1 Change of variables

After applying a function $g(\cdot)$ to a random variable X or random vector $\mathbf{X}$, the goal is to find the distribution of the transformed random variable $g(X)$, or the joint distribution of the random vector $g(\mathbf{X})$.

In the discrete case, we get the PMF of the transformed random variable $g(X)$ by translating the event $g(X) = y$ into an equivalent event involving X:

$$\mathsf{P}(g(X) = y) = \sum_{\{x:g(x)=y\}} \mathsf{P}(X = x).$$

In the continuous case, we start from the CDF of $g(X)$, and translate the event $g(X) \leq y$ into an equivalent event involving X.

Theorem (Change of variables in one dimension)**.** *Let X be a continuous random variable with PDF f_X, and let $Y = g(X)$, where g is differentiable, and strictly increasing or strictly decreasing. Then the PDF of Y is given by*

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|,$$

where $x = g^{-1}(y)$. The support of Y is all $g(x)$ with x in the support of X.

Proof. Let g be strictly increasing. The CDF of Y is

$$F_Y(y) = \mathsf{P}(Y \leq y) = \mathsf{P}(g(X) \leq y) = \mathsf{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(x).$$

The PDF of Y is obtain by differentiating both side using the chain rule. □

When applying the change of variables formula, we can choose whether to compute $\frac{dx}{dy}$, or $\frac{dy}{dx}$ and then take the reciprocal. Either way, in the end we should express the PDF of Y as a function of y.

Example: Log-Normal PDF
Let $X \sim \mathsf{N}(0, 1)$, and $Y = e^X$. Then $Y = g(X)$ with $g(x) = e^x$, a strictly increasing function. Let $y = e^x$, so $x = \log y$ and $\frac{dy}{dx} = e^x$. Then

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(x) \frac{1}{\left| \frac{dy}{dx} \right|} = \varphi(x) \frac{1}{e^x} = \varphi(\log y) \frac{1}{y}, \quad y > 0.$$

Example: Chi-Square PDF
Let $X \sim \mathsf{N}(0, 1)$, and $Y = X^2$. Then $Y = g(X)$ with $g(x) = x^2$. This function is not one-to-one, therefore we cannot apply the change of variables theorem directly. Instead, we start from the CDF:

$$F_Y(y) = \mathsf{P}(X^2 \leq y) = \mathsf{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1.$$

By differentiating on both sides we obtain

$$f_Y(y) = 2\varphi(\sqrt{y}) \cdot \frac{1}{2} y^{-1/2} = \varphi(\sqrt{y}) y^{-1/2}, \quad y > 0.$$

Theorem (Change of variables)**.** *Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *be a continuous random vector with joint PDF* $f_{\mathbf{X}}(\mathbf{x})$*, and let* $\mathbf{Y} = g(\mathbf{X})$ *where* g *is an invertible function from* $\mathbb{R}^n$ *to* $\mathbb{R}^n$*. Let* $\mathbf{y} = g(\mathbf{x})$ *which implies* $\mathbf{x} = g^{-1}(\mathbf{y})$*. We consider the Jacobian matrix which is the matrix of all the partial derivatives* $\frac{\partial x_i}{\partial y_j}$ *that are assumed to exist and be continuous:*

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \frac{\partial g^{-1}(\mathbf{y})}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}.$$

Then the joint PDF of $\mathbf{Y}$ *is*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|,$$

where the vertical bars mean: take the absolute value of the determinant of the Jacobian matrix $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$*. The following relation holds:*

$$\left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| = \left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right|^{-1},$$

so we can compute whichever of the two is easier, and then at the end express the joint PDF of $\mathbf{Y}$ *as a function of* $\mathbf{y}$*.*

The justification of this result comes from the fact that if A is a region in the support of $\mathbf{X}$ and $B = \{g(\mathbf{x}) : \mathbf{x} \in A\}$ is the corresponding region in the support of $\mathbf{Y}$, then the events $\{\mathbf{X} \in A\}$ and $\{\mathbf{Y} \in B\}$ are the same. Thus $\mathsf{P}(\mathbf{X} \in A) = \mathsf{P}(\mathbf{Y} \in B)$ which implies

$$\int_A f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = \int_B f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}.$$

Example: Bivariate Normal joint PDF

We find the PDF of the vector (Z, W) which follows a Bivariate Normal distribution with $\mathsf{N}(0,1)$ marginals and $\mathsf{Corr}(Z, W) = \rho \in (-1, 1)$. We construct (Z, W) by transforming another vector (X, Y) with $X, Y \sim \mathsf{N}(0,1)$, and X independent of Y as follows:

$$\begin{aligned} Z &= X \\ W &= \rho X + \tau Y, \end{aligned}$$

with $\tau = \sqrt{1-\rho^2}$. The inverse transformation that maps (Z, W) into (X, Y) is

$$\begin{aligned} X &= Z \\ Y &= -\frac{\rho}{\tau} Z + \frac{1}{\tau} W. \end{aligned}$$

The Jacobian matrix is

$$\frac{\partial(x, y)}{\partial(z, w)} = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\tau} & \frac{1}{\tau} \end{pmatrix}.$$

The absolute value of the determinant of the Jacobian matrix is $\frac{1}{\tau}$. The change of variable formula becomes:

$$\begin{aligned} f_{Z,W}(z,w) &= f_{X,Y}(x,y) \cdot \left| \frac{\partial(x,y)}{\partial(z,w)} \right|, \\ &= \frac{1}{2\pi\tau} \exp\left(-\frac{1}{2}(x^2+y^2) \right), \\ &= \frac{1}{2\pi\tau} \exp\left(-\frac{1}{2\tau^2}(z^2+w^2-2\rho zw) \right). \end{aligned}$$

2 Convolutions

A *convolution* is a sum of independent random variables.

Theorem (Convolution sums and integrals)**.** *If X and Y are independent discrete random variables, then the PMF of their sum $T = X + Y$ is*

$$\mathsf{P}(T=t) = \sum_x \mathsf{P}(Y=t-x)\mathsf{P}(X=x) = \sum_y \mathsf{P}(X=t-y)\mathsf{P}(Y=y).$$

If X and Y are independent continuous random variables, then the PDF of their sum $T = X + Y$ is

$$f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x) f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(t-y) f_Y(y)\,dy.$$

Proof. For the discrete case, we use LOTP, conditioning on x:

$$\begin{aligned} \mathsf{P}(T=t) &= \sum_x \mathsf{P}(X+Y=t \mid X=x)\mathsf{P}(X=x), \\ &= \sum_x \mathsf{P}(Y=t-x \mid X=x)\mathsf{P}(X=x), \\ &= \sum_x \mathsf{P}(Y=t-x)\mathsf{P}(X=x). \end{aligned}$$

For the continuous case, we first calculate the CDF of the convolution T using LOTP:

$$\begin{aligned} F_T(t) = \mathsf{P}(X+Y \le t) &= \int_{-\infty}^{\infty} \mathsf{P}(X+Y \le t \mid X=x) f_X(x)\,dx, \\ &= \int_{-\infty}^{\infty} \mathsf{P}(Y \le t-x) f_X(x)\,dx, \\ &= \int_{-\infty}^{\infty} F_Y(t-x) f_X(x)\,dx. \end{aligned}$$

To obtain the CDF we differentiate with respect to t, interchanging the order of differentiation and integration. We obtain

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

Another approach to obtain the PDF of T in the continuous case is to use the change of variable formula. We consider the transformation from (X, Y) to (T, W), where $T = X + Y$ and $W = X$. We let $t = x + y$ and $w = x$. Then

$$\frac{\partial(t, w)}{\partial(x, y)} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

We calculate the absolute value of the determinant of this Jacobian matrix:

$$\left| \frac{\partial(t, w)}{\partial(x, y)} \right| = 1.$$

Thus

$$\left| \frac{\partial(x, y)}{\partial(t, w)} \right| = \frac{1}{\left| \frac{\partial(t,w)}{\partial(x,y)} \right|} = 1.$$

Thus the joint PDF of T and W is

$$f_{T,W}(t, w) = f_{X,Y}(x, y) = f_X(x) f_Y(y) = f_X(w) f_Y(t - w),$$

and the marginal PDF of T is

$$f_T(t) = \int_{-\infty}^{\infty} f_{T,W}(t, w)\, dw = \int_{-\infty}^{\infty} f_X(x) f_Y(t - x)\, dx.$$

$\square$

Example: Uniform convolution
Let X, Y be i.i.d. $\mathsf{Unif}(0, 1)$. Find the distribution of $T = X + Y$.

Solution: The PDF of X and of Y is $g(x) = 1$ if $x \in (0, 1)$ and $g(x) = 0$, otherwise. The convolution formula gives:

$$f_T(t) = \int_{-\infty}^{\infty} f_X(x) f_Y(t - x)\, dx = \int_{-\infty}^{\infty} g(x) g(t - x)\, dx.$$

The integrand $g(x)g(t - x)$ is 1 if $0 < t - x < 1$ and $0 < x < 1$, and it is 0 otherwise. Thus

$$0 < x < t < 1 + x < 2,$$

which implies $t \in (0, 2)$. If $0 < t \leq 1$, we must have $x \in (0, t)$. If $t > 1$, we must have $x \in (t - 1, 1)$. Therefore the PDF of T is:

$$f_T(t) = \begin{cases} \int_0^t dx = t, & \text{for } 0 < t \leq 1, \\ \int_{t-1}^1 dx = 2 - t, & \text{for } 1 < t < 2. \end{cases}$$

3 The Beta distribution

The Beta distribution is a continuous distribution on the interval $(0, 1)$. It is a generalization of the $\mathsf{Unif}(0, 1)$ distribution, allowing the PDF to be non-constant on $(0, 1)$.

Definition. Beta distribution A random variable X is said to have the *Beta distribution* with parameters $a > 0$ and $b > 0$ if its PDF is

$$f(x) = \frac{1}{\beta(a, b)} x^{a-1}(1-x)^{b-1}, 0 < x < 1,$$

where the normalizing constant $\beta(a, b)$ is chose to make the PDF integrate to 1:

$$\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx.$$

We denote $X \sim \mathsf{Beta}(a, b)$. For $a = b = 1$, $\mathsf{Beta}(a, b)$ is $\mathsf{Unif}(0, 1)$. See Figure 1.

4 The Gamma distribution

The Gamma distribution is a continuous distribution on the positive real line. It is a generalization of the Exponential distribution. While an Exponential random variable represents the waiting time for the first success under conditions of memorylessness, a Gamma random variable represents the total waiting time for multiple successes.

Definition (Gamma function)**.** The *gamma function* Γ is defined by

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

for real numbers $a > 0$. The gamma function is such that $\Gamma(a + 1) = a\Gamma(a)$. This implies $\Gamma(n) = (n - 1)!$ for any $n \geq 1$ positive integer.

Definition (Gamma distribution)**.** A random variable Y is said to have the *Gamma distribution* with parameters a and λ, where $a > 0$ and $\lambda > 0$, if its PDF is

$$f(y) = \frac{1}{\Gamma(a)} (\lambda y)^a e^{-\lambda y} \frac{1}{y}, \quad y > 0.$$

We write $Y \sim \mathsf{Gamma}(a, \lambda)$.

The general $\mathsf{Gamma}(a, \lambda)$ distribution can be constructed from the $\mathsf{Gamma}(a, 1)$ by a scale transformation. The PDF of $X \sim \mathsf{Gamma}(a, 1)$ is

$$f_X(x) = \frac{1}{\Gamma(a)} x^a e^{-x} \frac{1}{x}, \quad x > 0.$$

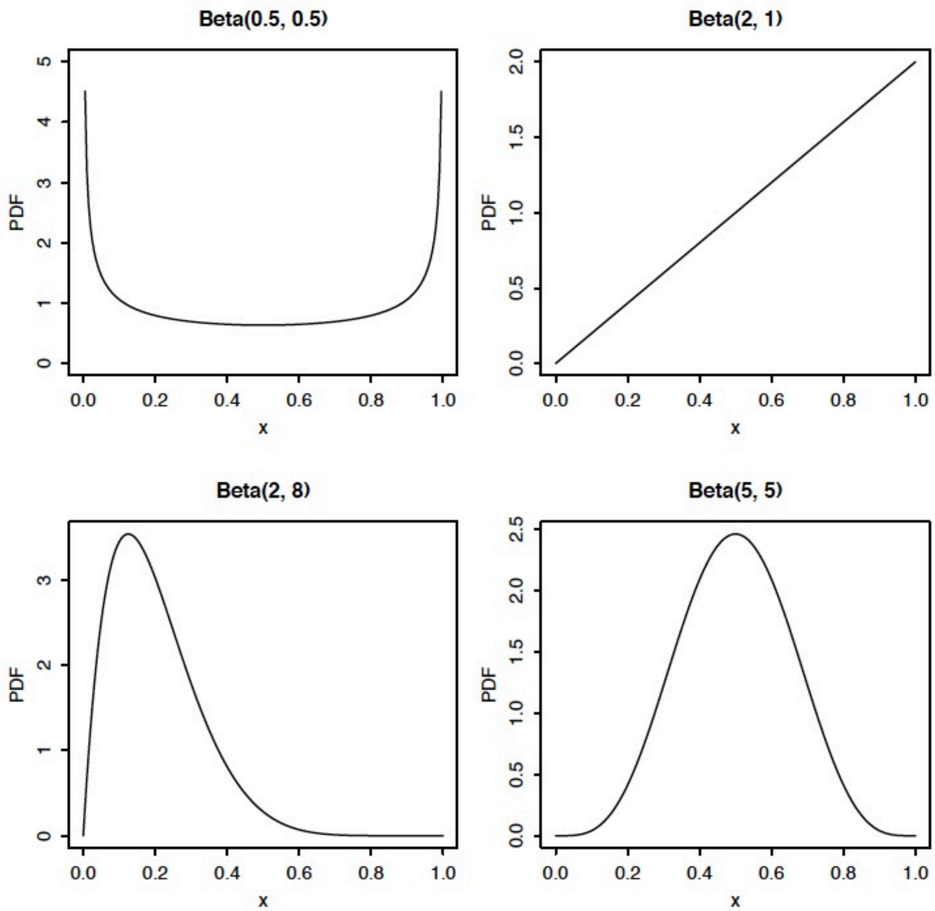

Figure 1: Beta PDFs for various values of a and b. **Top left panel**: $\text{Beta}(0.5, 0.5)$; if$a < 1$ and $b < 1$, the Beta PDF is U-shaped and opens upward. **Bottom right panel**: $\text{Beta}(5, 5)$; if $a > 1$ and $b > 1$, the Beta PDF opens down. If $a = b$, the Beta PDF is symmetric around $\frac{1}{2}$. **Top right panel**: $\text{Beta}(2, 1)$; if $a > b$, the Beta PDF favors larger values than $\frac{1}{2}$. **Bottom left panel**: $\text{Beta}(2, 8)$; if $a < b$, the Beta PDF favors values smaller than $\frac{1}{2}$.

Consider the random variable $Y = \frac{X}{\lambda}$ for some $\lambda > 0$. By the change of variables formula with $x = \lambda y$ and $\frac{dx}{dy} = \lambda$, the PDF of Y is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y}\frac{1}{\lambda y}\lambda = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y}\frac{1}{y}, \quad y > 0.$$

Thus $Y = \frac{X}{\lambda} \sim \mathsf{Gamma}(a, \lambda)$. Remark that the $\mathsf{Gamma}(1, \lambda)$ distribution coincides with the $\mathsf{Expo}(\lambda)$ distribution.

We determine the moments of $X \sim \mathsf{Gamma}(a, 1)$ using LOTUS:

$$\mathsf{E}(X^k) = \int_0^\infty \frac{1}{\Gamma(a)} x^{a+k-1} e^{-x}\, dx = \frac{\Gamma(a+k)}{\Gamma(a)}.$$

Thus $\mathsf{E}(X) = \frac{\Gamma(a+1)}{\Gamma(a)} = a$, $\mathsf{E}(X^2) = \frac{\Gamma(a+2)}{\Gamma(a)} = a(a+1)$, hence

$$\mathsf{Var}(X) = \mathsf{E}(X^2) - (\mathsf{E}X)^2 = a(a+1) - a^2 = a.$$

This implies that $\mathsf{E}(X) = \mathsf{Var}(X) = a$. Next we calculate the moments and the variance of $Y = \frac{X}{\lambda} \sim \mathsf{Gamma}(a, \lambda)$:

$$\begin{aligned}
\mathsf{E}(Y) &= \frac{1}{\lambda}\mathsf{E}(X) = \frac{a}{\lambda},\\
\mathsf{Var}(Y) &= \frac{1}{\lambda^2}\mathsf{Var}(X) = \frac{a}{\lambda^2},\\
\mathsf{E}(Y^k) &= \frac{1}{\lambda^k}\mathsf{E}(X^k) = \frac{1}{\lambda^k}\frac{\Gamma(a+k)}{\Gamma(a)}.
\end{aligned}$$

In the special case when a is an integer, we can represent a $\mathsf{Gamma}(a, \lambda)$ random variable as a convolution of $\mathsf{Expo}(\lambda)$ random variables.

Theorem. *Let $X_1, \ldots, X_n$ be i.i.d.* $\mathsf{Expo}(\lambda)$. *Then*

$$Y = X_1 + \ldots + X_n \sim \mathsf{Gamma}(n, \lambda).$$

Proof. By LOTUS, the MGF of Y is

$$\mathsf{E}\left(e^{tY}\right) = \int_0^\infty e^{ty}\frac{1}{\Gamma(n)}(\lambda y)^n e^{-\lambda y}\frac{dy}{y} = \frac{\lambda^n}{(\lambda - t)^n}\int_0^\infty \frac{1}{\Gamma(n)} e^{-(\lambda - t)y}((\lambda - t)y)^n \frac{dy}{y}.$$

The expression inside the integral is the $\Gamma(n, \lambda - t)$ PDF, for $t < \lambda$. Since the PDFs integrate to 1, we have

$$\mathsf{E}\left(e^{tY}\right) = \left(\frac{\lambda}{\lambda - t}\right)^n, \text{ for } t < \lambda.$$

Since the MGF of an $\mathsf{Expo}(\lambda)$ random variable is $\frac{\lambda}{\lambda - t}$ for $t < \lambda$, it follows that the MGF of the sum of n i.i.d. $\mathsf{Expo}(\lambda)$ random variables is equal with the MGF of Y. □

This result allows us to connect the Gamma distribution with Poisson processes. For a Poisson process of rate λ, the interarrival times are i.i.d. $\mathsf{Expo}(\lambda)$ random variables. But the total waiting time for the nth arrival is the sum of the first n interarrival times. Thus, from this theorem, $T_n \sim \Gamma(n, \lambda)$. The interarrival times in a Poisson process are Exponential random variables, while the raw arrival times are Gamma random variables. Note that the interarrival times are independent and identically distributed, while the raw arrival times are dependent and do not follow the same distribution. As such, $Y \sim \mathsf{Gamma}(a, \lambda)$ is the total waiting time for the ath arrival in a Poisson process with rate λ, and λ is the rate at which successes arrive.