halmaz elemeiből álló sorozat. Az i változó szerinti indukcióval definiáljuk az $u(x_i)$ kódszavakat úgy, hogy az $u(x_1), \ldots, u(x_i)$ kódszavak az $X_i = \{x_1, x_2, \ldots, x_i\}$ halmaz prefix kódját alkossák. Ha az $i-1$ indexre találtunk ilyen kódszavakat, akkor a prefix tulajdonság megőrzéséhez az indukció i-ik lépésében elég olyan $n(i)$ hosszúságú $u(x_i)$ kódszót találni, amely nem folytatása egyik $u(x_j)$, $1 \le j \le i-1$, kódszónak sem. Ez azt jelenti, hogy az $d^{n(i)}$ darab $n(i)$ hosszúságú, az $Y = \{1, \ldots, d\}$ halmaz elemeiből álló sorozat közül $\sum_{j=1}^{i-1} d^{n(i)-n(j)}$ sorozat választása van tiltva. Akkor tudunk egy kívánt

tulajdonságú $u(x_i)$ sorozatot (kódszót) választani, ha $\sum_{j=1}^{i-1} d^{n(i)-n(j)} < d^{n(i)}$, azaz, ha $\sum_{j=1}^{i-1} d^{-n(j)} < 1$. Az adott feltétel mellett ez a tulajdonság minden $2 \le i \le M$ indexre teljesül. Ezzel beláttuk a tétel második állítását is.

A következő eredmény azt állítja, hogy adott szóhosszúságú kódszavakkal rendelkező prefix kódok létezésének ugyanaz a feltétele, mint annak, hogy létezzen ilyen hosszúságú kódszavakkal rendelkező egyértelműen dekódolható kód.

Tétel egyértelműen dekódolható kódok létezésének szükséges feltételéről. *Legyen $X = \{x_1, \dots, x_M\}$ egy M elemű halmaz, és legyen $u(x_i)$ e halmaz egy egyértelműen dekódolható kódja az $Y = \{1, \dots, d\}$ d-elemű ABC-vel. Ekkor az $u(x_i)$ kódszavak $n(i)$ kódhosszai teljesítik a $\sum_{i=1}^{M} d^{-n(i)} \le 1$ egyenlőtlenséget.*

Bizonyítás. Jelölje ω_j azon $x_i \in X$, $1 \le i \le M$, pontok számát, amelyek $u(x_i)$ kódszavának a hossza $n(i) = j$, és legyen $r = \sup_{1 \le i \le M} n(i)$. Ezzel a jelöléssel $\sum_{i=1}^{M} d^{-n(i)} = \sum_{j=1}^{r} \omega_j d^{-j}$, és a bizonyítandó egyenlőtlenség $\sum_{j=1}^{r} \omega_j d^{-j} \le 1$ alakban is írható. Ezen egyenlőtlenség igazolása érdekében vegyünk egy pozitív egész p számot, és írjuk fel a

$$\left(\sum_{j=1}^{r} \omega_j d^{-j} \right)^p = (\omega_1 d^{-1} + \dots + \omega_r d^{-r})^p = \sum_{k=p}^{pr} N_k d^{-k}$$

azonosságot, ahol

$$N_k = \sum_{(i_1,\dots,i_p):\ i_1+\dots+i_p=k} \omega_{i_1} \dots \omega_{i_p}, \quad p \le k \le pr.$$

Azt állítom, hogy teljesül az $N_k \le d^k$ egyenlőtlenség minden $p \le k \le pr$ indexre. Valóban, az $\omega_{i_1} \dots \omega_{i_p}$ szorzat azon $u(x_{l_1}), \dots, u(x_{l_p})$ kódszósorozatok számával egyenlő, amelyekre $n(l_1) = i_1, \dots, n(l_p) = i_p$. Ezért N_k egyenlő azon $u(x_{l_1}), \dots, u(x_{l_p})$ kódszósorozatok számával, amelyek összhossza k-val egyenlő. Az egyértelmű dekódolhatóság miatt az összes előbb felsorolt kódszósorozat különböző, ezért számuk kisebb, mint az összes lehetséges k hosszúságú az $Y = \{1, \dots, d\}$ halmaz elemeit tartalmazó sorozat száma. Ezért $N_k \le d^k$, amint állítottuk.

A fenti összefüggésekből következik, hogy

$$\left(\sum_{j=1}^{r} \omega_j d^{-j} \right)^p \le \sum_{k=p}^{pr} d^k \cdot d^{-k} = (p(r-1)+1) \le pr,$$

ezért

$$\sum_{j=1}^{r} \omega_j d^{-j} \le (pr)^{1/p} \quad \text{minden } p = 1, 2, \dots \text{ számra.}$$

Innen $p \to \infty$ határátmenetet véve azt kapjuk, hogy

$$\sum_{j=1}^{r} \omega_j d^{-j} \le \lim_{p\to\infty} (pr)^{1/p} = 1,$$

ahonnan következik a tétel állítása.

A bizonyítás gondolatának jobb megértése érdekében érdemes megjegyezni, hogy a $\left(\sum_{j=1}^{r} \omega_j d^{-j}\right)^p$ kifejezés becslésében felhasznált $N_k \le d^k$ egyenlőtlenség indoklása azon alapult, hogy a $k \le pr$ hosszúságú kódszavak egyértelműen dekódolhatóak. A tekintett kód egyértelmű dekódolhatóságát, azt, hogy a hosszú $x_{l_1}, \dots, x_{l_p}$ sorozatok kódjai is egyértelműen dekódolhatók a $p \to \infty$ határátmenet alkalmazásakor használtuk ki.

Legyen adva egy ξ valószínűségi változó, amely értékeit egy $X = \{x_1, \dots, x_M\}$ halmazon veszi fel, és $P(\xi = x_i) = p(i)$, $1 \le i \le M$. Legyen $u(x_i)$, $x_i \in X$, az X halmaz elemeinek egy olyan kódja, amelyre $u(x_i)$ egy $n(i)$ hosszúságú az $Y = \{1, \dots, d\}$ halmaz elemeiből álló sorozat. Jelölje $|u(\xi)|$ a ξ valószínűségi változó $u(\xi)$ kódjának a hosszát, azaz, ha $\xi = x_i$, akkor bevezetve az $|u(x_i)| = n(i)$ jelölést azt írhatjuk, hogy $u(\xi) = u(x_i)$, és $|u(\xi)| = n(i)$. Az $u(\xi)$ véletlen kódszó $|u(\xi)|$ hosszának a várható értéke $E|u(\xi)| = \sum_{i=1}^{M} p(i)n(i)$. Az $u(\xi)$ véletlen kódszó hosszának a várható értékére kívánunk jó alsó becslést adni, ha $u(x_i)$, $x_i \in X$ prefix, illetve általánosabban, ha az egyértelműen dekódolható kód. Ezenkívül jó felső becslést is akarunk adni egy jól választott prefix kód hosszának a várható értékére.

Tudjuk, hogy akkor és csak akkor létezik $n(i)$, $1 \le i \le M$, hosszú, az $Y = \{1, \dots, d\}$ ABC-t használó kódszavakkal rendelkező prefix, illetve általánosabban egyértelműen dekódolható kód, ha teljesül a $\sum_{i=1}^{M} d^{-n(i)} \le 1$ egyenlőtlenség. Ezért problémánk ahhoz az egész-értékű szélsőérték feladathoz vezet, hogy keressük meg a $\sum_{i=1}^{M} p(i)n(i)$ kifejezés (majdnem) optimumát a $\sum_{i=1}^{M} d^{-n(i)} \le 1$ feltétel mellett. Ezen optimalizációs probléma vizsgálatában hasznos az alábbi becslés, amelyet I-divergencia típusú becslésnek fogok hívni. Ugyanis, mint egy megjegyzésben elmagyarázom, ez a becslés tekinthető, úgy mint az információelmélet egyik fontos, a később bevezetendő I-divergenciáról szóló becslésének a speciális esete.

Egy I-divergencia típusú becslést megfogalmazó lemma. *Legyen $a_1, a_2, \dots$ és $b_1, b_2, \dots$ két véges vagy végtelen, ugyanannyi elemet tartalmazó sorozat, amelyekre $a_i \ge 0$, $b_i \ge 0$ minden i indexre, és $a = \sum_i a_i < \infty$, $0 < b = \sum_i b_i < \infty$. Ekkor*

$$\sum_i a_i \log \frac{a_i}{b_i} \ge a \log \frac{a}{b}.$$

Egyenlőség akkor és csak akkor érvényes, ha elhagyva azon (a_i, b_i) párokat, amelyekre $a_i = b_i = 0$ $\frac{a_i}{b_i} = \frac{a}{b}$ minden i indexre. (Ezen egyenőtlenségben az $x \cdot \log \frac{0}{x} = 0$, ha $x \geq 0$, és $x \cdot \log \frac{x}{0} = \infty$, ha $x > 0$ konvenciót alkalmazzuk.)

Bizonyítás. Fel fogjuk használni, hogy az első fejezetben bevezetett és vizsgált $g(x) = x \log x$, $x \geq 0$, függvény konvex. A $g(x)$ függvény a konvexitását az $u_i = \frac{a_i}{b_i}$ koordináták, és $p_i = \frac{b_i}{\sum\limits_i b_i} = \frac{b_i}{b}$ súlyok választásával fogom alkalmazni. (Nyilván $p_i \geq 0$, és $\sum\limits_i p(i) =$ 1.) A $g(x)$ függvény konvexitását felhasználva azt kapjuk, hogy

$$\sum_i a_i \log \frac{a_i}{b_i} = b \sum_i p_i g(u_i) \geq bg\left(\sum_i p_i u_i\right) = bg\left(\frac{a}{b}\right) = a \log \frac{a}{b}.$$

A $g(\cdot)$ szigorú konvexitásából következik, hogy egyenlőség csak a lemmában megadott esetben van.

Megjegyzés. A fenti lemmát könnyen redukálni lehet arra a speciális esetre, amikor $\sum_i a_i = \sum_i b_i = 1$. (Azt használhatjuk ki, hogy a bizonyítandó egyenlőtlenség két oldalának különbsége az α illetve β paraméter homogén függvénye, ha az a_i számokat αa_i vagy a b_i számokat βb_i számokkal helyettesítjük valamely $\alpha > 0$ és $\beta > 0$ számmal.) Ebben az esetben az egyenlőtlenség jobboldalán 0 áll. Ez a redukált egyenlőtlenség felfogható az alábbi egyenlőtlenség speciális esetének. Legyen μ és ν két valószínűségi mérték ugyanazon az $(X, \mathcal{X})$ mérhető téren, és legyen a ν mérték abszolút folytonos a μ mértékre nézve. Jelölje $\frac{d\nu}{d\mu}$ a ν mértéknek a μ mérték szerinti Radon–Nikodym deriváltját. Ekkor $\int_X \log \frac{d\nu}{d\mu}(x)\, d\nu(x) \geq 0$. Ez az egyenlőtlenség tekinthető úgy, mint a később bevezetendő I-divergencia egy fontos tulajdonsága. Egyébként ez az egyenlőtlenség a lemmához hasonlóan bizonyítható.

A most ismertetett lemma segít az alábbi eredmény bizonyításában.

Tétel egyértelműen dekódolható és prefix kódok hosszának várható értékéről. *Vegye fel egy ξ valószínűségi változó értékeit egy $X = \{x_1, \ldots, x_M\}$ halmazon, amelynek az eloszlását a $P(\xi = x_i) = p(i)$, $1 \leq i \leq M$, képlet adja meg. Legyen $u(x_i)$, $x_i \in X$, az X halmaznak egy az $Y = \{1, \ldots, d\}$ halmaz segítségével definiált prefix, vagy általánosabban, egyértelműen dekódolható kódja, és jelölje $n(i)$ az $u(x_i)$ kódszó szóhosszát. Ekkor az $u(\xi)$ véletlen kódszó $|u(\xi)|$ hosszának a várható értéke teljesíti az*

$$E|u(\xi)| = \sum_{i=1}^{M} p(i)n(i) \geq \frac{H(\xi)}{\log d}$$

egyenlőtlenséget. Megfordítva, létezik az X halmaznak olyan prefix kódja, amelyre

$$E|u(\xi)| = \sum_{i=1}^{M} p(i)n(i) \leq \frac{H(\xi)}{\log d} + 1.$$

Bizonyítás. Legyen $u(x_i)$, $1 \le i \le M$, az X halmaznak az $Y = \{1, \dots, d\}$ ABC segítségével definiált egyértelműen dekódolható kódja, és legyen $n(i)$ az $u(x_i)$ kódszó kódhossza. Tudjuk, hogy ekkor $\sum_{i=1}^{M} d^{-n(i)} \le 1$. Alkalmazzuk az előző lemmában bizonyított egyenlőtlenséget $a_i = p(i)$ és $b_i = d^{-n(i)}$, $1 \le i \le M$, választással. Ekkor $a = \sum_{i=1}^{M} a_i = 1$, és $b = \sum_{i=1}^{M} b_i \le 1$, ahonnan

$$\sum_{i=1}^{M} p(i) \log(p(i) d^{n(i)}) = \sum_{i=1}^{M} a_i \log \frac{a_i}{b_i} \ge a \log \frac{a}{b} \ge 0.$$

Tehát $\log d \cdot \sum_{i=1}^{M} n(i)p(i) \ge -\sum_{i=1}^{M} p(i) \log p(i)$, azaz $\log d \cdot E|u(\xi)| \ge H(\xi)$, és ezt kellett belátni.

Láttuk, hogy a tétel második felének igazolása érdekében olyan $n(i)$, $1 \le i \le M$, pozitív egész számokat kell találnunk, amelyekre $\sum_{i=1}^{M} d^{-n(i)} \le 1$, és az $E|u(\xi)| = \sum_{i=1}^{M} p(i)n(i)$ várható érték viszonylag kicsi. Természetes olyan $n(i)$ számokat választani, amelyekre az $E|u(\xi)|$ várható értékre adott alsó becslés bizonyításában felhasznált I-divergencia típusú becslést megfogalmazó lemmában szereplő egyenlőtlenség majdnem egyenlőséggel teljesül. Ezért válasszunk olyan $n(i)$ egész számokat, amelyekre $d^{-n(i)} \le p(i)$, és a $d^{-n(i)}$ szám olyan közel van a $p(i)$ számhoz, amennyire ez a feltétel megengedi. Ennek alapján a következő választást tesszük. Legyen minden $1 \le i \le M$ indexre $n(i)$ az az egész szám, amelyre $\frac{p(i)}{d} \le d^{-n(i)} \le p(i)$. Ekkor $\sum_{i=1}^{M} d^{-n(i)} \le \sum_{i=1}^{M} p(i) = 1$, azaz létezik a kívánt hosszúságú kódszavakkal rendelkező prefix kód. Másrészt $n(i) \le -\frac{\log p(i)}{\log d} + 1$, és ezért $E|u(\xi)| = \sum_{i=1}^{M} p(i)n(i) \le -\sum_{i=1}^{M} \frac{p(i) \log p(i)}{\log d} + \sum_{i=1}^{M} p(i) = \frac{H(\xi)}{\log d} + 1$, tehát e prefix kód hosszának a várható értéke teljesíti a kívánt egyenlőtlenséget.

Az egyértelműen dekódolható és prefix kódok hosszának várható értékéről szóló tétel felső becslésének bizonyításában definiált prefix kód hossza közel van az optimumhoz, de nem feltétlenül egyenlő vele. Ugyanakkor ismert az optimális, úgynevezett Huffman kód konstrukciója is. (Lásd Robert Ash: Information Theory, Lemma 2.6.2), amely eléggé bonyolult. Ezért e kód tulajdonságai nehezen vizsgálhatóak, és jelentősége korlátozott. Emiatt mi a Huffman kódot nem tárgyaljuk.

Hasonló jelenséggel találkoztunk, amikor független valószínűségi változók n hosszúságú sorozataiból álló viszonylag kis elemszámú majdnem 1 valószínűségű, alkalmasan definiált halmazok elemszámát becsültük. Ott sem az optimális halmaz elemszámát

becsültük. Ez a legvalószínűbb sorozatok alkalmas elemszámú halmaza lett volna. Ehelyett egy olyan csak aszimptotikusan optimális halmazt tekintettünk, amelynek elemszámát a nagy számok törvénye segítségével jól tudtuk becsülni.

Megjegyzés. A most bizonyított eredmények viszonylag rövid, hibátlanul dekódolható forráskódolást biztosítanak prefix kódok segítségével. Gyakorlati szempontból azonban a most ismertetett prefix kódok eredeti formájukban nem jól használhatóak. A fő probléma az, hogy ha egy prefix kód dekódolása során egyszer hibáztunk, akkor ennek a hibának a következtében az összes további üzenet dekódolása hibás lehet. Ugyanis nem tudjuk, hogy az üzenet további dekódolandó jelei hol kezdődtek. Az ilyen problémák leküzdésére érdemes olyan kissé lassúbb módszereket kidolgozni, amelyekbe bizonyos javítási lehetőségeket építenek be. Az ilyen, úgynevezett 'error correcting codes' módszereknek külön elmélete van. Ezzel azonban itt nem foglalkozunk.

3. Csatorna kódolás és dekódolás.

E fejezet fő témája az a kérdés, hogy hogyan lehet egy a jeleket esetleg hibával közvetítő csatornán keresztül viszonylag biztosan és gyorsan üzeneteket átadni. A következő fejezetben tárgyalom azt a kérdést, hogy az ebben és az előző fejezetekben bizonyított eredmények segítségével hogyan lehet egy hírforrás közvetítését egy csatornán keresztül jól továbbítani.

Egy üzenetnek egy (hírközlési) csatorna segítségével végrehajtott továbbadása azt jelenti szemléletesen, hogy a csatorna bemeneti végén leadnak egy jelet, és ennek hatására valamilyen másik jel jelenik meg a csatorna másik, kimeneti végén. A kimeneti jel értéke függhet a véletlentől, és az, hogy a kimeneti oldalon milyen valószínűséggel milyen jel jelenik meg, attól függ, hogy mi volt a bemeneti jel. Egy felhasználó, aki a kimeneti jelet megismeri, megpróbál ennek alapján visszakövetkeztetni a leadott bemeneti jelre. Olyan eljárást akarunk kidolgozni a csatorna tulajdonságainak az ismeretében, amely lehetővé teszi, hogy a felhasználó viszonylag nagy valószínűséggel helyesen következtessen a leadott bemeneti jelre akkor is, ha a bemeneti jelek száma viszonylag nagy. Fogalmazzuk meg ezt a vázlatosan leírt problémát pontosabban. Ennek érdekében bevezetem először a (hírközlési) csatorna fogalmát.

(Hírközlési) csatorna és e csatorna által összekapcsolt valószínűségi változók definiciója. *Legyen adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálhatóan végtelen halmaz. Egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűség függvényt a V és $\tilde{V}$ halmaz közötti csatornának nevezünk. Az, hogy a $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$ függvény átmenetvalószínűség függvény azt jelenti, hogy $p(\tilde{v}_j|v_i) \geq 0$ minden $v_i \in V$ és $\tilde{v}_j \in \tilde{V}$ párra, és $\sum_{\tilde{v}_j \in \tilde{V}} p(\tilde{v}_j|v_i) = 1$ minden $v_i \in V$ elemre. A V halmazt a csatorna bemeneti, a $\tilde{V}$ halmazt a csatorna kimeneti oldalának fogjuk hívni, a $v_i \in V$ pontokat bemeneti, a $\tilde{v}_j \in \tilde{V}$ pontokat pedig kimeneti jeleknek fogjuk nevezni. Azt mondjuk hogy két η és $\tilde{\eta}$ valószínűségi változó össze van kapcsolva a $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, csatornával ha $\eta(\omega) \in V$, $\tilde{\eta}(\omega) \in \tilde{V}$ minden ω elemi eseményre, és $P(\tilde{\eta} = \tilde{v}_j|\eta = v_i) = p(\tilde{v}_j|v_i)$ minden $v_i \in V$ és $\tilde{v}_j \in \tilde{V}$ elempárra.*

Ha egy csatornán leadunk valamely $v_i \in V$ bemeneti jelet, akkor a felhasználó $p(\tilde{v}_j|v_i)$ valószínűséggel kapja a $\tilde{v}_j$ kimeneti jelet. Ezen kimeneti jel segítségével próbálja megtalálni a bemeneti jel értékét. Ennek érdekében természetes a következő típusú eljárás alkalmazása. A bemeneti oldalon alkalmas módon kiválasztunk néhány $v_{i_1} \in V$, $\dots$, $v_{i_N} \in V$ elemet bizonyos N elemszámmal, és ezen jelek valamelyikét adjuk le a csatornán. Ezeknek a bemeneti jeleknek a kiválasztását nevezzük csatorna kódolásnak. Úgy kívánjuk ezt a kiválasztást végrehajtani, hogy az egyes kiválasztott elemeket a csatornán leadva, azok nagy valószínűséggel különböző halmazokba essenek. Ez a következőt jelenti. Ha egy $v_i \in V$ jelet leadunk a csatornán, akkor jelölje az $\tilde{\eta}(v_i)$ valószínűségi változó a kimenő jel értékét. Olyan $v_{i_1} \in V$, $\dots$, $v_{i_N} \in V$ elemeket akarunk a kódolásban választani, amelyekhez találhatóak olyan diszjunt $B_1 \subset \tilde{V}$, $\dots$, $B_N \subset \tilde{V}$ diszjunkt halmazok, amelyekre a $P(\tilde{\eta}(v_{i_k}) \in B_k)$ valószínűségek minden $1 \le k \le N$ indexre viszonylag nagyok. Ha a kimeneti oldalon egy olyan $\tilde{v}_j$ jel jelent meg, amelyre $\tilde{v}_j \in B_k$, akkor tekintse a felhasználó a v_{i_k} jelet a bemeneti jelnek. A B_k, $1 \le k \le N$, halmazok kiválasztását és a $B_k \to v_{i_k}$ leképezés megadását csatorna dekódodolásnak nevezzük. E definiciót úgy adtam meg, hogy amennyiben $\tilde{v}_j \notin \bigcup_{i=k}^{N} B_k$ akkor az itt ismertetett eljárás szerint nem dekódoljuk a $\tilde{v}_j$ kimenetet. Az azonban, hogy ezt az elvet követjük-e, vagy olyan eljárást adunk, amelyikben mindig tudunk dekódolni csak apró ízlésbeli kérdés. A B_k halmazok kiterjesztésével ugyanis azt is elérhetjük, hogy ezek a halmazok ne csak diszjunktak legyenek, hanem egyben a $\tilde{V}$ halmaz egy particióját is szolgáltassák. Ilyen választással minden kimenetet tudunk dekódolni, és az eljárásban a jó dekódolás valószínűsége a B_k halmazok kiterjesztése által nem csökkent. Célunk viszonylag sok kódszó kiválasztása úgy, hogy a dekódolás nagy valószínűséggel jó legyen. A későbbi eredmények pontos megfogalmazásának az érdekében vezessük be a következő definiciót.

Egy csatorna által λ megkülönböztethető elemekből álló halmaz definiciója. *Legyen adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálhatóan végtelen halmaz és köztük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$ csatorna. Ha egy $v_{i_1} \in V$, $\dots$, $v_{i_N} \in V$ sorozat elemeihez léteznek olyan $B_1 \subset \tilde{V}$, $\dots$ $B_N \subset \tilde{V}$* diszjunkt *halmazok, amelyekre*

$$P(B_k|v_{i_k}) = \sum_{j \in B_k} p(\tilde{v}_j|v_{i_k}) \ge 1 - \lambda \quad \textit{minden } 1 \le k \le N \textit{ indexre},$$

akkor azt mondjuk, hogy az $A = \{v_{i_1}, \dots, v_{i_N}\}$ halmaz elemei λ megkülönböztethetőek.

Megjegyzés. Az előbbi definicióban felírt egyenlőtlenséget úgy is megfogalmazhatjuk, hogy amennyiben η és $\tilde{\eta}$ két a csatornával összekapcsolt valószínűségi változó, akkor

$$P(\tilde{\eta} \in B_k|\eta = v_{i_k}) \ge 1 - \lambda \quad \text{minden } 1 \le k \le N \text{ indexre.}$$

A későbbiekben többször ezt a jellemzését fogjuk használni az egy csatorna által λ megkülönböztethető elemekből álló halmazoknak.

Egy az előbb leírt módon definiált jó, azaz kis hibájú csatorna kódolás, dekódolás definiciója egy λ megkülönböztethető elemekből álló $v_{i_1} \in V, \dots, v_{i_N} \in V$ sorozat megadását jelenti kis $\lambda > 0$ paraméterrel az e sorozatokhoz tartozó $B_1 \subset \tilde{V}, \dots, B_N \subset \tilde{V}$ halmazokkal együtt. A minket érdeklő feladatokban valójában nem egy jelet, hanem egy jelsorozat elemeit adjuk le egymás után a csatornán, és célunk ennek a jelsorozatnak a minél pontosabb azonosítása. Minket az az eset érdekel elsősorban, amikor a jelsorozat egyes jelei egymástól függetlenül, és ugyanolyan törvényszerűségek szerint mennek át a csatornán. Ebben a jegyzetben csak ezt az esetet fogom tárgyalni. A probléma pontos megfogalmazása érdekében bevezetem az emlékezet nélküli csatorna fogalmát.

Emlékezet nélküli csatorna definiciója. *Legyen adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálható halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált csatorna. Az e csatorna által definiált n hosszúságú emlékezet nélküli csatorna olyan csatorna, amelynek bemeneti jelei a V^n, kimeneti jelei a $\tilde{V}^n$ halmaz elemei, azaz az n hosszúságú v_{i_k} illetve $\tilde{v}_{j_k}$, $1 \le k \le n$, elemekből álló sorozatok, átmenetvalószínűsége pedig $p((\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n})|(v_{i_1}, \dots, v_{i_n})) = \prod_{k=1}^{n} p(\tilde{v}_{j_k}|v_{i_k})$ tetszőleges $v = (v_{i_1}, \dots, v_{i_n}) \in V^n$, és $\tilde{v} = (\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n}) \in \tilde{V}^n$ sorozatokra.*

Az emlékezet nélküli csatorna természetes megfelelője a független, egyforma eloszlású valószínűségi változók sorozatainak. Azt a kérdést fogjuk vizsgálni, hogy egy n hosszúságú emlékezet nélküli csatorna bemeneti oldalán aszimptotikusan hány a csatorna által λ megkülönböztethető elemet tartalmazó halmazt lehet kiválasztani nagy n és rögzített $0 < \lambda < 1$ szám esetén. Be fogjuk látni, hogy nagyon általános feltételek mellett ez a szám $2^{Cn(1+o(1))}$, ahol a C szám, amelyet csatorna kapacitásnak fogunk hívni, a csatorna tulajdonságaitól függ. Ahhoz, hogy az eredményt pontosan megfogalmazzam definiálni kell a csatorna kapacitást. Ennek érdekében bevezetem először két valószínűségi változó kölcsönös információjának a fogalmát.

A kölcsönös információ fogalma. *Legyen η és $\tilde{\eta}$ két értékeiket egy $V = \{v_1, v_2, \dots\}$ illetve $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálható halmazon felvevő valószínűségi változó, és jelölje $r(v_i, \tilde{v}_j) = P(\eta = v_i, \tilde{\eta} = v_j)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, az együttes eloszlásukat. Vezessük be a $p(v_i) = P(\eta = v_i) = \sum_{\tilde{v}_j \tilde{V}} r(v_i, \tilde{v}_j)$ és $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j) = \sum_{v_i \in V} r(v_i, v_j)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, mennyiségeket is. Ezzel a jelöléssel az η és $\tilde{\eta}$ valószínűségi változók kölcsönös információja az*

$$I(\eta \wedge \tilde{\eta}) = \sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} r(v_i, \tilde{v}_j) \log \left(\frac{r(v_i, \tilde{v}_j)}{p(v_i) q(\tilde{v}_j)} \right)$$

kifejezéssel egyenlő. Az $I(\eta \wedge \tilde{\eta})$ kölcsönös információt kifejező összegben csak olyan $(v_i, \tilde{v}_j)$ párokra összegezünk, amelyekre $r(v_i, \tilde{v}_j) > 0$.

Megjegyzés. Az $I(\eta \wedge \tilde{\eta})$ kölcsönös információt kifejező összeg mindig értelmes, mert az összeg negatív értékű tagjainak az összege mindig konvergens. Ugyanis

$$I(\eta \wedge \tilde{\eta}) = \sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} p(v_i) q(\tilde{v}_j) g\left(\frac{r(v_i, \tilde{v}_j)}{p(v_i) q(\tilde{v}_j)}\right)$$

ahol $g(x) = x \log x$, és egyrészt ez a $g(x)$ függvény alulról korlátos a $[0, \infty)$ intervallumban, másrészt $\sum p(v_i) q(\tilde{v}_j) = 1$.

Tétel a kölcsönös információ viselkedéséről. *Érvényes az $I(\eta \wedge \tilde{\eta}) \geq 0$ egyenlőtlenség, és egyenlőség akkor és csak akkor áll fenn ebben a formulában, ha η és $\tilde{\eta}$ függetlenek.*

Bizonyítás. Alkalmazzuk az $r(v_i, \tilde{v}_j) = P(\eta = v_i, \tilde{\eta} = \tilde{v}_j)$, és $p(v_i) = \sum_{\tilde{v}_j \in \tilde{V}} r(v_i, \tilde{v}_j)$, $q(\tilde{v}_j) = \sum_{v_i \in V} r(v_i, \tilde{v}_j)$ $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, jelöléseket. Azt kapjuk, felhasználva a $g(x) = x \log x$ függvény szigorú konvexitását, hogy

$$\begin{aligned} I(\eta \wedge \tilde{\eta}) &= \sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} r(v_i, \tilde{v}_j) \log\left(\frac{r(v_i, \tilde{v}_j)}{p(v_i) q(\tilde{v}_j)}\right) = \sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} p(v_i) q(\tilde{v}_j) g\left(\frac{r(v_i, \tilde{v}_j)}{p(v_i) q(\tilde{v}_j)}\right) \\ &\geq g\left(\sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} r(v_i, \tilde{v}_j)\right) = g(1) = 0, \end{aligned}$$

Ebben a számolásban a $g(x)$ függvény konvexitását alkalmaztuk a $p(v_i) q(\tilde{v}_j)$ súlyfüggvénnyel. Vegyük észre, hogy $\sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} p(v_i) q(\tilde{v}_j) = 1$, és $p(v_i) q(\tilde{v}_j) \geq 0$ minden i és j indexre. Továbbá felhasználtuk azt is, hogy mivel a $g(x)$, $x \geq 0$, függvény alulról korlátos, ezért jogunk van a bizonyításban használt konvexitási tulajdonságot akkor is használni, ha végtelen sok osztópontot tekintünk. Egyenlőség a $g(x)$ függvény szigorú konvexitása miatt csak akkor teljesül, ha $\frac{r(v_i, \tilde{v}_j)}{p(v_i) q(\tilde{v}_j)} = \alpha$ egy az i és j indextől nem függő α számmal minden i és j indexre. (Ha $p(v_i) q(\tilde{v}_j) = 0$, akkor $r(v_i, \tilde{v}_j) = 0$, és ekkor ezt a törtet tetszőleges módon definiálhatjuk.) De mivel $\sum_{v_i, \tilde{v}_j} p(v_i) q(\tilde{v}_j) = \sum_{v_i, \tilde{v}_j} r(v_i, \tilde{v}_j) = 1$, ez csak akkor lehetséges, hogy $\alpha = 1$, azaz az η és $\tilde{\eta}$ valószínűségi változók függetlenek.

1. megjegyzés. Ha az $\tilde{\eta}$ valószínűségi változó entrópiája teljesíti a $H(\tilde{\eta}) < \infty$ feltételt (vagy $H(\eta) < \infty$), akkor az η és $\tilde{\eta}$ valószínűségi változók kölcsönös információja egyszerűbben is kifejezhető, és az előző tétel állítása következik az entrópia már bizonyított tulajdonságaiból is. Ekkor

$$\begin{aligned} I(\eta \wedge \tilde{\eta}) &= \sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} r(v_i, \tilde{v}_j) \log\left(\frac{r(v_i, \tilde{v}_j)}{p(v_i) q(\tilde{v}_j)}\right) \\ &= \sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} r(v_i, \tilde{v}_j) \log \frac{r(v_i, \tilde{v}_j)}{p(v_i)} - \sum_{v_i \in V, \tilde{v}_j \in \tilde{V}} r(v_i, \tilde{v}_j) \log q(\tilde{v}_j) \\ &= H(\tilde{\eta}) - H(\tilde{\eta} | \eta), \end{aligned}$$

ahonnan $(I\eta \wedge \tilde{\eta}) = H(\tilde{\eta}) - H(\tilde{\eta}|\eta) \geq 0$, és egyenlőség csak akkor áll fenn, ha η és $\tilde{\eta}$ független valószínűségi változók. Ha a $H(\eta) < \infty$ feltétel is teljesül, akkor $I(\eta \wedge \tilde{\eta}) = H(\eta) + H(\tilde{\eta}) - H(\eta, \tilde{\eta}) = H(\eta) - H(\eta|\tilde{\eta}) = H(\tilde{\eta}) - H(\tilde{\eta}|\eta)$.

2. megjegyzés. Érdemes felidézni az első fejezetben tárgyalt eredményeket és tárgyalni azok kapcsolatát a kölcsönös információ fogalmával. Először azt a példát tárgyaltuk, hogy ha egy mérkőzéssorozat $\xi_1, \dots, \xi_n$ eredményei egymástól független és egy ξ valószínűségi változóval azonos eloszlású valószínűségi változók, akkor körülbelül $2^{nH(\xi)}$ szelvényt kell kitölteni annak érdekében, hogy majdnem biztosan legyen telitalálatunk. Ha ismerjük egy másik mérkőzéssorozat $\eta_1, \dots, \eta_n$ eredményeit, és a $(\xi_1, \eta_1), \dots, (\xi_n, \eta_n)$ eredménypárok egymástól függetlenek, és egy (ξ, η) véletlen vektorral azonos eloszlású valószínűségi változók, akkor az $\eta_1, \dots, \eta_n$ mérkőzéssorozat eredményeinek ismeretében körülbelül $2^{nH(\xi|\eta)}$ szelvényt kell kitöltenünk a majdnem biztos telitalálat eléréséhez, tehát körülbelül $2^{-n(H(\xi)-H(\xi|\eta))} = 2^{-nI(\xi\wedge\eta)}$-szorosát annak amennyi szelvényt akkor kell kitöltenünk, ha az $\eta_1, \dots, \eta_n$ mérkőzéssorozat eredményeit nem ismerjük.

Ezt az eredményt heurisztikusan úgy is interpretálhatjuk, hogy az η_j valószínűségi változó ismerete $I(\xi \wedge \eta)$-vel csökkenti a ξ_j valószínűségi változó megismeréséhez szükséges információt. Az $I(\xi \wedge \eta) = H(\xi) - H(\xi|\eta) = H(\eta) - H(\eta|\xi)$ azonosság azt jelenti, hogy a kölcsönös információnak ebben az interpretációjában a ξ és η valószínűségi változók szerepe felcserélhető.

Bevezetem a csatorna kapacitás fogalmát.

A csatorna kapacitás fogalma. *Legyen adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálható halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált csatorna. A csatorna kapacitását a*

$$C = \sup_{\eta, \tilde{\eta}} I(\eta \wedge \tilde{\eta})$$

képlet adja meg, ahol a szuprémumban az összes a csatorna által összekapcsolt $(\eta, \tilde{\eta})$ valószínűségi változó párt tekintjük.

Megjegyzés. Láttuk, hogy egy csatorna kapacitása mindig nagyobb vagy egyenlő, mint nulla. Sőt, jellemezni lehet azt az esetet is, amikor a csatorna kapacitás nullával egyenlő. Ez akkor következik be, ha bármely két a csatornával összekapcsolt η és $\tilde{\eta}$ valószínűségi változó független. Nem nehéz belátni, hogy ez akkor és csak akkor lehetséges, ha a csatornát meghatározó $p(\tilde{v}_j|v_i)$ átmenetvalószínűségek nem függnek a v_i bemeneti jel értékétől. Ez azt jelenti, hogy bármely bemeneti jel leadása esetén ugyanolyan valószínűséggel kapjuk bármely kimeneti jelet. Ekkor a kimeneti jel ismerete semmilyen információt nem nyújt arról, hogy milyen bemeneti jelet adtak le.

E fejezet fő eredményei arra adnak becslést, hogy, milyen nagy, azaz hány λ megkülönböztethető elemet tartalmazó halmazt lehet konstruálni egy n hosszúságú emlékezet nélküli csatorna bemeneti oldalán. A következő eredményeket fogom belátni.

Csatorna kódolási tétel. *Legyen adva adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálható halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált $C < \infty$ csatorna kapacitású csatorna. Tekintsünk egy ezen csatorna által definiált n hosszúságú emlékezet nélküli csatornát. Minden $0 < \lambda < 1$ és $\varepsilon > 0$ számhoz létezik olyan $n_0 = n_0(\varepsilon, \lambda)$ küszöbindex úgy, hogy amennyiben $n \geq n_0$, akkor létezik $N \geq 2^{(1-\varepsilon)Cn}$ n hosszúságú $(v_{j_1}^{(k)}, \dots, v_{j_n}^{(k)}) \in V^n$, $1 \leq k \leq N$, alakú λ megkülönböztethető sorozatot tartalmazó halmaz ennek az n hosszúságú emlékezet nélküli csatornának a bemeneti oldalán.*

A fenti eredmény megfordítását, azaz egy λ megkülönböztethető halmaz elemszámára adott felső becslést csak véges állapotterű csatornákra fogom bizonyítani. Egy csatornát véges állapotterűnek nevezek, ha mind a bemeneti jelek V mind a kimeneti jelek $\tilde{V}$ halmaza véges elemszámú.

A csatorna kódolási tétel megfordítása. *Legyen adva két $V = \{v_1, \dots, v_m\}$ és $\tilde{V} = \{\tilde{v}_1, \dots, \tilde{v}_n\}$ véges halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált véges állapotterű $C < \infty$ csatorna kapacitású csatorna. Tekintsünk egy ezen csatorna által definiált n hosszúságú emlékezet nélküli csatornát. Legyen A a V^n halmaz egy λ megkülönböztethető sorozatokat tartalmazó részhalmaza valamely $0 < \lambda < 1$ számmal. Ekkor létezik olyan $n_0 = n_0(\varepsilon, \lambda)$ küszöbindex, hogy amennyiben $n \geq n_0$, akkor az A halmaz elemszáma kisebb, mint $2^{(1+\varepsilon)Cn}$. Sőt, igaz a következő élesebb becslés. Minden $n \geq 1$ számra az adott tulajdonságú A halmaz elemszáma kisebb, mint $\frac{2}{1-\lambda} 2^{Cn + K\sqrt{n}/\sqrt{1-\lambda}}$ egy alkalmas, a csatorna tulajdonságaitól függő K konstanssal.*

Megjegyzés. Az előbb megfogalmazott eredményt az irodalomban gyakran a a csatorna kódolási tétel erős megfordításának hívják. E tétel bizonyításában ki fogjuk használni, hogy a tekintett csatorna véges állapotterű. Általános, nem feltétlenül véges állapotterű csatornák esetében csak egy gyengébb állítást tudnak bizonyítani, amelyet a csatorna kódolási tétel gyenge megfordításának neveznek. Mi ezzel az eredménnyel nem fogunk foglalkozni. Megelégszünk a csatorna kódolási tétel erős megfordításának a bizonyításával abban a speciális esetben, amikor a csatorna véges állapotterű.

Mielőtt rátérnék a fenti tételek bizonyítására, heurisztikus magyarázatot adok arra, hogy miért természetes ilyen eredményeket várni.

Ha adva van egy csatornát meghatározó $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűség függvény, akkor tekintsünk egy ehhez az átmenetvalószínűséghez adaptált μ mértéket a $V \times \tilde{V}$ halmazon, azaz egy olyan μ valószínűségi mértéket, amelyre egy μ eloszlású $(\eta, \tilde{\eta})$ véletlen vektor teljesíti a $P(\tilde{\eta} = \tilde{v}_j | \eta_i = v_i) = p(\tilde{v}_j|v_i)$ azonosságot minden $v_i \in V$ és $\tilde{v}_j \subset \tilde{V}$ pontra. Válasszuk ezt a μ mértéket úgy, hogy egy μ eloszlású $(\eta, \tilde{\eta})$ párra $I(\eta \wedge \tilde{\eta}) = C$, vagy legalábbis $I(\eta \wedge \tilde{\eta})$ nagyon közel van a C csatorna kapacitáshoz. Tekintsük a μ mérték μ^n n-ik hatványát a $V^n \times \tilde{V}^n$ szorzattéren, és jelölje ν_1^n és ν_2^n a μ^n mérték vetületét a V^n illetve $\tilde{V}^n$ térre. Próbáljunk a λ megkülönböztethető $v_1 \in V^n$, $v_2 \in V_n$, $\dots$, sorozatokat a ν_1^n mérték szerint tipikus sorozatok közül kiválasztani, és válasszuk a v_k vektornak megfelelő B_k halmazt úgy, mint a $\mu^n(\cdot|v_k)$ feltételes mérték szerinti tipikus sorozatok halmazát, illetve ezen halmaz kis módosítását. Ezt a módosítást a B_k halmazok diszjunktságának a biztosítása

érdekében tesszük. Annak érdekében, hogy megbecsüljük hány ilyen (v_k, B_k) párt tudunk választani becsüljük meg a B_k halmazok ν_2^n mértékét. Az első fejezet eredményei alapján a B_k halmaz körülbelül $2^{nH(\tilde{\eta}|\eta)}$ sorozatból áll, és az egyes sorozatok ν_2^n mértéke körülbelül $2^{-nH(\tilde{\eta})}$. Ezért $\nu_2^n(B_k) \sim 2^{nH(\tilde{\eta}|\eta)-nH(\tilde{\eta})} = 2^{-nI(\eta\wedge\tilde{\eta})}$, és körülbelül $2^{nI(\eta\wedge\tilde{\eta})}$ B_k halmaz fedi le a $\tilde{V}^n$ teret. Viszont, ha ennél sokkal kevesebb, nevezetesen csak $2^{n(1-\varepsilon)C} \sim 2^{(1-\varepsilon)nI(\eta\wedge\tilde{\eta})}$ számú B_k halmazt választunk, akkor bízhatunk abban, hogy ily módon egy a csatorna kódolási tételt teljesítő rendszert kapunk. A csatorna kódolási tétel bizonyítása tekinthető úgy, mint a fenti heurisztikus okoskodás rendbe tétele.

Arra, hogy a fenti módon végzett konstrukció éles eredményt ad csak kevésbé meggyőző heurisztikus érvet tudok adni. Mindenesetre jegyezzük meg, hogy bár a v_k sorozatokat és a B_k hozzátartozó halmazokat ebben a konstrukcióban egy véletlen szorzatmérték szerint választottuk ki, ez a választás kevésbé speciális, mint ahogy az első pillanatban látszik. A függetlenségből csak azt használtuk ki, hogy a $(v_i, \tilde{v}_j)$ párok relatív gyakorisága elő van írva, és a tekintett μ mérték megválasztásával ezt a relatív gyakoriságot írtuk elő. Ha jól írjuk elő a kiválasztandó v_k vektorokban szereplő jelek relatív gyakoriságát akkor ezzel a megszorítással nem csökkentettük lényegesen a keresett λ megkülönböztethető sorozatokból álló A halmaz nagyságát. Ezért szép esetekben a véletlen választás jó eredményt ad.

A csatorna kódolási tétel bizonyítása két lemmán alapszik. Az első az emlékezet nélküli csatorna kapacitásáról szól.

Lemma az emlékezet nélküli csatorna kapacitásáról. *Legyen adva adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálhatóan végtelen számosságú halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált $C < \infty$ csatorna kapacitású csatorna. Adva két olyan η és $\tilde{\eta}$ valószínűségi változó, amelyek össze vannak kapcsolva ezzel a csatornával, tekintsük független, az $(\eta, \tilde{\eta})$ párral azonos eloszlású véletlen vektorok egy $(\eta_1, \tilde{\eta}_1)$, $\dots$, $(\eta_n, \tilde{\eta}_n)$ sorozatát, és definiáljuk az $\iota_k = \iota_{\eta_k\wedge\tilde{\eta}_k}$, $1 \le k \le n$, valószínűségi változókat a következő képlet segítségével: $\iota_k = \log \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)}$, ha $\eta_k = v_i$ és $\tilde{\eta}_k = \tilde{v}_j$, ahol $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$. Minden $\varepsilon > 0$ számhoz megadhatóak olyan a csatornával összekapcsolt η és $\tilde{\eta}$ valószínűségi változók, amelyekre $\frac{1}{n}\sum_{k=1}^{n} \iota_k = \frac{1}{n}\sum_{k=1}^{n} \iota_{\eta_k\wedge\tilde{\eta}_k} \Rightarrow I(\eta\wedge\tilde{\eta})$, ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl, és $I(\eta\wedge\tilde{\eta}) > (1-\varepsilon)C$.*

Ha feltesszük, hogy a tekintett csatorna olyan, hogy bármely e csatornával összekapcsolt η és $\tilde{\eta}$ valószínűségi változó párra teljesül a $H(\tilde{\eta}) < \infty$ reláció is, akkor az is igaz, hogy egy az e csatorna által definiált n hosszúságú emlékezet nélküli csatorna csatorna kapacitása nC-vel egyenlő.

Megjegyzés. Valójában, e lemmának csak az első, könnyen bizonyítható részére lesz szükségünk. A második részben megfogalmazott eredmény tárgyalásának inkább elvi okai vannak. Ezen eredmény szemléletes tartalma az, hogy egy emlékezet nélküli csatorna kapacitását, azaz két ezen csatornával összekapcsolt valószínűségi változó kölcsönös információját nem lehet blokkosítással növelni. A legjobb, amit tenni tudunk az, hogy az egyes koordinátáknak megfelelő bemeneti és kimeneti jeleket egymástól

függetlenül optimálisan választjuk. Ez egyébként azt is jelenti, hogy nem lehet a kódolási tétel becslését triviális blokkosítással javítani. A csatorna kódolási tétel megfordításának általunk megfogalmazott és később bizonyítandó alakjából következik, hogy ez nem lehetséges véges állapotterű csatornákban. A fent megfogalmazott lemma eredménye kizárja az ilyen típusú javítás lehetőségét általánosabb csatornák esetében is. A $H(\tilde{\eta}) < \infty$ feltétel szerepeltetésének e lemma megfogalmazásában technikai okai vannak. E feltétel teljesülése vizsgálatainkat egyszerűbbé teszi, mert ekkor elkerülünk bizonyos nem feltétlenül abszolút konvergens sorok átrendezésével kapcsolatos kényelmetlen számolásokat.

Bizonyítás. Válasszunk olyan a csatornával összekapcsolt η és $\tilde{\eta}$ valószínűségi változókat amelyekre $I(\eta \wedge \tilde{\eta}) \geq (1-\varepsilon)C$, ahol C a csatorna kapacitása. Ekkor

$$\begin{aligned} E\iota_k &= \sum_{v_i, \tilde{v}_j} P(\eta = v_i, \tilde{\eta} = \tilde{v}_j) \log \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)} = \sum_{v_i, \tilde{v}_j} r(v_i, \tilde{v}_j) \log \frac{r(v_i, \tilde{v}_j)}{p(v_i)q(\tilde{v}_j)} \\ &= I(\eta \wedge \tilde{\eta}) \geq (1-\varepsilon)C, \end{aligned}$$

ahol $r(v_i, \tilde{v}_j) = P(\eta = v_i, \tilde{\eta} = \tilde{v}_j)$, $p(v_i) = P(\eta = v_i)$, és $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$. Ezért a nagy számok gyenge törvénye alapján $\frac{1}{n} \sum_{k=1}^{n} \iota_k \Rightarrow I(\eta \wedge \tilde{\eta})$, ha $n \to \infty$, és $I(\eta \wedge \tilde{\eta}) \geq (1-\varepsilon)C$.

Egy n hosszúságú emlékezet nélküli csatorna kapacitásának a kiszámolása érdekében tekintsünk két ezen emlékezet nélküli csatorna által összekapcsolt $\eta = (\eta_1, \ldots, \eta_n)$ és $\tilde{\eta} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_n)$ véletlen vektort, és becsüljük meg az $I(\eta \wedge \tilde{\eta})$ kölcsönös információt. Ennek érdekében először megmutatom, hogy $H(\tilde{\eta}|\eta) = \sum_{k=1}^{n} H(\tilde{\eta}_k|\eta_k)$. (Nem tettem fel, hogy az η illetve $\tilde{\eta}$ vektor koordinátái függetlenek.)

Vezessük be a következő jelöléseket:

$$\begin{aligned} p(v_{i_1}, \ldots, v_{i_n}) &= P(\eta_1 = v_{i_1}, \ldots, \eta_n = v_{i_n}), \\ q(\tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n}) &= P(\tilde{\eta}_1 = \tilde{v}_{j_1}, \ldots, \tilde{\eta}_n = \tilde{v}_{j_n}), \\ r(v_{i_1}, \ldots, v_{i_n}, \tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n}) &= P(\eta_1 = v_{i_1}, \ldots, \eta_n = v_{i_n}, \tilde{\eta}_1 = \tilde{v}_{j_1}, \ldots, \tilde{\eta}_n = \tilde{v}_{j_n}). \end{aligned}$$

Ezekkel a jelölésekkel $r(v_{i_1}, \ldots, v_{i_n}, \tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n}) = p(v_{i_1}, \ldots, v_{i_n}) \prod_{k=1}^{n} p(\tilde{v}_{j_k}|v_{i_k})$, és

$$\begin{aligned} H(\tilde{\eta}|\eta) &= \sum_{(v_{i_1}, \ldots, v_{i_n}), (\tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n})} r(v_{i_1}, \ldots, v_{i_n}, \tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n}) \log \left(\prod_{k=1}^{n} p(\tilde{v}_{j_k}|v_{i_k}) \right) \\ &= \sum_{(v_{i_1}, \ldots, v_{i_n}), (\tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n})} r(v_{i_1}, \ldots, v_{i_n}, \tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n}) \left(\sum_{k=1}^{n} \log p(\tilde{v}_{j_k}|v_{i_k}) \right) \\ &= \sum_{k=1}^{n} \sum_{(v_{i_1}, \ldots, v_{i_n}), (\tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n})} r(v_{i_1}, \ldots, v_{i_n}, \tilde{v}_{j_1}, \ldots, \tilde{v}_{j_n}) \log p(\tilde{v}_{j_k}|v_{i_k}). \end{aligned}$$

Vezessük be a következő mennyiségeket is: $r_k(v_{i_k}, \tilde{v}_{j_k}) = P(\eta_k = v_{i_k}, \tilde{\eta}_k = \tilde{v}_{j_k})$, $1 \le k \le n$. Azt állítom, hogy ezzel a jelöléssel az utolsó azonosság jobb oldalán lévő kifejezés belső összegét a következő módon írhatjuk fel rögzített k indexre:

$$\sum_{(v_{i_1},\dots,v_{i_n}),(\tilde{v}_{j_1},\dots,\tilde{v}_{j_n})} r(v_{i_1},\dots,v_{i_n},\tilde{v}_{j_1},\dots,\tilde{v}_{j_n}) \log p(\tilde{v}_{j_k}|v_{i_k})$$
$$= \sum_{(v_{i_k},\tilde{v}_{j_k})} r_k(v_{i_k},\tilde{v}_{j_k}) \log p(\tilde{v}_{j_k}|v_{i_k}).$$

Valóban, rögzítve a tekintett összegzésben a v_{i_k}, és $\tilde{v}_{j_k}$ argumentumokat, és összegezve az összes többi argumentum szerint az $r_k(v_{i_k}, \tilde{v}_{j_k}) \log p(\tilde{v}_{j_k}|v_{i_k})$ kifejezést kapjuk, majd ezekre az argumentumokra is összegezve megkapjuk az előbb felírt azonosságot. Ezt az azonosságot összegezve a k változó szerint, és felhasználva az előző azonosságot azt kapjuk, hogy

$$H(\tilde{\eta}|\eta) = \sum_{k=1}^{n} \sum_{(v_{i_k},\tilde{v}_{j_k})} r_k(v_{i_k},\tilde{v}_{j_k}) \log p(\tilde{v}_{j_k}|v_{i_k}) = \sum_{k=1}^{n} H(\tilde{\eta}_k|\eta_k),$$

amint azt állítottam.

Vegyük észre, hogy feltételeink teljesülése esetében $H(\tilde{\eta}) \le \sum_{k=1}^{n} H(\tilde{\eta}_k) < \infty$, ezért $I(\eta \wedge \tilde{\eta}) = H(\tilde{\eta}) - H(\tilde{\eta}|\eta) \le \sum_{k=1}^{n} (H(\tilde{\eta}_k) - H(\tilde{\eta}_k|\eta_k)) = \sum_{k=1}^{n} I(\eta_k \wedge \tilde{\eta}_k)$. Továbbá, ha az $\eta = (\eta_1, \dots, \eta_n)$ és $\tilde{\eta} = (\tilde{\eta}_1, \dots, \tilde{\eta}_n)$ véletlen vektorok az emlékezet nélküli csatorna szerinti összekapcsolt valószínűségi változók, akkor ezek $(\eta_k, \tilde{\eta}_k)$, $1 \le k \le n$, koordinátái összekapcsolt valószínűségi változók azon kiinduló csatorna szerint, amelynek a segítségével az emlékezet nélküli csatornát definiáltuk. Ezért $I(\eta_k \wedge \tilde{\eta}_k) \le C$, és a bizonyított egyenlőtlenségből az is következik, hogy $I(\eta \wedge \tilde{\eta}) \le nC$. Mivel ez tetszőleges az emlékezet nélküli csatorna szerint összekapcsolt η és $\tilde{\eta}$ véletlen vektorokra igaz, innen következik, hogy egy n hosszúságú emlékezet nélküli csatorna csatorna kapacitása kisebb vagy egyenlő, mint nC.

Annak érdekében, hogy belássuk, hogy az n hosszúságú emlékezet nélküli csatorna kapacitása valójában egyenlő az nC mennyiséggel tekintsünk egymástól független, és az emlékezet nélküli csatornát meghatározó csatornával összekapcsolt $(\eta_k, \tilde{\eta}_k)$, $1 \le k \le n$, valószínűségi párokat. Ekkor az $\eta = (\eta_1, \dots, \eta_n)$ és $\tilde{\eta} = (\tilde{\eta}_1, \dots, \tilde{\eta}_n)$ véletlen vektorok összekapcsoltak az emlékezet nélküli csatorna által, és $I(\eta \wedge \tilde{\eta}) = \sum_{k=1}^{n} I(\eta_k \wedge \tilde{\eta}_k)$. Mivel minden $\varepsilon > 0$ számra az $(\eta_k, \tilde{\eta}_k)$ párokat úgy is választhatjuk, hogy az $I(\eta_k \wedge \tilde{\eta}_k) \ge C - \varepsilon$ reláció teljesüljön, innen következik, hogy egy n hosszúságú emlékezet nélküli csatorna csatorna kapacitása nagyobb vagy egyenlő, mint nC. A lemmát beláttuk.

A következő lemmában olyan alsó becslést adunk egy alkalmasan konstruált λ megkülönböztethető halmaz elemszámáról, amely lehetővé teszi, hogy az előző lemma segítségével bebizonyítsuk a csatorna kódolási tételt.

Alsó becslés alkalmasan konstruált λ megkülönböztethető elemeket tartalmazó halmaz elemszámáról. *Legyen adva adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálható halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált csatorna. Legyen η és $\tilde{\eta}$ két e csatornával összekapcsolt valószínűségi változó, és vezessük be a $\iota_{\eta\wedge\tilde{\eta}}$ valószínűségi változót is a következő képlet segítségével: $\iota_{\eta\wedge\tilde{\eta}} = \log \frac{p(v_i|\tilde{v}_j)}{q(\tilde{v}_j)}$, ha $\eta = v_i$, és $\tilde{\eta} = \tilde{v}_j$, ahol $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$. Tetszőleges $z > 0$ és $0 < \lambda < 1$ számokra létezik a bemeneti jelek V halmazának olyan λ megkülönböztethető elemekből álló részhalmaza, amelynek N elemszáma teljesíti az*

$$N \geq 2^z(\lambda - P(\iota_{\eta\wedge\tilde{\eta}} < z))$$

egyenlőtlenséget.

Mielőtt leírnám a bizonyítást ismertetem annak fő gondolatát. Egy $\{v_{i_1}, \dots, v_{i_N}\}$ λ megkülönböztethető elemekből álló halmazt keresünk viszonylag nagy N elemszámmal, valamint a v_{i_k} pontokhoz olyan diszjunkt B_k halmazokat akarunk társítani, amelyekre $\sum_{\tilde{v}_j \in B_k} p(\tilde{v}_j|v_{i_k}) \geq 1 - \lambda$. Érdemes olyan B_k halmazokat választani, amelyekre a $P(\tilde{\eta} \in B_k) = \sum_{\tilde{v}_j \in B_k} q(\tilde{v}_j)$ valószínűségek kicsik. Mivel $\sum_{\tilde{v}_j \in B_k} p(\tilde{v}_j|v_{i_k}) \geq 1-\lambda$. (egy heurisztikus okoskodásban feltehetjük, hogy az utolsó képletben egyenlőség van,) ez azt sugallja, hogy a v_{i_k} elem megválasztása után olyan B_k halmazt érdemes definiálni, amelynek $\tilde{v}_j \in B_k$ elemeire $\frac{q(\tilde{v}_j)}{p(\tilde{v}_j|v_{i_k})}$ kicsi. Ezért a B_k halmazt $B_k = \tilde{V} \setminus \{\tilde{v}_j\colon \frac{p(\tilde{v}_j|v_{i_k})}{q(\tilde{v}_j)} \leq 2^z\}$ alakban keressük, ahol a z számot a tekintett csatorna tulajdonságaitól függően alkalmasan választjuk meg. Úgy akarjuk a λ megkülönböztethető $v_{i_k} \in V$ pontokat választani, hogy a v_{i_k} pont a neki megfelelő B_k halmazzal együtt teljesítse a $\sum_{\tilde{v}_j \in B_k} p(\tilde{v}_j|v_{i_k}) \geq 1 - \lambda$ egyenlőtlenséget. Valójában kissé másképpen kell eljárni annak érdekében, hogy diszjunkt B_k halmazokat válasszunk. Ha a $v_{i_1}, \dots, v_{i_{k-1}}$ pontokat és $B_1, \dots, B_{k-1}$ halmazokat már kiválasztottuk, akkor próbálunk olyan $v_{i_k} \in V$ pontot és hozzátartozó $B_k = (\tilde{V} \setminus \bigcup_{j=1}^{k-1} B_j) \setminus \{\tilde{v}_j\colon \frac{p(\tilde{v}_j|v_{i_k})}{q(\tilde{v}_j)} \leq 2^z\}$ halmazt találni, amelyekre $\sum_{\tilde{v}_j \in B_k} p(\tilde{v}_j|v_{i_k}) \geq 1-\lambda$. Ezt az eljárást addig folytatjuk szukcesszive, amíg meg nem akadunk. A lemma arról szól, hogy ilyen módon mekkora λ megkülönböztethető elemekből álló halmazt tudunk konstruálni.

A lemma bizonyítása. Definiáljuk a $W \subset V \times \tilde{V}$ halmazt, mint

$$W = \left\{(v_i, \tilde{v}_j)\colon\ (v_i, \tilde{v}_j) \in V \times \tilde{V},\ \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)} < 2^z\right\},$$

és ennek a halmaznak a metszeit a $\{v_i\} \times \tilde{V}$ halmazzal minden $v_i \in V$ pontra, mint

$$W^{v_i} = \{\tilde{v}_j\colon\ \tilde{v}_j \in \tilde{V},\ (v_i, \tilde{v}_j) \in W\}, \quad v_i \in V,$$

Egy olyan $v_{i_1} \in V$ pontot választunk, amelyre $P(\tilde{\eta} \in \tilde{V} \setminus W^{v_{i_1}}|\eta = v_{i_1}) \geq 1 - \lambda$, feltéve, hogy ilyen pont létezik. Ebben az esetben legyen $B_1 = \tilde{V} \setminus W^{v_{i_1}}$. Ha nincs ilyen $v_{i_1} \in V$

pont, akkor a procedurát befejezzük egyetlen pont kiválasztása nélkül. A $v_{i_1} \in V, \dots,$ $v_{i_k} \in V$ pontokat szukcesszive választjuk egymás után, és a v_{i_p} ponthoz társított B_p halmazt a

$$B_1 = \tilde{V} \setminus W^{v_1}, \quad B_p = \left(\bigcap_{l=1}^{p-1} W^{v_{i_l}} \right) \setminus W^{v_{i_p}}, \qquad p = 2, 3, \dots .$$

képlettel definiáljuk. Ezek a $B_1, B_2, \dots$ halmazok diszjunktak. Ha a $v_{i_1} \in V, \dots,$ $v_{i_k} \in V$ pontok már ki vannak választva, akkor a $k+1$-k lépésben olyan $v_{i_{k+1}} \in V$ pontot keresünk, amelyre $P\left(\tilde{\eta} \in \left(\bigcap_{l=1}^{k} W^{v_{i_l}} \right) \setminus W^{v_{i_{k+1}}} \middle| \eta = v_{i_{k+1}} \right) \geq 1 - \lambda$, ugyanis ez jelenti azt, hogy $P(\tilde{\eta} \in B_{k+1} | \eta = v_{i_{k+1}}) \geq 1 - \lambda$. Ha ez a feltétel teljesül, akkor választunk egy kívánt tulajdonságú $v_{i_{k+1}} \in V$ pontot, ha nem teljesül, akkor a v_{i_p} pontok választását a k-ik lépésben befejezzük. A lemmában szereplő N számot választhatjuk úgy, mint a legnagyobb olyan k számot, amelyre választottunk v_{i_k} pontot. Ugyanis az $A = \{v_{i_1}, \dots, v_{i_N}\}$ halmaz pontjai a $B_1, \dots, B_N$ halmazokkal együtt λ megkülönböztethető pontok.

Azon k index nagyságára kell tehát jó alsó becslést adni, amelyikre még tudjuk folytatni az algoritmusunkat, és megfelelő tulajdonságú $v_{i_k} \in V$ pontot találni. A k-ik lépés után akkor és csak akkor tudunk megfelelő tulajdonságú $v_{i_{k+1}} \in V$ pontot találni, ha $\inf_{v_i \in V} P(\tilde{\eta} \in \tilde{V}(k) \cup W^{v_i} | \eta = v_i) < \lambda$, ahol $\tilde{V}(1) = \tilde{V}$, és $\tilde{V}(k) = \bigcup_{l=1}^{k} (\tilde{V} \setminus W^{v_{i_l}})$, $k = 1, 2, \dots$. Az alábbi becslések segítségével meg tudjuk mutatni, hogy ez az egyenlőtlenség teljesül bizonyos számunkra érdekes esetekben.

$$\begin{aligned}
\inf_{v_i \in V} P(\tilde{\eta} \in \tilde{V}(k) \cup W^{v_i} | \eta = v_i) &\leq \sum_{v_i \in V} P(\eta = v_i) P(\tilde{\eta} \in \tilde{V}(k) \cup W^{v_i} | \eta = v_i) \\
&= \sum_{v_i \in V} P(\eta = v_i, \tilde{\eta} \in \tilde{V}(k) \cup W^{v_i}) = P((\eta, \tilde{\eta}) \in (V \times \tilde{V}(k)) \cup W) \\
&\leq P(\tilde{\eta} \in \tilde{V}(k)) + P((\eta, \tilde{\eta}) \in W).
\end{aligned}$$

Másrészt $P((\eta, \tilde{\eta}) \in W) = P(\iota_{\eta \wedge \tilde{\eta}} < z)$, és a $\tilde{V}(k)$ halmaz definiciója alapján

$$\begin{aligned}
P(\tilde{\eta} \in \tilde{V}(k)) &\leq \sum_{l=1}^{k} P(\tilde{\eta} \in \tilde{V} \setminus W^{v_{i_l}}) = \sum_{l=1}^{k} \sum_{\tilde{v}_j : \frac{p(\tilde{v}_j | v_{i_l})}{q(\tilde{v}_j)} \geq 2^z} q(\tilde{v}_j) \\
&\leq \sum_{l=1}^{k} \sum_{\tilde{v}_j : \frac{p(\tilde{v}_j | v_{i_l})}{q(\tilde{v}_j)} \geq 2^z} 2^{-z} p(\tilde{v}_j | v_{i_l}) \leq \sum_{l=1}^{k} \sum_{\tilde{v}_j \in \tilde{V}} 2^{-z} p(\tilde{v}_j | v_{i_l}) = \sum_{l=1}^{k} 2^{-z} = k 2^{-z}.
\end{aligned}$$

A fenti becslésekből következik, hogy

$$\inf_{v_i \in V} P(\tilde{\eta} \in \tilde{V}(k) \cup W^{v_i} | \eta = v_i) \leq P(\iota_{\eta \wedge \tilde{\eta}} < z) + k 2^{-z},$$

ezért $\inf_{v_i \in V} P(\tilde{\eta} \in \tilde{V}(k) \cup W^{v_i} | \eta = v_i) < \lambda$, ha $k < 2^z(\lambda - P(\iota_{\eta\wedge\tilde{\eta}} < z))$. Ez azt jelenti, hogy a kívánt tulajdonságú $v_{i_k} \in V$ pontok választását tudjuk folytatni a $\bar{k} = [2^z(\lambda - P(\iota_{\eta\wedge\tilde{\eta}} < z))] + 1$ értékig, ahol $[x]$ az x szám egész részét jelöli. Innen következik a lemma állítása.

A csatorna kódolási tétel bizonyítása. Válasszunk olyan $(\eta', \tilde{\eta}')$ valószínűségi változó párt, amelynek tagjai össze vannak kapcsolva az emlékezet nélküli csatornát meghatározó csatornával, és $I(\eta' \wedge \tilde{\eta}') \ge (1 - \frac{\varepsilon}{4})C$, ahol C ennek a csatornának a kapacitása. Legyen $(\eta_1, \tilde{\eta}_1)$, $(\eta_2, \tilde{\eta}_2)$, $\dots$, $(\eta_n, \tilde{\eta}_n)$ független, az $(\eta', \tilde{\eta}')$ véletlen vektorral azonos eloszlású véletlen vektorok sorozata, és a ι_k valószínűségi változót definiálja a $\iota_k = \log \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)}$, ha $\eta_k = v_i$, és $\tilde{\eta}_k = \tilde{v}_j$ képlet. Vezessük be az $\eta = (\eta_1, \dots, \eta_n)$ és $\tilde{\eta} = (\tilde{\eta}_1, \dots, \tilde{\eta}_n)$ véletlen vektorokat, és definiáljuk segítségükkel a $\iota_{\eta\wedge\tilde{\eta}}$ valószínűségi változót úgy, mint az *Alsó becslés alkalmasan konstruált λ megkülönböztethető elemeket tartalmazó halmaz elemszámáról* eredmény megfogalmazásában tettük. Ekkor $\iota_{\eta\wedge\tilde{\eta}} = \sum_{k=1}^{n} \iota_k$, ezért az emlékezet nélküli csatorna kapacitásáról szóló lemma (első) eredménye alapján létezik olyan n_0 küszöbindex, amelyre igaz, hogy $P(\iota_{\eta\wedge\tilde{\eta}} < (1 - \frac{\varepsilon}{2})Cn) \le \frac{\lambda}{2}$, ha $n \ge n_0$. Ezért alkalmazva az *Alsó becslés alkalmasan konstruált λ megkülönböztethető elemeket tartalmazó halmaz elemszámáról* nevű lemma eredményét $z = (1 - \frac{\varepsilon}{2})Cn$ választással azt kapjuk, hogy $N \ge 2^z(\lambda - P(\iota_{\eta\wedge\tilde{\eta}} < z)) \ge 2^{(1-\varepsilon/2)Cn}\frac{\lambda}{2} \ge 2^{(1-\varepsilon)Cn}$, ha $n \ge n_0(\varepsilon, \lambda)$. A tétel bizonyítását befejeztük.

Megjegyzés. Véges állapotterű csatornák esetén érvényes a csatorna kódolási tétel becslésének a következő a csatorna kódolási tétel megfordításában szereplő becsléshez hasonló élesítése. Az n hosszúságú emlékezet nélküli csatornának létezik olyan λ megkülönböztethető sorozatokból álló halmaza, amelynek $N = N(n)$ elemszámára teljesül az $N \ge \frac{\lambda}{2}2^{Cn - K\sqrt{n}/\sqrt{\lambda}}$ egyenlőtlenség alkalmas $K > 0$ számmal, ahol C a csatorna kapacitása.

Valóban, ebben az esetben, mint látni fogjuk, létezik olyan $(\eta', \tilde{\eta}')$ a csatornával összekapcsolt valószínűségi változó pár, amelyre $I(\eta' \wedge \tilde{\eta}') = C$, és ha $\eta = (\eta_1, \dots, \eta_n)$ és $\tilde{\eta} = (\tilde{\eta}_1, \dots, \tilde{\eta}_n)$, ahol $(\eta_k, \tilde{\eta}_k)$, $\le n \le k$, az $(\eta, \tilde{\eta}')$ valószínűségi változó pár független példányai, akkor nem nehéz belátni a Csebisev egyenlőtlenség segítségével — felhasználva az $\iota_{\eta\wedge\tilde{\eta}} = \sum_{k=1}^{n} \iota_{\eta_k\wedge\tilde{\eta}_k}$ azonosságot, — hogy létezik olyan $K > 0$ szám, amelyre $P(\iota_{\eta\wedge\tilde{\eta}} \le Cn - K\sqrt{\frac{n}{\lambda}}) = P\left(\sum_{k=1}^{n}(\iota_{\eta_k\wedge\tilde{\eta}_k} - E\iota_{\eta_k\wedge\tilde{\eta}_k}) \le -K\sqrt{\frac{n}{\lambda}}\right) \le \frac{\lambda}{2}$. Ezért az *Alsó becslés alkalmasan konstruált λ megkülönböztethető elemeket tartalmazó halmaz elemszámáról* eredményéből az említett egyenlőtlenséget kapjuk $z = Cn - K\frac{\sqrt{n}}{\sqrt{\lambda}}$ választással.

A csatorna kódolási tétel megfordításának a bizonyítása véges állapotterű csatornákra azon alapul, hogy ebben az esetben a csatorna kapacitást definiáló szuprémum felvétetik, és explicit módon lehet jellemezni azokat a csatornával összekapcsolt valószínűségi változó párok eloszlását, amelyek kölcsönös információja egyenlő a csatorna kapacitással. Az alábbi eredményt fogom bebizonyítani.

Tétel véges állapotterű csatorna optimális bemenetének a jellemzéséről. *Legyen adva két $V = \{v_1, \dots, v_m\}$ és $\tilde{V} = \{\tilde{v}_1, \dots, \tilde{v}_n\}$ véges halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált véges állapotterű $C < \infty$ csatorna kapacitású csatorna. Létezik olyan e csatornával összekapcsolt η, $\tilde{\eta}$ valószínűségi változó pár, amelyre $I(\eta \wedge \tilde{\eta}) = C$. Egy e csatornával összekapcsolt η, $\tilde{\eta}$ valószínűségi változó párra akkor és csak akkor igaz az $I(\eta \wedge \tilde{\eta}) = C$ egyenlőség, ha a $p(v_i) = P(\eta = v_i)$, $v_i \in V$, és $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$, $\tilde{v}_j \in \tilde{V}$, valószínűségek teljesítik a*

$$\sum_{\tilde{v}_j \in \tilde{V}} p(\tilde{v}_j|v_i) \log \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)} \le C \quad \textit{minden } v_i \in V \textit{ pontban} \tag{3.1}$$

egyenlőtlenségeket, és ha $p(v_i) > 0$, akkor a v_i pontnak megfelelő reláció élesíthető, és egyenlőséget is írhatunk az egyenlőtlenség helyett.

Megjegyzés. A (3.1) képlet úgy értendő, hogy $p(\tilde{v}_j|v_i) \log \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)} = 0$, ha $p(\tilde{v}_j|v_i) = 0$. A tétel azt is állítja, hogy $q(\tilde{v}_j) > 0$, ha $p(\tilde{v}_j|v_i) > 0$ valamilyen v_i bemenő állapotra (egy az $I(\eta \wedge \tilde{\eta}) = C$ feltételt teljesítő a csatornával összekapcsolt $\eta, \tilde{\eta}$ pár által definiált $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$ valószínűségre). Ellenkező esetben nem teljesülne a (3.1) egyenlőtlenség, mert annak baloldala ∞ lenne, mivel tartalmazna egy $p(\tilde{v}_j|v_i) \log \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)} = \infty$ alakú összeadandót. Igaz a

$$\sum_{\tilde{v}_j \in \tilde{V}} p(\tilde{v}_j|v_i) \log \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)} \ge 0 \quad \text{minden } v_i \in V \text{ pontban}$$

egyenlőtlenség is. Ez következik például a második fejezetben bizonyított *Egy I-divergencia típusú lemmából* az $a_j = p(\tilde{v}_j|v_i)$ és $b_j = q(\tilde{v}_j)$ szereposztással felhasználva az $A = \sum_{\tilde{v}_j \in \tilde{V}} p(\tilde{v}_j|v_i) = 1$ és $B = \sum_{\tilde{v}_j \in \tilde{V}} q(\tilde{v}_j) = 1$ relációkat.

A tétel bizonyítása két függvények konvexitásával (vagy konkávításával) kapcsolatos eredményen alapul. Az első lemma arról szól, hogy egy függvény, amely a kölcsönös információt definiáló függvény természetes kiterjesztésének tekinthető a pozitív ortánsra konkáv.

Lemma egy függvény konkávitásáról. *Legyenek adva bizonyos $p(i,j)$, $1 \le i \le M$, $1 \le j \le N$ számok, amelyekre $0 \le p(i,j) \le 1$ minden $1 \le i \le M$ és $1 \le j \le N$ indexre, és $\sum_{j=1}^{N} p(i,j) = 1$ minden $1 \le i \le M$ indexre. Definiáljuk segítségükkel az*

$$F(u_1, \dots, u_M) = \sum_{i=1}^{M} \sum_{j=1}^{N} u_i p(i,j) \log \frac{p(i,j)}{y_j}, \quad u_i \ge 0 \textit{ minden } 1 \le i \le M \textit{ indexre}$$

függvényt, ahol $y_j = y_j(u_1, \dots, u_M) = \sum_{i=1}^{M} p(i,j) u_i$. Az $F(u_1, \dots, u_M)$ függvény folytonos és konkáv az $A = \{(u_1, \dots, u_M)\colon\ u_i \ge 0,\ 1 \le i \le M\}$ halmazon.

1. megjegyzés. Az F függvényt definiáló összeg úgy értendő, hogy csak azon (i,j) párokra összegzünk, amelyekre $p(i,j)>0$. Tehát az $u_i p(i,j)\log\frac{p(i,j)}{y_j}=0$ konvenciót fogjuk alkalmazni a $p(i,j)=0$ esetben akkor is, ha $y_j=0$. Szükséges még definiálni az $u_i p(i,j)\log\frac{p(i,j)}{y_j}$ kifejezést akkor, ha az $u_i=0$, mert ekkor is előfordulhat, hogy $y_j=0$. Ebben az esetben az $u_i p(i,j)\log\frac{p(i,j)}{y_j}=0$, ha $u_i=0$ konvenciót fogjuk alkalmazni. A következő számolásokban az egyszerűbb jelölés kedvéért a log természetes és nem 2 alapú logaritmust fog jelölni, de ennek nincs nagy jelentősége. A természetes logaritmusról a 2 alapú logaritmusra való áttérés csak $\log_2 e$-vel való szorzást jelent.

2. megjegyzés. Értsük meg, miért hasznos számunkra a fenti lemma. Célunk egy véges állapotterű csatorna csatorna kapacitásának a meghatározása. A lemmában szereplő $F(u_1,\dots,u_M)$ segítségével ezt a problémát természetes módon át tudjuk fogalmazni egy feltételes szélsőérték feladattá, ahol az $F(u_1,\dots,u_M)$ függvény maximumát keressük a $\sum_{i=1}^{M} u_i=1$ feltétel mellett.

Valóban, tekintsünk egy olyan csatornát, amelyre $p(\tilde{v}_j|v_i)=p(i,j)$, és legyen η egy olyan valószínűségi változó, amelyre $P(\eta=v_i)=u_i$, $1\le i\le M$, $(\sum_{i=1}^{M} u_i=1)$. Ha $\tilde{\eta}$ olyan valószínűségi változó, amelyre η és $\tilde{\eta}$ össze vannak kapcsolva ezzel a csatornával, akkor $P(\tilde{\eta}=\tilde{v}_j)=y_j$, $1\le j\le N$, és $F(u_1,\dots,u_M)=I(\eta\wedge\tilde{\eta})$. Innen látható, hogy a csatorna kapacitás meghatározása valóban a fent említett feltételes szélsőérték feladathoz vezet. Ilyen típusú feladatokat érdemes az úgynevezett Kuhn–Tucker tétel segítségével vizsgálni, és mi is ennek az eredménynek egy egyszerű speciális esetét fogjuk alkalmazni. Ahhoz azonban, hogy ezt megtehessük, tudnunk kell, hogy a lemmában definiált F függvény konkáv.

A lemma bizonyítása. Az F függvény folytonos, ha azt a határon úgy definiáljuk, ahogy az első megjegyzésben tettük. Ezen állítás egyetlen részletesebb indoklást igénylő része az, hogy az olyan (i,j) indexpárokra, amelyekre $p(i,j)>0$ az $u_i p(i,j)\log\frac{p(i,j)}{y_j}$ függvény folytonos az olyan u pontokban is, amelyek i-ik koordinátája $u_i=0$. Azt kell ellenőrizni, hogy ha $u_i^{(n)}\to 0$, akkor $u_i^{(n)}p(i,j)\log\frac{p(i,j)}{y_j^{(n)}}\to 0$. De ekkor $\frac{p(i,j)}{y_j^{(n)}}\le\frac{1}{u_i^{(n)}}$, ezért $u_i^{(n)}p(i,j)\log\frac{p(i,j)}{y_j^{(n)}}\le u_i^{(n)}p(i,j)\log\frac{1}{u_i^{(n)}}\to 0$, és ezt kellett belátni. Ha $u_i>0$ (és $p(i,j)>0$), akkor $y_j>0$, és az $u_i p(i,j)\log\frac{p(i,j)}{y_j}$ függvény folytonossága könnyen látható.

Az F függvény konkávitásának bizonyításához elég megmutatni, hogy a $\left(\frac{\partial^2 F}{\partial u_k\partial u_l}\right)$, $1\le k,l\le M$, mátrix negatív szemidefinit minden $(u_1,\dots,u_M)$, $u_i>0$, $1\le i\le M$ pontban. (Mivel az F függvény folytonos elég csak azokat a pontokat tekinteni, amelyek koordinátái szigorúan pozitívak.) Számoljuk ki e mátrix elemeit. Felírhatjuk, hogy

$$\frac{\partial F}{\partial u_k}=\sum_{j:\,p(k,j)>0}p(k,j)\log p(k,j)-\sum_{j:\,p(k,j)>0}p(k,j)\log y_j-\sum_{i=1}^{M}\sum_{j:\,p(k,j)>0}u_i p(i,j)\frac{p(k,j)}{y_j}.$$

Az utolsó kifejezésben szereplő kettős szumma valójában konstans. Ugyanis

$$\sum_{i=1}^{M} \sum_{j:\ p(k,j)>0} u_i p(i,j) \frac{p(k,j)}{y_j} = \sum_{j:\ p(k,j)>0} \frac{p(k,j)}{y_j} \left(\sum_{i=1}^{M} u_i p(i,j) \right) = \sum_{j=1}^{N} p(k,j) \frac{y_j}{y_j} = 1.$$

Innen

$$\frac{\partial F}{\partial u_k} = \sum_{j:\ p(k,j)>0} p(k,j) \log \frac{p(k,j)}{y_j} - 1, \tag{3.2}$$

és mivel $y_j > 0$ az $\{(u_1, \ldots, u_M)\colon\ u_i > 0,\ 1 \le i \le M\}$ halmazon

$$\frac{\partial^2 F}{\partial u_k \partial u_l} = -\frac{\partial}{\partial u_l} \left(\sum_{j=1}^{N} p(k,j) \log y_j \right) = -\sum_{j=1}^{N} \frac{p(k,j)p(l,j)}{y_j}.$$

Az F függvény konkávitásának a bizonyításához azt kell belátni, hogy tetszőleges $((\alpha(1), \ldots, \alpha(M))$ vektorra

$$\sum_{k=1}^{M} \sum_{l=1}^{M} \frac{\partial^2 F}{\partial u_k \partial u_l} \alpha(k)\alpha(l) \le 0.$$

Viszont

$$\begin{aligned} \sum_{k=1}^{M} \sum_{l=1}^{M} \frac{\partial^2 F}{\partial u_k \partial u_l} \alpha(k)\alpha(l) &= -\sum_{k=1}^{M} \sum_{l=1}^{M} \sum_{j=1}^{N} \frac{p(k,j)p(l,j)}{y_j} \alpha(k)\alpha(l) \\ &= -\sum_{j=1}^{N} \frac{1}{y_j} \sum_{k=1}^{M} \sum_{l=1}^{M} p(k,j)p(l,j)\alpha(k)\alpha(l) = -\sum_{j=1}^{N} \frac{1}{y_j} \left(\sum_{k=1}^{M} p(k,j)\alpha(k) \right)^2 \le 0. \end{aligned}$$

Szükségünk lesz még az előző lemma egy kiegészítésére is, amelyik a $\frac{\partial F}{\partial u_k}(u)$ parciális derivált viselkedését olyan u vektorokra is leírja, amelyeknek van 0 koordinátája. Amikor a k-ik koordináta szerinti parciális deriváltat tekintjük meg kell különböztetnünk azt az esetet, amikor az u vektor u_k koordinátája az $u_k > 0$ és amikor az $u_k = 0$ relációt teljesíti. Be fogjuk látni, hogy az első esetben a parciális derivált folytonos, és teljesíti a (3.2) formulát. A második esetben azt állítom, hogy ekkor is létezik (az egyoldali) parciális derivált, és az egyenlő a parciális derivált iránymenti határértékével. Ez az iránymenti határérék azonban végtelen is lehet.

Az, hogy a $\frac{\partial F}{\partial u_k}(u)$ függvénynek létezik iránymenti határértéke az $u = (u_1, \ldots, u_M)$ pontban azt jelenti, hogy létezik a $G_k(u) = \lim\limits_{t \to 0+} \frac{\partial F}{\partial u_k}(u + t\tilde{u})$ esetleg végtelennel egyenlő határérték minden olyan $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_M)$ vektorra, amelynek k-ik koordinátája $\tilde{u}_k = 1$, és $u + t\tilde{u} \in A$ elég kis $t > 0$ számokra. Továbbá ez a határérték nem függ az $\tilde{u}$ vektor választásától. A $\lim\limits_{t \to 0+}$ azt jelenti, hogy szigorúan pozitív $t > 0$ számokkal tekintjük ezt a

limeszt. (A most bevezetett iránymenti határértékhez hasonló fogalommal találkoztunk a komplex függvénytanban is, amikor egy hatványsor esetleges folytonosságát vizsgálják a hatványsor konvergenciakörének egy pontjában.) Nem nehéz megmutatni, hogy ha a $G_k(u)$ iránymenti határérték létezik egy olyan $u = (u_1, \dots, u_M)$ pontban, amelyre $u_k = 0$, akkor $G_k(u) = \frac{\partial F}{\partial u_k}(u)$, ahol $\frac{\partial F}{\partial u_k}(u)$ a megfelelő féloldali parciális deriváltat jelöli.

Kiegészítés az egy függvény konkávitásáról szóló lemmához. *A lemmában tekintett F függvény $\frac{\partial F}{\partial u_k}$ parciális deriváltja folytonos függvény az*

$$A_k = \{(u_1, \dots, u_M)\colon\ u_i \ge 0 \text{ minden } 1 \le i \le M \text{ indexre, és } u_k > 0\}$$

halmazon minden $1 \le k \le M$ indexre, és teljesíti a (3.2) formulát.

Ha az $u = (u_1, \dots, u_M) \in A$ pontban $u_k = 0$, akkor is létezik a $\frac{\partial F}{\partial u_k}(u)$ függvény $G_k(u)$ iránymenti határértéke az u pontban. Ha az ilyen u pontokban a $\frac{\partial F}{\partial u_k}(u)$ parciális deriváltat úgy definiáljuk, mint ezen iránymenti határértéket, akkor a (3.2) formula érvényes lesz minden $u \in A$ pontban. Ez az iránymenti határérték akkor és csak akkor véges, ha az $y_j = y_j(u)$ számra $y_j > 0$ minden olyan j indexre, amelyre $p(k, j) > 0$.

Bizonyítás. Mivel $u_k > 0$ egy $u = (u_1, \dots, u_k) \in A_k$ pontra, ezért az A_k halmaz pontjaiban $y_j = y_j(u) \ge u_k p(k, j) > 0$ minden olyan j indexre, amelyre $p(k, j) > 0$. Ezt felhasználva kapjuk, hogy a (3.2) formula, illetve az ezen azonosság bizonyításában felhasznált formulák az $u \in A_k$ pontokban is érvényesek. Ezután a (3.2) formula segítségével az is könnyen látható, hogy a $\frac{\partial F}{\partial u_k}$ parciális derivált folytonos az A_k halmaz pontjaiban.

Ha $u = (u_1, \dots, u_M) \in A$ olyan pont, amelyre $u_k = 0$, és az $\tilde{u} = (\tilde{u}_1, \dots, \tilde{u}_M)$ pontra $\tilde{u}_k = 1$, és $u + t\tilde{u} \in A$ elég kis $t > 0$ számra, akkor az $u + t\tilde{u} \in A_k$ reláció is teljesül kis $t > 0$ számokra. Ezért a $\frac{\partial F}{\partial u_k}(u + t\tilde{u})$ kifejezésre érvényes a (3.2) formula, és a $\frac{\partial F}{\partial u_k}$ derivált iránymenti határértékének létezéséhez az u pontban azt kell ellenőrizni, hogy az e formula jobb oldalán szereplő kifejezésnek van az $\tilde{u}$ vektortól független határértéke, amely egyenlő a (3.2) formula jobboldalán levő kifejezés értékével az u pontban. Ez azonban könnyen látható felhasználva, hogy $\lim_{t \to 0} y_j(u + t\tilde{u}) = y_j(u)$ minden $1 \le j \le M$ indexre. Továbbá ez a limesz akkor és csak akkor végtelen, ha létezik olyan j index, amelyre $p(k, j) > 0$, és $y_j = 0$.

A másik lemma, amely hasznos lesz az egy véges állapotterű csatorna optimális bemenetének a jellemzésében az úgynevezett Kuhn–Tucker tételnek egy speciális és egyszerűbben bizonyítható esete. A Kuhn–Tucker tétel a konvex programozás egyik alapvető eredménye. Bár a Kuhn–Tucker tétel általános alakjára nem lesz szükségünk, röviden, bizonyítás nélkül ismertetni fogom ezt az eredményt is. A bizonyítás megtalálható például Jiří Matoušek és Bernd Gärtner *Understanding and Using Linear Programming* című könyvében (Proposition 8.7.2). Bizonyítani csak a következő eredményt fogom.

A Kuhn–Tucker tétel egy speciális esete. *Legyen adva egy folytonos és konkáv $F(u_1, \dots, u_M)$ függvény az $A = \{(u_1, \dots, u_M)\colon\ u_i \ge 0,\ 1 \le i \le M\}$ halmazon, amelyre a $\frac{\partial F}{\partial u_k}$ parciális derivált létezik és folytonos az $A_k = \{(u_1, \dots, u_M)\colon\ u_i \ge 0,\ 1 \le$*

$i \le M$, és $u_k > 0\}$ halmazon minden $1 \le k \le M$ indexre, továbbá létezik a $\frac{\partial F}{\partial u_k}$ parciális deriváltiránymenti határértéke azon $u = (u_1, \dots, u_M) \in A$ pontokban is, amelyekre $u_k = 0$. Definiáljuk a $\frac{\partial F}{\partial u_k}$ parciális deriváltat ezekben a pontokban, mint ezt az iránymenti deriváltat, és keressük az $F(u_1, \dots, u_M)$ függvény maximumát a $\sum_{i=1}^{M} u_i = 1$ feltétel mellett, azaz az $A \cap B$ halmazon, ahol $B = \{(u_1, \dots, u_M)\colon \sum_{i=1}^{M} u_i = 1\}$. Egy $\bar{u} = (\bar{u}_1, \dots, \bar{u}_M)$, $\bar{u} \in A \cap B$, vektor akkor és csak akkor megoldása ennek a szélsőérték feladatnak, ha létezik olyan $D < \infty$ konstans, amelyre $\frac{\partial F}{\partial u_i}(\bar{u}) \le D$ minden $1 \le i \le M$ indexre, és ebben a relációban egyenlőség áll azon i indexekre, amelyekre $\bar{u}_i > 0$.

Bizonyítás. Legyen $\bar{u} = (\bar{u}_1, \dots, \bar{u}_M)$ a szélsőérték feladat megoldása. Válasszuk ki e vektor két $\bar{u}_i$ és $\bar{u}_j$ koordinátát úgy, hogy $\bar{u}_i > 0$. Jelölje $u(i,j)$ azt az M-dimenziós vektort, amelynek az i-ik koordinátája -1, a j-ik koordinátája 1, és az összes többi koordinátája 0. Ekkor az $\bar{u}(\vartheta) = \bar{u} + \vartheta u(i,j)$ vektorra $\bar{u}(\vartheta) \in A \cap B$, sőt $\bar{u}(\vartheta) \in A_i \cap A_j \cap B$, ha a $\vartheta > 0$ szám elég kicsi. Innen azt kapjuk, hogy a $g(\vartheta) = \frac{F(\bar{u}(\vartheta)) - F(\bar{u})}{\vartheta}$ függvény kis $\vartheta > 0$ paraméterrel teljesíti a

$$0 \ge g(\vartheta) = \frac{dF(u(\vartheta))}{d\vartheta}(\vartheta') = \left(-\frac{\partial F}{\partial u_i}(\bar{u} + \vartheta' \bar{u}(i,j)) + \frac{\partial F}{\partial u_j}(\bar{u} + \vartheta' \bar{u}(i,j)) \right)$$

relációt alkalmas $0 < \vartheta' < \vartheta$ számmal. A $\vartheta \to 0$ határátmenetet alkalmazva ebben a képletben azt kapjuk, hogy $\frac{\partial F}{\partial u_j}(\bar{u}) \le \frac{\partial F}{\partial u_i}(\bar{u})$. Legyen $D = \frac{\partial F}{\partial u_i}(\bar{u})$. Ekkor $\frac{\partial F}{\partial u_j}(\bar{u}) \le D$ minden $1 \le j \le M$ indexre. De mivel a fenti érvelésben tetszőleges olyan i indexet választhatunk, amelyre $\bar{u}_i > 0$ innen az is következik, hogy $D = \frac{\partial F}{\partial u_i}(\bar{u})$ minden olyan i indexre, amelyre $\bar{u}_i > 0$.

Megfordítva, legyen $\bar{u} = (\bar{u}_1, \dots, \bar{u}_M) \in A \cap B$ olyan vektor, amelyikre létezik olyan $D < \infty$ szám, amelyre $\frac{\partial F}{\partial u_i}(\bar{u}) \le D$ minden $1 \le i \le M$ indexre, és ebben a relációban egyenlőség van azon i indexekre, amelyekre $\bar{u}_i > 0$. Ekkor, mivel F konkáv függvény

$$F(\vartheta u + (1-\vartheta)\bar{u}) \ge \vartheta F(u) + (1-\vartheta) F(\bar{u})$$

minden $u \in A \cap B$ vektorra és $0 \le \vartheta \le 1$ számra, ami úgy is írható, hogy

$$\frac{F(\bar{u} + \vartheta(u - \bar{u})) - F(\bar{u})}{\vartheta} \ge F(u) - F(\bar{u}).$$

Másrészt, az $\bar{F}(s) = F(\bar{u} + s(u - \bar{u}))$, $0 \le s \le 1$, függvény segítségével azt írhatjuk, hogy

$$\frac{F(\bar{u} + \vartheta(u - \bar{u})) - F(\bar{u})}{\vartheta} = \frac{1}{\vartheta} \int_0^{\vartheta} \frac{d\bar{F}(s)}{ds}\, ds = \frac{1}{\vartheta} \int_0^{\vartheta} \sum_{i=1}^{M} \frac{\partial F}{\partial u_i}(\bar{u} + s(u - \bar{u}))(u_i - \bar{u}_i)\, ds.$$

Innen $\vartheta \to 0$ határátmenettel kapjuk felhasználva $\frac{\partial F}{\partial u_k}$ parciális deriváltak folytonossági tulajdonságait az $\bar{u}$ pontban, hogy

$$\sum_{i=1}^{M} \frac{\partial F}{\partial u_i}(\bar{u})(u_i - \bar{u}_i) \ge F(u) - F(\bar{u}).$$

Azt állítom, hogy $\sum_{i=1}^{M} \frac{\partial F}{\partial u_i}(\bar{u})(u_i - \bar{u}_i) \leq 0$. Valóban, abban a speciális esetben, amikor $\frac{\partial F}{\partial u_i}(\bar{u}) = D$ minden $1 \leq i \leq M$ indexre ez a reláció egyenlőséggel is érvényes, mert $\sum_{i=1}^{M} \bar{u}_i = \sum_{i=1}^{M} u_i = 1$ az $u \in B$ és $\bar{u} \in B$ feltétel miatt. A $\frac{\partial F}{\partial u_i}(\bar{u}) < D$ szigorú egyenlőtlenség csak olyan i indexekre érvényes, amelyekre $\bar{u}_i = 0$. Ezenkívül $u_i \geq 0$ az $u \in A$ reláció miatt, ezért ebben az esetben $\frac{\partial F}{\partial u_i}(\bar{u})(u_i - \bar{u}_i) \leq D(u_i - \bar{u}_i)$. Innen $\sum_{i=1}^{M} \frac{\partial F}{\partial u_i}(\bar{u})(u_i - \bar{u}_i) \leq \sum_{i=1}^{M} D(u_i - \bar{u}_i) = 0$, amint állítottam. Azt kaptuk, hogy $F(u) - F(\bar{u}) \leq 0$ minden $u \in A \cap B$ vektorra, tehát a $\bar{u}$ pont a vizsgált szélsőérték feladat megoldása. A tételt beláttuk.

Megfogalmazom a Kuhn–Tucker tétel eredeti alakját.

Kuhn–Tucker tétel. *Tekintsük a*

$$\min f(x_1, \ldots, x_N)$$
$$Ax^* = b^*$$
$$x_i \geq 0 \text{ } \textit{minden } 1 \leq i \leq N \text{ } \textit{indexre}$$

optimalizációs feladatot, ahol $f(x_1, \ldots, x_N)$ egy mindenütt differenciálható, konvex függvény, $b = (b_1, \ldots, b_M) \in R^M$, A egy $N \times M$ méretű mátrix, x^* az x vektor transzponáltját jelöli, és $x = (x_1, \ldots, x_N)$. Egy $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N) \in R^N$, $x_i \geq 0$, $1 \leq i \leq N$, $A\bar{x}^* = b^*$, vektor akkor és csak akkor megoldása ennek az optimalizációs feladatnak, ha létezik olyan $m = (m_1, \ldots, m_M) \in R^M$ vektor, amelyre

$$\frac{\partial f}{\partial x_j}(\bar{x}_1, \ldots, \bar{x}_N) + (m, a_j) \quad \begin{cases} = 0 & \textit{ha } \bar{x}_j > 0 \\ \geq 0 & \textit{ha } x_j = 0, \end{cases}$$

ahol a_j az A mátrix j-ik oszlopát, és (x, y) az x és y vektorok skalárszorzatát jelöli.

Az előzőleg tárgyalt konkáv függvény maximalizációs problémája átfogalmazható ilyen konvex optimalizációs problémává -1-gyel való szorzás segítségével. Az optimalizáció ott kimondott feltétele is átfogalmazható az e tételben kimondott alakra. Ebben az esetben az $1 \times N$ méretű A mátrix szerepét a csupa 1 számot tartalmazó vektor játssza, és az 1 dimenziós m vektor $-D$-vel egyenlő, a tételben szerelő D számmal. Az e jegyzetben bizonyított tétel, — ha eltekintünk az F függvényre tett simasági feltételektől — a Kuhn–Tucker tétel speciális esetének tekinthető ezzel a szereposztással.

A Kuhn–Tucker tétel lényeges újdonsága az általunk tárgyalt eredményhez képest az, hogy több lineáris feltétel megkövetelése esetén is jellemzi az optimális megoldásokat. A bizonyítás lényegesen új gondolatok felhasználását igényelte. A tétel igazolásának fő nehézsége a benne szereplő feltétel szükségességének, vagyis annak a ténynek a bizonyítása, hogy ha egy $x = (x_1, \ldots, x_N)$ vektor megoldása a minimum feladatnak, akkor az teljesíti a megadott egyenlőségekből és egyenlőtlenségekből álló feltételt. Különösen

fontos a feltételben szereplő m vektor megtalálása. Ezt meg lehet tenni a lineáris programozás dualitás tételének a segítségével. Az eredeti optimalizációs feladat megoldása teljesít egy lineáris optimalizációs feladatot (implicit módon definiált együtthatókkal), és az m vektor e lineáris programozási feladat duáljának a megoldása.

Maga a Kuhn–Tucker tétel jellege hasonlít a Lagrange-féle multiplikátor módszerre. A lényeges különbség a két eredmény között az, hogy a Kuhn–Tucker tétel az nemcsak szükséges, hanem elégséges feltételét is ad arra, hogy egy vektor az optimálizációs feladat megoldása legyen, és az optimum keresésekor a tartomány határpontjait is figyelembe veszi. Ennek viszont az az ára, hogy csak viszonylag speciális problémákat lehet ezzel a módszerrel vizsgálni. Így például a problémában szereplő kényszer feltételek lineárisak.

Az előzőleg igazolt eredmények segítenek az alábbi bizonyításban.

A véges állapotterű csatorna optimális bemenetének a jellemzéséről szóló tétel bizonyítása. Ha egy a a csatornával összekapcsolt η, $\tilde{\eta}$ valószínűségi változó pár $p(v_i) = P(\eta = v_i)$ valószínűségeit az egy függvény konkávitásáról szóló lemmában szereplő u_i, a csatorna $p(\tilde{v}_j|v_i)$ átmenetvalószínűségeit az e lemmában szereplő $p(i,j)$ számokkal azonosítjuk, akkor $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j) = y_j$ a szintén e lemmában szereplő y_j számokkal. Továbbá, ha η és $\tilde{\eta}$ két a csatornával összekapcsolt valószínűségi változó, és $P(\eta = v_i) = u_i$, $1 \le i \le M$, akkor $I(\eta \wedge \tilde{\eta}) = F(u_1, \dots, u_M)$ a lemmában definiált $F(\cdot)$ függvénnyel. Ezért egy olyan $\eta, \tilde{\eta}$ a csatornával összekapcsolt valószínűségi változó pár eloszlásának a jellemzése, amelyre $I(\eta \wedge \tilde{\eta}) = C$, ahol C a csatorna kapacitása, ekvivalens azzal a feladattal, hogy találjuk meg a lemmában definiált (konkáv) $F(u_1, \dots, u_M)$, $u_i \ge 0$, $1 \le i \le M$, függvény maximumát a $\sum_{i=1}^{M} u_i = 1$ kényszerfeltétel mellett. E feladatban egy folytonos függvény maximumát keressük egy kompakt halmazon, tehát ez a maximum létezik.

A keresett maximum megtalálása érdekében alkalmazhatjuk a *A Kuhn–Tucker tétel egy speciális esete* néven megfogalmazott állítást, mert a minket érdeklő feladatban e tétel feltételei teljesülnek. Továbbá a (3.2) formulában kiszámoltuk a $\frac{\partial F}{\partial u_k}$ parciális deriváltakat. E képlet és az előbb említett tétel azt adják, hogy egy a csatornával összekapcsolt $\eta, \tilde{\eta}$ valószínűségi változó párra akkor és csak akkor teljesül az $I(\eta \wedge \tilde{\eta}) = C$ reláció, ha az $\tilde{\eta}$ valószínűségi változó $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$, $1 \le j \le N$, eloszlása teljesíti a

$$\begin{aligned} \sum_{j=1}^{N} p(\tilde{v}_j|v_k) \log \frac{p(\tilde{v}_j|v_k)}{q(\tilde{v}_j)} &= D, \quad \text{ha } p(v_k) > 0, \\ \sum_{j=1}^{N} p(\tilde{v}_j|v_k) \log \frac{p(\tilde{v}_j|v_k)}{q(\tilde{v}_j)} &\le D, \quad \text{ha } p(v_k) = 0 \end{aligned} \tag{3.3}$$

relációt valamilyen $D < \infty$ számmal minden $1 \le k \le M$ indexre. (A (3.3) formulában szereplő összegek úgy értendőek, hogy $p(\tilde{v}_j|v_k) \log \frac{p(\tilde{v}_j|v_k)}{q(\tilde{v}_j)} = 0$, ha $p(\tilde{v}_j|v_k) = 0$.) A k-ik egyenletet vagy egyenlőtlenséget megszorozva $p(v_k)$-val egyenlőséget kapunk minden k

indexre. Ezeket összeadva azt kapjuk, hogy

$$\sum_{k=1}^{M}\sum_{j=1}^{N} r(v_k,\tilde{v}_j)\log\frac{p(\tilde{v}_j|v_k)}{q(\tilde{v}_j)} = D,$$

ahol $r(v_k,\tilde{v}_j) = P(\eta = v_k, \tilde{\eta} = \tilde{v}_j)$, $1 \le k \le M$, $1 \le j \le N$. Ennek az egyenletnek a baloldala egyenlő az $I(\eta \wedge \tilde{\eta}) = C$ számmal. Tehát egy a csatornával összekapcsolt $\eta, \tilde{\eta}$ valószínűségi változó párra akkor és csak akkor érvényes az $I(\eta \wedge \tilde{\eta}) = C$ reláció, ha az $\tilde{\eta}$ valószínűségi változó $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$, $1 \le j \le M$, eloszlása teljesíti a (3.3) relációt a $D = C$ számmal. A tételt beláttuk.

Megfogalmazok egy lemmát, amely segít a csatorna kódolási tétel megfordításának a bizonyításában.

Felső becslés egy λ megkülönböztethető elemeket tartalmazó halmaz elemszámáról. *Legyen adva adva két $V = \{v_1, v_2, \dots\}$ és $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges vagy megszámlálható halmaz, és közöttük egy $p(\tilde{v}_j|v_i)$, $v_i \in V$, $\tilde{v}_j \in \tilde{V}$, átmenetvalószínűségekkel definiált csatorna. Legyen η és $\tilde{\eta}$ két e csatornával összekapcsolt valószínűségi változó. Vezessük be a $q(\tilde{v}_j) = P(\tilde{\eta} = \tilde{v}_j)$, $\tilde{v}_j \in \tilde{V}$, valószínűségeket és az*

$$A(v_i) = \left\{\tilde{v}_j\colon\ \tilde{v}_j \in \tilde{V},\ \ \frac{p(\tilde{v}_j|v_i)}{q(\tilde{v}_j)} \ge 2^{\vartheta}\right\}, \quad v_i \in V,$$

halmazokat egy alkalmasan választott ϑ paraméterrel. Tegyük fel, hogy teljesül a

$$P(\tilde{\eta} \in A(v_i)|\eta = v_i) = \sum_{\tilde{v}_j\colon\, p(\tilde{v}_j|v_i) \ge 2^{\vartheta} q(\tilde{v}_j)} p(\tilde{v}_j|v_i) \le \gamma \tag{3.4}$$

egyenlőtlenség valamely $\gamma < 1$ számmal minden $v_i \in V$ pontra. Legyen $\gamma + \lambda < 1$ valamely $\lambda > 0$ számmal. Ekkor tetszőleges $A = \{v_{i_1}, \dots, v_{i_N}\} \subset V$ λ megkülönböztethető pontokból álló halmaz N elemszámára $N \le \frac{2^{\vartheta}}{1-\lambda-\gamma}$.

1. megjegyzés. A (3.4) formulában szereplő feltételes valószínűséget az ott felírt összegként definiáljuk abban az esetben is, ha $P(\eta = v_i) = 0$. Az $A(v_i)$ halmaz definiciójában nem egyértelmű, hogy azok a $\tilde{v}_j \in \tilde{V}$ pontok, amelyekre mind a $p(\tilde{v}_j|v_i) = 0$ mind a $q(\tilde{v}_j) = 0$ azonosság teljesül beletartoznak-e az $A(v_i)$ halmazba. Ennek azonban nincs jelentősége, mert az ilyen $\tilde{v}_j$ pontok hozadéka nulla a (3.4) képletben szereplő összegben. Kényelmi okokból azt fogom feltételezni, hogy az ilyen pontok nincsenek benne az $A(v_i)$ halmazban.

2. megjegyzés. Mind itt, mind a csatorna kódolási tételben tekintettünk egy a csatornával összekapcsolt $(\eta, \tilde{\eta})$ valószínűségi változó párt. A csatorna kódolási tételben olyan $(v_i, A(v_i))$, $v_i \in V$, $A(v_i) \subset \tilde{V}$ párokat kerestünk, amelyekre $A(v_i) = A(v_i, \vartheta) = \{\tilde{v}_j\colon\ \frac{q(\tilde{v}_j)}{p(\tilde{v}_j|v_i)} \le 2^{\vartheta}\}$ egy kis ϑ számmal, és a $P(\tilde{\eta} \in A(v_i)|\eta = v_i)$ feltételes valószínűség nagy. Így tudtunk ugyanis olyan (v_i, B_i) párokat találni alkalmas, (diszjunkt) $B_i \subset$

$A(v_i)$ halmazokkal, amelyekre $P(\tilde\eta \in B_i|\eta = v_i) \ge 1-\lambda$, és a B_i halmazok kicsik abban az értelemben, hogy a $P(\tilde\eta \in B_i)$ valószínűségek kicsik. A fenti, a csatorna kódolási tételt előkészítő eredményben azt állítjuk, hogy ha egy ellenkező irányú tulajdonság érvényes, nevezetesen, ha a $P(\tilde\eta \in A(v_i,\vartheta)|\eta = v_i)$ feltételes valószínűségek kicsik *minden* $v_i \in V$ feltétel esetén még viszonylag nagy ϑ számokra is, akkor csak viszonylag kevés λ megkülönböztethető $v_i \in V$ elem létezik. (Ez az állítás kissé eltérő, de ekvivalens formában van megfogalmazva a fenti eredményben.) Ennek oka, mint a bizonyításból látszódni fog, az, hogy ebben az esetben minden olyan $B \subset \tilde V$ halmazra, amelyre $P(\tilde\eta \in B|\eta = v_i) > 1-\lambda$, a $P(\tilde\eta \in B)$ valószínűség is nagy.

Bizonyítás. Adva egy $A = \{v_{i_1}, \dots, v_{i_N}\} \subset V$ λ megkülönböztethető pontokból álló halmaz válasszunk olyan $B_1 \in \tilde V, \dots, B_N \in \tilde V$ diszjunkt halmazokat, amelyekre $P(\tilde\eta \in B_k|\eta = v_{i_k}) \ge 1-\lambda$ minden $1 \le k \le N$ indexre. Ekkor

$$1-\lambda-\gamma \le P(\tilde\eta \in B_k \setminus A(v_{i_k})|\eta = v_{i_k}) = \sum_{\tilde v_j\colon\, \tilde v_j \in B_k,\, p(\tilde v_j|v_{i_k}) \le 2^\vartheta q(\tilde v_j)} p(\tilde v_j|v_{i_k})$$

$$\le 2^\vartheta \sum_{\tilde v_j\colon\, \tilde v_j \in B_k,\, p(\tilde v_j|v_{i_k}) \le 2^\vartheta q(\tilde v_j)} q(\tilde v_j) = 2^\vartheta P(\tilde\eta \in B_k \setminus A(v_{i_k})) \le 2^\vartheta P(\tilde\eta \in B_k)$$

minden $1 \le k \le N$ indexre. Összegezve ezeket az egyenlőtlenségeket minden $1 \le k \le N$ indexre és felhasználva, hogy a B_k halmazok diszjunktak azt kapjuk, hogy

$$N(1-\lambda-\gamma) \le 2^\vartheta.$$

A lemmát beláttuk.

A csatorna kódolási tétel megfordításának a bizonyítása. Legyen η' és $\tilde\eta'$ két olyan egy V véges és C kapacitású csatornával összekötött valószínűségi változó, amelyekre $I(\eta' \wedge \tilde\eta') = C$. Legyen $(\eta_1,\tilde\eta_1), \dots, (\eta_n,\tilde\eta_n)$ az $(\eta',\tilde\eta')$ véleten vektorral azonos eloszlású, független véletlen vektorok sorozata, amelyeket a V által meghatározott V^n emlékezet nélküli csatorna köt össze. Vezessük be az $\eta = (\eta_1,\dots,\eta_n)$ és $\tilde\eta = (\tilde\eta_1,\dots,\tilde\eta_n)$ jelölést. A csatorna kódolási tétel megfordításának a bizonyításában az előbb bizonyított *felső becslés* eredményét fogom alkalmazni *egy λ megkülönböztethető elemeket tartalmazó halmaz elemszámáról* az $\eta,\tilde\eta$ valószínűségi változó párra és V^n csatornára a $\vartheta = Cn + K\frac{\sqrt{n}}{\sqrt{1-\lambda}}$ paraméterrel, ahol K egy a csatornától függő elég nagy szám.

Ezen becslés alapján a csatorna kódolási tétel megfordításának a bizonyításához elég megmutatni, hogy

$$P((\tilde\eta_1,\dots,\tilde\eta_n) \in A((v_{i_1},\dots,v_{i_n}))|\eta_1 = v_{i_1},\dots,\eta_n = v_{i_n}) \le \frac{1-\lambda}{2} \tag{3.5}$$

minden $(v_{i_1},\dots,v_{i_n}) \in V^n$ vektorra, ahol

$$A((v_{i_1},\dots,v_{i_n})) = \left\{ (\tilde v_{j_1},\dots,\tilde v_{j_n})\colon\, (\tilde v_{j_1},\dots,\tilde v_{j_n}) \in \tilde V^n,\, \frac{\prod_{k=1}^n p(\tilde v_{j_k}|v_{i_k})}{\prod_{k=1}^n q(\tilde v_{j_k})} \ge 2^{Cn+K\sqrt{n}/\sqrt{1-\lambda}} \right\},$$

egy elég nagy $K > 0$ konstanssal, mely képletben $p(\tilde{v}_j|v_i)$ jelöli az emlékezet nélküli csatornát meghatározó csatorna átmenetvalószínűségeit, és $q(\tilde{v}_j) = P(\tilde{\eta}' = \tilde{v}_j)$. Ugyanis, ha ez az egyenlőtlenség igaz, akkor az előbb említett felső becslés $\vartheta = Cn + K\frac{\sqrt{n}}{\sqrt{1-\lambda}}$ és $\gamma = \frac{1-\lambda}{2}$ szereposztással a tételben megfogalmazott eredményt szolgáltatja.

A bizonyítandó egyenlőtlenségnek megadjuk egy jobban vizsgálható, ekvivalens átfogalmazását. Ennek érdekében rögzítünk valamilyen $v_{i_1} \in V, \dots, v_{i_n} \in V$ pontokat, és olyan $\zeta_1(v_{i_1}), \dots, \zeta_n(v_{i_n})$ független valószínűségi változókat definiálunk, amelyekre a $\zeta_k(v_{i_k})$ valószínűségi változó eloszlását a $P(\zeta_k(v_{i_k}) = \tilde{v}_j) = p(\tilde{v}_j|v_{i_k})$, $\tilde{v}_j \in \tilde{V}$, képlet adja meg minden $1 \le k \le n$ indexre. Mivel $P((\zeta_1(v_{i_1}), \dots, \zeta_n(v_{i_n})) = (\tilde{v}_{j_1}, \dots, \tilde{v}_{j_k})) = P((\tilde{\eta}_1, \dots, \tilde{\eta}_n) = (\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n}))|\eta_1 = v_{i_1}, \dots, \eta_n = v_{i_n})$ minden $(\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n})$ vektorra ezzel a jelöléssel a (3.5) egyenlőtlenség a következő alakban írható:

$$P((\zeta_1(v_{i_1}), \dots, \zeta_n(v_{i_n})) \in A((v_{i_1}, \dots, v_{i_n})) \\ = P\left(\prod_{i=1}^{n} \frac{p(\zeta_k(v_{i_k})|v_{i_k})}{q(\zeta_k(v_{i_k}))} \ge 2^{Cn+K\sqrt{n}/\sqrt{1-\lambda}}\right) \le \frac{1-\lambda}{2}.$$

Az utolsó formulában a valószínűségen belül logaritmust véve azt kapjuk, hogy a

$$P\left(\sum_{i=1}^{n} \log \frac{p(\zeta_k(v_{i_k})|v_{i_k})}{q(\zeta_k(v_{i_k}))} \ge Cn + \frac{K\sqrt{n}}{\sqrt{1-\lambda}}\right) \le \frac{1-\lambda}{2} \tag{3.6}$$

egyenlőtlenséget kell bizonyítanunk minden $(v_{i_1}, \dots, v_{i_n}) \in V^n$ vektorra.

Vegyük észre, hogy a (3.6) formula szummájában szereplő tagok várható értéke

$$E \log \frac{p(\zeta_k(v_{i_k})|v_{i_k})}{q(\zeta_k(v_{i_k}))} = \sum_{\tilde{v}_j \in \tilde{V}} p(\tilde{v}_j|v_{i_k}) \log \frac{p(\tilde{v}_j|v_{i_k})}{q(\tilde{v}_j)}.$$

Továbbá, mivel $I(\eta' \wedge \tilde{\eta}') = C$, ahol C az emlékezet nélküli csatornát meghatározó csatorna csatorna kapacitása, ezt az egyenletet összehasonlítva a *lemma az emlékezet nélküli csatorna kapacitásáról* eredményében szereplő (3.1) képlettel azt kapjuk, hogy a tekintett várható érték kisebb vagy egyenlő, mint C, és a C számmal egyenlő azon $v_{i_k} \in V$ pontokra, amelyekre $P(\eta' = v_{i_k}) > 0$. Továbbá létezik olyan $D < \infty$ konstans, amelyre

$$E\left(\log \frac{p(\zeta_k(v_{i_k})|v_{i_k})}{q(\zeta_k(v_{i_k}))}\right)^2 = \sum_{\tilde{v}_j \in \tilde{V}} p(\tilde{v}_j|v_{i_k}) \left(\log \frac{p(\tilde{v}_j|v_{i_k})}{q(\tilde{v}_j)}\right)^2 \le D \text{ minden } v_i \in V \text{ pontra,}$$

mert a fenti kifejezésben egy olyan véges tagszámú összeg szerepel, amelynek mindegyik tagja véges. (Tudjuk, hogy ha $p(\tilde{v}_j|v_i) > 0$, akkor $q(\tilde{v}_j) > 0$, és csak véges sok különböző összeget kell tekintenünk.)

A most bizonyított összefüggések és a Csebisev egyenlőtlenség a következő becslést adják (3.6) képletben szereplő független valószínűségi változók összegének az eloszlására.

$$
\begin{aligned}
&P\left(\sum_{i=1}^{n}\log\frac{p(\zeta_k(v_{i_k})|v_{i_k})}{q(\zeta_k(v_{i_k}))}\ge Cn+\frac{K\sqrt{n}}{\sqrt{1-\lambda}}\right)\\
&\quad\le P\left(\sum_{i=1}^{n}\left(\log\frac{p(\zeta_k(v_{i_k})|v_{i_k})}{q(\zeta_k(v_{i_k}))}-E\log\frac{p(\zeta_k(v_{i_k})|v_{i_k})}{q(\zeta_k(v_{i_k}))}\right)\ge\frac{K\sqrt{n}}{\sqrt{1-\lambda}}\right)\\
&\quad\le\frac{Dn(1-\lambda)}{K^2n}\le\frac{1-\lambda}{2},
\end{aligned}
$$

ha a K konstanst elég nagynak ($K^2>2D$) választjuk. Így a (3.6) formulát, és ezzel a tételt is bebizonyítottuk.

Érdemes lehet heurisztikus szinten áttekinteni, hogy milyen gondolatokra épül a csatorna kódolási tételnek illetve e tétel megfordításának a bizonyítása.

Legyen $(\eta,\tilde\eta)$ olyan a csatornával összekapcsolt valószínűségi változó pár, amelyre $I(\eta\wedge\tilde\eta)=C$, ahol C a csatorna kapacitása, vagy ha ilyen pár nincs akkor az $I(\eta\wedge\tilde\eta)$ szám nagyon közel van ehhez a C értékhez. Annak bizonyítása, hogy alkalmas feltételek mellett mind a két tétel érvényes azon állítás igazolásán alapul, hogy egy n hosszúságú emlékezet nélküli csatornában az

$$
A(v)=\left\{\tilde v=(\tilde v_{j_1},\dots,\tilde v_{j_n})\in\tilde V_n\colon\ \prod_{k=1}^{n}\frac{p(\tilde v_{j_k}|v_{i_k})}{q(\tilde v_{j_k})}\sim 2^{Cn}\right\}
$$

halmaz teljesíti a $P((\tilde\eta_1,\dots,\tilde\eta_n)\in A(v)|(\eta_1,\dots,\eta_n)=v)\sim 1$ relációt tipikus $v=(v_{i_1},\dots,v_{i_n})\in V^n$ pontokban, ahol $(\eta_1,\tilde\eta_1)$, $\dots$, $(\eta_n,\tilde\eta_n)$ az $(\eta,\tilde\eta)$ párral azonos eloszlású, független véletlen vektorok sorozata, és $q(\tilde v_j)=P(\tilde\eta=\tilde v_j)$. (A csatorna kódolási tétel bizonyításában a $\sim 2^{Cn}$ kifejezés helyett $\ge 2^{Cn(1-o(1))}$-t, a csatorna kódolási tétel megfordításának a bizonyításában pedig $\le 2^{Cn(1+o(1))}$-et érdemes írni az $A(v)$ halmaz definiciójában.) A most használt 'tipikus' kifejezés első közelítésben azt jelenti, hogy a reláció igaz a $v=(v_{i_1},\dots,v_{i_n})\in V^n$ vektorok majdnem egy valószínűségi halmazára azon valószínűségi mérték szerint, amelyet az $(\eta_1,\dots,\eta_n)$ vektor eloszlása határoz meg. A bizonyítás részletesebb vizsgálata során a 'tipikus' szó jelentését pontosítani kell. Érdemes megjegyezni, hogy az $A(v)$ halmazra felírt reláció úgy is írható, hogy $\sum_{k=1}^{n}\log\frac{p(\tilde\eta_k|v_{i_k})}{q(\tilde\eta_k)}\sim Cn$ majdnem 1 valószínűséggel ha $(\tilde\eta_1,\dots,\tilde\eta_n)$ eloszlását a $P(\tilde\eta_1=\tilde v_{j_1},\dots,\tilde\eta_n=\tilde v_{j_n})=\prod_{k=1}^{n}p(v_{i_k})p(\tilde v_{j_k}|v_{i_k})$ képlet adja meg. Továbbá $E\log\frac{p(\tilde\eta_k|\eta_k)}{q(\tilde\eta_k)}=\sum_{i,j}P(\eta_k=v_i,\tilde\eta_k=\tilde v_j)\log\frac{P(\eta_k=v_i,\tilde\eta_k=\tilde v_j)}{P(\eta_k=v_i)P(\tilde\eta_k=\tilde v_j)}=I(\eta_k\wedge\tilde\eta_k)=C$. Ezért az előbb felírt reláció olyan tényt fejez ki, hogy független valószínűségi változók összege közel van az összeg várható értékéhez.

Az $A(v)$ halmazra megfogalmazott tulajdonság egyik következménye az, hogy egy tipikus $v \in V^n$ pontra

$$P((\tilde{\eta}_1, \dots, \tilde{\eta}_n) \in A(v)) = \sum_{\tilde{v} \in A(v)} q(\tilde{v}) = \sum_{\tilde{v} \in A(v)} q(\tilde{v}) \frac{p(\tilde{v}|v)}{q(\tilde{v})} \frac{q(\tilde{v})}{p(\tilde{v}|v)}$$

$$\sim 2^{-Cn} \sum_{\tilde{v} \in A(v)} q(\tilde{v}) \frac{p(\tilde{v}|v)}{q(\tilde{v})} = 2^{-Cn} \sum_{\tilde{v} \in A(v)} p(\tilde{v}|v) \sim 2^{-Cn},$$

ahol $q(\tilde{v}) = \prod_{k=1}^{n} q(\tilde{v}_{j_k})$, és $p(\tilde{v}|v) = \prod_{k=1}^{n} p(\tilde{v}_{j_k}|v_{i_k})$. Ha egy olyan $B(v) \subset \tilde{V}^n$ halmazt keresünk egy $v \in V^n$ ponthoz, amelyre $P((\tilde{\eta}_1, \dots, \tilde{\eta}_n) \in B(v)|(\eta_1, \dots, \eta_n) = v) \geq 1 - \lambda$ valamely kis $\lambda > 0$ számra, akkor a $B(v)$ halmaz az előbb definiált $A(v)$ halmaz kis módosítása a $P(\cdot|(\eta_1, \dots, \eta_n) = v)$ mérték szerint. Sőt, néhány számunkra nem lényeges feltétel teljesülése esetén az is igaz, hogy az $A(v)$ halmazhoz hasonlóan a $B(v)$ halmaz teljesíti a $P((\tilde{\eta}_1, \dots, \tilde{\eta}_n)) \in B(v)) \sim 2^{-nC}$ relációt. Ezért, ha olyan $(v^{(1)}, B^{(1)})$, $\dots$, $(v^{(N)}, B^{(N)})$, $v^{(k)} \in V^n$, $B^{(k)} \subset \tilde{V}^n$, minden $1 \leq k \leq N$ indexre, párokat keresünk, amelyekre $P((\tilde{\eta}_1, \dots, \tilde{\eta}_n) \in B^{(k)}|(\eta_1, \dots, \eta_n) = v^{(k)}) > 1 - \lambda$ minden k indexre, és a $B^{(k)}$ halmazok diszjunktak, akkor legfeljebb $N = 2^{Cn(1+o(1))}$ ilyen párt választhatunk, és ezt mondja ki a csatorna kódolási tétel megfordítása. A csatorna kódolási tétel viszont az állítja, hogy ennyi $(v^{(k)}, B^{(k)})$ párt ki is lehet választani.

Természetesen, az előbb vázolt gondolatmenetek pontos kidolgozása az érvelés finomítását igényli több ponton. Ezek részleteit azonban itt nem tárgyalom, mert a már leírt bizonyítás tartalmazza azokat. Csak egy figyelemre méltó részletet említek. A csatorna kódolási tétel megfordításában az $A(v)$ halmazokra felírt aszimptotikus relációt nem elegendő csak 'tipikus' $v \in V^n$ pontokra belátni, azokat minden $v \in V^n$ pontra igazolni kell. Ez okozza a fő nehézséget e tétel bizonyításában. E probléma leküzdésében a *véges állapotterű csatorna optimális bemenetének a jellemzéséről* szóló tétel eredménye segít. Ez teszi lehetővé a vizsgálandó feltételes valószínűség becslését minden $v \in V^n$ vektorra.

4. Kis hibájú és viszonylag gyors információ továbbítás a forrás és csatorna kódolási tétel segítségével.

E fejezetben a következő problémával fogunk foglalkozni. Legyen adva egy információ forrás, azaz legyen adva egy értékeit egy véges vagy megszámlálhatóan végtelen $X = \{x_1, x_2 \dots\}$ halmazon felvevő ξ valószínűségi változó, és legyen $\xi_1, \xi_2, \dots$ független, és a ξ valószínűségi változóval azonos eloszlású valószínűségi változók sorozata, amit információ forrásnak fogunk nevezni. Legyen ezenkívül adva egy emlékezet nélküli csatorna, amelyet egy olyan csatorna határoz meg, amely valamely $V = \{v_1, v_2, \dots\}$ bemeneti jelek halmazát átviszi kimeneti jelek valamely $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ halmazába, és $p(\tilde{v}_j|v_i)$ annak a feltételes valószínűsége, hogy a csatorna $\tilde{v}_j$ kimeneti jelet közli, feltéve, hogy a v_i bemeneti jelet adtuk le. A forrás jeleit akarjuk közölni a felhasználóval úgy, hogy az (emlékezet nélküli) csatornán leadjuk bemeneti jelek egy sorozatát, aminek hatására a felhasználó kimeneti jelek valamilyen sorozatát kapja, és ennek alapján próbálja rekonstruálni az információ forrás $\xi_1, \xi_2, \dots$ értékeit.

Tegyük fel, hogy az információ forrás $\xi_1, \xi_2, \dots$ jelei egységnyi sebességgel érkeznek, és mi is egységnyi sebességgel tudjuk továbbítani a v_i jeleket a csatornán keresztül. Olyan módszert szeretnénk kidolgozni, amely lehetővé teszi, hogy a felhasználó a forrás minden jelét ε-nál kisebb hibával rekonstruálni tudja, ahol $\varepsilon > 0$ egy előre rögzített nagyon kicsi szám. Emellett azt szeretnénk, hogy a felhasználó minden jelet annak megérkezése után véges időn belül megismerjen. Pontosabban megfogalmazva azt követeljük meg, hogy bármilyen nagy n számra a felhasználó az n-ik időpontban ismerje az összes $1 \le j \le n - K$ időintervallumban leadott ξ_j jelet, ahol K egy rögzített szám, amely függhet az ε hibakorláttól, de nem függ az n időponttól.

A következő két a forrás és csatorna kódolásról szóló eredményeken alapuló tételben azt mutatom meg, hogy ilyen információ továbbítás lehetséges akkor, ha a csatorna kapacitása nagyobb, mint a forrás entrópiája, de nem lehetséges akkor, ha a forrás entrópiája nagyobb, mint a csatorna kapacitása.

Tétel a jó információ továbbítás lehetőségéről, ha a csatorna kapacitása nagyobb, mint a forrás entrópiája. *Legyen adva egy értékeit egy véges vagy megszámlálhatóan végtelen $X = \{x_1, x_2 \dots\}$ halmazon felvevő ξ valószínűségi változó és e ξ valószínűségi változóval azonos eloszlású, független $\xi_1, \xi_2, \dots$ valószínűségi változók egy sorozata. Legyen ezenkívül adva egy emlékezet nélküli csatorna, amelyet egy olyan csatorna határoz meg, amely valamely $V = \{v_1, v_2, \dots\}$ bemeneti jelek halmazát átviszi kimeneti jelek valamely $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ halmazába, és $p(\tilde{v}_j|v_i)$ annak a feltételes valószínűsége, hogy a csatorna a kimeneti oldalon a $\tilde{v}_j$ jelet közli, feltéve, hogy v_i volt a bemeneti jel. Tegyük fel, hogy a csatorna C csatorna kapacitása nagyobb, mint a ξ valószínűségi változó $H(\xi)$ entrópiája. Egy rögzített $\varepsilon > 0$ számra alkalmazzuk a következő módszert annak érdekében, hogy megismertessük a $\xi_1, \xi_2, \dots$ sorozat jeleit a felhasználóval.*

Az $\varepsilon > 0$ számhoz válasszunk először egy $n_0 = n_0(\varepsilon)$ küszöbindexet, majd definiáljuk az emlékezet nélküli csatorna n_0 hosszúságú bemeneti jeleiből álló V^{n_0} halmaznak egy $A = A(n_0) = \{v^{(n_0)}(l) = (v_1(l), \dots, v_{n_0}(l)),\ 1 \le l \le N(n_0)\} \subset V^{n_0}$ részhalmazát valamely $N = N(n_0)$ elemszámmal, és minden $v^{(n_0)}(l) \in A(n_0)$ vektorhoz adjunk meg az n_0 hosszúságú kimeneti jelek egy egy alkalmas $B_l \subset \tilde{V}^{n_0}$ részhalmazát úgy, hogy ezek a B_l, $1 \le l \le N(n_0)$, halmazok a $\tilde{V}^{n_0}$ halmaz egy particióját adják. Definiáljunk ezenkívül egy $f\colon X^{n_0} \to A(n_0)$ függvényt, amelyet kódoló, és egy $g\colon A(n_0) \to X^{n_0}$ függvényt, amelyet dekódoló függvénynek fogunk nevezni.

Tekintsük a $\xi_1, \xi_2, \dots$ sorozat egymást követő diszjunkt n_0 hosszúságú $\xi_{ln_0+1}, \dots, \xi_{(l+1)n_0}$ blokkjait minden $l = 0, 1, 2, \dots$ számra. Helyettesítsük be az l-ik blokk értékeit az $f(\cdot)$ kódfüggvénybe, azaz tekintsük a (véletlen) $f(\xi_{ln_0+1}, \dots, \xi_{(l+1)n_0}) \in A(n_0)$ sorozatot minden $l = 1, 2, \dots$ indexre, és küldjük a felhasználónak a csatornán keresztül ezt a sorozatot. Ő végezze a kapott (véletlen) $(\tilde{v}_{j_1}, \dots, \tilde{v}_{j_{n_0}}) \in \tilde{V}^{n_0}$ sorozat dekódolását a következő módon. Válassza ki azt a $B_l \in \tilde{V}^{n_0}$ halmazt, amelyre $(\tilde{v}_{j_1}, \dots, \tilde{v}_{j_{n_0}}) \in B_l$, és vegye a neki megfelelő $v_l^{(n_0)} = (v_{i_1}(l), \dots, v_{i_{n_0}}(l)) \in A(n_0)$ sorozatot. Alkalmazza erre a sorozatra a $g(\cdot)$ dekódoló függvényt, azaz vegye a $g(v_{i_1}(l), \dots, v_{i_{n_0}}(l)) \in X^{n_0}$ sorozatot, és válassza ezt a $\xi_{ln_0+1}, \dots, \xi_{(l+1)n_0}$ sorozatnak.

Az n_0 küszöbindexet, az $A = A(n_0) \subset V^{n_0}$ részhalmazt, a $\tilde{V}^{n_0}$ halmaznak a $v^{(n_0)}(l) \in A(n_0)$ vektoroknak megfelelő B_l, $1 \le l \le N(n_0)$, particióját, valamint az $f\colon X^{n_0} \to A(n_0)$ kódoló és a $g\colon A(n_0) \to X^{n_0}$ dekódoló függvényt alkalmasan választva elérhetjük, hogy a felhasználó ezen eljárás segítségével legalább $1-\varepsilon$ valószínűséggel a forrás által leadott $\xi_{ln_0+1}, \dots, \xi_{(l+1)n_0}$ sorozatot válassza a leadott sorozat l-ik blokkjának, azaz az l-ik blokkot legalább $1-\varepsilon$ valószínűséggel jól dekódolja.

A tétel bizonyítása. Ha a csatorna kapacitása nagyobb, mint a forrás entrópiája, akkor létezik olyan $\delta > 0$ szám, amelyre $C > H(\xi) + 2\delta$. Válasszunk egy ilyen $\delta > 0$ számot. Ekkor a csatorna kódolási tétel alapján van olyan $n_0 = n_0(\varepsilon, \delta)$ küszöbindex, hogy minden $n \ge n_0$ indexre létezik egy $2^{(C-\delta)n} \ge N(n) \ge 2^{(H(\xi)+\delta)n}$ elemszámú $A(n) \subset V^n$ halmaz, valamint a $\tilde{V}^n$ halmaznak egy olyan $B_1, \dots, B_{N(n)}$ particiója, amely teljesíti a $P(B_l|v^{(n)}(l)) = \sum_{\tilde{v}^{(n)} \in B_l} p(\tilde{v}^{(n)}|(v^{(n)}(l)) \ge 1 - \frac{\varepsilon}{2}$ egyenlőtlenséget a $v^{(n)}(l) = (v_{i_1}(l), \dots, v_{i_n}(l)) \in A(n)$ vektorra és a neki megfelelő $B_l \subset \tilde{V}^n$ halmazra minden $1 \le l \le N(n_0)$ indexre, ahol $p(\tilde{v}^{(n)}|(v^{(n)}(l))$ az emlékezet nélküli csatorna átmenetvalószínűsége, azaz $p(\tilde{v}^{(n)}|(v^{(n)}(l)) = \prod_{k=1}^{n} p(\tilde{v}_{j_k}|v_{i_k}(l))$ a $\tilde{v}^{(n)} = (\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n})$ és $v^{(n)}(l) = (v_{i_1}(l), \dots, v_{i_n}(l))$ jelöléssel. Továbbá, mivel $N(n) \ge 2^{(H(\xi)+\delta)n}$, a második fejezetben bizonyított *független, egyforma eloszlású valószínűségi változókból álló forrás kis hibájú kódolásáról és dekódolásáról* szóló tétel alapján van olyan $f\colon X^n \to A(n)$, kódoló és $g\colon A(n) \to X^n$ dekódoló függvény, amelyre $P(g(f(\xi_1, \dots, \xi_n)) = (\xi_1, \dots, \xi_n)) \ge 1 - \frac{\varepsilon}{2}$, ha $n \ge n_0(\varepsilon, \delta)$ egy esetleg nagyobb n_0 küszöbindexszel.

Válasszunk egy olyan n_0 indexet, amelyre létezik a kívánt tulajdonságú $A(n_0) \subset V^{n_0}$ halmaz a $\tilde{V}^{n_0}$ halmaz hozzátartozó $B_1, \dots, B_{N(n_0)}$ particiójával együtt, valamint létezik a megfelelő tulajdonságú f kódoló és g dekódoló függvény is. Ezzel a választással a tételben leírt dekódolási eljárás hibája kisebb, mint ε. Valóban, tekintsük az információ forrás által leadott $\xi_{ln_0+1}, \dots, \xi_{(l+1)n_0}$ sorozat $f(\xi_{ln_0+1}, \dots, \xi_{(l+1)n_0}) = v_l^{n_0} \in A(n_0)$ kódját. A felhasználó legalább $1-\frac{\varepsilon}{2}$ valószínűséggel azonosítani fogja ezt a jelet a csatornán keresztül kapott jel segítségével. Ezután a $g(\cdot)$ dekódoló függvény segítségével végzett dekódolás is legfeljebb $\frac{\varepsilon}{2}$ valószínűséggel fogja döntése hibáját növelni.

A második tételt, amely azt állítja, hogy ha $C < H(\xi)$ akkor nem lehetséges az adott módon kis hibájú, gyors adatátvitelt bizosítani csak abban az esetben bizonyítom, ha a tekintett csatorna véges állapotterű, mert csak ebben az esetben bizonyítottam a csatorna kódolási tétel megfordítását, amely fontos szerepet játszik ezen állítás igazolásában. Olyan állítást fogok bizonyítani, amely szerint a $C < H(\xi)$ esetben minden $1 > \varepsilon > 0$ számhoz megadható egy olyan $n_0 = n_0(\varepsilon)$ küszöbindex, hogy egy $n \ge n_0$ hosszúságú blokknak tetszőleges a csatornán kereszül történő továbbításának a hibája legalább $1-\varepsilon$. Azután egy következményben megmutatom, hogy a viszonylag rövid blokkoknak is bármely a csatornán keresztül történő továbbításának a hibája alulról becsülhető egy a forrástól és a csatornától függő pozitív számmal.

A tétel pontos megfogalmazása érdekében először azt definiálom, hogy mit jelent egy n hosszúságú sorozat továbbítása a csatornán keresztül. Az egyszerűbb jelölés érdekében csak a $\xi_1, \dots, \xi_n$ sorozat csatornán keresztül történő továbbításáról fogok

beszélni, bár a fogalmat hasonlóan definiálhatnánk és a megfelelő eredményt hasonlóan bizonyíthatnánk tetszőleges $\xi_{l+1}, \dots, \xi_{l+n}$, $l = 0, 1, 2, \dots$, sorozatra is.

Az $\xi_1, \dots, \xi_n$ sorozat egy a csatornán keresztül történő továbbítását a következő mennyiségek segítségével fogjuk definiálni. Vezessünk be egy $f\colon X^n \to V^n$ kódfüggvényt, amely az n hosszúságú $(x_1, \dots, x_n) \in X^n$ sorozatokat képezi a csatorna bemeneti jeleinek $(v_1, \dots, v_n) \in V^n$ sorozataiba. Definiáljuk a csatorna $\tilde{V}$ kimeneti jeleinek n hosszúságú sorozataiból álló $\tilde{V}^n$ halmaznak egy $N = N(n)$ elemű $B_1, \dots, B_N$ particióját. Ezenkívül rendeljük hozzá e partició mindegyik B_l, $1 \le l \le N$, eleméhez az X^n halmaz valamely $x^{(n)}(l) = (x_{i_1}(l), \dots, x_{i_n}(l)) \in X^n$ elemét úgy, hogy a partició különböző elemeihez különböző sorozatot rendelünk hozzá, azaz $x^{(n)}(l) \neq x^{(n)}(l')$, ha $l \neq l'$. Alkalmazzuk a következő információ továbbítási és dekódolási eljárást. Ha megérkezik a forrásból a $\xi_1, \dots, \xi_n$ sorozat, akkor alkalmazzuk rá az f kódfüggvényt. Így egy $f(\xi_1, \dots, \xi_n) = (v_{i_1}, \dots, v_{i_n}) \in V^n$ sorozatot kapunk. Ezt átengedjük az emlékezet nélküli csatornán és kapunk egy $(\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n}) \in \tilde{V}^n$ sorozatot, amelyet tartalmaz a $B_1, \dots, B_N$ partició egyik eleme. Ha $(\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n}) \in B_l$, $1 \le l \le N$, akkor legyen a dekódolt sorozat a B_l halmaznak megfeleltetett $x^{(n)}(l) = (x_{i_1}(l), \dots, x_{i_n}(l))$ sorozat. Jelöljük ezt az $x^{(n)}(l)$ 'dekódoló' (véletlen) sorozatot $(\zeta_1, \dots, \zeta_n)$-nel. Akkor tekintjük az információ továbbítási és dekódolási eljárást jónak, ha $(\xi_1, \dots, \xi_n) = (\zeta_1, \dots, \zeta_n)$.

A következő tételt fogom bizonyítani.

Tétel a jó információ továbbítás lehetőségeinek a korlátairól, ha a csatorna kapacitása kisebb, mint a forrás entrópiája. *Legyen adva egy értékeit egy véges vagy megszámlálhatóan végtelen $X = \{x_1, x_2 \dots\}$ halmazon felvevő ξ valószínűségi változó és e ξ valószínűségi változóval azonos eloszlású, független $\xi_1, \xi_2, \dots$ valószínűségi változók egy sorozata. Legyen ezenkívül adva egy emlékezet nélküli csatorna, amelyet egy olyan csatorna határoz meg, amely valamely $V = \{v_1, v_2, \dots\}$ bemeneti jelek véges halmazát átviszi kimeneti jelek valamely $\tilde{V} = \{\tilde{v}_1, \tilde{v}_2, \dots\}$ véges halmazába, és $p(\tilde{v}_j|v_i)$ annak a feltételes valószínűsége, hogy a csatorna kimeneti jele a $\tilde{v}_j$ pont, feltéve, hogy a bemeneti jele v_i volt. Tegyük fel, hogy a csatorna C csatorna kapacitása kisebb, mint a ξ valószínűségi változó $H(\xi)$ entrópiája. Ekkor minden rögzített $\varepsilon > 0$ számhoz létezik olyan $n_0 = n_0(\varepsilon)$ küszöbindex, amelyre igaz a következő állítás.*

Adva egy $n \ge n_0$ szám, tekintsünk egy $f\colon X^n \to V^n$ kódfüggvényt, és vegyük a $\tilde{V}^n$ halmaznak egy $N = N(n)$ elemű $B_1, \dots, B_N$ particióját. Ezenkívül rendeljük hozzá e partició mindegyik B_l, $1 \le l \le N$, eleméhez az X^n halmaz egyik $x^{(n)}(l) = (x_{i_1}(l), \dots, x_{i_n}(l)) \in X^n$ elemét úgy, hogy $x^{(n)}(l) \neq x^{(n)}(l')$, ha $l \neq l'$. Alkalmazzuk az ezen f kódfüggvény, $B_1, \dots, B_N$ partició és $B_l \to X^{(n)}$ megfeleltetés által meghatározott az e tétel előtt leírt információ továbbítási és dekódolási eljárást. Ezen információ továbbítási és dekódolási eljárás hibája legalább $1 - \varepsilon$, ha $n \ge n_0 = n_0(\varepsilon)$, ami azt jelenti, hogy $P((\xi_1, \dots, \xi_n) = (\zeta_1, \dots, \zeta_n)) \le \varepsilon$.

A tétel bizonyítása. Azt állítom, hogy ha $H(\xi) > C$, akkor minden $\varepsilon > 0$ számhoz létezik olyan $n_0 = n_0(\varepsilon)$ küszöbindex, hogy ha az előbb leírt eljárást alkalmazzuk $n \ge n_0$ hosszúságú sorozatokra, akkor bárhogy is választjuk az $f(\cdot)$ kódoló függvényt, a $\tilde{V}^n$ halmaz $B_1, \dots, B_N$ particióját és bárhogy is adjuk meg a B_l halmaz elemeinek a $B_l \to x^{(n)}(l) = (x_{i_1}(l), \dots, x_{i_n}(l))$ hozzárendelését az X^n halmaz valamely eleméhez,

a dekódolás legfeljebb ε valószínűséggel ad helyes eredményt, azaz $P((\xi_1,\dots,\xi_n) = (\zeta_1,\dots,\zeta_n)) \le \varepsilon$.

A bizonyítást indirekt módon végzem el. Feltételezem, hogy van olyan $n \ge n_0$ szám bármely n_0 küszöbindexre, amelyre létezik olyan $f(\cdot)$ kódfüggvény, a $\tilde{V}^n$ halmaz olyan $B_1,\dots,B_N$ particiója, illetve e partícióknak olyan $B_l \to x^{(n)}(l) = (x_{i_1}(l),\dots,x_{i_n}(l))$ megfeleltetése, amelyre a tétel megfogalmazása előtt leírt módon konstruált $(\zeta_1,\dots,\zeta_n)$ sorozatra $P((\xi_1,\dots,\xi_n) = (\zeta_1,\dots,\zeta_n)) \ge \varepsilon$. Megmutatom, hogy e feltevésből következik egy olyan állítás, ami ellentmond a csatorna kódolási tétel megfordításában megfogalmazott eredménynek. Olyan nagy N számra fogok N darab $v^{(n)}(l) \in V$ pontot és diszjunkt, a $p(B_{(u(l)}|v^{(n)}(l)) \ge \varepsilon$ tulajdonságot teljesítő $B_{u(l)} \in \tilde{V}$, $1 \le l \le N$, halmazt konstruálni, amekkora N szám esetén ez az említett eredmény szerint ez nem lehetséges.

Mivel $H(\xi) > C$, választhatunk olyan $\delta > 0$ számot, amelyre $H(\xi)(1-\delta)^4 > C$. Először azt mutatom meg, hogy feltevésünkből következik, hogy ha $n \ge n_0 = n_0(\varepsilon,\delta)$ valamely n_0 küszöbindexszel akkor létezik $N_1 = N_1(n) \ge 2^{(1-\delta)^2 H(\xi)n}$ darab olyan $x^{(n)}(l) = (x_{i_1}(l),\dots,x_{i_n}(l)) \in X^n$ sorozat, amelyekre

$$P((\xi_1,\dots,\xi_n) = (\zeta_1,\dots,\zeta_n)|(\xi_1,\dots,\xi_n) = x^{(n)}(l)) \ge \frac{\varepsilon}{2} \quad \text{minden } 1 \le l \le N_1 \text{ indexre.}$$

Ezt igazolandó, vegyük észre, hogy ha definiáljuk az

$$\begin{aligned} A = A(n) = \{&(x_{i_1},\dots,x_{i_n})\colon (x_{i_1},\dots,x_{i_n}) \in X^n, \\ &P((\xi_1,\dots,\xi_n) = (\zeta_1,\dots,\zeta_n)|(\xi_1,\dots,\xi_n) = (x_{i_1},\dots,x_{i_n})) \ge \tfrac{\varepsilon}{2}\} \end{aligned} \tag{4.1}$$

halmazt, akkor $P((\xi_1,\dots,\xi_n) \in A) \ge \frac{\varepsilon}{2}$. Ugyanis feltevésünk szerint

$$\varepsilon \le P((\xi_1,\dots,\xi_n) = (\zeta_1,\dots,\zeta_n)) \le P((\xi_1,\dots,\xi_n) \in A) + \frac{\varepsilon}{2}(1 - P((\xi_1,\dots,\xi_n) \in A)),$$

ahonnan következik ez az állítás. Viszont innen az is következik, hogy ha $n \ge n_0$ egy elég nagy n_0 számmal, akkor az $A(n)$ halmaz elemszáma nagyobb, mint $2^{(1-\delta)^2 H(\xi)n}$. Valóban, vezessük be az

$$\begin{aligned} A_1 = A_1(n) = \{&(x_{i_1},\dots,x_{i_n})\colon (x_{i_1},\dots,x_{i_n}) \in X^n, \\ &P((\xi_1,\dots,\xi_n) = (x_{i_1},\dots,x_{i_n})) < 2^{-(1-\delta)H(\xi)n}\} \end{aligned}$$

halmazt. Láttuk korábban, hogy $P((\xi_1,\dots,\xi_n) \in A_1) \ge 1 - \frac{\varepsilon}{4}$. Ezért $P((\xi_1,\dots,\xi_n) \in A \cap A_1) \ge \frac{\varepsilon}{4}$, ahonnan az $A \cap A_1$, következésképpen az $A = A(n)$ halmaz elemszáma nagyobb, mint $\frac{\varepsilon}{4}2^{(1-\delta)H(\xi)n} \ge 2^{(1-\delta)^2 H(\xi)n}$, és valójában az A halmaz elemszámára ilyen alsó becslést kívántunk adni.

Tekintsük a (4.1) formulában definiált $A(n)$ halmazt, és soroljuk fel az elemeit $A(n) = \{x^{(n)}(l),\ 1 \le l \le N(n)\}$ alakban. Láttuk, hogy $N(n) \ge 2^{(1-\delta)^2 H(\xi)n}$. Feleltessünk meg mindegyik $x^{(n)}(l) \in A(n)$ vektornak azt a $(v^{(n)}(l), B_{u(l)})$ párt, amelyre $v^{(n)}(l) = f(x^{(n)}(l)) \in V^n$ a tekintett modellben szereplő $f(\cdot)$ kódfüggvénnyel, és

$B_{u(l)} \subset \tilde{V}^{(n)}$ a $\tilde{V}^{(n)}$ halmaz $B_1, \dots, B_N$ particiójának az a B_r eleme, amelyre a tekintett modellben adott megfeleltetésben a $B_r \to x^{(n)}(l)$ reláció teljesül. (Létezik egy ilyen B_r halmaz a $P((\zeta_1, \dots, \zeta_n) = x^{(n)}(l) | (\xi_1, \dots, \xi_n) = x^{(n)}(l)) > 0$ tulajdonság miatt.) Vegyük észre, hogy

$$p(B_{u(l)} | v^{(n)}(l)) = \sum_{\tilde{v}^{(n)} \in B_{u(l)}} p(\tilde{v}^{(n)} | v^{(n)}(l)) \ge \frac{\varepsilon}{2},$$

ahol $p(\tilde{v}^{(n)} | v^{(n)}(l))$ az emlékezet nélküli csatorna átmenetvalószínűsége, azaz

$$p(\tilde{v}^{(n)} | v^{(n)}(l)) = \prod_{k=1}^{n} p(\tilde{v}_{j_k} | v_{i_k}(l)),$$

ha $\tilde{v}^{(n)} = (\tilde{v}_{j_1}, \dots, \tilde{v}_{j_n})$, és $v^{(n)}(l) = (v_{i_1}(l), \dots, v_{i_n}(l))$. A felírt egyenlőtlenség azért igaz, mert annak a valószínűségét tekintettük, hogy ha egy $x^{(n)}(l) \in A(n)$ sorozatot veszünk, tekintjük annak az $f(\cdot)$ leképezés szerinti $v^{(n)}(l) = f(x^{(n)}(l))$ képét, azt átengedjük a csatornán, majd a kapott jelet az általunk leírt módon dekódoljuk, akkor a kapott $(\zeta_1, \dots, \zeta_n)$ sorozat teljesíti a $(\zeta_1, \dots, \zeta_n) = x^{(n)}(l)$ azonosságot. Ennek valószínűsége pedig legalább $\frac{\varepsilon}{2}$. Vegyük észre azt is, hogy bár lehetséges, hogy $v^{(n)}(l) = v^{(n)}(l')$ akkor is, ha $l \neq l'$, azaz lehet két különböző $x^{(n)}(l) \in A(n)$ és $x^{(n)}(l') \in A(n)$ vektor, amelyekre $l \neq l'$, és $f(x^{(n)}(l)) = f(x^{(n)}(l'))$, viszont bármely $v^{(n)} \in V^n$ vektorra az $f(x^{(n)}(l)) = v^{(n)}$ reláció legfeljebb $\frac{2}{\varepsilon}$ különböző l indexre állhat fenn. Valóban, mivel a $B_{u(l)}$ halmazok diszjunktak különböző l indexekre, ezért

$$\sum_{l: f(x^{(n)}(l)) = v^{(n)}} p(B_{u(l)} | v^{(n)}(l)) = p\left(\bigcup_{l: f(x^{(n)}(l)) = v^{(n)}} B_{u(l)} \,\middle|\, v^{(n)} \right) \le 1,$$

és az összeg mindegyik tagjának az értéke legalább $\frac{\varepsilon}{2}$. Ezért igaz ez az állítás.

Tekintsük a $C = C(n) = \{v^{(n)}(l) = f(x^{(n)}(l))\colon\ x^{(n)}(l) \in A(n)\}$ halmazt, ahol $v^{(n)}(l) = v^{(n)}(l')$ esetén e két vektor közül csak az egyiket soroljuk fel a $C = C(n)$ halmaz definiciójában. Társítsuk mindegyik $v^{(n)}(l) \in C$ vektorhoz a neki megfelelő $B_{u(l)}$ halmazt a $\tilde{V}^n$ halmaz $B_1, \dots, B_N$ particiójából. Láttuk, hogy $p(B_{u(l)} | v^{(n)}(l)) \ge \frac{\varepsilon}{2}$, ami azt jelenti, hogy a C halmaz elemei a csatornára nézve $\frac{\varepsilon}{2}$ megkülönböztethető elemek. Másrészt azt is láttuk, hogy a C halmaz elemszáma nagyobb, mint $\frac{\varepsilon}{2} 2^{(1-\delta)^2 H(\xi) n} \ge 2^{(1-\delta)^3 H(\xi) n} \ge 2^{Cn/(1-\delta)}$, ha $n \ge n_0$. Ez viszont ellentmond a a csatorna kódolási tétel megfordításának. Ezért ilyen tulajdonságú legalább ε pontosságú információ továbbítás és dekódolás nem létezhet, ha $n \ge n_0(\varepsilon)$ egy elég nagy n_0 számmal.

Következmény. *Tekintsük azt az esetet, amikor teljesüljenek a jó információ továbbítás lehetőségeinek a korlátairól szóló tétel feltételei, speciálisan $H(\xi) > C$. Rögzítsünk egy tetszőleges $n \ge 1$ számot, definiáljunk egy $f\colon X^n \to V^n$ kódfüggvényt, és vegyük a $\tilde{V}^n$ halmaznak egy $N = N(n)$ elemű $B_1, \dots, B_N$ particióját. Ezenkívül rendeljük*

hozzá e partició mindegyik B_l, $1 \le l \le N$, eleméhez az X^n halmaz egyik $x^{(n)}(l) = (x_{i_1}(l), \dots, x_{i_n}(l)) \in X^n$ elemét úgy, hogy $x^{(n)}(l) \neq x^{(n)}(l')$, ha $l \neq l'$. Alkalmazzuk azt az ezen f kódfüggvény, $B_1, \dots, B_N$ partició és $B_l \to X^{(n)}$ megfeleltetés által meghatározott információ továbbítási és dekódolási eljárást, amelyet az előbbi tétel megfogalmazása előtt vezettem be. Létezik olyan a forrástól és csatornától függő, de az n számtól független $\alpha > 0$ szám, hogy ezen információ továbbítási és dekódolási eljárás hibája legalább α, azaz $P((\xi_1, \dots, \xi_n) = (\zeta_1, \dots, \zeta_n)) \le 1 - \alpha$.

A következmény indoklása. Azt kell megindokolni, hogy a $H(\xi) > C$ esetben kis n számokra sem lehet n hosszúságú blokkok segítségével nagyon jó információ továbbítást elérni. Beláttuk, hogy ha $H(\xi) > C$, akkor minden $\varepsilon > 0$ számhoz létezik olyan $n_0 = n_0(\varepsilon)$ küszöbindex, hogy az $n \ge n_0$ számokra minden n hosszúságú blokkokon alapuló információ továbbítás és dekódolás hibája legalább ε. Ezt az eredményt fogjuk alkalmazni $\varepsilon = \frac{1}{2}$ választással. Tekintsük az $n_0 = n_0(\frac{1}{2})$ küszöbindexet. Azt állítom, hogy az $n < n_0$ hosszúságú blokkokon alapuló információ továbbítás és dekódolás hibája nagyobb vagy egyenlő, mint $\frac{1}{2n_0}$. Innen adódik a Következmény állítása.

A bizonyítás alapgondolata az, hogy ha létezne olyan módszer, amely $\frac{1}{2n_0}$-nél kisebb dekódolási hibát biztosít, akkor ezt alkalmazva n_0 egymás utáni blokkra olyan információ továbbítási és dekódolási eljárást kapnánk valamely n_0-nál hosszabb blokkra, amelynek a hibája kisebb, mint $\frac{1}{2}$. Viszont tudjuk, hogy ez nem lehetséges.

Valóban, rögzítsünk egy $n < n_0$ számot. Egy n hosszúságú blokkokon alapuló információ továbbítást és dekódolást egy $f\colon X^n \to V^n$ kódoló függvény, a $\tilde{V}^n$ halmaz egy $B_1, \dots, B_N$ particiója valamint e partició elemeinek egy $B_l \to x^{(n)}(l)$, $1 \le l \le N$, leképezése az X^n térbe határoz meg. Definiáljunk e mennyiségeknek megfelelő objektumokat az $n_0 n$ hosszúságú sorozatok terén a következő módon. Definiáljuk az $\bar{f}$ kódolási függvényt, amely az $X^{n_0 n}$ teret a $V^{n_0 n}$ térbe képezi az $\bar{f}(x_1, \dots, x_{n_0 n}) = (f(x_{kn+1}, \dots, x_{(k+1)n}),\ k = 0, \dots, n_0 - 1)$ képlet segítségével, a $\tilde{V}^{n_0 n}$ halmaz N^{n_0} elemű particióját pedig a következő módon: E partició elemei a $B(l_{i(1)}, \dots, l_{i(n_0)}) = B_{l_{i(1)}} \times \dots \times B_{l_{i(n_0)}}$ halmazok, ahol $1 \le i(j) \le N$ minden $1 \le j \le n_0$ indexre. Végül e partició $B(l_{i(1)}, \dots, l_{i(n_0)})$ elemének a $x^{(n)}(l_{i(1)}) \times \dots \times x^{(n)}(l_{i(n_0)}) \in X^{n_0 n}$ vektort feleltetjük meg.

Nem nehéz belátni, hogy ha az eredeti n hosszúságú blokkokon alapuló információ továbbításban és dekódolásában a $\xi_{ln+1}, \dots, \xi_{(l+1)n}$, $0 \le l < n_0$, blokkok hibás dekódolásának a valószínűsége kisebb, mint $\frac{1}{2n_0}$, (a dekódolás hibája nem függ az l számtól), akkor az $n_0 n$ hosszúságú sorozatok 'szorzatterében' az új objektumok által meghatározott információ továbbítás és dekódolás hibájának a valószínűsége kisebb, mint $n_0 \frac{1}{2n_0} = \frac{1}{2}$. Ugyanis a 'szorzattérben' az új dekódolás valójában úgy működik, hogy az egyes $kn+1 \le \bar{n} \le (k+1)n$ blokkokat, $k = 0, \dots, n_0 - 1$, egymástól függetlenül az n hosszúsági blokkokon érvényes szabály szerint továbbítjuk a csatornán keresztül, majd dekódoljuk őket. Ha ezen n_0 blokk dekódolása mindegyik k-ra kevesebb, mint $\frac{1}{2n_0}$ valószínűséggel hibás, akkor igaz az említett becslés. De mivel az $n_0 n$ hosszú sorozatokkal végzett dekódolások hibája legalább $\frac{1}{2}$, innen következik az $n < n_0$ hosszú sorozatok hibájáról megfogalmazott állítás.

5. Az entrópia fogalmának Kolmogorov-féle általánosítása és e fogalom alkalmazása egy probléma vizsgálatában.

Ebben a fejezetben egy olyan problémát fogok tárgyalni, amelynek látszólag nincs köze az információelmélethez. Mégis, meglepő módon, e probléma megoldásában kulcsszerepet játszik az entrópia, pontosabban e fogalom egy alkalmas általánosítása. A vizsgálandó kérdés megfogalmazásának érdekében először felidézem a valószínűségszámításban gyakran használt Bernoulli rendszer definicióját.

A Bernoulli rendszer definiciójának megadása előtt ismertetem annak informális leírását. Veszünk egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt, és azon egy véges sok, mondjuk az $1, 2, \dots, r$ értékeket felvevő ξ valószínűségi változót. Minden egész l számra tekintjük ennek a rendszernek egy ezzel az l számmal indexelt példányát, és vesszük ezek direkt szorzatát. Ezután definiáljuk azt az eltolást ezen a szorzattéren, amelynek hatására a ξ_l valószínűségi változó a ξ_{l+1} változóba megy át. Alább egy olyan rendszert definiálunk, ahol ilyen valószínűségi változókat és azok eltoltjait természetes módon be lehet vezetni.

Bernoulli rendszer definiciója. *Legyen adva egy $r \geq 2$ egész szám, és olyan $p_j \geq 0$, $1 \leq j \leq r$, számok, amelyekre $\sum_{j=1}^{r} p_j = 1$. Az $r \geq 2$, és p_j, $1 \leq j \leq r$, számok által meghatározott Bernoulli rendszeren az alábbi $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt és az Ω halmazon definiált T úgynevezett shift (eltolás) transzformációt értjük. Az Ω halmaz elemei azon $\omega = (\dots, x_{-1}, x_0, x_1, \dots)$ sorozatok, amelyekre $x_j \in \{1, \dots, r\}$, minden $-\infty < j < \infty$ indexre. Az $\mathcal{A}$ σ-algebra az alábbi $A(k, j_{-k}, \dots, j_k) \subset \Omega$ úgynevezett hengerhalmazok által generált legszűkebb σ-algebra:*

$$A(k, j_{-k}, \dots, j_k) = \{\omega = (\dots, x_{-1}, x_0, x_1, \dots) \colon x_s = j_s, -k \leq s \leq k\},$$

ahol k tetszőleges pozitív egész szám, és $j_s \in \{1, \dots, r\}$ minden $-k \leq s \leq k$ indexre. A hengerhalmazok P valószínűségét a $P(A(k, j_{-k}, \dots, j_k) = \prod_{s=-k}^{k} p_{j_s}$ képlet adja meg, és a P mérték e valószínűség kiterjesztése a $\mathcal{A}$ σ-algebrára. Végül egy

$$\omega = (\dots, x_{-2}, x_{-1}, x_0, x_1, \dots) \in \Omega$$

elemi esemény $T\omega$ shiftje (eltoltja) a

$$T\omega = (\dots, x_{-1}, x_0, x_1, x_2 \dots) \in \Omega$$

sorozat, azaz az ω-t definiáló sorozat x_s, s-ik koordinátáját eggyel eltoljuk balra. Ez azt jelenti az x_s szám a $T\omega$-t definiáló sorozat $s-1$-ik koordinátájában jelenik meg.

Megjegyzés. A Bernoulli rendszerek definiciójában nem jelentek meg az e fogalom informális ismertetésében említett ξ_l, $l = 0, \pm 1, \dots$, független és egyforma eloszlású valószínűségi változók. De ilyen valószínűségi változókat egyszerű és természetes módon definiálhatunk egy Bernoulli rendszerben. Nevezetesen, legyen $\xi_l(\omega) = x_l$, $l = 0, \pm 1, \dots$, ha $\omega = (\dots, x_{-1}, x_0, x_1, \dots)$.

Azzal a kérdéssel fogunk foglalkozni, hogy két különböző Bernoulli rendszer mikor izomorf egy alább ismertetendő természetes izomorfia fogalom szerint, mely izomorfia szemléletesen a két dinamikus rendszer hasonlóságát fejezi ki. Érdemes ezt a kérdést általánosabban megfogalmazni. Bevezetem az (invertálható shift transzformációval rendelkező) dinamikus rendszerek fogalmát, és definiálom ezek izomorfiáját. A minket érdeklő kérdés arról szól, hogy bizonyos speciális dinamikus rendszerek mikor izomorfak.

(Invertálható) dinamikus rendszerek definiciója. *Egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt, és egy az Ω halmazt önmagába képező, mérhető T leképezést dinamikus rendszernek nevezünk, ha T mértéktartó transzformáció, azaz $P(T^{-1}(A)) = P(A)$ minden $A \in \mathcal{A}$ halmazra. Egy dinamikus rendszert invertálhatónak nevezünk, ha a T transzformáció automorfizmus, azaz minden $\omega \in \Omega$ pontra pontosan egy olyan $\tilde{\omega} \in \Omega$ pont van, amelyre $T\tilde{\omega} = \omega$.*

Nem nehéz belátni, hogy egy Bernoulli rendszer (az ott definiált) shift transzformációval invertálható dinamikus rendszer. A jobb érthetőség kedvéért mutatok egy a Bernoulli rendszerhez hasonló nem invertálható dinamikus rendszert, amelyet féloldali Bernoulli rendszernek fogok nevezni.

Féloldali Bernoulli rendszer definiciója. *Legyen adva egy $r \geq 2$ egész szám, és olyan $p_j \geq 0$, $1 \leq j \leq r$, számok, amelyekre $\sum_{j=1}^{r} p_j = 1$. Az $r \geq 2$, és p_j, $1 \leq j \leq r$, számok által meghatározott féloldali Bernoulli rendszeren az alábbi $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt és az Ω halmazon definiált T úgynevezett shift (eltolás) transzformációt értjük. Az Ω halmaz elemei azon $\omega = (x_0, x_1, \dots)$ sorozatok, amelyekre $1 \leq x_j \leq r$, és x_j egész szám minden $0 \leq j < \infty$ indexre. Az $\mathcal{A}$ σ-algebra az Ω halmaz az alábbi $A(k, j_0, \dots, j_k)$ úgynevezett hengerhalmazok által generált σ-algebra: $A(k, j_0, \dots, j_k) = \{\omega = (x_0, x_1, x_2, \dots)$: $x_s = j_s$, $0 \leq s \leq k\}$, $k = 1, 2, \dots$, $1 \leq j_s \leq r$ minden $0 \leq s \leq k$ indexre. A hengerhalmazok P valószínűségét a $P(A(k, j_0, \dots, j_k) = \prod_{s=0}^{k} p_{j_s}$ képlet adja meg, és a P mérték e valószínűség kiterjesztése a $\mathcal{A}$ σ-algebrára. Végül egy $\omega = (x_0, x_1, x_2, \dots) \in \Omega$ elemi esemény $T\omega$ shiftje a $T\omega = (x_1, x_2, x_3, \dots) \in \Omega$ sorozat, azaz az ω-t definiáló sorozat x_s, s-ik koordinátáját eggyel eltoljuk balra, és az x_0 koordináta 'elveszik'.*

Legyen adva két $(\Omega, \mathcal{A}, P, T)$ és $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ dinamikus rendszer. Természetesnek látszana ezek valamely φ izomorfiáját úgy definiálni, mint az Ω halmaznak olyan kölcsönösen egyértelmű, mértéktartó φ leképezését az $\tilde{\Omega}$ halmazba, amely a T shift transzformációt a $\tilde{T}$ shift transzformációba viszi, azaz $\varphi(T\omega) = \tilde{T}\varphi(\omega)$ minden $\omega \in \Omega$ pontban. Mint a következő példa mutatja, érdemes ezt a definiciót kissé finomítani. Lehetséges ugyanis, hogy valamelyik dinamikus rendszernek van egy olyan rossz null mértékű részhalmaza, amely kizárja az ilyen értelemben vett izomorfiát, de ha ezt a null mértékű halmazt elhagyjuk akkor minden rendben lesz.

Tekintsük a következő példát. Vegyük azt az $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt, amelyre $\Omega = \{1\}$, $\mathcal{A}$ az Ω halmaz összes részhalmaza, (ez valójában az $\{1\}$ halmaz és az üres

halmaz), és $P(\{1\}) = 1$. Definiáljuk a T shift transzformációt az Ω halmazon, mint az identitás transzformációt. Vezessük be az $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ valószínűségi mezőt, amelyre $\tilde{\Omega} = \{0, 1\}$, $\tilde{\mathcal{A}}$ az $\tilde{\Omega}$ halmaz összes részhalmaza, $\tilde{P}(\{1\}) = 1$, és $\tilde{P}(\{0\}) = 0$. Definiáljuk a $\tilde{T}$ shift transzformációt az $\tilde{\Omega}$ halmazon, mint az identitás transzformációt. Ekkor, mind $(\Omega, \mathcal{A}, P, T)$ mind $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ dinamikus rendszer egy invertálható shift transzformációval. A két rendszer nem izomorf az előbb vázolt értelemben, mert Ω egy, $\tilde{\Omega}$ pedig két elemből áll. Viszont a $\tilde{\Omega}$ halmazból kihagyva a null mértékű $\{0\}$ halmazt már két izomorf dinamikus rendszert kapunk. Ezért érdemes dinamikus rendszerek izomorfiáját az alább megadandó módon definiálni, mert az jobban kifejezi két dinamikus rendszer hasonlóságát. A definició szemléletes tartalma az, hogy két dinamikus rendszert akkor tekintünk izomorfnak, ha egy rossz null mértékű halmazt kihagyva mind a két dinamikus rendszerből olyan rendszereket kapunk, amelyek az előbb jelzett erősebb értelemben is izomorfak.

Dinamikus rendszerek izomorfiájának a definiciója. *Legyen adva két* $(\Omega, \mathcal{A}, P, T)$ *és* $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ *dinamikus rendszer. A két rendszer izomorf, ha létezik két olyan* $\Omega_0 \in \mathcal{A}$ *és* $\tilde{\Omega}_0 \in \tilde{\mathcal{A}}$ *halmaz és egy (mérhető) kölcsönösen egyértelmű* $\varphi\colon \Omega_0 \to \tilde{\Omega}_0$ *leképezés, amelyekre*

1.) $P(\Omega_0) = 1$, $\tilde{P}(\tilde{\Omega}_0) = 1$, *az* Ω_0, *és* $\tilde{\Omega}_0$ *halmazok invariánsak a* T *illetve* $\tilde{T}$ *shift transzformációra, azaz* $\Omega_0 \subset T^{-1}\Omega_0$, *és* $\tilde{\Omega}_0 \subset \tilde{T}^{-1}\tilde{\Omega}_0$. *Ez ekvivalensen úgy is megfogalmazható, hogy* $T\Omega_0 \subset \Omega_0$, *és* $T\tilde{\Omega}_0 \subset \tilde{\Omega}_0$.

2.) *A* $\varphi\colon\ \Omega_0 \to \tilde{\Omega}_0$ *leképezés mértéktartó, azaz ha* $A \subset \Omega_0$, $\tilde{A} \subset \tilde{\Omega}_0$, *és* $\tilde{A} = \varphi(A)$, *akkor* $A \in \mathcal{A}$ *akkor és csak akkor, ha* $\tilde{A} \in \tilde{\mathcal{A}}$, *és ebben az esetben* $P(A) = \tilde{P}(\tilde{A})$.

3.) *A* φ *leképezés felcserélhető a* T, $\tilde{T}$ *shift párral, azaz* $\varphi(T\omega) = \tilde{T}\varphi(\omega)$ *tetszőleges* $\omega \in \Omega_0$ *pontra.*

Ha a fenti tulajdonságok teljesülnek valamely Ω_0, $\tilde{\Omega}_0$ *párral és* φ *leképezéssel, akkor azt mondjuk, hogy az* $(\Omega, \mathcal{A}, P, T)$ *és* $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ *dinamikus rendszerek izomorfak az* $(\Omega_0, \tilde{\Omega}_0, \varphi)$ *hármason keresztül.*

Megjegyzés. Tetszőleges dinamikus rendszerek izomorfiáját definiáltuk, de a fő eredményekben csak invertálható dinamikus rendszerek izomorfiáját fogjuk vizsgálni. A minket érdeklő Bernoulli rendszerek invertálható dinamikus rendszerek, és vizsgálataink bizonyos részeiben ezt ki fogjuk használni.

Ha két rendszer izomorfiáját akarjuk vizsgálni valamilyen izomorfia fogalom szerint, akkor természetes az izomorfia invariánsait, azaz olyan tulajdonságokat és mennyiségeket keresni, amelyek nem változnak akkor, ha egy rendszerből egy másik vele izomorf rendszerbe térünk át. Minél több invariánst ismerünk annál jobban tudjuk az izomorfiát vizsgálni.

Tekintsünk invertálható dinamikus rendszereket, és vizsgáljuk ezek előbb bevezetett izomorfiáját. Először a következő nem triviális az izomorfiára invariáns tulajdonságot találták. Adva egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, természetes módon definiálhatjuk a következő Hilbert teret és rajta definiált unitér operátort. Álljon a Hilbert tér az $(\Omega, \mathcal{A}, P)$ téren értelmezett négyzetesen integrálható függvényekből a

szokásos L_2 normával, és vezessük be e téren az alábbi U operátort. Ha $\int f^2(\omega)P(\,d\omega) < \infty$, akkor definiáljuk az f függvény Uf képét az $Uf(\omega) = f(T\omega)$ képlettel. Nem nehéz belátni, hogy (a T shift transzformáció mértéktartó tulajdonsága és invertálhatósága miatt) az előbb definiált U operátor unitér. Továbbá, ha két invertálható dinamikus rendszer izomorf, akkor a nekik megfelelő Hilbert tér a rajtuk definiált U unitér operátorral izomorf. Felmerült a kérdés, hogy ez a tény milyen információt ad két Bernoulli rendszer izomorfiájáról.

Kiderült, hogy bármely két Bernoulli rendszerhez tartozó az előbbi módon bevezetett Hilbert tér a rajta definiált unitér operátorral együtt izomorf. Ezen eredmény bizonyítását ismertetem e fejezet kiegészítésében. Sokáig azt hitték, hogy ez az a lényeges izomorfiára invariáns tulajdonág, amely eldönti, hogy két Bernoulli rendszer izomorf-e. Ezért többen azt sejtették, hogy bármely két Bernoulli rendszer izomorf. Sőt, Paul Halmos bebizonyította, hogy ezen sejtés igazolásához elegendő lenne azt belátni, hogy az $r = 3$, $p_1 = p_2 = p_3 = \frac{1}{3}$ illetve $r = 4$, $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$ paraméterekkel meghatározott Bernoulli rendszerek izomorfak. Később Kolmogorov bebizonyította, hogy ez a sejtés hamis, mert létezik olyan további a dinamikus rendszerek izomorfiájára invariáns mennyiség, amelynek létezéséből következik például, hogy a fent említett Bernoulli rendszerek nem izomorfak.

Kolmogorov bevezette a Shannon-féle entrópia egy természetes általánosítását. Definiálta dinamikus rendszerek entrópiáját, és megmutatta, hogy egymással izomorf dinamikus rendszerek entrópiája egyenlő. Ezenkívül olyan eredményt bizonyított, amely segített az entrópia kiszámolásában bizonyos esetekben. Speciálisan megmutatta, hogy egy r, $p_1, \dots, p_r$, paraméterekkel meghatározott Bernoulli rendszer entrópiája $H = -\sum_{j=1}^{r} p_j \log p_j$, és az általa bevezetett entrópia tekinthető úgy, mint a Shannon-féle entrópia általánosítása. Később David Ornstein egy mély eredményben bebizonyította, hogy két Bernoulli rendszer, amelyeknek megegyezik az entrópiája, izomorf.

Ebben a fejezetben Kolmogorov eredményét és annak bizonyítását ismertetem. Nem fogom tárgyalni Ornstein eredményének a bizonyítását. Természetesen Bernoulli rendszerek izomorfiájának a problémája csak egy speciális, bár fontos része annak a kérdéskörnek, hogy két dinamikus rendszer mikor izomorf. Általános dinamikus rendszerek izomorfiájának a kérdésével azonban ebben a jegyzetben nem foglalkozom.

Kolmogorov eredményeinek tárgyalása előtt ismertetek néhány a dinamikus rendszerek izomorfiájával kapcsolatos tényt.

Észrevétel. *Dinamikus rendszerek izomorfiája ekvivalencia reláció.*

Bizonyítás. Nyilvánvaló, hogy egy dinamikus rendszer önmagával izomorf, azaz az izomorfia reflexiv. Ugyancsak könnyen látható, hogy az izomorfia szimmetrikus tulajdonság. Ha $(\Omega, \mathcal{A}, P, T)$ izomorf egy $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ dinamikus rendszerrel egy $(\Omega_0, \tilde{\Omega}_0, \varphi)$ hármason keresztül, akkor $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ izomorf az $(\Omega, \mathcal{A}, P, T)$ dinamikus rendszerrel az $(\tilde{\Omega}_0, \Omega_0, \varphi^{-1})$ hármason keresztül. Be kell még látni, hogy az izomorfia tranzitív tulajdonság.

Azt kell megmutatni, hogy ha $(\Omega, \mathcal{A}, P, T)$ izomorf egy $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ dinamikus

rendszerrel egy $(\Omega_0, \tilde{\Omega}_0, \varphi)$, és $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ izomorf egy $(\Omega', \mathcal{A}', P', T')$ dinamikus rendszerrel egy $(\tilde{\Omega}_1, \Omega'_1, \psi)$ hármason keresztül, akkor az $(\Omega, \mathcal{A}, P, T)$ és $(\Omega', \mathcal{A}', P', T')$ dinamikus rendszerek is izomorfak. Ennek érdekében először megmutatom azt, hogy az izomorfiákat biztosító hármasok választhatóak $(\Omega_2, \tilde{\Omega}_2, \varphi_2)$ és $(\tilde{\Omega}_2, \Omega'_2, \psi_2)$ alakban alkalmas Ω_2, $\tilde{\Omega}_2$, Ω'_2, φ_2 és ψ_2 mennyiségekkel. A lényeges pont ebben az állításban az, hogy a két hármasban ugyanaz az $\tilde{\Omega}_2$ halmaz szerepel.

Legyen $\tilde{\Omega}_2 = \tilde{\Omega}_0 \cap \tilde{\Omega}_1$. Ha $\Omega_2 = \varphi^{-1}\tilde{\Omega}_2$, φ_2 a φ függvény megszorítása az Ω_2 halmazra, akkor $(\Omega, \mathcal{A}, P, T)$ izomorf az $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ dinamikus rendszerrel az $(\Omega_2, \tilde{\Omega}_2, \varphi_2)$ hármason keresztül is. Hasonlóan, legyen $\Omega'_2 = \psi\tilde{\Omega}_2$, és ψ_2 a ψ függvény megszorítása az $\tilde{\Omega}_2$ halmazra. Ekkor $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ izomorf az $(\Omega', \mathcal{A}', P', T')$ dinamikus rendszerrel az $(\tilde{\Omega}_2, \Omega'_2, \psi_2)$ hármason keresztül. Ezt felhasználva kapjuk, hogy az $(\Omega, \mathcal{A}, P, T)$ és $(\Omega', \mathcal{A}', P', T')$ dinamikus rendszrek izomorfak az $(\Omega_2, \Omega'_2, \rho)$ hármason keresztül, ahol $\rho(\omega) = \psi_2(\varphi_2(\omega))$ minden $\omega \in \Omega_2$ pontban. (A ρ függvény definiciójában használtuk ki az $\tilde{\Omega}_2$ halmaz fent említett tulajdonságát.)

Szükségünk van még a következő eredményre is.

Lemma izomorf dinamikus rendszerek tulajdonságairól. *Legyen $(\Omega, \mathcal{A}, P, T)$ és $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ két izomorf dinamikus rendszer egy $(\Omega_0, \tilde{\Omega}_0, \varphi)$ hármason keresztül. Ekkor*

$$\varphi(T^n\omega) = \tilde{T}^n\varphi(\omega) \quad \textit{minden } n = 1, 2, \dots \textit{ számra, és minden } \omega \in \Omega_0 \textit{ pontban.} \tag{5.1}$$

Legyen adva k darab $\tilde{A}_j$ halmaz, amelyekre $\tilde{A}_j \subset \tilde{\Omega}_0$, és $\tilde{A}_j \in \tilde{\mathcal{A}}$, $1 \le j \le k$, és $n_j \ge 0$, $1 \le j \le k$, nem negatív egész számok egy sorozata. Ekkor

$$P(T^{-n_1}\varphi^{-1}(\tilde{A}_1) \cap \dots \cap T^{-n_k}\varphi^{-1}(\tilde{A}_k)) = \tilde{P}(\tilde{T}^{-n_1}\tilde{A}_1 \cap \dots \cap \tilde{T}^{-n_k}\tilde{A}_k). \tag{5.2}$$

A fenti lemma azt mondja ki, hogy bár dinamikus rendszerek izomorfiájának a definiciójában megengedtük bizonyos null mértékű halmazok kihagyását, a T illetve $\tilde{T}$ shift operátorok hatványai úgy viselkednek, mint abban az egyszerűbb esetben, amikor a null mértékű halmazok ezen kihagyását nem engedjük meg, azaz, ha $\Omega_0 = \Omega$, és $\tilde{\Omega}_0 = \tilde{\Omega}$.

A lemma bizonyítása. Az (5.1) formulát n szerinti teljes indukcióval láthatjuk be. $n =$ 1-re a formula igaz, és ha igaz n-re, akkor $\varphi(T^{n+1}\omega) = \varphi(T^n(T\omega)) = \tilde{T}^n\varphi(T\omega) = \tilde{T}^n(\tilde{T}(\varphi(\omega)) = \tilde{T}^{n+1}\varphi(\omega)$.

Az (5.2) reláció igazolása érdekében először mutassuk meg, hogy amennyiben $\tilde{A} \in \tilde{\Omega}_0$, akkor minden $n \ge 0$ számra $T^{-n}\varphi^{-1}(\tilde{A}) \cap \Omega_0 = \varphi^{-1}(\tilde{T}^{-n}(\tilde{A}) \cap \tilde{\Omega}_0)$. Valóban, $\omega \in T^{-n}\varphi^{-1}(\tilde{A}) \cap \Omega_0$ akkor és csak akkor, ha $\omega \in T^{-n}\varphi^{-1}(\tilde{A})$ és $\omega \in \Omega_0$, tehát $T^n\omega \in \varphi^{-1}(\tilde{A})$ és $\omega \in \Omega_0$, azaz $\varphi(T^n\omega) \in \tilde{A}$, és $\omega \in \Omega_0$.

Másrészt $\omega \in \varphi^{-1}(\tilde{T}^{-n}(\tilde{A}) \cap \tilde{\Omega}_0)$ azzal ekvivalens, hogy $\varphi(\omega) \in \tilde{T}^{-n}(\tilde{A}) \cap \tilde{\Omega}_0$, azaz $\tilde{T}^n\varphi(\omega) \in \tilde{A}$ és $\varphi(\omega) \in \tilde{\Omega}_0$. Ez viszont az (5.1) reláció szerint azzal ekvivalens, hogy $\varphi(T^n\omega) \in \tilde{A}$ és $\omega \in \Omega_0$. A felírt azonosság tehát érvényes.

Alkalmazva ezt az azonosságot mindegyik $\tilde{A}_j$, $1 \le j \le k$, halmazra n_j paraméterrel, és véve az azonosság két oldalán lévő halmazok metszetét azt kapjuk, hogy

$$T^{-n_1}\varphi^{-1}(\tilde{A}_1) \cap \dots \cap T^{-n_k}\varphi^{-1}(\tilde{A}_k) \cap \Omega_0 = \varphi^{-1}(\tilde{T}^{-n_1}\tilde{A}_1 \cap \dots \cap \tilde{T}^{-n_k}\tilde{A}_k \cap \tilde{\Omega}_0).$$

(Jegyezzük meg, hogy a φ függvény φ^{-1} inverzére $\varphi^1(A) \cap \varphi^{-1}(B) = \varphi^{-1}(A \cap B)$.) Ezért a fenti azonosság két oldalán levő halmaz P valószínűsége egyenlő. Az (5.2) azonosság következik ebből az azonosságból és a következő két észrevételből.

$$P(T^{-n_1}\varphi^{-1}(\tilde{A}_1) \cap \dots \cap T^{-n_k}\varphi^{-1}(\tilde{A}_k) \cap \Omega_0) = P(T^{-n_1}\varphi^{-1}(\tilde{A}_1) \cap \dots \cap T^{-n_k}\varphi^{-1}(\tilde{A}_k)),$$

mert $P(\Omega_0) = 1$. Másrészt

$$\begin{aligned} &P(\varphi^{-1}(\tilde{T}^{-n_1}\tilde{A}_1 \cap \dots \cap \tilde{T}^{-n_k}\tilde{A}_k \cap \tilde{\Omega}_0)) \\ &\quad = \tilde{P}(\tilde{T}^{-n_1}\tilde{A}_1 \cap \dots \cap \tilde{T}^{-n_k}\tilde{A}_k \cap \tilde{\Omega}_0) = \tilde{P}(\tilde{T}^{-n_1}\tilde{A}_1 \cap \dots \cap \tilde{T}^{-n_k}\tilde{A}_k) \end{aligned}$$

a φ transzformáció mértéktartó tulajdonsága és a $\tilde{P}(\tilde{\Omega}_0) = 1$ reláció miatt.

Bevezetem egy invertálható dinamikus rendszer entrópiájájának a fogalmát. De ezt csak *invertálható* dinamikus rendszerek esetében fogom megtenni. A definició bevezetése érdekében először bebizonyítok egy egyszerű lemmát. A lemma megfogalmazásában használni fogom a következő jelölést. Legyen $(\Omega, \mathcal{A}, P, T)$ egy invertálható dinamikus rendszer. Egy e dinamikus rendszerben definiált ξ valószínűségi változó $T^n\xi$ eltoltján a $T^n\xi(\omega) = \xi(T^n\omega)$ valószínűségi változót értjük minden $n = \dots, -1, 0, 1, \dots$ indexre. (Speciálisan $T^0\xi(\omega) = \xi(\omega)$.)

Lemma az entrópia egy tulajdonságáról. *Legyen $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, és legyen $\xi(\omega)$ egy $\mathcal{A}$ mérhető véges vagy megszámlálhatóan végtelen sok értéket felvevő valószínűségi változó. Ekkor létezik az*

$$\lim_{n\to\infty} H(\xi|T^{-1}\xi, \dots, T^{-n}\xi). \tag{5.3}$$

határérték. Ha $H(\xi) < \infty$ akkor ez a határérték véges, és

$$\begin{aligned} \lim_{n\to\infty} H(\xi|T^{-1}\xi, \dots, T^{-n}\xi) &= \lim_{n\to\infty} \frac{1}{n} H(\xi, T\xi, \dots, T^{n-1}\xi) \\ &= \lim_{n\to\infty} \frac{1}{2n-1} H(T^{-n+1}\xi, \dots, T^{-1}\xi, \xi, T\xi, \dots, T^{n-1}\xi). \end{aligned} \tag{5.4}$$

Bizonyítás. Az első fejezet eredményeiből következik, hogy a $H(\xi|T^{-1}\xi, \dots, T^{-k}\xi)$ feltételes entrópia sorozat a k paraméter monoton csökkenő függvénye. Ezért létezik a

$$\lim_{n\to\infty} H(\xi|T^{-1}\xi, \dots, T^{-n}\xi)$$

határérték, és az véges, ha $H(\xi) < \infty$. Ebben az esetben felírhatjuk a

$$\begin{aligned}\frac{1}{n}H(\xi, T\xi, \dots, T^{n-1}\xi) &= \frac{1}{n}\sum_{k=1}^{n-1} H(T^k\xi | T^{k-1}\xi, \dots, T^0\xi) + \frac{H(\xi)}{n} \\ &= \frac{1}{n}\sum_{k=1}^{n-1} H(\xi | T^{-1}\xi, \dots, T^{-k}\xi) + \frac{H(\xi)}{n}\end{aligned}$$

azonosságot. E formula első azonosságában felhasználtuk az entrópia és feltételes entrópia első fejezetben bizonyított tulajdonságait, a második azonosságban pedig azt a tényt, hogy a $(T^k\xi, T^{k-1}\xi, \dots, \xi)$, illetve $(\xi, T^{-1}\xi, \dots, T^{-k}\xi)$ vektorok azonos eloszlásúak, ezért $H(T^k\xi | T^{k-1}\xi, \dots, T^0\xi) = H(\xi | T^{-1}\xi, \dots, T^{-k}\xi)$ minden $k = 0, 1, \dots$ indexre. Innen azt kapjuk, hogy

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}H(\xi, T\xi, \dots, T^{n-1}\xi) &= \lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n-1} H(\xi | T^{-1}\xi, \dots, T^{-k}\xi) + \frac{H(\xi)}{n}\right) \\ &= \lim_{n\to\infty} H(\xi | T^{-1}\xi, \dots, T^{-1}\xi, \dots, T^{-n}\xi)\end{aligned}$$

azaz igaz az (5.4) formula első azonossága.

Mivel

$$H(T^{-n+1}\xi, \dots, T^{-1}\xi, \xi, T\xi, \dots, T^{n-1}\xi) = H(\xi, T\xi, \dots, T^{2n-2}\xi)$$

igaz az (5.4) formula második azonossága is.

Invertálható dinamikus rendszer entrópiájának a definiciója. *Legyen adva egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, és tekintsünk ezen egy olyan $\mathcal{A}$ mérhető véges vagy megszámlálhatóan végtelen sok értéket felvevő ξ valószínűségi változót. A ξ valószínűségi változó T shift transzformáció szerinti entrópiája a*

$$H(T, \xi) = \lim_{n\to\infty} H(\xi | T^{-1}\xi, \dots, T^{-n}\xi) \tag{5.5}$$

határérték. (Az előző lemma szerint ez a határérték létezik, és véges, ha $H(\xi) < \infty$.) A T shift transzformáció entrópiája a

$$H(T) = \sup_{\xi} H(T, \xi) \tag{5.6}$$

kifejezéssel egyenlő, ahol a szuprémumot az összes $\mathcal{A}$ mérhető és véges sok értéket felvevő ξ valószínűségi változóra vesszük.

Megjegyzés. Láttuk, hogy a $H(\xi) < \infty$ esetben a $H(T, \xi)$ mennyiséget a

$$H(T, \xi) = \lim_{n\to\infty}\frac{1}{n}H(T^0\xi, \dots, T^{n-1}\xi)$$

képlet segítségével is kifejezhetjük.

Egy invertálható dinamikus rendszer shift transzformációjának az entrópiáját az irodalomban gyakran kissé más, de ekvivalens módon írják le. Ismertetem ezt a definiciót is. Előtte emlékeztetek arra, hogy egy (véges vagy megszámlálható értéket felvevő) valószínűségi változó természetes módon meghatározza a valószínűségi mező egy partícióját. Nevezetesen, e partició elemei a valószínűségi mező azon (nívó)halmazai, ahol a valószínűségi változó egy előírt értéket vesz fel. Továbbá a valószínűségi változó entrópiája csak ezen particiótól függ. Másrészt egy $(\Omega, \mathcal{A}, P)$ valószínűségi mező minden $\mathcal{A}$ mérhető halmazokból álló particiójához létezik olyan $\mathcal{A}$ mérhető ξ valószínűségi változó, amely ezt a particiót határozza meg. Ez lehetővé teszi, hogy egy dinamikus rendszer shift transzformációjának az entrópiáját ne valószínűségi változók, hanem partíciók segítségével definiáljuk, és az irodalomban gyakran ezt teszik. Megadom ezt a definiciót is, illetve ismertetem kapcsolatát az előbb leírt definicióval. Az egyszerűség kedvéért csak olyan ξ valószínűségi változókkal, illetve nekik megfelelő $\mathcal{B}$ particiókkal fogok foglalkozni, amelyekre $H(\xi) < \infty$.

Legyen adva egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer. Vegyük észre, hogy amennyiben egy $\mathcal{A}$ mérhető ξ valószínűségi változónak az Ω halmaz egy $\mathcal{B}$ particiója felel meg, akkor a $T^n\xi$ valószínűségi változónak a $\mathcal{B}$ partíciónak az e partició B_j halmazainak a T^n transzformáció szerinti $T^{-n}B_j$ ősképeiből álló $T^{-n}\mathcal{B}$ particició felel meg. Ha adva vannak az Ω halmaz valamely $\mathcal{C}_1, \dots, \mathcal{C}_k$ particiói, akkor jelölje $\mathcal{C}_1 \wedge \dots \wedge \mathcal{C}_k$ e partíciók közös finomítását. Ez az összes $C_{j_1}(1) \cap \dots \cap C_{j_k}(k)$ alakú halmazból áll, ahol $C_{j_s}(s) \in \mathcal{C}_s$, $1 \le s \le k$. Ha a ξ, $H(\xi) < \infty$, valószínűségi változó a $\mathcal{B}$ particiót határozza meg, akkor ezzel a jelöléssel a $(T^{n_1}\xi, \dots, T^{n_k}\xi)$ vektornak a $T^{-n_1}\mathcal{B} \wedge \dots \wedge T^{-n_k}\mathcal{B}$ partició felel meg.

Ha adva van az Ω halmaz egy olyan $\mathcal{B} = \{B_1, B_2, \dots\}$ particiója, amelyet egy ξ valószínűségi változó határoz meg, akkor természetes a $\mathcal{B}$ partició T shift szerinti entrópióját a

$$H(T, \mathcal{B}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{B} \wedge \dots \wedge T^{-(n-1)}\mathcal{B})$$

és

$$H(\mathcal{B} \wedge \dots \wedge T^{-(n-1)}\mathcal{B}) = H(\xi, T\xi, \dots, T^{n-1}\xi) = -\sum_{j_1,\dots,j_n} g(P(B_{j_1} \cap \dots \cap T^{-(n-1)}B_{j_n}))$$

képletek segítségével definiálni, ahol $g(x) = x \log x$. Ekkor a

$$H(T) = \sup_{\mathcal{B}} H(T, \mathcal{B})$$

képlet, ahol a szuprémumot az Ω halmaz összes véges és $\mathcal{A}$ mérhető particiójára vesszük a T shift transzformáció előbb bevezetett entrópiáját adja.

Megfogalmazom az entrópia invariáns tulajdonságáról szóló eredményt.

Tétel izomorf dinamikus rendszerek entrópiájának egyenlőségéről. *Legyen* $(\Omega, \mathcal{A}, P, T)$ *és* $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{T})$ *két izomorf invertálható dinamikus rendszer. Ekkor a* T *és* $\tilde{T}$ *shift transzformációk entrópiája egyenlő, azaz* $H(T) = H(\tilde{T})$.

Bizonyítás. A jobb megértés érdekében tekintsük először azt a speciális esetet, amikor a két dinamikus rendszer az $(\Omega, \tilde{\Omega}, \varphi)$ hármason keresztül izomorf, ahol φ az Ω halmaz egy alkalmas kölcsönösen egyértelmű leképezése a $\tilde{\Omega}$ halmazba. Azaz azt az esetet tekintjük, amikor az izomorfia definicióját biztosító $(\Omega_0, \tilde{\Omega}_0, \varphi)$ hármasban $\Omega_0 = \Omega$ és $\tilde{\Omega}_0 = \tilde{\Omega}$ halmazokat választhatunk.

Ebben az esetben az Ω halmaz egy $\mathcal{A}$ mérhető $\mathcal{B} = \{B_1, \dots, B_r\}$ véges particiójának feleltessük meg az $\tilde{\Omega}$ halmaz $\tilde{\mathcal{A}}$ mérhető $\tilde{\mathcal{B}} = \{\varphi(B_1), \dots, \varphi(B_r)\}$ mérhető particióját. Az izomorfia tulajdonságai miatt ekkor tetszőleges n számra és $1 \le j_s \le r$, $1 \le s \le n$ indexekre $P(T^{-0}B_{j_0} \cap T^{-1}B_{j_1} \cap \dots \cap T^{-n}B_{j_n}) = \tilde{P}(\tilde{T}^{-0}\tilde{B}_{j_0} \cap \tilde{T}^{-1}\tilde{B}_{j_1} \cap \dots \cap \tilde{T}^{-n}\tilde{B}_{j_n})$. Innen következik, hogy tetszőleges $\mathcal{A}$ mérhető (véges sok értéket felvevő) ξ valószínűségi változóhoz létezik olyan $\tilde{\mathcal{A}}$ mérhető (véges sok értéket felvevő) η valószínűségi változó, amelyre $H(\xi, T\xi, \dots, T^n\xi) = H(\eta, \tilde{T}\eta, \dots, \tilde{T}^n\eta)$ minden n-re, ezért $H(T, \xi) = H(\tilde{T}, \eta)$. Hasonló állítás érvényes akkor is, ha az Ω halmaz és $\mathcal{A}$ mérhető partíciók illetve a $\tilde{\Omega}$ halmaz és $\tilde{\mathcal{A}}$ mérhető partíciók szerepét felcseréljük. Ezért $H(T) = H(\tilde{T})$.

Az általános esetben a fenti érvelést kissé finomítani kell. Szimmetria okokból elég azt belátni, hogy $H(\tilde{T}) \le H(T)$, és ehhez elég azt bebizonyítani, hogy ha η egy az $\tilde{\Omega}$ halmazon definiált véges sok értéket felvevő $\tilde{\mathcal{A}}$ mérhető valószínűségi változó, akkor létezik olyan az Ω halmazon definiált véges sok értéket felvevő $\mathcal{A}$ mérhető ξ valószínűségi változó, amelyre $H(T, \xi) = H(\tilde{T}, \eta)$. Sőt azt is feltehetjük, hogy az η valószínűségi változó nívóhalmazai az $\tilde{\Omega}$ halmaznak egy olyan $\{\tilde{B}_1, \dots, \tilde{B}_r, \tilde{B}_{r+1}\}$ particióját adják, amelyre $\bigcup_{j=1}^{r} \tilde{B}_j = \tilde{\Omega}_0$, és $\tilde{B}_{r+1} = \tilde{\Omega} \setminus \tilde{\Omega}_0$. Vezessük be az Ω halmaznak azt a $\{B_1, \dots, B_r, B_{r+1}\}$ particióját, amelyre $B_s = \varphi^{-1}(\tilde{B}_s)$, ha $1 \le s \le r$, és $B_{r+1} = \Omega \setminus \Omega_0$. Legyen ξ egy olyan valószínűségi változó, amelynek nívóhalmazai ezek a B_s, $1 \le s \le r+1$, halmazok. Azt állítom, hogy $H(T, \xi) = H(\tilde{T}, \eta)$. Sőt, az is igaz, hogy tetszőleges n számra $H(\xi, T\xi, \dots, T^n\xi) = H(\eta, \tilde{T}\eta, \dots, \tilde{T}^n\eta)$. Ehhez azt kell észrevenni, hogy az (5.2) azonosság miatt

$$P(\varphi^{-1}(\tilde{C}_{j_0}) \cap T^{-1}\varphi^{-1}(\tilde{C}_{j_1}) \cap \dots \cap T^{-n}\varphi^{-1}(\tilde{C}_{j_n})) = \tilde{P}(\tilde{C}_{j_0} \cap \tilde{T}^{-1}\tilde{C}_{j_1} \cap \dots \cap \tilde{T}^{-n}\tilde{C}_{j_n}),$$

ha $1 \le j_s \le r$ minden $1 \le s \le n$ indexre, és a bizonyítandó azonosságban szereplő két entrópia ezen valószínűségek függvénye. (Azokat a tagokat, amelyekben a $\tilde{C}_{r+1}$ vagy C_{r+1} események szerepelnek elhagyhatjuk a megfelelő entrópiák kiszámolásában, mert ezáltal nulla valószínűségű események függvényeit hagyjuk ki a megfelelő entrópiákat kifejező összegekből. A tételt beláttuk.

Annak érdekében, hogy a fenti tételt alkalmazni tudjuk szükségünk van olyan eredményre, amely lehetővé teszi azt, hogy egy invertálható dinamikus rendszer shift transzformációjának az entrópiáját kiszámoljuk. Egy ilyen eredmény megfogalmazásának az érdekében bevezetem a következő definiciót.

Egy valószínűségi változó eltoltjai által generált σ-algebra definiciója. *Legyen $(\Omega, \mathcal{A}, P, T)$ egy invertálható dinamikus rendszer, és ξ egy $\mathcal{A}$ mérhető valószínűségi változó. A ξ valószínűségi változó és a T shift által generált σ-algebrán a $T^j\xi$, $-\infty < j < \infty$, valószínűségi változók által generált $\sigma(T, \xi) = \sigma(T^j\xi, -\infty < j < \infty)$, σ-algebrát értjük.*

Tétel egy dinamikus rendszerben definiált valószínűségi változók entrópiájának az összehasonlításáról. *Legyen $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, ξ és η pedig két olyan $\mathcal{A}$ mérhető valószínűségi változó, amelyek közül ξ véges sok, η véges sok vagy megszámlálhatóan végtelen sok értéket vesz fel, $H(\eta) < \infty$, és ξ $\sigma(T, \eta)$ mérhető valószínűségi változó. Ekkor $H(T, \xi) \le H(T, \eta)$.*

Következmény. *Legyen adva egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, és legyen η olyan az Ω halmazon definiált véges sok értéket felvevő valószínűségi változó, amelyre $\sigma(T, \eta) = \mathcal{A}$. Ekkor $H(T) = H(T, \eta)$. Speciálisan, ha $(\Omega, \mathcal{A}, P, T)$ egy $r \ge 2$ egész számmal és $p_j \ge 0$, $1 \le j \le r$, $\sum_{j=1}^{r} p_j = 1$, paraméterekkel meghatározott Bernoulli rendszer, akkor $H(T) = -\sum_{j=1}^{r} p_j \log p_j$.*

A következmény bizonyítása. A tétel feltételeinek teljesülése esetén $H(T, \xi) \le H(T, \eta)$ minden véges sok értéket felvevő és $\mathcal{A}$ mérhető ξ valószínűségi változóra. Ezért a $H(T)$ entrópiának az (5.6) formulában megadott definiciója szerint $H(T, \eta) = H(T)$. Egy Bernoulli rendszer esetében definiáljuk az $\eta(\omega) = x_0$, ha $\omega = (\dots, x_{-1}, x_0, x_1, \dots)$ képlettel megadott valószínűségi változót. Ekkor $\sigma(T, \eta) = \mathcal{A}$, ezért $H(T) = H(T, \eta)$. Továbbá a Bernoulli rendszer definiciója miatt a $T^{-n}\eta$, $n = 0, \pm 1, \dots$, valószínűségi változók függetlenek (és egyforma eloszlásúak), ezért $H(\eta | T^{-1}\eta, \dots, T^{-n}\eta) = H(\eta)$, és $H(T) = h(T, \eta) = H(\eta) = -\sum_{j=1}^{r} p_j \log p_j$.

A fent megfogalmazott eredményekből következik, hogy két r, és $p_1, \dots, p_r$ illetve $\bar{r}$ és $\bar{p}_1, \dots, \bar{p}_r$ paraméterekkel definiált Bernoulli rendszer csak akkor lehet izomorf, ha az entrópiájuk egyenlő, azaz, ha $\sum_{j=1}^{r} p_j \log p_j = \sum_{j=1}^{\bar{r}} \bar{p}_j \log \bar{p}_j$. Később be fogom bizonyítani a következmény egy olyan általánosítását, amely lehetővé teszi a Bernoulli rendszerek izomorfiájáról szóló eredmény általánosítását olyan dinamikus rendszerekre is, amelyeket a Bernoulli rendszerekhez hasonlóan definiálunk, de megengedjük azt is, hogy a benne szereplő r paraméter $r = \infty$ legyen. Ezelőtt azonban a shift transzformáció által generált σ-algebráról szóló tétel bizonyítását ismertetem.

Először értsük meg e tétel szemléletes tartalmát. Az entrópia szemléletesen azt adja meg, hogy egy valószínűségi változó megismeréséhez mennyi információ szükséges. Első ránézésre azt várnánk, hogy ha adva van egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer és egy ξ $\mathcal{A}$ mérhető valószínűségi változó, akkor a ξ valószínűségi változó megismeréséhez sokkal kevesebb információ kell, mint például az $\eta = (\xi, T\xi, \dots, T^{1000}\xi)$ véletlen vektor megismeréséhez. Ez az elképzelés azonban téves, mert nem a $H(\xi)$ és $H(\eta)$, hanem a $H(T, \xi)$ és $H(T, \eta)$ entrópiákat kell összehasonlítanunk. Az utóbbi entrópiákat közelítő $\frac{1}{n}H(\xi, \dots, T^n\xi)$ és $\frac{1}{n}H(\xi, \dots, T^{n+1000}\xi)$ entrópiák pedig nagy n paraméterre már nagyon közel vannak egymáshoz. A bizonyítandó tétel az e példa által sugallt képnek felel meg. Azt állítja, hogy ahhoz, hogy egy ξ valószínűségi változóhoz tartozó $\xi, T\xi, \dots$ sorozat tagjainak megismeréséhez szükséges $H(T, \xi)$ információ ne legyen több, mint egy η valószínűségi változóhoz tartozó $\eta, T\eta, \dots$ sorozat tagjainak

megismeréséhez szükséges $H(T,\eta)$ információ elégséges azt feltenni, hogy $\xi \in \sigma(T,\eta)$, azaz szemléletesen azt előírni, hogy a $\dots, T^{-1}\eta, \eta, T\eta, \dots$ véletlen sorozat ismeretében ismerjük a ξ valószínűségi változót is. A bizonyításban szükségünk van egy olyan eredményre, amely azt biztosítja, hogy a $\xi \in \sigma(T,\eta)$ esetben a ξ valószínűségi változó nívóhalmazait jól tudjuk approximálni a $\sigma(T,\eta)$ σ-algebra bizonyos speciális és kényelmesen használható halmazaival. Ezért hasznos lesz számunkra a következő lemma.

Lemma σ-algebra elemeinek jó approximációjáról. *Legyen adva egy $(\Omega, \mathcal{A}, P)$ valószínűségi mező, és jelölje $A\Delta B = (A\setminus B)\cup(B\setminus A)$ két $A \in \mathcal{A}$ és $B \in \mathcal{A}$ halmaz szimmetrikus differenciáját. Vezessük be a $\rho(A,B) = P(A\Delta B)$, $A \in \mathcal{A}$, $B \in \mathcal{A}$, függvényt. Ekkor $\rho(A,B)$ pszeudo metrika, ($\rho(A,B) \geq 0$, de lehetséges, hogy $\rho(A,B) = 0$, noha $A \neq B$). Továbbá $\rho(A_1 \cup A_2, B_1 \cup B_2) \leq \rho(A_1,B_1) + \rho(A_2,B_2)$. Ha $\mathcal{B} \subset \mathcal{A}$ egy halmaz algebra, és $\mathcal{C} = \sigma(\mathcal{B})$ a $\mathcal{B}$ halmaz algebra által generált σ-algebra, akkor minden $\varepsilon > 0$ számhoz és $C \in \mathcal{C}$ halmazhoz létezik olyan $B = B(\varepsilon, C) \in \mathcal{B}$ halmaz, amelyre $\rho(B,C) \leq \varepsilon$.*

Megjegyzés: A következő észrevétel segíthet megérteni a lemmában bevezetett ρ (pszeudo) metrika jelentését. Azonosítsunk egy mérhető A halmazt az indikátorfüggvényével. Ekkor a ρ metrika megegyezik az $L_1(\Omega, \mathcal{A}, P)$ tér normája által indukált metrika megszorításával az indikátorfüggvényekből álló halmazok terére.

Bizonyítás. A $\rho(\cdot)$ függvény pszeudo metrika, mert nyilván $\rho(A,B) \geq 0$, $\rho(A,B) = \rho(B,A)$, és a $\rho(A,C) \leq \rho(A,B) + \rho(B,C)$ reláció is teljesül, mert mint könnyű ellenőrizni, $P(A \setminus C) \leq P(A \setminus B) + P(B \setminus C)$, és $P(C \setminus A) \leq P(B \setminus A) + P(C \setminus B)$. Hasonlóan, $\rho(A_1\cup A_2, B_1\cup B_2) \leq \rho(A_1,B_1)+\rho(A_2,B_2)$, mert $P((A_1\cup A_2)\setminus(B_1\cup B_2)) \leq P(A_1 \setminus B_1) + P(A_2 \setminus B_2)$, és $P((B_1 \cup B_2) \setminus (A_1 \cup A_2)) \leq P(B_1 \setminus A_1) + P(B_2 \setminus A_2)$. A lemma utolsó állításának igazolásához egy $C \in \mathcal{C}$ halmaznak egy $B \in \mathcal{B}$ halmazzal való jó approximálhatóságáról vezessük be az Ω halmaz részhalmazainak a következő $\mathcal{D}$ osztályát.

$$\mathcal{D} = \{D\colon\ D \in \mathcal{A}, \quad \text{minden } \varepsilon > 0 \text{ számhoz létezik olyan } B \in \mathcal{B} \text{ halmaz, amelyre } P(B\Delta D) \leq \varepsilon\}.$$

Azt kell megmutatni, hogy $\mathcal{C} \subset \mathcal{D}$. Mivel $\mathcal{B} \subset \mathcal{D}$, elég igazolni, hogy $\mathcal{D}$ σ-algebra, mert ez azt jelenti, hogy tartalmazza a $\mathcal{B}$ által generált σ-algebrát.

Nyilvánvaló, hogy $D \in \mathcal{D}$ esetén $\Omega \setminus D \in \mathcal{D}$, mert ha $A \in \mathcal{A}$ a D halmaznak jó közelítése a ρ távolság szerint, akkor a $P((\Omega \setminus D)\Delta(\Omega \setminus A)) = P(D\Delta A)$ azonosság miatt az $\Omega \setminus A \in \mathcal{A}$ halmaz jó közelítése az $\Omega \setminus D$ halmaznak a ρ távolság szerint. Azt kell még belátni, hogy amennyiben $D_n \in \mathcal{D}$, minden $n = 1, 2, \dots$ indexre akkor $D = \bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$.

Ennek bizonyítása érdekében tekintsünk egy rögzített $\varepsilon > 0$ számra egy olyan $N = N(\varepsilon)$ indexet, amelyre a $D^{(N)} = \bigcup_{n=1}^{N} D_n$ halmazra $P(D \setminus D^{(N)}) \leq \frac{\varepsilon}{2}$. Ezenkívül válasszunk minden D_n halmazhoz egy olyan $B_n \in \mathcal{B}$ halmazt, amelyre $P(D_n\Delta B_n) \leq$

$\frac{\varepsilon}{2^{n+1}}$. Ekkor a $B = \bigcup_{n=1}^{N} B_n$ halmazra $B \in \mathcal{B}$. Ezenkívül azt állítom, hogy $P(B\Delta D) \le \varepsilon$. Valóban,

$$P(B\Delta D) \le P(B\Delta D^{(N)}) + P(D \setminus D^{(N)}) \le \sum_{n=1}^{N} P(B_n \Delta D_n) + P(D \setminus D^{(N)}) \le \varepsilon.$$

Mivel ilyen konstrukció minden $\varepsilon > 0$-ra elvégezhető, innen következik, hogy $D \in \mathcal{D}$. A lemmát beláttuk.

A most bizonyított lemma segítségével belátjuk a következő eredményt.

Tétel véges sok értéket felvevő valószínűségi változó jó approximálhatóságáról. *Legyen adva egy $(\Omega, \mathcal{A}, P)$ valószínűségi mező, egy $\mathcal{B} \subset \mathcal{A}$ halmaz algebra, és az Ω halmaznak egy olyan $\mathcal{D}$ véges sok elemből álló partíciója, amely partíció elemei benne vannak a $\mathcal{B}$ algebra által generált $\mathcal{C} = \sigma(\mathcal{B})$ σ-algebrában. Ekkor minden $\varepsilon > 0$ számhoz létezik az Ω halmaznak egy olyan a $\mathcal{B}$ algebra véges sok eleméből álló $\mathcal{E}$ partíciója, amely jól közelíti a $\mathcal{D}$ partíciót a következő értelemben. Ha ξ olyan valószínűségi változó, amelynek adott értéket felvevő nívóhalmazai a $\mathcal{D}$ partíció elemei, ζ olyan valószínűségi változó, amelynek adott értéket felvevő nívóhalmazai az $\mathcal{E}$ partíció elemei, akkor a ξ valószínűségi változónak a ζ valószínűségi változó szerinti $H(\xi|\zeta)$ feltételes entrópiája teljesíti a $H(\xi|\zeta) \le \varepsilon$ egyenlőtlenséget.*

A tétel állításának jobb megértése érdekében tekintsük a következő példát. Legyen az $(\Omega, \mathcal{A}, P)$ valószínűségi mező a $[0, 1)$ intervallum, rajta a Borel σ-algebrával és a Lebesgue mértékkel, mint valószínűségi mértékkel. Tekintsük ezen a téren azt a $\mathcal{B}$ algebrát, amelynek elemei olyan halmazok, amelyek véges sok balról zárt, jobbról nyílt, racionális végpontú intervallum uniójaként állíthatóak elő. Vegyük ezenkívül $[0, 1)$ intervallum $D_1 = [0, \frac{\sqrt{2}}{2})$, $D_2 = [\frac{\sqrt{2}}{2}, 1)$ intervallumokból álló $\mathcal{D}$ partícióját. Ezen partíció elemei nincsenek benne a $\mathcal{B}$ algebrában, csak az általa generált $\mathcal{C}$ σ-algebrában. Ezért egy olyan ξ valószínűségi változó, amelynek nívóhalmazai a D_1 és D_2 halmaz nem tekinthető úgy, mint egy olyan valószínűségi változó, amelynek nívóhalmazai a $\mathcal{B}$ algebrában vannak. De az jól megközelíthető egy ilyen valószínűségi változóval a következő értelemben. Minden $\varepsilon > 0$ számhoz létezik az Ω halmaznak olyan $\mathcal{B}$-beli halmazokból álló véges (akár két elemű) partíciója úgy, hogy egy olyan ζ valószínűségi változóra, amelynek a nívóhalmazai ezen partíció elemei $H(\xi|\zeta) < \varepsilon$. A tétel azt állítja, hogy hasonló eredmény érvényes általánosabb esetben is.

A tétel bizonyítása. Álljon a $\mathcal{D}$ partíció valamely $D_1, \dots, D_r$ elemekből. Feltehetjük, hogy $P(D_i) > 0$ minden $1 \le i \le r$ indexre. Először azt bizonyítom be, hogy minden (az $\varepsilon > 0$, r és $P(D_i) > 0$, $1 \le i \le r$, számoktól függően) elég kicsi $\delta > 0$ számra igaz a következő állítás. Ha $E_1, \dots, E_r$ az Ω halmaz egy olyan partíciója, amelyre $P(D_i \Delta E_i) \le \delta$ minden $1 \le i \le r$ számra, akkor egy olyan ξ, ζ valószínűségi változó párra, amelyek közül a ξ valószínűségi változó nívóhalmazai a D_i, a ζ valószínűségi változó nívóhalmazai pedig az E_i halmazok, $1 \le i \le r$, teljesül a $H(\xi|\zeta) < \varepsilon$ egyenlőtlenség.

Ezen állítás igazolása érdekében vezessük be a $g(x) = x \log x$, ha $x > 0$, $g(0) = 0$, függvényt. Mivel $g(0) = g(1) = 0$, $g(x)$ folytonos függvény, $g(x) \le 0$, ha $0 \le x \le 1$, ezért létezik olyan $\delta_0 > 0$ szám, amelyre $-\frac{\varepsilon}{r} < g(x) \le 0$, ha $0 \le x \le \delta_0$ vagy $1 - \delta_0 \le x \le 1$. A bizonyítandó állítás indoklása azon az észrevételen fog alapulni, hogy ha a $P(D_i \Delta E_i)$ valószínűségek nagyon kicsik minden i indexre, akkor a $P(D_i|E_i)$ feltételes valószínűségek majdnem eggyel, és a $P(D_i|E_j)$, $i \neq j$, feltételes valószínűségek majdnem nullával egyenlőek. Ezért a $g(P(D_i|E_j))$ mennyiségek nagyon kicsik minden (i, j) párra. Mivel a minket érdeklő feltételes entrópia felírható ilyen kifejezések véges sok tagból álló lineáris kombinációjaként, innen könnyen levezethető a kívánt egyenlőtlenség.

A részletes bizonyításban tekintsük az Ω halmaz egy olyan $E_1, \dots, E_r$ particióját, amelyre $P(D_i \Delta E_i) \le \delta$ a $\delta = \frac{\delta_0}{2} \min\limits_{0 \le i \le r} P(D_i)$ számmal minden $1 \le i \le r$ indexre. Ekkor, mivel $P(D_i) \le P(E_i) + P(D_i \Delta E_i) \le P(E_i) + \delta \le P(E_i) + \frac{P(D_i)}{2}$, ezért $P(D_i) \le 2P(E_i)$ minden $1 \le i \le r$ indexre. Innen $P(E_i) - P(D_i \cap E_i) \le P(D_i \Delta E_i) \le \delta \le \delta_0 P(E_i)$, tehát $P(D_i|E_i) \ge 1 - \delta_0$ minden $1 \le i \le r$ indexre, és $P(D_j|E_i) \le 1 - P(D_i|E_i) \le \delta_0$, ha $i \neq j$. Ezért $-\frac{\varepsilon}{r} \le g(P(D_i|E_j)) \le 0$ minden $1 \le i, j \le r$ indexre, és két olyan ξ és ζ valószínűségi változóra, amelyeknek a D_i illetve E_i, $1 \le i \le r$, halmazok a nívóhalmazai

$$H(\xi|\zeta) = -\sum_{i=1}^{r} \sum_{j=1}^{r} P(E_j) g(P(D_i|E_j)) \le \sum_{i=1}^{r} \sum_{j=1}^{r} P(E_j) \frac{\varepsilon}{r} = \varepsilon.$$

Ezért elég belátni, hogy az Ω halmaznak létezik olyan $E_1, \dots, E_r$ a $\mathcal{B}$ algebra elemeiből álló particiója, amelyre $P(D_i \Delta E_i) \le \delta$ minden $1 \le i \le r$ indexre. Ezt igazolandó válasszunk először olyan $\bar{E}_i \in \mathcal{B}$, $1 \le i \le r$, halmazokat, amelyekre $P(D_i \Delta \bar{E}_i) \le \lambda$ minden $1 \le i \le r$ indexre egy később megválasztandó elég kis $\lambda > 0$ számmal. Ez lehetséges az előző lemma szerint. Definiáljuk az $N = \bigcup\limits_{1 \le i,j \le r,\, i \neq j} (\bar{E}_i \cap \bar{E}_j)$ halmazt, és legyen $E_i = \bar{E}_i \setminus N$, ha $1 \le i \le r-1$, és $E_r = \Omega \setminus (\bigcup\limits_{1 \le i \le r-1} E_i)$. Ekkor $E_1, \dots, E_r$ az Ω halmaz egy particiója a $\mathcal{B}$ algebra elemeivel. Továbbá, mivel $P(\bar{E}_i \cap \bar{E}_j) \le P(\bar{E}_i \Delta D_i) + P(\bar{E}_j \Delta D_j)) \le 2\lambda$, ha $i \neq j$, $P(N) \le r(r-1)\lambda$, ezért $P(E_i \Delta \bar{E}_i) \le P(N) \le r(r-1)\lambda$ minden $1 \le i \le r-1$ indexre, ahonnan $P(E_i \Delta D_i) \le P(E_i \Delta \bar{E}_i) + P(\bar{E}_i \Delta D_i) \le r(r-1)\lambda + \lambda$, ha $1 \le i \le r-1$, és $P(E_r \Delta D_r) = P((\Omega \setminus E_r) \Delta (\Omega \setminus D_r)) \le \sum\limits_{i=1}^{r-1} P(E_i \Delta D_i) \le (r-1)[r(r-1) + \lambda]$. Innen következik, hogy ha a $\lambda > 0$ számot elég kicsinek választjuk, akkor $E_1, \dots, E_r$ az Ω halmaz kívánt tulajdonságú particióját adja. A tételt beláttuk.

Következmény. *Legyen adva egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, azon két ξ és η valószínűségi változó, amelyek közül ξ véges sok η pedig vagy véges sok vagy megszámlálhatóan végtelen sok értéket vesz fel. Legyen ezenkívül a ξ valószínűségi változó mérhető az η valószínűségi változó és T shift operátor által generált $\sigma(T, \eta)$ σ-algebrára nézve. Ekkor minden $\varepsilon > 0$ számhoz létezik olyan $M = M(\varepsilon, \xi, \eta)$ pozitív egész szám, amelyre $H(\xi|T^{-M}\eta, \dots, T^0\eta, \dots, T^M\eta) \le \varepsilon$.*

Bizonyítás. A bizonyításban az előző tételt alkalmazzuk úgy, hogy az Ω halmaz $\mathcal{D}$ particiójának a ξ valószínűségi változó $\{\omega\colon \xi(\omega) = x_j\}$ alakú nívóhalmazait választjuk, és az alább definiált $\mathcal{B}$ halmaz algebrával dolgozunk.

Jelölje Y az η valószínűségi változó értékkészletét, és adva két $m \geq 1$ és $n \geq 1$ szám definiáljuk az $Y^{(m,n)} = \{(y_{j_{-m}}, \dots, y_{j_n})\colon\ y_{j_s} \in Y,\ -m \leq s \leq n\}$ halmazt. Adva egy $U \subset Y^{(m,n)}$ halmaz, definiáljuk az U halmaznak az $\eta(\omega)$ valószínűségi változó és a T shift operátor hatványai által meghatározott ősképét a

$$\mathbf{T}_{m,n}U = \{\omega\colon\ (T^{-m}\eta(\omega), \dots, T^n\eta(\omega)) \in U\}$$

képlet segítségével, és legyen $\mathcal{U}_{m,n} = \{\mathbf{T}_{m,n}U\colon\ U \subset Y^{(m,n)}\}$. A $\mathcal{B}$ halmazrendszert a $\mathcal{B} = \bigcup_{1 \leq m,n < \infty} \mathcal{U}_{m,n}$ képlettel definiáljuk. Nem nehéz belátni, hogy $\mathcal{B}$ halmaz algebra, és az általa generált $\sigma(\mathcal{B})$ σ-algebra megegyezik a $\sigma(T, \eta)$ σ-algebrával. Mivel feltételeink szerint ξ $\sigma(T, \eta)$ mérhető, ezért az előző tétel alapján minden $\varepsilon > 0$ számra létezik az Ω halmaznak olyan a $\mathcal{B}$ algebra elemeiből álló $\mathcal{E}$ véges particiója, amelyre igaz, hogy egy olyan ζ valószínűségi változóra, amelynek a a nívóhalmazai az $\mathcal{E}$ partició elemei $H(\xi|\zeta) \leq \varepsilon$.

Mivel ζ véges sok értéket vesz fel, és minden értékét egy olyan halmazon veszi fel, amely eleme a $\mathcal{U}_{m,n}$ halmazosztálynak, ha $m \geq m_0$ és $n \geq n_0$ alkalmas m_0 és n_0 számokkal, ezért létezik olyan M szám, és olyan $g(u_{-M}, \dots, u_0, \dots, u_M)$ függvény, amelyekre

$$\zeta = g(T^{-M}\eta, \dots, T^0\eta, \dots, T^M\eta).$$

Ezért, illetve a feltételes entrópia tulajdonságai alapján

$$\begin{aligned} \varepsilon \geq H(\xi|\zeta) &\geq H(\xi|\zeta, T^{-M}\eta, \dots, T^0\eta, \dots, T^M\eta) \\ &= H(\xi|T^{-M}\eta, \dots, T^0\eta, \dots, T^M\eta). \end{aligned}$$

Mivel ilyen konstrukciót minden $\varepsilon > 0$ számra tudunk csinálni a következményt beláttuk.

A most igazolt következmény segít az alábbi bizonyításban.

Az egy dinamikus rendszerben definiált valószínűségi változók entrópiájának összehasonlításáról szóló tétel bizonyítása. Rögzítsünk egy $\varepsilon > 0$ számot, és válasszunk egy olyan $M > 0$ egész számot, amelyre $H(\xi|T^{-M}\eta, \dots, T^0\eta, \dots, T^M\eta) \leq \varepsilon$. Az előbb megfogalmazott *Következmény* eredménye szerint ilyen M szám létezik. Ilyen választással érvényesek a következő becslések.

$$\begin{aligned} H(T^0\xi, \dots, T^{n-1}\xi) &\leq H(T^0\xi, \dots, T^{n-1}\xi, T^{-M}\eta, \dots, T^{n-1+M}\eta) \\ &= H(T^0\xi, \dots, T^{n-1}\xi|T^{-M}\eta, \dots, T^{n-1+M}\eta) + H(T^{-M}\eta, \dots, T^{n-1+M}\eta), \end{aligned}$$

és

$$\begin{aligned} H(T^0\xi, \dots, T^{n-1}\xi|T^{-M}\eta, \dots, T^{n-1+M}\eta) &\leq \sum_{j=0}^{n-1} H(T^j\xi|T^{-M}\eta, \dots, T^{n-1+M}\eta) \\ = \sum_{j=0}^{n-1} H(\xi|T^{-M-j}\eta, \dots, T^{n-1+M-j}\eta) &\leq nH(\xi|T^{-M}\eta, \dots, T^M\eta) \leq n\varepsilon. \end{aligned}$$

Innen

$$\frac{1}{n}H(T^0\xi,\dots,T^{n-1}\xi) \le \frac{1}{n}H(T^{-M}\eta,\dots,T^{n-1+M}\eta) + \varepsilon$$

minden $n \ge 0$ számra, ahonnan $n \to \infty$ határátmenettel kapjuk, hogy $H(T,\xi) \le H(T,\eta) + \varepsilon$. Mivel ez az egyenlőtlenség minden $\varepsilon > 0$ számra igaz, innen következik a tétel állítása.

Bebizonyítottuk egy dinamikus rendszerben definiált valószínűségi változók entrópiájának összehasonlításáról szóló tétel egy következményét, amely lehetővé tette bizonyos dinamikus rendszerek entrópiájának a kiszámolását. Ismertetem ennek az eredménynek egy enyhe általánosítását, amely hasonló eredményt fogalmaz meg nem feltétlenül véges sok értéket felvevő valószínűségi változók esetében. A bizonyításban szükségünk van az alábbi lemmára, amely egy dinamikus rendszer entrópiájának egy az eredeti definiciótól kissé eltérő jellemzését adja.

Lemma az entrópia jellemzéséről. *Egy $(\Omega,\mathcal{A},P,T)$ invertálható dinamikus rendszer T shift transzformációjának az entrópiáját ki lehet fejezni az (5.6) kifejezéshez hasonló módon, úgy mint*

$$H(T) = \sup_{\xi} H(T,\xi),$$

ahol a szuprémumot az összes olyan $\mathcal{A}$ mérhető és véges sok vagy megszámlálhatóan végtelen sok értéket felvevő ξ valószínűségi változóra vesszük, amelyekre $H(\xi) < \infty$.

Bizonyítás. Legyen ξ olyan véges vagy megszámlálhatóan végtelen sok értéket felvevő $\mathcal{A}$ mérhető valószínűségi változó, amelyre $H(\xi) < \infty$. A lemma bizonyításához elegendő belátni, hogy minden $\varepsilon > 0$ számhoz létezik olyan véges sok értéket felvevő $\eta = \eta(\varepsilon)$ valószínűségi változó, amelyre $H(T,\xi) \le H(T,\eta) + \varepsilon$. Azt mutatom meg, hogy létezik olyan véges sok értéket felvevő $\mathcal{A}$ mérhető η valószínűségi változó, amelyre

$$H(\xi|T\xi,\dots,T^n\xi) \le H(\eta|T\eta,\dots,T^n\eta) + \varepsilon$$

minden n számra, mert innen $n \to \infty$ határátmenettel megkapjuk a kívánt egyenlőtlenséget.

Ezen állítás igazolása érdekében jegyezzük meg, hogy mint az első fejezetben láttuk létezik olyan véges sok értéket felvevő $g(x)$ függvény, amelyre az $\eta = g(\xi)$ valószínűségi változó teljesíti a $H(\xi) \le H(\eta) + \varepsilon$ egyenlőtlenséget. Ebből az egyenlőtlenségből az is következik, hogy $H(\xi|\eta) = H(\xi,\eta) - H(\eta) = H(\xi) - H(\eta) \le \varepsilon$, és

$$\begin{aligned} H(\xi|T\xi,\dots,T^n\xi) &= H(\xi,\eta|T\xi,\dots,T^n\xi) = H(\xi|\eta,T\xi,\dots,T^n\xi) + H(\eta|T\xi,\dots,T^n\xi) \\ &\le H(\xi|\eta) + H(\eta|T\eta,\dots,T^n\eta) \le H(\eta|T\eta,\dots,T^n\eta) + \varepsilon, \end{aligned}$$

és ezt kellett belátnunk. E számolásban kihasználtuk, hogy mivel a $(T\eta,\dots,T^n\eta) = (g(T\xi),\dots,g(T^n\xi))$ véletlen vektor függvénye a $(T\xi,\dots,T^n\xi)$ véletlen vektornak, ezért $H(\eta|T\xi,\dots,T^n\xi) \le H(\eta|T\eta,\dots,T^n\eta)$. Hasonlóan $H(\xi|\eta,T\xi,\dots,T^n\xi) \le H(\xi|\eta)$. A lemmát beláttuk.

Megfogalmazom az egy dinamikus rendszerben definiált valószínűségi változók entrópiájának összehasonlításáról szóló tétel következményének az alábbi, az eredetinél kissé élesebb változatát.

Tétel az entrópia egy tulajdonságáról *Legyen ξ egy olyan véges vagy megszámlálhatóan végtelen sok értéket felvevő valószínűségi változó egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszerben, amelyre $H(\xi) < \infty$, és $\sigma(T, \xi) = \mathcal{A}$. Ekkor $H(T) = H(T, \xi)$.*

Bizonyítás. Az előző lemma alapján $H(T, \xi) \leq H(T)$. Másrészt azt is láttuk, hogy mivel $\sigma(T, \xi) = \mathcal{A}$, ezért $H(T, \eta) \leq H(T, \xi)$ tetszőleges véges sok értéket felvevő és $\mathcal{A}$ mérhető η valószínűségi változóra. Ezért $H(T, \xi) = H(T)$.

Megjegyzés. Az előző tételben megengedtük azt, hogy ξ végtelen sok értéket vegyen fel, de megköveteltük a $H(\xi) < \infty$ reláció teljesülését. E nélkül a feltétel nélkül az állítás nem igaz. Erre mutattak példát az irodalomban.

A fenti vizsgálatokban csak invertálható dinamikus rendszerekkel foglalkoztunk. Nem nehéz a most bizonyított tételhez hasonló eredményt bizonyítani általános, nem feltétlenül invertálható dinamikus rendszerekre is, de ezek jelentősége kisebb.

Általános esetben az *Egy dinamikus rendszerben definiált valószínűségi változók entrópiájának összehasonlításáról* szóló tételnek azt a feltételét, amely szerint a ξ valószínűségi változó $\sigma(T, \eta)$ mérhető újra kell értelmezni. Az általános esetben a $\sigma(T, \eta)$ σ-algebrát úgy definiáljuk, mint a $T^n\xi$, $n = 0, 1, 2 \ldots$, valószínűségi változók által generált σ-algebrát, hiszen a T^{-n}, $n = 1, 2, \ldots$ operátorokat nem mindig tudjuk definiálni. Ez erősebb megszorítást jelent, mint az invertálható dinamikus rendszerek esetében megfogalmazott feltétel.

Másrészt minden dinamikus rendszernek létezik úgynevezett természetes kiterjesztése, amely invertálható dinamikus rendszer. Az, hogy két kiterjesztett dinamikus rendszer természetes kiterjesztése izomorf legyen egymással az eredeti dinamikus rendszerek izomorfiájának szükséges, de nem elégséges feltétele. Így például, ha veszünk két féloldali Bernoulli rendszert, akkor ezek természetes kiterjesztése két Bernoulli rendszer ugyanazokkal a paraméterekkel. A féloldali Bernoulli rendszerek izomorfiájából azonnal következik azok természetes kiterjesztésének az izomorfiája, de ennek az állításnak a megfordítása nem igaz.

Kiegészítés. *Bernoulli rendszerek vizsgálatában felmerült unitér operátorok izomorfiájának a vizsgálata.*

Invertálható dinamikus rendszerek izomorfiájának vizsgálatában felmerült a következő gondolat. Ha adva van egy $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, akkor természetes módon definiálhatjuk az $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn definiált négyzetesen integrálható függvények által meghatározott $L_2(\Omega, \mathcal{A}, P)$ Hilbert téren a következő U unitér operátort: $Uf(x) = f(Tx)$ minden $f \in L_2(\Omega, \mathcal{A}, P)$ függvényre. Nem nehéz belátni, hogy Létezik az U^{-1} operátor, amit az $U^{-1}f(x) = f(T^{-1}(x)$ képlet határoz meg. Mivel T mértéktartó leképezés, ezért U normatartó, és unitér operátor. Továbbá, ha két dinamikus rendszer izomorf, akkor az általuk meghatározott U unitér operátorok

is izomorfak, azaz ha $(\Omega_1, \mathcal{A}_1, P_1, T_1)$ és $(\Omega_2, \mathcal{A}_2, P_2, T_2)$ két izomorf invertálható dinamikus rendszer, és U_1 és U_2 a nekik megfelelő unitér operátor, akkor az $L_2(\Omega_1, \mathcal{A}_1, P_1)$ és $L_2(\Omega_2, \mathcal{A}_2, P_2)$ Hilbert tereknek létezik egy olyan G izomorfiája, amelyre $G(U_1(f)) = U_2(G(f))$ minden $f \in L_2(\Omega_1, \mathcal{A}_1, P_1)$ függvényre.

Be lehet bizonyítani ezen eredmény segítségével, hogy bizonyos dinamikus rendszerek nem izomorfak. Felmerült az a kérdés, hogy ez az eredmény segítséget nyújt-e Bernoulli rendszerek izomorfiájának a vizsgálatában. Kiderült, hogy a válasz nemleges, mert bármely két Bernoulli rendszerre az általuk a fenti módon definiált U unitér operátorok izomorfak egymással. Ismertetem ennek az állításnak a bizonyítását.

Az állítás pontos megfogalmazása érdekében bevezetem a következő jelöléseket. Rögzítsünk egy pozitív egész r számot és r darab olyan $p_j > 0$, $1 \le j \le r$, számot, amelyekre $\sum_{j=1}^{r} p_j = 1$. Definiáljuk segítségükkel azt az $(\bar{X}, \mathcal{A}, \bar{P})$ valószínűségi mezőt, amelyre $\bar{X} = \{1, \dots, r\}$, $\mathcal{A}$ az $\bar{X}$ halmaz összes részhalmazából áll, és $\bar{P}(\{j\}) = p_j$, $1 \le j \le r$. Tekintsük ennek a valószínűségi mezőnek $(\bar{X}_l, \mathcal{A}_l, \bar{P}_l)$ példányait minden $l = 0, \pm 1, \pm 2, \dots$ egész számra, és vegyük ezek $(X, \mathcal{A}, P)$ direkt szorzatát. Azaz álljon az X halmaz az összes $x = (\dots, i_{-1}, i_0, i_1, \dots)$, $i_j \in \{1, \dots, r\}$ minden $-\infty < j < \infty$ indexre, sorozatból, legyen $\mathcal{A}$ a $\mathcal{A}_l$ σ-algebrák, P a $\bar{P}_l$ mértékek direkt szorzata az X téren, $l = \dots, -1, 0, 1, \dots$. Speciálisan minden $m \ge 0$, $n \ge 0$ számpárra $P(x\colon \{x = (\dots, i_{-1}, i_0, i_1, \dots)\colon\ i_l = j_l,$ ha $-m \le l \le n\}) = \prod_{l=-m}^{n} p_{j_l}$, ha $1 \le j_l \le r$ minden $-m \le l \le n$ indexre. Definiáljuk továbbá egy $x = (\dots, i_{-1}, i_0, i_1, i_2, \dots) \in X$ pont Tx eltoltját a $Tx = (\dots, i_{-2}, i_{-1}, i_0, i_1, \dots)$ képlettel, és adva egy $f(x)$ (mérhető) függvény az $(X, \mathcal{A}, P)$ téren legyen $Uf(x) = f(Tx)$. Nem nehéz belátni, hogy ha az U transzformáció értelmezési tartományát megszorítjuk az $(X, \mathcal{A}, P)$ téren négyzetesen integrálható függvényekre, akkor U egy unitér transzformáció az $L_2(X, \mathcal{A}, P)$ Hilbert térben. Igaz továbbá az alábbi tétel.

Tétel unitér operátorok izomorfiájáról. *Tekintsük minden $r = 1, 2, \dots$ számra és $p_j > 0$, $\sum_{j=1}^{r} p_j = 1$, $1 \le j \le r$, vektorra az előbb bevezetett $L_2(X, \mathcal{A}, P)$ Hilbert teret és a rajta definiált U unitér operátort. Ezek az operátorok izomorfak az r és p_j, $1 \le j \le r$, számok tetszőleges választása esetén.*

A bizonyítás lényeges része az $L_2(X, \mathcal{A}, P)$ Hilbert tér egy olyan ortonormált bázisának a megadása, amelyben az U operátor hatása jól látható. Olyan bázist fogunk konstruálni, amelyik segít abban, hogy az $L_2(X, \mathcal{A}, P)$ teret felbontsuk U-invariáns alterek direkt összegére. Ennek érdekében tekintsük az $(X, \mathcal{A}, P)$ valószínűségi mező definiciója során bevezetett $(\bar{X}, \mathcal{A}, \bar{P})$ teret, illetve az ezen téren értelmezett függvények r-dimenziós $K(r)$ Euklideszi terét a $(\varphi, \psi) = \sum_{j=1}^{r} p_j x_j y_j$, ha $\varphi = (x_1, \dots, x_r) \in K(r)$, $\psi = (y_1, \dots, y_r) \in K(r)$ skalár szorzatttal. Válasszunk a $K(r)$ Euklideszi térben egy olyan $(\varphi_1, \dots, \varphi_r)$ ortonormált bázist, amelynek első eleme $\varphi_1 = (1, \dots, 1)$, az $\{1, \dots, r\}$ halmazon definiált azonosan 1 függvény. Jelölje V az összes olyan $v = (v_l, -\infty < l < \infty)$ sorozat terét, amelyre $v_l \in \{1, \dots, r\}$ minden $-\infty < l < \infty$ indexre, és a v

sorozat elemei csak véges sok 1-től különböző koordinátát tartalmaznak. Bevezetem a következő u_v a $v \in V$ sorozatokkal indexelt az $(X, \mathcal{A}, P)$ téren definiált függvényeket: $u_v(\dots, i_{-1}, i_0, i_1, \dots) = \prod_{l=-\infty}^{\infty} \varphi_{v_l}(i_l)$. Ezek a szorzatok jól definiáltak, mert csak véges sok tényezőből állnak. Ugyanis $\varphi_{v_l}(i_l) = \varphi_1(i_l) = 1$ véges sok l index kivételével. A következő lemmát fogom bebizonyítani.

Lemma ortonormált bázisok létezéséről. *Az előbb definiált $u_v(\cdot)$, $v \in V$, függvények együttese teljes ortonormált rendszert alkot az $L_2(X, \mathcal{A}, P)$ térben.*

A lemma bizonyítása. Könnyen látható, hogy az $u_v(\cdot)$, $v \in V$, függvények ortonormáltak. Ugyanis véve két $v = (\dots, v_{-1}, v_0, v_1, \dots) \in V$ és $\bar{v} = (\dots, \bar{v}_{-1}, \bar{v}_0, \bar{v}_1, \dots) \in V$ vektort és egy $x \in X$ pontot, felírhatjuk az $u_v(x)u_{\bar{v}}(x) = \prod_{l=-\infty}^{\infty} \varphi_{v_l}(x(l))\varphi_{\bar{v}_l}(x(l))$ azonosságot, ahol $x(l) = i_l$, ha $x = (\dots, -i_1, i_0, i_0, \dots)$. Tehát az $u_v(x)u_{\bar{v}}(x)$ kifejezés faktorizálódik. Sőt ez a szorzat csak véges sok tagból áll, mert $\varphi_{v_l}(x) = \varphi_1(x) \equiv 1$, és $\varphi_{\bar{v}_l}(x) = \varphi_1(x) \equiv 1$ véges sok l index kivételével. A P valószínűségi mérték szintén faktorizálódik, és innen azt kapjuk, hogy

$$(u_v, u_{\bar{v}}) = \int u_v(x)u_{\bar{v}}(x)P(\,dx) = \prod_{l=-\infty}^{\infty} \int \varphi_{v_l}(x(l))\varphi_{\bar{v}_l}(x(l))\bar{P}_l(\,dx(l))$$

$$= \prod_{l=-\infty}^{\infty} \left(\sum_{i=1}^{r} p_i \varphi_{v_l}(i)\varphi_{\bar{v}_l}(i) \right) = \prod_{l=-\infty}^{\infty} \delta(v_l, \bar{v}_l) = \delta(v, \bar{v}),$$

ahol $\delta(i, j) = 0$, ha $i \neq j$, $\delta(i, j) = 1$, ha $i = j$, és hasonlóan $\delta(v, \bar{v}) = 0$, ha $v \neq \bar{v}$, és $\delta(v, \bar{v}) = 1$, ha $v = \bar{v}$.

Adva két $m > 0$ és $n > 0$ egész szám jelölje $Q_{m,n}$ az $L_2(X, \mathcal{A}, P)$ Hilbert tér azon alterét, amely az olyan $u(x)$, $x \in X$, függvényekből áll, amelyek az $x = (\dots, i_{-1}, i_0, i_1, \dots)$ argumentumnak csak az i_j, $-m \leq j \leq n$, koordinátáitól függnek, és jelölje $V_{m,n} \subset V$ azon $v = (v_l, -\infty < l < \infty) \in V$ sorozatok halmazát, amelyekre $v_l = 1$, ha $l < -m$ vagy $l > n$. Ekkor az $u_v(\cdot)$, $v \in V_{m,n}$ függvények (r^{n+m+1} elemből álló) rendszere egy teljes ortonormált rendszert alkot a $Q_{m,n}$ (r^{n+m+1} dimenziós) Euklideszi térben. Ezért annak bizonyításához,hogy az $u_v(\cdot)$, $v \in V$, függvények teljes ortonormált rendszert alkotnak az $L_2(X, \mathcal{A}, P)$ Hilbert térben elég azt megmutatni, hogy a $Q_{m,n}$ alterek $\bigcup_{0<m,n<\infty} Q_{m,n}$ uniója mindenütt sűrű halmaz az $L_2(X, \mathcal{A}, P)$ Hilbert térben. Sőt, ezt arra az állításra lehet redukálni, hogy a $\mathcal{B} = \bigcup_{0<m,n<\infty} \mathcal{A}_{m,n}$ halmaz, ahol $\mathcal{A}_{m,n}$ a $Q_{m,n}$ altér függvényeinek nívóhalmazai által generált σ-algebra, sűrű az $\mathcal{A}$ σ-algebrában. Ez azt jelenti, hogy minden $\varepsilon > 0$ számhoz és $A \in \mathcal{A}$ halmazhoz létezik olyan $B \in \mathcal{B}$ halmaz, amelyre $P(A\Delta B) < \varepsilon$. (Itt $A\Delta B$ az A és B halmaz szimmetrikus differenciáját jelöli.) Ugyanis innen következik, hogy véges sok $A_i \in \mathcal{A}$ halmaz indikátor függvényének a lineáris kombinációja benne van a $\bigcup_{0<m,n<\infty} Q_{m,n}$ függvényhalmaznak (az L_2 norma szerinti) lezártjában. De akkor az ilyen alakú függvények lezártja, ami egyenlő $L_2(X, \mathcal{A}, P)$ Hilbert térrel, szintén benne van ebben a függvényosztályban.

Viszont az, hogy a $\mathcal{B}$ halmazosztály sűrű a $\mathcal{A}$ σ-algebrában következik a *Lemma σ-algebra elemeinek jó approximációjáról* eredményéből. Ugyanis $\mathcal{B}$ halmaz algebra, és $\mathcal{A}$ az általa generált σ-algebra. A lemmát beláttuk.

Belátjuk a tételt ezen lemma segítségével.

Az unitér operátorok izomorfiájáról szóló tétel bizonyítása. Álljon a $V_0 \subset V$ halmaz azon $v = (v_l,\ -\infty < l < \infty) \in V$ sorozatokból, amelyekre $v_l = 1$, ha $l < 0$, és $v_0 \neq 1$. A V_0 halmaz megszámlálható, ezért megadható $V_0 = \{v^{(1)}, v^{(2)}, \dots\}$ alakban. Definiáljuk az $u_n = u_{v^{(n)}} \in L_2(X, \mathcal{A}, P)$, $n = 1, 2, \dots$, függvényeket, és ezenkívül az $u_0 = u_{v^{(0)}}$ függvényt, ahol $v^{(0)} = (v_l = 1,\ -\infty < l < \infty) \in V$, azaz a csupa 1 koordinátából álló $v^{(o)} \in V$ sorozat. Definiáljuk a $T^k v$ (shift) transzformációt minden $v \in V$ sorozatra és $-\infty < k < \infty$ számra a $T^k v = (v_{l-k},\ -\infty < l < \infty)$, ha $v = (v_l,\ -\infty < l < \infty)$ képlet segítségével. Továbbá vezessük be a következő jelöléseket. Definiáljuk az $u_{n,k} = u_{T^k v^{(n)}}$ függvényeket minden $n = 1, 2, \dots$ és $-\infty < k < \infty$ számpárra. (Tehát speciálisan $u_n = u_{n,0}$). Ekkor nem nehéz belátni, hogy mivel $V \setminus \{v^{(0)}\} = \bigcup_{k=-\infty}^{\infty} \{T^k v,\ v \in V_0\}$, és ebben az unióban minden $v \in V \setminus \{v^{(0)}\}$ vektor pontosan egyszer van felsorolva. Ezért az u_0, $u_{n,k}$, $1 \le n < \infty$, $-\infty < k < \infty$, függvényrendszer megegyezik az előző lemmában tekintett $u_v(\cdot)$, $v \in V$, függvények rendszerével, és teljes ortonormált rendszert alkot. Továbbá $Uu_0 = u_0$, és $Uu_{n,k} = u_{n,k+1}$ minden $1 \le n < \infty$ és $-\infty < k < \infty$ indexekre. Az U transzformáció ezen jellemzésének segítségével könnyen be tudjuk látni a tételt.

Tekintsünk egy H szeparábilis Hilbert teret valamely ortonormált bázissal, amelynek elemeit indexeljük az előzőleg vizsgált esethez hasonlóan úgy, hogy g_0, $g_{n,k}$, $1 \le n < \infty$, $-\infty < k < \infty$. Definiáljuk a H Hilbert téren a következő $\bar{U}$ operátort: $\bar{U}g_0 = g_0$, $\bar{U}g_{n,k} = g_{n,k+1}$, ha $1 \le n < \infty$, $-\infty < k < \infty$. Nem nehéz belátni, hogy $\bar{U}$ unitér operátor. ($\bar{U}^{-1}g_{n,k} = g_{n,k-1}$, és $\bar{U}^{-1}g_0 = g_0$.) Továbbá a G: $u_{n,k} \to g_{n,k}$, $1 \le n < \infty$, $-\infty < k < \infty$ és G: $u_0 \to g_0$ leképezés az $L_2(X, \mathcal{A}, P)$ térnek és a H Hilbert térnek egy olyan izomorfiáját definiálja, amely az U és $\bar{U}$ unitér operátorok izomorfiáját is biztosítja. Mivel az ebben az izomorfiában szereplő H Hilbert tér és $\bar{U}$ operátor megválasztása nem függött az $(X, \mathcal{A}, P, T)$ Bernoulli rendszer definiciójában szereplő r és p_j, $1 \le j \le r$, paraméterektől, innen következik a tétel állítása.

Megjegyzés. Kidolgozták a Hilbert téren definiált unitér (vagy önadjungált, vagy általánosabban úgynevezett normális) operátorok spektrál elméletét, amely jól leírja az ilyen operátorok szerkezetét. Értsük meg, hogyan írja le ezen elmélet az egy Bernoulli rendszerben előbb definiált és vizsgált U unitér operátort. Annak érdekében, hogy jobban megértsük egy Hilbert téren definiált operátor viselkedését, érdemes a Hilbert teret felbontani az operátor invariáns altereinek direkt összegére. Az előző tételben tulajdonképpen egy ilyen felbontást konstruáltunk.

Az $L_2(X, \mathcal{A}, P)$ teret felbontottuk $K_0, K_1, \dots$ ortogonális U-invariáns alterek összegére a következő módon. K_0 az u_0 vektor által generált (1 dimenziós) altér, K_n pedig az $u_{n,k}$, $k = 0, \pm 1, \pm 2, \dots$, vektorok által generált altér minden $n = 1, 2, \dots$ számra. Az U operátor megszorítását a K_n altérre az $Uu_{n,k} = u_{n,k+1}$, $k = 0, \pm 1, \pm 2 \dots$, képlet definiálja. Érdemes megjegyezni, hogy az U operátor megszorítása valamelyik K_n

altérre izomorf a következő $\bar{U}$ operátorral. Tekintsük a $G = L_2([0,1), \mathcal{B}, \lambda)$ Hilbert teret, ahol $\mathcal{B}$ a Borel σ-algebra, λ pedig a Lebesgue mérték a $[0,1)$ intervallumon. Definiáljuk az $\bar{U}$ operátort, mint az $f(x) = e^{i2\pi x}$ függvénnyel való szorzást a G Hilbert térben, azaz legyen $\bar{U}g(x) = e^{i2\pi x}g(x)$, ha $g(x) \in L_2([0,1), \mathcal{B}, \lambda)$. Ekkor felhasználva, hogy a $g_k(x) = e^{i2\pi kx}$, $k = 0, \pm 1, \dots$, függvények egy teljes ortonormált rendszert alkotnak az $L_2([0,1), \mathcal{B}, \lambda)$ térben, és $\bar{U}g_k = g_{k+1}$ minden $k = 0, \pm 1, \pm 2, \dots$ indexre, meg tudjuk mutatni, hogy az $u_{n,k} \to g_k$, $k = 0, \pm 1, \pm 2, \dots$, leképezés izomorfiát létesít az K_n és G Hilbert terek és a rajtuk definiált U és $\bar{U}$ unitér operátorok között. Ezt a tényt felhasználva a következő az $L_2(X, \mathcal{A}, P)$ Hilbert térrel és U unitér operátorral izomorf rendszert tudjuk definiálni. Vegyük az $L_2([0,\infty), \mathcal{B}, \lambda^+)$ Hilbert teret, ahol $\mathcal{B}$ a Borel σ-algebra $[0,\infty)$ félegyenesen, λ^+ a Lebesgue mérték a $[0,\infty)$ félegyenesen plusz a $\{0\}$ pontba koncentrált egység mérték. Akkor az $f(x) = e^{i2\pi x}$ függvénnyel való szorzás az $L_2([0,\infty), \mathcal{B}, \lambda^+)$ térben izomorf az U operátorral. Ez az állítás tekinthető úgy is, mint az U operátor implicit módon megadott spektrál előállítása.

6. A Shannon–McMillan–Breiman tétel.

Ebben a fejezetben a Shannon–McMillan–Breiman tételt, az információelmélet egyik klasszikus eredményét ismertetem. Ez az eredmény nagy n számokra hasznos jellemzést ad egy véges vagy megszámlálható sok értéket felvevő, diszkrét idejű stacionárius sztochasztikus folyamat n hosszúságú szeleteinek tipikus értékeire. A tétel megfogalmazása érdekében felidézek előbb néhány fontos fogalmat és eredményt.

Diszkrét idejű stacionárius sztochasztikus folyamat definiciója. *Legyen adva ξ_n, $-\infty < n < \infty$, valószínűségi változók egy sorozata egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Azt mondjuk, hogy ez a sorozat diszkrét idejű stacionárius sztochasztikus folyamat, ha minden $-\infty < n_1 < n_2 < \dots < n_k < \infty$ és $m \geq 1$ egész számokra a $(\xi_{n_1}, \xi_{n_2}, \dots, \xi_{n_k})$ és $(\xi_{n_1+m}, \xi_{n_2}, \dots, \xi_{n_k+m})$ véletlen vektorok eloszlása megegyezik.*

A diszkrét idejű stacionárius sztochasztikus folyamatok és az invertálható dinamikus rendszerek szoros kapcsolatban állnak egymással. Ha veszünk egy dinamikus rendszerben egy $\xi(\omega)$ valószínűségi változót és annak összes $\xi_n(\omega) = T^n\xi(\omega) = \xi(T^n(\omega))$, $n = 0, \pm 1, \pm 2, \dots$, eltoltját, akkor ezen eltoltak sorozata egy diszkrét idejű stacionárius sztochasztikus folyamat. Ennek igazolásához azt kell megérteni, hogy

$$\{\omega\colon (\xi(T^{n_1+m}(\omega), \dots, T^{n_k+m}(\omega)) \in A\} = T^{-m}\{\omega\colon (\xi(T^{n_1}(\omega), \dots, T^{n_k}(\omega)) \in A\},$$

és T mértéktartó leképezés. Azért, hogy az előző állítás megfordítását is megfogalmazzam, definiálni fogok úgynevezett speciális dinamikus rendszereket, amelyek invertálható dinamikus rendszerek, és definiálni fogok minden speciális dinamikus rendszerre $\bar{\xi}_n$, $\bar{\xi}_n = T^n\bar{\xi}_0$, $-\infty < n < \infty$, valószínűségi változóknak egy sorozatát, amelyet e rendszer által indukált sorozatnak fogok nevezni. A fenti képletben T^n a tekintett speciális dinamikus rendszer shift transzformációjának az n-ik hatványa. Meg fogom mutatni, hogy minden ξ_n, $-\infty < n < \infty$, diszkrét idejű stacionárius sztochasztikus folyamathoz tudunk egy olyan speciális dinamikus rendszert konstruálni, amelyre az általa indukált

$\bar{\xi}_n$, $\bar{\xi}_n = T^n\bar{\xi}_0$, valószínűségi változók sorozatának és az eredeti ξ_n, $-\infty < n < \infty$, valószínűségi változó sorozatnak az eloszlása megegyezik. Az állítás pontos megfogalmazásának az érdekében bevezetem a következő definiciót.

Speciális dinamikus rendszerek és általuk indukált valószínűségi változók sorozatának a definiciója. *Jelölje $R^{\pm\infty}$ az R számegyenes (pozitív, negatív vagy nulla) egész számokkal indexelt példányainak a direkt szorzatát, azaz az $x = (\dots, x_{-1}, x_0, x_1, \dots)$ két irányban végtelen, valós számokból álló sorozatok halmazát, és jelölje $\mathcal{B}^{\pm\infty}$ a Borel σ-algebrát az $R^{\pm\infty}$ halmazon. Vezessük be a (baloldali eltolást jelentő)*

$$Tx = T(\dots, x_{-1}, x_0, x_1, \dots) = (\dots, x_{-2}, x_{-1}, x_0, \dots), \quad x \in R^{\pm\infty}$$

shift transzformációt az $R^{\pm\infty}$ téren. Egy $\bar{P}$ valószínűségi mértéket az $(R^{\pm\infty}, \mathcal{B}^{\pm\infty})$ téren T invariánsnak nevezünk, ha $\bar{P}(T^{-1}(A)) = \bar{P}(A)$ minden $A \in \mathcal{B}^{\pm\infty}$ halmazra. Egy $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ rendszert a fent definiált $R^{\pm\infty}$, $\mathcal{B}^{\pm\infty}$, $\bar{P}$ és T menyiségekkel, ahol $\bar{P}$ T invariáns valószínűségi mérték az $(R^{\pm\infty}, \mathcal{B}^{\pm\infty})$ téren speciális dinamikus rendszernek nevezünk. Adva egy $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszer, a $\bar{\xi}_n(x) = x_n$, $-\infty < n < \infty$, $x = (\dots, x_{-1}, x_0, x_1, \dots)$, valószínűségi változók sorozatát a rendszer által indukált valószínűségi változók sorozatának fogjuk nevezni. (Nyilván $\bar{\xi}_n(x) = \bar{\xi}_0(T^n x)$) minden $x \in R^{\pm\infty}$ pontra és $n = 0, \pm 1, \dots$ számra.)

Nem nehéz látni, hogy egy speciális dinamikus rendszer invertálható dinamikus rendszer. Továbbá igaz a következő lemma.

Lemma diszkrét idejű stacionárius sztochasztikus folyamatok és invertálható dinamikus rendszerek kapcsolatáról. *Legyen $(\Omega, \mathcal{A}, P, T)$ invertálható dinamikus rendszer, és ξ egy az $(\Omega, \mathcal{A}, P)$ mezőn értelmezett valószínűségi változó. Ekkor a $\xi_n = T^n\xi$, $n = \dots, -1, 0, 1, \dots$, valószínűségi változók sorozata egy diszkrét idejű stacionárius sztochasztikus folyamat.*

Megfordítva, minden ξ_n, $-\infty < n < \infty$, diszkrét idejű stacionárius sztochasztikus folyamathoz létezik olyan $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszer, amelyre a speciális dinamikus rendszer által indukált $\bar{\xi}_n$, $-\infty < n < \infty$, valószínűségi változók sorozatának és a ξ_n, $-\infty < n < \infty$, valószínűségi változó sorozatnak az eloszlása megegyezik.

Bizonyítás. A lemma első fele nyilvánvaló. A lemma második felének a bizonyításában definiálni kell a lemma feltételeit kielégítő $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszert. Ebben a definicióban a $\bar{P}$ valószínűségi mértéket kell alkalmas módon megadni. Ennek érdekében tekintsük azt az $(\Omega, \mathcal{A}, P)$ valószínűségi mezőt, ahol a $\xi_n(\omega)$ valószínűségi változók vannak definiálva, és definiáljuk a következő U: $\Omega \to R^{\pm\infty}$ (mérhető) leképezést: $U(\omega) = (\dots, \xi_{-1}(\omega), \xi_0(\omega), \xi_1(\omega), \dots)$. Legyen $\bar{P}$ a P mérték ezen U transzformáció szerinti ősképe az $(R^{\pm\infty}, \mathcal{B}^{\pm\infty})$ téren, azaz legyen $\bar{P}(A) = P(\{\omega\colon U(\omega) \in A\})$ minden $A \in \mathcal{B}^{\pm\infty}$ halmazra. Meg fogjuk mutatni, hogy a ξ_n, $-\infty < n < \infty$, sorozat stacionaritásából következik, hogy a $\bar{P}$ mérték T invariáns, azaz $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ ezzel a $\bar{P}$ mértékkel valóban speciáis dinamikus rendszer. Továbbá a $\bar{P}$ mérték definiciójából

az is következik, hogy a $\bar{\xi}_n$, $-\infty < n < \infty$, és a ξ_n, $-\infty < n < \infty$, valószínűségi változók sorozatainak az együttes eloszlásai megegyeznek.

Be kell még látni, hogy a $\bar{P}$ mérték valóban T invariáns, azaz definiálva a $Q(A) = \bar{P}(T^{-1}A)$, $A \in \mathcal{B}^{\pm\infty}$, mértéket $\bar{P}(A) = Q(A)$ minden $A \in \mathcal{B}^{\pm\infty}$ halmazra. Ez az azonosság igaz a ξ_n, $n = \dots, -1, 0, 1, \dots$, valószínűségi változó sorozat stacionaritása miatt a következő speciális alakú $A \in \mathcal{B}^{\pm\infty}$ (henger)halmazokra.

$$A = A(n_1, \dots, n_k, B) = \{x = (\dots, x_1, x_0, x_1, \dots)\colon\ (x_{n_1}, \dots, x_{n_k}) \in B\},$$

ahol $n_1, \dots, n_k$ tetszőleges egész számok, és B tetszőleges Borel mérhető halmaz az R^k k-dimenziós Euklideszi térben. Ugyanis $\bar{P}(A) = P((\xi_{n_1}, \dots, \xi_{n_k}) \in B)$, és $Q(A) = P((\xi_{n_1+1}, \dots, \xi_{n_k+1}) \in B)$ ebben az esetben. Viszont az ilyen alakú halmazok egy olyan halmaz algebrát alkotnak, amely generálja a $\mathcal{B}^{\pm\infty}$ σ-algebrát. Mivel egy mérték kiterjesztése egy halmaz algebráról az általa generált σ-algebrára egyértelmű, innen következik, hogy $\bar{P}(A) = Q(A)$ minden $A \in \mathcal{B}^{\pm\infty}$ halmazra, amint azt állítottuk.

A fenti lemma lehetővé teszi, hogy a dinamikus rendszerek elméletének az eredményeit alkalmazzuk diszkrét idejű stacionárius sztochasztikus folyamatok vizsgálatában. A dinamikus rendszerek elméletének egyik legfontosabb eredménye az ergod tétel. Ezt kívánom megfogalmazni. Ez előtt be kell vezetni néhány definiciót.

Dinamikus rendszer invariáns halmazainak a definiciója. *Egy $(\Omega, \mathcal{A}, P, T)$ dinamikus rendszer valamely $A \in \mathcal{A}$ halmazát e rendszer invariáns halmazának nevezünk, ha $T^{-1}(A) = A$. Ezt az azonosságot úgy értjük, hogy a benne szereplő két halmaz szimmetrikus differenciájának nulla a P mérték szerinti valószínűsége.*

Szükségünk lesz a következő egyszerű lemmára.

Lemma az invariáns halmazok viselkedéséről. *Egy dinamikus rendszer invariáns halmazai σ-algebrát alkotnak, azaz, ha A invariáns halmaz akkor annak komplementere, $\Omega \setminus A$ is az, és ha $A_1, A_2, \dots$ invariáns halmazok, akkor a $\bigcup_{n=1}^{\infty} A_n$ és $\bigcap_{n=1}^{\infty} A_n$ halmazok is invariánsak.*

Bizonyítás. Az állítás egyszerű következménye a T^{-1} inverz transzformáció tulajdonságainak.

Adva egy $(\Omega, \mathcal{A}, P, T)$ dinamikus rendszer jelölje $\mathcal{I} \subset \mathcal{A}$ az invariáns halmazok σ-algebráját, és vezessük be a következő definiciót.

Ergodikus dinamikus rendszerek definiciója. *Egy $(\Omega, \mathcal{A}, P, T)$ dinamikus rendszert ergodikusnak nevezünk, ha e rendszer invariáns halmazainak $\mathcal{I}$ σ-algebrája triviális a következő értelemben. Minden $A \in \mathcal{I}$ halmazra $P(A) = 0$ vagy $P(A) = 1$.*

Ergod tétel. *Legyen $(\Omega, \mathcal{A}, P, T)$ egy dinamikus rendszer, $U(\omega)$ egy $\mathcal{A}$ mérhető valós értékű függvény, amelyre $\int_\Omega |U(\omega)| P(\,d\omega) < \infty$. Ekkor*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} U(T^j\omega) = E(U|\mathcal{I})(\omega) \quad \text{a } P \text{ mérték szerint majdnem minden } \omega \in \Omega \text{ pontra,}$$

ahol $\mathcal{I}$ az invariáns halmazok σ-algebrája, és $E(\cdot|\mathcal{I})$ az $\mathcal{I}$ σ-algebra szerinti feltételes várható értéket jelöli. Ha az $(\Omega, \mathcal{A}, P, T)$ dinamikus rendszer ergodikus, akkor a képlet egyszerűsödik, mert ebben az esetben $E(U|\mathcal{I})(\omega) = EU = \int U(\omega)P(\,d\omega)$.

A fentiekben dinamikus rendszerek ergodicitását definiáltuk. De diszkrét idejű stacionárius sztochasztikus folyamatok ergodicitását is természetes módon lehet definiálni. Annak érdekében, hogy megadjuk azt, hogy egy ξ_n, $-\infty < n < \infty$, diszkrét idejű stacionárius sztochasztikus folyamat mikor ergodikus tekintsük azt az $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszert, amelyre a speciális dinamikus rendszer által indukált $\bar{\xi}_n$, $-\infty < n < \infty$, valószínűségi változók sorozatának és a ξ_n, $-\infty < n < \infty$, valószínűségi változó sorozatnak az eloszlása megegyezik. (Láttuk, hogy ilyen speciális dinamikus rendszer létezik.) Akkor mondjuk, hogy a ξ_n, $-\infty < n < \infty$, diszkrét idejű stacionárius sztochasztikus folyamat ergodikus, ha a fenti tulajdonságú $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszer ergodikus. Mivel a ξ_n, $-\infty < n < \infty$, és a $\bar{\xi}_n$, $-\infty < n < \infty$, sorozatok eloszlása megegyezik, ez a két sorozat ugyanolyan valószínűségszámítási törvényeket teljesít. Ez lehetővé teszi, hogy diszkrét idejű stacionárius sztochasztikus folyamatok vizsgálatát visszavezessük invertálható dinamikus rendszerek vizsgálatára, ahol alkalmazhatjuk az ergod tételt is.

Tekintsünk egy olyan diszkrét idejű ξ_n, $-\infty < n < \infty$, stacionárius sztochasztikus folyamatot, amelyben a ξ_n valószínűségi változók értékeit egy véges vagy megszámlálhatóan végtelen X halmazban veszik fel, és definiáljuk ennek entrópiáját. Az egyszerűbb jelölés érdekében feltehetjük, hogy X a valós számok egy részhalmaza. Olyan definiciót fogunk adni, amely összhangban van a dinamikus rendszerek esetében definiált $H(T, \xi)$ entrópia definicióval valamint a diszkrét idejű stacionárius sztochasztikus folyamatok és az invertálható dinamikus rendszerek közötti kapcsolattal. Legyen

$$H(\xi_n, -\infty < n < \infty) = \lim_{n\to\infty} H(\xi_0|\xi_{-1}, \dots, \xi_{-n}), \tag{6.1}$$

ahol $H(\xi_0|\xi_{-1}, \dots, \xi_{-n})$ az első fejezetben bevezetett feltételes entrópia. Az, hogy a (6.1) formulában szereplő limeszek valóban léteznek hasonlóan mutatható meg, mint ahogy a $H(T, \xi)$ entrópia definiciójának a jogosságát indokoltuk az 5. fejezetben. Továbbá nem nehéz belátni, hogy $H(\xi_n, -\infty < n < \infty) = \lim_{n\to\infty} \frac{1}{n} H(\xi_0, \xi_{-1}, \dots, \xi_{-n+1})$, ha $H(\xi_1) < \infty$. (Azt is állítjuk, hogy a $H(\xi_1) < \infty$ esetben ez a véges határérték létezik.)

Egyébként szoros kapcsolat van a diszkrét idejű stacionárius sztochasztikus folyamatok és invertálható dinamikus rendszerek entrópiája között. Legyen ξ_n, $-\infty < n < \infty$, olyan diszkrét idejű stacionárius sztochasztikus folyamat, amelyben a ξ_n valószínűségi változók értékeiket egy véges vagy megszámlálhatóan végtelen X halmazban veszik fel. Tekintsük azt a $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszert, amelyre a speciális dinamikus rendszer által indukált $\bar{\xi}_n$, $-\infty < n < \infty$, valószínűségi változók sorozatának és a ξ_n, $-\infty < n < \infty$, valószínűségi változó sorozatnak az eloszlása megegyezik. Ekkor $H(\xi_n, -\infty < n < \infty) = H(T, \bar{\xi}_0)$. Pontosabban egy apró technikai kellemetlenség elkerülése végett érdemes a $\bar{\xi}_n$ valószínűségi változók definicióját kissé módosítani. A $\bar{\xi}_n$ valószínűségi változók ugyanis, — legalábbis formálisan, — nem véges vagy megszámlálhatóan sok értéket vesznek fel. E valószínűségi változók értékkészlete az R számegyenes. De mivel $P(\xi_n \in X) = 1$, ettől a kellemetlenségtől egyszerűen

meg tudunk szabadulni. Arra az $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszerre, amelyet most tekintettünk $\bar{P}(X^{\pm\infty}) = 1$, ahol $X^{\pm\infty} = \{(\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)\colon\ x_{j_n} \in X,\ -\infty < n < \infty\}$. Ezért a tekintett $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszert helyettesíthetjük az $(X^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ (invertálható) dinamikus rendszerrel, ahol $\mathcal{B}^{\pm\infty}$ a korábban definiált $\mathcal{B}^{\pm\infty}$ (Borel) σ-algebra, és $\bar{P}$ a korábban definiált $\bar{P}$ mérték megszorítása az $X^{\pm\infty}$ halmazra. Az e rendszer által indukált $\bar{\xi}_n = T^n\bar{\xi}_0$ valószínűségi változókat a korábbi esethez hasonlóan definiáljuk. Ezzel a módosítással közvetlenül látható, hogy $H(\xi_n, -\infty < n < \infty) = H(T, \bar{\xi}_0)$.

A Shannon–McMillan–Breiman tétel megfogalmazása előtt teszek egy rövid kitérőt. Gyakran valószínűségi változóknak olyan ξ_n, $n \ge 0$ sorozataival kell foglalkoznunk, amelyek hasonlóan viselkednek a diszkrét idejű stacionárius sztochasztikus folyamatokhoz, de csak nem negatív n indexekre vannak definiálva. Az ilyen sorozatokat féloldali diszkrét idejű stacionárius sztochasztikus folyamatoknak fogom nevezni. Megfogalmazom, hogy ez pontosan mit jelent, és megmutatom, hogy az ilyan sorozatok vizsgálata visszavezethető a hagyományos diszkrét idejű stacionárius sztochasztikus folyamatok vizsgálatához. Először bevezetem a következő fogalmat.

Féloldali diszkrét idejű stacionárius sztochasztikus folyamat definiciója. *Legyen adva ξ_n, $n = 0, 1, 2, \dots$, valószínűségi változók egy sorozata egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Azt mondjuk, hogy ez a sorozat féloldali diszkrét idejű stacionárius sztochasztikus folyamat, ha minden $0 \le n_1 < n_2 < \dots < n_k < \infty$ és $m \ge 1$ egész számokra a $(\xi_{n_1}, \xi_{n_2}, \dots, \xi_{n_k})$ és $(\xi_{n_1+m}, \xi_{n_2}, \dots, \xi_{n_k+m})$ véletlen vektorok eloszlása megegyezik.*

A következő lemma kapcsolatot teremt féloldali diszkrét idejű és diszkrét idejű stacionárius sztochasztikus folyamatok között.

Lemma féloldali diszkrét idejű és diszkrét idejű stacionárius sztochasztikus folyamatok kapcsolatáról. *Legyen ξ_n, $n = 0, 1, 2, \dots$, egy féloldali diszkrét idejű stacionárius sztochasztikus folyamat egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Létezik olyan diszkrét idejű $\bar{\xi}_n$, $-\infty < n < \infty$, stacionárius sztochasztikus folyamat egy alkalmas $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{P})$ valószínűségi mezőn, amelyre a ξ_n, $n = 0, 1, 2, \dots$, és $\bar{\xi}_n$, $n = 0, 1, 2, \dots$, sorozatok eloszlása megegyezik.*

A lemma bizonyítása. Legyen $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{P}) = (R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P})$ a korábban definiált $R^{\pm\infty}$ halmazzal és $\mathcal{B}^{\pm\infty}$ σ-algebrával és egy alkalmasan definiált $\bar{P}$ valószínűségi mértékkel. Legyen továbbá $\bar{\xi}_n(x) = x_n$, $-\infty < n < \infty$, ha $x = (\dots, x_{-1}, x_0, x_1, \dots) \in R^{\pm\infty}$. A $\bar{P}$ mérték definiálása érdekében vegyük észre, hogy $P(\xi_{n_1} \in B_1, \dots, \xi_{n_k} \in B_k) = P(\xi_{n_1+p} \in B_1, \dots, \xi_{n_k+p} \in B_k)$ nem negatív egész számok minden monoton növekvő $0 \le n_1 < \dots < n_k$ véges sorozatára, tetszőleges $p \ge -n_1$ egész számra és a számegyenesen Borel mérhető $B_1, \dots, B_k$ halmazokra. Definiáljuk a $\bar{P}$ mértéket először bizonyos speciális halmazokra a

$$\begin{aligned}\bar{P}((\dots, x_{-1}, x_0, x_1, \dots)\colon\ x_{n_1} \in B_1, \dots, x_{n_k} \in B_k) &= P(\xi_0 \in B_1, \dots, \xi_{n_k-n_1} \in B_k) \\ &= P(\xi_{n_1+p} \in B_1, \dots, \xi_{n_k+p} \in B_k)\end{aligned} \tag{6.2}$$

képlet segítségével. E képletben $-\infty < n_1 < n_2 < \cdots < n_k < \infty$ egész számok, $p \geq -n_1$, és $B_1, \dots, B_k$ Borel mérhető halmazok a számegyenesen. (Megengedjük, hogy $n_j < 0$ legyen bizonyos j indexekre.) Ha a fenti képletekben az x_j koordinátákat valószínűségi változóknak tekintjük, akkor (6.2) képletben ezek véges dimenziós eloszlásait konzisztens módon definiáltuk. Ezért Kolmogorov tétele alapján létezik (egyetlen) olyan $\bar{P}$ mérték az $(R^{\pm\infty}, \mathcal{B}^{\pm\infty})$ téren, amely teljesíti a (6.2) formulát tetszőleges $-\infty < n_1 < n_2 < \cdots < n_k < \infty$, $p > -n_1$ egész számokra és $B_1, \dots, B_k$ Borel mérhető halmazokra a számegyenesen. (A valószínűségszámítás alaptételének egy lehetséges megfogalmazását alkalmaztuk.) Nem nehéz belátni, hogy az így konstruált $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P})$ valószínűségi mező és $\bar{\xi}_n$, $-\infty < n < \infty$, valószínűségi változók teljesítik a lemma állítását.

Rátérek a Shannon–McMillan–Breiman tétel ismertetésére. Ezen eredmény két ekvivalens verzióját fogom megfogalmazni. Az első verzió invertálható dinamikus rendszerek, a második verzió diszkrét idejű stacionárius sztochasztikus folyamatok viselkedéséről fog szólni.

A Shannon–McMillan–Breiman tétel invertálható dinamikus rendszerekre. *Legyen $(\Omega, \mathcal{A}, P, T)$ egy ergodikus invertálható dinamikus rendszer, és legyen azon adva egy olyan ξ valószínűségi változó, amely értékeit egy véges vagy megszámlálható $X = \{x_1, x_2, \dots\}$ halmazon veszi fel, és $H(\xi) < \infty$. Vezessük be a $\xi_n = T^n\xi$, $-\infty < n < \infty$, valószínűségi változókat, és definiáljuk minden $n = 1, 2, \dots$ számra a*

$$p_n(x_{j_0}, \dots, x_{j_{n-1}}) = P(\xi_0 = x_{j_0}, \dots, \xi_{n-1} = x_{j_{n-1}})$$

függvényt, ahol $x_{j_s} \in X$ minden $1 \leq s \leq n-1$ indexre. Ekkor

$$\lim_{n\to\infty} -\frac{1}{n} \log p_n(\xi_0, \dots, \xi_{n-1}) = H(T, \xi) \quad 1 \text{ valószínűséggel}, \tag{6.3}$$

ahol a $H(T, \xi)$ entrópiát az (5.5) képletben definiáltuk.

A Shannon–McMillan–Breiman tétel diszkrét idejű stacionárius sztochasztikus folyamatokra. *Legyen ξ_n, $-\infty < n < \infty$, egy olyan diszkrét idejű, ergodikus stacionárius sztochasztikus folyamat, amelyre a ξ_n valószínűségi változók értékeiket egy véges vagy megszámlálható $X = \{x_1, x_2, \dots\}$ halmazon veszik fel, és $H(\xi_n) < \infty$. Vezessük be minden $n = 1, 2, \dots$ számra a $p_n(x_{j_0}, \dots, x_{j_{n-1}}) = P(\xi_0 = x_{j_0}, \dots, \xi_{n-1} = x_{j_{n-1}})$ függvényt, ahol $x_{j_s} \in X$ minden $1 \leq s \leq n-1$ indexre. Ekkor*

$$\lim_{n\to\infty} -\frac{1}{n} \log p_n(\xi_0, \dots, \xi_{n-1}) = H(\xi_n,\ -\infty < n < \infty) \quad 1 \text{ valószínűséggel},$$

ahol a $H(\xi_n,\ -\infty < n < \infty)$ entrópiát az (6.1) képletben definiáltuk.

Megjegyzés. A ξ_n sorozat stacionaritása miatt a $p_n(\cdot)$ függvények definicióját

$$p_n(x_{j_0}, \dots, x_{j_{n-2}}, x_{j_{n-1}}) = P(\xi_{-n-1} = x_{j_0}, \dots, \xi_{-1} = x_{j_{n-2}}, \xi_0 = x_{j_{n-1}})$$

alakban is írhattuk volna.

Az, hogy egy ξ_n, $-\infty < n < \infty$, értékeiket egy véges vagy megszámlálható X halmazon felvevő valószínűségi változókból álló sztochasztikus folyamat teljesíti a Shannon–McMillan–Breiman tételt heurisztikusan úgy interpretálható, hogy meg lehet adni a sztochasztikus folyamat értékeinek egy olyan 'tipikus sorozatokból álló' 1 valószínűségű $X_0^{\pm\infty} \subset X^{\pm\infty}$ részhalmazát úgy, hogy nagy n számokra jó aszimptotikus formula adható annak valószínűségére, hogy a sztochasztikus folyamat megszorítása a 0 és $n-1$ indexű koordináták közé megegyezik egy előírt tipikus sorozat megszorításával a 0 és $n-1$ koordináták közé. Ez a valószínűség minden tipikus $x \in X_0^{\pm\infty}$ sorozatra közelítőleg egyenlő; exponenciálisan kicsi, és logaritmusának a $-\frac{1}{n}$-szerese körülbelül a sztochasztikus folyamat entrópiájával egyenlő.

A diszkrét idejű stacionárius sztochasztikus folyamatokra megfogalmazott Shannon–McMillan–Breiman tétel egyszerűen következik az invertálható dinamikus rendszerekről szóló Shannon–McMillan–Breiman tételből. Valóban, adva egy olyan ξ_n, $-\infty < n < \infty$, diszkrét idejű, ergodikus stacionárius sztochasztikus folyamat, amelyre a ξ_n valószínűségi változók értékeiket egy véges vagy megszámlálható $X = \{x_1, x_2, \dots\}$ halmazon veszik fel, és $H(\xi_n,\ -\infty < n < \infty) < \infty$, tekintsük azt az $(R^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ speciális dinamikus rendszert, amelyre a speciális dinamikus rendszer által indukált $\bar{\xi}_n$, $-\infty < n < \infty$, valószínűségi változók sorozatának és a ξ_n, $-\infty < n < \infty$, valószínűségi változó sorozatnak az eloszlása megegyezik. Pontosabban, ezt a speciális dinamikus rendszert kissé módosítjuk, felhasználva, hogy olyan valószínűségi változókat tekintünk, amelyek értékeiket az X halmzaban veszik fel. Ezért az $(X^{\pm\infty}, \mathcal{B}^{\pm\infty}, \bar{P}, T)$ dinamikus rendszert vesszük, és ebben a $\bar{\xi}_n$, $-\infty < n < \infty$, sorozatra alkalmazzuk a Shannon–McMillan–Breiman tételt. Ezt felírhatjuk $P((\dots, \bar{\xi}_{-1}(\omega), \bar{\xi}_0(\omega), \bar{\xi}_1(\omega), \dots) \in D) = 1$ alakban, ahol

$$D = \{(\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots) \in X^{\pm\infty}\colon \lim_{n\to\infty} -\frac{1}{n} \log p_n(x_{j_0}, \dots, x_{j_{n-1}}) = H(T, \bar{\xi}_0)\}.$$

Mivel a $\bar{\xi}_n$, $-\infty < n < \infty$ és ξ_n, $-\infty < n < \infty$, véletlen sorozatok eloszlása megegyezik, és $H(T, \bar{\xi}_0) = H(\xi_n,\ -\infty < n < \infty)$ $P((\dots, \xi_{-1}(\omega), \xi_0(\omega), \xi_1(\omega), \dots) \in \bar{D}) = 1$, ahol a $\bar{D}$ halmazt hasonlóan definiáljuk a D halmazhoz, csak a $H(T, \bar{\xi}_0)$ mennyiséget a $H(\xi_n,\ -\infty < n < \infty)$ mennyiséggel helyettesítjük benne. Ez viszont azt jelenti, hogy a Shannon–McMillan–Breiman tétel diszkrét idejű, ergodikus stacionárius sztochasztikus folyamatokra is érvényes.

A Shannon–McMillan–Breiman tételt invertálható dinamikus rendszerekre fogjuk bizonyítani. Annak érdekében, hogy a bizonyítást jobban megértsük tekintsük először azt a két speciális esetet, amikor a $\xi_n = T^n\xi$, $-\infty < n < \infty$, sorozat vagy a) független, egyforma eloszlású valószínűségi változók sorozata vagy b) egy stacionárius Markov lánc. Az a) esetben $p_n(x_{j_0}, \dots, x_{j_{n-1}}) = \prod_{s=0}^{n-1} p(x_{j_s})$, ahol $p(x_{j_s}) = P(\xi = x_{j_s})$. Ezért $-\frac{1}{n} \log p_n(\xi_0, \dots, \xi_{n-1}) = -\frac{1}{n} \sum_{s=0}^{n-1} \log p(\xi_s)$. A nagy számok erős törvénye szerint ez az átlag 1 valószínűséggel konvergál a $-E \log p(\xi) = -\sum P(\xi = x_j) \log P(\xi = x_j) = H(\xi) = H(T, \xi)$ összeghez, ha $n \to \infty$, és az a) esetben ezt kellett belátni.

A b) esetben hasonló az indoklás, csak ekkor az ergod tételt kell alkalmazni a nagy számok törvénye helyett. Egy olyan ξ_n, $-\infty < n < \infty$, Markov láncot tekintünk, amely az $(X^{\pm\infty}, \mathcal{B}^{\pm\infty}, P, T)$ téren van definiálva alkalmas P valószínűségi mértékkel, a szokásos $T(\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots) = (\dots, x_{j_0}, x_{j_1}, x_{j_2}, \dots)$ shift transzformációval, és $\xi_n(x) = x_{j_n}$, $-\infty < n < \infty$, ha $x = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)$. Jelölje $q(x) = P(\xi_n = x)$, $x \in X$, a stacionárius Markov lánc egy dimenziós eloszlásait, és $r(\bar{x}|x) = P(\xi_{n+1} = \bar{x}|\xi_n = x)$, $x, \bar{x} \in X$, a Markov lánc átmenet valószínűségeit. Ekkor

$$p_n(x_{j_0}, \dots, x_{j_{n-1}}) = q(x_{j_0}) \prod_{s=0}^{n-2} r(x_{j_{s+1}}|x_{j_s}),$$

ahonnan

$$-\frac{1}{n} \log p_n(\xi_0, \dots, \xi_{n-1}) = -\frac{1}{n} \log q(\xi_0) - \frac{1}{n} \sum_{s=0}^{n-2} \log r(T^s\xi_1|T^s\xi_0).$$

Ezért az ergod tételből az $U(x) = -\log r(x_{j_1}|x_{j_0})$, ha $x = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)$ függvény választással azt kapjuk, hogy $\lim_{n\to\infty} -\frac{1}{n} \log p_n(\xi_0, \dots, \xi_{n-1}) = -E \log r(\xi_1|\xi_0)$ 1 valószínűséggel. A bizonyítás befejezéséhez az $-E \log r(\xi_1|\xi_0) = H(T, \xi_0)$ azonosságot kell még igazolni.

Viszont

$$\begin{aligned}
&H(\xi_n|\xi_{n-1}, \dots, \xi_0) \\
&\quad = -\sum_{x_{j_0}, \dots, x_{j_n}} P(\xi_0 = x_{j_0}, \dots \xi_n = x_{j_n}) \log P(\xi_n = x_{j_n}|\xi_{n-1} = x_{j_{n-1}}, \dots, \xi_0 = x_{j_0}) \\
&\quad = -\sum_{x_{j_0}, \dots, x_{j_n}} P(\xi_0 = x_{j_0}, \dots, \xi_n = x_{j_n}) \log P(\xi_n = x_{j_n}|\xi_{n-1} = x_{j_{n-1}}) \\
&\quad = -\sum_{x_{j_{n-1}}, x_{j_n}} P(\xi_{n-1} = x_{j_{n-1}}, \xi_n = x_{j_n}) \log P(\xi_n = x_{j_n}|\xi_{n-1} = x_{j_{n-1}}) \\
&\quad = -E \log r(\xi_n|\xi_{n-1}) = -E \log r(\xi_1|\xi_0)
\end{aligned}$$

minden $n \geq 1$ számra a Markov tulajdonság és a stacionaritás miatt. Innen

$$H(T, \xi_0) = \lim_{n\to\infty} H(\xi_n|\xi_{n-1}, \dots, \xi_0) = -E \log r(\xi_1|\xi_0).$$

Markov láncok esetében a Shannon–McMillan–Breiman bizonyítása azon alapult, hogy a $-\frac{1}{n} \log p_n(\xi_0, \dots, \xi_{n-1})$ kifejezést felbontottuk egy olyan összegre, amelyre alkalmazni lehetett az ergod tételt. Az általános eset bonyolultabb. Ekkor a vizsgált kifejezést egy hasonló összegre plusz egy elhanyagolhatóan kis hibatagra lehet felbontani. De ahhoz, hogy ezt a hibatagot jól meg tudjuk becsülni szükségünk van a martingálok elméletének néhány fontos eredményére. A kívánt felbontás megtalálásának az

érdekében vezessük be a tekintett $\xi_n(x)$, $-\infty < n < \infty$ valószínűségi változó sorozat következő függvényeit.

$$\begin{aligned} g_k(\omega) &= -\log \frac{p_{k+1}(\xi_{-k}(\omega), \dots, \xi_0(\omega))}{p_k(\xi_{-k}(\omega), \dots, \xi_{-1}(\omega))}, \quad k \geq 1, \\ g_0(\omega) &= -\log p_1(\xi_0(\omega)). \end{aligned} \tag{6.4}$$

Az egyértelmű definició érdekében definiáljuk a $g_k(\omega)$ függvényt, mint $g_k(\omega) = 0$, ha $p_k(\xi_{-k}(\omega), \dots, \xi_{-1}(\omega)) = 0$, és ezért $p_{k+1}(\xi_{-k}(\omega), \dots, \xi_{-1}(\omega), \xi_0(\omega)) = 0$. Mivel ennek az eseménynek nulla a valószínűsége, nincs különösebb jelentősége annak, hogy ebben az esetben hogyan definiáljuk a $g_k(\omega)$ függvényt. Hasonló megjegyzést lehet tenni a később definiálandó $f_k^j(\omega)$ függvényről is.

Ezzel a jelöléssel

$$\begin{aligned} &-\frac{1}{n}\log p_n(\xi_0(\omega), \dots, \xi_{n-1}(\omega)) \\ &\quad = -\frac{1}{n}\log p_1(\xi_0(\omega)) - \frac{1}{n}\sum_{k=1}^{n-1} \log \frac{p_{k+1}(\xi_0(\omega), \dots, \xi_k(\omega))}{p_k(\xi_0(\omega), \dots, \xi_{k-1}(\omega))} = \frac{1}{n}\sum_{k=0}^{n-1} g_k(T^k\omega). \end{aligned}$$

Be fogjuk látni a martingálelmélet segítségével, hogy $\lim_{k\to\infty} g_k(\omega) = g_\infty(\omega)$ 1 valószínűséggel egy alkalmas $g_\infty(\omega)$ függvénnyel, emelyre $Eg_\infty(\omega) = H(T, \xi)$. Ez azt sugallja, hogy

$$-\frac{1}{n}\log p_n(\xi_0(\omega), \dots, \xi_{n-1}(\omega)) = \frac{1}{n}\sum_{k=0}^{n-1} g_\infty(T^k\omega) + \text{elhanyagolhatóan kicsi hiba.} \tag{6.5}$$

Nem triviális érvek segítségével be lehet látni, hogy ez valóban így van. Az utolsó formulából és az ergod tételből következik a Shannon–McMillan–Breiman tétel. A pontos bizonyítás kidolgozásának az érdekében először felidézem a martingál elmélet számunkra legfontosabb eredményeit.

Martingál, szubmartingál és szupermartingál definiciója. *Legyen adva σ-algebrák $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots$ növekvő sorozata egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, amelyre teljesül az $\mathcal{F}_n \subset \mathcal{A}$ tulajdonság minden $n = 1, 2, \dots$ számra. Legyen adva ezen kívül $\mathcal{F}_n$ mérhető $\xi_n(\omega)$, $E|\xi_n(\omega)| < \infty$, valószínűségi változók sorozata. Azt mondjuk, hogy a $(\xi_n(\omega), \mathcal{F}_n)$, $n = 1, 2, \dots$, párok sorozata martingált alkot, ha teljesül az*

$$E(\xi_{n+1}(\omega)|\mathcal{F}_n) = \xi_n(\omega) \qquad \textit{1 valószínűséggel minden } n = 1, 2, \dots \textit{ számra}$$

azonosság. A fent definiált sorozat szubmartingál, ha

$$E(\xi_{n+1}(\omega)|\mathcal{F}_n) \geq \xi_n(\omega) \qquad \textit{1 valószínűséggel minden } n = 1, 2, \dots \textit{ számra,}$$

és szupermartingál, ha

$$E(\xi_{n+1}(\omega)|\mathcal{F}_n) \leq \xi_n(\omega) \qquad \textit{1 valószínűséggel minden } n = 1, 2, \dots \textit{ számra.}$$

Ha adva van $\xi_n(\omega)$, $E|\xi_n(\omega)| < \infty$, $n = 1, 2, \ldots$, valószínűségi változók sorozata egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, de nincsenek definiálva a $\mathcal{F}_n$ σ-algebrák, akkor e sorozatot martingálnak, szubmartingálnak illetve szupermartingálnak nevezzük, ha a $(\xi_n, \mathcal{F}_n)$ sorozat az $\mathcal{F}_n = \sigma(\xi_k,\ 1 \le k \le n)$ σ-algebra sorozat választással martingál, szubmartingál illetve szupermartingál.

1. megjegyzés. A fenti definicióban az $E|\xi_n(\omega)| < \infty$ feltételt azért tettük, hogy beszélhessünk a tekintett feltételes várható értékekről. Ezt a feltételt lehet gyengíteni, elég például azt megkövetelni, hogy $EX_n^- < \infty$, $n = 1, 2, \ldots$, ahol $x^- = -\min(x, 0)$.

2. megjegyzés. Ha a $(\xi_n, \mathcal{F}_n)$ sorozat martingál, szubmartingál, illetve szupermartingál, akkor a ξ_n sorozat martingál, szubmartingál illetve szupermartingál a fent megadott értelemben is, azaz akkor, ha az $\mathcal{F}_n$ σ-algebrákat a $\mathcal{G}_n = \sigma(\xi_k,\ 1 \le k \le n) \subset \mathcal{F}_n$ σ-algebrákkal helyettesítjük. Ez egyszerűen látható a feltételes várható érték alapvető tulajdonságainak a segítségével.

3. megjegyzés. A szubmartingál és szupermartingál elnevezés hátterében a martingálok kapcsolata van a harmonikus függvényekkel. A martingálok a harmonikus függvények természetes megfelelői. Egy függvény akkor harmonikus, ha értéke egyenlő e függvény körintegráljával tetszőleges a görbét közrefogó zárt görbén. Ha egyenlőtlenség helyett nagyobb vagy egyenlő áll, akkor szuperharmonikus függvényről beszélünk, és ez felel meg a szupermartingálnak. Hasonlóan, ha egyenlőség helyett kisebb vagy egyenlő áll, akkor szubharmonikus függvényről beszélünk, és ennek a szubmartingál felel meg.

Bizonyos 1 valószínűségű martingál konvergencia tételekre és martingál egyenlőtlenségekre lesz szükségünk, illetve olyan eredményekre, amelyek arról szólnak, hogy hogyan lehet martingálokból vagy szubmartingálokból konvex függvények segítségével szubmartingálokat konstruálni. A következő konvergencia tételt fogjuk használni.

Tétel martingálok és szubmartingálok 1 valószínűségi konvergenciájáról. *Legyen $(\xi_n(\omega), \mathcal{F}_n)$, $n = 1, 2, \ldots$, martingál egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Ekkor az $E|\xi_n(\omega)|$, $n = 1, 2, \ldots$, sorozat monoton növekszik. Ha ez a sorozat korlátos, azaz létezik olyan $K < \infty$ szám, amelyre $E|\xi_n(\omega)| \le K$ minden $n = 1, 2, \ldots$ számra, akkor 1 valószínűséggel létezik a $\xi_\infty(\omega) = \lim_{n\to\infty} \xi_n(\omega)$ határérték. Ezenkívül érvényes az $E|\xi_\infty(\omega)| \le K$ egyenlőtlenség is ugyanazzal a $K < \infty$ konstanssal.*

Ha $(\xi_n(\omega), \mathcal{F}_n)$, $n = 1, 2, \ldots$, olyan szubmartingál egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, amelyre $\sup_n E|\xi_n(\omega)| \le K$ alkalmas $K < \infty$ konstanssal, akkor 1 valószínűséggel létezik a $\xi_\infty(\omega) = \lim_{n\to\infty} \xi_n(\omega)$ határérték, és ez a határérték teljesíti az $E|\xi_\infty(\omega)| \le K$ egyenlőtlenséget.

Szubmartingálok szuprémuma teljesíti a következő momentum egyenlőtlenséget.

Tétel szubmartingálok szuprémumának a momentumairól. *Legyen $(\xi_n, \mathcal{F}_n)$, $P(\xi_n \ge 0) = 1$, $n \ge 1$, nem negatív szubmartingál. Ekkor*

$$E\left(\sup_{n\ge1} \xi_n\right)^r \le \left(\frac{r}{r-1}\right)^r \sup_{n\ge1} E\xi_n^r \quad \text{minden } r > 1 \text{ valós számra,}$$

és

$$E\left(\sup_{n\geq 1}\xi_n\right)\leq\frac{e}{e-1}+\frac{e}{e-1}\sup_{n\geq 1}E\xi_n\log^+\xi_n$$

az $r=1$ *esetben, ahol* $\log^+ x=\max(\log x,0)$.

Igaz a Jensen egyenlőtlenség következő, feltételes várható értékekről szóló alakja.

A Jensen egyenlőtlenség feltételes várható értékekről. *Legyen adva egy* $\xi(\omega)$ *valószínűségi változó és egy* $\Phi(x)$, $-\infty<x<\infty$, *konvex függvény, amelyekre teljesülnek az* $E|\xi(\omega)|<\infty$ *és* $E|\Phi(\xi(\omega)|<\infty$ *feltételek egy* $(\Omega,\mathcal{A},P)$ *valószínűségi mezőn, valamint egy* $\mathcal{F}\subset\mathcal{A}$ σ*-algebra. Ekkor*

$$E(\Phi(\xi(\omega)|\mathcal{F}))\geq\Phi\left(E(\xi(\omega)|\mathcal{F})\right)\quad 1\text{ valószínűséggel.}$$

Ez az egyenlőtlenség akkor is érvényes, ha a $\Phi(\cdot)$ *konvex függvény egy* $a\leq x\leq b$ *intervallumban van definiálva, és* $P(a\leq\xi\leq b)=1$, *ahol* $-\infty\leq a<b\leq\infty$ *tetszőleges valós számok.*

A feltételes várható értékekről szóló Jensen egyenlőtlenség érvényessége azon múlik, hogy a várható értékhez hasonlóan a feltételes várható érték is kiszámolható alkalmas valószínűségi mérték szerinti integrál segítségével, csak ebben az esetben egy úgynevezett reguláris feltételes eloszlásfüggvény szerint kell integrálni. Számunkra ez az eredmény az alábbi következménye miatt lesz érdekes.

Lemma martingálok, szubmartingálok és szupermartingálok konvex függvényeiről.

a) Ha $(\xi_n,\mathcal{F}_n)$, $n=1,2,\dots$, *martingál,* $\Phi(x)$ *konvex függvény, és* $E|\Phi(\xi_n)|<\infty$ *minden* $n=1,2,\dots$ *számra, akkor* $(\Phi(\xi_n),\mathcal{F}_n)$, $n=1,2,\dots$, *szubmartingál.*

b) Ha $(\xi_n,\mathcal{F}_n)$, $n=1,2,\dots$, *szubmartingál,* $\Phi(x)$, *monoton növekvő konvex függvény, és* $E|\Phi(\xi_n)|<\infty$ *minden* $n=1,2,\dots$ *számra, akkor* $(\Phi(\xi_n),\mathcal{F}_n)$, $n=1,2,\dots$, *szubmartingál.*

c) Ha $(\xi_n,\mathcal{F}_n)$, $n=1,2,\dots$, *szupermartingál,* $\Phi(x)$ *monoton csökkenő konvex függvény, és* $E|\Phi(\xi_n)|<\infty$ *minden* $n=1,2,\dots$ *számra, akkor* $(\Phi(\xi_n),\mathcal{F}_n)$, $n=1$ *,2, . . . , szubmartingál.*

A fenti állítások akkor is érvényesek, ha a $\Phi(\cdot)$ *függvény egy* $a\leq x\leq b$ *intervallumban van definiálva, és* $P(a\leq\xi_n\leq b)=1$ *minden* $n=1,2,\dots$ *számra, ahol a* $-\infty\leq a<b\leq\infty$ *számok tetszőlegesek.*

Bizonyítás. Az a) esetben $E(\Phi(\xi_{n+1})|\mathcal{F}_n)\geq\Phi(E(\xi_{n+1}|\mathcal{F}_n))=\Phi(\xi_n)$ 1 valószínűséggel a Jensen egyenlőtlenség és a martingál tulajdonság miatt. A b) esetben $E(\Phi(\xi_{n+1})|\mathcal{F}_n)\geq\Phi(E(\xi_{n+1}|\mathcal{F}_n))$ 1 valószínűséggel, és mivel $E(\xi_{n+1}|\mathcal{F}_n)\geq\xi_n$, és $\Phi(\cdot)$ monoton nővekvő függvény, ezért $\Phi(E(\xi_{n+1}|\mathcal{F}_n))\geq\Phi(\xi_n)$ 1 valószínűséggel. Ezekből az egyenlőtlenségekből következik a b) rész állítása. A c) rész bizonyítása hasonló.

Felidézek még egy eredményt arról, hogy hogyan lehet kiszámolni egy valószínűségi változó függvényének a várható értékét. Azért idézem fel ezt az eredményt, mert

később szükségünk lesz rá. Az ismertetendő formula valójában a Stieltjes integrálokra vonatkozó parciális integrálás egy alkalmazása.

Egy a várható érték kiszámolásáról szóló formula. *Legyen ξ olyan valószínűségi változó, amelyre $P(\xi \ge 0) = 1$. Jelölje $F(x) = P(\xi < x)$, $0 \le x < \infty$, a ξ valószínűségi változó eloszásfüggvényét, és legyen $G(x) = 1 - F(x) = P(\xi \ge x)$. Tekintsünk egy monoton, folytonos $H(x)$ függvényt az $x \ge 0$ félegyenesen, amelyre $H(0) = 0$. Ekkor*

$$EH(\xi) = \int_0^\infty H(x)F(\,dx) = -\int_0^\infty H(x)G(\,dx) = \int_0^\infty G(x)H(\,dx).$$

Rátérek a Shannon–McMillan–Breiman tétel bizonyítására. Először annak a következő gyengébb formáját bizonyítom.

A Shannon–McMillan–Breiman tétel egy gyengébb alakja. *Igaz az invertálható dinamikus rendszerekre korábban megfogalmazott Shannon–McMillan–Breiman tétel abban a speciális esetben, ha a tételben tekintett ξ valószínűségi változó értékeit egy véges $X = \{x_1, x_2, \dots, x_r\}$ halmazon veszi fel. Részletesebben megfogalmazva ebben az esetben a (6.4) formulában a $\xi_k(\omega)$ sorozat segítségével definiált $g_k(\omega)$, $k = 0, 1, 2, \dots$, valószínűségi változók teljesítik a következő két relációt.*

a) Majdnem minden $\omega \in \Omega$ pontban teljesül a $\lim_{k\to\infty} g_k(\omega) = g_\infty(\omega)$ reláció egy alkalmas $g_\infty(\omega)$ valószínűségi változóval.

b) $E \sup_{k\ge1} g_k(\omega) < \infty$.

Továbbá, ha adva van egy $(\Omega, \mathcal{A}, P, T)$ ergodikus invertálható dinamikus rendszer és azon egy olyan ξ valószínűségi változó, amely értékeit egy véges vagy megszámlálhatóan végtelen $X = \{x_1, x_2, \dots\}$ halmazon veszi fel akkor vezessük be a $\xi_k = T^k\xi$, $-\infty < k < \infty$, valószínűségi változókat. Ha az ezen ξ_k valószínűségi változók által a (6.4) formulában definiált $g_k(\omega)$, $k = 0, 1, 2, \dots$, valószínűségi változók teljesítik a fent megfogalmazott a) és b) relációkat akkor a ξ_n, $-\infty < n < \infty$, sorozat teljesíti a (6.3) formulát.

Feladat.

Bizonyítsuk be, hogy az a) és b) relációk teljesülése esetén a (6.3) formulának az a változata is igaz, amelyben az 1 valószínűségi konvergencia helyett L_1 normában való konvergenciát követelünk meg.

Az előzőleg megfogalmazott tétel lényegében a Shannon–McMillan–Breiman tétel eredeti, Breiman által bizonyított alakja. Ő eredetileg csak véges sok értéket felvevő valószínűségi változók sorozatára bizonyította be az állítást, mert csak ebben az esetben tudta igazolni az a) és b) relációt. (A fő nehézséget a b) reláció igazolása jelenti.) Először Kai Lai Chung publikált eredményt arról, hogy ez a b) tulajdonság, és így a Shannon–McMillan–Breiman tétel az általános esetben is érvényes (A note on the ergodic theorem of information theory. Ann. Math. Statist. 32, 612–614 (1961)), de az ő

bizonyítása hibás. Mi ehelyett Andrew R. Barron The strong ergodic theorem for densities: Generalized Shannon–McMillan–Breiman theorem, (The Annals of Probability (1985) Vol. 13 No.4 1292–1303) cikkének a segítségével fogjuk bizonyítani, hogy az a) és b) relációk és így a Shannon–McMillan–Breiman tétel érvényes az általános esetben is. Barron eredményének más érdekes következménye is van.

A Shannon–McMillan–Breiman tétel gyengébb alakjának a bizonyítása. Először az a) és b) relációt bizonyítjuk be abban az esetben, ha a ξ valószínűségi változó X értékkészlete véges halmaz. Ennek érdekében bevezetjük a következő mennyiségeket.

$$\begin{aligned} f_k^j(\omega) &= -\log \frac{p_{k+1}(\xi_{-k}(\omega), \dots, \xi_{-1}(\omega), x_j)}{p_k(\xi_{-k}(\omega), \dots, \xi_{-1}(\omega))} \\ &= -\log P(\xi_0 = x_j | \xi_{-k}(\omega), \dots, \xi_{-1}(\omega)), \quad k \geq 1,\ x_j \in X. \end{aligned}$$

Rögzítsünk egy j, $1 \leq j \leq r$, számot (az r szám az $X = \{x_1, \dots, x_r\}$ halmaz definiciójában jelent meg), és vezessük be az $\eta_k = \eta_k^j = P(\xi_0 = x_j | \xi_{-k}(\omega), \dots, \xi_{-1}(\omega))$, $k = 1, 2, \dots$, valószínűségi változókat és $\mathcal{F}_k = \sigma(\xi_{-1}, \dots, \xi_{-k})$ σ-algebrákat. Az $(\eta_k, \mathcal{F}_k)$, $k = 1, 2, \dots$ rendszer martingál. Ismertetem e tény elemi bizonyítását, de előtte egy megjegyzésben leírom, hogy hogyan következik ez a tény általánosabb, jól ismert és egyszerűen igazolható eredményekből.

Megjegyzés. Legyen $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ növekvő σ-algebráknak egy sorozata, és ξ, $E|\xi| < \infty$, egy valószínűségi változó egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Ekkor az $(E(\xi|\mathcal{F}_n), \mathcal{F}_n)$, $n = 1, 2, \dots$, rendszer martingál. Speciálisan, ha ξ egy $A \subset \mathcal{A}$ halmaz indikátorfüggvénye ($A = \{\omega\colon\ \xi_0 = x_j\}$ választással), akkor ez az eredmény speciális esetként tartalmazza az előbb megfogalmazott állítást.

Ha nem kívánunk hivatkozni a fenti eredményre, akkor az előbb definiált rendszer martingál tulajdonságát megkapjuk az alábbi számolások segítségével. Az $\{\omega\colon\ \xi_{-k}(\omega) = x_{j_k}, \dots, \xi_{-1}(\omega) = x_{j_1}\}$ halmazon

$$\begin{aligned} E(\eta_{k+1} | \xi_{-k} = x_{j_k}, \dots, \xi_{-1} = x_{j_1}) &= \sum_{x \in X} P(\xi_{-k-1} = x | \xi_{-k} = x_{j_k}, \dots, \xi_{-1} = x_{j_1}) \\ &\qquad P(\xi_0 = x_j | \xi_{-k-1} = x, \xi_{-k} = x_{j_k} \dots, \xi_{-1} = x_{j_1}) \\ &= \frac{P(\xi_0 = x_j, \xi_{-k} = x_{j_k}, \dots, \xi_{-1} = x_{j_1})}{P(\xi_{-k} = x_{j_k}, \dots, \xi_{-1} = x_{j_1})} = \eta_k. \end{aligned}$$

Továbbá $E\eta_k = E|\eta_k| \leq 1$. Ezért a martingál konvergenciatétel alapján 1 valószínűséggel létezik az $\eta_\infty^j(\omega) = \lim_{k\to\infty} \eta_k(\omega)$, és így a logaritmus függvény folytonossága miatt az $f_\infty^j(\omega) = \lim_{k\to\infty} f_k^j(\omega)$ határérték, bár nem tudjuk kizárni annak a lehetőségét, hogy az $f_\infty^j(\omega)$ határérték végtelen. Mivel $f_k^j(\omega) = g_k(\omega)$ az $\{\omega\colon\ \xi_0(\omega) = x_j\}$ halmazon, innen következik, hogy az esetleg végtelen $g_\infty(\omega) = \lim_{k\to\infty} g_k(\omega)$ határérték is létezik 1 valószínűséggel.

Annak érdekében, hogy belássuk a b) relációt, jó becslést adunk a $P(\sup_{k\ge1} g_k(\omega) > \lambda)$ valószínűségre minden $\lambda \ge 0$ számra. Írjuk fel a

$$P\left(\sup_{k\ge1} g_k(\omega) > \lambda\right) = \sum_{j=1}^{r} P\left(\left\{\omega\colon \sup_{k\ge1} f_k^j(\omega) > \lambda\right\} \cap \{\omega\colon \xi_0(\omega) = x_j\}\right)$$
$$= \sum_{j=1}^{r}\sum_{k=1}^{\infty} P(F_{j,k} \cap \{\omega\colon \xi_0(\omega) = x_j\})$$

azonosságot, ahol

$$F_{j,k} = \left\{\omega\colon \max_{1\le p<k} f_p^j(\omega) \le \lambda,\ f_k^j(\omega) > \lambda\right\}$$

Rögzített j számra az $F_{j,k}$ halmazok, $k = 1, 2, \dots$, diszjunktak, és mivel $F_{j,k} \in \mathcal{F}_k = \sigma(\xi_{-1}(\omega), \dots, \xi_{-k}(\omega))$

$$P(F_{j,k} \cap \{\omega\colon \xi_0(\omega) = x_j\}) = \int_{F_{j,k}} P(\xi_0(\omega) = x_j | \xi_{-1}(\omega), \dots, \xi_{-k}(\omega)) P(\,d\omega)$$
$$= \int_{F_{j,k}} e^{-f_k^j(\omega)} P(\,d\omega) \le \int_{F_{j,k}} e^{-\lambda} P(\,d\omega) = e^{-\lambda} P(F_{j,k}).$$

Innen

$$P\left(\sup_{k\ge1} g_k(\omega) > \lambda\right) = \sum_{j=1}^{r}\sum_{k=1}^{\infty} P(F_{j,k} \cap \{\omega\colon \xi_0(\omega) = x_j\})$$
$$\le e^{-\lambda} \sum_{j=1}^{r} \left(\sum_{k=1}^{\infty} P(F_{j,k})\right) \le r e^{-\lambda}$$

minden $\lambda > 0$ számra. Ebből az egyenlőtlenségből következik a b) reláció.

Rátérek a Shannon–McMillan–Breiman tétel bizonyítására az a) és b) reláció segítségével. Mivel $\lim_{k\to\infty} g_k(\omega) = g_\infty(\omega)$ 1 valószínűséggel, a b) reláció és a dominált konvergencia tétel (Lebesgue tétel) alapján azt kapjuk, hogy $Eg_\infty(\omega) = \lim_{k\to\infty} Eg_k(\omega) = \lim_{k\to\infty} H(\xi_0|\xi_{-1}, \dots, \xi_{-k}) = H(T, \xi)$. Ez speciálisan azt is jelenti, hogy $g_\infty(\omega)$ 1 valószínűséggel véges.

A (6.5) formulát pontosan megfogalmazva azt írhatjuk, hogy

$$-\frac{1}{n} \log p_n(\xi_0(\omega), \dots, \xi_{n-1}(\omega)) = \frac{1}{n}\sum_{k=0}^{n-1} g_\infty(T^k\omega) + \frac{1}{n}\sum_{k=0}^{n-1} (g_k(T^k\omega) - g_\infty(T^k(\omega)).$$

Továbbá az ergod tétel alapján

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g_\infty(T^k\omega) = Eg_\infty(\omega) = H(T, \xi) \quad 1 \text{ valószínűséggel.}$$

Ezért a tétel bizonyításának befejezéséhez elég megmutatni, hogy

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}(g_k(T^k\omega)-g_\infty(T^k(\omega))=0 \quad 1 \text{ valószínűséggel.}$$

Ennek érdekében vezessük be a $G_N(\omega)=\sup_{k\ge N}|g_k(\omega)-g_\infty(\omega)|$, $N=1,2,\dots$, valószínűségi változókat, és bizonyítsuk be, hogy $\lim_{N\to\infty} EG_N(\omega)=0$. Valóban, $\lim_{N\to\infty} G_N(\omega)=0$ 1 valószínűséggel, $G_N(\omega)\le \sup_{k\ge 1} g_k(\omega)+g_\infty(\omega)$ minden N indexre, és mivel $E[\sup_{k\ge 1} g_k(\omega)+g_\infty(\omega)]<\infty$ a dominált konvergencia tételből következik a kívánt állítás.

Ezért az ergod tétel segítségével a következő becslést tudjuk tenni. Vegyünk egy tetszőleges $N\ge 1$ egész számot. Ekkor

$$\begin{aligned}\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{k=0}^{n-1}(g_k(T^k\omega)-g_\infty(T^k(\omega))\right| &\le \limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|g_k(T^k\omega)-g_\infty(T^k(\omega))| \\ &\le \limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}G_N(T^k\omega)=EG_N(\omega)\end{aligned}$$

1 valószínűséggel bármely $N\ge 1$ számra. Mivel $\lim_{N\to\infty} EG_N(\omega)=0$, innen

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{k=0}^{n-1}(g_k(T^k\omega)-g_\infty(T^k(\omega))\right|=0 \quad 1 \text{ valószínűséggel,}$$

és ezt kellett belátnunk.

Rátérek a Shannon–McMillan–Breiman tétel általános alakjának a bizonyítására. Elég azt megmutatni, hogy a tétel gyengébb alakjának megfogalmazásában szereplő a) és b) reláció az általános esetben is érvényes, és nemcsak akkor, ha ξ véges sok értéket vesz fel. Ezt a következő eredmény segítségével fogom bizonyítani.

Becslés Radon–Nikodym deriváltak növekvő σ-algebrákra vonatkozó viselkedéséről. *Legyen adva egy $(X,\mathcal{A})$ mérhető tér és azon növekvő σ-algebrák $\mathcal{F}_1\subset\mathcal{F}_2\subset\cdots\subset\mathcal{A}$ sorozata. Jelölje $\mathcal{F}_\infty$ a $\mathcal{F}_n$, $n=1,2,\dots$, σ-algebrák uniója által generált σ-algebrát. Legyen P és Q két olyan valószínűségi mérték az $(X,\mathcal{F}_\infty)$ téren, amelyeknek az $\mathcal{F}_n$ σ-algebrákra vett P_n és Q_n megszorításaira P_n abszolút folytonos a Q_n mértékre nézve minden $n=1,2,\dots$ indexre, és jelölje $\rho_n=\frac{dP_n}{dQ_n}$ a P_n mérték Q_n mérték szerinti Radon–Nikodym deriváltját. (Nem tesszük fel, hogy P abszolút folytonos a Q mértékre nézve az $\mathcal{F}_\infty$ σ-algebrán is.) Ekkor létezik a $\rho_\infty(\omega)=\lim_{n\to\infty}\rho_n(\omega)$ határérték P majdnem minden $\omega\in X$ pontban. Az $E\log\rho_n$ függvény az n index monoton növekvő függvénye. Ha $\lim_{n\to\infty} E\log\rho_n<\infty$, akkor $\lim_{n\to\infty} E\log\rho_n=E\log\rho_\infty$, és*

$$E\sup_n|\log\rho_n|\le eE\log\rho_\infty+e+2=e\lim_{n\to\infty}E\log\rho_n+e+2. \tag{6.6}$$

A most megfogalmazott eredményben tekintett várható értékeket a P mérték szerint vettük.

Először bebizonyítom a Shannon–McMillan–Breiman tételt az általános esetben ezen eredmény segítségével. A bizonyítás fő lépése a Shannon–McMillan–Breiman tétel egy gyengébb alakja néven megfogalmazott eredményben szereplő b) tulajdonság igazolása az általános esetben. Itt azt kell belátni, hogy bizonyos valószínűségi változok szuprémumának a várható értéke véges. Ebben nyújt segítséget az előző eredmény. Ez ugyanis bizonyos esetekben lehetővé teszi azt, hogy jó becslést adjunk valószínűségi változók szuprémumának várható értékére akkor is, ha csak az egyes valószínűségi változók várható értékének a szuprémumára van jó becslésünk. A b) tulajdonságban megfogalmazott egyenlőtlenség bizonyításában alkalmazhatjuk ezt a módszert.

A Shannon–McMillan–Breiman tétel bizonyítása az előző Radon–Nikodym deriváltakról szóló becslés segítségével. Feltehetjük, hogy az $(X^{\pm\infty}, \mathcal{A}^{\pm\infty}, T, \bar{P})$ invertálható dinamikus rendszerben dolgozunk, ahol $X^{\pm\infty}$ az összes $\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)$, $x_{j_n} \in X$ minden $-\infty < n < \infty$ indexre, két irányban végtelen X halmazbeli elemekből álló sorozat, $\mathcal{A}^{\pm\infty}$ a Borel σ-algebra ezen a halmazon, a T shift transzformáció a baloldali eltolás az $X^{\pm\infty}$ téren, azaz egy $\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots) \in X^{\pm\infty}$ pontra $T\omega = (\dots, x_{j_0}, x_{j_1}, x_{j_2}, \dots)$, $\bar{P}$ egy alkalmas ergodikus valószínűségi mérték ezen a téren. A (6.3) relációt a következő képlettel definiált ξ_n, $-\infty < n < \infty$, valószínűségi változókra akarjuk bizonyítani: $\xi_n(\omega) = x_{j_n}$, $-\infty < n < \infty$, az $\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)$ pontban. Ezzel a jelöléssel $\xi(\omega) = \xi_0(\omega)$ a tétel megfogalmazásában.

Valóban, tekintve a tételben eredetileg vizsgált $(X, \mathcal{A}, T, P)$ dinamikus rendszert és a rajta definiált $\xi_n = T^n\xi$, $-\infty < n < \infty$, valószínűségi változókat vezessük be az Ω tér $U(\omega) = (\dots, T^{-1}\xi(\omega), T^0\xi(\omega), T^1\xi(\omega), \dots)$ leképezését az $X^{\pm\infty}$ térbe. Ezután definiáljuk a $\bar{P}$ valószínűségi mértéket, mint a P mértéknek az U transzformáció szerinti ősképét, azaz legyen $\bar{P}(A) = P(\{\omega\colon U(\omega) \in A\})$ minden $A \in \mathcal{A}^{\pm\infty}$ halmazra. Be lehet látni, hogy ily módon egy invertálható $(X^{\pm\infty}, \mathcal{A}^{\pm\infty}, T, P)$ dinamikus rendszert kapunk, amely ergodikus, ha az eredeti $(\Omega, \mathcal{A}, T, P)$ rendszer az volt, és az ebben a rendszerben definiált $\xi_n = T^n\xi_0$, $-\infty < n < \infty$, valószínűségi változók együttes eloszlása megegyezik az eredeti ξ_n, $-\infty < n < \infty$, valószínűségi változók együttes eloszlásával. Ezért elég a bizonyítandó (6.3) formulát ebben az új rendszerben belátni.

Elég azt megmutatni, hogy a (6.4) formulában a most bevezetett ξ_n valószínűségi változók segítségével definiált $g_k(\omega)$ függvények teljesítik a Shannon–McMillan–Breiman tétel gyengébb alakjának megfogalmazásában szereplő a) és b) relációkat. Ezt a *Becslés Radon–Nikodym deriváltak növekvő σ-algebrákra vonatkozó viselkedéséről* eredménye segítségével fogom igazolni a következő szereposztással.

Az $(X^{\pm\infty}, \mathcal{A}^{\pm\infty})$ mérhető térben fogunk dolgozni, és a $\mathcal{F}_n$ σ-algebrákat úgy fogjuk definiálni, mint az olyan (mérhető) halmazokból álló σ-algebrákat, amelyek az

$$\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)$$

pontoknak csak az $x_{j_{-n}}, x_{j_{-n+1}}, \dots, x_{j_0}$ koordinátáitól függnek. Részletesebben fogalmazva vezessük be az $X^n = \{(x_{j_0}, \dots, x_{j_n}) \dots\ x_{j_s} \in X, \text{minden } 0 \le s \le n \text{ indexre}\}$

halmazt, és definiáljuk minden $\bar{x} = (\bar{x}_{j_0}, \dots, \bar{x}_{j_n}) \in X^n$ pontra az

$$A(\bar{x}) = \{\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots) \colon x_{j_{-s}} = \bar{x}_{j_{n-s}},\ 0 \le s \le n\} \in \mathcal{A}^{\pm\infty}$$

halmazt. Az $\mathcal{F}_n$ σ-algebra megegyezik az ilyen $A(\bar{x})$ halmazok által generált σ-algebrával. Az $\mathcal{F}_n$, $n = 1, 2, \dots$, σ-algebrák által generált $\mathcal{F}_\infty$ σ-algebra az olyan $A \in \mathcal{A}^{\pm\infty}$ halmazokból áll, amelyekre az $\omega \in A$ reláció teljesülése vagy nem teljesülése egy $\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)$ pontra csak az ω pont x_{j_s}, $s \le 0$, koordinátáitól függ.

A P mértéket úgy definiálom a $\mathcal{F}_\infty$ σ-algebrán, mint a $\bar{P}$ mérték megszorítását erre a σ-algebrára, tehát a

$$\begin{aligned} P(\{\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots) \colon\ & x_{j_{-n}} = \bar{x}_{j_n}, \dots, x_{j_{-1}} = \bar{x}_{j_1}, x_{j_0} = \bar{x}_{j_0}\}) \\ &= \bar{P}(\xi_{-n} = \bar{x}_{j_n}, \dots, \xi_{-1} = \bar{x}_{j_1}, \xi_0 = \bar{x}_{j_0}) \end{aligned}$$

képlet érvényes minden $n = 1, 2, \dots$ számra, és minden $\bar{x}_{j_s} \in X$, $0 \le s \le n$, pontokból álló n hosszúságú sorozatra. Ez a képlet egyértelműen definiálja a P mértéket a $\mathcal{F}_\infty$ σ-algebrán.

A Q mértéket a $\mathcal{F}_\infty$ σ-algebrán

$$\begin{aligned} Q(\{\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots) \colon\ & x_{j_{-n}} = \bar{x}_{j_n}, \dots, x_{j_{-1}} = \bar{x}_{j_1}, x_{j_0} = \bar{x}_{j_0}\}) \\ &= \bar{P}(\xi_{-n} = \bar{x}_{j_n}, \dots, \xi_{-1} = \bar{x}_{j_1}) P(\xi_0 = \bar{x}_{j_0}) \end{aligned}$$

képlet definiálja, amely érvényes minden $n = 1, 2, \dots$ számra, és minden $\bar{x}_{j_s} \in X$, $0 \le s \le n$, pontokból álló sorozatra.

A P_n mérték, azaz a P mérték megszorítása az $\mathcal{F}_n$ σ-algebrára abszolút folytonos a Q_n mértékre, a Q mérték megszorítására az $\mathcal{F}_n$ σ-algebrára, és fel tudjuk írni a Radon–Nikodym deriváltját. Nevezetesen

$$\begin{aligned} \rho_n(\omega) = \frac{P_n(\,d\omega)}{Q_n(\,d\omega)} &= \frac{P(\xi_{-n} = x_{j_{-n}}, \dots, \xi_{-1} = x_{j_{-1}}, \xi_0 = x_{j_0})}{P(\xi_{-n} = x_{j_{-n}}, \dots, \xi_{-1} = x_{j_{-1}}) P(\xi_0 = x_{j_0})} \\ &= \frac{p_{n+1}(\xi_{-n}(\omega), \dots, \xi_{-1}(\omega), \xi_0(\omega))}{p_n(\xi_{-n}(\omega), \dots, \xi_{-1}(\omega)) p_1(\xi_0(\omega))}, \end{aligned}$$

ha $\omega = (\dots, x_{j_{-1}}, x_{j_0}, x_{j_1}, \dots)$.

Innen

$$\begin{aligned} E \log \rho_n &= E \log \frac{p_{n+1}(\xi_{-n}(\omega), \dots, \xi_{-1}(\omega), \xi_0(\omega))}{p_n(\xi_{-n}(\omega), \dots, \xi_{-1}(\omega))} - E \log p_1(\xi_0(\omega)) \\ &= -H(\xi_0 | \xi_{-1}, \dots, \xi_{-n}) + H(\xi_0), \end{aligned}$$

és ennek a kifejezésnek van egy az n indextől nem függő felső korlátja, ha $H(\xi_0) < \infty$. Ezért ebben az esetben érvényes a (6.6) becslés. Továbbá, mivel $g_k(\omega) = -\log \rho_k(\omega) - \log p_1(\xi_0(\omega))$ minden $k = 1, 2, \dots$ indexre, $E \sup_{k \ge 1} g_k(\omega) \le E \sup_{k \ge 1} |\log \rho_k(\omega)| + H(\xi_0)$, és

a (6.6) formulából következik a b) reláció a $H(\xi) < \infty$ esetben. Továbbá, $\lim_{k\to\infty} g_k(\omega) = -\lim_{k\to\infty} \log \rho_k(\omega) - \log p_1(\xi_0(\omega)) = -\log \rho_\infty(\omega) - \log p_1(\xi_0(\omega))$ 1 valószínűséggel, azaz az a) reláció is teljesül. A Shannon–McMillan–Breiman tételt a *Becslés Radon–Nikodym deriváltak növekvő σ-algebrákra vonatkozó viselkedéséről* eredménye segítségével beláttuk.

A *Becslés Radon–Nikodym deriváltak növekvő σ-algebrákra vonatkozó viselkedéséről* eredményének a bizonyításában a fő nehézség az $E \sup_n |\log \rho_n(\omega)|$ várható érték becslése. Vezessük be a $\log^- x = -\min(\log x, 0)$ és $\log^+ x = \max(\log x, 0)$ függvényeket. Felírhatjuk az

$$\begin{aligned} E \sup_n |\log \rho_n(\omega)| &\le E \sup_n \log^+ \rho_n(\omega) + E \sup_n \log^- \rho_n(\omega) \\ &= E \sup_n \log^- \frac{1}{\rho_n(\omega)} + E \sup_n \log^+ \frac{1}{\rho_n(\omega)} \end{aligned}$$

egyenlőtlenséget. Az ezen egyenlőtlenség jobboldalán levő két tagot fogom megbecsülni. E két tag becslése más módszereket igényel. Ez az eset szétválasztás természetes. Ugyanis az első tag becslésében kihasználhatjuk azt, hogy $0 \le \log \rho(\omega) \le \rho(\omega)$ a $\rho(\omega) \ge 1$ esetben, míg a második tag becslésében a $\rho(\omega) \le 1$, esetet kell nézni, amikor $\log \rho(\omega) < 0$, és a $|\log \rho(\omega)|$ mennyiség hatásának a vizsgálatában más módszer szükséges. A második tag becslésében hasznos a következő lemma.

Egy martingál tipusú egyenlőtlenség valószínűségi változók szuprémumának a várható értékéről. *Legyen Z_n, $Z_n \ge 0$, $n = 1, 2, \dots$, nem negatív szupermartingál. Ekkor*

$$E \sup_n \log^- Z_n \le e + e \sup_n E \log^- Z_n.$$

Bizonyítás. Rögzítsük egy $r > 1$ számot, és vezessük be az $Y_n = \phi(Z_n)$, $n = 1, 2, \dots$, valószínűségi változók sorozatát, ahol $\phi(x) = \phi_r(x) = \max(1, (\log^- x)^{1/r})$. Azt állítom, hogy az Y_n, $n = 1, 2, \dots$, sorozat szubmartingál.

Mivel $\phi(x)$, $x \ge 1$, monoton csökkenő függvény, a *Lemma martingálok, szubmartingálok és szupermartingálok konvex függvényeiről* eredménye alapján ennek igazolásához elég megmutatni, hogy a $\phi(x)$ függvény konvex. A $\phi(x)$ függvény speciális alakja miatt ehhez elegendő azt ellenőrizni, hogy $\frac{d^2\phi(x)}{dx^2} \ge 0$ a $0 < x < \frac{1}{e}$ intervallumon. Viszont ezen az intervallumon $\phi(x) = (-\log x)^{1/r}$, $\frac{d\phi(x)}{dx} = -\frac{1}{rx}(-\log x)^{(1-r)/r}$, és $\frac{d^2\phi(x)}{dx^2} = \frac{1}{rx^2}(-\log x)^{(1-2r)/r}[-\log x - \frac{r-1}{r}] \ge 0$. (E számolás utolsó lépésében felhasználtuk, hogy $-\log x \ge 1 > \frac{r-1}{r}$, ha $0 < x < \frac{1}{e}$.)

Mivel $\log^- Z_n \le \phi(Z_n)^r \le 1 + \log^- Z_n$ minden $n = 1, 2, \dots$ számra, ezért a *Tétel*

szubmartingálok szuprémumának a momentumairól eredménye alapján

$$E \sup_n \log^- Z_n \le E \sup_n \phi(Z_n)^r \le \left(\frac{r}{r-1}\right)^r \sup_n E\phi(Z_n)^r$$
$$\le \left(\frac{r}{r-1}\right)^r \sup_n (1 + E \log^- Z_n).$$

Innen $r \to \infty$ határátmenettel megkapjuk a lemma állítását.

A Radon–Nikodym deriváltak növekvő σ-algebrákra vonatkozó viselkedéséről szóló becslés bizonyítása. Először azt mutatom meg, hogy az $(\frac{1}{\rho_n}, \mathcal{F}_n)$, $n = 1, 2, \dots$, rendszer szupermartingál. (Az $\frac{1}{\rho_n(\omega)}$ valószínűségi változók 1 valószínűséggel definiálva vannak, mert $P(\{\omega\colon\ \rho_n(\omega) = 0\}) = 0$.) Ennek érdekében vegyük észre, hogy a $(\rho_n, \mathcal{F}_n)$, $n = 1, \dots$, rendszer martingál a Q mérték szerint. Az igazolandó martingál tulajdonság azt jelenti ugyanis, hogy $\int_A \rho_n(\omega) Q(\,d\omega) = \int_A \rho_{n+1}(\omega) Q(\,d\omega)$ minden $A \in \mathcal{F}_n$ halmazra. Ez az azonosság viszont igaz, mert annak mind a két oldala $P(A)$-val egyenlő. (Felhasználtuk, hogy $A \in \mathcal{F}_n$ esetén $A \in \mathcal{F}_{n+1}$.) Vezessük be a következő $g(x)$ függvényt: $g(x) = 1$, ha $x > 0$, és $g(0) = 0$. A $g(\cdot)$ függvény konkáv a $[0, \infty)$ félegyenesen, $\rho_n(\omega) \ge 0$ minden $\omega \in X$ pontban, és $n = 1, 2, \dots$ indexre. Továbbá $g(\rho_n(\omega)) = I(\{\rho_n(\omega) > 0\})$, ahol $I(A)$, $A \in \mathcal{A}$, az A halmaz indikátorfüggvényét jelöli. A fenti tulajdonságokból következik, hogy az $(I(\{\rho_n(\omega) > 0\}), \mathcal{F}_n)$, $n = 1, 2, \dots$, sorozat szupermartingál a Q mérték szerint.

Azt állítom, hogy abból, hogy az $(I(\{\rho_n(\omega) > 0\}), \mathcal{F}_n)$, $n = 1, 2, \dots$, sorozat szupermartingál a Q mérték szerint, következik,hogy az $(\frac{1}{\rho_n(\omega)}, \mathcal{F}_n)$, $n = 1, 2, \dots$, rendszer szupermartingál (a P mérték szerint). (Valójában a két állítás ekvivalens.) Ehhez elegendő megmutatni, hogy $\int_A \frac{1}{\rho_n(\omega)} P(\,d\omega) = \int_A I(\{\rho_n(\omega) > 0\} Q(\,d\omega) = \bar{Q}(A)$ minden $A \in \mathcal{F}_n$ halmazra, ahol $\bar{Q}(A) = Q(A \cap \{\omega\colon\ \rho_n(\omega) > 0\})$, és hasonlóan $\int_A \frac{1}{\rho_{n+1}(\omega)} P(\,d\omega) = \tilde{Q}(A)$ minden $A \in \mathcal{F}_n$ halmazra, ahol $\tilde{Q}(A) = Q(A \cap \{\omega\colon\ \rho_{n+1}(\omega) > 0\})$. Ugyanis az, hogy az $(\frac{1}{\rho_n(\omega)}, \mathcal{F}_n)$, $n = 1, 2, \dots$, rendszer szupermartingál úgy is megfogalmazható, hogy $\int_A \frac{1}{\rho_n(\omega)} P(\,d\omega) \ge \int_A \frac{1}{\rho_{n+1}(\omega)} P(\,d\omega)$ minden $A \in \mathcal{F}_n$ halmazra, míg az, hogy az $(I(\{\rho_n(\omega) > 0\}), \mathcal{F}_n)$, $n = 1, 2, \dots$, sorozat szupermartingál a Q mérték szerint azt jelenti, hogy $\bar{Q}(A) = \int_A I(\{\rho_n(\omega) > 0\}) Q(\,d\omega) \ge \int_A I(\{\rho_{n+1}(\omega) > 0\}) Q(\,d\omega) = \tilde{Q}(A)$ minden $A \in \mathcal{F}_n$ halmazra.

Viszont tudjuk, hogy $P(A) = \int_A \rho_n(\omega) \bar{Q}(\,d\omega)$ minden $A \in \mathcal{F}_n$ halmazra. Innen az is következik, hogy $\int u(\omega) P(\,d\omega) = \int u(\omega) \rho_n(\omega) \bar{Q}(\,d\omega)$ minden $\mathcal{F}_n$ mérhető, nem negatív $u(\cdot)$ függvényre. Alkalmazzuk ezt a formulát az $u(\omega) = \frac{I(A)(\omega)}{\rho_n(\omega)}$ függvényre valamely $A \in \mathcal{F}_n$ halmazzal. Azt kapjuk, hogy $\int_A \frac{1}{\rho_n(\omega)} P(\,d\omega) = \bar{Q}(A)$, és ez volt az első bizonyítandó állítás. A második bizonyítandó állítást egyszerűen megkapjuk az elsőből, ha azt az $n + 1$ indexre alkalmazzuk az n index helyett, és felhasználjuk azt, hogy $A \in \mathcal{F}_{n+1}$, ha $A \in \mathcal{F}_n$.

A fenti relációkból az is következik, hogy $E|\frac{1}{\rho_n}| = E\frac{1}{\rho_n} = Q(\omega\colon\ \rho_n(\omega) > 0) \le 1$. Ezért a martingál konvergenciatételt alkalmazhatjuk a $(-\frac{1}{\rho_n}, \mathcal{F}_n)$ szubmartingálra, és

azt kapjuk, hogy az $\frac{1}{\rho_n(\omega)}$ sorozat 1 valószínűséggel konvergál. Innen az is következik, hogy a $\rho_n(\omega)$ sorozat 1 valószínűséggel konvergál, de határértéke lehet végtelen is. Másrészt alkalmazhatjuk az *Egy martingál tipusú egyenlőtlenség valószínűségi változók szuprémumának a várható értékéről* eredményét az $(\frac{1}{\rho_n}, \mathcal{F}_n)$ szupermartingálra, és az a

$$E \sup_n \log^- \frac{1}{\rho_n(\omega)} \le e + e \sup_n E \log^- \frac{1}{\rho_n(\omega)} = e + e \sup_n E \log^+ \rho_n(\omega)$$

egyenlőtlenséget adja. Továbbá

$$E \log^+ \rho_n(\omega) = E \log \rho_n(\omega) + E \log^- \rho_n(\omega) = E \log \rho_n(\omega) + E_Q \rho_n(\omega) \log^- \rho_n(\omega),$$

ahol E_Q a Q mérték szerinti várható értéket jelöli. (A $0 \log 0 = 0$ konvenciót alkalmazzuk.) Mivel $x \log x \ge -\frac{1}{e}$, minden $x \ge 0$ számra $\rho_n(\omega) \log^- \rho_n(\omega) \le \frac{1}{e}$, és emiatt $E \log^+ \rho_n(\omega) \le E \log \rho_n(\omega) + \frac{1}{e}$. Ezért az előbb bizonyított szuprémum egyenlőtlenségnek igaz az alábbi következménye.

$$E \sup_n \log^- \frac{1}{\rho_n(\omega)} \le e + 1 + e \sup_n E \log \rho_n(\omega). \tag{6.7}$$

(A (6.7) képlet előtt végzett számolások célja az volt, hogy olyan egyenlőtlenséget bizonyítsunk, amelyben az $E \log \rho_n(\omega)$ és nem az $E \log^+ \rho_n(\omega)$ mennyiségek segítségével adunk felső becslést.)

Azt állítom, hogy igaz az

$$E \sup_n \log^+ \frac{1}{\rho_n(\omega)} \le 1 \tag{6.8}$$

egyenlőtlenség is. Ezt az alábbi Ionescu Tulceától származó érvelés segítségével bizonyítom be.

A (6.8) formula igazolása érdekében vezessük be a $G(t) = P(\sup_n \log^+ \frac{1}{\rho_n(\omega)} > t)$ függvényt, $t \ge 0$, és írjuk fel az $E \sup_n \log^+ \frac{1}{\rho_n(\omega)} = \int_0^\infty G(t)\, dt$ azonosságot. (Az *egy a várható érték kiszámolásáról szóló formula* eredményét alkalmazzuk a $H(x) = x$ függvény választásával.) Definiáljuk ezenkívül az $A_{n,t} = \{\omega\colon \log \frac{1}{\rho_n(\omega)} > t,\ \max_{k<n} \frac{1}{\rho_k(\omega)} \le t\}$ halmazokat minden $t > 0$ számra és $n = 1, 2, \dots$ indexre. Rögzített $t \ge 0$ számra az $A_{n,t}$ halmazok diszjunktak, uniójuk az $\{\omega\colon \sup_n \log^+ \frac{1}{\rho_n(\omega)} > t\}$ halmaz, ezért $G(t) = \sum_{n=1}^{\infty} P(A_{n,t})$. Ezenkívül $A_{n,t} \in \mathcal{F}_n$, ahonnan

$$P(A_{n,t}) = \int_{A_{n,t}} \rho_n(\omega) Q_n(\, d\omega) = \int_{A_{n,t}} \rho_n(\omega) Q(\, d\omega).$$

Mivel $\rho_n(\omega) < e^{-t}$ az $\omega \in A_{n,t}$ pontokban, innen

$$P(A_{n,t}) \le \int_{A_{n,t}} e^{-t} Q(\, d\omega) = e^{-t} Q(A_{n,t}),$$

és $G(t) = \sum_{n=1}^{\infty} P(A_{n,t}) \le e^{-t} \sum_{n=1}^{\infty} Q(A_{n,t}) \le e^{-t}$ minden $t > 0$ számra. Ezért

$$E \sup_n \log^+ \frac{1}{\rho_n(\omega)} \le \int_0^\infty e^{-t}\, dt = 1,$$

amint állítottuk.

A (6.7) és (6.8) formulák alapján

$$E \sup_n |\log \rho_n(\omega)| = E \sup_n \left| \log \frac{1}{\rho_n(\omega)} \right| \le e \sup_n E \log \rho_n(\omega) + e + 2. \tag{6.9}$$

Abból, hogy az $(\frac{1}{\rho_n(\omega)}, \mathcal{F}_n)$, $n = 1, 2, \dots$, rendszer szupermartingál, és $-\log x$ monoton csökkenő konvex függvény következik, hogy a $(\log \rho_n(\omega), \mathcal{F}_n)$, $n = 1, 2, \dots$, rendszer szubmartingál. Speciálisan az $E \log \rho_n(\omega)$, $n = 1, 2, \dots$, sorozat monoton nő. Ha $\lim_{n\to\infty} E \log \rho_n(\omega) < \infty$, akkor a (6.9) formulából és a dominált konvergencia tételből következik, hogy $\lim_{n\to\infty} E \log \rho_n(\omega) = E \log \rho_\infty(\omega)$, ahol $\rho_\infty(\omega) = \lim_{n\to\infty} \rho_n(\omega)$. (Ez a $\rho_\infty(\omega)$ határérték 1 valószínűséggel létezik.) Innen és a (6.9) relációból következik a (6.6) formula. A tételt beláttuk.

A Shannon–McMillan–Breiman tétel jó becsést ad annak valószínűségére, hogy egy véges vagy megszámlálható sok értéket felvevő valószínűségi változókból álló ergodikus, diszkrét idejű stacionárius sztochasztikus folyamat egy hosszú szelete egy előírt tipikus értéket vesz fel. Hasonló eredményeket várhatunk akkor is, ha olyan ergodikus, diszkrét idejű stacionárius sztochasztikus folyamatokat tekintünk, amelyek olyan valószínűségi változókból állnak, amelyek értékeiket egy általános térben veszik fel. Természetes azt várni, hogy nagyon általános feltételek mellett az ilyen diszkrét idejű sztochasztikus folyamatok többváltozós sűrűségfüggvényei hasonló viselkedést mutatnak, mint az előbb tekintett speciális diszkrét idejű stacionárius sztochasztikus folyamatok szeleteinak az eloszlása. Annak érdekében, hogy pontosabban értsük, hogy mit jelent ez az állítás, megfogalmazok egy ilyen jellegű tényt kifejező eredményt.

Legyen $(X, \mathcal{A}, \mu)$ egy valószínűségi mező, ahol X egy teljes szeparábilis metrikus tér, és $\mathcal{A}$ a Borel σ-algebra ezen a téren. Vegyük e valószínűségi mezőnek az $n = \dots, -1, 0, 1, \dots$ egész számokkal indexelt $(X_n, \mathcal{A}_n, \mu_n)$ példányait, és definiáljuk ezek $(X^{\pm\infty}, \mathcal{A}^{\pm\infty}, \mu^\infty)$ direkt szorzatát. Vezessük be ezenkívül azon $\mathcal{F}_n \subset \mathcal{A}^{\pm\infty}$, $n = 1, 2, \dots$, σ-algebrákat az $(X^{\pm\infty}, \mathcal{A}^{\pm\infty}, \mu^\infty)$ valószínűségi mezőn, amelyek az

$$x = (\dots, x_{-1}, x_0, x_1, \dots) \in X^{\pm\infty}$$

pontok $x_{-n+1}, \dots, x_{-1}, x_0$ koordinátáitól függő hengerhalmazokból állnak. Azaz $\mathcal{F}_n$ az $\{x = (\dots, x_{-1}, x_0, x_1, \dots) \in X^{\pm\infty}\colon\ (x_{-n+1}, \dots, x_0) \in B\}$ alakú halmazokból áll egy $B \in \mathcal{A}^n$ halmazzal. E képletben $\mathcal{A}^n$ az $(X, \mathcal{A})$ tér $(X^n, \mathcal{A}^n)$ n-ik hatványában szereplő $\mathcal{A}^n$ σ-algebra. jelölje μ^n a μ^∞ mérték megszorítását a $\mathcal{F}_n$ σ-algebrára.

Legyen adva egy $\xi_n(x) = \xi_n(x_n)$, ha $x = (\dots, x_{-1}, x_0, x_1, \dots)$, $-\infty < n < \infty$, ergodikus, diszkrét idejű stacionárius sorozat az $(X^{\pm\infty}, \mathcal{A}^{\pm\infty}, \mu^\infty)$ valószínűségi mezőn.

A következő tételben a $(\xi_1, \dots, \xi_n)$ vektor sűrűségfüggvényének az aszimptotikájára adunk jó becslést nagy n számokra alkalmas feltételek teljesülése esetén.

Diszkrét idejű stacionárius sztochasztikus folyamat véges dimenziós sűrűségfüggvényeinek egy Shannon–McMillan–Breiman tétel tipusú becslése. *Tekintsünk egy $(X^{\pm\infty}, \mathcal{A}^{\pm\infty}, \mu^\infty)$ valószínűségi mezőt és azon egy az előbb bevezetett alakú $\xi_n(x) = \xi_n(x_n)$, $x \in X^{\pm\infty}$, $\infty < n < \infty$, ergodikus, diszkrét idejű stacionárius sztochasztikus folyamatot. Jelölje P ezen $(\xi_n(x),\ -\infty < n < \infty)$ sztochasztikus folyamat eloszlását az $(X^{\pm\infty}, \mathcal{A}^{\pm\infty})$ téren, és legyen P_n az $\mathcal{F}_n$ mérhető $(\xi_{-n+1}, \dots, \xi_0)$ vektor eloszlása az $(X^{\pm\infty}, \mathcal{F}_n)$ téren. Tegyük fel, hogy minden $n = 1, 2, \dots$ indexre a P_n mérték abszolút folytonos a μ^n mértékre nézve, és jelölje $p_n(x) = p_n(x_{-n+1}, \dots, x_0) = \frac{P_n(\,dx)}{\mu^n(\,dx)}$, $x = (\dots, x_{-1}, x_0, x_1, \dots)$, a P_n mértéknek a μ^n mérték szerinti Radon–Nikodym deriváltját. Tegyük fel azt is, hogy a $H = -\int \log p_1(x)\mu(\,dx) < \infty$ reláció teljesül. Ekkor létezik a*

$$\lim_{n\to\infty} -\int \log \frac{p_n(x_1, \dots, x_n)}{p_{n-1}(x_1, \dots, x_{n-1})} \mu^n(\,dx_1, \dots,\, dx_n) = H(P, \mu)$$

határérték, és $0 \le H(P, \mu) \le H$. Továbbá

$$\lim_{n\to\infty} -\frac{1}{n} \log p_n(\xi_1(x), \dots, \xi_n(x)) = H(P, \mu) \quad P \text{ majdnem minden } x \in X^{\pm\infty} \text{ pontban.}$$

E tétel bizonyítását, amely nagyon hasonlít az eredeti Shannon–McMillan–Breiman tétel bizonyításához, elhagyom. Csak annyit jegyzek meg, hogy a bizonyításban fontos szerepet játszik a *Becslés Radon–Nikodym deriváltak növekvő σ-algebrákra vonatkozó viselkedéséről* eredménye, és egy olyan (P, Q) valószínűségi mértékpárt kell választani, amellyel érdemes ezt az eredményt alkalmazni. Egyébként a Shannon–McMillan–Breiman tétel bizonyításában alkalmazott mértékpárhoz hasonló (P, Q) mértékpárt érdemes választani. Andrew R. Barron az ezen jegyzetben is említett *The strong ergodic theorem for densities: Generalized Shannon–McMillan–Breiman theorem* Probability (1985) Vol. 13 No.4 1292–1303) cikkében az előbb megfogalmazott eredmény lehetséges általánosításaival foglalkozik. Azt a kérdést vizsgálja, hogy milyen általánosabb μ^∞ dominálo mértékek esetében marad érvényben a tétel fő állítása.

Kiegészítés. *Az információelméleti előadásban használt martingálegyenlőtlenség bizonyítása.*

Szubmartingálok szuprémuma teljesíti a következő momentum egyenlőtlenséget.

Tétel szubmartingálok szuprémumának a momentumairól. *Legyen* $(\xi_k, \mathcal{F}_k)$, $1 \le k \le n$, *nem negatív (azaz a* $P(\xi_k \ge 0) = 1$, $1 \le k \le n$, *feltételt teljesítő) szubmartingál. Ekkor*

$$E\left(\max_{1\le k\le n} \xi_k\right)^r \le \left(\frac{r}{r-1}\right)^r E\xi_n^r \quad \textit{minden } r > 1 \textit{ valós számra,}$$

és

$$E\left(\max_{1\le k\le n} \xi_k\right) \le \frac{e}{e-1} + \frac{e}{e-1} E\xi_n \log^+ \xi_n$$

az $r = 1$ *esetben, ahol* $\log^+ x = \max(\log x, 0)$.

Ez a tétel következik az alábbi két lemmából.

Lemma 1. *Legyen* $(\xi_k, \mathcal{F}_k)$, $1 \le k \le n$, *szubmartingál. Ekkor*

$$\lambda P\left(\max_{1\le k\le n} \xi_k(\omega) \ge \lambda\right) \le \int_{\{\max\limits_{1\le k\le n} \xi_k(\omega)\ge\lambda\}} \xi_n(\omega) P(\,d\omega)$$

minden λ *valós számra.*

Lemma 2. *Ha* ξ *és* η *két olyan nem negatív valószínűségi változó, amelyekre*

$$P(\eta(\omega) \ge \lambda) \le \frac{1}{\lambda} \int_{\{\eta(\omega)\ge\lambda\}} \xi(\omega) P(\,d\omega) \tag{1}$$

minden $\lambda > 0$ *számra, akkor*

$$E\eta^r \le \left(\frac{r}{r-1}\right)^r E\xi^r \quad \textit{minden } r > 1 \textit{ valós számra,}$$

és

$$E\eta \le \frac{e}{e-1} + \frac{e}{e-1} E\xi \log^+ \xi$$

az $r = 1$ *esetben, ahol* $\log^+ x = \max(\log x, 0)$.

Lemma 1 bizonyítása. Legyen $\Lambda_1 = \{\omega\colon\ \xi_1(\omega) \ge \lambda\}$, $\Lambda_k = \{\omega\colon\ \xi_j(\omega) < \lambda,\ 1 \le j < k,\ \xi_k(\omega) \ge \lambda\}$, $1 < k \le n$, és $\Lambda = \bigcup\limits_{k=1}^{n} \Lambda_k$. Ekkor $\Lambda = \{\omega\colon\ \max\limits_{1\le k\le n} \xi_k(\omega) \ge \lambda\}$, és

$$\int_\Lambda \xi_n\,dP = \sum_{k=1}^n \int_{\Lambda_k} \xi_n\,dP = \sum_{k=1}^n \int_{\Lambda_k} E(\xi_n|\mathcal{F}_k)\,dP \ge \sum_{k=1}^n \int_{\Lambda_k} \xi_k\,dP$$
$$\ge \lambda \sum_{k=1}^n P(\Lambda_k) = \lambda P(\Lambda),$$

ahonnan következik a lemma állítása.

Lemma 2 bizonyítása. Legyen $\Psi(\lambda)$, $\lambda > 0$, monoton növekvő függvény, amelyre $\Psi(0) = 0$. Ekkor a Lebesgue–Stieltjes integrálokra vonatkozó parciális integrálás majd az (1) egyenlőtlenség, végül a Fubini tétel alapján

$$E\Psi(\eta) = -\int_0^\infty \Psi(\lambda)\,dP(\eta(\omega) \ge \lambda) \le \int_0^\infty P(\eta(\omega) \ge \lambda)\,d\Psi(\lambda)$$
$$\le \int_0^\infty \frac{1}{\lambda} \left(\int_{\eta(\omega)\ge\lambda} \xi(\omega) P(\,d\omega) \right) d\Psi(\lambda) = \int_\Omega \xi(\omega) \left(\int_0^{\eta(\omega)} \frac{d\psi(\lambda}{\lambda} \right) P(\,d\omega).$$

$\Psi(\lambda) = \lambda^r$, $r > 1$, választással felhasználva, hogy ezzel a választással $\frac{d\Psi(\lambda)}{\lambda} = r\lambda^{r-2}\,d\lambda$, majd alkalmazva a Hölder egyenlőtlenséget, azt kapjuk, hogy

$$E\eta^r \le \int_\Omega \xi(\omega) \left(\int_0^{\eta(\omega)} r\lambda^{r-2}\,d\lambda \right) P(\,d\omega)$$
$$= \frac{r}{r-1} E\xi\eta^{r-1} \le \frac{r}{r-1} \left(E\xi^r\right)^{1/r} \left(E\eta^r\right)^{1-1/r},$$

ahonnan

$$\left(E\eta^r\right)^{1/r} \le \frac{r}{r-1} \left(E\xi^r\right)^{1/r}.$$

$\Psi(\lambda) = \lambda$, ha $\lambda \ge 1$, és $\Psi(\lambda) = 0$, ha $0 \le \lambda < 1$ választással azt kapjuk, hogy

$$E(\eta - 1) \le E\Psi(\eta) \le \int_{\{\eta(\omega)\ge 1\}} \xi(\omega) \log \eta(\omega) P(\,d\omega),$$

és mivel $a \log b \le a \log^+ a + \frac{b}{e}$, ha $a > 0$, $b \ge 1$, (ugyanis az $a > 1$ esetben $a \log b = a \log^+ a + b(\frac{a}{b} \log \frac{b}{a}) \le a \log^+ a + \frac{b}{e}$, és az $a \le 1$, esetben $a \log b \le \log b \le \frac{b}{e}$), innen

$$E(\eta - 1) \le \int_{\{\eta(\omega)\ge 1\}} \xi(\omega) \log^+ \xi(\omega) P(\,d\omega) + \int_{\{\eta(\omega)\ge 1\}} \frac{\eta(\omega)}{e} P(\,d\omega) \le E\xi \log^+ \xi + \frac{E\eta}{e},$$

azaz

$$\left(1 - \frac{1}{e}\right) E\eta \le 1 + E\xi \log^+ \xi.$$

A fenti egyenlőtlenségekből következik a Lemma 2 állítása.

A tételben megfogalmazott egyenlőtlenségek következnek a Lemma 1 és Lemma 2 állításaiból, ha a Lemma 2-t $\eta = \max\limits_{1\le k\le n} \xi_k$ és $\xi = \xi_n$ szereposztással alkalmazzuk, ahol ξ_k, $1 \le k \le n$, a vizsgált szubmartingál.