Probability Theory

Problem set #10

De Moivre – Laplace Central limit theorem

Homework problems to be handed in:11.2, 11.4, 11.7, 11.10, 11.12

Bonus problems for extra credit: 11.13, 11.14

Due date: May 18

For the the numerical values of the function Φ use the table on the back!

11.1 Using the Gaussian approximation of binomial distribution determine the approximate value of

$$\binom{3600}{2376} 0.64^{2376} 0.36^{1224}.$$

11.2 Find the probability that among 10.000 random digits the digit 7 appears not more than 971 times.

11.3 Find an approximation of the probability that the number of sixes obtained when throwing a fair die 12.000 times between 1850 and 2100.

11.4 Find a number k such that the probability is about 0.5 that the number of heads obtained in 1000 coin tosses (with a fair coin) will be between 485 and k.

11.5 In 10.000 coin tosses fell heads 5400 times. Is it reasonable to assume that the coin is biased?

11.6 Estimate the probability that among 100.000 poker hands the number of full houses is between 128 and 158.

11.7 How many times do we have to flip a fair coin so that with probability exceeding 0.99 the number of the heads is between 49% and 51% of the number of experiments?

11.8 Peter plays roulette in a casino in Las Vegas; he always bets 10$ on red. After 100 games, he has lost 300$ and he suspects that the casino is cheating. Is his suspicion justified?

In the following two problems the sample is taken 'with replacements'. I.e. it may happen by chance that somebody is asked more than once. (Though, this event has negligibly small probability.)

11.9 Suppose that people from the population of a country are smokers with an *unknown* probability p independently of one another. This probability p is to be determined by observing relative frequencies: n people are chosen at random, the number of smokers among them is k. (We know from the Law of Large Numbers that for n large, the relative frequency $p' = k/n$ is close to p with overwhelming probability.) How large is n to be chosen in order that the relative frequency $p' = k/n$ of smokers in the observed sample to approximate the unknown p within an error less than 0.005, with probability larger than 0.95. That is: give a lower bound n_0, such that for any $n > n_0$ and any $p \in (0, 1)$

$$\mathbf{P}(|p' - p| \leq 0.005) \geq 0.95.$$

11.10 Before a presidential election, the Gallup organization wants to estimate the percentage of Democratic voters in the states of New Hampshire and Texas. It is known that this percentage is between 40% and 60% in both states but Gallup wants to estimate the unknown percentages with an accuracy of 2% with probability exceeding 0.99. There are 1.2 million eligible voters in New Hampshire and ten times as many voters, i.e. 12 million in the state of Texas. On the basis of these figures, their statistician claims that they need an approximately ten times larger sample in Texas than in New Hampshire. Is this correct? If not, how much larger sample is needed in Texas than in New Hampshire?

11.11 In a state there are 5.000.000 eligible voters, each of them votes for either party A or party B. Suppose that each individual chooses randomly with probability $1/2 - 1/2$ between the two parties (independently of the others). Find the probability that in the end the difference of votes between party A and party B is less than 300.

11.12 In a state there are 8.000.000 eligible voters, each of them votes for either party A or party B. Each individual chooses randomly with probability $1/2 - 1/2$ between the two parties (independently of the others), the party with the majority of the votes may form the state's government after the elections. Although it is not entirely legal, party B has the opportunity to 'import' some voters from the neighboring states who will vote for party B (with probability 1). How many of these voters should party B import to have at least 90% probability of winning the elections?

11.13 We have a (possibly) unfair coin and we want to check the validity of the hypothesis that the probability of tails for this coin is at least $3/4$. We are going to toss the coin 30.000 times and use the following rule. We choose a number k (before the coin-tosses) and accept the hypothesis if there were at least k tails among the coin-tosses. How should we choose k if we want that if a coin satisfies the hypothesis then with probability at least 90% we will actually accept the hypothesis?

11.14 (Approximating the Poisson distribution with a Gaussian.)
Prove that as $\lambda \to \infty$:

$$\sqrt{\lambda} p([\lambda + x\sqrt{\lambda}]; \lambda) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) + \mathcal{O}(\lambda^{-1/2}),$$

and the error term on the right is uniformly small if x is bounded. As a corollary, prove

$$\sum_{\lambda+\alpha\sqrt{\lambda}<k<\lambda+\beta\sqrt{\lambda}} p(k; \lambda) \to \int_{\alpha}^{\beta} \varphi(y) dy = \Phi(\beta) - \Phi(\alpha)$$

as $\lambda \to \infty$.

Further recommended exercises: Feller VII. 7., pages 193-195