Continuous random variables

(A very sketchy overview)

Introduction to Probability Theory, BSM

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is called a random variable, if for every $C \in \mathbb{R}$ the set $\{\omega : X(\omega) < c\} = \{X < c\}$ is an event, i.e. we can measure its probability with $\mathbf{P}$. If Ω is discrete then every function $X : \Omega \to \mathbb{R}$ is a random variable which can only take values from a discrete set $\{x_1, x_2, \dots\}$. We have seen that in this case the distribution of X may be described with the probabilities $\mathbf{P}(X = x_1), \mathbf{P}(X = x_2), \dots$

The distribution function

We would also like to describe random variables which are not necessary discrete. (The simplest such r.v. may be the position of a uniformly chosen random point from $[0, 1]$.) Since for these random variables it may happen, that the actual probability of being equal to a given number from their range is 0 (see the previous example), we need to find a new way to describe them. It turns out that instead of considering the probabilities $\mathbf{P}(X = c)$, the right thing is to look at $\mathbf{P}(X < c)$ for every $c \in \mathbb{R}$. This is the motivation for the following definition:
If X is a random variable defined on $(\Omega, \mathcal{A}, \mathbf{P})$, then its distribution function is defined as

$$F_X(x) := \mathbf{P}(\{\omega : X(\omega) < x\}) = \mathbf{P}(\{X < x\}) = \mathbf{P}(X \in (\infty, x)).$$

Since X is a random variable, the probability on the r.h.s. is defined. The subscript X is sometimes omitted. The distribution function $F(x)$ is sometimes called as cumulative distribution function or c.d.f.

Basic properties of the distribution function:

0. $F : \mathbb{R} \to [0, 1]$,
1. monotone non-decreasing;
2. left continuous:
$$F(x) = F(x-) := \lim_{y \to x^-} F(y)$$
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

(There are people who define $F(x)$ as $\mathbf{P}(X \leq x)$. In that case property 2 is changed to right-continuity.)
The proof of 0. is trivial: $F(x)$ is defined as a probability of an event. The next property follows from the fact that if $x_1 < x_2$ then $\{X < x_1\} \subset \{X < x_2\}$ and thus $\mathbf{P}(X < x_1) \leq \mathbf{P}(X < x_2)$. The other properties are simple consequences of the following lemma:

Lemma. Suppose that $A_1 \subset A_2 \subset A_3 \subset \dots$ are events on a probability space. Then $\mathbf{P}(\cup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} \mathbf{P}(A_n)$. Similarly, if $A_1 \supset A_2 \supset A_3 \supset \dots$ are events on a probability space then $\mathbf{P}(\cap_{n=1}^{\infty} A_n) = \lim_{n \to \infty} \mathbf{P}(A_n)$.

E.g. to prove the left-continuity, you need to consider the events $\{X < y\}$ as $y \nearrow x$. These are increasing, the union of these is $\{X < x\}$, applying the lemma we get that $F(x = \lim_{y \to x^-} F(y)$.

Some basic facts:

- If X is a discrete random variable, then its distribution function is a *step function*, i.e. it has some (finite or countably infinite) jump points, but between those it is constant.
- If X is discrete then $F(x)$ is not continuous. In general: F is continuous at x if and only if $\mathbf{P}(X = x)$.
- Using F we may express the probability of the event that X falls into a specific interval. E.g. $F(b) - F(a) = \mathbf{P}(X \in [a, b))$, $F(b) - F(a-) = \mathbf{P}(X \in (a, b))$, $1 - F(a) = \mathbf{P}(X \in [a, \infty))$ etc.
- If a function $F(x)$ satisfies the properties 0.-3. then there exists a random variable whose distribution function is exactly F.

The probability density function

Let $F(x)$ be a distribution function. It is called *absolutely continuous*, if there exists a (nice) function f such that

$$F(x) = \int_{-\infty}^{x} f(y)dy.$$

In that case f is called the probability density function (p.d.f.).
Basic properties, facts:

- $f \geq 0$, $\int_{-\infty}^{\infty} f(y)dy = 1$.
- For almost every $x \in \mathbb{R}$ the derivative $F'(x)$ exists and almost everywhere $F'(x) = f(x)$.
- f is not a probability!! (It may be larger than 1.) Its meaning is the following: if ε is small, then $\mathbf{P}(X \in (x, x + \varepsilon)) \approx \varepsilon f(x)$.
- For any $A \subset \mathbb{R}$ he probability of the event that X falls into the set A is $\mathbf{P}(X \in A) = \int_A f(y)dy$. In particular: $\mathbf{P}(X \in (a, b)) = \int_a^b f(y)dy$.

Examples

1. $U(a, b)$: uniform distribution on the interval $[a, b]$.

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & b < x \end{cases} \qquad f(x) = \begin{cases} 0 & x \leq a \\ \frac{1}{b-a} & a < x \leq b \\ 0 & b < x \end{cases}$$

E.g. the position of a uniformly chosen random point from $[a, b]$ is a random variable with distribution $U(a, b)$.

2. $EXP(\lambda)$: exponential distribution with parameter $\lambda > 0$.

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & 0 < x \end{cases} \qquad f(x) = \begin{cases} 0 & x \le 0 \\ \lambda e^{-\lambda x} & 0 < x \end{cases}$$

The exponential distribution is 'everfresh': if X has distribution $EXP(\lambda)$ then for any $x, y > 0$

$$\mathbf{P}(X > x + y | X > x) = \mathbf{P}(X > y).$$

This may be used to model the (random) life-length of a light bulb. The previous equation means that if the light bulb is working after x units of time, then it is as good as new: the probability that it will work for at least y more time units is the same we would get for a brand new light bulb.

3. $N(m, \sigma)$: normal or Gaussian distribution with parameters $m \in \mathbb{R}, \sigma > 0$.

Only the density function has a closed form:

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

its distribution function is $F_{m,\sigma}(x) = \int_{\infty}^{x} f_{m,\sigma}(y)dy$.

If $m = 0, \sigma = 1$ then the distribution is called *standard normal distribution*. In that case we use the notations $\varphi(x) = f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for the density and $\Phi(x) = F_{0,1}(x) = \int_{\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ for the distribution function. It is easy to check, that we can get the general normal distribution from the standard one with the following transformations:

$$f_{m,\sigma}(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right), \quad F_{m,\sigma}(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

This means that if $X \sim N(0,1)$, then the random variable defined as $\sigma X + m$ has distribution $N(m, \sigma)$, and vice versa: if $Y \sim N(m, \sigma)$ then $\frac{Y-m}{\sigma}$ is standard normal. This is a very useful relation, this way when we have to work with general normal distributions, we can always deduce the problem to one dealing with the standard normal distribution.

Functions of random variables

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution function $F(x)$ and density $f(x)$. Suppose $\Psi : \mathbb{R} \to \mathbb{R}$ is a 'nice' (usually differentiable) function. Then $Y = \Psi(X)$ is also a random variable on Ω, denote its distribution function with $G(y)$, and its density (if exists) by $g(y)$. Our goal is to express G and g using F, f and the function Ψ.

Transformation of the distribution function: Clearly,

$$G(y) = \mathbf{P}(Y < y) = \mathbf{P}(\Psi(X) < y).$$

If Ψ is increasing and one-to-one then from the previous equation:

$$G(y) = \mathbf{P}(X < \Psi^{-1}(y)) = F(\Psi^{-1}(y)).$$

If Ψ is not invertible we have to be careful, and understand the structure of the event $\{\Psi(X) < y\}$. E.g. if $\Psi(y) = y^2$ then (if $y > 0$ and F is cont.):

$$G(y) = \mathbf{P}(X^2 < y) = \mathbf{P}(-\sqrt{y} < X < \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y}).$$

Transformation of the density function:

If we have an expression for $G(y)$ then differentiating it (and hoping that the derivative exists) we get $g(y)$. If Ψ is monotone increasing and one-to-one then $G(y) = F(\Psi^{-1}(y))$, thus

$$g(y) = \frac{d}{dy}F(\Psi^{-1}(y)) = F'(\Psi^{-1}(y))\frac{d}{dy}\Psi^{-1}(y) = \frac{f(\Psi^{-1}(y))}{\Psi'(\Psi^{-1}(y))}.$$

If Ψ is not invertible then the formula is more complicated. If Ψ is differentiable and doesn't have horizontal line segments in its graph, then we have the following expression for $g(y)$:

$$g(y) = \sum_{x\in\Psi^{-1}(y)} \frac{x}{|\Psi'(x)|}.$$

IMPORTANT: in the formula $\Psi^{-1}(y)$ is the set of all numbers x for which $\Psi(x) = y$.

Expectation of abs. cont. random variables

If X is a random variable with density function $f(x)$, then if the integral $\int_{-\infty}^{\infty} |x|\, f(x)dx$ is finite, then its expectation is defined as

$$\mathbf{E}X = \int_{-\infty}^{\infty} xf(x)dx,$$

it is usually denoted by m. This definition is the natural extension of the one we used for defining the expectation of a discrete random variable.

Examples:

- $U(a,b)$:

$$m = \int_a^b x\frac{1}{b-a}dx = \frac{a+b}{2}$$

- $EXP(\lambda)$:

$$m = \int_0^\infty x\lambda e^{-\lambda x}dx = \frac{1}{\lambda}.$$

(partial integration...)

- $N(m,\sigma)$:
We need to compute the integral

$$\int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

Since $N(m,\sigma)$ may be written as a linear transformation of $N(0,1)$, it is enough to compute the integral for $m = 0, \sigma = 1$, which is

$$\int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = 0$$

by symmetry. (The integral on the positive side cancels out the integral on the negative side.) If $X \sim N(0,1)$ then $\sigma X + m \sim N(m,\sigma)$ thus using the linearity of the expectation we have that the expectation of $N(m,\sigma)$ is $\sigma \cdot 0 + m = m$.

- Example of a distribution where expectation is not defined:
  The standard Cauchy distribution is defined with its density function: $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
  We need the finiteness of the integral

$$\int_{-\infty}^{\infty} |x|\, f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{|x|}{1+x^2}dx.$$

  But for large x $\frac{|x|}{1+x^2} \approx \frac{1}{x}$ and the integral of $\frac{1}{x}$ is infinite on $[c,\infty]$ for every $c > 0$. Thus for the Cauchy distribution the expectation is not defined.

Variance of abs. cont. random variables

If X is abs. cont. with density function $f(x)$ then

$$\mathbf{Var}X = \int_{-\infty}^{\infty} (x-m)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - (\int_{-\infty}^{\infty} x f(x)dx)^2.$$

This is defined only if $\int_{-\infty}^{\infty} x^2 f(x)dx$ is finite. The variance is usually denoted by σ^2, σ is the standard deviation.

- $U(a,b)$:

$$\sigma^2 = \int_a^b x^2 \frac{1}{b-a}dx - m^2 = \frac{(b-a)^3}{12}$$

  (Check the calculations!)

- $EXP(\lambda)$:

$$\sigma^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx - m^2 = \frac{1}{\lambda^2}$$

  (Check the calculations using partial integration...)

- $N(m,\sigma)$:

$$\int_{-\infty}^{\infty} (x-m)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \sigma^2$$

  (Try partial integration with $(x-m)$ and $(x-m)e^{-\frac{(x-m)^2}{2\sigma^2}}$ as the two functions.) We may simplify the problem by first calculating the integral for $m = 0, \sigma = 1$ and then using the fact that we can get the general case by using linear transformation.

Expectations of functions of random variables

If X is abs. cont. with density function $f(x)$ and consider the random variable $Y = \Psi(X)$. It would be possible to compute $\mathbf{E}Y$ by first determining $g(y)$ (the density of Y) and then using the definition: $\mathbf{E}Y = \int_{-\infty}^{\infty} y g(y) dy$. This would be quite messy, as $g(y)$ may become quite complicated.
Fortunately, there is a much easier way, as

$$\mathbf{E}Y = \int_{-\infty}^{\infty} \Psi(x) f(x) dx.$$

Examples:

- Higher moments:
  k^{th} absolute moment: $\mathbf{E}\,|X|^k = \int_{-\infty}^{\infty} |x|^k f(x) dx$,
  k^{th} moment: $\mathbf{E}X^k = \int_{-\infty}^{\infty} x^k f(x) dx$.
  (It is defined only if the respective absolute moment is finite.)

- Exponential moments or momentum-generating function:
  This is a function $H : \mathbb{R} \to \mathbb{R}$ defined as

  $$H(t) = \mathbf{E}e^{tX} = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

  if it is finite. It is useful because if it is defined in a non-empty interval around 0, then we can obtain all moments by differentiation:

  $$\frac{d^k H}{dt^k}\Big|_{t=0} = \mathbf{E}X^k.$$

- Characteristic function:
  This is a complex function $\psi : \mathbb{R} \to \mathbb{C}$ defined as

  $$\psi(t) = \mathbf{E}e^{itX} = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

  where $i = \sqrt{-1}$. Since e^{itx} is bounded (its absolute value is 1) thus the integral is always defined for every t. This function is very useful because it has nice analytic properties and it contains all information about the distribution of X.

Some tips for calculating expectations:

- Use the definition, maybe the integral is easy to calculate.

- Use symmetry, if you can, as in the expectation of the standard normal.

- Try to use integrals whose values we have already calculated.
  E.g. Calculate $\mathbf{E}e^{tX}$ where X is standard normal. By the definition:

  $$\mathbf{E}e^{tX} = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

  Transforming the function under the integral (remember that t is a constant!):

  $$\int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} e^{t^2/2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2+tx-t^2/2} dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} dx$$

  But the integral on the right side is 1, since it is the integral of a density function ($N(t,1)$) on the whole line. This means that $\mathbf{E}e^{tX} = e^{t^2/2}$ if $X \sim N(0,1)$.