

# Statistics Background

L. R. Schaeffer, March 1999

## Random Variables

A *random variable* is a real-valued function which exists within the domain of a defined sample space. A random variable is designated by a capital letter, say  $Y$ , and the value of  $Y$ , depending on the outcome of the experiment, is denoted by a small letter, say  $y$ . The sample space is the range of values that the value of  $Y$  can be assigned.

Random variables can be either discrete or continuous. A discrete random variable can assume only a finite number of distinct values, such as zero or one for example. A continuous random variable can assume any value within the range of the sample space.

## Discrete Random Variables

In the discrete case, the probability that  $Y$  takes the value  $y$ , is defined as the sum of the probabilities of all sample points that are assigned the value  $y$ . That is,

$$P(Y = y) = p(y).$$

The *probability distribution* of  $Y$  lists the probabilities for each value of  $y$ . Suppose  $Y$  can take on four values with the following probabilities:

$y$	$p(y)$
0	1/8
1	1/4
2	1/4
3	3/8

Any other values of  $y$  are assumed to have  $p(y) = 0$ , and the sum of the probabilities is 1. The expected value of a discrete random variable is defined as

$$E(Y) = \sum_y y p(y).$$

For the example above,

$$E(Y) = (0(1/8) + 1(1/4) + 2(1/4) + 3(3/8)) = 1.875.$$

Similarly, the expected value of a function of  $Y$ , say  $g(Y)$  is given by

$$E(g(Y)) = \sum_y g(y) p(y).$$

Suppose  $g(Y) = Y^2$ , then

$$E(Y^2) = (0(1/8) + 1(1/4) + 4(1/4) + 9(3/8)) = 4.625.$$

The variance of discrete random variable  $Y$  is

$$Var(Y) = E[(Y - E(Y))^2] = E(Y^2) - [E(Y)]^2.$$

For the example,

$$\begin{aligned} Var(Y) &= (-1.875)^2(1/8) + (-.875)^2(1/4) + (.125)^2(1/4) + (1.125)^2(3/8) \\ &= 4.625 - (1.875)^2 \\ &= 1.109375 \end{aligned}$$

## Binomial Distribution

A common discrete distribution is the *binomial* distribution. A binomial event can take on only two possible outcomes, success or failure, zero or one, heads or tails, diseased or not diseased, and so on. The probability of one outcome is  $q$  and the probability of the other outcome is  $1 - q$ . Trials, or a succession of binomial events, are assumed to be independent. The random variable  $Y$  is the number of successes. The probability distribution is given by

$$p(y) = \binom{n}{y} q^y (1 - q)^{n-y},$$

for  $y = 0, 1, 2, \dots, n$  and  $0 \leq q \leq 1$ . The number of trials is  $n$ . The expected value and variance of the binomial distribution are

$$\begin{aligned} E(Y) &= n q \\ Var(Y) &= n q (1 - q). \end{aligned}$$

## Poisson Distribution

A Poisson probability distribution provides a good model for the probability distribution of the number  $Y$  of rare events that occur in a given space, time, volume, or any other dimension, and  $\lambda$  is the average value of  $Y$ . An example in animal breeding might be the number of quality embryos produced by a cow during superovulation, which can range from 0 to 20 (or more). The Poisson probability distribution is given by

$$p(y) = \frac{\lambda^y}{y!} \exp -\lambda,$$

for  $y = 0, 1, 2, \dots$  and  $\lambda > 0$ . Also,

$$\begin{aligned} E(Y) &= \lambda \\ Var(Y) &= \lambda. \end{aligned}$$

# General Results

## Expectations

If  $Y$  represents a random variable from some defined population, then the expectation of  $Y$  is denoted by

$$E(Y) = \mu$$

where  $E(\cdot)$  means *expected value*. Now let  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  represent vectors of random variables (or random vector variables), let  $k$  represent a scalar constant, and let  $\mathbf{K}$  be a matrix of constants, then

1.  $E(kY) = k E(Y) = k\mu.$

2.  $E(\mathbf{Y}_1) = \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$

3.  $E(k \mathbf{Y}_1) = k \mu.$

4.  $E(\mathbf{K}\mathbf{Y}_1) = \mathbf{K}\mu.$

5.  $E(\mathbf{Y}_1 + \mathbf{Y}_2) = E(\mathbf{Y}_1) + E(\mathbf{Y}_2) = \mu_1 + \mu_2$ , and if  $\mu_1 = \mu_2$  then  $E(\mathbf{Y}_1 + \mathbf{Y}_2) = 2\mu.$

The mean of a population is also known as the *first moment* of the distribution. The exact form of the distribution for a random variable will determine the form of the estimator of the mean and of other parameters of the distribution.

## Variance-Covariance Matrices

The variance of a scalar random variable,  $Y$ , is defined as

$$Var(Y) = E(Y^2) - E(Y)E(Y) = E(Y - E(Y))^2$$

and is commonly represented as  $\sigma_Y^2$ . With two scalar random variables, say  $Y_1$  and  $Y_2$ , then the covariance between them is defined as

$$Cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$$

and is represented as  $\sigma_{Y_1Y_2}$ . These definitions can be extended to a vector of random variables,  $\mathbf{Y}$ , to give a *variance-covariance* matrix of order equal to the length of  $\mathbf{Y}$ .

$$\begin{aligned} Var(\mathbf{Y}) &= E(\mathbf{Y}\mathbf{Y}') - E(\mathbf{Y})E(\mathbf{Y}') \\ &= \begin{pmatrix} \sigma_{Y_1}^2 & \sigma_{Y_1Y_2} & \cdots & \sigma_{Y_1Y_n} \\ \sigma_{Y_1Y_2} & \sigma_{Y_2}^2 & \cdots & \sigma_{Y_2Y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{Y_1Y_n} & \sigma_{Y_2Y_n} & \cdots & \sigma_{Y_n}^2 \end{pmatrix} = \mathbf{V}. \end{aligned}$$

A variance-covariance (VCV) matrix of a random vector contains variances on the diagonals and covariances on the off-diagonals. A VCV matrix is square, symmetric and should always be positive definite or positive semi-definite. Another commonly used name for VCV matrix is *dispersion* matrix.

Let  $\mathbf{A}$  be a matrix of constants conformable for multiplication with the vector  $\mathbf{Y}$ , then

$$\begin{aligned} \text{Var}(\mathbf{AY}) &= E(\mathbf{A}\mathbf{Y}\mathbf{Y}'\mathbf{A}') - E(\mathbf{AY})E(\mathbf{Y}'\mathbf{A}') \\ &= \mathbf{A}E(\mathbf{Y}\mathbf{Y}')\mathbf{A}' - \mathbf{A}E(\mathbf{Y})E(\mathbf{Y}')\mathbf{A}' \\ &= \mathbf{A}(E(\mathbf{Y}\mathbf{Y}') - E(\mathbf{Y})E(\mathbf{Y}'))\mathbf{A}' \\ &= \mathbf{A}\text{Var}(\mathbf{Y})\mathbf{A}' = \mathbf{A}\mathbf{V}\mathbf{A}'. \end{aligned}$$

If we have two sets of functions of  $\mathbf{Y}$ , say  $\mathbf{AY}$  and  $\mathbf{BY}$ , then

$$\text{Cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{A}\mathbf{V}\mathbf{B}'.$$

Similarly, if we have two functions of different random vectors, say  $\mathbf{AY}$  and  $\mathbf{MZ}$ , and  $\text{Cov}(\mathbf{Y}, \mathbf{Z}) = \mathbf{C}$ , then

$$\text{Cov}(\mathbf{AY}, \mathbf{MZ}) = \mathbf{A}\mathbf{C}\mathbf{M}'.$$

## Continuous Distributions

Consider measuring the amount of milk given by a dairy cow at a particular milking. Even if a machine of perfect accuracy was used, the amount of milk would be a unique point on a continuum of possible values, such as 32.35769842... kg of milk. As such it is mathematically impossible to assign a nonzero probability to all of the infinite possible points in the continuum. Thus, a different method of describing a probability distribution of a continuous random variable must be used. The sum of the probabilities (if they could be assigned) through the continuum is still assumed to sum to 1. The *cumulative distribution function* of a random variable is

$$F(y) = P(Y \leq y),$$

for  $-\infty < y < \infty$ . As  $y$  approaches  $-\infty$ , then  $F(y)$  approaches 0. As  $y$  approaches  $\infty$ , then  $F(y)$  approaches 1. Thus,  $F(y)$  is said to be a nondecreasing function of  $y$ . If  $a < b$ , then  $F(a) < F(b)$ .

If  $F(y)$  is the cumulative distribution function of  $Y$ , then the *probability density function* of  $Y$  is given by

$$f(y) = \frac{dF(y)}{dy} = F'(y),$$

wherever the derivative exists. Always for  $f(y)$  being a probability density function,

$$\int_{-\infty}^{\infty} f(y) dy = 1.$$

Conversely,

$$F(y) = \int_{-\infty}^y f(t) dt.$$

The expected value of a continuous random variable  $Y$  is

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

provided that the integral exists. If  $g(Y)$  is a function of  $Y$ , then

$$E(g(Y)) = \int_{-\infty}^{\infty} g(y) f(y) dy$$

provided that the integral exists. Finally,

$$Var(Y) = E(Y^2) - [E(Y)]^2.$$

### The Uniform Distribution

The basis for the majority of random number generators is a uniform distribution. A random variable  $Y$  has a continuous uniform probability distribution on the interval  $(\theta_1, \theta_2)$  if and only if the density function of  $Y$  is

$$f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2 \\ 0, & \text{elsewhere.} \end{cases}$$

The parameters of the density function are  $\theta_1$  and  $\theta_2$ . Also,

$$\begin{aligned} E(Y) &= (\theta_1 + \theta_2)/2 \\ Var(Y) &= (\theta_2 - \theta_1)^2/12. \end{aligned}$$

### The Normal Distribution

A random variable  $Y$  has a normal probability distribution if and only if

$$f(y) = (\sigma(2\pi)^{.5})^{-1} \exp(-.5(y - \mu)^2 \sigma^{-2})$$

for  $-\infty < x < +\infty$ , where  $\sigma^2$  is the variance of  $Y$  and  $\mu$  is the expected value of  $Y$ .

For the random vector,  $\mathbf{Y}$ , the multivariate normal density function is

$$f(\mathbf{y}) = (2\pi)^{-.5n} |\mathbf{V}|^{-.5} \exp(-.5(\mathbf{y} - \boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}))$$

denoted as  $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$  where  $\mathbf{V}$  is the variance-covariance matrix of  $\mathbf{Y}$ . Note that the determinant of  $\mathbf{V}$  must be positive, otherwise the density function is undefined.

## Chi-Square Distribution.

- If  $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I})$ , then  $\mathbf{y}'\mathbf{y} \sim \chi_n^2$ , where  $\chi_n^2$  is a *central chi-square* distribution with  $n$  degrees of freedom and  $n$  is the length of the vector  $\mathbf{y}$ . The mean of the central chi-square distribution is  $n$ , and the variance is  $2n$ . The probability distribution function is

$$f(y) = (y)^{(n/2)-1} \exp -0.5y/[2^{0.5n}\Gamma(0.5n)],$$

for  $y^2 > 0$ .

- If  $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{I})$ , then  $\mathbf{y}'\mathbf{y} \sim \chi_{n,\lambda}^2$  where  $\lambda$  is the noncentrality parameter which is equal to  $.5\boldsymbol{\mu}'\boldsymbol{\mu}$ . The mean of a noncentral chi-square distribution is  $n + 2\lambda$  and the variance is  $2n + 8\lambda$ .
- If  $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ , then  $\mathbf{y}'\mathbf{Q}\mathbf{y}$  has a noncentral chi-square distribution only if  $\mathbf{Q}\mathbf{V}$  is idempotent, i.e.  $\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V} = \mathbf{Q}\mathbf{V}$ . The noncentrality parameter is  $\lambda = .5\boldsymbol{\mu}'\mathbf{Q}\mathbf{V}\mathbf{Q}\boldsymbol{\mu}$  and the mean and variance of the distribution are  $tr(\mathbf{Q}\mathbf{V}) + 2\lambda$  and  $2tr(\mathbf{Q}\mathbf{V}) + 8\lambda$ , respectively.
- If there are two quadratic forms of  $\mathbf{y}$ , say  $\mathbf{y}'\mathbf{Q}\mathbf{y}$  and  $\mathbf{y}'\mathbf{P}\mathbf{y}$ , and both quadratic forms have chi-square distributions, then the two quadratic forms are independent if  $\mathbf{Q}\mathbf{V}\mathbf{P} = \mathbf{0}$ . Independence of quadratic forms is necessary for the construction of valid tests of hypotheses.

The chi-square distribution is used in hypothesis testing, for example in contingency tables. It is also used as a prior distribution for variances in Bayesian analyses. Chi-square variables are components of the following two distributions.

### The $t$ -distribution.

The  $t$ -distribution is based on the ratio of two independent random variables. The first is from a univariate normal distribution, and the second is from a central chi-square distribution. Let  $x \sim N(0, 1)$  and  $u \sim \chi_n^2$  with  $x$  and  $u$  being independent, then

$$\frac{x}{(u/n)^{.5}} \sim t_n.$$

The mean of a  $t$ -distribution is the mean of the  $x$  variable, and the variance is  $n/(n - 2)$ , and  $n$  is the degrees of freedom of the distribution.

### The F-distribution.

The central F-distribution is based on the ratio of two independent central chi-square variables. Let  $u \sim \chi_n^2$  and  $w \sim \chi_m^2$  with  $u$  and  $w$  being independent, then

$$\frac{(u/n)}{(w/m)} \sim F_{n,m}.$$

The mean of the F-distribution is  $m/(m - 2)$  and the variance is

$$\frac{2m^2(n + m - 2)}{n(m - 2)^2(m - 4)}.$$

Tables of F-values have been constructed for various probability levels as criteria to test if the numerator chi-square variable has a noncentral chi-square distribution. If the calculated F-value is greater than the value in the tables, then  $u$  is implied to have a noncentral chi-square distribution, otherwise we assume that  $u$  has a central chi-square distribution.

The square of a  $t$ -distribution variable gives a variable that has an F-distribution with 1 and  $n$  degrees of freedom.

Noncentral F-distributions exist depending on whether the numerator or denominator variables have noncentral chi-square distributions. Tables for noncentral F-distributions generally do not exist because of the difficulty in predicting the noncentrality parameters. However, using random chi-square generators it is possible to numerically calculate an expected noncentral F value for specific situations. When both the numerator and denominator chi-square variables are from noncentral distributions, then their ratio follows a doubly noncentral F-distribution.

## Quadratic and Bilinear Forms

### Quadratic Forms

A quadratic form is a sum of squares of elements of a vector. The general form is  $\mathbf{y}'\mathbf{Q}\mathbf{y}$ , where  $\mathbf{y}$  is a vector of random variables, and  $\mathbf{Q}$  is a regulator matrix. The regulator matrix can take on various forms and values depending on the situation. Usually  $\mathbf{Q}$  is a symmetric matrix. Examples of different  $\mathbf{Q}$  matrices are as follows:

1.  $\mathbf{Q} = \mathbf{I}$ , then  $\mathbf{y}'\mathbf{Q}\mathbf{y} = \mathbf{y}'\mathbf{y}$  which is a total sum of squares of the elements in  $\mathbf{y}$ .
2.  $\mathbf{Q} = \mathbf{J}(1/n)$ , then  $\mathbf{y}'\mathbf{Q}\mathbf{y} = \mathbf{y}'\mathbf{J}\mathbf{y}(1/n)$  where  $n$  is the length of  $\mathbf{y}$ . Note that  $\mathbf{J} = \mathbf{1}\mathbf{1}'$ , so that  $\mathbf{y}'\mathbf{J}\mathbf{y} = (\mathbf{y}'\mathbf{1})(\mathbf{1}'\mathbf{y})$  and  $(\mathbf{1}'\mathbf{y})$  is the sum of the elements in  $\mathbf{y}$ .
3.  $\mathbf{Q} = (\mathbf{I} - \mathbf{J}(1/n))/(n - 1)$ , then  $\mathbf{y}'\mathbf{Q}\mathbf{y}$  gives the variance of the elements in  $\mathbf{y}$ .

The expected value of a quadratic form is

$$E(\mathbf{y}'\mathbf{Q}\mathbf{y}) = E(\text{tr}(\mathbf{y}'\mathbf{Q}\mathbf{y})) = E(\text{tr}(\mathbf{Q}\mathbf{y}\mathbf{y}')) = \text{tr}(\mathbf{Q}E(\mathbf{y}\mathbf{y}')).$$

However,

$$\text{Var}(\mathbf{y}) = E(\mathbf{y}\mathbf{y}') - E(\mathbf{y})E(\mathbf{y}')$$

so that

$$E(\mathbf{y}\mathbf{y}') = \text{Var}(\mathbf{y}) + E(\mathbf{y})E(\mathbf{y}'),$$

then

$$E(\mathbf{y}'\mathbf{Q}\mathbf{y}) = tr(\mathbf{Q}(Var(\mathbf{y}) + E(\mathbf{y})E(\mathbf{y}'))).$$

If we let  $Var(\mathbf{y}) = \mathbf{V}$  and  $E(\mathbf{y}) = \mu$ , then

$$\begin{aligned} E(\mathbf{y}'\mathbf{Q}\mathbf{y}) &= tr(\mathbf{Q}(\mathbf{V} + \mu\mu')) \\ &= tr(\mathbf{Q}\mathbf{V}) + tr(\mathbf{Q}\mu\mu') \\ &= tr(\mathbf{Q}\mathbf{V}) + \mu'\mathbf{Q}\mu. \end{aligned}$$

The expectation of a quadratic form does not depend on the distribution of  $\mathbf{y}$ . However, the variance of a quadratic form requires that  $\mathbf{y}$  follows a multivariate normal distribution. Without showing the derivation, the variance of a quadratic form, assuming  $\mathbf{y}$  has a multivariate normal distribution, is

$$Var(\mathbf{y}'\mathbf{Q}\mathbf{y}) = 2tr(\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V}) + 4\mu'\mathbf{Q}\mathbf{V}\mathbf{Q}\mu.$$

The quadratic form,  $\mathbf{y}'\mathbf{Q}\mathbf{y}$ , has a chi-square distribution if

$$tr(\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V}) = tr(\mathbf{Q}\mathbf{V}), \text{ and } \mu'\mathbf{Q}\mathbf{V}\mathbf{Q}\mu = \mu'\mathbf{Q}\mu,$$

or the single condition that  $\mathbf{Q}\mathbf{V}$  is idempotent. Then if

$$m = tr(\mathbf{Q}\mathbf{V}) \text{ and } \lambda = .5\mu'\mathbf{Q}\mu,$$

the expected value of  $\mathbf{y}'\mathbf{Q}\mathbf{y}$  is  $m + 2\lambda$  and the variance is  $2m + 8\lambda$ , which are the usual results for a noncentral chi-square variable.

The covariance between two quadratic forms, say  $\mathbf{y}'\mathbf{Q}\mathbf{y}$  and  $\mathbf{y}'\mathbf{P}\mathbf{y}$ , is

$$Cov(\mathbf{y}'\mathbf{Q}\mathbf{y}, \mathbf{y}'\mathbf{P}\mathbf{y}) = 2tr(\mathbf{Q}\mathbf{V}\mathbf{P}\mathbf{V}) + 4\mu'\mathbf{Q}\mathbf{V}\mathbf{P}\mu.$$

The covariance is zero if  $\mathbf{Q}\mathbf{V}\mathbf{P} = \mathbf{0}$ , then the two quadratic forms are said to be independent.

## Bilinear Forms

A bilinear form is represented as  $\mathbf{x}'\mathbf{B}\mathbf{y}$ , where  $\mathbf{x}$  and  $\mathbf{y}$  are two different random vectors, possibly of different lengths, and  $\mathbf{B}$  is the regulator matrix. If  $E(\mathbf{x}) = \alpha$  and  $E(\mathbf{y}) = \mu$  with  $Cov(\mathbf{x}, \mathbf{y}) = \mathbf{C}$ , then

$$E(\mathbf{x}'\mathbf{B}\mathbf{y}) = tr(\mathbf{B}\mathbf{C}') + \alpha'\mathbf{B}\mu,$$

and if  $\mathbf{V} = Var(\mathbf{y})$  and  $\mathbf{T} = Var(\mathbf{x})$ , then

$$\begin{aligned} Var(\mathbf{x}'\mathbf{B}\mathbf{y}) &= tr(\mathbf{B}\mathbf{C}'\mathbf{B}\mathbf{C}') + tr(\mathbf{B}\mathbf{V}\mathbf{B}'\mathbf{T}) + \alpha'\mathbf{B}\mathbf{V}\mathbf{B}'\alpha \\ &\quad + \mu'\mathbf{B}'\mathbf{T}\mathbf{B}\mu + 2\alpha'\mathbf{B}\mathbf{C}'\mathbf{B}\mu. \end{aligned}$$

Usually, the lengths of  $\mathbf{x}$  and  $\mathbf{y}$  are equal and  $\mathbf{B}$  is symmetric. If  $\mathbf{B}$  is also positive definite, then  $\mathbf{x}'\mathbf{B}\mathbf{y}$  is a sum of cross-products. Bilinear forms occur in the estimation of covariances. Note that a bilinear form may be written as a quadratic form as follows: Let

$$\mathbf{z}' = \begin{pmatrix} \mathbf{x}' & \mathbf{y}' \end{pmatrix},$$

then

$$\mathbf{x}'\mathbf{B}\mathbf{y} = \begin{pmatrix} \mathbf{x}' & \mathbf{y}' \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix},$$

and

$$Var(\mathbf{z}) = \begin{pmatrix} \mathbf{T} & \mathbf{C} \\ \mathbf{C}' & \mathbf{V} \end{pmatrix}.$$