1. SIMULATION - II

Drawing random numbers from a pdf: Clearly, the procedure used for generating from a pmf is inapplicable here. First start with two examples. As before, U is a Unif($[0,1]$) random variable.

Example 1. Suppose we want to draw from the Unif($[3,7]$) distribution. Set $X = 4U + 3$. Clearly,

$$F(t) := \mathbf{P}\{X \le t\} = \mathbf{P}\{U \le \frac{t-3}{4}\} = \begin{cases} 0 & \text{if } t < 3 \\ (t-3)/4 & \text{if } 3 \le t \le 7 \\ 1 & \text{if } t > 7. \end{cases}$$

This is precisely the CDF of Unif($[3,7]$) distribution.

Example 2. Here, let us do the opposite, just take some function of a uniform variable and see what CDF we get. Let $\psi(t) = t^3$ and let $X = \varphi(U) = U^3$. Then,

$$F(t) := \mathbf{P}\{X \le t\} = \mathbf{P}\{U \le t^{1/3}\} = \begin{cases} 0 & \text{if } t < 0 \\ t^{1/3} & \text{if } 0 \le t \le 1 \\ 1 & \text{if } t > 1. \end{cases}$$

Differentiating the CDF, we get the density

$$f(t) = F'(t) = \begin{cases} \frac{1}{3}t^{-2/3} & \text{if } 0 < t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The derivative does not exist at 0 and 1, but as remarked earlier, it does not matter if we change the value of the density at finitely many points (as the integral over any interval will remain the same). Anyway, we notice that the density is that of Beta($1/3, 1$). Hence, $X \sim$ Beta($1/3, 1$).

This gives us the idea that to generate random number from a continuous CDF F, we should find a function $\psi : [0,1] \to \mathbb{R}$ such that $X := \psi(U)$ has the distribution F. How to find the distribution of X?

Lemma 3. *Let $\psi : (0,1) \to \mathbb{R}$ be a strictly increasing function with $a = \psi(0+)$ and $b = \psi(1-)$. Let $X = \psi(U)$ with $U \sim U(0,1)$. Then, X has CDF*

$$F(t) = \begin{cases} 0 & \text{if } t \le a \\ \psi^{-1}(t) & \text{if } a < t < b \\ 1 & \text{if } t \ge b. \end{cases}$$

If ψ is also differentiable and the derivative does not vanish anywhere (or, vanishes at finitely many points only), then X has pdf

$$f(t) = \begin{cases} \left(\psi^{-1}\right)'(t) & \text{if } a < t < b \\ 0 & \text{if } t \notin (a, b). \end{cases}$$

Proof. Since ψ is strictly increasing, $\psi(u) \le t$ if and only if $u \le \psi^{-1}(t)$. Hence,

$$F(t) = \mathbf{P}\{X \le t\} = \mathbf{P}\{U \le \psi^{-1}(t)\} = \begin{cases} 0 & \text{if } t \le a \\ \psi^{-1}(t) & \text{if } a < t < b \\ 1 & \text{if } t \ge b. \end{cases}$$

■

From this lemma, we immediately get the following rule for generating random numbers from a density.

How to simulate from a CDF: Let F be a CDF that is strictly increasing on an interval $[A, B]$, where $F(A) = 0$ and $F(B) = 1$ (it is allowed to take $A = -\infty$ and/or $B = +\infty$). Then, define $\psi : (0, 1) \to (A, B)$ as $\psi(u) = F^{-1}(u)$. Let $U \sim$ Unif([0, 1]) and let $X = \psi(U)$. Then, X has CDF equal to F.

This follows from the lemma because ψ is defined as the inverse of F, and hence F (restricted to (A, B)) is the inverse of ψ. Further, as the inverse of a strictly increasing function, the function ψ is also strictly increasing.

Example 4. Consider the Exponential distribution with parameter λ whose CDF is

$$F(t) = \begin{cases} 0 & \text{if } t \le 0 \\ 1 - e^{-\lambda t} & \text{if } t > 0. \end{cases}$$

Take $A = 0$ and $B = +\infty$. Then F is increasing on $(0, \infty)$ and its inverse is the function $\psi(u) = -\frac{1}{\lambda}\log(1-u)$. Thus to simulate a random number from Exp(λ) distribution, we set $X = -\frac{1}{\lambda}\log(1-U)$.

When the CDF is NOT explicitly available as a function, we can still adopt the above procedure, but only numerically. Consider an example.

Example 5. Suppose $F = \Phi$, the CDF of $N(0, 1)$ distribution. Then, we do not have an explicit form for either Φ or for its inverse Φ^{-1}. With a computer we can do the following. Pick a large number of closely placed points, for example divide the interval $[-5, 5]$ into 1000 equal intervals of length 0.01 each. Let the endpoints of these intervals be labelled $t_0 < t_1 < \cdots < t_{1000}$. For each

i, calculate $\Phi(t_i) = \int_{-\infty}^{t_i} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ using numerical methods for integration, say the numerical value obtained is w_i. This is done only once, and create the table of values

$$\begin{array}{cccccc} t_0 & t_1 & t_2 & \dots & \dots & t_{1000} \\ w_0 & w_1 & w_2 & \dots & \dots & w_{1000} \end{array}.$$

Now, draw a uniform random number U. Look up the table and find the value of i for which $w_i < U < w_{i+1}$. Then set $X = t_i$. If it so happens that $U < w_0$, set $X = t_0 = -5$ and if $U > w_{1000}$ set $X = t_{1000} = 5$. But since $\Phi(-5) < 0.00001$ and $\Phi(5) > 0.99999$, it is highly unlikely that the last two cases will occur. The random variable X has a distribution close to $N(0, 1)$.

Exercise 6. Give an explicit method to draw random numbers from the following densities.

(1) Cauchy distribution with density $\frac{1}{\pi(1+x^2)}$.

(2) Beta($\frac{1}{2}, \frac{1}{2}$) density $\frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$ on $[0, 1]$ (and zero elsewhere).

(3) Pareto(α) distribution which by definition has the density

$$f(t) = \begin{cases} \alpha t^{-\alpha-1} & \text{if } t \geq 1, \\ 0 & \text{if } t < 1. \end{cases}$$

Check the file 'Distributions_continuous.R'.

Remark 7. We have conveniently skipped the question of how to draw random numbers from the uniform distribution in the first place. This is a difficult topic and various results, proved and unproved, are used in generating such numbers.

2. JOINT DISTRIBUTIONS

In many situations we study several random variables at once. In such a case, knowing the individual distributions is not sufficient to answer all relevant questions. This is like saying that knowing $\mathbf{P}(A)$ and $\mathbf{P}(B)$ is insufficient to calculate $\mathbf{P}(A \cap B)$ or $\mathbf{P}(A \cup B)$ etc.

Definition 8 (Joint distribution)**.** Let $X_1, X_2, \ldots, X_m$ be random variables on the same probability space. We call $\mathbf{X} = (X_1, \ldots, X_m)$ a *random vector*, as it is just a vector of random variables. The CDF of $\mathbf{X}$, also called the joint CDF of $X_1, \ldots, X_m$ is the function $F_{\mathbf{X}} : \mathbb{R}^m \to [0,1]$ defined as

$$F_{\mathbf{X}}(t_1, \ldots, t_m) = \mathbf{P}\{X_1 \le t_1, \ldots, X_m \le t_m\} = \mathbf{P}\left\{\bigcap_{i=1}^{m}\{X_i \le t_i\}\right\}.$$

Exercise 9. Consider two events A and B in the probability space and let $X = \mathbf{1}_A$ and $Y = \mathbf{1}_B$ be their indicator random variables. Find their joint CDF.

Properties of joint CDFs: The following properties of the joint CDF $F_{\mathbf{X}} : \mathbb{R}^m \to [0,1]$ are analogous to those of the 1-dimensional CDF and the proofs are similar.

(1) $F_{\mathbf{X}}$ is increasing in each co-ordinate, i.e., if $s_1 \le t_1, \ldots, s_m \le t_m$, then $F_{\mathbf{X}}(s_1, \ldots, s_m) \le F_{\mathbf{X}}(t_1, \ldots, t_m)$.

(2) $\lim F_{\mathbf{X}}(t_1, \ldots, t_m) = 0$ if $\max\{t_1, \ldots, t_m\} \to -\infty$ (i.e., one of the t_i goes to $-\infty$).

(3) $\lim F_{\mathbf{X}}(t_1, \ldots, t_m) = 1$ if $\min\{t_1, \ldots, t_m\} \to +\infty$ (i.e., all of the t_i goes to $+\infty$).

(4) $F_{\mathbf{X}}$ is right continuous in each co-ordinate. That is $F_{\mathbf{X}}(t_1+h_1, \ldots, t_m+h_m) \to F_{\mathbf{X}}(t_1, \ldots, t_m)$ as $h_i \to 0^+$ for $i = 1, \ldots, m$.

Conversely, any function having these four properties is the joint CDF of some random variable(s).

Remark 10. Recall that the increasing property in $\mathbb{R}$ is $F(t_1) - F(s_1) = \mathbf{P}(s_1 < X_1 \le t_1) \ge 0$ for $s_1 \le t_1$. We draw a similar analogy to generalize this to $\mathbb{R}^2$, and need to verify the following:

$$\mathbf{P}(s_1 < X_1 \le t_1, s_2 < X_2 \le t_2) = F(t_1, t_2) - F(s_1, t_2) - F(t_1, s_2) + F(s_1, s_2) \ge 0$$

for $s_1 \le t_1, s_2 \le t_2$. A general expression can be derived for $\mathbb{R}^m$, but it is more complicated.

From the joint CDF, it is easy to recover the individual CDFs. Indeed, if $F_{\mathbf{X}} : \mathbb{R}^m \to [0,1]$ is the CDF of $\mathbf{X} = (X_1, \ldots, X_m)$, then the CDF of X_1 is given by $F_1(t) := F_{\mathbf{X}}(t, +\infty, \ldots, +\infty) := \lim F_{\mathbf{X}}(t, s_2, \ldots, s_m)$ as $s_i \to +\infty$ for each $i = 2, \ldots, m$. This is true because if $A_n := \{X_1 \le t\} \cap \{X_2 \le n\} \cap \cdots \cap \{X_m \le n\}$, then the events A_n increase to the event $A = \{X_1 \le t\}$ as $n \to \infty$. Hence, $\mathbf{P}(A_n) \to \mathbf{P}(A)$. But, $\mathbf{P}(A_n) = F_{\mathbf{X}}(t, n, \ldots, n)$ and $\mathbf{P}(A) = F_1(t)$. Thus, we see that $F_1(t) := F_{\mathbf{X}}(t, +\infty, \ldots, +\infty)$.

More generally, we can recover the joint CDF of any subset of $X_1, \ldots, X_n$. For example, the joint CDF of $X_1, \ldots, X_k$ is just $F_{\mathbf{X}}(t_1, \ldots, t_k, +\infty, \ldots, +\infty)$.

Joint pmf and pdf: Just like in the case of one random variable, we can consider the following two sub-classes of vector of random variables.

(1) Distributions with a pmf. These are CDFs for which there exist points $\mathbf{t}_1, \mathbf{t}_2, \ldots$ in $\mathbb{R}^m$ and non-negative numbers w_i such that $\sum_i w_i = 1$ (often we write $f(\mathbf{t}_i)$ in place of w_i). For every $\mathbf{t} \in \mathbb{R}^m$, we have

$$F(\mathbf{t}) = \sum_{i\,:\,\mathbf{t}_i \le \mathbf{t}} w_i,$$

where $\mathbf{s} \le \mathbf{t}$ means that each co-ordinate of $\mathbf{s}$ is less than, or equal to the corresponding co-ordinate of $\mathbf{t}$.

(2) Distributions with a pdf. These are CDFs for which there is a non-negative function (may assume piecewise continuous for convenience) $f : \mathbb{R}^m \to \mathbb{R}_+$ such that for every $\mathbf{t} \in \mathbb{R}^m$ we have

$$F(\mathbf{t}) = \int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_m} f(u_1, \ldots, u_m)\, du_1 \ldots du_m.$$

We give two examples, one of each kind.

Example 11. (Multinomial distribution). Fix parameters r, m (two positive integers) and $p_1, \ldots, p_m$ (positive numbers that add to 1). The *multinomial pmf* with these parameters is given by

$$f(k_1, \ldots, k_{m-1}) = \frac{r!}{k_1! k_2! \cdots k_{m-1}! (r - \sum_{i=1}^{m-1} k_i)!} p_1^{k_1} \ldots p_{m-1}^{k_{m-1}} p_m^{r - \sum_{i=1}^{m-1} k_i},$$

if $k_i \ge 0$ are integers such that $k_1 + \cdots + k_{m-1} \le r$. [1] One situation where this distribution arises is when r balls are randomly placed in m bins, with each ball going into the jth bin with probability p_j, and we look at the random vector $(X_1, \ldots, X_{m-1})$, where X_k is the number of balls that fell into the kth bin. This random vector has the multinomial pmf.

In this case, the marginal distribution of X_k is Bin(r, p_k). More generally, $(X_1, \ldots, X_\ell)$ has multinomial distribution with parameters $r, \ell, p_1, \ldots, p_\ell, p_0$, where $p_0 = 1 - (p_1 + \cdots + p_\ell)$. This is easy to prove, but even easier to see from the balls in bins interpretation (just think of the last $n - \ell$ bins as one).

[1] In some books, the distribution of $(X_1, \ldots, X_m)$ is called the multinomial distribution. This has the pmf

$$g(k_1, \ldots, k_m) = \frac{r!}{k_1! k_2! \cdots k_{m-1}! k_m!} p_1^{k_1} \cdots p_{m-1}^{k_{m-1}} p_m^{k_m},$$

where k_i are non-negative integers such that $k_1 + \cdots + k_m = r$. We have chosen our convention so that the binomial distribution is a special case of the multinomial.

Example 12. (Bivariate normal distribution). Consider a density function on $\mathbb{R}^2$ given by

$$f(x,y) = \frac{\sqrt{ab-c^2}}{2\pi} e^{-\frac{1}{2}\left[a(x-\mu)^2+b(y-\nu)^2+2c(x-\mu)(y-\nu)\right]},$$

where μ, ν, a, b, c are real parameters. We shall impose the conditions that $a > 0$, $b > 0$ and $ab - c^2 > 0$ (otherwise the above does not give a density, as we shall see).

The first thing is to check that this is indeed a density. We recall the one-dimensional Gaussian integral

$$\int_{-\infty}^{+\infty} e^{-\frac{\tau}{2}(x-a)^2} dx = \sqrt{2\pi}\frac{1}{\sqrt{\tau}} \text{ for any } \tau > 0 \text{ and any } a \in \mathbb{R}. \tag{1}$$

We shall take $\mu = \nu = 0$ (how do you compute the integral if they are not?). Then, the exponent in the density has the form

$$ax^2 + by^2 + 2cxy = b\left(y + \frac{cx}{b}\right)^2 + \left(a - \frac{c^2}{b}\right)x^2.$$

Therefore,

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[ax^2+by^2+2cxy\right]} dy &= e^{-\frac{1}{2}(a-\frac{c^2}{b})x^2} \int_{-\infty}^{\infty} e^{-\frac{b}{2}(y+\frac{cx}{b})^2} dy \\ &= e^{-\frac{1}{2}(a-\frac{c^2}{b})x^2} \frac{\sqrt{2\pi}}{\sqrt{b}} \end{aligned}$$

by (1) but ony if $b > 0$. Now, we integrate over x and use (1) again (also the fact that $a - \frac{c^2}{b} > 0$) to get

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[a(x-\mu)^2+b(y-\nu)^2+2c(x-\mu)(y-\nu)\right]} dy dx &= \frac{\sqrt{2\pi}}{\sqrt{b}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(a-\frac{c^2}{b})x^2} dx \\ &= \frac{\sqrt{2\pi}}{\sqrt{b}} \frac{\sqrt{2\pi}}{\sqrt{a-\frac{c^2}{b}}} = \frac{2\pi}{\sqrt{ab-c^2}}. \end{aligned}$$

This completes the proof that $f(x,y)$ is indeed a density. Note that $b > 0$ and $ab - c^2 > 0$ also implies that $a > 0$.

Matrix form of writing the density: Let $\Sigma^{-1} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$. Then, $\det(\Sigma) = \frac{1}{\det(\Sigma^{-1})} = \frac{1}{ab-c^2}$. Hence, we may re-write the density above as (let $\mathbf{u}$ be the column vector with co-ordinates x, y)

$$f(x,y) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}\mathbf{u}^t\Sigma^{-1}\mathbf{u}}.$$

The conditions $a > 0, b > 0, ab - c^2 > 0$ translate precisely to what is called positive-definiteness. One way to say it is that Σ is a symmetric matrix and all its eigenvalues are strictly positive.

Final form: We can now introduce an extra pair of parameters μ_1, μ_2 and define a density

$$f(x,y) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{u}-\mu)^t\Sigma^{-1}(\mathbf{u}-\mu)},$$

where μ is a column vector with co-ordinates μ_1, μ_2. This is the full bi-variate normal density. Along similar lines, we can talk of the m-variate normal density.

Example 13. (A class of examples). Let $f_1, f_2, \ldots, f_m$ be one-variable densities. In other words, $f_i : \mathbb{R} \to \mathbb{R}_+$ and $\int_{-\infty}^{\infty} f_i(x)dx = 1$. Then, we can make a multivariate density as follows. Define $f : \mathbb{R}^m \to \mathbb{R}_+$ by

$$f(x_1, \ldots, x_m) = f_1(x_1) \cdots f_m(x_m).$$

Then, f is a density.

If X_i are random variables on a common probability space and the joint density of $\mathbf{X} = (X_1, \ldots, X_m)$ is $f(x_1, \ldots, x_m)$, then we say that X_i are *independent random variables*. It is easy to see that the marginal density of X_i is f_i for $i = 1, \ldots, m$. It is also the case that the joint CDF factors as

$$F_{\mathbf{X}}(x_1, \ldots, x_m) = F_{X_1}(x_1) \cdots F_{X_m}(x_m).$$

Example 14. (A second class of examples). Let g be a pdf. In other words, $g : \mathbb{R} \to \mathbb{R}_+$ and $\int_{-\infty}^{\infty} g(x)dx = 1$. Then, we can make a multivariate density as follows. Define $f : \mathbb{R}^m \to \mathbb{R}_+$ by

$$f(x_1, \ldots, x_m) = g(x_1) \cdots g(x_m).$$

Then, f is a density.

If X_i are random variables on a common probability space and the joint density of $\mathbf{X} = (X_1, \ldots, X_m)$ is $f(x_1, \ldots, x_m)$, then we say that X_i are *independent and identically distributed (i.i.d.) random variables*. It is easy to see that the marginal density of X_i is g for $i = 1, \ldots, m$. It is also the case that the joint CDF factors as

$$F_{\mathbf{X}}(x_1, \ldots, x_m) = G(x_1) \cdots G(x_m).$$

CDF technique: Let $\mathbf{X} = (X_1, \dots, X_m)$ be a random vector and let g be a function such that $g : \mathbb{R}^m \to \mathbb{R}$. The distribution $Y = g(X_1, \dots, X_m)$ can be determined by computing the distribution function

$$F_Y(y) = \mathbf{P}\left(\{g(X_1, \dots, X_m) \le y\}\right), \quad -\infty < y < \infty.$$

Example 15. Let X_1, X_2 be i.i.d. from the $U(0,1)$ distribution. Find the distribution function of $Y = X_1 + X_2$. Hence, find the pdf of Y.

The joint pdf of (X_1, X_2) is given by

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) \\ &= \begin{cases} 1, & \text{if } 0 < x_1 < 1, 0 < x_2 < 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Therefore, the distribution function of Y is given by

$$\begin{aligned} F_Y(x) &= \mathbf{P}\left(\{X_1 + X_2 \le x\}\right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) \mathbf{1}_{(-\infty,x]}(x_1 + x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \int_0^1 \int_0^1 \mathbf{1}_{(0,x]}(x_1 + x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{2} \times x \times x, & \text{if } 0 \le x < 1 \\ 1 - \frac{1}{2} \times (2 - x) \times (2 - x), & \text{if } 1 \le x < 2 \\ 1, & \text{if } x \ge 2 \end{cases} \\ &= \begin{cases} 0, & \text{if } x < 0 \\ \frac{x^2}{2}, & \text{if } 0 \le x < 1 \\ \frac{4x - x^2 - 2}{2}, & \text{if } 1 \le x < 2 \\ 1, & \text{if } x \ge 2. \end{cases} \end{aligned}$$

Clearly, F_Y is differentiable everywhere except on a finite set $\{0, 1, 2\}$. Let

$$g(x) = \begin{cases} x, & \text{if } 0 < x < 1 \\ 2 - x, & \text{if } 1 < x < 2 \\ 0, & \text{otherwise,} \end{cases}$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}x} F_Y(x) = g(x) \; \forall\, x \in \mathbb{R} \setminus \{0, 1, 2\} \text{ and } \int_{-\infty}^{\infty} g(x) \mathrm{d}x = 1.$$

Change of variable for discrete distributions: The idea is quite similar to the univariate case, and will be skipped. One example is discussed.

Example 16. Let $\mathbf{X} = (X_1, X_2, X_3)$ be a discrete type random vector with pmf

$$f_{\mathbf{X}}(x_1, x_2, x_3) = \begin{cases} \frac{2}{9}, & \text{if } (x_1, x_2, x_3) \in \{(1,1,0), (1,0,1), (0,1,1)\} \\ \frac{1}{3}, & \text{if } (x_1, x_2, x_3) = (1,1,1) \\ 0, & \text{otherwise.} \end{cases}$$

Define $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Find the marginal pmf of Y_1 and Y_2.

We have

$$\mathbf{P}(\{Y_1 = y\}) = \mathbf{P}(\{X_1 + X_2 = y\}) = 0, \quad \text{if } y \notin \{1, 2\}.$$

$$\begin{aligned} \mathbf{P}(\{Y_1 = 1\}) &= \mathbf{P}(\{X_1 + X_2 = 1\}) \\ &= \mathbf{P}(\{(X_1, X_2, X_3) \in \{(1,0,1), (0,1,1)\}\}) \\ &= \mathbf{P}(\{(X_1, X_2, X_3) = (1,0,1)\}) + \mathbf{P}(\{(X_1, X_2, X_3) = (0,1,1)\}) \\ &= \tfrac{4}{9}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{P}(\{Y_1 = 2\}) &= \mathbf{P}(\{X_1 + X_2 = 2\}) \\ &= \mathbf{P}(\{(X_1, X_2, X_3) = (1,1,0)\}) + \mathbf{P}(\{(X_1, X_2, X_3) = (1,1,1)\}) \\ &= \tfrac{5}{9}. \end{aligned}$$

Therefore,

$$f_{Y_1}(y) = \begin{cases} \frac{4}{9}, & \text{if } y = 1 \\ \frac{5}{9}, & \text{if } y = 2 \\ 0, & \text{otherwise.} \end{cases}$$

By symmetry of $f_{\mathbf{X}}$, we get

$$f_{Y_2}(y) = \begin{cases} \frac{4}{9}, & \text{if } y = 1 \\ \frac{5}{9}, & \text{if } y = 2 \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 17. Find the joint pmf of $\underline{Y} = (Y_1, Y_2)$. Are Y_1 and Y_2 independent?

Change of variable for continuous distributions: The idea is quite similar to the univariate case, and will be discussed for a *special* class of functions.

Let $\mathbf{X} = (X_1, \ldots, X_m)$ be a random vector with density $f(t_1, \ldots, t_m)$. Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be a *one-one function which is continuously differentiable* (some exceptions can be made, as remarked later).

Let $\mathbf{Y} = T(\mathbf{X})$. In co-ordinates, we may write $\mathbf{Y} = (Y_1, \ldots, Y_m)$ and $Y_1 = T_1(X_1, \ldots, X_m)$, ..., $Y_m = T_m(X_1, \ldots, X_m)$, where $T_i : \mathbb{R}^m \to \mathbb{R}$ are the components of T.

Question: What is the joint density of $Y_1, \ldots, Y_m$?

The change of variable formula: In the setting described above, the joint density of $Y_1, \ldots, Y_m$ is given by

$$g(\mathbf{y}) = f\left(T^{-1}\mathbf{y}\right) \bigl| J[T^{-1}](\mathbf{y}) \bigr|,$$

where $|J[T^{-1}](\mathbf{y})|$ is the Jacobian determinant of the function T^{-1} at the point $\mathbf{y} = (y_1, \ldots, y_m)$.

Enlarging the applicability of the change of variable formula: The change of variable formula is applicable in greater generality than we stated above.

(1) Firstly, T does not have to be defined on all of $\mathbb{R}^m$. It is sufficient if it is defined on the range of $\mathbf{X}$ (i.e., if $f(t_1, \ldots, t_m) = 0$ for $(t_1, \ldots, t_m) = \mathbb{R}^m \setminus A$), then it is enough if T is defined on A.

(2) Similarly, the differentiability of T is required only on a subset, outside of which $\mathbf{X}$ has probability 0 of falling. For example, finitely many points, on a line (if $m \geq 2$), or on a plane (if $m \geq 3$), etc.

(3) One-one property of T is important, but there are special cases which can be dealt with by a slight modification. For example, if $T(x) = x^2$ or $T(x_1, x_2) = (x_1^2, x_2^2)$, where we can split the space into parts on each of which T is one-one.

Example 18. Let X_1, X_2 be independent Exp(λ) random variables. Let $T(x_1, x_2) = (x_1 + x_2, \frac{x_1}{x_1+x_2})$. This is well-defined on $\mathbb{R}_+^2$ (and note that $\mathbf{P}\{(X_1, X_2) \in \mathbb{R}_+^2\} = 1$) and its range is $\mathbb{R}_+ \times (0, 1)$. The inverse function is $T^{-1}(y_1, y_2) = (y_1 y_2, y_1(1 - y_2))$. Its Jacobian determinant is

$$|J[T^{-1}](y_1, y_2)| = \det \begin{bmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{bmatrix} = -y_1.$$

(X_1, X_2) has density $f(x_1, x_2) = \lambda^2 e^{-\lambda(x_1+x_2)}$ for $x_1, x_2 > 0$ (henceforth, if not mentioned explicitly, it will be a convention that the density is zero except where we specify it). Hence, the random variables $Y_1 = X_1 + X_2$ and $Y_2 = \frac{X_1}{X_1+X_2}$ have joint density

$$g(y_1, y_2) = f(y_1 y_2, y_1(1 - y_2))|J[T^{-1}](y_1, y_2)| = \lambda^2 e^{-\lambda(y_1 y_2 + y_1(1-y_2))} y_1 = \lambda^2 y_1 e^{-\lambda y_1}$$

for $y_1 > 0$ and $y_2 \in (0,1)$.

In particular, we see that $Y_1 = X_1 + X_2$ has density $h_1(t) = \int_0^1 \lambda^2 t e^{-\lambda t} ds = \lambda^2 t e^{-\lambda t}$ (for $t > 0$) which means that $Y_1 \sim \text{Gamma}(2, \lambda)$. Similarly, $Y_2 = \frac{X_1}{X_1+X_2}$ has density $h_2(s) = \int_0^\infty \lambda^2 t e^{-\lambda t} dt = 1$ (for $s \in (0,1)$) which means that Y_2 has Unif$(0,1)$ distribution. In fact, Y_1 and Y_2 are also independent since $g(u,v) = h_1(u)h_2(v)$.

Exercise 19. Let $X_1 \sim \text{Gamma}(\nu_1, \lambda)$ and $X_2 \sim \text{Gamma}(\nu_2, \lambda)$ (note that the shape parameter is the same) and assume that they are independent. Find the joint distribution of $X_1 + X_2$ and $\frac{X_1}{X_1+X_2}$.

Example 20. Suppose we are given that X_1 and X_2 are independent and each has Exp(λ) distribution. What is the distribution of the random variable $X_1 + X_2$?

The change of variable formula works for transformations from $\mathbb{R}^m$ to $\mathbb{R}^m$ whereas here we have two random variables X_1, X_2 and our interest is in one random variable $X_1 + X_2$. To use the change of variable formula, we must introduce an *auxiliary* variable. For example, we take $Y_1 = X_1 + X_2$ and $Y_2 = X_1/(X_1 + X_2)$. Then as in the first example, we find the joint density of (Y_1, Y_2) using the change of variable formula and then integrate out the second variable to get the density of Y_1.

Let us emphasize the point that if our interest is only in Y_1, then we have a lot of freedom in choosing the auxiliary variable. The only condition is that from Y_1 and Y_2 we should be able to recover X_1 and X_2. Let us repeat the same using $Y_1 = X_1 + X_2$ and $Y_2 = X_2$. Then, $T(x_1, x_2) = (x_1 + x_2, x_2)$ maps $\mathbb{R}_+^2$ onto $Q := \{(y_1, y_2) : y_1 > y_2 > 0\}$ in a one-one manner. The inverse function is $T^{-1}(y_1, y_2) = (y_1 - y_2, y_2)$. It is easy to see that $|J[T^{-1}](y_1, y_2)| = 1$ (check!). Hence, by the change of variable formula, the density of (Y_1, Y_2) is given by

$$\begin{aligned} g(y_1, y_2) &= f(y_1 - y_2, y_2) \cdot 1 \\ &= \lambda^2 e^{-\lambda(y_1 - y_2)} e^{-\lambda y_2} \quad \text{(if } y_1 > y_2 > 0\text{)} \\ &= \lambda^2 e^{-\lambda y_1} \mathbf{1}_{y_1 > y_2 > 0}. \end{aligned}$$

To get the density of Y_1, we integrate out the second variable. The density of Y_1 is

$$\begin{aligned} h(u) &= \int_{-\infty}^{\infty} \lambda^2 e^{-\lambda y_1} \mathbf{1}_{y_1 > y_2 > 0} dy_2 \\ &= \lambda^2 e^{-\lambda y_1} \int_0^{y_1} dy_2 \\ &= \lambda^2 y_1 e^{-\lambda y_1} \end{aligned}$$

which agrees with what we found before.

Example 21. Suppose $R \sim \text{Exp}(\lambda)$ and $\Theta \sim \text{Unif}(0, 2\pi)$ and the two are independent. Define $X = \sqrt{R}\cos(\Theta)$ and $Y = \sqrt{R}\sin(\Theta)$. We want to find the distribution of (X, Y). For this, we first write the joint density of (R, Θ) which is given by

$$f(r, \theta) = \frac{1}{2\pi}\lambda e^{-\lambda r} \quad \text{for } r > 0, \theta \in (0, 2\pi).$$

Define the transformation $T : \mathbb{R}_+ \times (0, 2\pi) \to \mathbb{R}^2$ by $T(r, \theta) = (\sqrt{r}\cos\theta, \sqrt{r}\sin\theta)$. The image of T consists of all $(x, y) \in \mathbb{R}^2$ with $y \neq 0$. The inverse is $T^{-1}(x, y) = (x^2 + y^2, \arctan(y/x))$, where $\arctan(y/x)$ is defined so as to take values in $(0, \pi)$ when $y > 0$ and to take values in $(\pi, 2\pi)$ when $y < 0$. Thus

$$|J[T^{-1}](x, y)| = \det \begin{bmatrix} 2x & 2y \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{bmatrix} = 2.$$

Therefore, (X, Y) has joint density

$$g(x, y) = 2f(x^2 + y^2, \arctan(y/x)) = \frac{\lambda}{\pi}e^{-\lambda(x^2+y^2)}.$$

This is for $(x, y) \in \mathbb{R}^2$ with $y \neq 0$. Since $g(x, y)$ separates into a function of x and a function of y, X, Y are independent $N(0, \frac{1}{2\lambda})$.

Remark 22. Relationships between random variables derived by the change of variable formulas can be used for simulation too. For instance, the CDF of $N(0, 1)$ is not explicit and hence simulating from that distribution is difficult (must resort to numerical methods). However, we can easily simulate it as follows. Simulate an $\text{Exp}(1/2)$ random variable R (easy, as the distribution function can be inverted) and simulate an independent $\text{Unif}(0, 2\pi)$ random variable Θ. Then, set $X = \sqrt{R}\cos(\Theta)$ and $Y = \sqrt{R}\sin(\Theta)$. These are two independent $N(0, 1)$ random numbers. Here, it should be noted that the random numbers in $(0, 1)$ given by a random number generator are supposed to be independent uniform random numbers.

4. Independence of random variables

Definition 23. Let $\mathbf{X} = (X_1, \ldots, X_m)$ be a random vector (this means that X_i are random variables on a common probability space). We say that X_is are *independent* if

$$F_{\mathbf{X}}(t_1, \ldots, t_m) = F_1(t_1) \cdots F_m(t_m) \text{ for all } t_1, \ldots, t_m.$$

Remark 24. Recalling the definition of independence of events, the equality $F_{\mathbf{X}}(t_1, \ldots, t_m) = F_1(t_1) \cdots F_m(t_m)$ is just saying that the events $\{X_1 \le t_1\}, \ldots, \{X_m \le t_m\}$ are independent. More generally, it is true that $X_1, \ldots, X_m$ are independent if and only if $\{X_1 \in A_1\}, \ldots, \{X_m \in A_m\}$ are independent events for any $A_1, \ldots, A_m \subseteq \mathbb{R}$.

Remark 25. In case $X_1, \ldots, X_m$ have a joint pmf or a joint pdf (which we denote by $f(t_1, \ldots, t_m)$), the condition for independence is equivalent to

$$f(t_1, \ldots, t_m) = f_1(t_1) \cdots f_m(t_m),$$

where f_i is the marginal density (or pmf) of X_i. This fact can be derived from the definition easily. For example, in the case of densities, observe that

$$\begin{aligned}
f(t_1, \ldots, t_m) &= \frac{\partial^m}{\partial t_1 \ldots \partial t_m} F(t_1, \ldots, t_m) \quad \text{(true for any joint density)} \\
&= \frac{\partial^m}{\partial t_1 \ldots \partial t_m} F_1(t_1) \cdots F_m(t_m) \quad \text{(by independence)} \\
&= F_1'(t_1) \cdots F_m'(t_m) \\
&= f_1(t_1) \cdots f_m(t_m).
\end{aligned}$$

When we turn it around, this gives us a quicker way to check independence.

Fact: Let $X_1, \ldots, X_m$ be random variables with joint pdf $f(t_1, \ldots, t_m)$. Suppose we can write this pdf as

$$f(t_1, \ldots, t_m) = c g_1(t_1) \cdots g_m(t_m),$$

where c is a constant and g_i are some functions of one-variable. Then, $X_1, \ldots, X_m$ are independent. Further, the marginal density of X_k is $c_k g_k(t)$, where $c_k = \frac{1}{\int_{-\infty}^{+\infty} g_k(s) ds}$. An analogous statement holds when $X_1, \ldots, X_m$ have a joint pmf instead of pdf.

Example 26. Let $\Omega = \{0,1\}^n$ with $p_{\underline{\omega}} = p^{\sum \omega_k} q^{n - \sum \omega_k}$. Define $X_k : \Omega \to \mathbb{R}$ by $X_k(\underline{\omega}) = \omega_k$. In words, we are considering the probability space corresponding to n tosses of a fair coin and X_k is the result of the kth toss. We claim that $X_1, \ldots, X_n$ are independent. Indeed, the joint pmf of $X_1, \ldots, X_n$ is

$$f(t_1, \ldots, t_n) = p^{\sum t_k} q^{n - \sum t_k}, \quad \text{where } t_i = 0 \text{ or } 1 \text{ for each } i \le n.$$

Clearly, $f(t_1, \ldots, t_n) = g(t_1)g(t_2)\cdots g(t_n)$, where $g(s) = p^s q^{1-s}$ for $s = 0$ or 1 (this is just a terse way of saying that $g(s) = p$ if $s = 1$ and $g(s) = q$ if $s = 0$). Hence, $X_1, \ldots, X_n$ are independent and X_k has pmf g (i.e., $X_k \sim \text{Ber}(p)$).

Example 27. Let (X, Y) have the bivariate normal density

$$f(x,y) = \frac{\sqrt{ab - c^2}}{2\pi} e^{-\frac{1}{2}(a(x-\mu_1)^2 + b(y-\mu_2)^2 + 2c(x-\mu_1)(y-\mu_2))}.$$

If $c = 0$, we observe that

$$f(x,y) = C_0 e^{-\frac{a(x-\mu_1)^2}{2}} e^{-\frac{b(y-\mu_2)^2}{2}} \qquad (C_0 \text{ is a constant, exact value unimportant})$$

from which we deduce that X and Y are independent, and $X \sim N(\mu_1, \frac{1}{a})$ while $Y \sim N(\mu_2, \frac{1}{b})$.

Exercise 28. Can you argue that if $c \neq 0$, then X and Y are not independent?

Example 29. Let $\mathbf{X} = (X_1, X_2, X_3)$ be a random vector of absolutely continuous type with pdf

$$f_{\mathbf{X}}(x_1, x_2, x_3) = \begin{cases} \frac{1}{x_1 x_2}, & \text{if } 0 < x_3 < x_2 < x_1 < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Are X_1, X_2 and X_3 independent random variables?

We have

$$f_{X_1}(x_1) = \begin{cases} \int_0^{x_1} \int_0^{x_2} \frac{1}{x_1 x_2}\, \mathrm{d}x_3\, \mathrm{d}x_2 = 1, & \text{if } 0 < x_1 < 1 \\ 0, & \text{other wise.} \end{cases}$$

$$f_{X_2}(x_2) = \begin{cases} \int_{x_2}^{1} \int_0^{x_2} \frac{1}{x_1 x_2}\, \mathrm{d}x_3\, \mathrm{d}x_1 = -\ln x_2, & \text{if } 0 < x_2 < 1 \\ 0, & \text{otherwise.} \end{cases}$$

$$f_{X_3}(x_3) = \begin{cases} \int_{x_3}^{1} \int_{x_2}^{1} \frac{1}{x_1 x_2}\, \mathrm{d}x_1\, \mathrm{d}x_2 = \frac{(\ln x_3)^2}{2}, & \text{if } 0 < x_3 < 1 \\ 0, & \text{otherwise.} \end{cases}$$

So,

$$f_{\mathbf{X}}(x_1, x_2, x_3) \neq f_{X_1}(x_1)\, f_{X_2}(x_2)\, f_{X_3}(x_3)\, \forall\, (x_1, x_2, x_3) \in \mathbb{R}^3,$$

and therefore X_1, X_2 and X_3 are not independent.

A very useful (and intuitively acceptable!) fact about independence is as follows.

Fact: Suppose $X_1, \ldots, X_n$ are independent random variables. Let $k_1 < k_2 < \cdots < k_m = n$. Let $Y_1 = h_1(X_1, \ldots, X_{k_1})$, $Y_2 = h_2(X_{k_1+1}, \ldots, X_{k_2}), \ldots, Y_m = h_m(X_{k_{m-1}}, \ldots, X_{k_m})$. Then, $Y_1, \ldots, Y_m$ are also independent.

Remark 30. In a previous section, we had defined independence of events and now we have defined independence of random variables. How are they related? We leave it to you to check that events $A_1, \ldots, A_n$ are independent (according to the definition in the previous section) if and only if the random variables $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_m}$ are independent (according to the definition in this section).