ECOM40006/ECOM90013 Econometrics 3
Department of Economics
University of Melbourne

The Normal and Related Distributions

Semester 1, 2025

Version: March 14, 2025

Contents

Abstract

These discuss about the Normal distribution and some distributions related to it. They are by no means exhaustive, although they probably far exceed the needs of most readers. They are provided as a resource that may be of use beyond the needs of this subject. I hope that you find them useful.

1 Introduction

Most books on statistics and econometrics will provide some level of treatment on probability theory and probability distributions. Some do nothing else, for example, Johnson, Kotz, and Kemp (2005), Johnson, Kotz, and Balakrishnan (1994, 1995, 1997, 2000) and Gupta and Nagar (2000). Others introduce probabilistic concepts on the way to other things, see inter alios Mood, Graybill, and Boes (1974), Wackerly, Mendenhall, and Scheaffer (2008), Stuart and Ord (1987), Rao (1973) and Muirhead (1982).[1] Easily the single most important distribution that we will encounter is the Normal distribution. The Normal distribution is term used to describe a family of continuous distributions that arise naturally in various aspects of probability theory. They, in turn, are special cases of a number of different families of probability distributions, such as the Pearson and exponential families. They also arise naturally in asymptotic arguments using central limit theorems but more on that at a later time. Suffice to say that Normal distributions are important and it is helpful to have a good understanding of how they work.

The structure of what follows is that we shall start with the simplest univariate cases and then extend our analysis to allow for bivariate and then more general multivariate cases. We shall then turn our attention to the distributions of some commonly encountered functions of Normal random variables, before returning to very briefly introduce

[1]The treatments in statistics books tend to be more complete and more useful as a consequence, which is why they dominate the references given here.

the sufficient statistics for Normal distributions, which are important in the theory of statistical inference. Some more advanced ideas are then explored in various appendices. Finally, Appendix D offers various tables related to the main distributions explored in these notes.

2 The Univariate Normal Distribution

2.1 The Standard Normal Distribution and Its Properties

Let Z denote a random variable with a standard Normal distribution, which is often written compactly as $Z \sim \mathrm{N}(0, 1)$. This statement tells us that Z has a continuous distribution function and that the corresponding density function is given by

$$\phi_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}. \tag{1}$$

Graphically, this density function is represented by the darker curve in Figure 1, the one labelled $\sigma^2 = 1$. The distribution function, sadly, has no closed form expression and so we can write no more than[2]

$$\Phi_Z(c) = \Pr(Z \leq c) = \int_{-\infty}^{c} \phi_Z(z)\, \mathrm{d}z, \quad \text{for fixed } c. \tag{2}$$

That said, even though we have no closed form expression for this function, much is known about it as it occurs in many different branches of mathematics where is is known as an error function (or the Gauss error function).[3] Note that the symbols Z (for the random variable), $\Phi_Z(\cdot)$ (for the distribution function evaluated at some point), and $\phi_Z(z) \equiv \mathrm{d}\Phi_Z(c)/\,\mathrm{d}c$ (for the density function evaluated at some point c) are commonly used in the context of the standard Normal distribution.

Although it is difficult ex ante to see why this particular expression should be a function of particular interest, the fact that it is one of those naturally occurring functions makes it so. It has been extensively studied over the years, dating back at least as far as the times of Gauss and Laplace,[4] Let us look at some of its key properties.

- The density function is symmetric about zero. That is, for all $z \geq 0$, $f(z) = f(-z)$. This means that equal amounts of their probability content must lie to either side of this point of symmetry and so the median of the distribution must also be zero.
- Both the distribution and density functions have infinite support. That is, they are well-defined for all real values of the their arguments.
- The density function is unimodal.

[2]Observe that because Z is a continuous random variable, $\Pr(Z = c) = 0$ and so $\Pr(Z \leq c) = \Pr(Z < c)$, meaning that either of these two expressions might equally have been used in (2).

[3]Carl Friedrich Gauss (1777–1855) is incredibly important in the history of the Normal distribution, and indeed in many other areas of mathematics, which is sometimes referred to as the Gauss or Gaussian distribution.

[4]Pierre-Simon Laplace (1749–1827) made many contributions to mathematics and, in particular, to the theory of probability. Amongst his claims to fame, he is credited with having developed the Bayesian interpretation of probability and of having established an early central limit theorem.

- The distribution function is differentiable $\Phi_Z(c)$ is differentiable at all real numbers c, which is a necessary requirement for the density function $\phi_Z(z)$ to be differentiable at all real values of z. Moreover, the density function is infinitely differentiable, with the nth derivative being of the form

$$\phi_z(z)^{(n)} = (-1)^n \mathrm{He}_n(z)\phi_z(z),$$

  where $\mathrm{He}_n(z)$ denotes the nth Hermite polynomial.

- The area under the density function must integrate to unity. This must be true for any proper density function and comes about through appropriate choice of normalizing constant. Specifically, any function that is non-negative over its support can be used as a density function. It becomes a *proper* density function when it is scaled to integrate/sum to unity. Let us illustrate this in the context of the standard Normal distribution. The kernel of the density function is the that part which is actually a function of the random variable. The *normalizing constant* is everything else and is there for the sole purpose of ensuring that the density integrates to unity. Thus the kernel of the standard Normal density function is given by $\exp\left\{-\frac{1}{2}z^2\right\}$ and the normalizing constant is $(2\pi)^{-1}$ (which we establish below). The support of the random variable is $\{-\infty, \infty\}$, so let's see what happens when we integrate the kernel with respect to z over that interval. The relevant (improper) integral is

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z = 2\int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z,$$

  where the right-hand side follows from the symmetry of the density function.[5] From here there is a bunch of things that one might do. As we will need this later on, we

[5] Alternatively, if you don't find that argument compelling, one might write

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z = \int_{-\infty}^{0} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z + \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z.$$

Change variable from z to $-w$, and write $\mathrm{d}z/\mathrm{d}w = -1$ so that $\mathrm{d}z = -\mathrm{d}w$. Noting that when $z = -\infty$, $w = \infty$ and when $z = 0$, $w = 0$ we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z &= \int_{\infty}^{0} \exp\left\{-\frac{(-w)^2}{2}\right\} (-\,\mathrm{d}w) + \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z \\ &= -\int_{\infty}^{0} \exp\left\{-\frac{w^2}{2}\right\} \,\mathrm{d}w + \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z.\end{aligned}$$

Reversing the direction of integration, which changes the sign of the value of the integral, we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z &= -\left(-\int_{0}^{\infty} \exp\left\{-\frac{w^2}{2}\right\} \,\mathrm{d}w\right) + \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z \\ &= \int_{0}^{\infty} \exp\left\{-\frac{w^2}{2}\right\} \,\mathrm{d}w + \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z.\end{aligned}$$

The final result follows on noting that the value of the integral is no longer a function of the variable that the integrand is integrated with respect to. That is, the value of the integral with respect to w is not a function of w, nor is the value of the integrand integrated with respect to z a function of z. Consequently, we can use any symbol on that role that we wish, as it will have absolutely no impact on the value of the integral. Let us replace the symbol w by z to obtain

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z = \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z + \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z = 2\int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z,$$

shall proceed as follows. Make the change of variables $t = \frac{1}{2}z^2$, or $z = \sqrt{2t}$, so now

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\mathrm{d}z}{\mathrm{d}(2t)} \times \frac{\mathrm{d}(2t)}{\mathrm{d}t} = \frac{1}{2\sqrt{2t}} \times 2 = \frac{1}{\sqrt{2t}}.$$

Moreover, when $z = 0$, $t = 0$ and when $z \to \infty$, $t \to \infty$. Making these substitutions we see that

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = 2\int_0^{\infty} e^{-t}\frac{1}{\sqrt{2t}}\,\mathrm{d}t = \sqrt{2}\int_0^{\infty} e^{-t}t^{-1/2}\,\mathrm{d}t.$$

The reason for this substitution is that the final integral is something extremely well-known, namely the integral definition of the Gamma function.[6] Looking to the definition of the Gamma function in Footnote 6, we see that

$$\int_0^{\infty} e^{-t}t^{-1/2}\,\mathrm{d}t = \int_0^{\infty} e^{-t}t^{1/2-1}\,\mathrm{d}t = \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

and so we have established that

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = \sqrt{2\pi},$$

---

as required. Note that the change of symbol, which changes nothing of importance, is a different operation to the change of variable, which is a substantive change to the integrand the thing being integrated). Clearly, the symmetry argument was easier!

[6]Originally studied by the great Swiss mathematician Daniel Bernoulli (1700–1782), it is also known as the Euler integral of the second kind, after another great Swiss mathematician Leonhard Euler (1707–1783). Specifically, we use the notation

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t}t^{\alpha-1}\,\mathrm{d}t, \quad \alpha > 0.$$

The definition can be extended to include non-integer negative values of α but that is unimportant to us. All that you need to know about Gamma functions is (i) that $\Gamma(1) = 1$, (ii) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, and (iii) that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, so that if α is an integer then

$$\Gamma(\alpha+1) = \alpha \times (\alpha-1) \times (\alpha-2) \times \ldots \times 2 \times 1 = \alpha!,$$

which is why Gamma functions are sometimes referred to a generalized factorial functions, and

$$\Gamma\left(\alpha+\frac{1}{2}\right) = \frac{(2\alpha-1)}{2} \times \frac{(2\alpha-3)}{2} \times \frac{(2\alpha-5)}{2} \times \ldots \times \frac{1}{2} \times \sqrt{\pi}.$$

For example, on setting $\alpha = 4$, we have $\Gamma(4) = 3! = 3 \times 2 \times 1 = 6$ and

$$\Gamma\left(\frac{9}{2}\right) = \frac{7}{2} \times \frac{5}{2} \times \frac{3}{2} \times \ldots \times \frac{1}{2} \times \Gamma\left(\frac{1}{2}\right) = \frac{105}{16}\sqrt{\pi} \approx 11.63.$$

As an aside, the Euler integral of the first kind is the Beta function, defined according to

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \quad \alpha > 0, \beta > 0.$$

We note in passing that changes of variables provide a variety of other integral representations of both of these functions. Moreover, both functions can be defined in terms of complex variables although, for our purposes, it is sufficient to assume that α and β are both real. Finally, we note that for all $\alpha > 0$ and $\beta > 0$, $B(\alpha,\beta) = B(\beta,\alpha)$, which is most easily seen from the definition of the Beta function in terms of Gamma functions.

which implies that

$$\int_{-\infty}^{\infty} \phi_z(z)\,\mathrm{d}z = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}z = 1,$$

as required.

- The reason for going to the trouble of establishing that the standard normal density integrated to unity is because we will use essentially the same approach to obtain moments for the distribution. Our initial focus will be on *raw moments* and then we will consider a *central moment*. Raw moments of any random variable Z are defined according to $\mu'_r = \mathrm{E}[Z^r]$. Note the notation: the prime indicates a raw moment and the r subscript denotes the order of the moment. In our case, Z is a standard Normal random variable and so

$$\mu'_r = \mathrm{E}[Z^r] = \int_{-\infty}^{\infty} \phi_z(z) z^r\,\mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}z^2\right\} z^r\,\mathrm{d}z. \tag{3}$$

In order to proceed we will again appeal to the symmetry result that we used earlier but, before we can do so, there is one subtlety that we must take care of. Suppose that we represented z^r as $(z^2)^{r/2}$ and wrote

$$\mu'_r = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}z^2\right\} (z^2)^{\alpha-1}\,\mathrm{d}z, \tag{4}$$

where $\alpha = (r+2)/2 > 1$ for all $r > 0$. If r is even then $z^r > 0$, regardless of the sign of z. However, if r is odd then z^r will have the same sign as z; in particular, $z^r < 0$ if $z < 0$. But $(z^2)^{\alpha-1} \geq 0$ regardless of the sign of z and so the representation used in (4) cannot be valid when $z < 0$. That is, the integral in (4) cannot be equal to the integrals in (3) over the interval $z \in \{-\infty, 0\}$. Let us treat that case more carefully by following the argument of Footnote 5. Specifically, for r odd, from (3) we need to explore the behaviour of

$$\mathcal{I} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \exp\left\{-\tfrac{1}{2}z^2\right\} z^r\,\mathrm{d}z.$$

Make the change of variable $z = -w$, so that $\mathrm{d}z = -\,\mathrm{d}w$. Observe that because z is negative over the support of this integral, w is everywhere positive. Hence,

$$\begin{aligned}
\mathcal{I} &= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{0} \exp\left\{-\tfrac{1}{2}(-w)^2\right\} (-w)^r(-\,\mathrm{d}w) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{0} \exp\left\{-\tfrac{1}{2}w^2\right\} w^r\,\mathrm{d}w
\end{aligned}$$

because $(-1)^r = -1$ on account of r being an odd integer. If we reverse the direction of integration we see that

$$\mathcal{I} = -\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \exp\left\{-\tfrac{1}{2}w^2\right\} w^r\,\mathrm{d}w.$$

That is, for all r odd,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \exp\left\{-\tfrac{1}{2}z^2\right\} z^r\,\mathrm{d}z = -\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \exp\left\{-\tfrac{1}{2}z^2\right\} z^r\,\mathrm{d}z,$$

and so $\mu'_r = 0$ for all r odd. That is, all the odd order raw moments a random variable with a standard Normal distribution are zero, including the mean, $\mu'_1 = \mathrm{E}[Z] = 0$, and the third raw moment, μ'_3, which measures skewness.

When r is an even number, the representation in (4) is valid and we can again appeal to the symmetry argument used in Footnote 5 to write

$$\mu'_r = \sqrt{\frac{2}{\pi}} \int_0^\infty \exp\left\{-\tfrac{1}{2}z^2\right\} (z^2)^{\alpha-1}\,\mathrm{d}z.$$

Similarly, making the change of variable $z = \sqrt{2t}$, for which $\mathrm{d}z = (2t)^{-1/2}\,\mathrm{d}t$, reduces the expression to

$$\mu'_r = \sqrt{\frac{2}{\pi}} \int_0^\infty e^{-t}(2t)^{\alpha-1}(2t)^{-1/2}\,\mathrm{d}t = \frac{2^{\alpha-1}}{\sqrt{\pi}} \int_0^\infty e^{-t}t^{(\alpha-1/2)-1}\,\mathrm{d}t. = \frac{2^{\alpha-1}}{\sqrt{\pi}}\Gamma\left(\alpha - \tfrac{1}{2}\right).$$

We again have the integral expression for the Gamma function appearing which is readily resolved to yield

$$\mu'_r = \frac{2^{\alpha-1}}{\sqrt{\pi}}\Gamma\left(\alpha - 1/2\right) = \frac{2^{r/2}}{\sqrt{\pi}}\Gamma\left(\tfrac{r+1}{2}\right)$$

For r equal to either 0 or an even positive integer, the Gamma function will have a half-integer argument. Noting that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, we see that

r	$\Gamma\left(\frac{r+1}{2}\right)$	μ'_r
0	$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$	1
2	$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$	1
4	$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3\sqrt{\pi}}{4}$	3
6	$\Gamma\left(\frac{7}{2}\right) = \frac{5}{2}\Gamma\left(\frac{5}{2}\right) = \frac{15\sqrt{\pi}}{8}$	15
$\vdots$	$\vdots$	$\vdots$

- Central moments about some fixed point c are of the form $\mu_r(c) = \mathrm{E}[(X-c)^r]$. We typically set $c = \mu'_1 = \mathrm{E}[X] = \mu$, say, so that $\mu_r(\mu) \equiv \mu_r = \mathrm{E}[(X-\mu)^r]$. In the case of the standard Normal distribution, we have established that $\mu = 0$. Hence, for this distribution, all central moments are the same as the corresponding raw moments. In particular, all odd order central moments are zero. Moreover, we see that the second central moment is $\mathrm{E}[(Z-\mu)^2] = \mathrm{Var}[Z] = \mu'_2 = 1$ and the fourth central moment is $\mathrm{E}[(Z-\mu)^2] = \kappa_Z = \mu'_4 = 3$, where κ_Z is a measure of kurtosis. More generally, the relationships between raw and central moments become increasingly complicated as the order of the moment increases. For an excellent discussion of these relationships see, for example, Stuart and Ord (1987, Chapter 3).

2.2 Non-Standard Normal Distributions

One reason the standard Normal distribution has proved important is because, being a singleton it is possible to tabulate probabilities and they have been, extensively. Indeed, most every introductory statistics textbook will include such a table. It is likely that other distributions will be tabulated too, but none as extensively as the standard Normal because those distributions will vary with parameters. These days, there is much more emphasis on generating empirical probabilities via simulation and so the tabulation is not as important as it was. Moreover, random variables with zero mea and unit variance are not always helpful in practical situations.

The Laws of Expectation tell us that we can change the scale and location of a random variable through multiplication by a scale factor, σ say, and addition of a possible negative) constant amount, μ say. In particular, if $Z \sim \mathrm{N}(0,1)$, then we could generate a new random variable, X say, with non-zero mean and and non-unit variable via the transformation $X = \sigma Z + \mu$, or $Z = (X-\mu)/\sigma$. We see that $\mathrm{E}[X] = \sigma\,\mathrm{E}[Z] + \mu = \mu$, as $\mathrm{E}[Z] = 0$, and $\mathrm{Var}[X] = \sigma^2\,\mathrm{Var}[Z] = \sigma^2$, as $\mathrm{Var}[Z] = 1$. Noting that $\mathrm{d}z = \sigma^{-1}\,\mathrm{d}x$ we see, on substitution into $\phi_Z(z)\,\mathrm{d}z$,

$$\begin{aligned} f_X(x;\mu,\sigma^2) = \phi_Z\left(\sigma^{-1}(x-\mu)\right)\,\mathrm{d}z &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\sigma^{-1}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\,\mathrm{d}x. \end{aligned}$$

From this we conclude that the density of $X$is given by that the density of X is given by

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \tag{5}$$

where x can take any value in the interval $-\infty < x < \infty$ and the function is well defined provided that the parameters values in the intervals $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$, respectively.[7] This density of completely specified by knowledge of just two parameters, the mean μ and variance σ^2.

We write $X \sim \mathrm{N}(\mu,\sigma^2)$ to denote a non-standard Normal random variable with mean μ and variance σ^2. As we saw with standard Normal random variables, non-standard Normals are continuous random variable with a symmetric bell shaped probability density function (pdf) centred at μ; see, for example, Figure 1.[8]

The standard Normal distribution is that special case corresponding to $\mu = 0$ and $\sigma^2 = 1$. Any normally distributed random variable, $X \sim \mathrm{N}(\mu,\sigma^2)$ say, can be mapped into a standard Normal variable, $Z \sim \mathrm{N}(0,1)$ according to the standardizing transformation $Z = (X-\mu)/\sigma$. For this reason we can find any probability $P(x_0 \le X \le x_1)$ from a table of standard Normal probabilities according to

$$P(x_0 \le X \le x_1) = P\left(\frac{x_0-\mu}{\sigma} \le Z \le \frac{x_1-\mu}{\sigma}\right).$$

This is why the standard Normal tables have proved so important over the years. It will be assumed throughout the course that you remember how to do this.

[7] Observe that the way I have written the density function takes care to specify the set of values that the random variable can take and also the possible parameter values for which the function is well defined. This information can be very important, although it is often missing, especially in econometrics. Try to think about these issues when you see expressions for density functions.

[8] Many, many texts deal with distribution theory. For a different, and truly excellent, approach you might want to take a look at the book by Rose and Smith (2002) and the accompanying Mathematica notebook. Many of the figures presented here, including this one, either come directly from the notebook or were inspired by figures presented there.

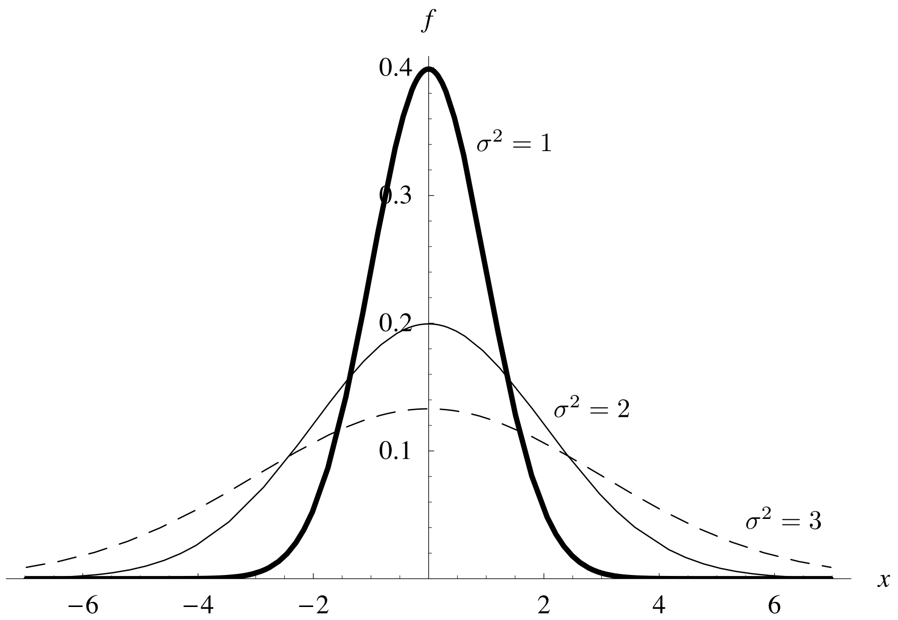


Figure 1: The pdf of a Normal random variable, when $\mu = 0$ and $\sigma^2 = 1$, 2 and 3

People sometimes wish to explore the normality of a population. One way to think about doing this is to look at functions of higher order moments of the distribution to see if they are consistent with the predictions of normality. The easiest approach involves the third and fourth moments about the mean which yield measures of the *skewness* and *kurtosis* of the distribution.[9] A common measure of skewness is[10]

$$\gamma_1 = \frac{\mu_3}{\sigma^3}, \qquad \text{where } \mu_3 = \mathrm{E}\left[(X-\mu)^3\right],$$

and a common measure of kurtosis is

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3, \qquad \text{where } \mu_4 = \mathrm{E}\left[(X-\mu)^4\right].$$

For a normally distributed random variable the distribution is symmetric so that $\mu_3 = 0$, and hence $\gamma_1 = 0$. If $\gamma_1 < 0$ it implies that the distribution is skewed to the left whereas $\gamma_1 > 0$ it implies that the distribution is skewed to the right. Kurtosis can be thought of as a measure is of the degree of flatness of a density near its centre. Positive values of γ_2 suggest that a density is more peaked around its centre than is a Normal distribution, with negative values suggesting a density which is less peaked. For a Normal distribution $\mu_4 = 3\sigma^4$ so that $\gamma_2 = 0$. The joint hypothesis $H_0 : \gamma_1 = \gamma_2 = 0$ is then a testable proposition. This all becomes operational by replacing the population quantities by their sample analogues, e.g., $\hat{\mu}_4 = n^{-1}\sum_{i=1}^{n}(x_i - \overline{x})^4$.

[9] The mean is the first moment (about zero) and variance is the second moment about the mean.

[10] Another common measure of skewness is $s = (\mu - M^d)/\sigma$, where M^d denotes the median of the distribution; it can be shown that $-1 \leq s \leq 1$. Yet another measure involves replacing M^d by the mode in the definition of s, so you can see that there is often no single measure of a quantity and which you use is often governed by circumstance. In all these examples observe how the standard deviation σ has been used to generate a unitless measure, which should be easier to interpret than otherwise.

3 The Bivariate Normal Distribution

3.1 Deriving the Bivariate Normal Density Function

Suppose that we have two independent standard Normal random variables, Z_1 and Z_2 say. Then, by their independence, their joint density function is simply the product of their marginal distributions. That is,

$$\begin{aligned}\phi_{Z_1,Z_2}(z_1,z_2) = \phi_{Z_1}(z_1)\phi_{Z_2}(z_2) &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}z_1^2\right\} \times \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}z_2^2\right\} \\ &= \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(z_1^2+z_2^2\right)\right\} \end{aligned} \tag{6}$$

We can extend this idea to the joint density of two independent non-standard Normal random variables in much the same way. It will be a slightly messier expression, especially if the variables have different variances, but it misses a key feature of multivariate distributions, which comes from the fact that the variables may be correlated with each other.

To illustrate the idea, suppose that Z_1 and Z_2 are as described above and that we define two new variables $X_1 = aZ_1 + bZ_2$ and $X_2 = cZ_1 + dZ_2$, for constants a, b, c, and d, such that $ad - bc \neq 0$. Let us explore the properties of X_1 and X_2:

$$\begin{aligned}\mathrm{E}[X_1] &= a\,\mathrm{E}[Z_1] + b\,\mathrm{E}[Z_2] = 0, \\ \mathrm{E}[X_2] &= c\,\mathrm{E}[Z_1] + d\,\mathrm{E}[Z_2] = 0.\end{aligned}$$

By independence, $\mathrm{Cov}[Z_1, Z_2] = 0$, hence

$$\begin{aligned}\mathrm{Var}[X_1] &= a^2\,\mathrm{Var}[Z_1] + b^2\,\mathrm{Var}[Z_2] = a^2 + b^2 \equiv \sigma_1^2, \\ \mathrm{Var}[X_2] &= c^2\,\mathrm{Var}[Z_1] + d^2\,\mathrm{Var}[Z_2] = c^2 + d^2 \equiv \sigma_2^2.\end{aligned}$$

Finally,

$$\begin{aligned}\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1X_2] &= \mathrm{E}[(aZ_1 + bZ_2)(cZ_1 + dZ_2)] \\ &= ac\,\mathrm{E}\left[Z_1^2\right] + bd\,\mathrm{E}\left[Z_2^2\right] + (ad+bc)\,\mathrm{E}[Z_1Z_2] \\ &= ac\,\mathrm{Var}[Z_1] + bd\,\mathrm{Var}[Z_2] + (ad+bc)\,\mathrm{E}[Z_1]\,\mathrm{E}[Z_2] = ac + bd \equiv \sigma_{12}.\end{aligned}$$

Next, observe that we can write

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \implies \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

That is,

$$\begin{aligned}Z_1 &= (dX_1 - bX_2)/(ad - bc) \\ Z_2 &= (aX_2 - cX_1)/(ad - bc).\end{aligned}$$

In the past when we have changed variables, there has been just a single derivative to work with. Here, in order to see how the volume element $\mathrm{d}z_1\,\mathrm{d}z_2$ changes when we transform from Z_1 and Z_2 to X_1 and X_2 there will be four derivatives that we need to take account of; namely, $\partial Z_1/\partial X_1$, $\partial Z_1/\partial X_2$, $\partial Z_2/\partial X_1$, and $\partial Z_2/\partial X_2$. Section 4.6.1 provides a more complete treatment of what we are about to do. For now we will simply state that

$$\mathrm{d}z_1\,\mathrm{d}z_2 = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}\right)\mathrm{d}x_1\,\mathrm{d}x_2 = \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)^{-1}\mathrm{d}x_1\,\mathrm{d}x_2 = \frac{1}{(ad-bc)}\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

Consequently, on substitution into (6), we see that

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \frac{1}{2(ad-bc)\pi}\exp\left\{-\frac{1}{2(ad-bc)^2}\left((dx_1-bx_2)^2+(ax_2-cx_1)^2\right)\right\} \\ &= \frac{1}{2(ad-bc)\pi}\exp\left\{-\frac{1}{2(ad-bc)^2}\left((c^2+d^2)x_1^2+(a^2+b^2)x_2^2-2(ac+bd)x_1x_2\right)\right\} \\ &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2-\sigma_{12}^2}}\exp\left\{-\frac{1}{2(\sigma_1^2\sigma_2^2-\sigma_{12}^2)}\left(\sigma_2^2x_1^2+\sigma_1^2x_2^2-2\sigma_{12}x_1x_2\right)\right\} \end{aligned}$$

We can readily extend this result to obtain the joint density function of two linear combinations of two non-standard Normal variables. For example, suppose $X_1 = aW_1 + bW_2$ and $X_2 = cW_1 + dW_2$, where $W_1 \sim \mathrm{N}(\mu_{W_1}, \sigma^2_{W_1})$ and $W_2 \sim \mathrm{N}(\mu_{W_2}, \sigma^2_{W_2})$. Following the steps above we will obtain new values for the various parameters of the joint density of X_1 and X_2, but the final expression will reduce to[11]

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2;\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\sigma_{12}) &= (2\pi)^{-1}[\sigma_1^2\sigma_2^2-\sigma_{12}^2]^{-1/2} \\ &\times\exp\left\{-\frac{\sigma_2^2(x_1-\mu_1)^2-2\sigma_{12}(x_1-\mu_1)(x_2-\mu_2)+\sigma_1^2(x_2-\mu_2)^2}{2[\sigma_1^2\sigma_2^2-\sigma_{12}^2]}\right\}, \end{aligned} \tag{7}$$

where $-\infty < x_1, x_2 < \infty$, $-\infty < \mu_1, \mu_2 < \infty$, $\sigma_1^2 > 0$, $\sigma_2^2 > 0$ and $\sigma_1^2\sigma_2^2 - \sigma_{12}^2 > 0$. There is not a lot that we can take out of this other than to recognize that the density depends on five parameters. As we shall see μ_1 and μ_2 are the means of X_1 and X_2, respectively; σ_1 and σ_2 are the variances of X_1 and X_2, respectively; and σ_{12} denotes the covariance between X_1 and X_2. Any two variables with this joint density function are said to be jointly Normally distributed.

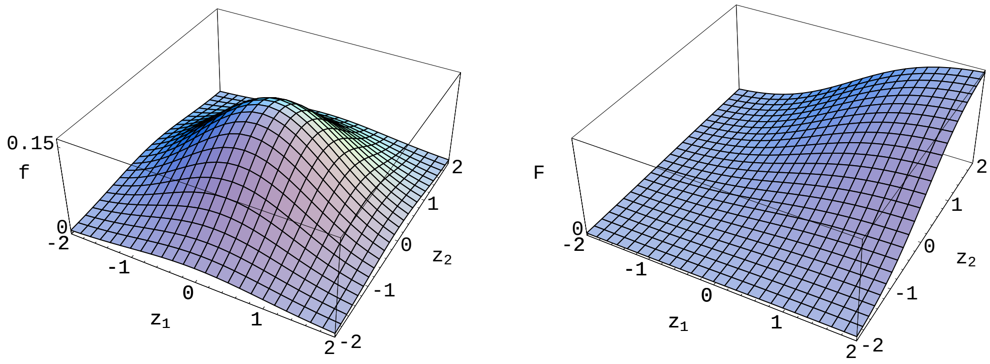
3.2 Describing the Bivariate Normal Distribution

Equation (7) looks a horrible mess and, frankly, it is. Some rearrangement yields a little bit of simplification (although not a lot):[12]

$$\begin{aligned} &f_{X_1,X_2}(x_1,x_2;\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho) \\ &\quad= \left[2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}\right]^{-1}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right.\right. \\ &\qquad\left.\left.-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}. \end{aligned} \tag{8}$$

[11] This is an awful lot easier in matrix notation but we will leave that to the discussion of the multivariate Normal distribution in Section 4.

[12] I am hoping that if you look hard you might be able to see bits and pieces that look like the density functions of univariate Normal random variables.



(a) Joint Probability Density Function

(b) Joint Cumulative Distribution Function

Figure 2: The Bivariate Standard Normal Distribution ($\rho = 0$)

The change that has been made in (8) relative to (7) is to set $\rho = \sigma_{12}/(\sigma_1\sigma_2)$. That is, covariance between X_1 and X_2 has been replaced by correlation between these two random variables. The latter has the advantage of being independent of the scale of both X_1 and X_2, and also being bounded, so that $-1 \leq \rho \leq 1$. These features combine to make ρ easier to think about than σ_{12}.

The nice thing about equation (8), and why it is the representation given in so many textbooks, is that it makes it easy to see a number of things. First, by analogy with the univariate case, we can think about standard forms. For example, if we let $Z_1 = (X_1 - \mu_1)/\sigma_1$ and $Z_2 = (X_2 - \mu_2)/\sigma_2$, so that $\mathrm{E}[Z_1] = \mathrm{E}[Z_2] = 0$ and $\mathrm{Var}[Z_1] = \mathrm{Var}[Z_2] = 1$, then we obtain[13]

$$f_{Z_1,Z_2}(z_1, z_2; 0, 0, 1, 1, \rho) = \left[2\pi(1-\rho^2)^{1/2}\right]^{-1} \exp\left\{-\frac{Z_1^2 - 2\rho Z_1 Z_2 + Z_2^2}{2(1-\rho^2)}\right\}. \tag{9}$$

Let us have a look at some pictures to see what this equation is telling us. Figure 2 illustrates the joint pdf and cumulative distribution function (cdf) for the special case of $\rho = 0$. Figure 3 illustrate how the pdf changes as ρ is allowed to vary. The important feature to notice is how the sign of the correlation between Z_1 and Z_2 changes the shape and orientation of the density function.

Another way of exploring the impact of ρ is through the use of of a contour diagram. To help understand contour plots consider first Figure 4 which plots 1000 observations from a bivariate Normal distribution with zero means and unit variances but with a correlation coefficient of 0.6. Superimposed on these points are ellipses which variously contain theoretical probability contents of 15% (the inner ring), 90%, and 99% (the outer ring).[14] These ellipses correspond to the contours found in a contour diagram. Figure 5 provides contour diagrams for a variety of different density functions. This diagram should be read in much the same way as one reads a contour map, or a weather map for that matter. Each contour encloses a set of points for which the density function takes the same value. The closer together are the contours the steeper is the slope of

[13]Observe that in (9) the normalizing constant, the $\left[2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}\right]^{-1}$ bit, has lost a factor of $(\sigma_1\sigma_2)^{-1}$. This is because when we make the substitutions that we have we need to adjust the density by a scale factor known as the *Jacobian of transformation*, which in this case is equal to $\sigma_1\sigma_2$. This is further discussed in Section 4.4.

[14]Strictly speaking these ellipses are probability quantiles or percentiles.

the surface of the density function. In our case the inner contours correspond to larger values of the density function, which always takes its largest value at the origin. Observe that for the case of the $\rho = 0$ the contours are perfectly circular. This illustrates that all outcomes equal distances form the origin are equally likely. Conversely, as ρ moves further away from zero the contours become increasingly elliptical, with the slope of the principle diagonal of the ellipse depending on the sign of ρ. For instance, when $\rho = -0.98$ it is more likely to observe Z_1 and Z_2 taking opposite signs than otherwise.

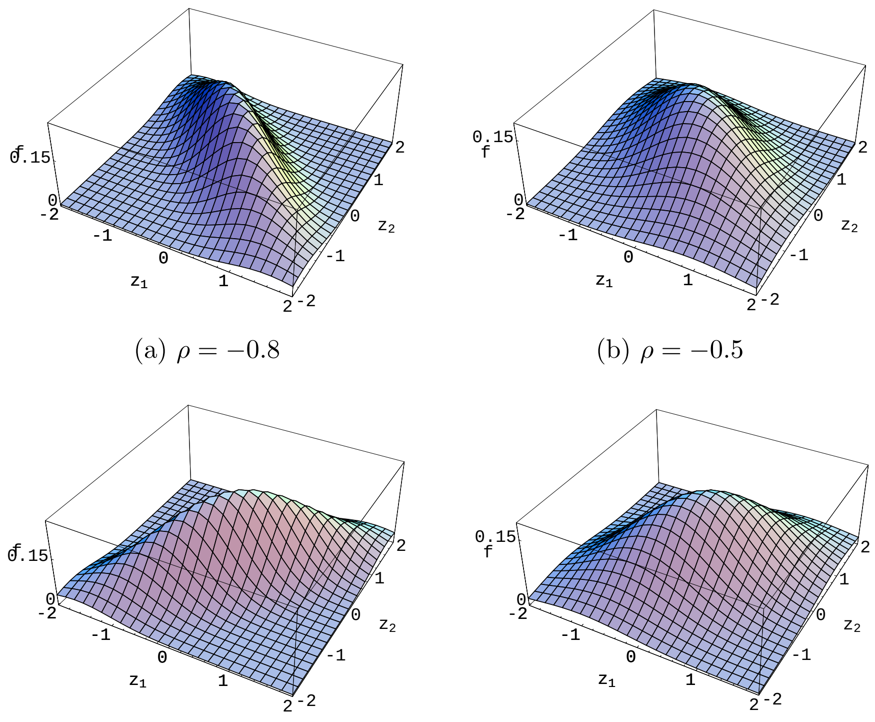
3.3 Properties of the Bivariate Normal Distribution

So much for the shape of the distribution, what else can we learn? If, in equation (9), it is also the case that Z_1 and Z_2 are uncorrelated, so that $\rho = 0$,[15] then we have the further simplification:

$$\begin{aligned} f_{Z_1,Z_2}(z_1, z_2; 0, 0, 1, 1, 0) &= (2\pi)^{-1} \exp\left\{-\tfrac{1}{2}\left[Z_1^2 + Z_2^2\right]\right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{Z_1^2}{2}\right\} \times \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{Z_2^2}{2}\right\} \\ &= f_{Z_1}(z_1; 0, 1) \times f_{Z_2}(z_2; 0, 1), \end{aligned} \tag{10}$$

where we see that both Z_1 and Z_2 have standard Normal distributions.

[15] It is always true that if two random variables are independent then they are uncorrelated; only for the Normal distribution is the converse true. That is, if two random variables are uncorrelated and they are jointly normally distributed then it follows that they are also independent.



(a) $\rho = -0.8$ (b) $\rho = -0.5$

(c) $\rho = 0.8$ (d) $\rho = 0.5$

Figure 3: The Bivariate Normal Distribution with non-zero ρ

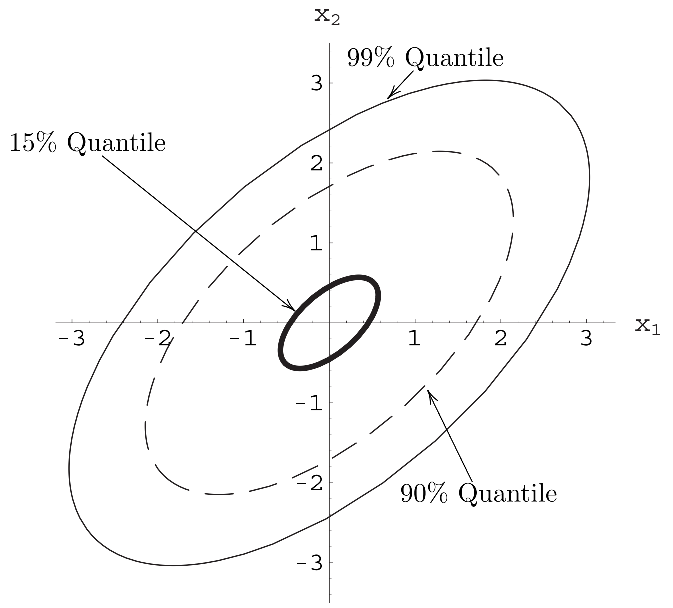


Figure 4: Scatter Plot of 1000 Drawings from a Bivariate Normal Distribution Against 15%, 90% and 99% Quantiles

Equation (10) illustrates a very important characteristic of jointly normally distributed sets of random variables; namely:[16]

> If a set of variables $X_1, \ldots, X_n$ is jointly normally distributed then each of the random variables has a Normal marginal distribution.[17] In particular, if X_1 and X_2 have density
> $$f_{X_1,X_2}(x_1, x_2; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$$
> as defined in equation (7), then $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$, and $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$.

We can obtain conditional density functions on dividing a joint density by the marginal density of the conditioning variate. For example, suppose I want the conditional density of X_1 given X_2 from a bivariate Normal distribution. Then[18]

$$\frac{f_{X_1,X_2}(x_1, x_2; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})}{f_{X_2}(x_2; \mu_2, \sigma_2^2)} = f_{X_1|X_2}\left(x_1; \mu_1 + \frac{\sigma_{12}(x_2 - \mu_2)}{\sigma_2^2}, \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right).$$

Observe that, because $\sigma_{12}^2/\sigma_2^2 \geq 0$ (why?),[19] the variance of the conditional distribution is never greater than that of the unconditional distribution. Equally, we have

$$\frac{f_{X_1,X_2}(x_1, x_2; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})}{f_{X_1}(x_1; \mu_1, \sigma_1^2)} = f_{X_2|X_1}\left(x_2; \mu_2 + \frac{\sigma_{12}(x_1 - \mu_1)}{\sigma_1^2}, \sigma_2^2 - \frac{(\sigma_{12}^1)^2}{\sigma_1^2}\right).$$

[16] We won't actually prove this result as that would require some tedious integration which is beyond the scope of what we are trying to achieve here.

[17] Surprisingly the converse statement is not true. Just because a bunch of random variables each have Normal marginal distributions does not imply that they will be jointly normally distributed.

[18] Convince yourself that this is so. That is, write out the formulae and divide one density by another. Because it involves exponential functions it is not that hard to do. Later on we will derive a variant of this result by factorizing the joint density to obtain the product of the conditional density and the marginal density of the conditioning variate.

[19] To be honest, I hate it when books do this. If you know how it works then you don't need the prompt and if you don't know how it works the prompt just makes you feel bad. Anyway both the numerator and denominator involve squares and so they are non-negative as must be their ratio. Note we have assumed that our variances are non-zero so the ratio is well defined.

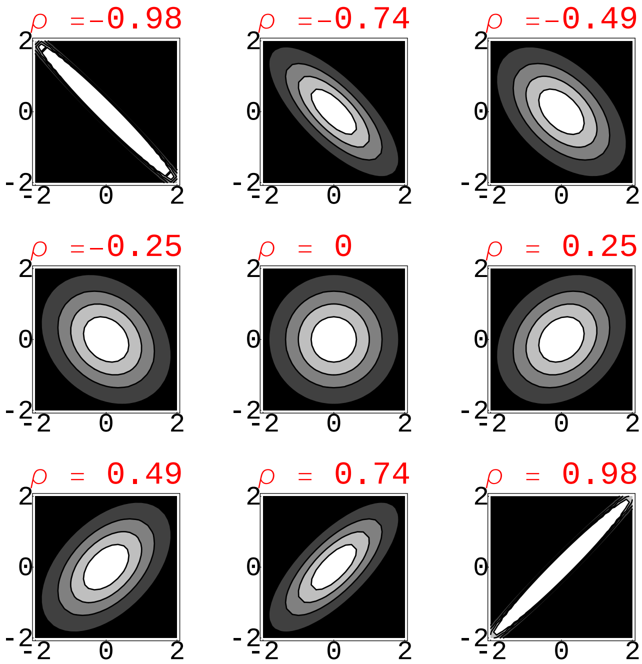


Figure 5: Contour Plots of the Bivariate Standard Normal Distribution

One thing to notice about these two expressions is that, not only are the marginal distributions Normal, but so too are the conditional distributions. This is a remarkable property of the Normal distribution, of which the bivariate distribution is but a special case. If you wish to see a plot of a conditional Normal distribution look back to the densities depicted in Figure 2a and Figure 3. Each of the densities is overlaid with a grid as a visual device to show the varying heights of the density functions. As it happens each one of the 'lines' that go to make up the grid is a conditional distribution of one variable given the other.[20] For any line that you choose the conditioning variable is the one whose axis is perpendicular to the line you are looking at. The axis that is parallel to the line you are looking at is that of the variable for which the line represents the conditional probability density function.

The final thing that we will comment on here is another property of the Normal distribution that is quite remarkable. Suppose I have a set of jointly normally distributed random variables X_1,...,X_n. Next suppose that I form a weighted sum, or linear combination, of these random variables, i.e., suppose I construct a sum of the form

$$S = c_1 X_1 + \ldots + c_n X_n,$$

where c_1, ..., c_n is a set of constants with respect to the random variables.[21] We know from the rules of expectation that

$$\mathrm{E}\left[S\right] = c_1 \,\mathrm{E}\left[X_1\right] + \ldots + c_n \,\mathrm{E}\left[X_n\right],$$

and that

[20]Line is not really the right word here as these 'lines' curve all over the place, although all of their curvature occurs in a plane that is perpendicular to the plane described by the pair of axes, perpendicular to one of the axes and parallel to the other.

[21]Recall from the earlier handout on probability that this may means that c_1, ..., c_n are actually constants but it may simply mean that they are random variables that we have conditioned on.

$$\begin{aligned}\mathrm{Var}[S] &= c_1^2\,\mathrm{Var}[X_1] + \ldots + c_n^2\,\mathrm{Var}[X_n] \\ &\quad + 2c_1c_2\,\mathrm{Cov}[X_1, X_2] + 2c_1c_3\,\mathrm{Cov}[X_1, X_3] + \ldots + 2c_{n-1}c_n\,\mathrm{Cov}[X_{n-1}, X_n]\,.\end{aligned}$$

It can also be shown that S is normally distributed. Given that the mean and variance completely characterize a Normal distribution, we have just specified the distribution of S. Why should we care? Determining the distributions of weighted sums of normally distributed random variables is a problem that we shall encounter repeatedly throughout the course. The easiest example that you will have encountered so far is simply that of the mean of a simple sample of size n from a population which is $\mathrm{N}(\mu, \sigma^2)$. In this case, $c_1 = \ldots = c_n = \frac{1}{n}$ and all of the covariances are zero. Applying these rules leads immediately to the result that $\overline{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$.

4 The Multivariate Normal Distribution

Everything that has been said about the bivariate Normal distribution applies to the multivariate Normal distribution. The only thing that would be in any way difficult to port across from the previous section is the figures, the other results carry over directly. In the remaining parts of this section we will explore the multivariate Normal distribution along with some of its properties and application.

4.1 The Bivariate Normal Distribution Re-Visited

As a first step, let us take another look at the density function of a bivariate Normal distribution as specified in (7). Suppose that we were to define the 2×1 vector of jointly Normal variates $X = [X_1\ X_2]'$, with corresponding parameters $\mathrm{E}[X] \equiv \mu = [\mu_1\ \mu_2]'$ and

$$\mathrm{Var}[X] \equiv \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

We will need the results

$$\det(\Sigma) = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 \tag{11a}$$

$$\Sigma^{-1} = \frac{1}{\det(\Sigma)}\begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix} = \frac{1}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}\begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix}. \tag{11b}$$

If we try to gather the various elements of (7) in matrix terms then we obtain

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12}) &= (2\pi)^{-1}[\sigma_1^2\sigma_2^2 - \sigma_{12}^2]^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)'\left(\frac{1}{[\sigma_1^2\sigma_2^2 - \sigma_{12}^2]}\begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix}\right)\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)\right\} \\ &= (2\pi)^{-1}\left(\det(\Sigma)\right)^{-1/2}\exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\} \\ &= f_X(x; \mu, \Sigma) \quad \text{(say)}. \end{aligned} \tag{12}$$

4.2 The Multivariate Normal Density Function

The extension from (12) to the multivariate normal density follows almost immediately because we already have an expression in matrix form, almost all that we need to do is to allow the matrices to contain more elements than the two for the bivariate case. In particular, suppose that the n-vectors X and μ are such that $-\infty < X_i, \mu_i < \infty$, $i = 1, \ldots, n$, and the $n \times n$ matrix $\Sigma > 0$. Note that when we write $\Sigma > 0$ we mean that Σ is a positive definite (symmetric) matrix, which implies that it is invertible and that its inverse will also be positive definite (and symmetric). The only thing left to do is to generalize the power of the '2π' term. If we look at the univariate density (5) then we see that the '2π' term has power $-\frac{1}{2}$. Looking at the bivariate density (7) we see that it has power -1. When we consider the n-variate case then the power '2π' term has power $-\frac{n}{2}$. Thus, final form of the density function in the multivariate case is[22]

$$f_X(x; \mu, \Sigma) = \underbrace{(2\pi)^{-n/2}(\det(\Sigma))^{-1/2}}_{\text{normalizing constant}} \underbrace{\exp\left\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}}_{\text{kernel of density}} \tag{13}$$

Given that proper density functions, such as this one, integrate to unity, we have the identity

$$\begin{aligned}
&\int_X (2\pi)^{-n/2}(\det(\Sigma))^{-1/2} \exp\left\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}\,\mathrm{d}x = 1 \\
\Longrightarrow\ &(2\pi)^{-n/2}(\det(\Sigma))^{-1/2} \int_X \exp\left\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}\,\mathrm{d}x = 1 \\
\Longrightarrow\ &\int_X \exp\left\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\} dx = (2\pi)^{n/2}(\det(\Sigma))^{1/2}. \qquad (14)
\end{aligned}$$

This result is known as Aitken's integral.

Just to be clear about the notation used in (14), we are actually evaluating an n-fold integral and so the way to read the integral is

$$\int_X g(x)\,\mathrm{d}x \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

The product $\mathrm{d}x_1 \ldots \mathrm{d}x_n$ is often called the volume element of the integral, which is a generalization of the differential element of a univariate integral.

4.3 Marginal and Conditional Densities

In this section we shall take a set of jointly Normal random variables, X say, that we will break into two subsets, such that $X = [X_1', X_2']'$. Our concerned is then finding marginal and conditional density functions for X_1 and X_1 given X_2, respectively. We will exploit the following results about partitioned matrices: Let $n = n_1 + n_2$, and partition x, μ, and Σ conformably, so that

$$x = \begin{array}{cc} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \end{array}, \qquad \begin{array}{cc} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \end{array}, \qquad \text{and} \qquad \begin{array}{cc} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \end{array} \\ \begin{array}{cc} n_1 & n_2 \end{array} & \end{array}.$$

Because we have assumed that $\Sigma > 0$, this, in turn, implies that $\Sigma_{11} > 0$, $\Sigma_{22} > 0$, and $\Sigma_{21} = \Sigma_{12}'$. Then:

$$\det(\Sigma) = \det(\Sigma_{11}) \times \det(\Sigma_{22\cdot 1}) \tag{15a}$$

[22] We will establish this result properly in Section 4.6.1.

and

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1} \\ -\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22\cdot 1}^{-1} \end{bmatrix}, \tag{15b}$$

where $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Note that the results of (11) are special cases of (15) when $n_1 = n_2 = 1$.

As most of the action will be in the quadratic form that is the argument of the exponential function, we will address that first. Let $Q = (x-\mu)'\Sigma^{-1}(x-\mu)$ then, from (15),

$$\begin{aligned} Q =&(x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1) \\ &+ (x_1-\mu_1)'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1) - (x_1-\mu_1)'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}(x_2-\mu_2) \\ &- (x_2-\mu_2)'\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1) + (x_2-\mu_2)'\Sigma_{22\cdot 1}^{-1}(x_2-\mu_2) \\ =&(x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1) \\ &+ \left[(x_2-\mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1)\right]'\Sigma_{22\cdot 1}^{-1}\left[(x_2-\mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1)\right]. \end{aligned}$$

Note that the second term of the final line has exploited the symmetry of Σ. Substituting this result, together with (15a), into (13) yields

$$\begin{aligned} f(x) =&(2\pi)^{-n_1/2}\left(\det\left(\Sigma_{11}\right)\right)^{-1/2}\exp\left\{-\tfrac{1}{2}(x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1)\right\} \times (2\pi)^{-n_2/2}\left(\det\left(\Sigma_{22\cdot 1}\right)\right)^{-1/2} \\ &\times \exp\left\{-\tfrac{1}{2}\left[x_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1)\right]'\Sigma_{22\cdot 1}^{-1}\left[x_2-\mu_2-\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1)\right]\right\}. \end{aligned}$$

which we can see to be the product of two Normal density functions:[23]

$$f(x) = f(x_1)f(x_2 \mid x_1), \tag{16}$$

where $X_1 \sim \mathrm{N}(\mu_1, \Sigma_{11})$ and $X_2 \mid X_1 \sim \mathrm{N}\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1), \Sigma_{22\cdot 1}\right)$.

As an aside, because $\Sigma_{11} > 0$ it follows that $\Sigma_{11}^{-1} > 0$ and that $\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \geq 0$.[24] Consequently, it follows that $\Sigma_{22} - \Sigma_{22\cdot 1} = \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \geq 0$. That is, Σ_{22} exceeds $\Sigma_{22\cdot 1}$ by a positive definite matrix *provided that* $\Sigma_{21} \neq 0$. This is another way of saying that conditioning on a set of jointly Normal random variables on a correlated set of jointly Normal random variables leads to a reduction in variance. However, conditioning a set of jointly Normal random variables on a set of uncorrelated Normal random variable has no such effect. Moreover, in the event that $\Sigma_{21} = 0$, we see that $f(x_2 \mid x_1) = f(x_2)$ and, as now $f(x) = f(x_1)f(x_2)$, we conclude that uncorrelated jointly Normal random variables are also independent Normal random variables as their joint distribution factors into the product of their marginals.

As a second aside, the mean of the conditional distribution, $\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1)$, is called a *regression function*.

Finally, the symmetry of the problem is such that we might equally write

$$f(x) = f(x_2)f(x_1 \mid x_2),$$

where $X_2 \sim \mathrm{N}(\mu_2, \Sigma_{22})$ and $X_1 \mid X_2 \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2), \Sigma_{11\cdot 2}\right)$.

[23]Observe that we are being pretty sloppy in our notation as the density function f are not indexed by the appropriate random variables. However, in this case, those random variables are obvious from the context.

[24]If the $n \times n$ matrix $S > 0$ then for an $m \times n$ matrix $A \neq 0$, $ASA' \geq 0$. If A is of full row rank and if $m \leq n$ then $ASA' > 0$. However, if $m > n$ or if A has less than full row rank then ASA' is only positive semi-definite, that is, $ASA' \geq 0$. Similar reasoning can be applied if $S < 0$, with signs simply reversed relative to the $S > 0$ case.

4.4 Jointly Normal Random Variables and Invertible Affine Transformations

An affine transformation is one where we change both scale and location. In the multivariate case that means pre-multiplication by a matrix and addition of a constant vector. That is, we seek the properties of $Y = AX + b$, where A and b are matrices/vectors conformable with X under matrix multiplication and matrix addition, respectively.[25] In this section we will assume with A is a non-singular $n \times n$ matrix and b an n-vector. This means that there are as many random variables in the vector Y as there are in X. Armed with these definitions, observe that

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}[AX+b] = A\,\mathrm{E}[X] + b = A\mu + b = \mu_y \text{ (say)} && (17a)\\ \mathrm{Var}[Y] &= \mathrm{E}[(Y-\mathrm{E}[Y])(Y-\mathrm{E}[Y])'] \\ &= \mathrm{E}\left[(AX+b-(A\mu+b))\,(AX+b-(A\mu+b))'\right] \\ &= A\,\mathrm{E}[(X-\mu)(X-\mu)']\,A' \\ &= A\Sigma A' = \Omega \text{ (say)}. && (17b) \end{aligned}$$

Note that if A had more rows than X then $\mathrm{Var}[Y]$ would be singular and life would become rather more complicated. We won't go down that path. We shall also note in passing that the support of each element of X is the set of real numbers and that that remains true for each element of Y.

To obtain the density of Y we will proceed much as we did in the univariate case (which is why we spent so much time on the univariate case). Recall that the steps were as follows. First, we need the inverse transformation, namely $X \equiv X(Y) = A^{-1}(Y-b)$. Second, we need to adjust the volume element. This is somewhat trickier that before, where only a single transformation was being performed. Now we have n transformations happening at once and they each, potentially, involve all n elements of X.

From the theory of integration, when you perform a multivariate change of variables you need to scale the volume element by a quantity known as the *Jacobian of transformation*.[26] Without going into the detail of how this is derived, it can be shown that, in our case,

$$dX = (\det(A))^{-1} dY.$$

More generally, the Jacobian of transformation is defined to be $\mathrm{abs}(\det(J))$, where $J = \partial X'/\partial Y$ denotes the so-called Jacobian matrix, being the matrix of partial derivatives of X with respect to Y; that is, the partial derivatives of the inverse of the transformation from X to Y. In our case, where the inverse transformation is simply $X = A^{-1}(Y-b)$, $J = \partial X'/\partial Y = A^{-1}$, as required. As a more general statement of this result, if $f_X(x)$ denotes the density function of X, and if $Y = u^{-1}(X)$ so that $X = u(Y)$ is the inverse of the transformation from $X \to Y$, where both u and u^{-1} are one-to-one in the supports of both X and Y, then the density function of Y is given by

$$f_Y(y) = f_X(u(y))\,\mathrm{abs}(\det(J)). \tag{18}$$

Note that the Jacobian of transformation is different for each different transformation that you consider. As a practical matter, it may be easier to obtain the Jacobian of

[25] A *linear transformation* changes just scale, so that $b = 0$.

[26] Named for the German mathematician Carl Gustav Jacob Jacobi (10 December 1804 — 18 February 1851). For (just a little) more discussion, see Section 9.6 of the Matrices handout.

transformation by noting $J^{-1} = \partial Y'/\partial X$. That is, obtain J by inverting the Jacobian matrix $\partial Y'/\partial X$ rather than having to figure out the inverse transformations u. The results can be extended to allow for situations where relationships are not one-to-one but that is a story for another day. We will leave this here for now, but it becomes extremely important in the theory of probability distributions and in the theory of integration more broadly.

Returning to our problem, we can apply (18) to (13) on writing $x(y) = A^{-1}(y-b)$ to obtain

$$\begin{aligned} f_Y(y) = f_{X(Y)}(x(y); \mu, \Sigma) &= (2\pi)^{-n/2}(\det(\Sigma))^{-1/2} \\ &\times \exp\left\{-\frac{1}{2}\left(A^{-1}(y-b)-\mu\right)'\Sigma^{-1}\left(A^{-1}(y-b)-\mu\right)\right\}\operatorname{abs}\left((\det(A))^{-1}\right). \end{aligned} \tag{19}$$

Recall that, by assumption, A is invertible, which means that $\det(A) \neq 0$. Because the Jacobian of transformation is specified to be the absolute value of the determinant of the Jacobian matrix, we know that the quantity is positive, given that it can't be zero. Observe that for any real number w, $\operatorname{abs}(w) = \sqrt{w^2}$ because, by definition, square roots are positive. For example, both $(-2)^2 = 4$ and $2^2 = 4$ but $\sqrt{4} = 2$ and, in particular, -2 is a not a square root of 4. As $(\det(A))^{-1} = \det(A^{-1})$, we see that

$$\operatorname{abs}\left((\det(A))^{-1}\right) = \left((\det(A^{-1}))^2\right)^{1/2} = \left((\det(A))^2\right)^{-1/2}.$$

Combining all the determinants in (19) we have

$$\begin{aligned} (\det(\Sigma))^{-1/2}\left((\det(A))^2\right)^{-1/2} &= \left(\det(\Sigma)(\det(A))^2\right)^{-1/2} = (\det(A)\det(\Sigma)\det(A))^{-1/2} \\ &= (\det(A)\det(\Sigma)\det(A'))^{-1/2} = (\det(A\Sigma A'))^{-1/2} = (\det(\Omega))^{-1/2}, \end{aligned} \tag{20}$$

where we have used the results (i) the determinant of a matrix is equal to the determinant of the transpose of that matrix, and (ii) the product of determinants of non-singular matrices is equal to the determinant of the product of the matrices.

Turning next to the exponential term, we can take out a common factor of A^{-1}.

$$\begin{aligned} &\exp\left\{-\frac{1}{2}\left(A^{-1}(y-b)-\mu\right)'\Sigma^{-1}\left(A^{-1}(y-b)-\mu\right)\right\} \\ &\quad = \exp\left\{-\frac{1}{2}\left((y-b)-A\mu\right)'\left(A^{-1}\right)'\Sigma^{-1}A^{-1}\left((y-b)-A\mu\right)\right\} \\ &\quad = \exp\left\{-\frac{1}{2}\left(y-(A\mu+b)\right)'\left(A^{-1}\right)'\Sigma^{-1}A^{-1}\left(y-(A\mu+b)\right)\right\} \\ &\quad = \exp\left\{-\frac{1}{2}\left(y-(A\mu+b)\right)'\left(A\Sigma A'\right)^{-1}\left(y-(A\mu+b)\right)\right\}, \end{aligned} \tag{21}$$

where we have used the results (i) $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$, for non-singular matrices A, B, and C, and (ii) $(A')^{-1} = (A^{-1})'$. Finally, combining the results (20) and (21) in (19), and using the definitions of μ_y and Ω in (17), we obtain the final result

$$f_Y(y) = (2\pi)^{-n/2}(\det(\Omega))^{-1/2}\exp\left\{-\frac{1}{2}(y-\mu_y)'\,\Omega^{-1}(y-\mu_y)\right\}, \tag{22}$$

which is the density function of a multivariate Normal random variable with mean μ_y and covariance matrix Ω. That an arbitrary affine transformation of a set of jointly Normal random variables yields a different set of jointly Normal random variables is something that is unique to multivariate Normal distributions. This is a very important property and is, in fact, one way of characterizing the distribution.

4.5 Jointly Normal Random Variables and Non-Invertible Affine Transformations

Now consider the problem of finding the properties of

$$Y = AX + b, \tag{23}$$

where the $m \times n$ matrix A is of full row rank, so that $m \leq n$ and b is an m-vector, with both A and b non-stochastic. Because A is not square it cannot be invertible. Results for the mean and variance of Y are identical to those of (17). Distributional results are slightly more difficult to tackle directly. We might observe that fixed linear combinations of jointly Normal random variables are themselves normal and so we might simply write down the density function given that we know the mean and variance, although this does abstract away from the problem of whether the elements of Y remain jointly Normal.

Alternatively, we might observe that it is always possible to find a matrix B also non-stochastic and of full row rank such that the matrix $[A', B']'$ is nonsingular. Moreover, if we let $d = (b', 0')'$ be an n-vector then we can write

$$\begin{bmatrix} Y \\ W \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} X + d \sim \mathrm{N}\left(\theta, \Omega\right),$$

where

$$\theta = \begin{bmatrix} A \\ B \end{bmatrix} \mu + d \qquad \text{and} \qquad \Omega = \begin{bmatrix} A \\ B \end{bmatrix} \Sigma \begin{bmatrix} A \\ B \end{bmatrix}'.$$

Given that Y and W are jointly Normal it follows from our earlier results that Y has a marginal Normal distribution, that is, $Y \sim \mathrm{N}(A\mu + b, A\Sigma A')$.

4.6 Applications

In this section we provide a couple of applications of the properties of the multivariate Normal distribution that were developed earlier. The first application is to derive (13) and the second involves establishing the independence of the least squares estimator and the residual sum of squares in the classical linear regression model (with a normality assumption).

4.6.1 The Joint Density Function

Suppose that $Z_1, \ldots, Z_n$ denotes a set of independent standard Normal random variables.[27]. That is, they each have density (from (5))

$$f_{Z_i}(z_i) = (2\pi)^{-1/2} \exp\left\{-\tfrac{1}{2} z_i^2\right\}, \qquad i = 1, \ldots, n.$$

As they are independent, by definition the joint density of $Z = [Z_1, \ldots, Z_n]'$ is then

$$f_Z(z) = \prod_{i=1}^{n} f_{Z_i}(z_i) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\} = (2\pi)^{-n/2} \exp\left\{-\tfrac{1}{2} z'z\right\}.$$

[27] Given that $Z_1, \ldots, Z_n$ are both independent and identically distributed we frequently speak of them as being iid, although that is less informative than and does not imply that the the identical distribution that they all share is a standard Normal.

We can obtain a non-standard Normal distribution via the invertible affine transformation $X = u^{-1}(Z) = \Sigma^{1/2}Z + \mu$, where $\Sigma^{1/2}$ is defined such that $\Sigma^{1/2}(\Sigma^{1/2})' = \Sigma > 0$, $X = [X_1, \ldots, X_n]$, and $\mu = [\mu_1, \ldots, \mu_n]'$. In terms of the notation of Section 4.4, $Z = u(X) = \Sigma^{-1/2}(X - \mu)$, with the density of X given by

$$f_X(x) = f_Z(u(x)) \,\text{abs}\,(\det(J)),$$

where

$$J = \left[\frac{\partial Z'}{\partial X}\right] = \begin{bmatrix} \frac{\partial Z_1}{\partial X_1} & \cdots & \frac{\partial Z_n}{\partial X_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z_1}{\partial X_n} & \cdots & \frac{\partial Z_n}{\partial X_n} \end{bmatrix}.$$

If we let $\sigma^{i,j}$ denote the (i,j)th element of $\Sigma^{-1/2}$ then we see that

$$Z_i = \sum_{j=1}^{n} \sigma^{i,j}(X_j - \mu_j).$$

Clearly,

$$\frac{\partial Z_i}{\partial X_j} = \sigma^{i,j}$$

so that $J = \Sigma^{-1/2}$. Making the relevant substitutions yields

$$\begin{aligned} f_X(x) &= f_Z(u(\Sigma^{-1/2}(x-\mu)))\,\text{abs}\left(\det(\Sigma^{-1/2})\right) \\ &= (2\pi)^{-n/2}\,\text{abs}\left((\det(\Sigma))^{-1/2}\right)\exp\left\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\} \\ &= (2\pi)^{-n/2}(\det(\Sigma))^{-1/2}\exp\left\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}, \end{aligned}$$

because $\Sigma > 0 \implies \det(\Sigma) > 0 \implies (\det(\Sigma))^{-1/2} > 0$, making the absolute value operator redundant. Compare this result with (13).

4.6.2 Some Properties of Least Squares

One reason for thinking about the Normal distribution in the way that we have is that it yields useful results in the classical linear regression model which, in turn, is a very important device in asymptotic analysis. In particular, if we have a vector of data Y that we seek to model in terms of k regressors that can be stack into an $n \times k$ data matrix according to $X = [X_1, \ldots, X_k]$, which we usually assume to be of full column rank, where we typically have $X_1 = \imath$, a column of ones. The linear model in its most basic form is then

$$Y = X\beta + u, \quad u \sim \text{N}(0, \sigma^2 I_n)$$

and the OLS estimator is that value $\hat{\beta}$ that minimizes the sum of squared residuals

$$\hat{\beta} = \arg\min_{\beta} S(\beta; Y, X) = \arg\min_{\beta}(Y - X\beta)'(Y - X\beta).$$

We know that this yields the solution

$$\hat{\beta} = (X'X)^{-1}X'Y$$

with fitted values

$$\widehat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = P_X Y$$

and residuals

$$e = Y - X\hat{\beta} = Y - P_X Y = (I - P_X)Y = M_X Y.$$

Now the matrices P_X and M_X are very special matrices. First, observe that $P_X X = X$ and that $X'P_X = X'$. Similarly, $M_X X = 0$ and $X'M_X = 0'$.

Second, the matrices P_X and M_X are sometimes called orthogonal projectors. This names comes from the ideal that P_X projects the line from the origin to the point Y in $\mathbb{R}^n$ orthogonally onto the space spanned by the columns X. In essence this means that it comes at right angles to the X-hyper plane. Equally, because $M_X P_X = P_X M_X = 0$,[28] it can be shown that M_X project the line joining the origin to Y onto a space orthogonal to that spanned by the columns of X. An important application of this result is that

$$\widehat{Y}'e = Y'P_X'M_X Y = Y'P_X M_X Y = 0$$

where the second equality follows from the symmetry of P_X. Equally, we might have exploited the symmetry of M_X to demonstrate that $e'\widehat{Y} = 0$. In any event, the key result is that, by construction, the OLS residuals are orthogonal to the fitted values.

Third, and equally as important, is the recognition that both P_X and M_X are symmetric, idempotent. This can manifest itself in lots of different ways. First, idempotent matrices refer to matrices that are idempotent under matrix multiplication. Specifically, for any idempotent matrix J, $JJ = J'J = J$. Because the matrices are symmetric, all of their eigenvalues are real and, indeed, take only the values 0 or 1 (because the matrix J is idempotent). The number of non-zero eigenvalues is equal to the rank of the matrix. In the case of P_X, it has rank k, whereas M_X has rank $n - k$. Of particular interest is the fact that all $n \times n$ symmetric matrices S of rank p admit full rank decompositions of the form $S = CC'$, where C is $n \times p$. Moreover, in the case of idempotent matrices, it can also be shown that $C'C = I_p$. Observe that

$$M_X X = 0 \implies CC'X = 0 \implies C'CC'X = C'0 = 0 \implies I_p C'X = 0 \implies C'X = 0.$$

Next, given that $Y \sim \mathrm{N}(X\beta, \sigma^2 I_n)$, observe that if $G = [X(X'X)^{-1}, C]'$ then $GY \sim \mathrm{N}(GX\beta, \sigma^2 GG')$, where

$$GX\beta = \begin{bmatrix} (X'X)^{-1}X'X\beta \\ C'X\beta \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$$

and

$$\begin{aligned} GG' &= \begin{bmatrix} (X'X)^{-1}X' \\ C' \end{bmatrix} [X(X'X)^{-1}, C] = \begin{bmatrix} (X'X)^{-1}X'X(X'X)^{-1} & (X'X)^{-1}X'C \\ C'X(X'X)^{-1} & C'C \end{bmatrix} \\ &= \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & I_p \end{bmatrix}. \end{aligned}$$

Because the covariances between $\hat{\beta} = (X'X)^{-1}X'Y$ and $C'Y$ are zero, given that these two variables are jointly normally distributed it follows that $\hat{\beta}$ and $C'Y$ are also independent.

[28] For instance, $M_X P_X = (I - P_X)P_X = P_X - X(X'X)^{-1}X'X(X'X)^{-1}X' = P_X - P_X = 0$.

Hence, it follows that $\hat{\beta}$ is independent of functions of $C'Y$, specifically, $\hat{\beta}$ is independent of $Y'CC'Y = Y'M_XY = u'M_xu = e'e$. That is, the OLS estimator of β is independent of the residual sum of squares, which is a very important result in the construction of numerous test statistics. More on this later.

4.7 Some References

Most people have a favourite book on matrix algebra. (Ask your parents what their's is.) It is usually where you encountered this stuff for the first time. For me, the best is Searle (1982). Other papers that I like that are somewhat relevant to the partitioned matrix stuff include Henderson and Searle (1981) and Ouellette (1981).

5 Distributions Related to the Normal

5.1 Introduction

The wonderful properties of the Normal distribution make it particularly attractive to work with in lots of situations because it is relatively easy to do so. We can obtain the exact sampling distributions of various functions of jointly normally distributed random variables, which makes certain types of inference (learning about the population from a sample) relatively easy. Our aim here is to consider some particularly important functions that are commonly encountered in the theory of sampling from a Normal population. Throughout what follows we will assume that we are working with a simple random sample of n observations $X_1, \dots, X_n$ from a Normal population with mean μ and variance σ^2, i.e., the population is $X \sim \mathrm{N}(\mu, \sigma^2)$. One implication of this assumption is that $X_1, \dots, X_n$ are jointly normally distributed but independent, so that the covariance between any pair of observations is zero.

5.2 The Chi-Squared Distribution

We frequently estimate variances using statistics of the form

$$\tilde{\sigma}^2 = \frac{1}{\nu}\sum_{i=1}^{n}(X_i - \tilde{\mu})^2,$$

where ν is a scale factor, $\tilde{\mu}$ is an estimator of some mean, and $\tilde{\sigma}^2$ is an estimator of the variance about that mean σ^2. Two common examples are:

Unconditional Variance If we set $\nu = n - 1$ and $\tilde{\mu} = \overline{X}$, then we obtain an unbiased estimator of the unconditional variance of a random variable; that is $\tilde{\sigma}^2 = s^2$.

Conditional variance Suppose that we fit a linear regression model of the form

$$\widehat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i.$$

We can then define residuals $e_i = Y_i - \widehat{Y}_i$. An unbiased estimator of the conditional variance of Y given X is

$$\tilde{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2.$$

The study of variance estimators leads us naturally to the study of the chi-squared distribution.[29] There are two important results that you need to know, which will be stated without proof

1. If $Z \sim \mathrm{N}(0,1)$ then $Q = Z^2 \sim \chi^2_1$, where the notation $Q \sim \chi^2_1$ should be read as 'the random variable Q follows a chi-squared distribution with one degree of freedom'.

2. If $Q_1 \sim \chi^2_p$ and $Q_2 \sim \chi^2_q$ are independent random variables, with p and q positive integers, then $Q_1 + Q_2 \sim \chi^2_{p+q}$.

The first of these results defines a chi-squared random variable with one degree of freedom. The second result tells us that when we combine independent chi-squared random variables we get another chi-squared random variable, and it tells us how to calculate the degrees of freedom too. One immediate consequence of these two results is that if Z_1,...,Z_n is a set of independent random variables, each with standard Normal marginal distributions,[30] then

$$Z_1^2 + \ldots + Z_n^2 \sim \chi^2_n. \tag{24}$$

This result is the foundation of inference about the population variance.[31] Let us turn now to the properties of the chi-squared distribution.

The density function of a chi-squared random variable is[32]

$$f(x) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{2}\right)^{\nu/2} x^{(\nu-2)/2} e^{-x/2}, \qquad \begin{array}{l} x \geq 0, \\ \nu > 0. \end{array}$$

We see from the density function that the random variable can only take non-negative values ($x > 0$), and that its distributions depends on the one parameter ν which is called the degrees of freedom.[33] Figure 6 illustrates the chi-squared distribution for various degrees of freedom. From Figure 6 we see that it is a uni-modal distribution which is skewed to the right; as the degrees of freedom increase there is greater probability in the right-hand tail of the distribution, so there is a greater chance of observing large values of the random variable. Also, the mode of the distribution moves to the right as the degrees of freedom increase, and the height of that mode decreases. As an side, observe how similar to a Normal distribution is the chi-squared distribution starting to look by the time that $\nu = 20$. Indeed, it can be shown that if $Q \sim \chi^2_\nu$ then the distribution

[29] The chi-squared distribution can be traced back to the work of Irénée-Jules Bienaymé (Bienaymé, 1838).

[30] When discussing the Normal distribution it was pointed out that marginal Normal distributions did not necessarily imply that a set of variables is jointly normally distributed. However, if we add the additional requirement that the variables are independent then it will be the case that they are also jointly normally distributed.

[31] Appendix B provides a derivation of the sampling distribution of s^2, where the result (24) is exploited; in particular, see equation (40).

[32] Recall the Gamma function discussed in Footnote 6. Note that the chi-squared distribution is a special case of the Gamma distribution, which can be traced back to the work of P. S. Laplace in 1836, although this is a topic for another time.

[33] It is common for the chi-squared distribution to be defined with ν a positive integer, i.e. $\nu = 1, 2, 3, \ldots$ Although this is the form in which it is most commonly encountered, everything is well-behaved provided that $\nu > 0$.

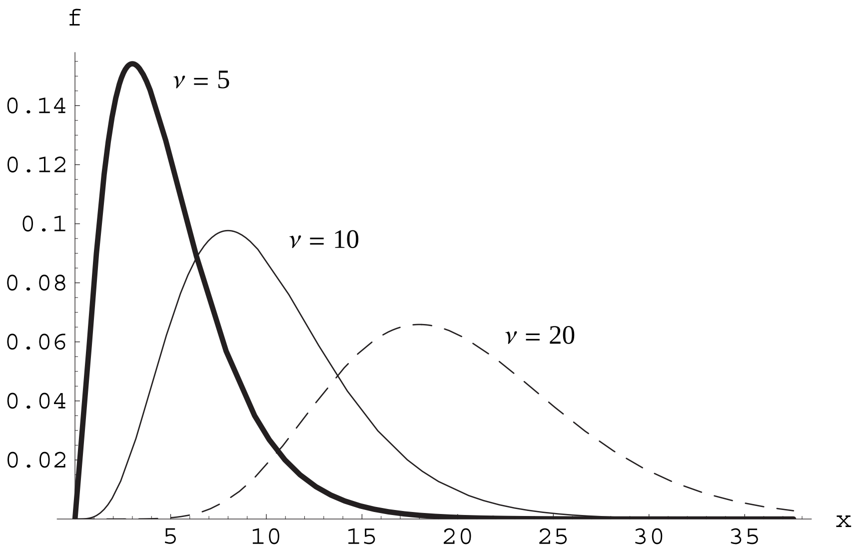


Figure 6: The Chi-Squared Distribution with ν Degrees of Freedom

of $(Q-\nu)/\sqrt{(2\nu)}$ looks increasingly like a standard Normal distribution as ν becomes larger.[34]

Our main use of the chi-squared distribution is in relation to the sampling distribution of s^2. The key result is

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}, \tag{25}$$

which is established in Appendix B. We can exploit this result to construct confidence intervals about σ^2 as follows. If a random variable $Q \sim \chi^2_\nu$ then we can always choose numbers $\theta_{\nu,\alpha/2}$ and $\theta_{\nu,1-\alpha/2}$, such that $P\left(Q \leq \theta_{\nu,\beta}\right) = \beta$ for $0 \leq \beta \leq 1$, to make the following statement true[35]

$$P\left(\theta_{\nu,\alpha/2} \leq Q \leq \theta_{\nu,1-\alpha/2}\right) = 1-\alpha. \tag{26}$$

So $\theta_{\nu,\alpha/2}$ and $\theta_{\nu,1-\alpha/2}$ are simply critical values from a chi-squared distribution which cut off lower-tail and upper-tail probabilities of $\alpha/2$, respectively. If equation (26) is true then so too is[36]

$$P\left(\frac{1}{\theta_{\nu,\alpha/2}} \geq \frac{1}{Q} \geq \frac{1}{\theta_{\nu,1-\alpha/2}}\right) = 1-\alpha,$$

or, on rearranging,

$$P\left(\frac{1}{\theta_{\nu,1-\alpha/2}} \leq \frac{1}{Q} \leq \frac{1}{\theta_{\nu,\alpha/2}}\right) = 1-\alpha.$$

[34]The transformation of Q used here should remind you of the standardizing transformation used in the central limit theorem that you will have seen in connection with the sample mean; namely, random variable minus its mean, all divided by the standard deviation of the random variable. Indeed, this result is simply another application of the central limit theorem.

[35]Observe that I am constructing a confidence interval with equal tail probabilities. There is no reason for this. Indeed, it is common to be interested in one-sided events of the form $P(Q > c)$.

[36]Note that when taking reciprocals the direction of the inequalities is reversed. For example, $5 < 6$ but $1/5 > 1/6$.

If we set $Q = (n-1)s^2/\sigma^2$, and $\nu = n-1$, then we obtain

$$P\left(\frac{1}{\theta_{n-1,1-\alpha/2}} \leq \frac{\sigma^2}{(n-1)s^2} \leq \frac{1}{\theta_{n-1,\alpha/2}}\right) = 1-\alpha,$$

or, on rearranging,

$$P\left(\frac{(n-1)s^2}{\theta_{n-1,1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\theta_{n-1,\alpha/2}}\right) = 1-\alpha.$$

So a $(1-\alpha)100\%$ confidence interval for σ^2 is[37]

$$\left[\frac{(n-1)s^2}{\theta_{n-1,1-\alpha/2}}, \frac{(n-1)s^2}{\theta_{n-1,\alpha/2}}\right].$$

For example, suppose that $n = 10$ and $s^2 = 1$. Convince yourself that a 90% confidence interval for σ^2 is approximately $[0.532, 2.707]$.[38]

5.2.1 An Aside About Degrees of Freedom

> *The notion of degrees of freedom is quite difficult and is not really well understood by many professionals. Here is one way of thinking about what is meant by degrees of freedom that you may find useful. Don't spend too much time worrying about degrees of freedom if you find this material difficult to follow.*

Degrees of freedom can be thought of as the number of pieces of information that you have available to learn about a population. Sometimes this equates to the number of observations that you have in a sample but not always. We will try to illustrate this idea with a very simple example and then show how it works in the context of estimating a variance.

A Simple Example Suppose that you are concerned with two random variables, X and Y say. Further suppose that I can observe sample values for the variables, x and y respectively. In general, there is no reason to believe that telling you the value of either x or y will convey any information about the value of the other. In particular, if you know x then y remains free to take any value that it wants. Equally, if you know y then x remains free, or unconstrained. We say that there are two degrees of freedom and the two observations x and y contain two pieces of information.[39]

Now suppose that I tell you that the two random variables are related according to $Y = 1 - X$.[40] In this case x and y are no longer both free, so they cannot contain two pieces of information. If I tell you the value of x then you can work out the value of y, namely $y = 1 - x$. And so, me revealing y to you will provide you with no additional

[37]Remember, it is the limits of the interval that are random, they are the functions of data through s^2. Consequently our confidence is in the ability of this method for constructing an interval (in repeated samples) to cover the true parameter value σ^2. Parameters, like σ^2, are characteristics of the population and so are not random!

[38]Hint: You will need to find the appropriate critical values; namely $\theta_{9,0.05} \approx 3.325$ and $\theta_{9,0.95} \approx 16.919$.

[39]More generally, if we had a sample of n observations that were chosen freely then we would have n degrees of freedom or n pieces of information.

[40]All that matters here is that we have a one-to-one mapping between the variables, so that knowledge of one implies knowledge of the other.

information over and above what you already have from knowing x. Equally, if I tell you the value of y then you can determine x according to $x = y + 1$, and so x provides no information in addition to that contained in y. In this case there is only a single piece of information contained in the two observations, or one degree of freedom. That is, only one observation can be chosen freely, with the other being given once the first is known.

5.2.2 Estimating a Variance (Intuition)

The big clue here is that the word 'variance' is actually an abbreviation for the expression 'variance about the mean'. In practice we never know the true population mean and so we replace it with a sample mean, e.g.,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

The original sample $X_1, \ldots, X_n$ contains n pieces of information but we use one piece of that information to estimate the mean. In the previous example we saw that it was the relationship $Y = 1 - X$ that meant knowledge of X (say) implied knowledge of Y, or vice verse. Here the relationship that consumes a piece of information is

$$\overline{X} = (X_1 + \ldots + X_n)/n.$$

Once we know $\overline{X}$, only $n - 1$ of the sample observations can be chosen freely, with the remaining observation then fixed, hence the remaining observation contains no additional information over what we already have. For example,

$$X_n = n\overline{X} - (X_1 + \ldots + X_{n-1})$$

or

$$X_3 = n\overline{X} - (X_1 + X_2 + X_4 + \ldots + X_{n-1}).$$

Given that we have to calculate a sample mean *before* we can calculate a sample variance it means that we can only be using the remaining $n-1$ pieces of information to estimate the population variance. In summary, a sample variance about the mean will be associated with $n - 1$ degrees of freedom.

The preceding argument provides some intuition as to why $(n-1)s^2/\sigma^2$ is distributed chi-squared *with* $n - 1$ *degrees of freedom* but doesn't constitute any sort of proof. A modern textbook is likely to require quite a bit of knowledge of matrix algebra to establish the degrees of freedom associated with a variance; such proofs seem relatively easy because they implicitly incorporate proofs of lots of other results. For those who are interested, Appendix B provides a derivation of the sampling distribution of s^2 that doesn't require matrix algebra. Beware! It is not for the faint-hearted.

5.3 Student's t Distribution

In our discussion of the Normal distribution we introduced the notion of a standardizing transformation where, for $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $Z = (X - \mu_X)/\sigma_X \sim \mathrm{N}(0, 1)$. The most important application of this result, from our perspective relate to the sampling distribution of the sample mean. If you have a simple random sample of size n ($X_1, \ldots, X_n$)

from a Normal population $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$ then $\overline{X} \sim \mathrm{N}(\mu_X, \sigma_X^2/n)$ and the standardizing transformation is

$$Z = \frac{\sqrt{n}\,(\overline{X} - \mu_X)}{\sigma_X} \sim \mathrm{N}(0,1). \tag{27}$$

Equation (27) is the foundation of all inference about the population mean, both confidence intervals and hypothesis testing, because it provides a probabilistic link between a sample statistic ($\overline{X}$) and the parameter of interest (μ). The only problem with this relationship is the presence of the *nuisance parameter* σ^2.[41] Simply put, if you don't know μ then you won't know σ^2 and so application of equation (27) is likely to be problematic.[42]

An obvious response to the presence of a nuisance parameter is to replace it with some estimator and an obvious estimator for σ^2 is the unbiased estimator s^2. The transformation

$$T = \frac{\sqrt{n}\,(\overline{X} - \mu_X)}{s} \sim t_{n-1} \tag{28}$$

is known as a *studentizing* transformation.[43] We see that the T statistic has a t distribution with $n-1$ degrees of freedom.[44] Equation (28) provides an operational procedure on which to base inference about the population mean μ. It works exactly as it does for the Normal distribution with known σ^2. For example, we can choose critical values from Student's t-distribution so that the following probability statement is true

$$P\left(\tau_{n-1,1-\alpha/2} \leq \frac{\sqrt{n}\,(\overline{X} - \mu_X)}{s} \leq \tau_{n-1,\alpha/2}\right) = 1 - \alpha,$$

where, in stark contrast to the corresponding treatment for the chi-squared distribution, I will define $\tau_{\nu,\beta}$ to be that value which cuts off an *upper*-tail probability of β from a t distribution with $n-1$ degrees of freedom; that is, if $T \sim t_\nu$ then $P(T \geq \tau_{\nu,\beta}) = \beta$.[45] Simple rearrangement yields

$$\begin{aligned}
1 - \alpha &= P\left(\tau_{n-1,1-\alpha/2} \leq \frac{\sqrt{n}\,(\overline{X} - \mu_X)}{s} \leq \tau_{n-1,\alpha/2}\right) \\
&= P\left(-\overline{X} + \frac{\tau_{n-1,1-\alpha/2}s}{\sqrt{n}} \leq -\mu_X \leq -\overline{X} + \frac{\tau_{n-1,\alpha/2}s}{\sqrt{n}}\right) \\
&= P\left(\overline{X} - \frac{\tau_{n-1,1-\alpha/2}s}{\sqrt{n}} \geq \mu_X \geq \overline{X} - \frac{\tau_{n-1,\alpha/2}s}{\sqrt{n}}\right) \\
&= P\left(\overline{X} - \frac{\tau_{n-1,\alpha/2}s}{\sqrt{n}} \leq \mu_X \leq \overline{X} - \frac{\tau_{n-1,1-\alpha/2}s}{\sqrt{n}}\right) \qquad (29)
\end{aligned}$$

[41] A nuisance parameter is simply one that is not of direct interest but which must be known in order to conduct inference about the parameter of interest.

[42] Formally we would say that Z is not *operational* without knowledge of σ^2.

[43] The transformation is named for William Sealy Gosset (1876–1937). Gosset was employed by the Guinness brewery which had stipulated that he not publish under his own name. He therefore wrote under the pen name 'Student.' He first derived the distribution of the t statistic in 1908. His interest in t statistics arose from the analysis of agricultural experiments conducted by the brewery. Indeed, much statistical development at this time was driven by sciences associated with agriculture, biology and genetics.

[44] Observe that I have used a capital letter T to denote the random variable. When we evaluate the statistic, on the basis of some observed sample, it will be denoted by a lowercase t, which is probably the symbol you are used to seeing.

[45] The reason for this vaguely odd definition will become clear in a moment, but reflects the way t critical values are commonly defined.

We might stop there but there is one further piece of information that we will have available and so we shall return to equation (29) shortly. In order to discover this final piece of information we need to explore the properties of Student's t distribution.

To see where Student's t distribution comes from reconsider the studentizing transformation (28). Dividing both numerator and denominator by σ, we can write

$$T = \frac{\sqrt{n}\,(\overline{X} - \mu_X)/\sigma}{s/\sigma} \sim t_{n-1}. \tag{30}$$

Now, we know that

$$Z = \frac{\sqrt{n}\,(\overline{X} - \mu_X)}{\sigma} \sim N(0,1).$$

From (25) we also know that

$$Q = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

The denominator of equation (30) can be written

$$\frac{s}{\sigma} = \sqrt{\frac{Q}{n-1}}.$$

Hence,

$$T = \frac{Z}{\sqrt{Q/(n-1)}} \sim t_{n-1}. \tag{31}$$

One further thing to note is that the standard Normal random Z and Q are independent random variables, which follows from the developments at the end of Appendix B.[46] This provides us with a general characterization of Student's t distribution:

> If $Z \sim N(0,1)$ and $Q \sim \chi^2_\nu$ are independent random variables then $T = Z/\sqrt{Q/\nu}$ has Student's t distribution with ν degrees of freedom, denoted t_ν.[47]

The density function for Student's t distribution with ν degrees of freedom is given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{(1+\nu^{-1}t^2)^{(\nu+1)/2}}, \qquad \begin{array}{l} -\infty < t < \infty, \\ \nu = 1,2,3,\ldots \end{array}$$

which is not very revealing, although we do see that, like the chi-squared distribution, Student's t-distribution depends on just one parameter, the degrees of freedom ν. Plots of the density function for various values of ν are presented in Figure 7. The main feature to observe is that, like the Normal distribution, Student's t is a uni-modal distribution which is symmetric about zero. This observation is what is required to complete our treatment of confidence intervals for the sample mean in equation (29). Because of this symmetry about zero it follows that the value which cuts off an upper-tail probability

[46] In the notation of the Appendix, $\overline{X}$ is a function of U_n whereas s^2 is a function of $U_1, \ldots, U_{n-1}$. Because the U_i are mutually independent, it follows that so too are Z and Q.

[47] I find it helpful to remember

$$T = \frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_\nu/\nu}} \sim t_\nu.$$

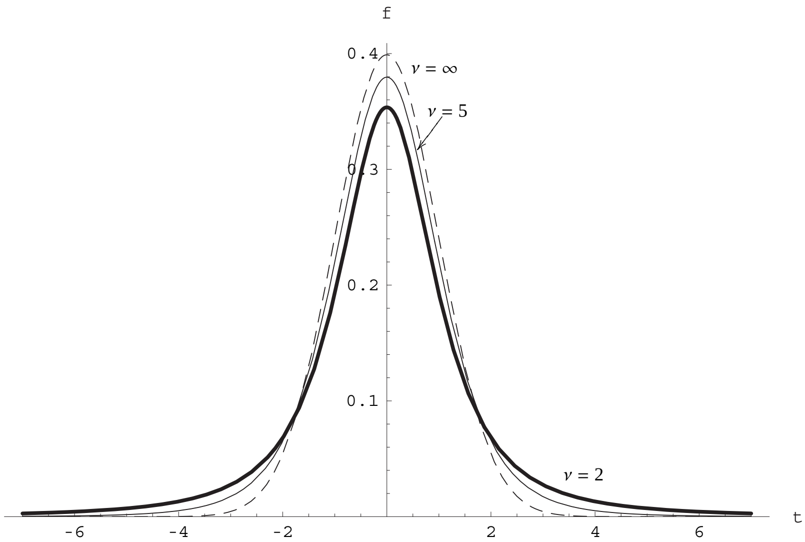


Figure 7: Various Student's t Density Functions

of $1 - \alpha/2$ is minus that value which cuts off an upper-tail probability of $\alpha/2$; that is, $\tau_{n-1,1-\alpha/2} = -\tau_{n-1,\alpha/2}$. If we make this substitution then we obtain[48]

$$P\left(\overline{X} - \frac{\tau_{n-1,\alpha/2}S}{\sqrt{n}} \leq \mu_X \leq \overline{X} + \frac{\tau_{n-1,\alpha/2}S}{\sqrt{n}}\right) = 1 - \alpha.$$

Student's t distribution has a couple of other interesting properties that deserve comment. First, as the degrees of freedom become large the density function looks increasingly like a standard Normal distribution which, in fact, is the limiting case when $\nu = \infty$. Figure 7 includes this limiting case which enables us to see that the height of the density at the mode increases with ν, so that the density of a standard Normal distribution is higher at zero (the mode of the random variable) than any Student's t distribution with finite degrees of freedom. Conversely, the tails of Students's t distribution become increasingly thick as the degrees of freedom reduce, which means that there is a greater probability of observing values a long way from the means than is the case for a random variable with a standard Normal distribution.[49] If $T \sim t_\nu$ then

$$\mathrm{E}[T] = 0, \quad \text{provided that } \nu > 1,$$
$$\mathrm{Var}[T] = \frac{\nu}{\nu - 2}, \quad \text{provided that } \nu > 2.$$

The requirements that there be sufficient degrees of freedom in order to define a mean or variance is something new. If these requirements are not met then these expectations are unbounded or infinitely large. We see that if $\nu = 1$ then not even the mean of the distribution is well defined. This special case of Student's t distribution is known as the Cauchy distribution and is sometimes encountered in finance theory, where its very fat

[48]The format of this statement is then in keeping with common practice where the critical value is that which cuts off the desired upper tail probability.

[49]If you like, imagine that there is a weight sitting on top of the density at the mode which becomes increasingly heavy as the degrees of freedom decrease; the density at the mode slowly collapses under this increasing weight. But the area under the curve must always be unity and so, as the height of the density at the mode reduces, the probability has to go somewhere. Clearly it gets pushed out towards the tails of the distribution, making them fatter.

tails are meant to capture the volatility sometimes seen in financial data. This is not very satisfactory and we will not pursue it here.[50]

5.4 Fisher's F Distribution

The final distribution that we will consider is Fisher's F distribution, also know as Snedecor's F distribution and the Fisher-Snedecor F distribution.[51] As it is most commonly encountered in the context of hypothesis testing that is how I will introduce it here.

Suppose that you have a sample $X_1, \ldots, X_n$ from a Normal population $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$ and that you wish to test the null hypothesis that $\mathcal{H}_0 : \mu = 0$ against the two-sided alternative $\mathcal{H}_1 : \mu \neq 0$. You could use a t statistic of the form $T = \sqrt{n}\,\overline{X}/s$ which, if the null hypothesis were true would have a t distribution with $n-1$ degrees of freedom. The evidence will support the null hypothesis if the observed value of T, denoted t, lies within the *acceptance region* or, conversely, the evidence will not support the null hypothesis if t lies in the *critical*, or rejection, region. If you wish to conduct the test at an α level of significance then the acceptance region is the set of all those values of T which satisfy $\left\{-\tau_{n-1,\alpha/2} \leq T \leq \tau_{n-1,\alpha/2}\right\}$. The critical region is then the union of two disjoint sets, namely $\left\{T < -\tau_{n-1,\alpha/2}\right\} \cup \left\{T > \tau_{n-1,\alpha/2}\right\}$.

The symmetry of Student's t distribution allows me to make these statements a little bit more simply. Define a new statistic $F = T^2$. Observe that, because it is a square $0 \leq F < \infty$. Every value in the acceptance region maps into one of the values in the set $\left\{0 \leq F \leq \tau^2_{n-1,\alpha/2}\right\}$. Equally, every value in the critical region maps into the set $\left\{F > \tau^2_{n-1,\alpha/2}\right\}$. This means that for testing $\mathcal{H}_0$ against $\mathcal{H}_1$ it makes no difference whether you use T as the test statistic or F, *you will always reach exactly the same conclusion.* The only factor that might currently enter your choice is that you have critical values for Student's t distribution but not for the distribution of F. Perhaps not surprisingly, F has Fisher's F distribution.

Let us look more carefully at the structure of the F statistic

$$F = T^2 = \left[\frac{\sqrt{n}\,\overline{X}}{s}\right]^2 = \left[\frac{\overline{X}/\left(\sigma/\sqrt{n}\right)}{s/\sigma}\right]^2 = \frac{\left[\overline{X}/\left(\sigma/\sqrt{n}\right)\right]^2}{s^2/\sigma^2}.$$

Looking first to the denominator, we know that

$$Q_2 = (n-1)s^2/\sigma^2 \sim \chi^2_{n-1}.$$

[50]The general result is that Student's t distribution possesses finite moments of order one less than its degrees of freedom. As a Normal distribution is the special case of an infinite number of degrees of freedom it possesses all moments, whereas the Cauchy has none.

[51]Professor Sir Ronald Aylmer Fisher, FRS, (1890-1962) is arguably the most important statistician of all. He made enormous contributions to the development almost all areas of statistics. A man of strong opinion, he was, amongst other things, a friend of Gosset's who was discussed in Footnote 43. Coincidently, he was married to a Guinness. Late in his life he took a position at the University of Adelaide. His ashes lie in St. Peter's Cathedral in Adelaide, where there is both a memorial pew and, on the floor beside it, a brass plaque. George W. Snedecor (1881–1974) was an American statistician who founded the first department of statistics in the USA, at the University of Iowa. The American Statistical Association awards the George W. Snedecor Award to a statistician for contribution in biometry, which is named in his honour.

That is, $s^2/\sigma^2 = Q_2/(n-1)$ can be thought of as a chi-squared random variable divided by its degrees of freedom. With this structure in mind we see that the numerator has the form

$$\left[\overline{X}/\left(\sigma/\sqrt{n}\right)\right]^2 = Z^2,$$

where $Z \sim \mathrm{N}(0,1)$. But $Z^2 \sim \chi_1^2$. If we write $Q_1 = Q_1/1 = Z^2 \sim \chi_1^2$ we see that the numerator can also be written as a chi-squared random variable divided by its degrees of freedom. Furthermore, we know these two chi-squared random variables to be independent as a consequence of the arguments presented following equation (31). With this structure in mind, we have the following characterization of Fisher's F distribution.

> If $Q_1 \sim \chi^2_{\nu_1}$ and $Q_2 \sim \chi^2_{\nu_2}$ are independent random variables then $F = (Q_1/\nu_1)/(Q_2/\nu_2)$ has Fisher's F distribution, denoted $F(\nu_1, \nu_2)$, with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom.[52]

It is not entirely obvious how our motivating example would lead to this general characterization of Fisher's F distribution. However, if attention is focussed on the denominator, it becomes clear that the statistic that most naturally gives rise to the F distribution is a ratio of variance estimators. So-called variance ratios occur commonly in the treatment of regression models so you can rest assured that you will see them again. For now, recognize that our motivating example was a special case where $T^2 \sim F(1, n-1)$.

Let us explore the distribution further. The density function for Fisher's F distribution is

$$g(f) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{f^{(\nu_1-2)/2}}{(1+(\nu_1/\nu_2)f)^{(\nu_1+\nu_2)/2}}, \qquad \begin{array}{c} f \geq 0, \\ \nu_1, \nu_2 = 1, 2, 3, \ldots \end{array}$$

and the mean and variance of the distribution are given by

$$\begin{aligned} \mathrm{E}\left[F\right] &= \frac{\nu_2}{\nu_2 - 2}, \quad \text{provided that } \nu_2 > 2, \\ \mathrm{Var}\left[F\right] &= \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}, \quad \text{provided that } \nu_2 > 4. \end{aligned}$$

There is not an awful lot that can be gained from inspection of these formulae and so we will learn more from examination of Figure 8. Prior to doing that, however, it is worth pointing out a couple of features of these results. The most obvious feature is that Fisher's F distribution is a two parameter distribution, depending on both the numerator and the denominator degrees of freedom. It is important that you know which is which because the distribution is not symmetric in these parameters — mess them up and you mess everything up! Also, it is somewhat surprising to observe that the mean of the distribution is not affected by ν_1. For $\nu_2 < 2$ the mean does not exist (it is unbounded) and as ν_2 increases the mean approaches unity. Similarly, the variance of the distribution is largely controlled by the behaviour of the denominator of the variance ratio. This is quite a common characteristic of sampling distributions of ratios. A more important property of Fisher's F distribution stems from the following remarkable observation. Suppose that a random variable $F \sim F(\nu_1, \nu_2)$. Now suppose that we are interested in a random variable of the form $G = 1/F$. Observe that if

[52] I find it helpful to remember

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} \sim F(\nu_1, \nu_2).$$

$F = (Q_1/\nu_1)/(Q_2/\nu_2)$ then $G = (Q_2/\nu_2)/(Q_1/\nu_1) \sim F(\nu_2, \nu_1)$. Consequently we see that, if $F \sim F(\nu_1, \nu_2)$,

$$P(F < c) = P\left(\frac{1}{F} > \frac{1}{c}\right) = P\left(G > \frac{1}{c}\right),$$

where $G \sim F(\nu_2, \nu_1)$ and c is just some constant. Hence, critical values that cut off certain *lower*-tail probabilities in an $F(\nu_1, \nu_2)$ distribution can be obtained as critical values which cut off *upper*-tail probabilities in an $F(\nu_2, \nu_1)$ distribution. For this reason many tables of Fisher's F distribution only provide critical values for *upper*-tail probabilities. Of course, this device is becoming less important as computer packages incorporate functions which will calculate any probabilities from an F distribution that you may desire.

Recall from Footnote 6 that the relationship

$$\mathrm{B}(\nu_1, \nu_2) = \frac{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}$$

defines a beta function. Not surprisingly, it can be shown that Fisher's F distribution is intimately related to the Beta distribution but that, like so many things, remains a story for another day.

Let us turn to Figure 8, which is comprised of four pairs of graphs labelled (a)–(d), respectively. Each pair of graphs corresponds to a given denominator degrees of freedom; namely, $\nu_2 = 5, 10, 20, 50$, respectively. Within a given pair of graphs the numerator degrees of freedom are varied. In each case the left-side graph plots the density function for values $\nu_1 = 1, 2, 3$ and the right-side graph plots the density function for values $\nu_1 = 3, 5, 20, 50$.[53] The patterns that emerge are remarkably similar as the parameters vary.

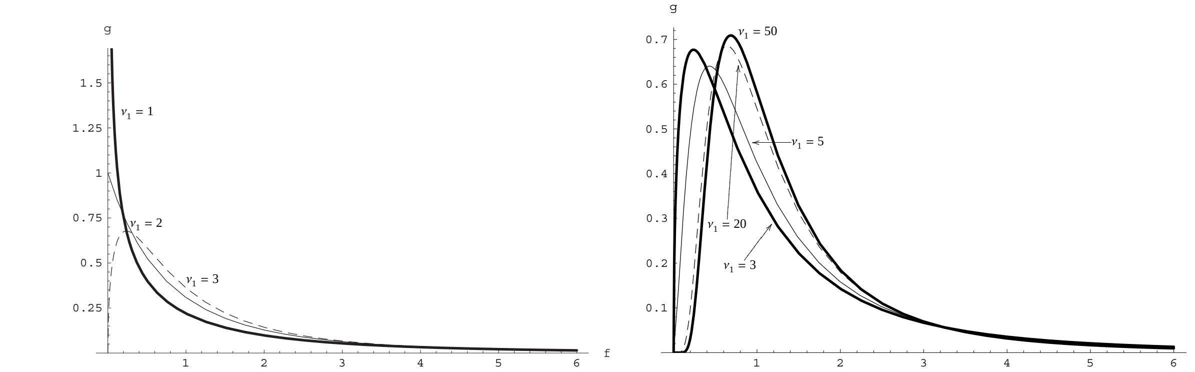
1. The density function is always uni-modal and skewed to the right. Indeed, as ν_1 increases the density function looks increasingly similar in shape to a chi-square density function.[54]

2. If one compares graphs for given value of $\nu_1 > 2$ as ν_2 varies we see that the height of the density at the mode increases with ν_2. For $\nu_1 \leq 2$, the height of the density at the mode doesn't change with ν_2, occurring at $f = 0$ in both cases. When $\nu_1 = 1$ the density function diverges, whereas for $\nu_1 = 2$ density is everywhere bounded with a mode of unity.

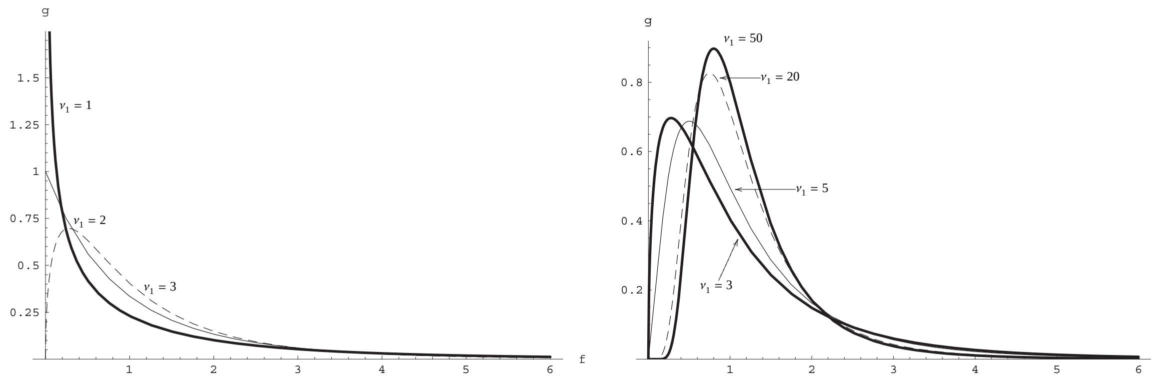
3. The shape of the density is monotonically decreasing in f for $\nu_1 \leq 2$. For $\nu_1 > 2$ the mode of the distribution occurs at an interior point of the support of f, moving further to the right as ν_1 increases. Note that the height of the density at the mode initially decreases as $\nu_1 = 1, 2, 3$ but then begins to increase with ν_1. The value of ν_1 at which the height of the density at the mode starts to increase gets smaller as ν_2 increases. So, for example, for $\nu_2 = 5$ the height of the density at the mode for $\nu_1 = 5$ is less than that for $\nu_1 = 3$ whereas the reverse is true when $\nu_2 = 50$.

[53] Note that the scales of the different graphs differ and so $\nu_1 = 3$ is included in both graphs of each pair to make it easier to compare the results.

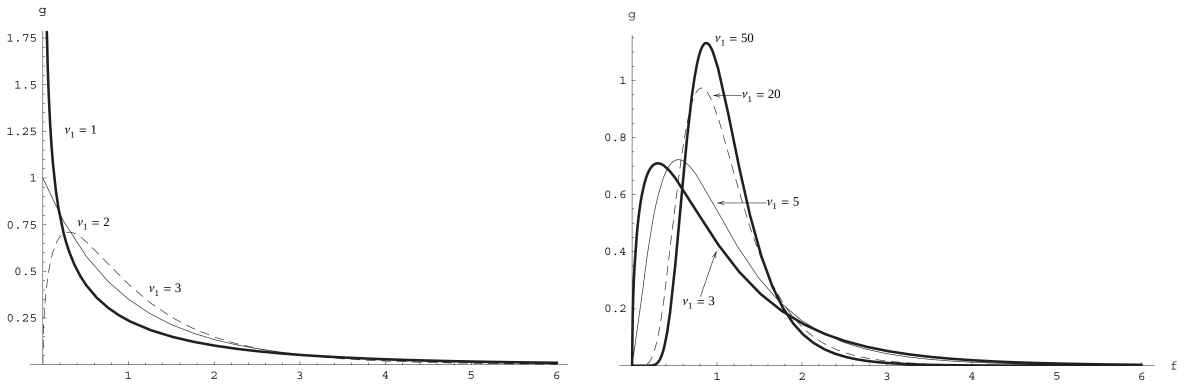
[54] It can actually be shown that if $F \sim F(\nu_1, \nu_2)$ then as $\nu_2 \to \infty$, then density function of $\nu_1 F$ becomes increasing like that of $\chi^2_{\nu_1}$. If you think about it, this is essentially the same result as Student's t distribution with an infinite degrees of freedom approaching a standard Normal distribution.



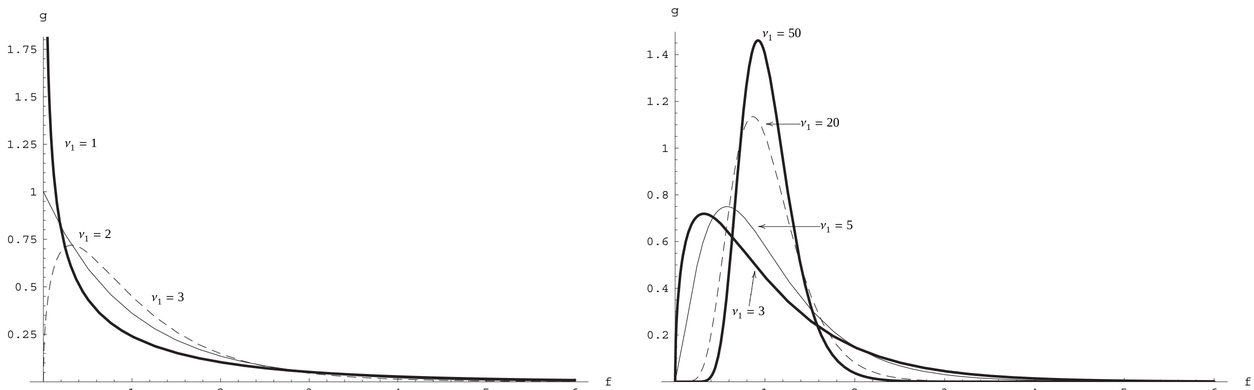
(a) $\nu_2 = 5$



(b) $\nu_2 = 10$



(c) $\nu_2 = 20$



(d) $\nu_2 = 50$

Figure 8: Various Fisher's F Density Functions

In summary, Fisher's F distribution mostly looks something like a chi-squared distribution, although it is not the same. However, the shape can be dramatically affected by the relative magnitudes of the numerator and denominator degrees of freedom. The most atypical density shapes correspond to low numerator degrees of freedom which are, in fact, the situations most likely to be encountered in practice.

6 Normality and the Exponential Family

The exponential family of distributions holds a special place in statistical theory because of a set of very general results that are applicable to all members of the family. While we won't pursue any of these results in this handout, it is not unreasonable to explore the exponential family here because, as luck would have it, the distributions encountered here fall into this very important family. Let us begin with a definition.

Definition (The Exponential Family)**.** The (one-parameter) exponential family (or class) of densities includes any density function that can be written in the general form

$$f(x;\theta) = a(\theta)b(x)\exp\{c(\theta)d(x)\}$$

for $-\infty < x < \infty$, for all $\theta \in \Theta$ (whatever Θ may be), and for suitable choices of functions $a(\cdot)$, $b(\cdot)$, $c(\cdot)$, and $d(\cdot)$. The extension to the k-parameter exponential family is those densities $f(\cdot;\theta_1,\ldots,\theta_k)$ that can be expressed as

$$f(\cdot;\theta_1,\ldots,\theta_k) = a(\theta_1,\ldots,\theta_k)b(x)\exp\left\{\sum_{j=1}^{k} c_j(\theta_1,\ldots,\theta_k)d_j(x)\right\}$$

for suitable choices of functions $a(\theta_1,\ldots,\theta_k)$, $b(x)$, $c_j(\theta_1,\ldots,\theta_k)$, and $d_j(\theta_1,\ldots,\theta_k)$, for $j = 1,\ldots,k$. □

Let us look at some examples of members of this family of distributions. First, the exponential distribution, which is a one-parameter distribution. Here, $f(x;\theta) = \theta e^{-\theta x}$, with $x \geq 0$ and $\theta > 0$. To see that this belongs to the exponential family of distributions set $a(\theta) = \theta$, $b(x) = I_{[0,\infty)}(x)$, $c(\theta) = -\theta$, and $d(x) = x$.[55] As a second example, consider the Poisson distribution with density function

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,3,\ldots;\lambda > 0.$$

Here $a(\lambda) = e^{-\lambda}$ and $b(x) = I_{\{0,1,2,3,\ldots\}}(x)/x!$. Observe that $\lambda^x = \exp\{\log\lambda^x\} = \exp\{x\log\lambda\}$, and so we see that $c(\lambda) = \log\lambda$ and $d(x) = x$. Our final example is, of course, that of the Normal density. Here the support of the random variable is $-\infty < x < \infty$ and the parameter support is $-\infty < \mu < \infty$ and $\sigma^2 > 0$, and the density function is

$$\begin{aligned} f(x;\mu,\sigma^2) &= (2\pi\sigma^2)^{-1/2}\exp\left\{-\tfrac{1}{2\sigma^2}(x-\mu)^2\right\} \\ &= (2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}\exp\left\{\frac{\mu x}{\sigma^2}\right\}. \end{aligned}$$

[55]Note that $I_{[0,\infty)}(x)$ is an indicator function that takes the value unity of $x \geq 0$ and zero otherwise. In general, the functions are of the form $I_{condition}(argument)$ and they take the value unity when the argument satisfies the condition and zero otherwise.

This is slightly more complicated that the previous two examples because here we have a two parameter problem. Nevertheless, it is straightforward to see that

$$a(\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}, \; b(x) = 1,$$
$$c_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2}, \; c_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}, \; d_1(x) = x^2, \text{ and } d_2(x) = x.$$

Note that because there are two parameters, μ and σ^2, there are two c functions and two d functions. The chi-squared distribution is also a member of the exponential family (convince yourself that this is so), as are the Pareto distributions, but the t distribution is not, nor is either the F distribution or the hypergeometric distribution. So we see that the exponential family is applicable in many circumstances but certainly not all.

Remark. The really important feature of the exponential family of distributions is that the functions

$$\sum_{i=1}^{n} d_1(X_i), \ldots, d_k(X_i)$$

constitutes a set of jointly sufficient statistics for $\theta_1, \ldots, \theta_k$.

Sets of jointly sufficient statistics contain the same amount of information about the parameters as does the original sample. (Another way of saying this is that the original sample is itself a set of jointly sufficient statistics.) The obvious attraction is that sufficiency provides as guide to data reduction in that we may be able to find sets of statistics that have smaller dimension than the sample size but which contain as much information about the parameters as does the sample itself. Indeed, the smallest set of statistics that are sufficient for the parameters are called *minimal sufficient statistics.* As alluded to at the beginning of this section, we won't be exploring sufficient statistics at any length but they play an important role in the classical theories of estimation and hypothesis testing. (It should be clear that arguments can be constructed suggesting that efficient estimators or powerful test statistics should be functions of minimal sufficient statistics, if possible. Moreover, estimators or test statistics not satisfying this criterion can quite possibly be dominated, in some sense, by procedures that do. But we leave such discussion for another time.)

7 Final Remark

It is important that you have a feel for what the Normal and related distributions entail because, hereafter, they will be central to almost everything that we do.

Bibliography

Bienaymé, I.-J. (1838). Sur la probabilitié des résultat moyens des observations; demonstration directe de la règle de Laplace. *Mémoires Des Savants Étrangers, Académie des Sciences, Paris* 513–558. 25

Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distributions*. Chapman & Hall/CRC, Boca Raton. 2

Henderson, H. V. and S. R. Searle (1981). On deriving the inverse of a sum of matrices. *SIAM Review* 23(1), 53–60, ISSN 00361445. 24

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions*, volume 1. John Wiley & Sons, Inc., New York, second edition. 2

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions*, volume 2. John Wiley & Sons, Inc., New York, second edition, ISBN 0-471-58494-0. 2

Johnson, N. L., S. Kotz, and N. Balakrishnan (1997). *Discrete Multivariate Distributions*. John Wiley & Sons, Inc., New York. 2

Johnson, N. L., S. Kotz, and N. Balakrishnan (2000). *Continuous Multivariate Distributions. Volume 1: Models and Applications*. John Wiley & Sons, Inc., New York, second edition. 2

Johnson, N. L., S. Kotz, and A. W. Kemp (2005). *Univariate Discrete Distributions*. John Wiley & Sons, Inc., Hoboken, New Jersey, third edition. 2

Mood, A. M., F. A. Graybill, and D. C. Boes (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, Inc., New York, third edition. 2

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, Inc., New York. 2

Ouellette, D. V. (1981). Schur complements and statistics. *Linear Algebra and Its Applications* 36, 187——295. 24

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley & Sons, Inc., New York, second edition. 2

Rao, C. R. and S. K. Mitra (1971). *Generalized Inverse of Matrices and Its Applications*. John Wiley & Sons, Inc., New York. 48

Rose, C. and M. D. Smith (2002). *Mathematical Statistics with Mathematica*. Springer-Verlag, New York. 8

Searle, S. R. (1982). *Matrix Algebra Useful For Statistics*. John Wiley & Sons, New York. 24, 46

Stuart, A. and J. K. Ord (1987). *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory*. Charles Griffin & Company, London, fifth edition. 2, 7

Wackerly, D. D., W. Mendenhall, and R. L. Scheaffer (2008). *Mathematical Statistics With Applications*. Thomson Brooks/Cole, Belmont, CA, 7th edition. 2

A The Normal Distribution and Moment Generating Functions

Moment generating functions are a device that allows one to do a lot of different sorts of calculations under certain conditions. They are beyond the scope of this subject and are presented only for completeness.

A.1 Moment Generating Functions for Normal Random Vectors

If $Z = [Z_1, \ldots, Z_p]'$ denotes a p-vector of independent standard Normal random variables, so that $Z \sim \mathrm{N}(0, I_p)$, then, as was shown in lectures, they each have moment generating function

$$\Phi_{Z_i}(t_i) = \mathrm{E}\left[e^{t_i Z_i}\right] = \exp\left\{\tfrac{1}{2}t_i^2\right\}, \text{ valid for all real } t_i, \quad i = 1, \ldots, p.$$

The joint moment generating function of Z is then

$$\Phi_Z(t) = \prod_{i=1}^{p} \Phi_{Z_i}(t_i) = \exp\left\{\tfrac{1}{2}\sum_{i=1}^{p} t_i^2\right\} = e^{t't/2}, \text{ valid for all real } t = [t_1, \ldots, t_p]',$$

where the first equality follows from the independence of the elements of Z. Armed with this information, it is possible to build up a moment generating function for X.

We know, that we can add a covariance structure by pre-multiplying Z by Ω, e.g. $W = \Omega Z \sim \mathrm{N}(\mathbf{0}, \Omega\Omega')$. Provided that Ω has full row rank then $\Omega\Omega' > 0$. We will assume that there exists an Ω such that $\Omega\Omega' = \Sigma$.[56] In order to obtain the moment generating function for W observe that

$$\Phi_W(t) = \Phi_{\Omega Z}(t) = \mathrm{E}_Z\left[e^{t'(\Omega Z)}\right] = \mathrm{E}_Z\left[e^{(\Omega' t)' Z}\right] = e^{(\Omega' t)'(\Omega' t)/2} = e^{t'\Sigma t/2}.$$

We know that if we add a constant to a random variable then its moment generating function must be multiplied by a constant amount. For example, suppose that we add c_i to each element of Z, according to

$$U_i = Z_i + \mu_i, \quad i = 1, \ldots, p.$$

Then, the moment generating function of each U_i is

$$\Phi_{U_i}(t_i) = \mathrm{E}_{Z_i}\left[e^{t_i U_i}\right] = \mathrm{E}_{Z_i}\left[e^{t_i(Z_i + c_i)}\right] = e^{t_i c_i}\,\mathrm{E}_{Z_i}\left[e^{t_i Z_i}\right] = e^{t_i c_i}\Phi_{Z_i}(t_i), \quad i = 1, \ldots, p.$$

Moreover, the joint moment generating function of $\mathbf{U} = [U_1, \ldots, U_p]'$ is then

$$\Phi_{\mathbf{U}}(t) = e^{t'c}\Phi_Z(t) = \exp\left\{t'c + t't/2\right\}.$$

Clearly, if we define $X = W + \mu$, for some p-vector μ, then we know that $X \sim \mathrm{N}(\mu, \Sigma)$ and we also know that the corresponding moment generating function is

$$\Phi_X(t) = \exp\left\{t'\mu + t'\Sigma t/2\right\}. \tag{32}$$

[56] There is actually a number of different definitions of Ω that will fit this bill. For now, the exact definition doesn't matter. All that matters is the existence of such an Ω, which we will take as a given.

A.2 Marginals Via Moment Generating Functions

The device that we used of replacing the dummy vector t by something that is more convenient is extremely powerful and will be the approach adopted here. In particular, let

$$B = \begin{bmatrix} \mathbf{I}_{p_1} & \mathbf{0} \\ -\Sigma_{21}\Sigma_{11}^{-1} & \mathbf{I}_{p_2} \end{bmatrix},$$

where definitions of X, Σ and p are as given in previous sections. Then

$$\Phi_{BX}(t) = \mathrm{E}_X\left[e^{t'BX}\right] = \exp\left\{t'B\mu + t'B\Sigma B't/2\right\}.$$

Observe that

$$BX = \begin{bmatrix} X_1 \\ X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1 \end{bmatrix}, \quad B\mu = \begin{bmatrix} \mu_1 \\ \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1 \end{bmatrix}, \quad \text{and} \quad B\Sigma B' = \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22\cdot 1} \end{bmatrix}.$$

Partitioning t conformably with X, we see that

$$\begin{aligned} \Phi_{BX}(t) &= \Phi_{X_1, X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1} \\ &= \exp\left\{t_1'\mu_1 + t_2'(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1) + t_1'\Sigma_{11}t_1 + t_2'\Sigma_{22\cdot 1}t_2\right\} \\ &= \exp\left\{t_1'\mu_1 + t_1'\Sigma_{11}t_1\right\} \times \exp\left\{t_2'(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1) + t_2'\Sigma_{22\cdot 1}t_2\right\} \\ &= \Phi_{X_1}(t_1)\Phi_{X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1}(t_2). \end{aligned} \tag{33}$$

We see that $X_1 \sim \mathrm{N}(\mu_1, \Sigma_{11})$ and $X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1 \sim \mathrm{N}\left(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22\cdot 1}\right)$ are independent random vectors. Because of this independence, it follows that the conditional distribution of $X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1$ given X_1 is the same as its marginal distribution. By implication, the respective moment generating functions must also be the same. We can then recover the moment generating function of X_2 given X_1 by transformation, i.e. adding $\Sigma_{21}\Sigma_{11}^{-1}X_1$ to $X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1$. That is

$$\begin{aligned} \Phi_{X_2|X_1}(t_2) &= \Phi_{[(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1) + \Sigma_{21}\Sigma_{11}^{-1}X_1]|X_1}(t_2) \\ &= \exp\{\Sigma_{21}\Sigma_{11}^{-1}X_1\}\Phi_{X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1}(t_2) \\ &= \exp\left\{t_2'(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)) + t_2'\Sigma_{22\cdot 1}t_2\right\}, \end{aligned} \tag{34}$$

which implies that $X_2 \mid X_1 \sim \mathrm{N}\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22\cdot 1}\right)$. The two densities obtained via moment generating functions — being the marginal density for X_1 given in (33) and the conditional density for X_2 given X_1 given in (34) — can be seen to be the same as those in (16), obtained from working with the joint density function directly.

B An Old-Fashioned Derivation of the Sampling Distribution of s^2

The discussion of this appendix is only for those who are interested. It is written to be read but will require some work on the part of the reader. Please feel free to try it but if it is not to your taste then move on to something else.

B.1 Introduction

As suggested by the title, the aim of this appendix is to derive the sampling distribution of s^2. The derivation shall be done in a series of small steps. First, we will transform the sample observations $X_1, \ldots, X_n$ into a new set of variables $U_1, \ldots, U_n$; this transformation is known as Helmert's transformation. This is done so that we can work with a set of random variables that have particular properties. The next step of our development is to explore the properties of $U_1, \ldots, U_n$. Once we have done that we show how s^2 can be written in terms of the U_i and then exploit that result to the sampling distribution of s^2.

B.2 Helmert's Transformation

First, suppose that we have a simple random sample of n observations $X_1, \ldots, X_n$ from a population $X \sim \mathrm{N}(\mu, \sigma^2)$.[57] Now, without discussing why it might be a good idea, define

[57]Remember that the important feature of a simple random sample is that the observations are independent of one another.

a new set of variables according to[58]

$$\left.\begin{aligned}
U_1 &= (X_1 - X_2)/\sqrt{2} \\
U_2 &= (X_1 + X_2 - 2X_3)/\sqrt{6} \\
U_3 &= (X_1 + X_2 + X_3 - 3X_4)/\sqrt{12} \\
\ldots &\qquad \ldots \qquad \ldots \qquad \ldots \\
U_{n-1} &= (X_1 + \ldots + X_{n-1} - (n-1)X_n)/\sqrt{n(n-1)} \\
U_n &= (X_1 + \ldots + X_n)/\sqrt{n}
\end{aligned}\right\} \tag{35}$$

Consider first the term U_{n-1}, which has a structure typical of $U_1, \ldots, U_{n-1}$.

$$\begin{aligned}
\mathrm{E}\left[U_{n-1}\right] &= \left(\mathrm{E}\left[X_1\right] + \ldots + \mathrm{E}\left[X_{n-1}\right] - (n-1)\,\mathrm{E}\left[X_n\right]\right)/\sqrt{n(n-1)} \\
&= (\underbrace{\mu + \ldots + \mu}_{(n-1)\text{ terms}} - (n-1)\mu)/\sqrt{n(n-1)} \\
&= 0 \\
\mathrm{Var}\left[U_{n-1}\right] &= \mathrm{Var}\left[\left(X_1 + \ldots + X_{n-1} - (n-1)X_n\right)/\sqrt{n(n-1)}\right] \\
&= \big(\mathrm{Var}\left[X_1\right] + \ldots + \mathrm{Var}\left[X_{n-1}\right] + (n-1)^2\,\mathrm{Var}\left[X_n\right]\big)/[n(n-1)] \\
&= (\underbrace{\sigma^2 + \ldots + \sigma^2}_{(n-1)\text{ terms}} + (n-1)^2\sigma^2)/[n(n-1)] \\
&= \frac{(n-1)^2 + (n-1)}{n(n-1)}\,\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

Note that the second equality in the development of the variance exploits the fact that covariances amongst as set of mutually independent random variables are zero.[59] The only term that looks at all different is the term U_n for which:

$$\begin{aligned}
\mathrm{E}\left[U_n\right] &= \mathrm{E}\left[\sqrt{n}\,\overline{X}\right] = \sqrt{n}\,\mu \\
\mathrm{Var}\left[U_n\right] &= \mathrm{Var}\left[\sqrt{n}\,\overline{X}\right] = n \times \frac{\sigma^2}{n} = \sigma^2,
\end{aligned}$$

where we have exploited known properties of the mean of a simple random sample.[60]

The next thing to observe is that each of the U_i can be written as a weighted sum of Normal random variables; that is,

$$U_i = c_{i,1}X_1 + c_{i,2}X_2 + \ldots + c_{i,n}X_n, \qquad i = 1, \ldots, n. \tag{36}$$

For example, for U_1, we see that $c_{11} = 1/\sqrt{2}$, $c_{12} = -1/\sqrt{2}$, and $c_{13} = \ldots = c_{1n} = 0$. Similarly, for U_n, $c_{n1} = \ldots = c_{1n} = 1/\sqrt{n}$, and so on. Hence, from the properties of the Normal distribution, each of the U_i are also normally distributed.

[58]This transformation can be dated back to the work of F. R. Helmert in 1876. In case it matters to you, it is easy to show that the Jacobian of this transformation is unity. If you don't know about Jacobians of transformation then just forget that I even mentioned it.

[59]There has also been a couple of applications of the rule $\mathrm{Var}\left[\alpha X\right] = \alpha^2\,\mathrm{Var}\left[X\right]$, but you knew that already.

[60]These properties are, of course, that if $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^2, and if $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$, then $\mathrm{E}\left[\overline{X}\right] = \mu$ and $\mathrm{Var}\left[\overline{X}\right] = n^{-1}\sigma^2$.

We need to establish one more result. Before actually working it out, let's see an example of it in action. Consider the covariance between U_1 and U_2:

$$\begin{aligned}\operatorname{Cov}[U_1, U_2] &= \operatorname{Cov}\left[(X_1 - X_2)/\sqrt{2}, (X_1 + X_2 - 2X_3)/\sqrt{6}\right] \\ &= \frac{1}{\sqrt{2} \times \sqrt{6}} \left(\operatorname{Var}[X_1] + \operatorname{Cov}[X_1, X_2] - 2\operatorname{Cov}[X_1, X_3]\right. \\ &\qquad \left. - \operatorname{Cov}[X_2, X_1] - \operatorname{Var}[X_2] + 2\operatorname{Cov}[X_2, X_3]\right).\end{aligned}$$

Noting that all of the covariance terms are zero, from the independence of the X_i, and that $\operatorname{Var}[X_1] = \operatorname{Var}[X_2] = \sigma^2$

$$\operatorname{Cov}[U_1, U_2] = \frac{1}{\sqrt{12}} \left(\operatorname{Var}[X_1] - \operatorname{Var}[X_2]\right) = 0.$$

And so we see that, because they are normally distributed, U_1 and U_2 are independent.[61] The result that we want to establish is that the U_i are all mutually independent; that is, $\operatorname{Cov}[U_i, U_j] = 0$, $1 \le i < j \le n$.

Our starting point will be equation (36), so that

$$\operatorname{Cov}[U_i, U_j] = \operatorname{Cov}\left[c_{i,1}X_1 + c_{i,2}X_2 + \ldots + c_{i,n}X_n, c_{j,1}X_1 + c_{j,2}X_2 + \ldots + c_{j,n}X_n\right].$$

Recognizing, yet again, that all of the covariance terms between the X_i are zero and that all the variances are equal to σ^2, we obtain

$$\begin{aligned}\operatorname{Cov}[U_i, U_j] &= c_{i,1}c_{j,1}\operatorname{Var}[X_1] + c_{i,2}c_{j,2}\operatorname{Var}[X_2] + \ldots + c_{i,n}c_{j,n}\operatorname{Var}[X_n]. \\ &= \left(c_{i,1}c_{j,1} + c_{i,2}c_{j,2} + \ldots + c_{i,n}c_{j,n}\right)\sigma^2. \end{aligned} \tag{37}$$

If we look back to (35) we see that our remaining coefficients have a remarkable structure which enables us to evaluate this covariance. For notational clarity we shall distinguish between two cases, (i) $i < n-1$ and (ii) $i = n-1$ (which implies that $j = n$).

Case 1: $i < n-1$ In this case $c_{i,(i+2)} = \ldots = c_{i,n} = 0$, so that equation (37) reduces to

$$\operatorname{Cov}[U_i, U_j] = \left(c_{i,1}c_{j,1} + c_{i,2}c_{j,2} + \ldots + c_{i,i+1}c_{j,i+1}\right)\sigma^2.$$

Now $c_{j,1} = \ldots = c_{j,(i+1)} = 1/\kappa_j$, where

$$\kappa_j = \begin{cases}\sqrt{j(j+1)}, & \text{if } j < n, \\ \sqrt{n}, & \text{if } j = n,\end{cases}$$

which provides a further reduction to

$$\operatorname{Cov}[U_i, U_j] = \frac{1}{\kappa_j}\left(c_{i,1} + c_{i,2} + \ldots + c_{i,i+1}\right)\sigma^2.$$

Recognizing that

$$c_{i,k} = \begin{cases}1/\sqrt{i(i+1)}, & k = 1, \ldots, i, \\ -i/\sqrt{i(i+1)}, & k = i+1,\end{cases}$$

we obtain

$$\operatorname{Cov}[U_i, U_j] = \frac{1}{\kappa_j\sqrt{i(i+1)}}\left(\underbrace{1 + 1 + \ldots + 1}_{i \text{ terms}} - i\right)\sigma^2 = 0,$$

as required.

[61] If they weren't normally distributed a zero covariance would not imply independence.

Case 2: $i = n-1$ This case is distinguished from the previous one simply because none of the coefficients are equal to zero. Here $j = n$, so we need $c_{n,1} = \ldots = c_{n,n} = 1/\sqrt{n}$, and, as $i = n-1$, we also need

$$c_{n-1,k} = \begin{cases} 1/\sqrt{n(n-1)}, & k = 1, \ldots, n-1, \\ -(n-1)/\sqrt{n(n-1)}, & k = n. \end{cases}$$

If we make these substitutions then equation (37) becomes

$$\mathrm{Cov}\left[U_{n-1}, U_n\right] = \frac{1}{\sqrt{n} \times \sqrt{n(n-1)}} \left(\underbrace{1 + 1 + \ldots + 1}_{n-1 \text{ terms}} - (n-1) \right) \sigma^2 = 0,$$

as required.

In summary, we have established that the $U_1, \ldots, U_n$ are mutually independent because they have zero covariance and are normally distributed. Transformations like Helmert's transformations are called *orthogonal* transformations. They are characterized by two conditions on the coefficients in equation (36):

$$\begin{aligned} \sum_{i=1}^{n} c_{j,i}^2 &= 1, \qquad j = 1, \ldots, n, \\ \sum_{i=1}^{n} c_{j,i} c_{k,i} &= 0, \qquad j, k = 1, \ldots, n. \end{aligned} \tag{38}$$

One feature of such transformations is that they generate independent sets of variables while preserving variance. If you look back at how we established the variance and covariance results, the derivations essentially involved establishing that these conditions (38) hold. Other orthogonal transformations exist, for which the weights $c_{j,i}$ differ but for which these two conditions will still hold.

B.3 The Sample Variance

The point of the previous section becomes clear once we recognize that

$$S(n) = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n-1} U_i^2, \quad \text{where} \quad \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{39}$$

There is no easy way of showing this, so here goes. We shall establish the result by induction. First, we will show that $S(n) - S(n-1) = U_{n-1}^2$. Second, we shall show that $S(2) = U_1^2$, where $n = 2$ is the smallest sample size for which $S(n) > 0$. In combination these two facts establish the desired result.

To begin, write

$$U_{n-1} = \frac{1}{\sqrt{n(n-1)}} \left(\sum_{i=1}^{n} X_i - nX_n \right) = \frac{n}{\sqrt{n(n-1)}} \left(\overline{X}_n - X_n \right),$$

Then

$$U_{n-1}^2 = \frac{n}{n-1} \left(\overline{X}_n^2 - 2X_n \overline{X}_n + X_n^2 \right).$$

Next,

$$\begin{aligned}
S(n)-S(n-1) &= \sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2-\sum_{i=1}^{n-1}\left(X_i-\overline{X}_{n-1}\right)^2 \\
&= \left(\sum_{i=1}^{n}X_i^2-n\overline{X}_n^2\right)-\left(\sum_{i=1}^{n-1}X_i^2-(n-1)\overline{X}_{n-1}^2\right) \\
&= X_n^2-\left(n\overline{X}_n^2-(n-1)\overline{X}_{n-1}^2\right) \\
&= X_n^2+\left(n\overline{X}_n^2-2nX_n\overline{X}_n+X_n^2\right)/(n-1) \\
&= \frac{n}{n-1}\left(\overline{X}_n^2-2X_n\overline{X}_n+X_n^2\right) \\
&= U_{n-1}^2,
\end{aligned}$$

as required.[62] Finally,

$$S(2)=\sum_{i=1}^{2}X_i^2-2\overline{X}_2^2=\frac{X_1^2-2X_1X_2+X_2^2}{2}=U_1^2,$$

which completes the proof of equation (39).

Let us now turn to the sampling distribution of the sample variance. Suppose that we have a simple random sample of size n from the population $X\sim \mathrm{N}\left(\mu,\sigma^2\right)$. Then the unbiased estimator of the sample variance is

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\frac{1}{n-1}\sum_{i=1}^{n-1}U_i^2,$$

where the $U_i\sim \mathrm{N}\left(0,\sigma^2\right)$ are mutually independently distributed, $i=1,\ldots,n-1$. Dividing each of the U_i by σ transforms them into standard Normal random variables. Now

$$s^2=\frac{\sigma^2}{n-1}\sum_{i=1}^{n-1}\left(\frac{U_i}{\sigma}\right)^2=\frac{\sigma^2}{n-1}Q,$$

where $Q\sim\chi^2_{n-1}$. That is, s^2 is proportional to a chi-squared random variable with $n-1$ degrees of freedom or, conversely,

$$\frac{(n-1)s^2}{\sigma^2}\sim\chi^2_{n-1}, \tag{40}$$

which is the result that all of this has been about.

One final point to notice. We see that s^2 is a function of $U_1,\ \ldots,\ U_{n-1}$. Recall that $U_n=\overline{X}\sqrt{n}$ is independent of $U_1,\ \ldots,\ U_{n-1}$. It follows immediately that s^2 and $\overline{X}$ are independent random variables. This is a very important result that will be exploited in Section 5.3.

[62] Note that the fourth equality has used the result $\overline{X}_{n-1}=\left(n\overline{X}_n-X_n\right)/(n-1)$.

C A Distributional Cheat Sheet

C.1 The Multivariate Normal and Related Distributions

Consider a set of jointly distributed random variables y_1, y_2, ..., y_n. Suppose that these variables have means $\mathrm{E}[y_i] = \mu_i$, $i = 1, \ldots, n$, and have a variance covariance structure

$$\mathrm{E}[(y_i - \mu_i)(y_j - \mu_j)] = \begin{cases} \mathrm{Var}[y_i] = \sigma_i^2, & i = 1, \ldots, n; i = j \\ \mathrm{Cov}[y_i, y_j] = \sigma_{ij}, & i, j = 1, \ldots, n; i \neq j. \end{cases}$$

We shall put this in matrix notation.

Let $y = [y_1, \ldots, y_n]'$ be an $n \times 1$ vector. We shall define the expected value of a random vector according to $\mathrm{E}[y] = [\mu_1, \ldots, \mu_n]' = \mu$, where μ is also an $n \times 1$ vector. Further, we shall define the variance of a vector according to

$$\mathrm{Var}[y] = \mathrm{E}[(y - \mu)(y - \mu)'] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{n1} & \ldots & \ldots & \sigma_n^2 \end{bmatrix} = \Sigma.$$

The matrix Σ is often called the *variance-covariance matrix*. I shall typically either refer to it as a variance matrix or a covariance matrix but almost never as a variance-covariance matrix. Covariance matrices have lots of properties that are important. First, in any situation that we are likely to encounter Σ will be positive definite (a multivariate notion of positive). This implies that it is non-singular, which means that it can be inverted (a multivariate version of not zero!). Also Σ is symmetric, i.e., $\sigma_{ij} = \sigma_{ji}$. Symmetric matrices have all sorts of useful properties. For instance, we know that there exists an non-singular upper-triangular matrix T, with all the elements on the leading diagonal positive, such that $\Sigma = TT'$. This is known as a Cholesky decomposition. One way of thinking about T in this context is as a multivariate generalization of a standard deviation, in the same way as a covariance matrix is a generalization of variance to a set of random variables.

The joint density function for y is

$$f(y; \mu, \Sigma) = (2\pi)^{-n/2} (\det \Sigma)^{-1/2} \exp\left\{-\frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\right\}.$$

We write $y \sim N(\mu, \Sigma)$.

There are certain results involving the multivariate normal distribution that occur repeatedly throughout the theory of the linear regression model.[63]

1. For any $p \times n$ matrix A and $n \times 1$ vector c we have $Ay \sim N(A\mu + c, A\Sigma A')$. In particular, $T^{-1}(y - \mu) \sim N(0, I_n)$ yields a generalization the standard normal distribution.

2. If $z \sim N(0, I_n)$ then $z'z \sim \chi^2_n$, where χ^2_n denotes a chi-squared distribution with n degrees of freedom. In particular, if $z \sim \mathrm{N}(0, 1)$ then $z^2 \sim \chi^2_1$.

3. If $y \sim N(\mu, \Sigma)$ then $(y - \mu)'\Sigma^{-1}(y - \mu) \sim \chi^2_n$.

[63] If you are unfamiliar with matrices I have always found the book by Searle (1982) especially useful.

4. If Q is an $n \times n$ idempotent matrix of rank q and $z \sim N(0, I_n)$ then $z'Qz \sim \chi^2_q$.[64]

5. Suppose that the $n \times 1$ vector $x \sim N(0, \Sigma)$. Now consider the two quadratic forms $x'Ax$ and $x'Bx$, where A and B are both $n \times n$ symmetric matrices. The two quadratic forms are independent if and only if $A\Sigma B = 0$.

6. Suppose that the $n \times 1$ vector $z \sim N(0, I_n)$. The linear and quadratic forms Az and $z'Bz$, where A is $p \times n$ and B is an $n \times n$ symmetric matrix, are independent if and only if $AB = 0$.

7. Suppose that the $n \times 1$ vector $x \sim N(0, \Sigma)$. Further suppose that $x'Ax \sim \chi^2_r$ and $x'Bx \sim \chi^2_s$ are independent quadratic forms. Then

$$\frac{x'Ax/r}{x'Bx/s} \sim F_{r,s},$$

where $F_{r,s}$ denotes an F distribution with r numerator and s denominator degrees of freedom.

C.1.1 Miscellaneous Other Distributional Results

1. If $x_1 \sim \chi^2_1$ and $x_1 \sim \chi^2_1$ are independent random variables then

$$x_1 + x_2 \sim \chi^2_2.$$

2. If $x_1 \sim \chi^2_p$ and $x_1 \sim \chi^2_q$ are independent random variables then

$$x_1 + x_2 \sim \chi^2_{p+q}.$$

3. If $z \sim N(0, 1)$ and $x \sim \chi^2_k$ are independent random variables then

$$\frac{z}{\sqrt{x/k}} \sim t_k.$$

4. If $t \sim t_k$ then, as $k \to \infty$, $t \underset{a}{\sim} N(0, 1)$.

5. If $x_1 \sim \chi^2_p$ and $x_1 \sim \chi^2_q$ are independent random variables then

$$\frac{x_1/p}{x_2/q} \sim F_{p,q}.$$

6. If $t \sim t_k$ then $t^2 \sim F_{1,k}$.

7. If $F \sim F_{r,s}$ then, as $s \to \infty$, $rF \underset{a}{\sim} \chi^2_r$.

[64] An idempotent matrix is always symmetric and satisfies the relationship $QQ = Q$, so it is like a multivariate version of the number one. The only idempotent matrix with full rank is the identity matrix.

C.2 Quadratic Forms Involving Generalized Inverses

Let A^- denote the generalized inverse of the matrix A, so that $AA^-A = A$. Generalize inverses are discussed in Section 5 of the Matrices handout. We note in passing that the matrices AA^- and A^-A are idempotent, so that $(AA^-)(AA^-) = AA^-$ and $(A^-A)(A^-A) = A^-A$ from property (i) above. An important property of idempotent matrices, which are necessarily square matrices, is that their trace is equal to their rank. We will exploit this fact in the statement of Corollary 2 where we see that $\mathrm{tr}(\Omega^-\Omega)$ is equal to the rank of $\Omega^-\Omega$ which, in turn, is equal to the rank of Ω, denoted ρ_Ω, because it can be shown that $\rho_{\Omega^-} \geq \rho_\Omega$ (Rao and Mitra, 1971, Lemma 2.2.3).

Our interest in generalized inverses lies in the following theorem and its two corollaries:

Theorem 1. *Rao and Mitra (1971, Theorem 9.2.1) Let $Y \sim \mathrm{N}(\mu, \Omega)$ be a random p-vector. Additionally, let A be a $p \times p$ symmetric matrix, b a p-vector and c a scalar. Then*

$$Y'AY + 2b'Y + c \sim \chi^2_{j,\delta},$$

where $\chi^2_{j,\delta}$ denotes a non-central chi-squared distribution with $j = \mathrm{tr}(A\Omega)$ degrees of freedom and non-centrality parameter $\delta = (A\mu + b)'\Omega A\Omega(A\mu + b)$, if and only if one of the following equivalent conditions is satisfied: (i) $\Omega A\Omega A\Omega = \Omega A\Omega$ or (ii) $(A\mu + b)'\Omega(A\mu + b) = \mu'A\mu + 2b'\mu + c$.

In the statement of Theorem 1 the notation $\mathrm{tr}(A\Omega)$ denotes the trace of the matrix $A\Omega$. Theorem 1 is what we would need to be able to study the power properties of the various test statistics. We will not pursue this further. For our purposes it is sufficient to restrict ourselves to central cases, so that $\mu = 0$, and simple quadratic forms where $b = 0$ and $c = 0$. This yields the following special case.

Corollary 1 (Corollary to Theorem 1). Let $Y \sim \mathrm{N}(0, \Omega)$ be a random p-vector and let A be a $p \times p$ symmetric matrix. Then, if and only if $\Omega A\Omega A\Omega = \Omega A\Omega$,

$$Y'AY \sim \chi^2_j, \quad \text{where } j = \mathrm{tr}(A\Omega).$$

In the special case where $A = \Omega^-$ we obtain further simplification:

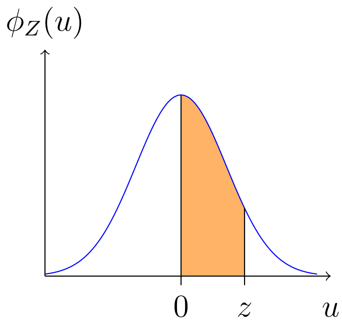
Corollary 2. Rao and Mitra (1971, Theorem 9.2.2) Let $Y \sim \mathrm{N}(0, \Omega)$ be a random p-vector. Then $Y'\Omega^-Y \sim \chi^2_k$ for any choice of Ω^-, where j is equal to the rank of Ω.

Note that the statement of this corollary has exploited the result $\mathrm{tr}(\Omega^-\Omega) = \rho_\Omega$ which was mentioned prior to the statement of Theorem 1 on page 48.

D Probability Tables

Tables provided:

Table D.1: Areas under the Standard Normal Density Function



Shaded Area: $\Phi_Z(z) = \int_0^z \phi_Z(u)\,du$, where $\phi_Z(u)$ denotes a standard Normal density function.

For example,

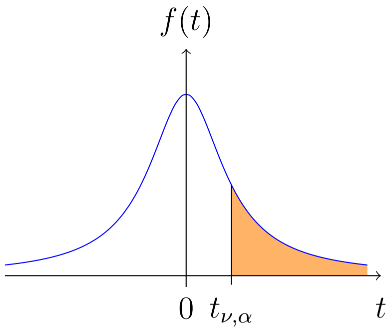
$P(Z \le 1.05) = P(Z \le 0) + P(0 < Z \le 1.05) = 0.5 + 0.3531 = 0.8531.$

By symmetry, if $z < 0$ then $\Phi_Z(z) = 1 - \Phi_Z(-z)$

Note that z is the sum of terms in the first row and column.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.40	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.50	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.60	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.70	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.80	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.90	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.00	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.10	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.20	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.30	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.40	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.50	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.60	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.70	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.80	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.90	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.00	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.10	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.20	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.30	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.40	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.50	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.60	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.70	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.80	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.90	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.00	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.10	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.20	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.30	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.40	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.50	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.60	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.70	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.80	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.90	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Table D.2: Critical Values for Upper Tail Probabilities of Student's t distribution



Shaded Area $= P(t \geq t_{\nu,\alpha}) = \alpha$

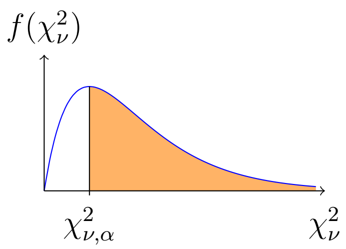
For example,

$$P(t \geq t_{3,0.05}) = P(t \geq 2.3534) = 0.05.$$

Note that ν denotes the degrees of freedom of the distribution.

ν	0.1	0.05	0.025	0.01	0.005	ν	0.1	0.05	0.025	0.01	0.005
1	3.0777	6.3137	12.7062	31.8210	63.6559	28	1.3125	1.7011	2.0484	2.4671	2.7633
2	1.8856	2.9200	4.3027	6.9645	9.9250	29	1.3114	1.6991	2.0452	2.4620	2.7564
3	1.6377	2.3534	3.1824	4.5407	5.8408	30	1.3104	1.6973	2.0423	2.4573	2.7500
4	1.5332	2.1318	2.7765	3.7469	4.6041	31	1.3095	1.6955	2.0395	2.4528	2.7440
5	1.4759	2.0150	2.5706	3.3649	4.0321	32	1.3086	1.6939	2.0369	2.4487	2.7385
6	1.4398	1.9432	2.4469	3.1427	3.7074	33	1.3077	1.6924	2.0345	2.4448	2.7333
7	1.4149	1.8946	2.3646	2.9979	3.4995	34	1.3070	1.6909	2.0322	2.4411	2.7284
8	1.3968	1.8595	2.3060	2.8965	3.3554	35	1.3062	1.6896	2.0301	2.4377	2.7238
9	1.3830	1.8331	2.2622	2.8214	3.2498	36	1.3055	1.6883	2.0281	2.4345	2.7195
10	1.3722	1.8125	2.2281	2.7638	3.1693	37	1.3049	1.6871	2.0262	2.4314	2.7154
11	1.3634	1.7959	2.2010	2.7181	3.1058	38	1.3042	1.6860	2.0244	2.4286	2.7116
12	1.3562	1.7823	2.1788	2.6810	3.0545	39	1.3036	1.6849	2.0227	2.4258	2.7079
13	1.3502	1.7709	2.1604	2.6503	3.0123	40	1.3031	1.6839	2.0211	2.4233	2.7045
14	1.3450	1.7613	2.1448	2.6245	2.9768	45	1.3007	1.6794	2.0141	2.4121	2.6896
15	1.3406	1.7531	2.1315	2.6025	2.9467	50	1.2987	1.6759	2.0086	2.4033	2.6778
16	1.3368	1.7459	2.1199	2.5835	2.9208	60	1.2958	1.6706	2.0003	2.3901	2.6603
17	1.3334	1.7396	2.1098	2.5669	2.8982	70	1.2938	1.6669	1.9944	2.3808	2.6479
18	1.3304	1.7341	2.1009	2.5524	2.8784	80	1.2922	1.6641	1.9901	2.3739	2.6387
19	1.3277	1.7291	2.0930	2.5395	2.8609	90	1.2910	1.6620	1.9867	2.3685	2.6316
20	1.3253	1.7247	2.0860	2.5280	2.8453	100	1.2901	1.6602	1.9840	2.3642	2.6259
21	1.3232	1.7207	2.0796	2.5176	2.8314	120	1.2886	1.6576	1.9799	2.3578	2.6174
22	1.3212	1.7171	2.0739	2.5083	2.8188	140	1.2876	1.6558	1.9771	2.3533	2.6114
23	1.3195	1.7139	2.0687	2.4999	2.8073	160	1.2869	1.6544	1.9749	2.3499	2.6069
24	1.3178	1.7109	2.0639	2.4922	2.7970	180	1.2863	1.6534	1.9732	2.3472	2.6034
25	1.3163	1.7081	2.0595	2.4851	2.7874	200	1.2858	1.6525	1.9719	2.3451	2.6006
26	1.3150	1.7056	2.0555	2.4786	2.7787	∞	1.2816	1.6449	1.9600	2.3263	2.5758
27	1.3137	1.7033	2.0518	2.4727	2.7707						

Table D.3: Critical Values for Upper-Tail Probabilities of the Chi-square Distribution



Shaded Area: $P(\chi^2_\nu \geq \chi^2_{\nu,\alpha}) = \alpha$.

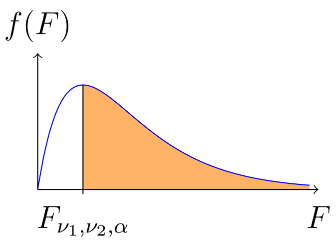
For example, $P(\chi^2_3 \geq \chi^2_{3,0.95}) = P(\chi^2_\nu \geq 0.3518) = 0.95$.

Note that ν denotes the degrees of freedom of the distribution.

	Upper-Tail Probabilities (α)						
ν	0.9950	0.9900	0.9750	0.9500	0.9000	0.7500	0.5000
1	3.927×10^{-5}	0.0002	0.0010	0.0039	0.0158	0.1015	0.4549
2	0.0100	0.0201	0.0506	0.1026	0.2107	0.5754	1.3863
3	0.0717	0.1148	0.2158	0.3518	0.5844	1.2125	2.3660
4	0.2070	0.2971	0.4844	0.7107	1.0636	1.9226	3.3567
5	0.4117	0.5543	0.8312	1.1455	1.6103	2.6746	4.3515
6	0.6757	0.8721	1.2373	1.6354	2.2041	3.4546	5.3481
7	0.9893	1.2390	1.6899	2.1673	2.8331	4.2549	6.3458
8	1.3444	1.6465	2.1797	2.7326	3.4895	5.0706	7.3441
9	1.7349	2.0879	2.7004	3.3251	4.1682	5.8988	8.3428
10	2.1559	2.5582	3.2470	3.9403	4.8652	6.7372	9.3418
11	2.6032	3.0535	3.8157	4.5748	5.5778	7.5841	10.3410
12	3.0738	3.5706	4.4038	5.2260	6.3038	8.4384	11.3403
13	3.5650	4.1069	5.0088	5.8919	7.0415	9.2991	12.3398
14	4.0747	4.6604	5.6287	6.5706	7.7895	10.1653	13.3393
15	4.6009	5.2293	6.2621	7.2609	8.5468	11.0365	14.3389
16	5.1422	5.8122	6.9077	7.9616	9.3122	11.9122	15.3385
17	5.6972	6.4078	7.5642	8.6718	10.0852	12.7919	16.3382
18	6.2648	7.0149	8.2307	9.3905	10.8649	13.6753	17.3379
19	6.8440	7.6327	8.9065	10.1170	11.6509	14.5620	18.3377
20	7.4338	8.2604	9.5908	10.8508	12.4426	15.4518	19.3374
21	8.0337	8.8972	10.2829	11.5913	13.2396	16.3444	20.3372
22	8.6427	9.5425	10.9823	12.3380	14.0415	17.2396	21.3370
23	9.2604	10.1957	11.6886	13.0905	14.8480	18.1373	22.3369
24	9.8862	10.8564	12.4012	13.8484	15.6587	19.0373	23.3367
25	10.5197	11.5240	13.1197	14.6114	16.4734	19.9393	24.3366
26	11.1602	12.1981	13.8439	15.3792	17.2919	20.8434	25.3365
27	11.8076	12.8785	14.5734	16.1514	18.1139	21.7494	26.3363
28	12.4613	13.5647	15.3079	16.9279	18.9392	22.6572	27.3362
29	13.1211	14.2565	16.0471	17.7084	19.7677	23.5666	28.3361
30	13.7867	14.9535	16.7908	18.4927	20.5992	24.4776	29.3360
35	17.1918	18.5089	20.5694	22.4650	24.7967	29.0540	34.3356
40	20.7065	22.1643	24.4330	26.5093	29.0505	33.6603	39.3353
50	27.9907	29.7067	32.3574	34.7643	37.6886	42.9421	49.3349
60	35.5345	37.4849	40.4817	43.1880	46.4589	52.2938	59.3347
70	43.2752	45.4417	48.7576	51.7393	55.3289	61.6983	69.3345
80	51.1719	53.5401	57.1532	60.3915	64.2778	71.1445	79.3343
100	67.3276	70.0649	74.2219	77.9295	82.3581	90.1332	99.3341

	Upper Tail Probabilities (α)					
ν	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050
1	1.3233	2.7055	3.8415	5.0239	6.6349	7.8794
2	2.7726	4.6052	5.9915	7.3778	9.2103	10.5966
3	4.1083	6.2514	7.8147	9.3484	11.3449	12.8382
4	5.3853	7.7794	9.4877	11.1433	13.2767	14.8603
5	6.6257	9.2364	11.0705	12.8325	15.0863	16.7496
6	7.8408	10.6446	12.5916	14.4494	16.8119	18.5476
7	9.0371	12.0170	14.0671	16.0128	18.4753	20.2777
8	10.2189	13.3616	15.5073	17.5345	20.0902	21.9550
9	11.3888	14.6837	16.9190	19.0228	21.6660	23.5894
10	12.5489	15.9872	18.3070	20.4832	23.2093	25.1882
11	13.7007	17.2750	19.6751	21.9200	24.7250	26.7568
12	14.8454	18.5493	21.0261	23.3367	26.2170	28.2995
13	15.9839	19.8119	22.3620	24.7356	27.6882	29.8195
14	17.1169	21.0641	23.6848	26.1189	29.1412	31.3193
15	18.2451	22.3071	24.9958	27.4884	30.5779	32.8013
16	19.3689	23.5418	26.2962	28.8454	31.9999	34.2672
17	20.4887	24.7690	27.5871	30.1910	33.4087	35.7185
18	21.6049	25.9894	28.8693	31.5264	34.8053	37.1565
19	22.7178	27.2036	30.1435	32.8523	36.1909	38.5823
20	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968
21	24.9348	29.6151	32.6706	35.4789	38.9322	41.4011
22	26.0393	30.8133	33.9244	36.7807	40.2894	42.7957
23	27.1413	32.0069	35.1725	38.0756	41.6384	44.1813
24	28.2412	33.1962	36.4150	39.3641	42.9798	45.5585
25	29.3389	34.3816	37.6525	40.6465	44.3141	46.9279
26	30.4346	35.5632	38.8851	41.9232	45.6417	48.2899
27	31.5284	36.7412	40.1133	43.1945	46.9629	49.6449
28	32.6205	37.9159	41.3371	44.4608	48.2782	50.9934
29	33.7109	39.0875	42.5570	45.7223	49.5879	52.3356
30	34.7997	40.2560	43.7730	46.9792	50.8922	53.6720
35	40.2228	46.0588	49.8018	53.2033	57.3421	60.2748
40	45.6160	51.8051	55.7585	59.3417	63.6907	66.7660
50	56.3336	63.1671	67.5048	71.4202	76.1539	79.4900
60	66.9815	74.3970	79.0819	83.2977	88.3794	91.9517
70	77.5767	85.5270	90.5312	95.0232	100.4252	104.2149
80	88.1303	96.5782	101.8795	106.6286	112.3288	116.3211
100	109.1412	118.4980	124.3421	129.5612	135.8067	140.1695

Table D.4: Critical Values for 5% Upper Tail Probabilities of the F Distribution



Shaded Area: $P(F \geq F_{\nu_1,\nu_2,\alpha}) = \alpha$.

For example, $P(F_{3,4} \geq F_{3,4,0.95}) = P(F_{3,4} \geq 6.59) = 0.95$.

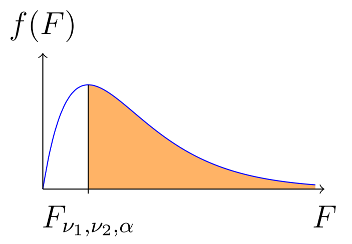
Note that ν_1 and ν_2 denote the numerator and denominator degrees of freedom of the distribution, respectively.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69	245.36	245.95
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.80	1.78	1.75
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.72	1.69	1.67

Critical Values for 5% Upper Tail Probabilities of the F Distribution (Table D.4 Continued)

$\nu_2 \backslash \nu_1$	16	17	18	19	20	25	30	40	60	∞
1	246.46	246.92	247.32	247.69	248.01	249.26	250.10	251.14	252.20	254.31
2	19.43	19.44	19.44	19.44	19.45	19.46	19.46	19.47	19.48	19.50
3	8.69	8.68	8.67	8.67	8.66	8.63	8.62	8.59	8.57	8.53
4	5.84	5.83	5.82	5.81	5.80	5.77	5.75	5.72	5.69	5.63
5	4.60	4.59	4.58	4.57	4.56	4.52	4.50	4.46	4.43	4.36
6	3.92	3.91	3.90	3.88	3.87	3.83	3.81	3.77	3.74	3.67
7	3.49	3.48	3.47	3.46	3.44	3.40	3.38	3.34	3.30	3.23
8	3.20	3.19	3.17	3.16	3.15	3.11	3.08	3.04	3.01	2.93
9	2.99	2.97	2.96	2.95	2.94	2.89	2.86	2.83	2.79	2.71
10	2.83	2.81	2.80	2.79	2.77	2.73	2.70	2.66	2.62	2.54
11	2.70	2.69	2.67	2.66	2.65	2.60	2.57	2.53	2.49	2.40
12	2.60	2.58	2.57	2.56	2.54	2.50	2.47	2.43	2.38	2.30
13	2.51	2.50	2.48	2.47	2.46	2.41	2.38	2.34	2.30	2.21
14	2.44	2.43	2.41	2.40	2.39	2.34	2.31	2.27	2.22	2.13
15	2.38	2.37	2.35	2.34	2.33	2.28	2.25	2.20	2.16	2.07
16	2.33	2.32	2.30	2.29	2.28	2.23	2.19	2.15	2.11	2.01
17	2.29	2.27	2.26	2.24	2.23	2.18	2.15	2.10	2.06	1.96
18	2.25	2.23	2.22	2.20	2.19	2.14	2.11	2.06	2.02	1.92
19	2.21	2.20	2.18	2.17	2.16	2.11	2.07	2.03	1.98	1.88
20	2.18	2.17	2.15	2.14	2.12	2.07	2.04	1.99	1.95	1.84
21	2.16	2.14	2.12	2.11	2.10	2.05	2.01	1.96	1.92	1.81
22	2.13	2.11	2.10	2.08	2.07	2.02	1.98	1.94	1.89	1.78
23	2.11	2.09	2.08	2.06	2.05	2.00	1.96	1.91	1.86	1.76
24	2.09	2.07	2.05	2.04	2.03	1.97	1.94	1.89	1.84	1.73
25	2.07	2.05	2.04	2.02	2.01	1.96	1.92	1.87	1.82	1.71
26	2.05	2.03	2.02	2.00	1.99	1.94	1.90	1.85	1.80	1.69
27	2.04	2.02	2.00	1.99	1.97	1.92	1.88	1.84	1.79	1.67
28	2.02	2.00	1.99	1.97	1.96	1.91	1.87	1.82	1.77	1.65
29	2.01	1.99	1.97	1.96	1.94	1.89	1.85	1.81	1.75	1.64
30	1.99	1.98	1.96	1.95	1.93	1.88	1.84	1.79	1.74	1.62
40	1.90	1.89	1.87	1.85	1.84	1.78	1.74	1.69	1.64	1.51
60	1.82	1.80	1.78	1.76	1.75	1.69	1.65	1.59	1.53	1.39
120	1.73	1.71	1.69	1.67	1.66	1.60	1.55	1.50	1.43	1.25
∞	1.64	1.62	1.60	1.59	1.57	1.51	1.46	1.39	1.32	1.00

Table D.5: Critical Values for 1% Upper Tail Probabilities of the F Distribution



Shaded Area: $P(F \geq F_{\nu_1,\nu_2,\alpha}) = \alpha$.

For example, $P(F_{3,4} \geq F_{3,4,0.99}) = P(F_{3,4} \geq 16.69) = 0.99$.

Note that ν_1 and ν_2 denote the numerator and denominator degrees of freedom of the distribution, respectively.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6083.32	6106.32	6125.86	6142.67	6157.28
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.42	99.43	99.43
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.13	27.05	26.98	26.92	26.87
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.31	14.25	14.20
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.82	9.77	9.72
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01	4.96
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12	3.07	3.03
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02	2.98
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	3.02	2.97	2.93
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.98	2.93	2.89
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89	2.85
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.90	2.86	2.81
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.87	2.82	2.78
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.84	2.79	2.75
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87	2.81	2.77	2.73
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.28	2.23	2.19
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.13	2.08	2.04

Critical Values for 1% Upper Tail Probabilities of the F Distribution (Table D.5 continued)

$\nu_2 \backslash \nu_1$	16	17	18	19	20	25	30	40	60	∞
1	6170.10	6181.43	6191.53	6200.58	6208.73	6239.83	6260.65	6286.78	6313.03	6365.86
2	99.44	99.44	99.44	99.45	99.45	99.46	99.47	99.47	99.48	99.50
3	26.83	26.79	26.75	26.72	26.69	26.58	26.50	26.41	26.32	26.13
4	14.15	14.11	14.08	14.05	14.02	13.91	13.84	13.75	13.65	13.46
5	9.68	9.64	9.61	9.58	9.55	9.45	9.38	9.29	9.20	9.02
6	7.52	7.48	7.45	7.42	7.40	7.30	7.23	7.14	7.06	6.88
7	6.28	6.24	6.21	6.18	6.16	6.06	5.99	5.91	5.82	5.65
8	5.48	5.44	5.41	5.38	5.36	5.26	5.20	5.12	5.03	4.86
9	4.92	4.89	4.86	4.83	4.81	4.71	4.65	4.57	4.48	4.31
10	4.52	4.49	4.46	4.43	4.41	4.31	4.25	4.17	4.08	3.91
11	4.21	4.18	4.15	4.12	4.10	4.01	3.94	3.86	3.78	3.60
12	3.97	3.94	3.91	3.88	3.86	3.76	3.70	3.62	3.54	3.36
13	3.78	3.75	3.72	3.69	3.66	3.57	3.51	3.43	3.34	3.17
14	3.62	3.59	3.56	3.53	3.51	3.41	3.35	3.27	3.18	3.00
15	3.49	3.45	3.42	3.40	3.37	3.28	3.21	3.13	3.05	2.87
16	3.37	3.34	3.31	3.28	3.26	3.16	3.10	3.02	2.93	2.75
17	3.27	3.24	3.21	3.19	3.16	3.07	3.00	2.92	2.83	2.65
18	3.19	3.16	3.13	3.10	3.08	2.98	2.92	2.84	2.75	2.57
19	3.12	3.08	3.05	3.03	3.00	2.91	2.84	2.76	2.67	2.49
20	3.05	3.02	2.99	2.96	2.94	2.84	2.78	2.69	2.61	2.42
21	2.99	2.96	2.93	2.90	2.88	2.79	2.72	2.64	2.55	2.36
22	2.94	2.91	2.88	2.85	2.83	2.73	2.67	2.58	2.50	2.31
23	2.89	2.86	2.83	2.80	2.78	2.69	2.62	2.54	2.45	2.26
24	2.85	2.82	2.79	2.76	2.74	2.64	2.58	2.49	2.40	2.21
25	2.81	2.78	2.75	2.72	2.70	2.60	2.54	2.45	2.36	2.17
26	2.78	2.75	2.72	2.69	2.66	2.57	2.50	2.42	2.33	2.13
27	2.75	2.71	2.68	2.66	2.63	2.54	2.47	2.38	2.29	2.10
28	2.72	2.68	2.65	2.63	2.60	2.51	2.44	2.35	2.26	2.06
29	2.69	2.66	2.63	2.60	2.57	2.48	2.41	2.33	2.23	2.03
30	2.66	2.63	2.60	2.57	2.55	2.45	2.39	2.30	2.21	2.01
40	2.48	2.45	2.42	2.39	2.37	2.27	2.20	2.11	2.02	1.80
60	2.31	2.28	2.25	2.22	2.20	2.10	2.03	1.94	1.84	1.60
120	2.15	2.12	2.09	2.06	2.03	1.93	1.86	1.76	1.66	1.38
∞	2.00	1.97	1.93	1.90	1.88	1.77	1.70	1.59	1.47	1.00