12 — TRANSFORMING BIVARIATE DENSITY FUNCTIONS

Having seen how to transform the probability density functions associated with a single random variable, the next logical step is to see how to transform bivariate probability density functions.

Integration with two Independent Variables

Consider $f(x_1, x_2)$, a function of two independent variables. Using cartesian coordinates, $f(x_1, x_2)$ might be represented by a surface above the x_1–x_2 plane; the value of $f(x_1, x_2)$ at any point (x_1, x_2) corresponds to the height above the point.

With two independent variables, integration is normally expressed with a double-integral sign and integration is over some range R, a specified area in the x_1–x_2 plane:

$$\iint_R f(x_1, x_2) \, dx_1 \, dx_2$$

Such an integral represents a volume. If R is a circle then this integral corresponds to the volume of a cylinder standing on R whose upper end is cut by the surface $f(x_1, x_2)$.

The following integral gives the volume of a cone whose height is h and whose base is a circle of radius a centred on the origin $(0, 0)$, this circle being the region R:

$$\iint_R h \left(1 - \frac{\sqrt{x_1^2 + x_2^2}}{a} \right) dx_1 \, dx_2$$

Note that at the centre of the circle $\sqrt{x_1^2 + x_2^2} = 0$ and the value of the integrand is h. At the edge of the circle $\sqrt{x_1^2 + x_2^2} = a$ and the value of the integrand is 0.

In principle, two integrations are carried out in turn:

$$4 \int_0^a \left[\int_0^{\sqrt{a^2 - x_2^2}} h \left(1 - \frac{\sqrt{x_1^2 + x_2^2}}{a} \right) dx_1 \right] dx_2 \qquad (12.1)$$

In this rearrangement, integration is over one quadrant of the circle and the result is multiplied by 4. For a given value of x_2 integration is along a strip one end of which is at $x_1 = 0$ and the other end of which is at $x_1 = \sqrt{a^2 - x_2^2}$. This accounts for the limits on the inner integration.

Already, a seemingly simple example of integration with two independent variables is beginning to become uncomfortably hard!

There are better ways of determining the volume of a cone but by judicious substitution of both independent variables even the present approach can be greatly simplified.

Integration by Substitution of two new Variables

The general formula for integration by substitution of a new variable was given as (11.1):

$$\int_a^b f(x) \, dx = \int_{y(a)}^{y(b)} f\big(x(y)\big) \frac{dx}{dy} \, dy$$

The transformation function is $y(x)$ and its inverse is $x(y)$.

The equivalent formula when there are two independent variables is:

$$\iint_{R_x} f(x_1, x_2)\, dx_1\, dx_2 = \iint_{R_y} f\big(x_1(y_1, y_2), x_2(y_1, y_2)\big) \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\, dy_1\, dy_2 \qquad (12.2)$$

There are *two* transformation functions, $y_1(x_1, x_2)$ and $y_2(x_1, x_2)$, and their inverses are $x_1(y_1, y_2)$ and $x_1(y_1, y_2)$.

The regions R_x and R_y are identical subject to the first being specified in the x_1–x_2 plane and the second being specified in the y_1–y_2 plane.

The item $\dfrac{\partial(x_1, x_2)}{\partial(y_1, y_2)}$ is called a Jacobian and is defined as:

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\[2ex] \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}$$

To simplify (12.1) above, use the transformation functions:

$$y_1 = \sqrt{x_1^2 + x_2^2} \qquad\qquad\qquad x_1 = y_1 \cos y_2$$
$$\text{and their inverses}$$
$$y_2 = \tan^{-1}\left(\frac{x_2}{x_1}\right) \qquad\qquad\qquad x_2 = y_1 \sin y_2$$

Note that:

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \cos y_2 & -y_1 \sin y_2 \\ \sin y_2 & y_1 \cos y_2 \end{vmatrix} = y_1$$

Using (12.2), the integration in (12.1) becomes:

$$4 \int_0^{\frac{\pi}{2}} \left[\int_0^a h\left(1 - \frac{y_1}{a}\right) y_1\, dy_1 \right] dy_2$$

This is, of course, simply a transformation from cartesian coordinates to polar coordinates. In the first, a small element of area is $\delta x_1.\delta x_2$ whereas in the second a small element of area is $\delta y_1.y_1\delta y_2$.

Integration is again over one quadrant of the circle. The inner integration is along a radius and, in the outer integration, this radius is swept through an angle of $90°$.

Continuing:

$$4 \int_0^{\frac{\pi}{2}} h\left(\frac{a^2}{2} - \frac{a^3}{3a}\right) dy_2 = 4 \int_0^{\frac{\pi}{2}} h\, \frac{a^2}{6}\, dy_2 = 4h\, \frac{a^2}{6}\, \frac{\pi}{2} = \frac{\pi a^2 h}{3}$$

The result is the familiar formula for the volume of a cone.

Application to Bivariate Probability Density Functions

Formula (12.2) has direct application to the process of transforming bivariate probability density functions...

Suppose X_1 and X_2 are two random variables whose bivariate probability density function is $f(x_1, x_2)$. It is common practice to represent a given pair of values of the two random variables X_1 and X_2 as a point in the x_1–x_2 plane.

By definition:

$$\text{P}(X_1, X_2 \text{ lies in a specified region } R_x) = \iint_{R_x} f(x_1, x_2)\, dx_1\, dx_2 \qquad (12.3)$$

Any function of a random variable (or indeed of two or more random variables) is itself a random variable. If y_1 and y_2 are taken as transformation functions, both $y_1(X_1, X_2)$ and $y_2(X_1, X_2)$ will be derived random variables. Let $Y_1 = y_1(X_1, X_2)$ and $Y_2 = y_2(X_1, X_2)$.

Take R_y as identical to the region R_x but specified in the y_1–y_2 plane. Necessarily:

$$\text{P}(Y_1, Y_2 \text{ lies in a specified region } R_y) = \text{P}(X_1, X_2 \text{ lies in a specified region } R_x)$$

From this and by (12.3) and (12.2):

$$\text{P}(Y_1, Y_2 \text{ lies in a specified region } R_y) = \iint_{R_y} f\big(x_1(y_1, y_2), x_2(y_1, y_2)\big) \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\, dy_1\, dy_2$$

Notice that the integrand is expressed wholly in terms of y_1 and y_2.

Calling this integrand $g(y_1, y_2)$:

$$\text{P}(Y_1, Y_2 \text{ lies in a specified region } R_y) = \iint_{R_y} g(y_1, y_2)\, dy_1\, dy_2$$

This demonstrates that $g(y_1, y_2)$ is the probability density function associated with the two random variables Y_1 and Y_2.

The requirements for f and g to be single valued and non-negative are just as in the one-variable case and it is customary for the relationship between a probability density function $f(x_1, x_2)$, the inverses $x_1(y_1, y_2)$ and $x_2(y_1, y_2)$ of a pair of transformation functions, and the derived probability density function $g(y_1, y_2)$ to be written:

$$g(y_1, y_2) = f\big(x_1(y_1, y_2), x_2(y_1, y_2)\big) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right|$$

This is directly analogous to relationship (11.4) given for the transformation of a single random variable into another.

Summary — Single Variable and Bivariate Transformations

In this section, a summary of the single variable case and a summary of the bivariate case are presented together so that the correspondence between the two can readily be seen.

Transformation of a single random variable:

- Start with a random variable X.

- Assume the associated probability density function is $f(x)$.

- Choose a transformation function $y(x)$.

- Let the derived random variable be $Y = y(X)$.

- Assume the associated probability density function is $g(y)$.

- Assume the inverse of the transformation function is $x(y)$.

- The relationship between $f(x)$ and $g(y)$ is:

$$g(y) = f\big(x(y)\big) \left| \frac{dx}{dy} \right|$$

- As a special case, if $f(x)$ corresponds to a uniform distribution, the relationship is:

$$g(y) = \left| \frac{dx}{dy} \right|$$

Transformation of a pair of random variables:

- Start with two random variables X_1 and X_2.

- Assume the associated bivariate probability density function is $f(x_1, x_2)$.

- Choose two transformation functions $y_1(x_1, x_2)$ and $y_2(x_1, x_2)$.

- Let the derived random variables be $Y_1 = y_1(X_1, X_2)$ and $Y_2 = y_2(X_1, X_2)$.

- Assume the associated bivariate probability density function is $g(y_1, y_2)$.

- Assume the inverses of the two transformation functions are $x_1(y_1, y_2)$ and $x_2(y_1, y_2)$.

- The relationship between $f(x_1, x_2)$ and $g(y_1, y_2)$ is:

$$g(y_1, y_2) = f\big(x_1(y_1, y_2), x_2(y_1, y_2)\big) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right|$$

- As a special case, if $f(x_1, x_2)$ corresponds to a uniform distribution, the relationship is:

$$g(y_1, y_2) = \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \tag{12.4}$$

Example — The Box–Muller Transformation

An earlier attempt to transform a uniform distribution into a normal distribution proved unsuccessful. Fortunately the difficulties can be overcome by starting with the bivariate equivalent of the uniform distribution.

Suppose X_1 and X_2 are two independent random variables each distributed Uniform(0,1). Bringing these together leads to the following bivariate probability density function:

$$f(x_1, x_2) = \begin{cases} 1, & \text{if } 0 \leqslant x_1, x_2 < 1 \\ 0, & \text{otherwise} \end{cases}$$

Informally, the function $f = 1$ when (x_1, x_2) lies in a unit square which has one corner at the origin but $f = 0$ if (x_1, x_2) lies outside this square.

This is the uniform distribution assumed in relationship (12.4).

Suppose that the transformation functions are:

$$y_1 = \sqrt{-2\ln(x_1)}\,\cos(2\pi x_2) \qquad \text{and} \qquad y_2 = \sqrt{-2\ln(x_1)}\,\sin(2\pi x_2) \qquad (12.5)$$

First, derive the inverse functions:

$$x_1 = e^{-\frac{1}{2}(y_1^2 + y_2^2)} \qquad \text{and} \qquad x_2 = \frac{1}{2\pi}\tan^{-1}\left(\frac{y_2}{y_1}\right)$$

Next, evaluate the Jacobian:

$$\begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\[2mm] \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} -y_1\, e^{-\frac{1}{2}(y_1^2+y_2^2)} & -y_2\, e^{-\frac{1}{2}(y_1^2+y_2^2)} \\[2mm] \dfrac{1}{2\pi}\dfrac{-y_2/y_1^2}{1+(y_2/y_1)^2} & \dfrac{1}{2\pi}\dfrac{1/y_1}{1+(y_2/y_1)^2} \end{vmatrix} = -\frac{1}{2\pi}\frac{e^{-\frac{1}{2}(y_1^2+y_2^2)}}{1+(y_2/y_1)^2}\left[1+\left(\frac{y_2}{y_1}\right)^2\right]$$

From (12.4):

$$g(y_1, y_2) = \left|\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right| = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_1^2} \times \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_2^2}$$

Recall (10.3) and note that this bivariate probability density function corresponds to two independent random variables Y_1 and Y_2 which are each distributed Normal(0,1).

It is now clear how to transform a uniform distribution into a normal distribution:

- Start with two independent random variables X_1 and X_2 which are each distributed Uniform(0,1).

- From the transformation functions y_1 and y_2 derive two new random variables being $Y_1 = y_1(X_1, X_2)$ and $Y_2 = y_2(X_1, X_2)$.

- The derived random variables will each independently be distributed Normal(0,1).

This process is known as the Box–Muller transformation.

Box–Muller Refinement

The following procedure, written in a hypothetical programming language, makes use of the Box–Muller transformation; repeated calls of this procedure will return random numbers which are distributed Normal(0,1):

```
PROCEDURE normal
  X1 = uniform(0,1)
  X2 = uniform(0,1)
  Y1 = sqrt(-2*ln(X1))*cos(2*pi*X2)
  Y2 = sqrt(-2*ln(X1))*sin(2*pi*X2)
  RETURN Y1
END
```

It is assumed that `uniform`, `sqrt`, `ln`, `cos`, and `sin` are library procedures which have the obvious effects. In particular, repeated calls of `uniform(0,1)` will return random numbers which are distributed Uniform(0,1).

Mathematically the procedure is fine but it is not altogether satisfactory from a Computer Science point of view. Most obviously, the value `Y2` is computed but never used. It would be better to arrange for the procedure to have two states. In one state, both `Y1` and `Y2` would be evaluated and the value of `Y1` returned. The value of `Y2` would be retained so that it could be returned on the next call when the procedure would be in the alternate state in which no evaluation would be necessary.

The procedure also makes two calls of `sqrt`, two calls of `ln` and one each of `cos` and `sin`. All six of these calls are quite expensive in computer time and it is possible to be much more efficient.

Instead of starting with the two random variables X_1 and X_2 which are each distributed Uniform(0,1) a better approach is to begin with two other random variables W_1 and W_2 whose values represent the (cartesian) coordinates of a point in a unit circle centred on the origin. All points in the circle are equally likely, just as in the raindrops and pond example discussed earlier.

Assuming the values of W_1 and W_2 are w_1 and w_2 respectively, the random variables X_1 and X_2 are then given values:

$$x_1 = w_1^2 + w_2^2 \qquad \text{and} \qquad x_2 = \frac{1}{2\pi} \tan^{-1}\left(\frac{w_2}{w_1}\right) \tag{12.6}$$
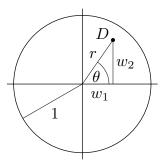
It will be demonstrated shortly that these derived random variables X_1 and X_2 are both distributed Uniform(0,1) and can therefore be used as before.

At this stage, the introduction of the two random variables W_1 and W_2 hardly seems to have led to an improvement but it will be shown that, by their use, the number of expensive procedure calls can be greatly reduced.

To appreciate how this revised approach works and why it leads to greater efficiency, it is necessary to revisit the circular pond. . .

In the figure below, the coordinates of the point D are shown as (w_1, w_2), these being the values of the random variables W_1 and W_2:

From the figure, r is the distance of point D from the centre and $r^2 = w_1^2 + w_2^2$ but, from (12.6), $x_1 = w_1^2 + w_2^2$.

Hence:

$$x_1 = r^2 \qquad \text{or} \qquad r = \sqrt{x_1} \tag{12.7}$$

The value of the derived random variable X_1 is therefore the *square* of the distance r of D from the centre and, from the experience of the raindrops and pond example, it is distributed Uniform(0,1).

From the figure, $\theta = \tan^{-1}\left(\dfrac{w_2}{w_1}\right)$ but, from (12.6), $x_2 = \dfrac{1}{2\pi} \tan^{-1}\left(\dfrac{w_2}{w_1}\right)$.

Hence:

$$x_2 = \frac{\theta}{2\pi} \qquad \text{or} \qquad \theta = 2\pi\, x_2 \tag{12.8}$$

Assuming a two-argument inverse-tangent function is used (such as `ATAN2` in Excel), θ will be uniformly distributed over the range 0 to 2π. This ensures that the value of the derived random variable X_2 is distributed Uniform(0,1).

It is now clear that both X_1 and X_2 are distributed Uniform(0,1).

From the figure and from (12.7) and (12.8):

$$w_1 = r\cos\theta = \sqrt{x_1}\,\cos(2\pi x_2) \qquad \text{so} \qquad \cos(2\pi x_2) = \frac{w_1}{\sqrt{x_1}}$$

and:

$$w_2 = r\sin\theta = \sqrt{x_1}\,\sin(2\pi x_2) \qquad \text{so} \qquad \sin(2\pi x_2) = \frac{w_2}{\sqrt{x_1}}$$

The transformation functions (12.5) can therefore be rewritten:

$$y_1 = \sqrt{\frac{-2\ln(x_1)}{x_1}}\, w_1 \qquad \text{and} \qquad y_2 = \sqrt{\frac{-2\ln(x_1)}{x_1}}\, w_2 \tag{12.9}$$

The procedure written in the hypothetical programming language can now be modified to accommodate the revised approach:

```
PROCEDURE normal
  REPEAT
    W1 = uniform(-1,+1)
    W2 = uniform(-1,+1)
    X1 = W1*W1+W2*W2
  UNTIL X1<1
  FACTOR = sqrt(-2*ln(X1)/X1)
  Y1 = FACTOR*W1
  Y2 = FACTOR*W2
  RETURN Y1
END
```

The first two assignment statements in the REPEAT–UNTIL loop give values to the random variables W_1 and W_2 but these values are each in the range -1 to $+1$. The coordinates (w_1, w_2) represent a point which is guaranteed to lie inside a 2×2 square centred on the origin but is not guaranteed to lie inside the unit circle.

A preliminary value $w_1^2 + w_2^2$ is assigned to the derived random variable X_1; this is the square of the distance from the origin. This value is acceptable if it is less than one. If not, the loop is repeated and new values are determined for W_1 and W_2 and the derived random variable X_1.

The value assigned to the identifier FACTOR is the value of the factor common to both expressions in (12.9). Multiplying this factor by w_1 and w_2 respectively provides values for the derived random variables Y_1 and Y_2 which are both distributed Normal(0,1).

Notice that no value is computed for the derived random variable X_2 since x_2 does not feature in the expressions in (12.9).

This procedure is more efficient that its predecessor and makes only a single call of sqrt and a single call of ln and there are no calls of cos or sin. Nevertheless, the procedure still makes no use of Y2. A little extra programming could save the value of Y2 for the next call of the procedure.

Another improvement would be to enhance the procedure so that it had two arguments MEAN and STDEV and returned a value which is distributed Normal(MEAN,STDEV^2) instead of Normal(0,1).

Nervous readers might be alarmed at what appears to be a negative argument for the sqrt function. Remember that x_1 is in the range 0 to 1 so $\ln(x_1)$ is guaranteed to be negative which ensures that $-2\ln(x_1)$ is positive.

There is a more serious cause for concern in that x_1, the argument of ln, could in principle be zero. This possibility can be trapped by modifying the condition after UNTIL to 0<X1<1 so that x_1 has to be strictly greater than zero.