Chapter 6

Random Variables (Continuous Case)

Thus far, we have purposely limited our consideration to random variables whose ranges are countable, or discrete. The reason for that is that distributions on countable spaces can be specified by means of the point distribution; the distribution is uniquely defined by specifying it only for elementary events. The construction of a distribution on an uncountable space is only done rigorously within the framework of measure theory. Here, we will only provide limited tools which will allow us to operate with such variables.

6.1 Basic definitions

Definition 6.1 *Let $(\Omega, \mathscr{F}, P)$ be a probability space. A real-valued function $\Omega \to \mathbb{R}$ is called a continuous random variable, if there exists a non-negative real-valued integrable function $f_X(x)$, such that*

$$P(\{\omega : X(\omega) \le a\}) = F_X(a) = \int_{-\infty}^{a} f_X(x)\, dx.$$

Moreover, f_X is normalized,

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1.$$

The function f_X is called the probability density function *(PDF) (צפיפות ההתפלגות) of X.*

Comment: Recall that a random variable has a σ-algebra of events $\mathscr{F}_X$ associated with its range (here $\mathbb{R}$), and we need $X^{-1}(A) \in \mathscr{F}$ for all $A \in \mathscr{F}_X$. What is a suitable σ-algebra for $\mathbb{R}$? These are precisely the issues that we sweep under the rug.

Thus, a continuous random variable is defined by its PDF. Since a random variable is by definition defined by its distribution P_X, we need to show that the PDF defines the distribution uniquely. Since we don't really know how to define distributions when we don't even know the set of events, this cannot really be achieved in this course.

The cumulative distribution function F_X defines the distribution P_X for the following type of events:

1. For every segment $a < b$,
$$(a, b] = (-\infty, b] \smallsetminus (-\infty, a],$$
hence by additivity,
$$P_X((a, b]) = F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx.$$

2. For every $A \subset \mathbb{R}$ expressible as a disjoint countable union,
$$A = \bigcup_{j=1}^{\infty} (a_j, b_j],$$
we have
$$P_X(A) = \int_A f_X(x)\, dx = \sum_{j=1}^{\infty} \int_{a_j}^{b_j} f_X(x)\, dx.$$

3. For every $a \in \mathbb{R}$, and for every $n \in \mathbb{N}$,
$$\left(a - \frac{1}{n+1}, a + \frac{1}{n+1}\right] \subset \left(a - \frac{1}{n}, a + \frac{1}{n}\right],$$
and
$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n}\right].$$
By the continuity of probability for decreasing events,
$$P_X(\{a\}) = P_X\left(\bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n}\right]\right) = \lim_{n\to\infty} \int_{a-1/n}^{a+1/n} f_X(x)\, dx = 0.$$

4. By the additivity of probability for disjoint events,

$$P_X([a,b]) = P_X(\{a\}) + P_X((a,b)) + P_X(\{b\}) = P_X((a,b)).$$

Comment: We may consider discrete random variables as having a PDF which is a sum of δ-functions.

Example: The random variable X has a PDF of the form

$$f_X(x) = \begin{cases} 2C(2x - x^2) & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}.$$

What is the value of the constant C and what is the probability that $X(\omega) > 1$?

The constant is obtained by normalization,

$$1 = 2C \int_0^2 (2x - x^2)\, dx = 2C \left(4 - \frac{8}{3}\right) = \frac{8C}{3}.$$

Then,

$$P(X > 1) = 2C \int_1^2 (2x - x^2)\, dx = \frac{1}{2}.$$

▲ ▲ ▲

6.2 The uniform distribution

Just as for discrete random variables, we will encounter next a collection continuous random variables that are sufficiently recurrent in applications to deserve a spacial name.

Definition 6.2 A random variable X is called uniformly distributed in $[a,b]$ *(מתפלג באופן אחיד), denoted $X \sim \mathcal{U}(a,b)$, if its* PDF *is given by*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \textit{otherwise} \end{cases}.$$

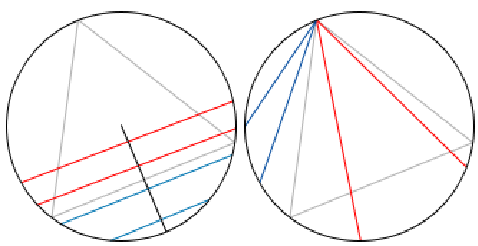
Example: Buses are arriving at a station every 15 minutes. A person arrives at the station at a random time, uniformly distributed between 7:00 and 7:30. What is the probability that he has to wait less than 5 minutes?

Let $X(\omega)$ be the arrival time (in minutes past 7:00), and $Y(\omega)$ the time he has to wait. We know that $X \sim \mathcal{U}(0, 30)$. Then,

$$P(Y < 5) = P(\{X = 0\} \cup \{10 \le X < 15\} \cup \{25 \le X < 30\}) = 0 + \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$$

▲ ▲ ▲

Example: *Bertrand's paradox*: consider a random chord of a circle. What is the probability that the chord is longer than the side of an equilateral triangle inscribed in that circle?



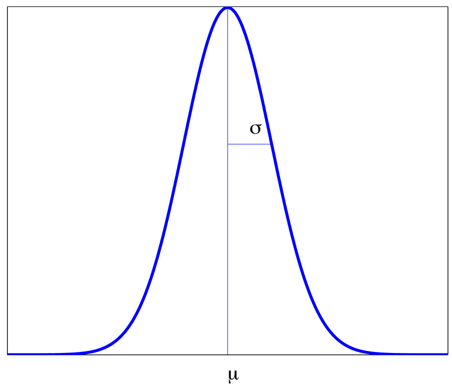
The "paradox" stems from the fact that the answer depends on the way the random chord is selected. One possibility is to take the distance of the chord from the center of the circle r to be $\mathcal{U}(0, R)$. Since the chord is longer than the side of the equilateral triangle when $r < R/2$, the answer is $1/2$. A second possibility is to take the angle θ between the chord and the tangent to the circle to be $\mathcal{U}(0, \pi)$. The chord is longer than the side of the triangle when $\pi/3 < \theta < 2\pi/3$, in which case the answer is $1/3$. ▲ ▲ ▲

6.3 The normal distribution

Definition 6.3 A random variable X is said to be normally distributed *with parameters μ, σ^2, denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$, if its* PDF *is*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

X is called a standard normal variable *if $\mu = 0$ and $\sigma^2 = 1$.*



Proposition 6.1 *This is a* PDF*, i.e., f_X is non-negative, integrable and*

$$\int_{-\infty}^{\infty} f_X(\xi)\, d\xi = 1.$$

Proof: Non-negativity is obvious and every continuous function is integrable. It remains to prove the normalization condition. Define $x = (\xi - \mu)/\sigma$. Since $d\xi = \sigma\, dx$, we need to prove that

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

i.e., that the PDF for $X \sim \mathcal{N}(0, 1)$ is normalized. Since the integrand is symmetric we need to show that

$$\int_0^{\infty} e^{-x^2/2}\, dx = \sqrt{\frac{\pi}{2}}.$$

Taking the square of this equation, we get

$$\int_0^{\infty} e^{-x^2/2}\, dx \int_0^{\infty} e^{-y^2/2}\, dy = \frac{\pi}{2}.$$

We turn the left-hand side into a double integral,

$$\int_0^{\infty} \left(\int_0^{\infty} e^{-(x^2+y^2)/2}\, dx \right) dy.$$

Changing variables $x = yt$,

$$\int_0^\infty \left(\int_0^\infty y e^{-y^2(1+t^2)/2} \, dt \right) dy.$$

Using Fubini's theorem, we interchange the order of integration,

$$\int_0^\infty \left(\int_0^\infty y e^{-y^2(1+t^2)/2} \, dy \right) dt.$$

Changing variables once again, $y^2(1+t^2)/2 = s$, so that $ds = y(1+t^2)\, dy$, we get

$$\int_0^\infty \left(\int_0^\infty \frac{e^{-s}}{1+t^2} \, ds \right) dt.$$

Integrating over s and then over t we get

$$\int_0^\infty \frac{dt}{1+t^2} = \tan^{-1} t\Big|_0^\infty = \frac{\pi}{2}.$$

$\blacksquare$

Proposition 6.2 *Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and set $Y = aX + b$, where $a > 0$. Then*

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2).$$

That is, the collection of normal random variables is closed unit linear transformations.

Proof: Consider the cumulative distribution function of Y,

$$\begin{aligned}
F_Y(y) &= P(Y \le y) = P(X \le a^{-1}(y-b)) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a^{-1}(y-b)} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) dx \\
&= \frac{1/a}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} \exp\left(-\frac{(a^{-1}(u-b)-\mu)^2}{2\sigma^2} \right) du \\
&= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \int_{-\infty}^{y} \exp\left(-\frac{(u-b-a\mu)^2}{2a^2\sigma^2} \right) du,
\end{aligned}$$

where we have changed variables, $x = a^{-1}(u-b)$. $\blacksquare$

Corollary 6.1 If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $(X - \mu)/\sigma$ is a standard normal variable.

Notation: The cumulative distribution function of a standard normal variable will be denoted by $\Phi(x)$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$

(This function is closely related to *Gauss' error function*).

A table of the values of $\Phi(x)$ is all that is needed to compute probabilities for general normal variables. Indeed, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$F_X(x) = P(X \le x) = P\left(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Because of the central importance of normal random variables, tables of Φ were given in many reference books before the appearance of computers.

Example: The duration of a normal pregnancy (in days) is a normal variable $\mathcal{N}(270, 100)$. What is the probability that a pregnancy lasts less than 240 days or more than 290 days?

Let X be the actual duration of the pregnancy. The question is

$$P(\{X > 290\} \cup \{X < 240\}) = ?,$$

which we solve as follows,

$$\begin{aligned} P(\{X > 290\} \cup \{X < 240\}) &= 1 - P(240 < X < 290) \\ &= 1 - P\Bigg(\underbrace{\frac{240 - 270}{10}}_{(-3)} < \underbrace{\frac{X - 270}{10}}_{\sim\mathcal{N}(0,1)} < \underbrace{\frac{290 - 270}{10}}_{2}\Bigg) \\ &= 1 - (\Phi(2) - \Phi(-3)) \approx 0.241. \end{aligned}$$

▲ ▲ ▲

The importance of the normal distribution stems from the *central limit theorem*, which we will encounter later on. The following theorem is an instance of the central limit theorem for a particular case:

Theorem 6.1 (DeMoivre-Laplace) Let (X_n) be a sequence of independent Bernoulli variables with parameter p and set

$$Y_n = \frac{X_n - p}{\sqrt{p(1-p)}}.$$

(The variables Y_n have zero expectation and unit variance.) Set then

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k.$$

Then S_n tends, as $n \to \infty$, to a standard normal variable in the sense that for every $a < b$,

$$\lim_{n\to\infty} P(a \le S_n \le b) = \Phi(b) - \Phi(a).$$

Comment: This theorem states that the sequence of random variables S_n converges to a standard normal variable *in distribution*, or *in law*.

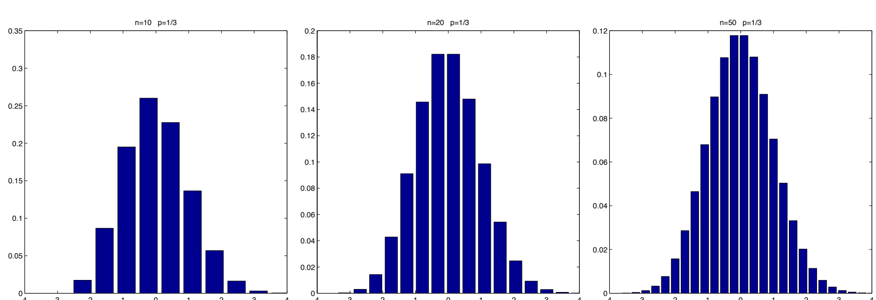
Proof: The event $\{a \le S_n \le b\}$ can be written as

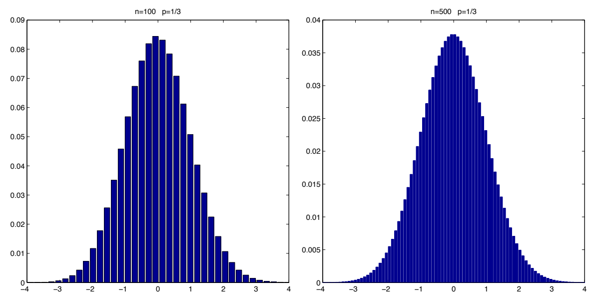
$$\left\{ np + \sqrt{np(1-p)}\, a \le \sum_{k=1}^{n} X_k \le np + \sqrt{np(1-p)}\, b \right\}.$$

The sum over X_k is a binomial variable $\mathscr{B}(n,p)$, so that we are trying to prove that

$$\frac{\mathscr{B}(n,p) - \mathbb{E}[\mathscr{B}(n,p)]}{\sigma(\mathscr{B}(n,p))} \xrightarrow{n\to\infty} \mathcal{N}(0,1)$$

The plots below show the (discrete!!!) distributions of "normalized" binomial variables for $p = 1/3$ and various values of n,





By the properties of the binomial distribution,

$$P\left(a \le S_n \le b\right) = \sum_{k=np+\sqrt{np(1-p)}\,a}^{np+\sqrt{np(1-p)}\,b} \binom{n}{k} p^k (1-p)^{n-k}.$$

(We will ignore the fact that limits are integer as the correction is negligible when $n \to \infty$.) As n becomes large (while p remains fixed), n, k, and $n-k$ become large, hence we use Stirling's approximation,

$$\binom{n}{k} \sim \frac{\sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi k}\, k^k e^{-k}\, \sqrt{2\pi(n-k)}\,(n-k)^{n-k} e^{-(n-k)}},$$

and so

$$\binom{n}{k} p^k (1-p)^{n-k} \sim \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k},$$

where, as usual, the $\sim$ relation means that the ratio between the two sides tends to one as $n \to \infty$. The summation variable k takes values that are of order $O(\sqrt{n})$ around np. This suggests a change of variables, $k = np + \sqrt{np(1-p)}\,m$, where m varies from a to b in units of $\Delta m = [np(1-p)]^{-1/2}$. Thus,

$$\begin{aligned} P\left(a \le S_n \le b\right) &= \frac{1}{\sqrt{2\pi}} \sum_{m=a}^{b} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k} \\ &= \frac{1}{\sqrt{2\pi}} \sum_{m=a}^{b} \underbrace{\sqrt{\frac{n^2 p(1-p)}{k(n-k)}}}_{I} \underbrace{\left(\frac{np}{k}\right)^k}_{II} \underbrace{\left(\frac{n(1-p)}{n-k}\right)^{n-k}}_{III} \Delta m. \end{aligned}$$

Consider the first term in the above product. As $n \to \infty$,

$$\lim_{n\to\infty} I = \lim_{n\to\infty} \frac{\sqrt{p(1-p)}}{\sqrt{(k/n)(1-k/n)}} = 1.$$

Consider the second term, which we can rewrite as

$$\left(\frac{np}{k}\right)^k = \left(\frac{np}{np + r\sqrt{n}}\right)^{np+r\sqrt{n}},$$

where $r = \sqrt{p(1-p)}\,m$. To evaluate the $n \to \infty$ limit it is easier to look at the logarithm of this expression, whose limit we evaluate using Taylor's expansion,

$$\begin{aligned}
\log\left(\frac{np}{k}\right)^k &= (np + r\sqrt{n}) \log\left(1 + \frac{r}{p} n^{-1/2}\right)^{-1} \\
&= -(np + r\sqrt{n}) \log\left(1 + \frac{r}{p} n^{-1/2}\right) \\
&= -(np + r\sqrt{n}) \left(\frac{r}{p} n^{-1/2} - \frac{r^2}{2p^2} n^{-1}\right) + \text{l.o.t} \\
&= -r\sqrt{n} - \frac{r^2}{2p} + \text{l.o.t} = -r\sqrt{n} - \frac{1}{2}(1-p)m^2 + \text{l.o.t},
\end{aligned}$$

where l.o.t stands for lower-order terms. Similarly,

$$\log\left(\frac{n(1-p)}{n-k}\right)^{n-k} = r\sqrt{n} - \frac{1}{2}pm^2 + \text{l.o.t.}$$

Combining the two together we have

$$\lim_{n\to\infty} \log\left[\left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k}\right] = -\frac{1}{2}m^2.$$

By the continuity of the exponential function,

$$\lim_{n\to\infty} P\left(a \le S_n \le b\right) = \lim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{m=a}^{b} e^{-m^2/2}\, \Delta m = \Phi(b) - \Phi(a),$$

where we have used the fact that an integral is a limit of Riemann sums. This concludes the proof. ■

Example: A fair coin is tossed 40 times. What is the probability that the number of Heads equals exactly 20?

Since the number of heads is a binomial variable, the answer is

$$\binom{40}{20}\left(\frac{1}{2}\right)^{20}\left(\frac{1}{2}\right)^{20} = 0.1268...$$

We can also approximate the answer using the DeMoivre-Laplace theorem,

$$\frac{X - 40 \times \frac{1}{2}}{\sqrt{40 \times \frac{1}{2} \times (1 - \frac{1}{2})}} \approx \mathcal{N}(0, 1).$$

The number of heads is a discrete variable, whereas the normal distribution refers to a continuous one. We will approximate the probability that the number of heads be 20 by the probability that it is, in a continuous context, between 19.5 and 20.5, i.e., that

$$-\frac{1}{2\sqrt{10}} \le \frac{X - 20}{\sqrt{10}} \le \frac{1}{2\sqrt{10}}.$$

Finally,

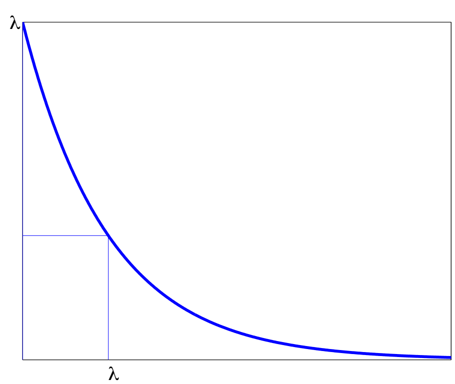
$$P\left(-\frac{1}{2\sqrt{10}} \le \mathcal{N}(0,1) \le \frac{1}{2\sqrt{10}}\right) \approx 2\left(\Phi\left(\frac{1}{2\sqrt{10}}\right) - \Phi(0)\right) = 0.127...$$

▲▲▲

6.4 The exponential distribution

Definition 6.4 A random variable X is said to be exponentially distributed *with parameter λ, denoted $X \sim \mathcal{E}xp(\lambda)$, if its* PDF *is*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}.$$



The corresponding cumulative distribution function is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

An exponential distribution is a suitable model in many situations, like the time until the next earthquake, or the time a hitchhiker has to wait for a car to stop (the the time a fisherman has to wait for the fish to bite).

Example: Suppose that the duration of a phone call (in minutes) is a random variable $\mathcal{Exp}(1/10)$. What is the probability that a given phone call lasts more than 10 minutes? The answer is

$$P(X > 10) = 1 - F_x(10) = e^{-10/10} \approx 0.368.$$

Suppose we know that a phone call has already lasted 10 minutes. What is the probability that it will last at least 10 more minutes. The perhaps surprising answer is

$$P(X > 20 \,|\, X > 10) = \frac{P(X > 20, X > 10)}{P(X > 10)} = \frac{e^{-2}}{e^{-1}} = e^{-1}.$$

More generally, we can show that for every $t > s$,

$$P(X > t \,|\, X > s) = P(X > t - s).$$

A random variable satisfying this property is called *memoryless*. ▲▲▲

Proposition 6.3 A random variable that satisfies

$$P(X > t \,|\, X > s) = P(X > t - s) \qquad \textit{for all } t > s > 0$$

is exponentially distributed.

Proof: It is given that

$$\frac{P(X > t, X > s)}{P(X > s)} = P(X > t - s),$$

or in terms of the cumulative distribution function,

$$\frac{1 - F_X(t)}{1 - F_X(s)} = 1 - F_X(t - s).$$

Let $g(t) = 1 - F_X(t)$, then for all $t > s$,

$$g(t) = g(s)g(t - s),$$

and the only family of functions satisfying this property is the exponentials. That is, there exists a number a, such that

$$g(t) = a^t = e^{t \log a},$$

or

$$F(t) = 1 - e^{-\log(1/a)\, t}.$$

■

6.5 The Gamma distribution

Recall that the Gamma function is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$

for $x > 0$; fwe have seen that or $n \in \mathbb{N}$,

$$\Gamma(n + 1) = n!,$$

A random variable X is said to be Gamma-distributed with parameters r, λ if it assumes positive values and

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}.$$

We denote it by $X \sim \mathcal{G}amma(r, \lambda)$. This is a normalized PDF since

$$\int_0^\infty f_X(x)\, dx = \frac{1}{\Gamma(r)} \int_0^\infty (\lambda x)^{r-1} e^{-\lambda x}\, d(\lambda x) = 1.$$

Note that for $r = 1$ we get the PDF of an exponential distribution, i.e.,

$$\mathcal{G}amma(1, \lambda) \sim \mathcal{E}xp(\lambda).$$

The significance of the Gamma distribution will be seen later on in this chapter.

6.6 The Beta distribution

A random variable assuming values in $[0,1]$ is said to have the Beta distribution with parameters $K, L > 0$, i.e., $X \sim \mathcal{B}eta(K, L)$, if it has the PDF

$$f_X(x) = \frac{\Gamma(K+L)}{\Gamma(K)\Gamma(L)} x^{K-1}(1-x)^{L-1}.$$

6.7 Functions of random variables

In this section we consider the following problem: let X be a continuous random variable with PDF $f_X(x)$. Let g be a real-valued function and let $Y(\omega) = g(X(\omega))$. What is the distribution of Y?

Example: Let $X \sim \mathcal{U}(0,1)$. What is the distribution of $Y = X^n$?
The random variable Y, like X, assumes values in the interval $[0,1]$. Now,

$$F_Y(y) = P(Y \le y) = P(X^n \le y) = P(X \le y^{1/n}) = F_X(y^{1/n}),$$

where we used the monotonicity of the power function for positive arguments. In the case of a uniform distribution,

$$F_X(x) = \int_{-\infty}^{x} f_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}.$$

Thus,

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ y^{1/n} & 0 \le y \le 1 \\ 1 & y > 1 \end{cases}.$$

Differentiating,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{1}{n} y^{1/n-1} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

▲▲▲

Example: Let X be a continuous random variable with PDF $f_X(x)$. What is the distribution of $Y = X^2$.

The main difference with the previous exercise is that X may possibly assume both positive and negative values, in which case the square function is non-monotonic. Thus, we need to proceed with more care,

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Differentiating, we get the PDF

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}.$$

▲▲▲

With these preliminaries, we can formulate the general theorem:

Theorem 6.2 Let X be a continuous random variable with PDF $f_X(x)$*. Let g be a strictly monotonic, differentiable function and set $Y(\omega) = g(X(\omega))$. Then the random variable Y has a* PDF

$$f_Y(y) = \begin{cases} |(g^{-1})'(y)|\, f_X(g^{-1}(y)) & y \text{ is in the range of } g(X) \\ 0 & \text{otherwise} \end{cases}.$$

Comment: If g is non-monotonic then $g^{-1}(y)$ may be set-valued and the above expression has to be replaced by a sum over all "branches" of the inverse function:

$$\sum_{g^{-1}(y)} \left|(g^{-1})'(y)\right| f_X(g^{-1}(y)).$$

Proof: Consider the case where g is strictly increasing. Then, g^{-1} exists, and

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y)),$$

and upon differentiation,

$$f_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = (g^{-1})'(y)\, f_X(g^{-1}(y)).$$

The case where g is strictly decreasing is handled similarly. ∎

The inverse transformation method An application of this formula is the following. Suppose that you have a computer program that generates a random variable $X \sim \mathcal{U}(0,1)$. How can we use it to generate a random variable with cumulative distribution function Ψ? The following method is known as the *inverse transformation method*.

If Ψ is strictly increasing (we know that it is at least non-decreasing), then we can define

$$Y(\omega) = \Psi^{-1}(X(\omega)).$$

Note that Ψ^{-1} maps $[0,1]$ onto the entire real line, while X has range $[0,1]$. Moreover, $F_X(x)$ is the identity on $[0,1]$. By the above formula,

$$F_Y(y) = F_X(\Psi(y)) = \Psi(y).$$

Example: Suppose we want to generate an exponential variable $Y \sim \mathcal{E}xp(\lambda)$, in which case $\Psi(y) = 1 - e^{-\lambda y}$. The inverse function is $\Psi^{-1}(x) = -\frac{1}{\lambda}\log(1-x)$, i.e., an exponential variable is generated by setting

$$Y(\omega) = -\frac{1}{\lambda}\log(1 - X(\omega)).$$

In fact, since $1 - X$ has the same distribution as X, we may equally well take $Y = -\lambda^{-1}\log X$.

▲▲▲

6.8 Multivariate distributions

We proceed to consider joint distributions of multiple random variables. The treatment is fully analogous to that for discrete variables.

Definition 6.5 *A pair of random variables X, Y over a probability space $(\Omega, \mathscr{F}, P)$ is said to have a* continuous joint distribution *if there exists an integrable non-negative bi-variate function $f_{X,Y}(x, y)$ (the* joint PDF*) (צפיפות התפלגות משותפת) such that for every (measurable) set $A \subseteq \mathbb{R}^2$,*

$$P_{X,Y}(A) = P(\{\omega : (X(\omega), Y(\omega)) \in A\}) = \iint_A f_{X,Y}(x, y)\, dxdy.$$

Note that in particular,

$$F_{X,Y}(x, y) = P(\{\omega : X(\omega) \le x, Y(\omega) \le y\}) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x, y)\, dxdy,$$

and consequently,

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y).$$

Furthermore, if X, Y are jointly continuous, then each random variable is continuous as a single variable. Indeed, for all $A \subseteq \mathbb{R}$,

$$P_X(A) = P_{X,Y}(A \times \mathbb{R}) = \int_A \left[\int_{\mathbb{R}} f_{X,Y}(x, y)\, dy \right] dx,$$

from which we identify the *marginal* PDF of X,

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy,$$

with an analogous expression for $f_Y(y)$. The generalization to *multivariate distributions* is straightforward.

Example: Consider a uniform distribution inside a circle of radius R,

$$f_{X,Y}(x, y) = \begin{cases} C & x^2 + y^2 \le R^2 \\ 0 & \text{otherwise} \end{cases}.$$

(1) What is C? (2) What is the marginal distribution of X? (3) What is the probability that the Euclidean norm of (X, Y) is less than a?

(1) The normalization condition is

$$\int_{x^2+y^2 \le R^2} C\, dxdy = \pi R^2 C = 1.$$

(2) For $|x| \le R$ the marginal PDF of X is given by

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy = \frac{1}{\pi R^2}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2}.$$

Finally,

$$P(X^2+Y^2 \le a^2) = \frac{a^2}{R^2}.$$

▲▲▲

Independence We next consider how does independence reflect in the joint PDF. Recall that X, Y are said to be independent if for all $A, B \subseteq \mathbb{R}$,

$$P_{X,Y}(A \times B) = P_X(A) P_Y(B).$$

For continuous distributions, this means that for all A, B,

$$\int_A \int_B f_{X,Y}(x,y)\,dxdy = \int_A f_X(x)\,dx \int_B f_Y(y)\,dy,$$

and since this should hold for every pair of sets A and B, we conclude that

$$f_{X,Y}(x,y) = f_X(x) f_Y(y).$$

Similarly, n random variables with continuous joint distribution are independent if their joint PDF equals to the product of their marginal PDFs.

Example: Let X, Y, Z be independent variables all being $\mathcal{U}(0,1)$. What is the probability that $X > YZ$?

The joint distribution of X, Y, Z is

$$f_{X,Y,Z}(x,y,z) = f_X(x) f_Y(y) f_Z(z) = \begin{cases} 1 & x,y,z \in [0,1] \\ 0 & \text{otherwise} \end{cases}.$$

Now,

$$\begin{aligned} P(X > YZ) &= \iint_{x>yz} dxdydz = \int_0^1 \int_0^1 \left(\int_{yz}^1 dx \right) dydz \\ &= \int_0^1 \int_0^1 (1 - yz)\,dydz = \int_0^1 \left(1 - \tfrac{z}{2}\right) dz = \tfrac{3}{4}. \end{aligned}$$

▲▲▲

Sums of independent random variables Let X, Y be independent continuous random variables. What is the distribution of $X + Y$?

We proceed by examining the cumulative distribution function of the sum

$$\begin{aligned} F_{X+Y}(z) = P(X+Y \le z) &= \iint_{x+y\le z} f_X(x) f_Y(y)\, dxdy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y)\, dxdy \\ &= \int_{-\infty}^{\infty} F_X(z-y) f_Y(y)\, dy. \end{aligned}$$

Differentiating, we obtain,

$$f_{X+Y}(z) = \frac{d}{dz} F_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy,$$

i.e., the PDF of a sum is the *convolution* of the PDFS, $f_{X+Y} = f_X * f_Y$.

Example: What is the distribution of $X+Y$ when $X, Y \sim \mathcal{U}(0,1)$ are independent? We have

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy = \int_{z-1}^{z} f_X(w)\, dw.$$

The integral vanishes if $z < 0$ and if $z > 2$. Otherwise,

$$f_{X+Y}(z) = \begin{cases} z & 0 \le z \le 1 \\ 2-z & 1 < z \le 2 \end{cases}.$$

▲ ▲ ▲

We conclude this section with a general formula for variable transformations. Let $\boldsymbol{X} = (X_1, X_2)$ be two random variables with joint PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$, and set

$$\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X}).$$

What is the joint PDF of $\boldsymbol{Y} = (Y_1, Y_2)$? We will assume that these relations are invertible, i.e., that

$$\boldsymbol{X} = \boldsymbol{g}^{-1}(\boldsymbol{Y}).$$

Furthermore, we assume that $\boldsymbol{g}$ is differentiable. Then,

$$F_{\boldsymbol{Y}}(\boldsymbol{y}) = \iint_{\boldsymbol{g}(\boldsymbol{x})\le \boldsymbol{y}} f_{\boldsymbol{X}}(\boldsymbol{x})\, dx_1 dx_2.$$

We change variables, $\boldsymbol{x} = \boldsymbol{g}^{-1}(\boldsymbol{u})$, and get

$$F_Y(\boldsymbol{y}) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f_X(\boldsymbol{g}^{-1}(\boldsymbol{u}))\, |J(\boldsymbol{u})|\, du_1 du_2,$$

where $J(\boldsymbol{y}) = \partial(\boldsymbol{x})/\partial(\boldsymbol{y})$ is the Jacobian of the transformation. Differentiating twice with respect to y_1, y_2 we obtain the joint PDF,

$$f_Y(\boldsymbol{y}) = |J(\boldsymbol{y})|\; f_X(\boldsymbol{g}^{-1}(\boldsymbol{u})).$$

✎ *Exercise 6.1* Let X_1, X_2 be two independent random variables with distribution $\mathcal{U}(0,1)$ (i.e., the variables that two subsequent calls of the `rand()` function on a computer would return). Define,

$$Y_1 = \sqrt{-2\log X_1}\, \cos(2\pi X_2)$$
$$Y_2 = \sqrt{-2\log X_1}\, \sin(2\pi X_2).$$

Show that Y_1 and Y_2 are independent and distributed $\mathcal{N}(0,1)$. This is the standard way of generating normally-distributed random variables on a computer. This change of variables is called the *Box-Muller transformation* (G.E.P. Box and M.E. Muller, 1958).

Example: Suppose that $X \sim \mathcal{Gamma}(K,1)$ and $Y \sim \mathcal{Gamma}(L,1)$ are independent, and consider the variables

$$V = \frac{X}{X+Y} \qquad \text{and} \qquad W = X + Y.$$

The reverse transformation is

$$X = VW \qquad \text{and} \qquad Y = W(1-V).$$

Since $X, Y \in [0,\infty)$ it follows that $V \in [0,1]$ and $W \in [0,\infty)$.
The Jacobian is

$$|J(v,w)| = \begin{vmatrix} w & -w \\ v & 1-v \end{vmatrix} = w.$$

Thus,

$$\begin{aligned} f_{V,W}(v,w) &= \frac{(vw)^{K-1}e^{-vw}}{\Gamma(K)} \frac{[w(1-v)]^{L-1}e^{-w(1-v)}}{\Gamma(L)} w \\ &= \frac{w^{K+L-1}e^{-w}}{\Gamma(K+L)} \times \frac{\Gamma(K+L)}{\Gamma(K)\Gamma(L)} v^{K-1}(1-v)^{L-1}. \end{aligned}$$

This means that

$$V \sim \mathcal{B}eta(K, L) \qquad \text{and} \qquad W \sim \mathcal{G}amma(K + L, 1).$$

Moreover, they are independent. ▲▲▲

Example: We now develop a general formula for the PDF *of ratios*. Let X, Y be random variables, not necessarily independent, and set

$$V = X \qquad \text{and} \qquad W = X/Y.$$

The inverse transformation is

$$X = V \qquad \text{and} \qquad Y = V/W.$$

The Jacobian is

$$|J(v,w)| = \begin{vmatrix} 1 & 1/w \\ 0 & -v/w^2 \end{vmatrix} = \left| \frac{v}{w^2} \right|.$$

Thus,

$$f_{V,W}(v,w) = f_{X,Y}\left(v, \frac{v}{w}\right) \left| \frac{v}{w^2} \right|,$$

and the uni-variate distribution of W is given by

$$f_W(w) = \int f_{X,Y}\left(v, \frac{v}{w}\right) \left| \frac{v}{w^2} \right| dv.$$

▲▲▲

✎ *Exercise 6.2* Find the distribution of X/Y when $X, Y \sim \mathcal{E}xp(1)$ are independent.

Example: Let $X \sim \mathcal{U}(0,1)$ and let Y be any (continuous) random variable independent of X. Define

$$W = X + Y \mod 1.$$

What is the distribution of W?

Clearly, W assumes value in $[0,1]$. We need to express the set $\{W \le c\}$ in terms of X, Y. If we decompose $Y = N + Z$, where $Z = Y \mod 1$, then

$$\begin{aligned} \{W \le c\} &= \{Z \le c\} \cap \{0 \le X \le c - Z\} \\ &\cup \{Z \le c\} \cap \{1 - Z \le X \le 1\} \\ &\cup \{Z > c\} \cap \{1 - Z \le X \le 1 - (Z - c)\} \end{aligned}$$

It follows that

$$P(W \le c) = \sum_{n=-\infty}^{\infty} \int_0^1 f_Y(n+y)\,[c\,I_{z\le c} + c\,I_{z>c}]\,dz = c,$$

i.e., no matter what Y is, $W \sim \mathcal{U}(0,1)$. ▲▲▲

6.9 Conditional distributions and conditional densities

Remember that in the case of *discrete* random variables we defined

$$p_{X\,|\,Y}(x|y) = P(X = x \,|\, Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

Since the PDF is, in a sense, the continuous counterpart of the point distribution, the following definition seems most appropriate:

Definition 6.6 The conditional probability density function *(CPDF) of X given Y is*

$$f_{X\,|\,Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

The question is what is the *meaning* is this conditional density? First, we note that viewed as a function of x, with y fixed, it is a density, as it is non-negative, and

$$\int_{\mathbb{R}} f_{X\,|\,Y}(x|y)\,dx = \frac{\int_{\mathbb{R}} f_{X,Y}(x,y)\,dx}{f_Y(y)} = 1.$$

Thus, it seems natural to speculate that the integral of the CPDF over a set A is the probability that $X \in A$ given that $Y = y$,

$$\int_A f_{X\,|\,Y}(x|y)\,dx \stackrel{?}{=} P(X \in A \,|\, Y = y).$$

The problem is that the right hand side is not defined, since the condition $(Y = y)$ has probability zero!

A heuristic way to resolve the problem is the following (for a rigorous way we need again measure theory): construct a sequence of sets $B_n \subset \mathbb{R}$, such that $B_n \to$

$\{y\}$ and each of the B_n has finite measure (for example, $B_n = (y - 1/n, y + 1/n)$), and define

$$P(X \in A \,|\, Y = y) = \lim_{n\to\infty} P(X \in A \,|\, Y \in B_n).$$

Now, the right-hand side is well-defined, provided the limit exists. Thus,

$$\begin{aligned} P(X \in A \,|\, Y = y) &= \lim_{n\to\infty} \frac{P(X \in A, Y \in B_n)}{P(Y \in B_n)} \\ &= \lim_{n\to\infty} \frac{\int_A \int_{B_n} f_{X,Y}(x,u)\, du dx}{\int_{B_n} f_Y(u)\, du} \\ &= \int_A \lim_{n\to\infty} \frac{\int_{B_n} f_{X,Y}(x,u)\, du}{\int_{B_n} f_Y(u)\, du}\, dx \\ &= \int_A \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dx, \end{aligned}$$

where we have used something analogous to l'Hopital's rule in taking the limit. This is precisely the identity we wanted to obtain.

What is the CPDF good for? We have the identity

$$f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y).$$

In many cases, it is more natural to define models in terms of conditional densities, and our formalism tells us how to convert this data into joint distributions.

Example: Let the joint PDF of X, Y be given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{y} e^{-x/y} e^{-y} & x, y \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

What is the CPDF of X given Y, and what is the probability that $X(\omega) > 1$ given that $Y = y$?

For $x, y \geq 0$ the CPDF is

$$f_{X|Y}(x|y) = \frac{\frac{1}{y} e^{-x/y} e^{-y}}{\int_0^\infty \frac{1}{y} e^{-x/y} e^{-y}\, dx} = \frac{1}{y} e^{-x/y},$$

and

$$P(X > 1 \,|\, Y = y) = \int_1^\infty f_{X|Y}(x|y)\, dx = \frac{1}{y} \int_1^\infty e^{-x/y}\, dx = e^{-1/y}.$$

▲ ▲ ▲

6.10 Expectation

Recall our definition of the expectation for discrete probability spaces,

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\, p(\omega),$$

where $p(\omega)$ is the point probability in Ω, i.e., $p(\omega) = P(\{\omega\})$. We saw that an equivalent definition was

$$\mathbb{E}[X] = \sum_{x\in S_x} x\, p_X(x).$$

In a more general context, the first expression is the integral of the function $X(\omega)$ over the probability space $(\Omega, \mathscr{F}, P)$, whereas the second equation is the integral of the identity function $X(x) = x$ over the probability space $(S_x, \mathscr{F}_X, P_X)$. We now want to generalize these definitions for uncountable spaces.

The definition of the expectation in the general case relies unfortunately on integration theory, which is part of measure theory. The expectation of X is defined as

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, P(d\omega),$$

but this is not supposed to make much sense to us. On the other hand, the equivalent definition,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\, P_X(dx),$$

does make sense if we identify $P_X(dx)$ with $f_X(x)\, dx$. That is, our definition of the expectation for continuous random variables is

$$\mathbb{E}[X] = \int_R x\, f_X(x)\, dx.$$

Example: For $X \sim \mathcal{U}(a,b)$,

$$\mathbb{E}[X] = \frac{1}{b-a}\int_a^b x\, dx = \frac{a+b}{2}.$$

▲ ▲ ▲

Example: For $X \sim \mathcal{E}xp(\lambda)$,

$$\mathbb{E}[X] = \int_0^\infty x\,\lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

▲ ▲ ▲

Example: For $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} x\, e^{-(x-\mu)^2/2\sigma^2}\, dx = \mu.$$

▲▲▲

Lemma 6.1 Let Y be a continuous random variable with PDF $f_Y(y)$*. Then*

$$\mathbb{E}[Y] = \int_0^\infty [1 - F_Y(y) - F_Y(-y)]\, dy.$$

Proof: Note that the lemma states that

$$\mathbb{E}[Y] = \int_0^\infty [P(Y > y) - P(Y \le -y)]\, dy.$$

We start with the first expression

$$\begin{aligned} \int_0^\infty P(Y > y)\, dy &= \int_0^\infty \int_y^\infty f_Y(u)\, du dy \\ &= \int_0^\infty \int_0^u f_Y(u)\, dy du \\ &= \int_0^\infty u\, f_Y(u)\, du, \end{aligned}$$

where the passage from the first to the second line involves a change in the order of integration, with the corresponding change in the limits of integration. Similarly,

$$\begin{aligned} \int_0^\infty P(Y \le -y)\, dy &= \int_0^\infty \int_{-\infty}^{-y} f_Y(u)\, du dy \\ &= \int_{-\infty}^0 \int_0^{-u} f_Y(u)\, dy du \\ &= -\int_{-\infty}^0 u\, f_Y(u)\, du. \end{aligned}$$

Subtracting the two expressions we get the desired result. ■

Theorem 6.3 (The unconscious statistician) *Let X be a continuous random variable and let $g : \mathbb{R} \to \mathbb{R}$. Then,*

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)\, dx.$$

Proof: In principle, we could write the PDF of $g(X)$ and follow the definition of its expected value. The fact that g does not necessarily have a unique inverse complicates the task. Thus, we use instead the previous lemma,

$$\begin{aligned}\mathbb{E}[g(X)] &= \int_0^\infty P(g(X) > y)\, dy - \int_0^\infty P(g(X) \le -y)\, dy \\ &= \int_0^\infty \int_{g(x)>y} f_X(x)\, dxdy - \int_0^\infty \int_{g(x)\le -y} f_X(x)\, dxdy.\end{aligned}$$

We now exchange the order of integration. Note that for the first integral,

$$\{0 < y < \infty, g(x) > y\} \qquad \text{can be written as} \qquad \{x \in \mathbb{R}, 0 < y < g(x)\}$$

whereas for the second integral,

$$\{0 < y < \infty, g(x) < -y\} \qquad \text{can be written as} \qquad \{x \in \mathbb{R}, 0 < y < -g(x)\}$$

Thus,

$$\begin{aligned}\mathbb{E}[g(X)] &= \int_{\mathbb{R}} \int_0^{\max(0,g(x))} f_X(x)\, dydx - \int_{\mathbb{R}} \int_0^{\max(0,-g(x))} f_X(x)\, dydx \\ &= \int_{\mathbb{R}} [\max(0, g(x)) - \max(0, -g(x))]\, f_X(x)\, dx \\ &= \int_{\mathbb{R}} g(x) f_X(x)\, dx.\end{aligned}$$

■

Example: What is the variance of $X \sim \mathcal{N}(\mu, \sigma^2)$?

$$\begin{aligned}\text{Var}[X] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}\, dx \\ &= \frac{\sigma^3}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} u^2 e^{-u^2/2}\, du = \sigma^2.\end{aligned}$$

▲▲▲

✎ *Exercise 6.3* Let $X \sim \mathcal{N}(0,\sigma^2)$. Calculate the moments $\mathbb{E}[X^k]$ (hint: consider separately the cases of odd and even k's).

The law of the unconscious statistician is readily generalized to multiple random variables, for example,

$$\mathbb{E}[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y) f_{X,Y}(x,y)\,dxdy.$$

✎ *Exercise 6.4* Show that if X and Y are independent continuous random variables, then for every two functions f, g,

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)].$$

6.11 The moment generating function

As for discrete variables the moment generating function is defined as

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} f_X(x)\,dx,$$

that is, it is the *Laplace transform* of the PDF. Without providing a proof, we state that the transformation $f_X \mapsto M_X$ is invertible (it is one-to-one), although the formula for the inverse is complicated and relies on complex analysis.

Comment: A number of other generating functions are commonly defined: first the *characteristic function*,

$$\varphi_X(t) = \mathbb{E}[e^{\imath tX}] = \int_{\mathbb{R}} e^{\imath tx} f_X(x)\,dx,$$

which unlike the moment generating function is always well defined for every t. Since its use relies on complex analysis we do not use it in this course. Another used generating function is the *probability generating function*

$$g_X(t) = \mathbb{E}[t^X] = \sum_x t^x p_X(x).$$

Example: What is the moment generating function of $X \sim \mathcal{N}(\mu, \sigma^2)$?

$$\begin{aligned} M_X(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{tx} e^{-(x-\mu)^2/2\sigma^2}\, dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp\left[-\frac{x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\mu^2/2\sigma^2} e^{(\mu+\sigma^2 t)^2/2\sigma^2} \int_{\mathbb{R}} \exp\left[-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right] dx \\ &= \exp\left[\mu t + \frac{\sigma^2}{2} t^2\right]. \end{aligned}$$

From this we readily obtain, say, the first two moments,

$$\mathbb{E}[X] = M_X'(0) = (\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2}\Big|_{t=0} = \mu,$$

and

$$\mathbb{E}[X^2] = M_X''(0) = \left[(\mu + \sigma^2 t)^2 + \sigma^2\right] e^{\mu t + \frac{1}{2}\sigma^2 t^2}\Big|_{t=0} = \sigma^2 + \mu^2,$$

as expected. ▲▲▲

Example: Recall the Gamma-distribution whose PDF is

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}.$$

To calculate its moments it is best to use the moment generating function,

$$M_X(t) = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty e^{tx} x^{r-1} e^{-\lambda x}\, dx = \frac{\lambda^r}{(\lambda - t)^r},$$

defined only for $t < \lambda$. We can then calculate the moment, e.g.,

$$\mathbb{E}[X] = M_X'(0) = \lambda^r\, r(\lambda - t)^{-(r+1)}|_{t=0} = \frac{r}{\lambda},$$

and

$$\mathbb{E}[X^2] = M_X''(0) = \lambda^r\, r(r+1)(\lambda - t)^{-(r+2)}|_{t=0} = \frac{r(r+1)}{\lambda^2},$$

from which we conclude that

$$\mathrm{Var}[X] = \frac{r}{\lambda^2}.$$

▲▲▲

From the above discussion it follows that the moment generating function embodies the same information as the PDF. A nice property of the moment generating function is that it converts convolutions into products. Specifically,

Proposition 6.4 Let f_X and f_Y be probability densities functions and let $f = f_X * f_Y$ be their convolution. If M_X, M_Y and M are the moment generating functions associated with f_X, f_Y and f, respectively, then $M = M_X M_Y$.

Proof: By definition,

$$\begin{aligned} M(t) &= \int_{\mathbb{R}} e^{tx} f(x)\, dx = \int_{\mathbb{R}} e^{tx} \int_{\mathbb{R}} f_X(y) f_Y(x-y)\, dy\, dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{ty} f_X(y) e^{t(x-y)} f_Y(x-y)\, dy\, d(x-y) \\ &= \int_{\mathbb{R}} e^{ty} f_X(y)\, dy \int_{\mathbb{R}} e^{tu} f_Y(u)\, du = M_X(t) M_Y(t). \end{aligned}$$

■

Example: Here is an application of the above proposition. Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent variables. We have already calculated their moment generating function,

$$\begin{aligned} M_X(t) &= \exp\left[\mu_1 t + \frac{\sigma_1^2}{2} t^2\right] \\ M_Y(t) &= \exp\left[\mu_2 t + \frac{\sigma_2^2}{2} t^2\right]. \end{aligned}$$

By the above proposition, the generating function of their sum is the product of the generating functions,

$$M_{X+Y}(t) = \exp\left[(\mu_1 + \mu_2) t + \frac{\sigma_1^2 + \sigma_2^2}{2} t^2\right],$$

from which we conclude at once that

$$X + Y \sim \mathcal{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right),$$

i.e., sums of independent normal variables are normal. ▲▲▲

Example: Consider now the sum of n independent exponential random variables $X_i \sim \mathcal{E}xp(\lambda)$. Since $\mathcal{E}xp(\lambda) \sim \mathcal{G}amma(1,\lambda)$ we know that

$$M_{X_i}(t) = \frac{\lambda}{\lambda - t}.$$

The PDF of the sum of n independent random variables,

$$Y = \sum_{i=1}^{n} X_i$$

is the n-fold convolution of their PDFS, and its generating function is the product of their generating functions,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \frac{\lambda^n}{(\lambda - t)^n},$$

which we identify as the generating function of the $\mathcal{G}amma(n,\lambda)$ distribution. Thus *the Gamma distribution with parameters (n,λ) characterizes the sum of n independent exponential variables with parameter λ.* ▲▲▲

✎ *Exercise 6.5* What is the distribution of $X_1 + X_2$ where $X_1 \sim \mathcal{G}amma(r_1,\lambda)$ and $X_2 \sim \mathcal{G}amma(r_2,\lambda)$ are independent?

Example: A family of distributions that have an important role in statistics are the χ^2_ν distributions with $\nu = 1, 2 \dots$. A random variable Y has the χ^2_ν-distribution if it is distributed like

$$Y \sim X_1^2 + X_2^2 + \cdots + X_\nu^2,$$

where $X_i \sim \mathcal{N}(0,1)$ are independent.

The distribution of X_1^2 is obtained by the change of variable formula,

$$f_{X_1^2}(x) = \frac{f_{X_1}(\sqrt{x}) + f_{X_1}(-\sqrt{x})}{2\sqrt{x}} = 2\frac{\frac{1}{\sqrt{2\pi}}e^{-x/2}}{2\sqrt{x}} = \frac{1}{\sqrt{2\pi x}}e^{-x/2}.$$

The moment generating function is

$$M_{X_1^2}(t) = \int_0^\infty e^{tx}\frac{1}{\sqrt{2\pi x}}e^{-x/2}\,dx = \frac{2}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{2}(1-2t)y^2}\,dy = (1-2t)^{-1/2},$$

and by the addition rule, the moment generating function of the χ^2_ν-distribution is

$$M_Y(t) = (1-2t)^{-\nu/2} = \frac{(1/2)^{\nu/2}}{(1/2 - t)^{\nu/2}}.$$

We identify this moment generating function as that of $\mathcal{G}amma(\nu/2, 1/2)$. ▲▲▲

6.12 Other distributions

We conclude this section with two distributions that have major roles in statistics. Except for the additional exercise in the change of variable formula, the goal is to know the definition of these very useful distributions.

Definition 6.7 Let $X \sim \chi^2_r$ and $Y \sim \chi^2_s$ be independent. A random variable that has the same distribution as

$$W = \frac{X/r}{Y/s}$$

is said to have the Fischer $F_{r,s}$ distribution.

Since, by the previous section

$$f_{X/r}(x) = \frac{(r/2)^r (rx)^{r/2-1} e^{-\frac{1}{2}rx}}{\Gamma(\frac{r}{2})} r$$

$$f_{Y/s}(y) = \frac{(s/2)^s (sy)^{s/2-1} e^{-\frac{1}{2}sy}}{\Gamma(\frac{s}{2})} s,$$

it follows from the distribution of ratios formula that

$$\begin{aligned} f_W(w) &= \int_0^\infty \frac{\left(\frac{1}{2}\right)^{r/2} (rv)^{r/2-1} e^{-\frac{1}{2}rv}}{\Gamma(\frac{r}{2})} r \frac{\left(\frac{1}{2}\right)^{s/2} (sv/w^2)^{s/2-1} e^{-\frac{1}{2}sv/w^2}}{\Gamma(\frac{s}{2})} s \frac{v}{w^2}\, dv \\ &= \frac{1}{\Gamma(\frac{r}{2})\Gamma(\frac{s}{2})} \frac{(\frac{1}{2}r)^{r/2}(\frac{1}{2}s)^{s/2}}{w^s} \int_0^\infty v^{r/2+s/2-1} e^{-\frac{1}{2}v(r+s/w^2)}\, dv. \end{aligned}$$

Changing variables we get

$$\begin{aligned} f_W(w) &= \frac{1}{\Gamma(\frac{r}{2})\Gamma(\frac{s}{2})} \frac{(\frac{1}{2}r)^{r/2}(\frac{1}{2}s)^{s/2}}{w^s} \left[\frac{1}{2}(r+s/w^2)\right]^{-(r/2+s/2)} \int_0^\infty \xi^{r/2+s/2-1} e^{-\xi}\, d\xi \\ &= \frac{\Gamma(\frac{r}{2}+\frac{s}{2})}{\Gamma(\frac{r}{2})\Gamma(\frac{s}{2})} \frac{(\frac{1}{2}r)^{r/2}(\frac{1}{2}s)^{s/2}}{w^s} \left[\frac{1}{2}(r+s/w^2)\right]^{-(r/2+s/2)} \\ &= \frac{\Gamma(\frac{r}{2}+\frac{s}{2})}{\Gamma(\frac{r}{2})\Gamma(\frac{s}{2})} \frac{r^{r/2}s^{s/2}}{w^s(r+s/w^2)^{\frac{1}{2}(r+s)}}. \end{aligned}$$

Definition 6.8 Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2_\nu$ be independent. A random variable that has the same distribution as

$$W = \frac{X}{\sqrt{Y/\nu}}$$

is said to have the Student's t_ν distribution.

✎ *Exercise 6.6* Find the PDF of the Student's t_ν distribution.