Lecture 12: Multivariate transformation

We have considered transformations of a single random variable.
We now consider a vector of transformations of a random vector.
First, we consider the sum of two random variables.

The pdf of the sum of two random variables (convolution)

Let X and Y be random variables having joint pdf $f(x,y)$.
For any $t \in \mathscr{R}$,

$$P(X+Y \le t) = \int\int_{x+y\le t} f(x,y)dxdy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{t-x} f(x,y)dy\right]dx$$

Hence, the pdf of the sum $X+Y$ is

$$f_{X+Y}(t) = \frac{d}{dt}P(X+Y\le t) = \int_{-\infty}^{\infty}\frac{d}{dt}\left[\int_{-\infty}^{t-x} f(x,y)dy\right]dx = \int_{-\infty}^{\infty} f(x,t-x)dx$$

Similarly,

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f(t-y,y)dy$$

The pdf f_{X+Y} is called a convolution.
The interchange of differentiation and integration can be justified.

If X and Y are independent and f_X and f_Y are their marginal pdf's, then

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x) f_Y(t-x)dx = \int_{-\infty}^{\infty} f_X(t-y) f_Y(y)dy$$

Example

If $X \sim \textit{uniform}(0,1)$ and $Y \sim \textit{exponential}(0,1)$ are independent, what is the pdf of $X+Y$?

$$\begin{aligned} f_{X+Y}(t) &= \int_{-\infty}^{\infty} f_X(x) f_Y(t-x)dx = \int_{0<x<1, x<t} e^{-(t-x)}dx \\ &= \begin{cases} \int_0^1 e^{-(t-x)}dx = (e-1)e^{-t} & t > 1 \\ \int_0^t e^{-(t-x)}dx = 1 - e^{-t} & 0 < t \leq 1 \\ 0 & t \leq 0 \end{cases} \end{aligned}$$

When f_X or f_Y is 0 in some regions, we need to be careful about the integration limits.

Sometimes, one of the two formulas is easier to work with.

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(t-y) f_Y(y)dy = \int_{0<t-y<1, y>0} e^{-y}dy = ?$$

Example

Suppose that (X, Y) has the following bivariate normal pdf:

$$f(x,y) = \frac{\exp\left(-\frac{x^2}{2(1-\rho^2)} + \frac{\rho xy}{(1-\rho^2)} - \frac{y^2}{2(1-\rho^2)}\right)}{2\pi\sqrt{1-\rho^2}} \qquad (x,y) \in \mathscr{R}^2$$

What is the pdf of $X+Y$?

$$\begin{aligned}
f_{X+Y}(t) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2(1-\rho^2)} + \frac{\rho x(t-x)}{(1-\rho^2)} - \frac{(t-x)^2}{2(1-\rho^2)}\right) dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2+2\rho x^2+x^2-2\rho tx-2tx+t^2}{2(1-\rho^2)}\right) dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-t/2)^2}{1-\rho} - \frac{t^2}{4(1+\rho)}\right) dx \\
&= \frac{1}{2\sqrt{\pi(1+\rho)}} \exp\left(-\frac{t^2}{4(1+\rho)}\right)
\end{aligned}$$

which is $N(0, 2(1+\rho))$.

The pmf of the sum of two discrete random variables

We now turn to discrete random variables X and Y with joint pmf $f(x,y)$.

By definition, the pmf of $X+Y$ is

$$f_{X+Y}(t) = P(X+Y=t) = \sum_{x+y\leq t} f(x,y) = \sum_x f(x,t-x) = \sum_y f(t-y,y)$$

and if X and Y are independent with marginal pmf's f_X and f_Y, then

$$f_{X+Y}(t) = \sum_x f_X(x) f_Y(t-x) = \sum_y f_X(t-y) f_Y(y)$$

However, in some cases it is more convenient to work out the probability $P(X+Y=t)$ directly.

Example 4.3.1.

Let $X \sim Poisson(\theta)$ and $Y \sim Poisson(\lambda)$ be independent Poisson random variables so that the joint pmf of (X,Y) is

$$f(x,y) = \frac{\theta^x e^{-\theta}}{x!}\frac{\lambda^y e^{-\lambda}}{y!}, \quad x = 0,1,2,...,y = 0,1,2,...$$

What is the pmf of $U = X + Y$?
Using the formula, for $u = 0, 1, 2, ...$

$$\begin{aligned} f_U(u) &= \sum_y f(u-y, y) = \sum_{y=0}^{u} \frac{\theta^{u-y} e^{-\theta}}{(u-y)!} \frac{\lambda^y e^{-\lambda}}{y!} \\ &= \frac{e^{-(\theta+\lambda)}}{u!} \sum_{y=0}^{u} \binom{u}{y} \lambda^y \theta^{u-y} = \frac{e^{-(\theta+\lambda)}(\theta+\lambda)^u}{u!} \end{aligned}$$

This is the pmf of $Poisson(\theta + \lambda)$.
What is the joint pmf of $U = X + Y$ and Y.

$$\begin{aligned} f_{U,Y}(u,y) &= P(U = u, Y = y) = P(X + Y = u, Y = y) \\ &= P(X = u - y, Y = v) = \frac{\theta^{u-y} e^{-\theta}}{(u-y)!} \frac{\lambda^y e^{-\lambda}}{y!} \end{aligned}$$

for $u = y, y+1, y+2, ...,$ and $y = 0, 1, 2, ...$
U and Y are not independent, because

$$P(U = 0, Y = 1) = 0 \neq P(U = 0)P(Y = 1) = e^{-(\theta+\lambda)} \lambda e^{-\lambda}$$

To deal with a sum of independent random variables, it is convenient to use mgf's.

Theorem 4.2.12.

If $X_1,...,X_n$ are independent random variables with mgf's $M_{X_1}(t)$,...,$M_{X_n}(t)$, then the mgf of the sum $T = X_1 + \cdots + X_n$ is

$$M_T(t) = M_{X_1}(t) \cdots M_{X_n}(t)$$

and $M_T(t) < \infty$ iff $M_{X_i}(t) < \infty$, $i = 1,...,n$.

Proof.

By definition,

$$\begin{aligned} M_T(t) &= E(e^{tT}) = E(e^{t(X_1+\cdots+X_n)}) = E(e^{tX_1} \cdots e^{tX_n}) \\ &= E(e^{tX_1}) \cdots E(e^{tX_n}) = M_{X_1}(t) \cdots M_{X_n}(t) \end{aligned}$$

where the 4th equality follows from the independence of $X_1,...,X_n$ and Theorem 4.6.12.

This result and the uniqueness theorem (Theorem 2.3.11) lead to many useful results concerning the sum of independent random variables.

Theorem 4.2.14.

If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\gamma, \tau^2)$ are independent, then $T = X + Y \sim N(\mu + \gamma, \sigma^2 + \tau^2)$.

This is easy to show, because the mfg's of X and Y are $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$ and $M_Y(t) = e^{\gamma t + \tau^2 t^2/2}$, and thus the mgf of T is

$$M_X(t) M_Y(t) = e^{\mu t + \sigma^2 t^2/2} e^{\gamma t + \tau^2 t^2/2} = e^{(\mu+\gamma)t + (\sigma^2+\tau^2)t^2/2}$$

which is the mgf of $N(\mu + \gamma, \sigma^2 + \tau^2)$.
By the uniqueness theorem, we must have $T \sim N(\mu + \gamma, \sigma^2 + \tau^2)$.

Note that this result does not apply to the bivariate normal distribution example with $\rho \neq 0$, since in that example X and Y are not independent.

The previous result can be generalized to more than two independent random variables having normal distributions: If $X_1, ..., X_k$ are independent random variables, $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, ..., k$, then the sum $X_1 + \cdots + X_k \sim N(\mu_1 + \cdots + \mu_k, \sigma_1^2 + \cdots + \sigma_k^2)$.

This type of results is called additivity of distributions; i.e., if X and Y have distributions in a class of distributions and $X+Y$ has a distribution also in that class, then the distributions in this class are called additive distributions.

Additivity of the chi-square distributions

In Chapter 2 we showed that if $Z \sim N(0,1)$, then $Z^2 \sim$ the chi-square distribution with degree of freedom 1.

A more general result is as follows.

If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, ..., k$, are independent, then the distribution of

$$Y = \left(\frac{X_1 - \mu_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{X_k - \mu_k}{\sigma_k}\right)^2$$

is the chi-square distribution with degrees of freedom k.

First, $Z_i \sim N(0,1)$, $i = 1, ..., k$, and Z_i's are independent.

Second, each $Z_i^2 \sim$ chi-square with degree of freedom 1.

Third, the chi-square distribution with degrees of freedom r is the *gamma*$(r/2, 2)$ distribution.

Finally, the result follows from the next result.

Additivity of the gamma distributions

If $X_i \sim gamma(\alpha_i, \beta)$, $i = 1, ..., k$, are independent, then the sum $T = X_1 + \cdots + X_k \sim gamma(\alpha_1 + \cdots + \alpha_k, \beta)$.
This is because the mgf of X_i is $(1-\beta t)^{-\alpha_i}$ when $t < \beta^{-1}$.
By Theorem 4.2.12, the mgf of T is

$$(1-\beta t)^{-\alpha_1} \cdots (1-\beta t)^{-\alpha_k} = (1-\beta t)^{-(\alpha_1+\cdots+\alpha_k)} \qquad t < \beta^{-1}$$

which is the mgf of $gamma(\alpha_1 + \cdots + \alpha_k, \beta)$.

A non-additive class of distributions

The class of all uniform distributions is not additive.
If $X \sim uniform(0,1)$ and $Y \sim uniform(0,1)$ are independent and $f(x)$ is the indicator function of the interval $(0,1)$, then the pdf of $X+Y$ is

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f(t-y)f(y)dy = \int_{0<t-y<1, 0<y<1} dy$$

$$= \begin{cases} \int_{t-1}^{1} dy = 2-t & 1 < t < 2 \\ \int_0^t dy = t & 0 < t < 1 \\ 0 & \text{otherwise} \end{cases}$$

which is not a uniform pdf.

The next result is useful if the mgf is not finite in a neighborhood of 0.

Theorem 4.2.12A.

If $X_1, ..., X_n$ are independent random variables with chf's $\phi_{X_1}(t)$,...,$\phi_{X_n}(t)$, then the chf of the sum $T = X_1 + \cdots + X_n$ is

$$\phi_T(t) = \phi_{X_1}(t) \cdots \phi_{X_n}(t) \qquad t \in \mathscr{R}$$

Additivity of the Cauchy distribution

If U and V are independent random variables having the $Cauchy(0,1)$ distribution, what is the distribution of $U+V$?

Since the mgf of a Cauchy distribution is not finite except at 0, we apply Theorem 4.2.12A.

Note that the chf of $Cauchy(0,1)$ is $e^{-|t|}$, $t \in \mathscr{R}$.

By Theorem 4.2.12A, the chf of $U+V$ is $e^{-|t|}e^{-|t|} = e^{-2|t|}$, $t \in \mathscr{R}$, which is the chf of $Cauchy(0,2)$.

By the uniqueness theorem, $U+V \sim Cauchy(0,2)$.

Of course, the same result can be derived using the convolution, but the argument is much more complicated (exercise).

We now consider a more general transformation, a vector of transformations of a random vector.

The pdf of a multivariate transformation

Let X be a k-dimensional random vector with a joint pdf f_X and let $Y = g(X)$, where g is a Borel function from $\mathscr{R}^k$ to $\mathscr{R}^k$ (so that Y is a k-dimensional random vector). Let $A_1, \ldots, A_m$ be disjoint sets in $\mathscr{B}^k$ such that $P(X \in A_1 \cup \cdots \cup A_m) = 1$ and g on A_j is one-to-one with a nonvanishing Jacobian, i.e., the determinant $|\partial g(x)/\partial x| \neq 0$ on A_j, $j = 1, \ldots, m$. Then Y has the following joint pdf

$$f_Y(x) = \sum_{j=1}^{m} \Big| |\partial h_j(x)/\partial x| \Big| f_X\left(h_j(x)\right),$$

where h_j is the inverse function of g on A_j, and $\big||\partial h_j(x)/\partial x|\big|$ is the absolute value of the determinant of the matrix $\partial h_j(x)/\partial x$, $j = 1, \ldots, m$.

Example

Let $X = (X_1, X_2)$ be a 2-dimensional random vector with joint pdf f_X. Consider first the transformation $u = g(x_1, x_2) = (u_1, u_2) = (x_1, x_1 + x_2)$.

$$\frac{\partial u}{\partial x} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \qquad \left|\frac{\partial u}{\partial x}\right| = 1 \qquad \left|\frac{\partial x}{\partial u}\right| = 1$$

Using the transformation theorem, the joint pdf of $U = g(X)$ is

$$f_U(u_1, u_2) = f_X(u_1, u_2 - u_1),$$

The marginal pdf of $U_2 = X_1 + X_2$ is the same as that previously derived

$$f_{U_2}(u_2) = \int_{-\infty}^{\infty} f_X(u_1, u_2 - u_1) du_1.$$

Next, consider the transformation $v = g(x_1, x_2) = (v_1, v_2) = (x_1/x_2, x_2)$.

$$\frac{\partial v}{\partial x} = \begin{pmatrix} \frac{\partial v_1}{\partial x_1} & \frac{\partial v_1}{\partial x_2} \\ \frac{\partial v_2}{\partial x_1} & \frac{\partial v_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{x_2} & -\frac{x_1}{x_2^2} \\ 0 & 1 \end{pmatrix} \qquad \left|\frac{\partial v}{\partial x}\right| = \frac{1}{x_2} \qquad \left|\frac{\partial x}{\partial v}\right| = v_2$$

Using the transformation theorem, the joint pdf of $V = g(X)$ is

$$f_V(v_1, v_2) = |v_2| f_X(v_1 v_2, v_2),$$

The marginal pdf of $V_1 = X_1/X_2$ is

$$f_{V_1}(v_1) = \int_{-\infty}^{\infty} |v_2| f_X(v_1 v_2, v_2) dv_2.$$

In particular, if X_1 and X_2 are independent, then

$$f_{V_1}(z) = \int_{-\infty}^{\infty} |v_2| f_{X_1}(v_1 v_2) f_{X_2}(v_2) dv_2$$

Finally, consider the transformation $w = g(x_1, x_2) = (w_1, w_2) = (x_1 x_2, x_2)$.

$$\frac{\partial w}{\partial x} = \begin{pmatrix} \frac{\partial w_1}{\partial x_1} & \frac{\partial w_1}{\partial x_2} \\ \frac{\partial w_2}{\partial x_1} & \frac{\partial w_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} x_2 & x_1 \\ 0 & 1 \end{pmatrix} \qquad \left| \frac{\partial w}{\partial x} \right| = x_2 \qquad \left| \frac{\partial x}{\partial w} \right| = \frac{1}{w_2}$$

Using the transformation theorem, the joint pdf of $W = g(X)$ is

$$f_W(w_1, w_2) = |w_2|^{-1} f_X(w_1 w_2^{-1}, w_2),$$

The marginal pdf of $W_1 = X_1 X_2$ is

$$f_{W_1}(w_1) = \int_{-\infty}^{\infty} |w_2|^{-1} f_X(w_1 w_2^{-1}, w_2) dw_2.$$

In particular, if X_1 and X_2 are independent, then

$$f_{W_1}(w_1) = \int_{-\infty}^{\infty} |w_2|^{-1} f_{X_1}(w_1 w_2^{-1}) f_{X_2}(w_2) dw_2$$

A number of results can be derived from these results.

Example 4.3.3 (product of two beta random variables)

Let $X \sim beta(\alpha, \beta)$ and $Y \sim beta(\alpha+\beta, \gamma)$ be independent so that the joint pdf is

$$f_{X,Y}(y,x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)} y^{\alpha+\beta-1}(1-y)^{\gamma-1}$$

$$0 < x < 1, \quad 0 < y < 1$$

Consider $U = XY$ and $V = X$.
An application of the early result for product gives that

$$f_U(u) = \int_{-\infty}^{\infty} |v|^{-1} f(uv^{-1}, v) dv$$

But we need to be careful when the pdf is 0 in some intervals.
Note that the range of (U,V) is the region $0 < u < v < 1$ (why?).
Thus,

$$f_U(u) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} \int_u^1 v^{\alpha-1}(1-v)^{\beta-1} \left(\frac{u}{v}\right)^{\alpha+\beta-1} \left(1-\frac{u}{v}\right)^{\gamma-1} \frac{1}{v} dv$$

for $0 < u < 1$ ($f_U(u) = 0$ when $u < 0$ or $u > 1$.)

Making the change of variable $s = (u/v - u)/(1-u)$ so that $dy = -u/[v^2(1-u)]$, $s = 0$ when $v = 1$ and $s = 1$ when $v = u$, we obtain that

$$\begin{aligned} f_U(u) &= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \int_u^1 \left(\frac{u}{v} - u\right)^{\beta-1} \left(1 - \frac{u}{v}\right)^{\gamma-1} \frac{u}{v^2} dv \\ &= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \int_0^1 s^{\beta-1}(1-s)^{\gamma-1} ds \\ &= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)} \\ &= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \qquad 0 < u < 1 \end{aligned}$$

Thus, the product $XY \sim beta(\alpha, \beta+\gamma)$.

Example

If $X_i \sim N(0,1)$, $i = 1,...,n$, are independent, we want to show that, for every $i = 1,...,n$, the random variables $U = X_1^2 + \cdots + X_n^2$ and $V_i = X_i^2/U$ are independent, and we want to find the pdf of V_i.

The joint pdf of $X_1,...,X_n$ is

$$f_X(x_1,...,x_n) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2}\sum_{i=1}^n x_i^2\right) \qquad x_i \in \mathscr{R},\ i = 1,...,n$$

For $Y_i = X_i^2$, $i = 1,...,n$, an application of the transformation theorem shows that the joint pdf of $Y_1,..., Y_n$ is

$$f_Y(y_1,...,y_n) = \frac{1}{2^{n/2}\sqrt{y_1\cdots y_n}} \exp\left(-\frac{1}{2}\sum_{i=1}^n y_i\right) \qquad y_i > 0,\ i = 1,...,n.$$

Let $U = Y_1 + \cdots + Y_n$ and $V_i = Y_i/U$, $i = 1,...,n$.

Then $Y_i = UV_i$, $i = 1,...,n-1$, and $Y_n = UV_n = U(1 - V_1 - \cdots - V_{n-1})$.

$$\left|\frac{\partial(y_1,...,y_n)}{\partial(u,v_1,...,v_{n-1})}\right| = \begin{vmatrix} v_1 & u & 0 & \cdots & 0 \\ v_2 & 0 & u & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ v_{n-1} & 0 & 0 & \cdots & u \\ v_n & 0 & 0 & \cdots & 0 \end{vmatrix} = v_n u^{n-1}$$

where $v_n = 1 - v_1 - \cdots - v_{n-1}$.

Since $Y_i > 0$, $i = 1,...,n$, $U > 0$ and $0 < V_i < 1$, $i = 1,...,n$.

The joint pdf for U and $V_1, ..., V_{n-1}$ is

$$f_Y(uv_1, ..., uv_n) v_n u^{n-1} = \frac{1}{2^{n/2}} u^{n/2-1} e^{-u/2} \sqrt{\frac{1 - v_1 - \cdots - v_{n-1}}{v_1 \cdots v_{n-1}}}$$

Therefore, we know that U and $(V_1, ..., V_{n-1})$ are independent and, hence, U and V_i are independent for any $i = 1, ..., n-1$.

Since $V_n = 1 - V_1 - \cdots - V_{n-1}$, U and V_n are independent.

If $W = U - X_i^2$, then

$$V_i = \frac{X_i^2}{U} = \frac{X_i^2}{W + X_i^2} = \frac{X_i^2/W}{1 + X_i^2/W}$$

Previously we showed that $Z = X_i^2/W$ has pdf

$$\frac{\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)} \frac{z^{-1/2}}{(1+z)^{(n-1)/2}}, \qquad z > 0$$

Then the transformation $V_i = Z/(1+Z)$ gives the following pdf for V_i:

$$\frac{\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)} v^{-1/2}(1-v)^{(n-3)/2}, \qquad 0 < v < 1$$

which is *beta*$(1/2, (n-1)/2)$.