MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018
Lecture 12

DERIVED DISTRIBUTIONS

Contents

Given a random variable X with density f_X, and a measurable function g, we are often interested in the distribution (CDF, PDF, or PMF) of the random variable $Y = g(X)$. For the case of a discrete random variable X, this is straightforward:

$$p_Y(y) = \sum_{\{x \mid g(x)=y\}} p_X(x).$$

However, the case of continuous random variables is more complicated. Note that even if X is continuous, $g(X)$ is not necessarily a continuous random variable, e.g., if the range of the function g is discrete. However, in many cases, Y is continuous and its PDF can be found by following a systematic procedure.

1 FUNCTIONS OF A SINGLE RANDOM VARIABLE

The principal method for deriving the PDF of $g(X)$ is the following two-step approach.

Calculation of the PDF of a Function $Y = g(X)$ of a Continuous Random Variable X

(a) Calculate the CDF F_Y of Y using the formula

$$F_Y(y) = \mathbb{P}\big(g(X) \le y\big) = \int_{\{x \,|\, g(x) \le y\}} f_X(x)\, dx.$$

(b) Differentiate to obtain the PDF of Y:

$$f_Y(y) = \frac{dF_Y}{dy}(y).$$

Example. Let $Y = g(X) = X^2$, where X is a continuous random variable with known PDF. For any $y > 0$, we have

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

and therefore, by differentiating and using the chain rule,

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}), \qquad y > 0.$$

Example. Suppose that X is a nonnegative random variable and that $Y = \exp(X^2)$. Note that $F_Y(y) = 0$ for $y < 1$. For $y \ge 1$, we have

$$F_Y(y) = \mathbb{P}\big(e^{X^2} \le y\big) = \mathbb{P}(X^2 \le \log y) = \mathbb{P}(X \le \sqrt{\log y}).$$

By differentiating and using the chain rule, we obtain

$$f_Y(y) = f_X(\sqrt{\log y}) \frac{1}{2y\sqrt{\log y}}, \qquad y > 1.$$

1.1 The case of monotonic functions

The calculation in the last example can be generalized as follows. Assume that the range of the random variable X contains an open interval A. Suppose that g is strictly monotone (say increasing), and also differentiable on the interval

A. Let B the set of values of $g(x)$, as x ranges over A. Let g^{-1} be the inverse function of g, so that $g(g^{-1}(y)) = y$, for $y \in B$. Then, for $y \in B$, and using the chain rule in the last step, we have

$$f_Y(y) = \frac{d}{dy}\mathbb{P}(g(X) \le y) = \frac{d}{dy}\mathbb{P}(X \le g^{-1}(y)) = f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy}.$$

Recall from calculus that the derivative of an inverse function satisfies

$$\frac{dg^{-1}}{dy}(y) = \frac{1}{g'(g^{-1}(y))},$$

where g' is the derivative of g. Therefore,

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

When g is strictly monotone decreasing, the only change is a minus sign in front of $g'(g^{-1}(y))$. Thus, the two cases can be summarized in the single formula:

$$f_Y(y) = f_X(g^{-1}(y))\frac{1}{|g'(g^{-1}(y))|}, \qquad y \in B. \tag{1}$$

An easy mnemonic for remembering (and also understanding this formula) is

$$f_Y(y)\,|dy| = f_X(x)\,|dx|,$$

where x and y are related by $y = g(x)$, and since $dy = |g'(x)| \cdot |dx|$,

$$f_Y(y)\,|g'(x)| = f_X(x).$$

1.2 Linear functions

Consider now the special case where $g(x) = ax + b$, i.e., $Y = aX + b$. We assume that $a \neq 0$. Then, $g'(x) = a$ and $g^{-1}(y) = (y - b)/a$. We obtain

$$f_Y(y) = \frac{1}{|a|}f_X\Big(\frac{y-b}{a}\Big).$$

Example. (A linear function of a normal random variable is normal) Suppose that $X \stackrel{d}{=} N(0,1)$ and $Y = aX + b$. Then,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a|}e^{-\frac{(y-b)^2}{2a^2}},$$

so that Y is $N(b, a^2)$. More generally, if $X \stackrel{d}{=} N(\mu, \sigma)$, then the same argument shows that $Y = ax + b \stackrel{d}{=} N(a\mu + b, a^2\sigma^2)$. We conclude that a linear (more precisely, affine) function of a normal random variable is normal.

1.3 The general case

Definition 1. *Consider two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ and a measurable function $g : \Omega_1 \to \Omega_2$. The **pushforward** $g_*\mu$ of measure μ on $(\Omega_1, \mathcal{F}_1)$ along g is a measure on $(\Omega_2, \mathcal{F}_2)$ defined as*

$$g_*\mu(E) \triangleq \mu(g^{-1}(E)) \qquad \forall E \in \mathcal{F}_2 \, .$$

To relate this general notion to the discussion above, we notice that

$$Y = g(X) \Rightarrow \mathbb{P}_Y = g_*\mathbb{P}_X$$

To emphasize the fact that g "carries" μ into $\rho = g_*\mu$ we may schematically denote

$$d\mu \xrightarrow{g} d\rho \tag{2}$$

We establish some simple facts of pushforwards:

Proposition 1. *Let $\rho = g_*\mu$. Then*

(i) (Change of variable formula) For any measurable $f : \Omega_2 \to \mathbb{R}$ we have

$$\int_{\Omega_2} f(\omega_2) d\rho = \int_{\Omega_1} f(g(\omega_1)) d\mu \tag{3}$$

and both integrals exist or do not exist simultaneously.

(ii) For arbitrary non-negative $\phi(\omega_2)$ measure $\phi(g(\omega_1))d\mu$ pushes forward to $\phi(\omega_2)d\rho$. Schematically:

$$\phi \circ g \, d\mu \xrightarrow{g} \phi \, d\rho \tag{4}$$

Proof: (i) is easy: for $f = 1_E$ this follows from the definition of $g_*\mu$. Linearity of integration implies the statement for simple functions. Finally, splitting $f = f^+ - f^-$ and approximating both f^+ and f^- by simple functions we conclude by invoking the MCT.

(ii) follows from the application of (3) with $f = 1_E \cdot \phi$:

$$\int_E \phi(\omega_2) d\rho = \int_{g^{-1}E} \phi \circ g(\omega_1) d\mu$$

□

Note: Consider a special case with $\Omega_1 = \Omega_2 = \mathbb{R}$ and $d\mu = dx$ – Lebesgue measure. Then whenever g – continuously differentiable with non-vanishing derivative on an open set U we have[1]

$$1_U(x)|g'(x)|\, dx \xrightarrow{g} 1_{g(U)}(y)\, dy \tag{5}$$

This result will be established in multiple dimensions below. For now we apply (4) with $\phi(x) = \frac{f_X \circ g^{-1}}{|g' \circ g^{-1}|}$ to get:

$$f_X(x)\, dx \xrightarrow{g} \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y)|}\, dy$$

which exactly means the formula for f_Y given in (1).

2 MULTIVARIATE TRANSFORMATIONS

Suppose now that $X = (X_1, \dots, X_n)$ is a vector of random variables that are jointly continuous, with joint PDF $f_X(x) = f_X(x_1, \dots, x_n)$. Consider a function $g : \mathbb{R}^n \to \mathbb{R}^n$, and the random vector $Y = (Y_1, \dots, Y_n) = g(X)$. Let g_i be the components of g, so that $Y_i = g_i(X) = g_i(X_1, \dots, X_n)$. Suppose that the function g is continuously differentiable on some open set $A \subset \mathbb{R}^n$. Let $B = g(A)$ be the image of A under g. Assume that the inverse function g^{-1} is well-defined on B; that is, for every $y \in B$, there is a unique $x \in \mathbb{R}^n$ such that $g(x) = y$.

The formula that we develop here is an extension of the formula $f_Y(y)\,|g'(x)| = f_X(x)$ that we derived for the one-dimensional case, with the derivative $g'(x)$ being replaced by a matrix of partial derivatives, and with the absolute value being replaced by the absolute value of the determinant. It can be justified by appealing to the change of variables theorem from multivariate calculus, but we provide here a more transparent argument.

[1] Here and below we abuse notation and write dx instead of $\mathrm{Leb}(dx)$.

2.1 Linear functions

Let us first assume that g is a linear function, of the form $g(x) = Mx$, for some $n \times n$ matrix M. Fix some $x \in A$ and some $\delta > 0$. Consider the cube $C = [x, x + \delta]^n$, and assume that δ is small enough so that $C \subset A$. The image $D = \{Mx \mid x \in C\}$ of the cube C under the mapping g is a parallelepiped. Furthermore, the volume of D is known to be equal to $|M| \cdot \delta^n$, where we use $|\cdot|$ to denote the absolute value of the determinant of a matrix.

Having fixed x, let us also fix $y = Mx$. Assuming that $f_X(x)$ is continuous at x, we have

$$\mathbb{P}(X \in C) = \int_C f_X(t)\,dt = f_X(x)\delta^n + o(\delta^n) \approx f_X(x)\delta^n,$$

where $o(\delta^n)$ stands for a function such that $\lim_{\delta \downarrow 0} o(\delta^n)/\delta^n = 0$, and where the symbol $\approx$ indicates that the difference between the two sides is $o(\delta^n)$. Thus,

$$\begin{aligned}
f_X(x) \cdot \delta^n &\approx \mathbb{P}(X \in C) \\
&= \mathbb{P}(g(X) \in g(C)) \\
&= \mathbb{P}(Y \in D) \\
&\approx f_Y(y) \cdot \text{vol}(D) \\
&= f_Y(y) \cdot |M| \cdot \delta^n.
\end{aligned}$$

Dividing by δ^n, and taking the limit as $\delta \downarrow 0$, we obtain $f_X(x) = f_Y(y) \cdot |M|$. Let us now assume that the matrix M is invertible, so that its determinant is nonzero. Using the relation $y = Mx$ and the fact $\det(M^{-1}) = 1/\det(M)$,

$$f_Y(y) = \frac{f_X(M^{-1}y)}{|M|} = f_X(M^{-1}y) \cdot |M^{-1}|.$$

Note that if M is not invertible, the random variable Y takes values in a proper subspace S of $\mathbb{R}^n$. Then, Y is not jointly continuous (cannot be described by a joint PDF). On the other hand, if we restrict our attention to S, and since S is isomorphic to $\mathbb{R}^m$ for some $m < n$, we could describe the distribution of Y in terms of a joint PDF on $\mathbb{R}^m$.

2.2 The general case: heuristic

Let us now generalize to the case where g is continuously differentiable at x. We define $M(x)$ as the Jacobian matrix $(\partial g/\partial x)(x)$, with entries $(\partial g_i/\partial x_j)(x)$.

The image $D = g(C)$ of the cube C is not a parallelepiped. However, from a first order Taylor series expansion, g is approximately linear in the vicinity of x. It can then be shown that the set D has volume $|M(x)| \cdot \delta^n + o(\delta^n)$. It then follows, as in the linear case, that

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|M(g^{-1}(y))|} = f_X(g^{-1}(y)) \cdot |M^{-1}(g^{-1}(y))|.$$

We note a useful fact from calculus that sometimes simplifies the application of the above formula. If we define $J(y)$ as the Jacobian (the matrix of partial derivatives) of the mapping $g^{-1}(y)$, and if some particular x and y are related by $y = g(x)$, then $J(y) = M^{-1}(x)$. Therefore,

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |J(y)|. \tag{6}$$

2.3 The general case: formal results

Our goal is to prove (6). First we state assumptions.

Assumption A. *Let U be an open set in $\mathbb{R}^n$ and $g : U \to \mathbb{R}^n$ be continuously differentiable injection with non-vanishin Jacobian* $\frac{\partial g}{\partial \mathbf{x}} \neq 0$.

It is well known that if $g : U \to g(U)$ satisfies Assumption A, then $g(U)$ is open and the inverse map g^{-1} satisfies Assumption A on $g(U)$.

We first remind the following important result from calculus: *For any continuous $f : g(U) \to \mathbb{R}$ with compact support we have*

$$\mathcal{R}\int_{g(U)} f(\mathbf{y})\, d\mathbf{y} = \mathcal{R}\int_U f(g(\mathbf{x})) \left| \frac{\partial g}{\partial \mathbf{x}} \right| d\mathbf{x} \tag{7}$$

where $\mathcal{R}\int$ denotes the Riemann integral.

We now use this fact to establish the following:

Theorem 1 (Jacobian formula)**.** *If U and $g : U \mapsto g(U)$ satisfy Assumption A then*

$$1_U(\mathbf{x}) \left| \frac{\partial g}{\partial \mathbf{x}} \right| d\mathbf{x} \xrightarrow{g} 1_{g(U)}(\mathbf{y})\, d\mathbf{y}$$

Applying Theorem 1 with (3) and (4) we can establish a number of consequences (for simplicity we assume $U = \mathbb{R}^n$):

$$\left|\frac{\partial g}{\partial \mathbf{x}}\right| d\mathbf{x} \xrightarrow{g} d\mathbf{y} \tag{8}$$

$$d\mathbf{x} \xrightarrow{g} \left|\frac{\partial g^{-1}}{\partial \mathbf{y}}\right| d\mathbf{y} \tag{9}$$

$$f_X(\mathbf{x})\, d\mathbf{x} \xrightarrow{g} f_X(g^{-1}(\mathbf{y})) \left|\frac{\partial g^{-1}}{\partial \mathbf{y}}\right| d\mathbf{y} \tag{10}$$

where we also used the fact that $\left|\frac{\partial g^{-1}}{\partial \mathbf{y}}\right| = \left|\frac{\partial g}{\partial \mathbf{x}}\right|^{-1}$ for $\mathbf{y} = g(\mathbf{x})$. In particular (10) implies (6). When g satisfies Assumption A on disjoint U_1 and U_2 then we have

$$1_{U_1 \cup U_2}(\mathbf{x}) \left|\frac{\partial g}{\partial \mathbf{x}}\right| d\mathbf{x} \xrightarrow{g} \left\{1_{g(U_1)}(\mathbf{y}) + 1_{g(U_2)}(\mathbf{y})\right\} d\mathbf{y}$$

which may be useful to find pushforwards along many-to-one maps.

Proof (optional). Let μ be a measure defined via

$$\mu(E) = \int_E 1_U(\mathbf{x}) \left|\frac{\partial g}{\partial \mathbf{x}}\right| d\mathbf{x}$$

and $\rho = g_*\mu$. We need to show that

$$\rho(E) = \mathrm{Leb}(E \cap g(U))$$

for every Borel set E. Or, equivalently, that $\rho(E) = \mathrm{Leb}(E)$ for every Borel $E \subset g(U)$. Under conditions of the theorem it is easy to see that $g(U)$ is an open set (since g is continuously invertible and hence every point $y \in g(U)$ has an open ball around itself contained in $g(U)$).

First, we have indicated before (and proved in dimension 1) that Riemann integral, when it exists, coincides with Lebesgue integral over Lebesgue measure. Thus from (7) we conclude that for every continuous function f with compact support contained inside $g(U)$ we have

$$\int_{\mathbb{R}^n} f(\mathbf{y})\, d\mathbf{y} = \int_{\mathbb{R}^n} f(g(\mathbf{y})) d\mu$$

where this time integrals are Lebesgue, and as before $d\mathbf{y}$ stands for $\mathrm{Leb}(d\mathbf{y})$. By (3) this implies that for all such functions f we have

$$\int_{\mathbb{R}^n} f(\mathbf{y})\, d\mathbf{y} = \int_{\mathbb{R}^n} f(\mathbf{y})\, d\rho$$

Now consider an arbitrary open ball B contained in $g(U)$. Although the indicator $1_B(\mathbf{y})$ is not a continuous function, it can be easily approximated by a sequence continuous functions from below:

$$0 \le f_n(\mathbf{y}) \nearrow 1_B(\mathbf{y})$$

Thus by the MCT we conclude that

$$\rho(B) = \mathrm{Leb}(B)$$

for every open ball contained in $g(U)$. Now every open set $V \subset g(U)$ can be written as a union of countably many open balls. Thus by continuity of measure $\rho(V) = \mathrm{Leb}(V)$. But two σ-finite measures coinciding on every open set must be identical (since open sets are a generating p-system). □

2.4 The bivariate normal in polar coordinates

Let X and Y be independent standard normal random variables. Let g be the mapping that transforms Cartesian coordinates to polar coordinates, and let $(R, \Theta) = g(X, Y)$. The mapping g is either undefined or discontinuous at $(x, y) = (0, 0)$. So, strictly speaking, in order to apply the multivariate transformation formula (6), we should work with $A = \mathbb{R}^2 \setminus \{(0, 0)\}$. The inverse mapping g^{-1} is given by $(x, y) = g^{-1}(r, \theta) = (r\cos\theta, r\sin\theta)$. Its Jacobian matrix is of the form

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{bmatrix},$$

and therefore, $|J(r, \theta)| = r\cos^2\theta + r\sin^2\theta = r$. From the multivariate transformation formula, we obtain

$$f_{R,\Theta}(r, \theta) = \frac{r}{2\pi} e^{-(r^2\cos^2\theta + r^2\sin^2\theta)/2} = \frac{1}{2\pi} r e^{-r^2/2}, \qquad r > 0.$$

We observe that the joint PDF of R and Θ is of the form $f_{R,\Theta} = f_R f_\Theta$, where

$$f_R(r) = r e^{-r^2/2},$$

and

$$f_\Theta(\theta) = \frac{1}{2\pi}, \qquad \theta = [0, 2\pi].$$

In particular, R and Θ are independent. The random variable R is said to have a **Rayleigh** distribution.

We can also find the density of $Z = R^2$. For the mapping g defined by $g(r) = r^2$, we have $g^{-1}(z) = \sqrt{z}$, and $g'(r) = 2r$, which leads to $1/g'(g^{-1}(z)) = 1/2\sqrt{z}$. We conclude that

$$f_Z(z) = \sqrt{z}e^{-z/2}\frac{1}{2\sqrt{z}} = \frac{1}{2}e^{-z/2}, \qquad z > 0$$

which we recognize as an exponential PDF with parameter $1/2$.

An interesting consequence of the above results is that in order to simulate normal random variables, it suffices to generate two independent random variables, one uniform and one exponential. Furthermore, an exponential random variable is easy to generate using a nonlinear transformation of another independent uniform random variable.

3 A SINGLE FUNCTION OF MULTIPLE RANDOM VARIABLES

Suppose that $X = (X_1, \ldots, X_n)$ is a vector of jointly continuous random variables with known PDF f_X. Consider a function $g_1 : \mathbb{R}^n \to \mathbb{R}$ and the random variable $Y_1 = g_1(X)$. Note that the formulas from Section 2 cannot be used directly. In order to find the PDF of Y, one possibility is to calculate the multi-dimensional integral

$$F_Y(y) = \mathbb{P}(g(X) \le y) = \int_{\{x \mid g(x) \le y\}} f_X(x)\, dx,$$

and then differentiate.

Another possibility is to introduce additional functions $g_2, \ldots, g_n : \mathbb{R}^n \to R$, and define $Y_i = g_i(X)$, for $i \ge 2$. As long as the resulting function $g = (g_1, \ldots, g_n)$ is invertible, we can appeal to our earlier formula to find the joint PDF of Y, and then integrate to find the marginal PDF of Y_1.

The simplest choice in the above described method is to let $Y_i = X_i$, $i \neq 1$, so that $g(x) = (g_1(x), x_2, \ldots, x_n)$. Let $h : \mathbb{R}^n \to \mathbb{R}$ be a function that corresponds to the first component of g^{-1}. That is, if $y = g(x)$, then $x_1 = h(y)$. Then, the inverse mapping g^{-1} is of the form

$$g^{-1}(y) = (h(y), y_2, \ldots, y_n),$$

and its Jacobian matrix is of the form

$$J(y) = \begin{bmatrix} \frac{\partial h}{\partial y_1}(y) & \frac{\partial h}{\partial y_2}(y) & \cdots & \frac{\partial h}{\partial y_n}(y) \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

It follows that $|J(y)| = |\frac{\partial h}{\partial y_1}(y)|$, and

$$f_Y(y) = f_X(h(y), y_2, \ldots, y_n)\left|\frac{\partial h}{\partial y_1}(y)\right|.$$

Integrating, we obtain

$$f_{Y_1}(y_1) = \int f_X(h(y), y_2, \ldots, y_n)\left|\frac{\partial h}{\partial y_1}(y)\right| dy_2 \cdots dy_n.$$

Example. Let X_1 and X_2 be positive, jointly continuous, random variables, and suppose that we wish to derive the PDF of $Y_1 = g(X_1, X_2) = X_1X_2$. We define $Y_2 = X_2$. From the relation $x_1 = y_1/x_2$ we see that $h(y_1, y_2) = y_1/y_2$. The partial derivative $\partial h/\partial y_1$ is $1/y_2$. We obtain

$$f_{Y_1}(y_1) = \int f_X(y_1/y_2, y_2)\frac{1}{y_2}\, dy_2 = \int f_X(y_1/x_2, x_2)\frac{1}{x_2}\, dx_2.$$

For a special case, suppose that $X_1, X_2 \overset{d}{=} U(0,1)$ are independent. Their common PDF is $f_{X_i}(x_i) = 1$, for $x_i \in [0,1]$. Note that $f_{Y_1}(y_1) = 0$ for $y \notin (0,1)$. Furthermore, $f_{X_1}(y_1/x_2)$ is positive (and equal to 1) only in the range $x_2 \geq y_1$. Also $f_{X_2}(x_2)$ is positive, and equal to 1, iff $x_2 \in (0,1)$. In particular,

$$f_X(y_1/x_2, x_2) = f_{X_1}(y_1/x_2)f_{X_2}(x_2) = 1, \qquad \text{for } x_2 \geq y_1.$$

We then obtain

$$f_{Y_1}(y_1) = \int_{y_1}^{1} \frac{1}{x_2}\, dx_2 = -\log y, \qquad y_1 \in (0,1).$$

The direct approach to this problem would first involve the calculation of $F_{Y_1}(y_1) = \mathbb{P}(X_1X_2 \leq y_1)$. It is actually easier to calculate

$$\begin{aligned} 1 - F_{Y_1}(y_1) = \mathbb{P}(X_1X_2 \geq y_1) &= \int_{y_1}^{1}\int_{y_1/x_1}^{1} dx_2\, dx_1 \\ &= \int_{y_1}\left(1 - \frac{y_1}{x_1}\right) dx_1 \\ &= (x_1 - y_1 \log x_1)\Big|_{y_1}^{1} = (1 - y_1) + y_1 \log y_1. \end{aligned}$$

Thus, $F_{Y_1}(y_1) = y_1 - y_1 \log y_1$. Differentiating, we find that $f_{Y_1}(y_1) = -\log y_1$.

An even easier solution for this particular problem (along the lines of the stick example in Lecture 9) is to realize that conditioned on $X_1 = x_1$, the random variable $Y_1 = X_1X_2$ is uniform on $[0, x_1]$, and using the total probability theorem,

$$f_{Y_1}(y_1) = \int_{y_1}^{1} f_{X_1}(x_1)f_{Y_1|X_1}(y_1 \mid x_1)\, dx_1 = \int_{y_1}^{1} \frac{1}{x_1}\, dx_1 = -\log y_1.$$

4 MAXIMUM AND MINIMUM OF RANDOM VARIABLES

Let $X_1, \dots, X_n$ be independent random variables, and let $X^{(1)} \leq X^{(2)} \leq \dots \leq X^{(1)}$ denote the corresponding *order statistics*. Namely, $X^{(1)} = \min_j X_j$, $X^{(2)}$ is the second smallest of the values $X_1, \dots, X_n$, and $X^{(n)} = \max_j X_j$. We would like to find the joint distribution of the order statistics and specifically the distribution of $\min_j X_j$ and $\max_j X_j$. Note that

$$\begin{aligned}\mathbb{P}(\max_j X_j \leq x) &= \mathbb{P}(X_1, \dots, X_n \leq x) = \mathbb{P}(X_1 \leq x) \cdots \mathbb{P}(X_n \leq x) \\ &= F_{X_1}(x) \cdots F_{X_n}(x).\end{aligned}$$

For the minimum, we have

$$\begin{aligned}\mathbb{P}(\min_j X_j \leq x) &= 1 - \mathbb{P}(\min_j X_j > x) \\ &= 1 - \mathbb{P}(X_1, \dots, X_n > x) \\ &= 1 - (1 - F_{X_1}(x)) \cdots (1 - F_{X_n}(x)).\end{aligned}$$

Let us consider the special case where the $X_1, \dots, X_n$ are i.i.d., with common CDF F and PDF f. For simplicity, assume that F is differentiable everywhere. Then,

$$\mathbb{P}(\max_j X_j \leq x) = F^n(x), \qquad \mathbb{P}(\min_j X_j \leq x) = 1 - (1 - F(x))^n,$$

implying that

$$f_{\max_j X_j}(x) = nF^{n-1}(x)f(x), \qquad f_{\min_j X_j}(x) = n(1 - F(x))^{n-1}f(x).$$

Exercise 1. Assuming that $X_1, \dots, X_n$ are independent with common density function f, establish that the joint distribution of $X^{(1)}, \dots, X^{(n)}$ is given by

$$f_{X^{(1)}, \dots, X^{(n)}}(x_1, \dots, x_n) = n! f(x_1) \cdots f(x_n), \qquad x_1 < x_2 < \dots < x_n,$$

and $f_{X^{(1)}, \dots, X^{(n)}}(x_1, \dots, x_n) = 0$, otherwise. Use this to derive the densities for $\max_j X_j$ and $\min_j X_j$.

5 SUM OF INDEPENDENT RANDOM VARIABLES – CONVOLUTION

If X and Y are independent discrete random variables, the PMF of $X+Y$ is easy to find:

$$\begin{aligned} p_{X+Y}(z) &= \mathbb{P}(X+Y=z) \\ &= \sum_{\{(x,y)|x+y=z\}} \mathbb{P}(X=x, Y=y) \\ &= \sum_x \mathbb{P}(X=x, Y=z-x) \\ &= \sum_x p_X(x) p_Y(z-x). \end{aligned}$$

When X and Y are independent and jointly continuous, an analogous formula can be expected to hold. We derive it in two different ways.

A first derivation involves plain calculus. Let $f_{X,Y}$ be the joint PDF of X and Y. Then.

$$\mathbb{P}(X+Y \le z) = \int_{\{x,y \mid x+y \le z\}} f_{X,Y}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x,y)\, dy\, dx.$$

Introduce the change of variables $t = x+y$. Then,

$$\mathbb{P}(X+Y \le z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_{X,Y}(x, t-x)\, dt\, dx,$$

which gives

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\, dx.$$

In the special case where X and Y are independent, we have $f_{X,Y}(x,y) = f_X(x) f_Y(y)$, resulting in

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx.$$

If we were to use instead our general tools, we could proceed as follows. Consider the linear function g that maps (X,Y) to $(X, X+Y)$. It is easily seen that the associated determinant is equal to 1. Thus, with $Z = X+Y$, we have

$$f_{X,Z}(x,z) = f_{X,Y}(x, z-x) = f_X(x) f_Y(z-x).$$

We then integrate over all x, to obtain the marginal PDF of Z.

Exercise 2. Suppose that $X \stackrel{d}{=} N(\mu_1, \sigma_1^2), X_2 \stackrel{d}{=} N(\mu_2, \sigma_2^2)$ and independent. Establish that $X_1 + X_2 \stackrel{d}{=} N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Exercise 3. Establish the semigroup properties for Gamma and Cauchy distributions mentioned in Lecture 10.