

## PROBABILITY DISTRIBUTIONS: (continued)

**The change of variables technique.** Let  $x \sim f(x)$  and let  $y = y(x)$  be a monotonic transformation of  $x$  such that  $x = x(y)$  exists.

Let  $A$  be an event defined in terms of  $x$ , and let  $B$  be the equivalent event defined in terms of  $y$  such that if  $x \in A$ , then  $y = y(x) \in B$  and *vice versa*.

Then,  $P(A) = P(B)$  and we can find the the p.d.f of  $y$  denoted by  $g(y)$ .

**The continuous case.** If  $x, y$  are continuous random variables, then

$$\int_{y \in B} g(y) dy = \int_{x \in A} f(x) dx.$$

If we write  $x = x(y)$  in the second integral, then the change of variable technique gives

$$\int_{y \in B} g(y) dy = \int_{x \in B} f\{x(y)\} \frac{dx}{dy} dy.$$

If  $y = y(x)$  is a monotonically decreasing transformation, then  $dx/dy < 0$ , and  $f\{x(y)\} > 0$ , and so  $f\{x(y)\} dx/dy < 0$  cannot represent a p.d.f since  $g(y) \geq 0$  is necessary.

The recourse is to change the sign on the  $y$ -axis. When  $dx/dy < 0$ , it is replaced by its modulus  $|dx/dy| > 0$ . In general,

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

**The Standard Normal Distribution.** Consider the function  $g(z) = e^{-z^2/2}$ , where  $-\infty < z < \infty$ . There is  $g(z) > 0$  for all  $z$  and also

$$\int e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}}.$$

It follows that

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

constitutes a p.d.f., described as the standard normal and denoted by  $N(z; 0, 1)$ . In general, the normal distribution is denoted by  $N(x; \mu, \sigma^2)$ ; so, in this case, there are  $\mu = 0$  and  $\sigma^2 = 1$ .

**The General Normal Distribution.** This can be derived via the change of variables technique. Let

$$z \sim N(0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = f(z),$$

and let

$$y = z\sigma + \mu \quad \text{so that} \quad z = (y - \mu)/\sigma$$

is the inverse function and  $dz/dy = \sigma^{-1}$  is its derivative. Then,

$$\begin{aligned} g(y) = f\{z(y)\} \left| \frac{dz}{dy} \right| &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{y - \mu}{\sigma} \right)^2 \right\} \frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \frac{(y - \mu)^2}{\sigma^2} \right\}. \end{aligned}$$

**The discrete case.** Matters are simpler when  $x \sim f(x)$  is discrete and  $y = y(x)$  is a monotonic transformation. Then, if  $y \sim g(y)$ , it must be the case that

$$\sum_{y \in B} g(y) = \sum_{x \in A} f(x) = \sum_{y \in B} f\{x(y)\},$$

whence  $g(y) = f\{x(y)\}$ .

**Example.** Let

$$x \sim b(n = 3, p = 2/3) = \frac{3!}{x!(3-x)!} \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{3-x},$$

with  $x = 0, 1, 2, 3$ , and let  $y = x^2$  so that  $x(y) = \sqrt{y}$ . Then

$$y \sim g(y) = \frac{3!}{(\sqrt{y})!(3-\sqrt{y})!} \left(\frac{2}{3}\right)^{\sqrt{y}} \left(\frac{1}{3}\right)^{3-\sqrt{y}},$$

where  $y = 0, 1, 4, 9$ .

**Example.** Let  $x \sim f(x) = 2x; 0 \leq x \leq 1$ . Let  $y = 8x^3$ , which implies that  $x = (y/8)^{1/3} = y^{1/3}/2$  and  $dx/dy = y^{-2/3}/6$ . Then,

$$g(y) = f\{x(y)\} \left| \frac{dx}{dy} \right| = 2 \left( \frac{y^{1/3}}{2} \right) \left| \frac{y^{-2/3}}{6} \right| = \frac{y^{-1/3}}{6}.$$

**Expectations of a random variable.** If  $x \sim f(x)$ , then the expected value of  $x$  is

$$E(x) = \int_x x f(x) dx \quad \text{if } x \text{ is continuous, and}$$

$$E(x) = \sum_x x f(x) \quad \text{if } x \text{ is discrete.}$$

**The expectations operator.** We can economise on notation by defining the expectations operator  $E$ , which is subject to a number of simple rules. They are as follows:

(a) If  $x \geq 0$ , then  $E(x) \geq 0$ .

(b) If  $a$  is a constant, then  $E(a) = a$ .

(c) If  $a$  is a constant and  $x$  is a random variable, then

$$E(ax) = aE(x).$$

(d) If  $x, y$  are random variables, then

$$E(x + y) = E(x) + E(y).$$

(e) The expectation of a sum is the sum of the expectations. By combining (c) and (d), we get:

$$\text{If } E(ax + by) = aE(x) + bE(y).$$

Thus, the expectation operator is a linear operator and

$$E(\sum_i a_i x_i) = \sum_i a_i E(x_i).$$

**Example.** The expected value of the binomial distribution  $b(x; n, p)$  is

$$E(x) = \sum_x x b(x; n, p) = \sum_{x=0}^n x \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}.$$

We factorise  $np$  from the expression under the summation and we begin the summation at  $x = 1$ . On defining  $y = x - 1$ , which means setting  $x = y + 1$  in the expression above, we get

$$\begin{aligned} E(x) &= \sum_{x=0}^n x b(x; n, p) \\ &= np \sum_{y=0}^{n-1} \frac{(n-1)!}{([n-1]-y)!y!} p^y (1-p)^{[n-1]-y} \\ &= np \sum_{y=0}^{n-1} b(y; n-1, p) = np, \end{aligned}$$

where the final equality follows from the fact that we are summing the values of the binomial distribution  $b(y; n-1, p)$  over its entire domain to obtain a total of unity.

**Example.** Let  $x \sim f(x) = e^{-x}; 0 \leq x < \infty$ . Then

$$E(x) = \int_0^{\infty} x e^{-x} dx.$$

This must be evaluated by integrating by parts:

$$\int u \frac{dv}{dx} dx = uv - \int v \frac{du}{dx} dx.$$

With  $u = x$  and  $dv/dx = e^{-x}$ , this formula gives

$$\begin{aligned} \int_0^{\infty} x e^{-x} dx &= [-x e^{-x}]_0^{\infty} + \int_0^{\infty} e^{-x} dx \\ &= [-x e^{-x}]_0^{\infty} - [e^{-x}]_0^{\infty} = 0 + 1 = 1. \end{aligned}$$

Observe that, since this integral is unity and since  $x e^{-x} > 0$  over the domain of  $x$ , it follows that  $f(x) = x e^{-x}$  is also a valid p.d.f.

**Expectation of a function of random variable.** Let  $y = y(x)$  be a function of  $x \sim f(x)$ . The value of  $E(y)$  can be found without first determining the p.d.f  $g(y)$  of  $y$ . Quite simply, there is

$$E(y) = \int_x y(x)f(x)dx.$$

If  $y = y(x)$  is a monotonic transformation of  $x$ , then it follows that

$$\begin{aligned} E(y) &= \int_y yg(y)dy = \int_y yf\{x(y)\} \left| \frac{dx}{dy} \right| dy \\ &= \int_x y(x)f(x)dx, \end{aligned}$$

which establishes a special case of the result. However, the result is not confined to monotonic transformations. It is equally valid for functions that are piecewise monotonic, i.e. ones that have monotonic segments, which includes virtually all of the functions that we might consider.

**The moments of a distribution.** The  $r$ th raw moment of  $x$  relative to the datum  $a$  is defined by the expectation

$$E\{(x - a)^r\} = \int_x (x - a)^r f(x) dx \quad \text{if } x \text{ is continuous, and}$$

$$E\{(x - a)^r\} = \sum_x (x - a)^r f(x) \quad \text{if } x \text{ is discrete.}$$

The variance, which is a measure of the dispersion of  $x$  is the second moment of about the mean. This measure is minimised by the choice of  $a = E(x)$ :

$$\begin{aligned} V(x) &= E[\{x - E(x)\}^2] = E[x^2 - 2xE(x) + \{E(x)\}^2] \\ &= E(x^2) - \{E(x)\}^2. \end{aligned}$$

We can also define the variance operator  $V$ .

- (a) If  $x$  is a random variable, then  $V(x) > 0$ .
- (b) If  $a$  is a constant, then  $V(a) = 0$ .
- (c) If  $a$  is a constant and  $x$  is a random variable, then  $V(ax) = a^2V(x)$ .

To confirm the latter, we may consider

$$\begin{aligned} V(ax) &= E\{[ax - E(ax)]^2\} \\ &= a^2E\{[x - E(x)]^2\} = a^2V(x). \end{aligned}$$

If  $x, y$  are independently distributed random variables, then  $V(x + y) = V(x) + V(y)$ . But this is not true in general.



**The variance of the binomial distribution.** Consider a sequence of Bernoulli trials with  $x_i \in \{0, 1\}$  for all  $i$ . The p.d.f of the generic trial is

$$f(x_i) = p^{x_i}(1 - p)^{1-x_i}.$$

Then

$$E(x_i) = \sum_{x_i} x_i f(x_i) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

It follows that, in  $n$  trials, the expected value of the total score  $x = \sum_i x_i$  is

$$E(x) = \sum_i E(x_i) = np.$$

This is the expected value of the binomial distribution.

To find the variance of the Bernoulli trial, we use the formula  $E(x) = E(x^2) - \{E(x)\}^2$ . For a single trial, there is

$$E(x_i^2) = \sum_{x_i=0,1} f(x_i) = p,$$

$$V(x_i) = p - p^2 = p(1 - p) = pq, \quad \text{where} \quad q = 1 - p.$$

The outcome of the binomial random variable is the sum of a set of  $n$  independent and identical Bernoulli trials. Thus, the variance of the sum is the sum of the variances, and we have

$$V(x) = \sum_{i=1}^n V(x_i) = npq.$$