Appendix A: Some Basic Results in Probability and Statistics Subsections A.1-A.5

From Applied Linear Statistical Models

January 22, 2020

This appendix contains some basic results in probability and statistics. It is intended as a reference to which you may refer as you read the book.

1 Summation and Product Operators

Summation Operator

The summation operator $\sum$ is defined as follows:

(A.1) $$\sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \cdots + Y_n$$

Some important properties of this operator are:

(A.2a) $$\sum_{i=1}^{n} k = nk \quad \text{where } k \text{ is a constant}$$

(A.2b) $$\sum_{i=1}^{n} (Y_i + Z_i) = \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} Z_i$$

(A.2c) $$\sum_{i=1}^{n} (a + cY_i) = na + c\sum_{i=1}^{n} Y_i \quad \text{where } a \text{ and } c \text{ are constants}$$

The double summation operator $\sum\sum$ is defined as follows:

(A.3) $$\begin{aligned}\sum_{i=1}^{n}\sum_{k=1}^{m} Y_{ij} &= \sum_{i=1}^{n} (Y_{i1} + \cdots + Y_{im}) \\ &= Y_{11} + \cdots + Y_{1m} + Y_{21} + \cdots + Y_{2m} + \cdots + Y_{nm}\end{aligned}$$

An important property of the double summation operator is:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} Y_{ij} \tag{A.4}$$

Product Operator

The product operator $\prod$ is defined as follows:

$$\prod_{i=1}^{n} Y_i = Y_1 \cdot Y_2 \cdot Y_3 \cdots Y_n \tag{A.5}$$

2 Probability

Addition Theorem

Let A_i and A_j be two events defined on a sample space. Then:

$$P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j) \tag{A.6}$$

where $P(A_i \cup A_j)$ denotes the probability of either A_i or A_j or both occurring; $P(A_i)$ and $P(A_j)$ denote, respectively, the probability of A_i and the probability of A_j; and $P(A_i \cap A_j)$ denotes the probability of both A_i and A_j occurring.

Multiplication Theorem

Let $P(A_i|A_j)$ denote the conditional probability of A_i occurring, given that A_j has occurred, and let $P(A_j|A_i)$ denote the conditional probability of A_j occurring, given that A_i has occurred. These conditional probabilities are defined as follows:

$$P(A_i|A_j) = \frac{P(A_i \cap A_j)}{P(A_j)} \quad P(A_j) \neq 0 \tag{A.7a}$$

$$P(A_j|A_i) = \frac{P(A_i \cap A_j)}{P(A_i)} \quad P(A_i) \neq 0 \tag{A.7b}$$

The multiplication theorem states:

$$P(A_i \cap A_j) = P(A_i)\,P(A_j|A_i) = P(A_j)\,P(A_i|A_j) \tag{A.8}$$

Complementary Events

The complementary event of A_i is denoted by $\bar{A}_i$. The following results for complementary events are useful:

(A.9)
$$P\left(\bar{A}_i\right) = 1 - P\left(A_i\right)$$

(A.10)
$$P\left(\overline{A_i \cup A_j}\right) = P\left(\bar{A}_i \cap \bar{A}_j\right)$$

3 Random Variables

Throughout this section, expect as noted, we assume that the random variable Y assumes a finite number of outcomes.

Expected Value

Let the random variable Y assume the outcomes $Y_1, \ldots, Y_k$ with probabilities given by the probability function:

(A.11)
$$f\left(Y_s\right) = P\left(Y = Y_s\right) \quad s = 1, \ldots, k$$

The expected value of Y, denoted by $E\left\{Y\right\}$, is defined by:

(A.12)
$$E\left\{Y\right\} = \sum_{s=1}^{k} Y_s f\left(Y_s\right)$$

$E\{\ \}$ is called the *expectation operator*.

An important property of the expectation operator is:

(A.13)
$$E\left\{a + cY\right\} = a + cE\left\{Y\right\} \quad \text{where } a \text{ and } c \text{ are constants}$$

Special cases of this are:

(A.13a)
$$E\left\{a\right\} = a$$

(A.13b)
$$E\left\{cY\right\} = cE\left\{Y\right\}$$

(A.13c)
$$E\left\{a + Y\right\} = a + E\left\{Y\right\}$$

Note

If the random variable Y is continuous, with density function $f(Y)$, $E\{Y\}$ is defined as follows:

$$E\{Y\} = \int_{-\infty}^{\infty} Y f(Y)\, dY \tag{A.14}$$

Variance

The variance of the random variable Y is denoted by $\sigma^2\{Y\}$ and is defined as follows:

$$\sigma^2\{Y\} = E\left\{(Y - E\{Y\})^2\right\} \tag{A.15}$$

An equivalent expression is:

$$\sigma^2\{Y\} = E\left\{Y^2\right\} - (E\{Y\})^2 \tag{A.15a}$$

$\sigma^2\{\ \}$ is called the *variance operator*.

The variance of a linear function of Y is frequently encountered. We denote the variance of $a + c + Y$ by $\sigma^2\{a + cY\}$ and have:

$$\sigma^2\{a + cY\} = c^2\sigma^2\{Y\} \quad \text{where } a \text{ and } c \text{ are constants} \tag{A.16}$$

Special cases of this result are:

$$\sigma^2\{a + Y\} = \sigma^2\{Y\} \tag{A.16a}$$

$$\sigma^2\{cY\} = c^2\sigma^2\{Y\} \tag{A.16b}$$

Note

If Y is continuous, $\sigma^2\{Y\}$ is defined as follows:

$$\sigma^2\{Y\} = \int_{-\infty}^{\infty} (Y - E\{Y\})^2 f(Y)\, dY \tag{A.17}$$

Joint, Marginal, and Conditional Probability Distributions

Let the joint probability function for the two random variables Y and Z be denoted by $g(Y, Z)$:

(A.18) $$g(Y_s, Z_t) = P(Y = Y_s \cap Z = Z_t) \quad s = 1, \ldots, k; t = 1, \ldots, m$$

The marginal probability function of Y, denoted by $f(Y)$, is:

(A.19a) $$f(Y_s) = \sum_{t=1}^{m} g(Y_s, Z_t) \quad s = 1, \ldots, k$$

and the marginal probability function of Z, denoted by $h(Z)$, is:

(A.19b) $$h(Z_t) = \sum_{s=1}^{k} g(Y_s, Z_t) \quad t = 1, \ldots, m$$

The conditional probability function of Y, given $Z = Z_t$, is:

(A.20a) $$f(Y_s|Z_t) = \frac{g(Y_s, Z_t)}{h(Z_t)} \quad h(Z_t) \neq 0;\ s = 1, \ldots, k$$

and the conditional probability function of Z, given $Y = Y_s$, is:

(A.20b) $$h(Z_t|Y_s) = \frac{g(Y_s, Z_t)}{f(Y_s)} \quad f(Y_s) \neq 0;\ t = 1, \ldots, m$$

Covariance

The covariance of Y and Z is denoted by $\sigma\{Y, Z\}$ and is defined by:

(A.21) $$\sigma\{Y, Z\} = E\{(Y - E\{Y\})(Z - E\{Z\})\}$$

An equivalent expression is:

(A.21a) $$\sigma\{Y, Z\} = E\{YZ\} - (E\{Y\})(E\{Z\})$$

$\sigma\{\ ,\ \}$ is called the *covariance operator*.

The covariance of $a_1 + c_1Y$ and $a_2 + c_2Z$ is denoted by $\sigma\{a_1 + c_1Y, a_2 + c_2Z\}$, and we have:

(A.22) $$\sigma\{a_1 + c_1Y, a_2 + c_2Z\} = c_1c_2\sigma\{Y, Z\} \quad \text{where } a_1,\ a_2,\ c_1, \text{ and } c_2 \text{ are constants}$$

Special cases of this are:

(A.22a) $$\sigma\{c_1Y, c_2Z\} = c_1c_2\sigma\{Y, Z\}$$

(A.22b) $$\sigma\{a_1 + Y, a_2 + Z\} = \sigma\{Y, Z\}$$

By definition, we have:

(A.23) $$\sigma\{Y, Y\} = \sigma^2\{Y\}$$

where $\sigma^2\{Y\}$ is the variance of Y.

Coefficient of Correlation

The standardized form of a random variable Y, whose mean and variance are $E\{Y\}$ and $\sigma^2\{Y\}$, respectively, is as follows:

(A.24) $$Y' = \frac{Y - E\{Y\}}{\sigma\{Y\}}$$

where Y' denotes the *standardized random variable* form of the random variable Y.

The coefficient of correlation between random variables Y and Z, denoted by $\rho\{Y, Z\}$, is the covariance between the standardized variables Y' and Z':

(A.25) $$\rho\{Y, Z\} = \sigma\left\{Y', Z'\right\}$$

Equivalently, the coefficient of correlation can be expressed as follows:

(A.25a) $$\rho\{Y, Z\} = \frac{\sigma\{Y, Z\}}{\sigma\{Y\}\,\sigma\{Z\}}$$

$\rho\{\ ,\ \}$ is called the *correlation operator*.

The coefficient of correlation can take on values between -1 and 1:

(A.26) $$-1 \leq \rho\{Y, Z\} \leq 1$$

When $\sigma\{Y, Z\} = 0$, it follows from (A.25a) that $\rho\{Y, Z\} = 0$ and Y and Z are said to be uncorrelated.

Independent Random Variables

The independence of two discrete random variables is defined as follows:

(A.27) Random variables Y and Z are independent if and only if:
$$g(Y_s, Z_t) = f(Y_s)\,h(Z_t) \quad s = 1, \ldots, k;\ t = 1, \ldots, m$$

If Y and Z are independent variables:

(A.28) $$\sigma\{Y, Z\} = 0 \text{ and } \rho\{Y, Z\} = 0 \quad \text{when } Y \text{ and } Z \text{ are independent}$$

(In the special case where Y and Z are jointly normally distributed, $\sigma\{Y, Z\} = 0$ implies that Y and Z are independent.)

Functions of Random Variables

Let $Y_1, \ldots, Y_n$ be n random variables. Consider the function $\sum a_i Y_i$, where the a_i are constants. We then have:

(A.29a)
$$E\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n} a_i E\{Y_i\} \quad \text{where the } a_i \text{ are constants}$$

(A.29b)
$$\sigma^2\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma\{Y_i, Y_j\} \quad \text{where the } a_i \text{ are constants}$$

Specifically, we have for $n = 2$:

(A.30a)
$$E\{a_1 Y_1 + a_2 Y_2\} = a_1 E\{Y_1\} + a_2 E\{Y_2\}$$

(A.30b)
$$\sigma^2\{a_1 Y_1 + a_2 Y_2\} = a_1^2 \sigma^2\{Y_1\} + a_2^2 \sigma^2\{Y_2\} + 2 a_1 a_2 \sigma\{Y_1, Y_2\}$$

If the random variables Y_i are independent, we have:

(A.31)
$$\sigma^2\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n} a_i^2 \sigma^2\{Y_i\} \quad \text{when the } Y_i \text{ are independent}$$

Special cases of this are:

(A.31a)
$$\sigma^2\{Y_1 + Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} \quad \text{when the } Y_1 \text{ and } Y_2 \text{ are independent}$$

(A.31b)
$$\sigma^2\{Y_1 - Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} \quad \text{when the } Y_1 \text{ and } Y_2 \text{ are independent}$$

When the Y_i are independent random variables, the covariance of two linear functions $\sum a_i Y_i$ and $\sum c_i Y_i$ is:

(A.32)
$$\sigma\left\{\sum_{i=1}^{n} a_i Y_i, \sum_{i=1}^{n} c_i Y_i\right\} = \sum_{i=1}^{n} a_i c_i \sigma^2\{Y_i\} \quad \text{when the } Y_i \text{ are independent}$$

Central Limit Theorem

The central limit theorem is basic for much of statistical inference.

(A.33) If $Y_1, \ldots, Y_n$ are independent random observations from a population with probability function $f(Y)$ for which $\sigma^2\{Y\}$ is finite, the sample mean $\bar{Y}$:

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

is approximately normally distributed when the sample size n is reasonably large, with mean $E\{Y\}$ and variance $\sigma^2\{Y\}/n$.

4 Normal Probability Distribution and Related Distributions

Normal Probability Distribution

The density function for a normal random variable Y is:

(A.34) $$f(Y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{Y-\mu}{\sigma}\right)^2\right] \quad -\infty < Y < \infty$$

where μ and σ are the two parameters of the normal distribution and $\exp(a)$ denotes e^a. The mean and variance of a normal random variable of a normal random variable Y is:

(A.35a) $$E\{Y\} = \mu$$

(A.35b) $$\sigma^2\{Y\} = \sigma^2$$

Linear Function of a Normal Random Variable. A linear function of a normal random variable Y has the following property:

(A.36) If Y is a normal random variable, the transformed variable $Y' = a + cY$ (a and c are constants) is normally distributed, with mean $a + cE\{Y\}$ and variance $c^2\sigma^2\{Y\}$

Standard Normal Random Variable. The standard normal random variable z:

(A.37) $$z = \frac{Y-\mu}{\sigma} \quad \text{where } Y \text{ is a normal random variable}$$

is normally distributed, with mean 0 and variance 1. We denote this as follows:

$$z \text{ is } N(0,1)$$

Table B.1 in Appendix B contains the cumulative probabilities A for percentiles $z(A)$ where:

(A.39) $$P\{z \leq z(A)\} = A$$

For instance, when $z(A) = 2.00$, $A = 0.9772$. Because the normal distribution is symmetrical about 0, when $z(A) = -2.00$, $A = 1 - 0.9772 = 0.0228$.

Linear Combinations of Independent Normal Random Variables. Let $Y_1, \ldots, Y_n$ be independent normal random variables. We then have:

(A.40) When $Y_1, \ldots, Y_n$ are independent normal random variables, the linear combination $a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$ is normally distributed, with mean $\sum a_i E\{Y_i\}$ and variance $\sum a_i^2 \sigma^2 \{Y_i\}$.

χ^2 Distribution

Let $z_1, \ldots, z_\nu$ be ν independent standard normal random variables. We then define a chi-square random variable as follows:

(A.41) $$\chi^2(\nu) = z_1^2 + z_2^2 + \cdots + z_\nu^2 \quad \text{where the } z_i \text{ are independent}$$

The χ^2 distribution has one parameter, ν, which is called the *degrees of freedom (df)*. The mean of the χ^2 distribution with ν degrees of freedom is:

(A.42) $$E\left\{\chi^2(\nu)\right\} = \nu$$

Table B.3 in Appendix B contains percentiles of various χ^2 distributions. We define $\chi^2(A;\nu)$ as follows:

(A.43) $$P\left\{\chi^2(\nu) \leq \chi^2(A;\nu)\right\} = A$$

Suppose $\nu = 5$. The 90th percentile of the χ^2 distribution with 5 degrees of freedom is $\chi^2(0.90; 5) = 9.24$.

t Distribution

Let z and $\chi^2(\nu)$ be independent random variables (standard normal and χ^2, respectively). We then define a t random variable as follows:

(A.44) $$t(\nu) = \frac{z}{\left[\frac{\chi^2(\nu)}{\nu}\right]^{1/2}} \quad \text{where } z \text{ and } \chi^2(\nu) \text{ are independent}$$

The t distribution has one parameter, the *degrees of freedom* ν. The mean of the t-distribution with ν degrees of freedom is:

$$E\{t(\nu)\} = 0 \tag{A.45}$$

Table B.2 in Appendix B contains percentiles of various t distributions. We define $t(A;\nu)$ as follows:

$$P\{t(\nu) \leq t(A;\nu)\} = A \tag{A.46}$$

Suppose $\nu = 10$. The 90th percentile of the t distribution with 10 degrees of freedom is $t(0.90;10) = 1.372$. Because the t distribution is symmetrical about 0, we have $t(0.10;10) = -1.372$.

F Distribution

Let $\chi^2(\nu_1)$ and $\chi^2(\nu_2)$ be two independent χ^2 random variables. We then define an F random variable as follows:

$$F(\nu_1,\nu_2) = \frac{\chi^2(\nu_1)}{\nu_1} \div \frac{\chi^2(\nu_2)}{\nu_2} \quad \text{where } \chi^2(\nu_1) \text{ and } \chi^2(\nu_2) \text{ are independent} \tag{A.47}$$

The F distribution has two parameters, the *numerator degrees of freedom* and the *denominator degrees of freedom*, here ν_1 and ν_2, respectively.

Table B.4 in Appendix B contains percentiles of various F distributions. We define $F(A;\nu_1,\nu_2)$ as follows:

$$P\{F(\nu_1,\nu_2) \leq F(A;\nu_1,\nu_2)\} = A \tag{A.48}$$

Suppose $\nu_1 = 2$ and $\nu_2 = 3$. The 90th percentile of the F distribution with 2 and 3 degrees of freedom, respectively, in the numerator and denominator is $F(0.90;2,3) = 5.46$.

Percentiles below 50 percent can be obtained by utilizing the relation:

$$F(A;\nu_1,\nu_2) = \frac{1}{F(1-A;\nu_2,\nu_1)} \tag{A.49}$$

Thus, $F(0.10;3,2) = 1/F(0.90;2,3) = 1/5.46 = 0.183$.

The following relation exists between the t and F random variables:

$$\left[t(\nu)^2\right] = F(1,\nu) \tag{A.50a}$$

and the percentiles of the t and F distributions are related as follows:

$$[t(0.5+A/2;\nu)]^2 = F(A;1,\nu) \tag{A.50b}$$

Note

Throughout this text, we consider $z(A)$, $\chi^2(A;\nu)$, $t(A,\nu)$, and $F(A;\nu_1,\nu_2)$ as $A(100)$ percentiles. Equivalently, they can be considered as A fractiles.

5 Statistical Estimation

Properties of Estimators

Four important properties of estimators are as follows:

(A.51) An estimator $\hat{\theta}$ of the parameter θ is *unbiased* if:

$$E\left\{\hat{\theta}\right\} = \theta$$

(A.52) An estimator $\hat{\theta}$ is a *consistent estimator* of θ if:

$$\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| > \epsilon\right) = 0 \quad \text{for any } \epsilon > 0$$

(A.53) An estimator $\hat{\theta}$ is a *sufficient estimator* of θ if the conditional joint probability function of the sample observations, given $\hat{\theta}$, does not depend on the parameter θ.

(A.54) An estimator $\hat{\theta}$ is a *minimum variance estimator* of θ if for any other estimator $\hat{\theta}^*$

$$\sigma^2\left\{\hat{\theta}\right\} \le \sigma^2\left\{\hat{\theta}^*\right\} \quad \text{for all } \hat{\theta}^*$$

Maximum Likelihood Estimators

The method of maximum likelihood is a general method of finding estimators. Suppose we are sampling a population whose probability function $f(Y;\theta)$ involves one parameter, θ. Given independent observations $Y_1, \ldots, Y_n$, the joint probability function of the sample observations is:

(A.55a)
$$g(Y_1, \ldots, Y_n) = \prod_{i=1}^{n} f(Y_i; \theta)$$

When this joint probability function is viewed as a function of θ, with the observations given, it is called the *likelihood function* $L(\theta)$:

(A.55b)
$$L(\theta) = \prod_{i=1}^{n} f(Y_i; \theta)$$

Maximizing $L(\theta)$ with respect to θ yields the maximum likelihood estimator of θ. Under quite general conditions, maximum likelihood estimators are consistent and sufficient.

Least Squares Estimators

The method of least squares is another general method of finding estimators. The sample observations are assumed to be of the form (for the case of a single parameter θ):

(A.56)
$$Y_i = f_i(\theta) + \epsilon_i \quad i = 1, \ldots, n$$

where $f_i(\theta)$ is a known function of the parameter θ and the ϵ_i are random variables, usually assumed to have expectation $\mathbb{E}[\epsilon_i] = 0$.

With the method of least squares, for the given sample observations, the sum of squares:

(A.57)
$$Q = \sum_{i=1}^{n} [Y_i - f_i(\theta)]^2$$

is considered a function of θ. The least squares estimator of θ is obtained by minimizing Q with respect to θ. In many instances, least squares estimators are unbiased and consistent.