Bivariate Transformations

November 4 and 6, 2008

Let $X$ and $Y$ be jointly continuous random variables with density function $f_{X,Y}$ and let $g$ be a one to one transformation. Write $(U, V) = g(X, Y)$. The goal is to find the density of $(U, V)$.

1 Transformation of Densities

Above the rectangle from $(u, v)$ to $(u + \Delta u, v + \Delta v)$ we have the joint density function $f_{U,V}(u, v)$ and probability

$$f_{U,V}(u, v)\Delta u\Delta v$$

Write $(x, y) = g^{-1}(u, v)$, then this probability is equal to the area of image of the rectangle from $(u, v)$ to $(u + \Delta u, v + \Delta v)$ under the map $g^{-1}$ times the density $f_{X,Y}(x, y)$.

The linear approximations for $g^{-1}$ give, in vector form, two sides in the parallelogram that approximates the image of the rectangle.

$$g^{-1}(u + \Delta u, v) \approx g^{-1}(u, v) + \frac{\partial}{\partial u} g^{-1}(u, v)\Delta u = g^{-1}(u, v) + \left( \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u} \right) \Delta u,$$

and

$$g^{-1}(u, v + \Delta v) \approx g^{-1}(u, v) + \frac{\partial}{\partial v} g^{-1}(u, v)\Delta v = g^{-1}(u, v) + \left( \frac{\partial x}{\partial v}, \frac{\partial y}{\partial v} \right) \Delta v.$$  

The area of the rectangle is given by the norm of cross product

$$\left| \left( \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u} \right) \times \left( \frac{\partial x}{\partial v}, \frac{\partial y}{\partial v} \right) \right| \Delta u \Delta v.$$  

This is computed using the determinant of the Jacobian matrix

$$J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Thus,

$$f_{U,V}(u, v)\Delta u\Delta v \approx f_{X,Y}(g^{-1}(u, v)) |J(u, v)| \Delta u \Delta v$$

with the approximation improving as $\Delta y, \Delta v \to 0$. Thus, give the formula for the transformation of bivariate densities.

$$f_{U,V}(u, v) = f_{X,Y}(g^{-1}(u, v)) |J(u, v)|.$$
Example 1. If $A$ is a one-to-one linear transformation and $(U, V) = A(X, Y)$, then
\[ f_{U,V}(u, v) = f_{X,Y}(A^{-1}(u, v)) |\det(A^{-1})| = \frac{1}{\det(A)} f_{X,Y}(A^{-1}(u, v)). \]

2 Convolution

Example 2 (convolution). Let
\[ u = x + y, \quad v = x. \]
Then,
\[ x = v, \quad y = u - v \]
and
\[ J(u, v) = \det \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = -1 \]
This yields
\[ f_{U,V}(u, v) = f_{X,Y}(v, u - v). \]
The marginal distribution for $v$ can be found by taking an integral
\[ f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(v, u - v) \, dv. \]
If $X$ and $Y$ are independent, then
\[ f_U(u) = \int_{-\infty}^{\infty} f_X(v) f_Y(u - v) \, dv. \]
This is called the convolution is often written $f_{X+Y} = f_X * f_Y$.

Example 3. Let $X$ and $Y$ be independent random variables uniformly distributed on $[0, 1]$. Then $U = X + Y$ can take values from 0 to 2.
\[ f_U(u) = \int_{-\infty}^{\infty} I_{[0,1]}(v) I_{[0,1]}(u - v) \, dv = \int_0^1 I_{[0,1]}(u - v) \, dv. \]
Now
\[ 0 < u - v < 1 \quad \text{or} \quad u - 1 < v < u. \]
In addition, $0 < v < 1$. If $0 < u < 1$, then combining the two restrictions gives $0 < v < u$ and
\[ f_U(u) = \int_0^1 I_{[0,1]}(u - v) \, dv = \int_0^u dv = u. \]
If $1 < u < 2$, then combining the two restrictions gives $u < v < 1$ and
\[ f_U(u) = \int_0^1 I_{[0,1]}(u - v) \, dv = \int_u^1 dv = 1 - u. \]
Combining, we write
\[ f_{X+Y}(u) = \begin{cases} u & \text{if } 0 < u < 1, \\
1 - u & \text{if } 1 < u < 2. \end{cases} \]
Example 4. For $X$ and $Y$ be independent standard normal random variables. Then

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \exp^{-\frac{y^2}{2}} = \frac{1}{2\pi} \exp^{-\frac{x^2 + y^2}{2}}.$$ 

and change to polar coordinates. Here, we know the inverse transformation $g^{-1}(x, y)$

$$x = r \cos \theta, \quad y = r \sin \theta.$$ 

The Jacobian matrix has determinant

$$J(u, v) = \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \det \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} = r \cos^2 \theta + r \sin^2 \theta = r.$$ 

Thus,

$$f_{R, \Theta}(r, \theta) = \frac{1}{2\pi} r \exp^{-\frac{r^2}{2}}.$$ 

As a consequence, $R$ and $\Theta$ are independent and $\Theta$ is uniform on $[0, 2\pi)$. In addition, this transformation explains the constant $1/\sqrt{2\pi}$ in the density for the standard normal. We can use this transformation and the probability transform to simulate a pair of independent standard normal random variables.

The cumulant distribution function for $R$, known as the Rayleigh distribution $F_R(r) = 1 - \exp^{-\frac{r^2}{2}}$.

Thus, $F^{-1}(w) = \sqrt{-2 \log(1 - w)}$. If $U$ and $W$ are independent random variables uniformly distributed on $[0, 1]$, then so are $U$ and $V = 1 - W$. We can represent the random variables $R$ and $\Theta$ by

$$R = \sqrt{-2 \log V} \quad \text{and} \quad \Theta = 2\pi U.$$ 

In turn, we can represent the random variables $X$ and $Y$ by

$$X = \sqrt{-2 \log V} \cos(2\pi U) \quad \text{and} \quad X = \sqrt{-2 \log V} \sin(2\pi U).$$ 

Finally,

$$\theta = \begin{cases} \tan^{-1} \frac{y}{x}, & \text{if } x > 0, \\ \pi + \tan^{-1} \frac{y}{x}, & \text{if } x < 0. \end{cases}$$ 

The density of $U = \tan \Theta$.

$$f_U(u) = 2 \cdot \frac{1}{2\pi} \cdot \frac{1}{1 + u^2} = \frac{1}{\pi} \frac{1}{1 + u^2}.$$ 

The factor of 2 arises because the map $(x, y, \tan^{-1}(y/x))$ is 2 to 1. Thus, the Cauchy distribution arises from the ratio of independent normal random variables.

For discrete random variables, we can write the convolution

$$f_{X+Y}(u) = \sum_v f_X(v) f_Y(u - v).$$ 

Example 5. If $X$ and $Y$ are independent Poisson random variables with respective parameters $\lambda$ and $\mu$, then

$$f_{X+Y}(u) = \sum_{v=0}^{u} \frac{\lambda^v}{v!} e^{-\lambda} \frac{\mu^{u-v}}{(u-v)!} e^{-\mu} = \frac{1}{u!} e^{-(\lambda + \mu)} \sum_{v=0}^{u} \frac{u!}{v!(u-v)!} \lambda^v \mu^{u-v} = \frac{(\lambda + \mu)^u}{u!} e^{-(\lambda + \mu)}.$$
3 Tower Property

Again, if we write \( a(x) = E[g(Y)|X = x] \). Then,

\[
E[h(X)a(X)] = \int_{-\infty}^{\infty} h(x)a(x)f_X(x) dx = \int_{-\infty}^{\infty} h(x) \left( \int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x) dy \right) f_X(x) dx
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x)g(y)f_{Y|X}(y|x)f_X(x) dydx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x)g(y)f_{X,Y}(x,y) dydx
\]

\[
= Eh(X)g(Y).
\]

In summary, \( E[h(X)E[g(Y)|X]] = E[h(X)g(Y)] \). A similar gives the identity for discrete random variables.

4 Law of Total Variance

\[
\text{Var}(Y) = E[Y^2] - (EY)^2
\]

\[
= E[E[Y^2|X]] - (E[E[Y|X]])^2
\]

\[
= E[\text{Var}(Y|X) + (E[Y|X]^2 - (E[E[Y|X]])^2)]
\]

\[
= E[\text{Var}(Y|X)] + \text{Var}(E[Y|X]).
\]

The second term is variance in \( Y \) due to the variation in \( X \). The first term is the variation in \( Y \) given the value of \( X \).

Example 6. For the bivariate standard normal,

\[ E[Y|X] = \rho X, \text{ so } \text{Var}(E[Y|X]) = \rho^2 \]

and

\[ \text{Var}(Y|X) = 1 - \rho^2, \text{ so } E[\text{Var}(Y|X)] = 1 - \rho^2 \]

giving

\[ \text{Var}(Y) = (1 - \rho^2) + \rho^2 = 1. \]

5 Hierarchical Models

Because

\[ f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x). \]

we can introduce a bivariate density function by given the density for \( X \) and the conditional density for \( Y \) given the value for \( X \). We then recover the density for \( Y \) by taking an integral. A similar statement holds for discrete random variables.

Example 7. Let \( X \) be a Poisson random variable with parameter \( \lambda \) and consider \( Y \), the number of successes in \( X \) Bernoulli trials. Then,

\[ f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad f_{Y|X}(y|x) = \binom{x}{y} p^y (1-p)^{x-y}. \]
In particular, the conditional mean $E[Y|X] = pX$ and $EY = E[E[Y|X]] = E[pX] = p\lambda$. The conditional variance $\text{Var}(Y|X) = p(1-p)X$. Consequently, by the law of total variance,

$$\text{Var}(Y) = E[\text{Var}(Y|X)] + \text{Var}(E[Y|X]) = E[p(1-p)X] + Var(pX) = p(1-p)\lambda + p^2\lambda = p\lambda.$$ 

The joint density,

$$f_{X,Y}(x, y) = \binom{x}{y} p^y (1-p)^{x-y} \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \geq y$$

and

$$f_Y(y) = \sum_{x=y}^{\infty} f_{X,Y}(x, y) = \sum_{x=y}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} \frac{\lambda^x}{x!} e^{-\lambda} = \frac{(p\lambda)^y}{y!} \sum_{x=y}^{\infty} \frac{((1-p)\lambda)^{x-y}}{(x-y)!} e^{-\lambda} = \frac{(p\lambda)^y}{y!} e^{(1-p)\lambda} e^{-\lambda} = \frac{(p\lambda)^y}{y!} e^{-p\lambda}.$$ 

Thus, we see that $Y$ is a Poisson random variable with parameter $p\lambda$.

## 6 Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

- For a $d$-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_\mathbf{X}(\mathbf{x}) = P\{X = \mathbf{x}\}$.
  - For all $\mathbf{x}$, $f_\mathbf{X}(\mathbf{x}) \geq 0$ and $\sum_{\mathbf{x}} f_\mathbf{X}(\mathbf{x}) = 1$.
  - $P\{X \in A\} = \sum_{\mathbf{x} \in A} f_\mathbf{X}(\mathbf{x})$ and $Eg(X) = \sum_{\mathbf{x}} g(\mathbf{x}) f_\mathbf{X}(\mathbf{x})$
  - For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint mass function $f_{X,Y}(\mathbf{x}, \mathbf{y}) = P\{X = \mathbf{x}, Y = \mathbf{y}\}$, marginal mass function $f_\mathbf{X}(\mathbf{x}) = \sum_{\mathbf{y}} f_{X,Y}(\mathbf{x}, \mathbf{y})$, and conditional mass function $f_{Y|X}(\mathbf{y}|\mathbf{x}) = P\{Y = \mathbf{y}|X = \mathbf{x}\} = f_{X,Y}(\mathbf{x}, \mathbf{y})/f_\mathbf{X}(\mathbf{x})$.
  - $E[g(Y)|X = \mathbf{x}] = \sum_{\mathbf{y}} g(\mathbf{y}) f_{Y|X}(\mathbf{y}|\mathbf{x})$.

- For a $d$-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_\mathbf{X}(\mathbf{x})$.
  - For all $\mathbf{x}$, $f_\mathbf{X}(\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}^d} f_\mathbf{X}(\mathbf{x}) \, d\mathbf{x} = 1$.
  - $P\{X \in A\} = \int_A f_\mathbf{X}(\mathbf{x}) \, d\mathbf{x}$ and $Eg(X) = \int_{\mathbb{R}^d} g(\mathbf{x}) f_\mathbf{X}(\mathbf{x}) \, d\mathbf{x}$
  - For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint density function $f_{X,Y}(\mathbf{x}, \mathbf{y})$, marginal density function $f_\mathbf{X}(\mathbf{x}) = \int_{\mathbb{R}^c} f_{X,Y}(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}$, and conditional density function $f_{Y|X}(\mathbf{y}|\mathbf{x}) = f_{X,Y}(\mathbf{x}, \mathbf{y})/f_\mathbf{X}(\mathbf{x})$.
  - $E[g(Y)|X = \mathbf{x}] = \int_{\mathbb{R}^c} g(\mathbf{y}) f_{Y|X}(\mathbf{y}|\mathbf{x}) \, d\mathbf{y}$. 

5
Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $A_1, A_2, \ldots, A_d$,

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_d \in A_d\} = P\{X_1 \in A_1\} P\{X_2 \in A_2\} \cdots P\{X_d \in A_d\}.$$  

- For either mass functions or density functions, the joint mass or density function is the product of the 1-dimensional marginals.

$$f_X(x) = f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_d}(x_d).$$

- $E[g_1(X_1)g_2(X_2) \cdots g_d(X_d)] = E[g_1(X_1)]E[g_2(X_2)] \cdots E[g_d(X_d)].$

- For discrete random variables, the probability generating function of the sum is the product of 1-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z) \cdots \rho_{X_d}(z).$$

- For continuous random variables, the moment generating function of the sum is the product of 1-dimensional probability generating functions.

$$M_{X_1+X_2+\cdots+X_d}(t) = M_{X_1}(t)M_{X_2}(t) \cdots M_{X_d}(t).$$

- For $g : B \to \mathbb{R}^d$ is one-to-one, write $U = g(X)$ and let $J(u)$ denote the Jacobian matrix for $g^{-1}$. Then the $ij$-th entry in matrix

$$J_{ij}(u) = \frac{\partial x_i}{\partial u_j}$$

The density

$$f_U(u) = f_X(g^{-1}(u))|\det(J(u))|.$$