Lecture 1. Transformation of Random Variables

Suppose we are given a random variable X with density $f_X(x)$. We apply a function g to produce a random variable $Y = g(X)$. We can think of X as the input to a black box, and Y the output. We wish to find the density or distribution function of Y. We illustrate the technique for the example in Figure 1.1.

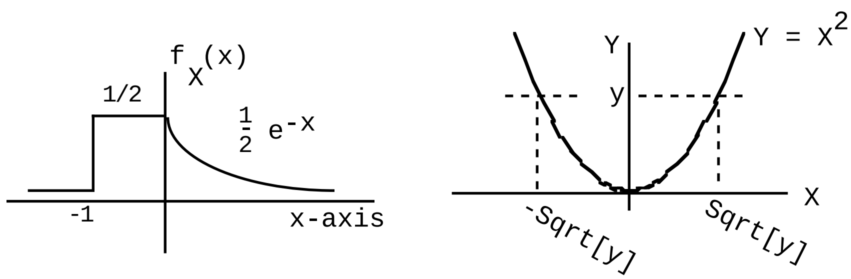


Figure 1.1

The **distribution function method** finds F_Y directly, and then f_Y by differentiation. We have $F_Y(y) = 0$ for $y < 0$. If $y \geq 0$, then $P\{Y \leq y\} = P\{-\sqrt{y} \leq x \leq \sqrt{y}\}$.

Case 1. $0 \leq y \leq 1$ (Figure 1.2). Then

$$F_Y(y) = \frac{1}{2}\sqrt{y} + \int_0^{\sqrt{y}} \frac{1}{2}e^{-x}\,dx = \frac{1}{2}\sqrt{y} + \frac{1}{2}(1 - e^{-\sqrt{y}}).$$

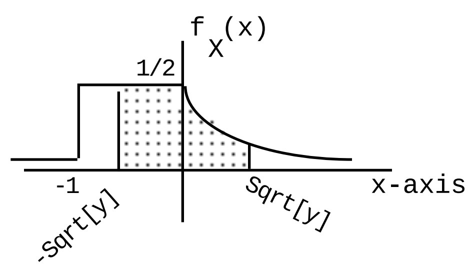


Figure 1.2

Case 2. $y > 1$ (Figure 1.3). Then

$$F_Y(y) = \frac{1}{2} + \int_0^{\sqrt{y}} \frac{1}{2}e^{-x}\,dx = \frac{1}{2} + \frac{1}{2}(1 - e^{-\sqrt{y}}).$$

The density of Y is 0 for $y < 0$ and

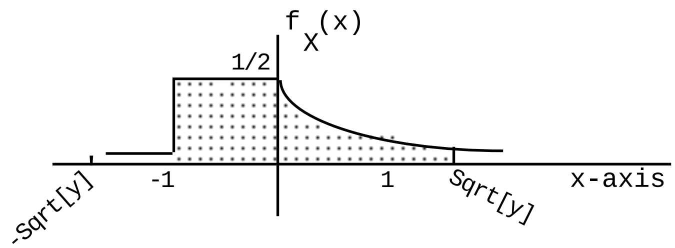


Figure 1.3

$$f_Y(y) = \frac{1}{4\sqrt{y}}(1 + e^{-\sqrt{y}}), \quad 0 < y < 1;$$

$$f_Y(y) = \frac{1}{4\sqrt{y}} e^{-\sqrt{y}}, \quad y > 1.$$

See Figure 1.4 for a sketch of f_Y and F_Y. (You can take $f_Y(y)$ to be anything you like at $y = 1$ because $\{Y = 1\}$ has probability zero.)

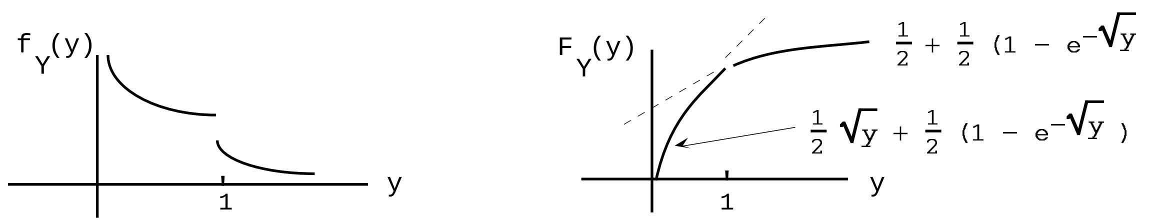


Figure 1.4

The **density function method** finds f_Y directly, and then F_Y by integration; see Figure 1.5. We have $f_Y(y)|dy| = f_X(\sqrt{y})dx + f_X(-\sqrt{y})dx$; we write $|dy|$ because probabilities are never negative. Thus

$$f_Y(y) = \frac{f_X(\sqrt{y})}{|dy/dx|_{x=\sqrt{y}}} + \frac{f_X(-\sqrt{y})}{|dy/dx|_{x=-\sqrt{y}}}$$

with $y = x^2, \quad dy/dx = 2x$, so

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}.$$

(Note that $|-2\sqrt{y}| = 2\sqrt{y}$.) We have $f_Y(y) = 0$ for $y < 0$, and:
Case 1. $0 < y < 1$ (see Figure 1.2).

$$f_Y(y) = \frac{(1/2)e^{-\sqrt{y}}}{2\sqrt{y}} + \frac{1/2}{2\sqrt{y}} = \frac{1}{4\sqrt{y}}(1 + e^{-\sqrt{y}}).$$

Case 2. $y > 1$ (see Figure 1.3).

$$f_Y(y) = \frac{(1/2)e^{-\sqrt{y}}}{2\sqrt{y}} + 0 = \frac{1}{4\sqrt{y}}e^{-\sqrt{y}}$$

as before.

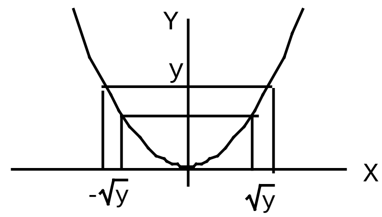


Figure 1.5

The distribution function method generalizes to situations where we have a single output but more than one input. For example, let X and Y be independent, each uniformly distributed on $[0, 1]$. The distribution function of $Z = X + Y$ is

$$F_Z(z) = P\{X + Y \le z\} = \int\int_{x+y\le z} f_{XY}(x, y)\,dx\,dy$$

with $f_{XY}(x, y) = f_X(x)f_Y(y)$ by independence. Now $F_Z(z) = 0$ for $z < 0$ and $F_Z(z) = 1$ for $z > 2$ (because $0 \le Z \le 2$).

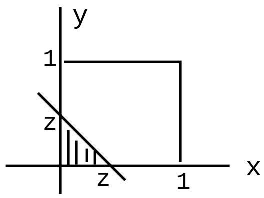
Case 1. If $0 \le z \le 1$, then $F_Z(z)$ is the shaded area in Figure 1.6, which is $z^2/2$.

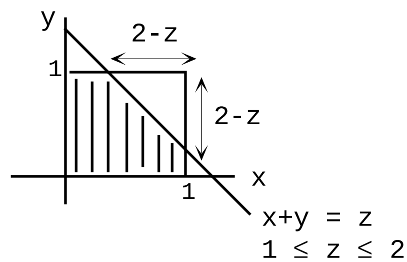
Case 2. If $1 \le z \le 2$, then $F_Z(z)$ is the shaded area in Figure 1.7, which is $1-[(2-z)^2/2]$. Thus (see Figure 1.8)

$$f_Z(z) = \begin{cases} z, & 0 \le z \le 1 \\ 2 - z & 1 \le z \le 2 \\ 0 & \text{elsewhere.} \end{cases}.$$

Problems

1. Let X, Y, Z be independent, identically distributed (from now on, abbreviated iid) random variables, each with density $f(x) = 6x^5$ for $0 \le x \le 1$, and 0 elsewhere. Find the distribution and density functions of the maximum of X, Y and Z.
2. Let X and Y be independent, each with density $e^{-x}, x \ge 0$. Find the distribution (from now on, an abbreviation for "Find the distribution or density function") of $Z = Y/X$.
3. A discrete random variable X takes values $x_1, \dots, x_n$, each with probability $1/n$. Let $Y = g(X)$ where g is an arbitrary real-valued function. Express the probability function of Y ($p_Y(y) = P\{Y = y\}$) in terms of g and the x_i.





Figures 1.6 and 1.7

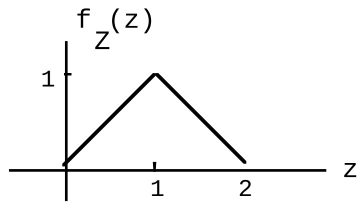


Figure 1.8

4. A random variable X has density $f(x) = ax^2$ on the interval $[0, b]$. Find the density of $Y = X^3$.
5. The *Cauchy density* is given by $f(y) = 1/[\pi(1 + y^2)]$ for all real y. Show that one way to produce this density is to take the tangent of a random variable X that is uniformly distributed between $-\pi/2$ and $\pi/2$.

Lecture 2. Jacobians

We need this idea to generalize the density function method to problems where there are k inputs and k outputs, with $k \geq 2$. However, if there are k inputs and $j < k$ outputs, often extra outputs can be introduced, as we will see later in the lecture.

2.1 The Setup

Let $X = X(U, V), Y = Y(U, V)$. Assume a one-to-one transformation, so that we can solve for U and V. Thus $U = U(X, Y), V = V(X, Y)$. Look at Figure 2.1. If u changes by du then x changes by $(\partial x/\partial u)\, du$ and y changes by $(\partial y/\partial u)\, du$. Similarly, if v changes by dv then x changes by $(\partial x/\partial v)\, dv$ and y changes by $(\partial y/\partial v)\, dv$. The small rectangle in the $u-v$ plane corresponds to a small parallelogram in the $x-y$ plane (Figure 2.2), with $A = (\partial x/\partial u, \partial y/\partial u, 0)\, du$ and $B = (\partial x/\partial v, \partial y/\partial v, 0)\, dv$. The area of the parallelogram is $|A \times B|$ and

$$A \times B = \begin{vmatrix} I & J & K \\ \partial x/\partial u & \partial y/\partial u & 0 \\ \partial x/\partial v & \partial y/\partial v & 0 \end{vmatrix} du\, dv = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix} du\, dvK.$$

(A determinant is unchanged if we transpose the matrix, i.e., interchange rows and columns.)

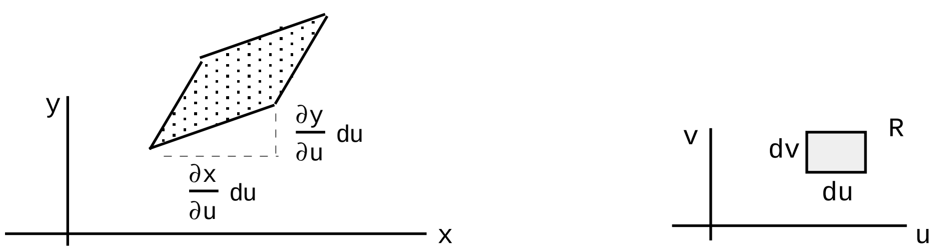


Figure 2.1

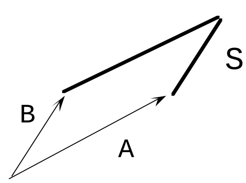


Figure 2.2

2.2 Definition and Discussion

The *Jacobian* of the transformation is

$$J = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix}, \quad \text{written as} \quad \frac{\partial(x, y)}{\partial(u, v)}.$$

Thus $|A \times B| = |J|\,du\,dv$. Now $P\{(X,Y) \in S\} = P\{(U,V) \in R\}$, in other words, $f_{XY}(x,y)$ times the area of S is $f_{UV}(u,v)$ times the area of R. Thus

$$f_{XY}(x,y)|J|\,du\,dv = f_{UV}(u,v)\,du\,dv$$

and

$$f_{UV}(u,v) = f_{XY}(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|.$$

The absolute value of the Jacobian $\partial(x,y)/\partial(u,v)$ gives a magnification factor for area in going from $u-v$ coordinates to $x-y$ coordinates. The magnification factor going the other way is $|\partial(u,v)/\partial(x,y)|$. But the magnification factor from $u-v$ to $u-v$ is 1, so

$$f_{UV}(u,v) = \frac{f_{XY}(x,y)}{|\partial(u,v)/\partial(x,y)|}.$$

In this formula, we must substitute $x = x(u,v), y = y(u,v)$ to express the final result in terms of u and v.

In three dimensions, a small rectangular box with volume $du\,dv\,dw$ corresponds to a parallelepiped in xyz space, determined by vectors

$$A = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \end{pmatrix} du,\ B = \begin{pmatrix} \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{pmatrix} dv,\ C = \begin{pmatrix} \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w} & \frac{\partial z}{\partial w} \end{pmatrix} dw.$$

The volume of the parallelepiped is the absolute value of the dot product of A with $B \times C$, and the dot product can be written as a determinant with rows (or columns) A, B, C. This determinant is the Jacobian of x, y, z with respect to u, v, w [written $\partial(x,y,z)/\partial(u,v,w)$], times $du\,dv\,dw$. The volume magnification from uvw to xyz space is $|\partial(x,y,z)/\partial(u,v,w)|$ and we have

$$f_{UVW}(u,v,w) = \frac{f_{XYZ}(x,y,z)}{|\partial(u,v,w)/\partial(x,y,z)|}$$

with $x = x(u,v,w), y = y(u,v,w), z = z(u,v,w)$.

The Jacobian technique extends to higher dimensions. The transformation formula is a natural generalization of the two and three-dimensional cases:

$$f_{Y_1 Y_2 \cdots Y_n}(y_1, \dots, y_n) = \frac{f_{X_1 \cdots X_n}(x_1, \dots, x_n)}{|\partial(y_1, \dots, y_n)/\partial(x_1, \dots, x_n)|}$$

where

$$\frac{\partial(y_1, \dots, y_n)}{\partial(x_1, \dots, x_n)} = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ & \vdots & \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix}.$$

To help you remember the formula, think $f_Y(y)\,dy = f_X(x)dx$.

2.3 A Typical Application

Let X and Y be independent, positive random variables with densities f_X and f_Y, and let $Z = XY$. We find the density of Z by introducing a new random variable W, as follows:

$$Z = XY, \quad W = Y$$

($W = X$ would be equally good). The transformation is one-to-one because we can solve for X, Y in terms of Z, W by $X = Z/W, Y = W$. In a problem of this type, we must always pay attention to the range of the variables: $x > 0, y > 0$ is equivalent to $z > 0, w > 0$. Now

$$f_{ZW}(z,w) = \frac{f_{XY}(x,y)}{|\partial(z,w)/\partial(x,y)|_{x=z/w,y=w}}$$

with

$$\frac{\partial(z,w)}{\partial(x,y)} = \begin{vmatrix} \partial z/\partial x & \partial z/\partial y \\ \partial w/\partial x & \partial w/\partial y \end{vmatrix} = \begin{vmatrix} y & x \\ 0 & 1 \end{vmatrix} = y.$$

Thus

$$f_{ZW}(z,w) = \frac{f_X(x)f_Y(y)}{w} = \frac{f_X(z/w)f_Y(w)}{w}$$

and we are left with the problem of finding the marginal density from a joint density:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{ZW}(z,w)\,dw = \int_0^{\infty} \frac{1}{w} f_X(z/w)f_Y(w)\,dw.$$

Problems

1. The joint density of two random variables X_1 and X_2 is $f(x_1, x_2) = 2e^{-x_1}e^{-x_2}$, where $0 < x_1 < x_2 < \infty$; $f(x_1, x_2) = 0$ elsewhere. Consider the transformation $Y_1 = 2X_1, \quad Y_2 = X_2 - X_1$. Find the joint density of Y_1 and Y_2, and conclude that Y_1 and Y_2 are independent.
2. Repeat Problem 1 with the following new data. The joint density is given by $f(x_1, x_2) = 8x_1x_2, \quad 0 < x_1 < x_2 < 1; \quad f(x_1, x_2) = 0$ elsewhere; $Y_1 = X_1/X_2, \quad Y_2 = X_2$.
3. Repeat Problem 1 with the following new data. We now have three iid random variables $X_i, i = 1, 2, 3$, each with density $e^{-x}, x > 0$. The transformation equations are given by $Y_1 = X_1/(X_1 + X_2), \quad Y_2 = (X_1 + X_2)/(X_1 + X_2 + X_3), \quad Y_3 = X_1 + X_2 + X_3$. As before, find the joint density of the Y_i and show that Y_1, Y_2 and Y_3 are independent.

Comments on the Problem Set

In Problem 3, notice that $Y_1Y_2Y_3 = X_1, \quad Y_2Y_3 = X_1 + X_2$, so $X_2 = Y_2Y_3 - Y_1Y_2Y_3, \quad X_3 = (X_1 + X_2 + X_3) - (X_1 + X_2) = Y_3 - Y_2Y_3$.

If $f_{XY}(x,y) = g(x)h(y)$ for all x, y, then X and Y are independent, because

$$f(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{g(x)h(y)}{g(x)\int_{-\infty}^{\infty} h(y)\,dy}$$

which does not depend on x. The set of points where $g(x) = 0$ (equivalently $f_X(x) = 0$) can be ignored because it has probability zero. It is important to realize that in this argument, "for all x, y" means that x and y must be allowed to vary independently of each other, so the set of possible x and y must be of the rectangular form $a < x < b, c < y < d$. (The constants a, b, c, d can be infinite.) For example, if $f_{XY}(x, y) = 2e^{-x}e^{-y}, 0 < y < x$, and 0 elsewhere, then X and Y are *not* independent. Knowing x forces $0 < y < x$, so the conditional distribution of Y given $X = x$ certainly depends on x. Note that $f_{XY}(x, y)$ is *not* a function of x alone times a function of y alone. We have

$$f_{XY}(x, y) = 2e^{-x}e^{-y}I[0 < y < x]$$

where the *indicator* I is 1 for $0 < y < x$ and 0 elsewhere.

In Jacobian problems, pay close attention to the range of the variables. For example, in Problem 1 we have $y_1 = 2x_1, y_2 = x_2 - x_1$, so $x_1 = y_1/2, x_2 = (y_1/2) + y_2$. From these equations it follows that $0 < x_1 < x_2 < \infty$ is equivalent to $y_1 > 0, y_2 > 0$.

Lecture 3. Moment-Generating Functions

3.1 Definition

The *moment-generating function* of a random variable X is defined by

$$M(t) = M_X(t) = E[e^{tX}]$$

where t is a real number. To see the reason for the terminology, note that $M(t)$ is the expectation of $1 + tX + t^2X^2/2! + t^3X^3/3! + \cdots$. If $\mu_n = E(X^n)$, the n-th moment of X, and we can take the expectation term by term, then

$$M(t) = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \cdots + \frac{\mu_n t^n}{n!} + \cdots .$$

Since the coefficient of t^n in the Taylor expansion is $M^{(n)}(0)/n!$, where $M^{(n)}$ is the n-th derivative of M, we have $\mu_n = M^{(n)}(0)$.

3.2 The Key Theorem

If $Y = \sum_{i=1}^n X_i$ where $X_1, \dots, X_n$ are independent, then $M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$.

Proof. First note that if X and Y are independent, then

$$E[g(X)h(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)\,dx\,dy.$$

Since $f_{XY}(x,y) = f_X(x)f_Y(y)$, the double integral becomes

$$\int_{-\infty}^{\infty} g(x)f_X(x)\,dx \int_{-\infty}^{\infty} h(y)f_Y(y)\,dy = E[g(X)]E[h(Y)]$$

and similarly for more than two random variables. Now if $Y = X_1 + \cdots + X_n$ with the X_i's independent, we have

$$M_Y(t) = E[e^{tY}] = E[e^{tX_1}\cdots e^{tX_n}] = E[e^{tX_1}]\cdots E[e^{tX_n}] = M_{X_1}(t)\cdots M_{X_n}(t). \quad ♣$$

3.3 The Main Application

Given independent random variables $X_1, \dots, X_n$ with densities $f_1, \dots, f_n$ respectively, find the density of $Y = \sum_{i=1}^n X_i$.

Step 1. Compute $M_i(t)$, the moment-generating function of X_i, for each i.

Step 2. Compute $M_Y(t) = \prod_{i=1}^n M_i(t)$.

Step 3. From $M_Y(t)$ find $f_Y(y)$.

This technique is known as a *transform method*. Notice that the moment-generating function and the density of a random variable are related by $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$. With t replaced by $-s$ we have a *Laplace transform*, and with t replaced by it we have a *Fourier transform*. The strategy works because at step 3, the moment-generating function determines the density uniquely. (This is a theorem from Laplace or Fourier transform theory.)

3.4 Examples

1. *Bernoulli Trials.* Let X be the number of successes in n trials with probability of success p on a given trial. Then $X = X_1 + \cdots + X_n$, where $X_i = 1$ if there is a success on trial i and $X_i = 0$ if there is a failure on trial i. Thus

$$M_i(t) = E[e^{tX_i}] = P\{X_i = 1\}e^{t1} + P\{X_i = 0\}e^{t0} = pe^t + q$$

with $p + q = 1$. The moment-generating function of X is

$$M_X(t) = (pe^t + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{tk}.$$

This could have been derived directly:

$$M_X(t) = E[e^{tX}] = \sum_{k=0}^{n} P\{X = k\}e^{tk} = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{tk} = (pe^t + q)^n$$

by the binomial theorem.

2. *Poisson.* We have $P\{X = k\} = e^{-\lambda}\lambda^k/k!, \quad k = 0, 1, 2, \ldots$. Thus

$$M(t) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} e^{tk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = \exp(-\lambda)\exp(\lambda e^t) = \exp[\lambda(e^t - 1)].$$

We can compute the mean and variance from the moment-generating function:

$$E(X) = M'(0) = [\exp(\lambda(e^t - 1))\lambda e^t]_{t=0} = \lambda.$$

Let $h(\lambda, t) = \exp[\lambda(e^t - 1)]$. Then

$$E(X^2) = M''(0) = [h(\lambda, t)\lambda e^t + \lambda e^t h(\lambda, t)\lambda e^t]_{t=0} = \lambda + \lambda^2$$

hence

$$\operatorname{Var} X = E(X^2) - [E(X)]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

3. *Normal(0,1).* The moment-generating function is

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Now $-(x^2/2) + tx = -(1/2)(x^2 - 2tx + t^2 - t^2) = -(1/2)(x - t)^2 + (1/2)t^2$ so

$$M(t) = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp[-(x - t)^2/2]\, dx.$$

The integral is the area under a normal density (mean t, variance 1), which is 1. Consequently,

$$M(t) = e^{t^2/2}.$$

4. *Normal*(μ, σ^2). If X is normal(μ, σ^2), then $Y = (X - \mu)/\sigma$ is normal(0,1). This is a good application of the density function method from Lecture 1:

$$f_Y(y) = \frac{f_X(x)}{|dy/dx|_{x=\mu+\sigma y}} = \sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2}.$$

We have $X = \mu + \sigma Y$, so

$$M_X(t) = E[e^{tX}] = e^{t\mu} E[e^{t\sigma Y}] = e^{t\mu} M_Y(t\sigma).$$

Thus

$$M_X(t) = e^{t\mu} e^{t^2\sigma^2/2}.$$

Remember this technique, which is especially useful when $Y = aX + b$ and the moment-generating function of X is known.

3.5 Theorem

If X is normal(μ, σ^2) and $Y = aX + b$, then Y is normal$(a\mu + b, a^2\sigma^2)$.
Proof. We compute

$$M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = e^{bt} M_X(at) = e^{bt} e^{at\mu} e^{a^2t^2\sigma^2/2}.$$

Thus

$$M_Y(t) = \exp[t(a\mu + b)] \exp(t^2a^2\sigma^2/2). \ \clubsuit$$

Here is another basic result.

3.6 Theorem

Let $X_1, \dots, X_n$ be independent, with X_i normal (μ_i, σ_i^2). Then $Y = \sum_{i=1}^n X_i$ is normal with mean $\mu = \sum_{i=1}^n \mu_i$ and variance $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.
Proof. The moment-generating function of Y is

$$M_Y(t) = \prod_{i=1}^n \exp(t\mu_i + t^2\sigma_i^2/2) = \exp(t\mu + t^2\sigma^2/2). \ \clubsuit$$

A similar argument works for the Poisson distribution; see Problem 4.

3.7 The Gamma Distribution

First, we define the *gamma function* $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy, \quad \alpha > 0$. We need three properties:

(a) $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, the *recursion formula*;

(b) $\Gamma(n + 1) = n!, n = 0, 1, 2, \dots$;

(c) $\Gamma(1/2) = \sqrt{\pi}$.

To prove (a), integrate by parts: $\Gamma(\alpha) = \int_0^\infty e^{-y} d(y^\alpha/\alpha)$. Part (b) is a special case of (a). For (c) we make the change of variable $y = z^2/2$ and compute

$$\Gamma(1/2) = \int_0^\infty y^{-1/2} e^{-y}\, dy = \int_0^\infty \sqrt{2} z^{-1} e^{-z^2/2} z\, dz.$$

The second integral is $2\sqrt{\pi}$ times half the area under the normal(0,1) density, that is, $2\sqrt{\pi}(1/2) = \sqrt{\pi}$.

The *gamma density* is

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

where α and β are positive constants. The moment-generating function is

$$M(t) = \int_0^\infty [\Gamma(\alpha)\beta^\alpha]^{-1} x^{\alpha-1} e^{tx} e^{-x/\beta}\, dx.$$

Change variables via $y = (-t + (1/\beta))x$ to get

$$\int_0^\infty [\Gamma(\alpha)\beta^\alpha]^{-1} \left(\frac{y}{-t + (1/\beta)} \right)^{\alpha-1} e^{-y} \frac{dy}{-t + (1/\beta)}$$

which reduces to

$$\frac{1}{\beta^\alpha} \left(\frac{\beta}{1 - \beta t} \right)^\alpha = (1 - \beta t)^{-\alpha}.$$

In this argument, t must be less than $1/\beta$ so that the integrals will be finite.

Since $M(0) = \int_{-\infty}^\infty f(x)\, dx = \int_0^\infty f(x)\, dx$ in this case, with $f \geq 0$, $M(0) = 1$ implies that we have a legal probability density. As before, moments can be calculated efficiently from the moment-generating function:

$$E(X) = M'(0) = -\alpha(1 - \beta t)^{-\alpha-1}(-\beta)|_{t=0} = \alpha\beta;$$

$$E(X^2) = M''(0) = -\alpha(-\alpha - 1)(1 - \beta t)^{-\alpha-2}(-\beta)^2|_{t=0} = \alpha(\alpha + 1)\beta^2.$$

Thus

$$\operatorname{Var} X = E(X^2) - [E(X)]^2 = \alpha\beta^2.$$

3.8 Special Cases

The *exponential density* is a gamma density with $\alpha = 1 : f(x) = (1/\beta)e^{-x/\beta}, x \geq 0$, with $E(X) = \beta, \quad E(X^2) = 2\beta^2, \quad \operatorname{Var} X = \beta^2$.

A random variable X has the *chi-square density* with r *degrees of freedom* ($X = \chi^2(r)$ for short, where r is a positive integer) if its density is gamma with $\alpha = r/2$ and $\beta = 2$. Thus

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}, \quad x \geq 0$$

and

$$M(t) = \frac{1}{(1-2t)^{r/2}}, \quad t < 1/2.$$

Therefore $E[\chi^2(r)] = \alpha\beta = r, \quad \mathrm{Var}[\chi^2(r)] = \alpha\beta^2 = 2r.$

3.9 Lemma

If X is normal(0,1) then X^2 is $\chi^2(1)$.

Proof. We compute the moment-generating function of X^2 directly:

$$M_{X^2}(t) = E[e^{tX^2}] = \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Let $y = \sqrt{1-2t}x$; the integral becomes

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \frac{dy}{\sqrt{1-2t}} = (1-2t)^{-1/2}$$

which is $\chi^2(1)$. ♣

3.10 Theorem

If $X_1, \dots, X_n$ are independent, each normal (0,1), then $Y = \sum_{i=1}^n X_i^2$ is $\chi^2(n)$.

Proof. By (3.9), each X_i^2 is $\chi^2(1)$ with moment-generating function $(1-2t)^{-1/2}$. Thus $M_Y(t) = (1-2t)^{-n/2}$ for $t < 1/2$, which is $\chi^2(n)$. ♣

3.11 Another Method

Another way to find the density of $Z = X + Y$ where X and Y are independent random variables is by the *convolution formula*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)\, dy.$$

To see this intuitively, reason as follows. The probability that Z lies near z (between z and $z + dz$) is $f_Z(z)\, dz$. Let us compute this in terms of X and Y. The probability that X lies near x is $f_X(x)\, dx$. Given that X lies near x, Z will lie near z if and only if Y lies near $z - x$, in other words, $z - x \leq Y \leq z - x + dz$. By independence of X and Y, this probability is $f_Y(z-x)\, dz$. Thus $f_Z(z)dz$ is a sum over z of terms of the form $f_X(x)\, dx\, f_Y(z-x)\, dz$. Cancel the dz's and replace the sum by an integral to get the result. A formal proof can be given using Jacobians.

3.12 The Poisson Process

This process occurs in many physical situations, and provides an application of the gamma distribution. For example, particles can arrive at a counting device, customers at a serving counter, airplanes at an airport, or phone calls at a telephone exchange. Divide the time interval $[0, t]$ into a large number n of small subintervals of length dt, so that $n\,dt = t$. If $I_i, i = 1, \dots, n$, is one of the small subintervals, we make the following assumptions:

(1) The probability of exactly one arrival in I_i is $\lambda\,dt$, where λ is a constant.

(2) The probability of no arrivals in I_i is $1 - \lambda\,dt$.

(3) The probability of more than one arrival in I_i is zero.

(4) If A_i is the event of an arrival in I_i, then the $A_i, i = 1, \dots, n$ are independent.

As a consequence of these assumptions, we have $n = t/dt$ Bernoulli trials with probability of success $p = \lambda dt$ on a given trial. As $dt \to 0$ we have $n \to \infty$ and $p \to 0$, with $np = \lambda t$. We conclude that the number $N[0, t]$ of arrivals in $[0, t]$ is Poisson (λt):

$$P\{N[0, t] = k\} = e^{-\lambda t}(\lambda t)^k/k!, k = 0, 1, 2, \dots .$$

Since $E(N[0, t]) = \lambda t$, we may interpret λ as the *average number of arrivals per unit time.*

Now let W_1 be the waiting time for the first arrival. Then

$$P\{W_1 > t\} = P\{\text{no arrival in [0,t]}\} = P\{N[0, t] = 0\} = e^{-\lambda t}, t \geq 0.$$

Thus $F_{W_1}(t) = 1 - e^{-\lambda t}$ and $f_{W_1}(t) = \lambda e^{-\lambda t}, t \geq 0$. From the formulas for the mean and variance of an exponential random variable we have $E(W_1) = 1/\lambda$ and $\operatorname{Var} W_1 = 1/\lambda^2$.

Let W_k be the (total) waiting time for the k-th arrival. Then W_k is the waiting time for the first arrival plus the time after the first up to the second arrival plus $\cdots$ plus the time after arrival $k - 1$ up to the k-th arrival. Thus W_k is the sum of k independent exponential random variables, and

$$M_{W_k}(t) = \left(\frac{1}{1 - (t/\lambda)}\right)^k$$

so W_k is gamma with $\alpha = k, \beta = 1/\lambda$. Therefore

$$f_{W_k}(t) = \frac{1}{(k-1)!}\lambda^k t^{k-1} e^{-\lambda t}, t \geq 0.$$

Problems

1. Let X_1 and X_2 be independent, and assume that X_1 is $\chi^2(r_1)$ and $Y = X_1 + X_2$ is $\chi^2(r)$, where $r > r_1$. Show that X_2 is $\chi^2(r_2)$, where $r_2 = r - r_1$.
2. Let X_1 and X_2 be independent, with X_i gamma with parameters α_i and $\beta_i, i = 1, 2$. If c_1 and c_2 are positive constants, find convenient sufficient conditions under which $c_1 X_1 + c_2 X_2$ will also have a gamma distribution.
3. If $X_1, \dots, X_n$ are independent random variables with moment-generating functions $M_1, \dots, M_n$, and $c_1, \dots, c_n$ are constants, express the moment-generating function M of $c_1 X_1 + \cdots + c_n X_n$ in terms of the M_i.

4. If $X_1, \ldots, X_n$ are independent, with X_i Poisson$(\lambda_i), i = 1, \ldots, n$, show that the sum $Y = \sum_{i=1}^n X_i$ has the Poisson distribution with parameter $\lambda = \sum_{i=1}^n \lambda_i$.
5. An unbiased coin is tossed independently n_1 times and then again tossed independently n_2 times. Let X_1 be the number of heads in the first experiment, and X_2 the number of *tails* in the second experiment. Without using moment-generating functions, in fact without any calculation at all, find the distribution of $X_1 + X_2$.

Lecture 4. Sampling From a Normal Population

4.1 Definitions and Comments

Let $X_1, \dots, X_n$ be iid. The *sample mean* of the X_i is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the *sample variance* is

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

If the X_i have mean μ and variance σ^2, then

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n\mu = \mu$$

and

$$\operatorname{Var}\overline{X} = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}X_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0 \quad \text{as} \quad n \to \infty.$$

Thus $\overline{X}$ is a good estimate of μ. (For large n, the variance of $\overline{X}$ is small, so $\overline{X}$ is concentrated near its mean.) The sample variance is an average squared deviation from the sample mean, but it is a biased estimate of the true variance σ^2:

$$E[(X_i - \overline{X})^2] = E[(X_i - \mu) - (\overline{X} - \mu)]^2 = \operatorname{Var}X_i + \operatorname{Var}\overline{X} - 2E[(X_i - \mu)(\overline{X} - \mu)].$$

Notice the *centralizing technique*: We subtract and add back the mean of X_i, which will make the cross terms easier to handle when squaring. The above expression simplifies to

$$\sigma^2 + \frac{\sigma^2}{n} - 2E[(X_i - \mu)\frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)] = \sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n}E[(X_i - \mu)^2].$$

Thus

$$E[(X_i - \overline{X})^2] = \sigma^2(1 + \frac{1}{n} - \frac{2}{n}) = \frac{n-1}{n}\sigma^2.$$

Consequently, $E(S^2) = (n-1)\sigma^2/n$, not σ^2. Some books define the sample variance as

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

where S^2 is our sample variance. This adjusted estimate of the true variance is unbiased (its expectation is σ^2), but *biased does not mean bad*. If we measure performance by asking for a small mean square error, the biased estimate is better in the normal case, as we will see at the end of the lecture.

4.2 The Normal Case

We now assume that the X_i are normally distributed, and find the distribution of S^2. Let $y_1 = \overline{x} = (x_1 + \cdots + x_n)/n, \quad y_2 = x_2 - \overline{x}, \dots, y_n = x_n - \overline{x}$. Then $y_1 + y_2 = x_2, \quad y_1 + y_3 = x_3, \dots, y_1 + y_n = x_n$. Add these equations to get $(n-1)y_1 + y_2 + \cdots + y_n = x_2 + \cdots + x_n$, or

$$ny_1 + (y_2 + \cdots + y_n) = (x_2 + \cdots + x_n) + y_1. \tag{1}$$

But $ny_1 = n\overline{x} = x_1 + \cdots + x_n$, so by cancelling $x_2, \dots, x_n$ in (1), $x_1 + (y_2 + \cdots + y_n) = y_1$. Thus we can solve for the x's in terms of the y's:

$$\begin{aligned} x_1 &= y_1 - y_2 - \cdots - y_n \\ x_2 &= y_1 + y_2 \\ x_3 &= y_1 + y_3 \\ &\vdots \\ x_n &= y_1 + y_n \end{aligned} \tag{2}$$

The Jacobian of the transformation is

$$d_n = \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} = \begin{vmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

To see the pattern, look at the 4 by 4 case and expand via the last row:

$$\begin{vmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{vmatrix} = (-1) \begin{vmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} + \begin{vmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix}$$

so $d_4 = 1 + d_3$. In general, $d_n = 1 + d_{n-1}$, and since $d_2 = 2$ by inspection, we have $d_n = n$ for all $n \geq 2$. Now

$$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum (x_i - \overline{x} + \overline{x} - \mu)^2 = \sum (x_i - \overline{x})^2 + n(\overline{x} - \mu)^2 \tag{3}$$

because $\sum (x_i - \overline{x}) = 0$. By (2), $x_1 - \overline{x} = x_1 - y_1 = -y_2 - \cdots - y_n$ and $x_i - \overline{x} = x_i - y_1 = y_i$ for $i = 2, \dots, n$. (Remember that $y_1 = \overline{x}$.) Thus

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = (-y_2 - \cdots - y_n)^2 + \sum_{i=2}^{n} y_i^2. \tag{4}$$

Now

$$f_{Y_1 \cdots Y_n}(y_1, \dots, y_n) = n f_{X_1 \cdots X_n}(x_1, \dots, x_n).$$

By (3) and (4), the right side becomes, in terms of the y_i's,

$$n\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[\frac{1}{2\sigma^2}\left(-\,(\sum_{i=2}^n y_i)^2 - \sum_{i=2}^n y_i^2 - n(y_1-\mu)^2\right)\right].$$

The joint density of $Y_1,\dots,Y_n$ is a function of y_1 times a function of $(y_2,\dots,y_n)$, so Y_1 and $(Y_2,\dots,Y_n)$ are independent. Since $\overline{X}=Y_1$ and [by (4)] S^2 is a function of $(Y_2,\dots,Y_n)$,

$$\boxed{\overline{X} \quad \text{and} \quad S^2 \quad \text{are independent}}$$

Dividing Equation (3) by σ^2 we have

$$\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{nS^2}{\sigma^2} + \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2.$$

But $(X_i-\mu)/\sigma$ is normal (0,1) and

$$\boxed{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \quad \text{is normal (0,1)}}$$

so $\chi^2(n)=(nS^2/\sigma^2)+\chi^2(1)$ with the two random variables on the right independent. If $M(t)$ is the moment-generating function of nS^2/σ^2, then $(1-2t)^{-n/2}=M(t)(1-2t)^{-1/2}$. Therefore $M(t)=(1-2t)^{-(n-1)/2}$, i.e.,

$$\boxed{\frac{nS^2}{\sigma^2} \quad \text{is} \quad \chi^2(n-1)}$$

The random variable

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n-1}}$$

is useful in situations where μ is to be estimated but the true variance σ^2 is unknown. It turns out that T has a "T distribution", which we study in the next lecture.

4.3 Performance of Various Estimates

Let S^2 be the sample variance of iid normal (μ,σ^2) random variables $X_1,\dots,X_n$. We will look at estimates of σ^2 of the form cS^2, where c is a constant. Once again employing the centralizing technique, we write

$$E[(cS^2-\sigma^2)^2]=E[(cS^2-cE(S^2)+cE(S^2)-\sigma^2)^2]$$

which simplifies to

$$c^2 \operatorname{Var} S^2 + (cE(S^2)-\sigma^2)^2.$$

Since nS^2/σ^2 is $\chi^2(n-1)$, which has variance $2(n-1)$, we have $n^2(\operatorname{Var} S^2)/\sigma^4 = 2(n-1)$. Also $nE(S^2)/\sigma^2$ is the mean of $\chi^2(n-1)$, which is $n-1$. (Or we can recall from (4.1) that $E(S^2) = (n-1)\sigma^2/n$.) Thus the mean square error is

$$\frac{c^2 2\sigma^4(n-1)}{n^2} + \big(c\frac{(n-1)}{n}\sigma^2 - \sigma^2\big)^2.$$

We can drop the σ^4 and use n^2 as a common denominator, which can also be dropped. We are then trying to minimize

$$c^2 2(n-1) + c^2(n-1)^2 - 2c(n-1)n + n^2.$$

Differentiate with respect to c and set the result equal to zero:

$$4c(n-1) + 2c(n-1)^2 - 2(n-1)n = 0.$$

Dividing by $2(n-1)$, we have $2c + c(n-1) - n = 0$, so $c = n/(n+1)$. Thus the best estimate of the form cS^2 is

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

If we use S^2 then $c = 1$. If we use the unbiased version then $c = n/(n-1)$. Since $[n/(n+1)] < 1 < [n/(n-1)]$ and a quadratic function decreases as we move toward its minimum, we see that the biased estimate S^2 is better than the unbiased estimate $nS^2/(n-1)$, but neither is optimal under the minimum mean square error criterion. Explicitly, when $c = n/(n-1)$ we get a mean square error of $2\sigma^4/(n-1)$ and when $c = 1$ we get

$$\frac{\sigma^4}{n^2}\big[2(n-1) + (n-1-n)^2\big] = \frac{(2n-1)\sigma^4}{n^2}$$

which is always smaller, because $[(2n-1)/n^2] < 2/(n-1)$ iff $2n^2 > 2n^2 - 3n + 1$ iff $3n > 1$, which is true for every positive integer n.

For large n all these estimates are good and the difference between their performance is small.

Problems

1. Let $X_1, \dots, X_n$ be iid, each normal (μ, σ^2), and let $\overline{X}$ be the sample mean. If c is a constant, we wish to make n large enough so that $P\{\mu - c < \overline{X} < \mu + c\} \ge .954$. Find the minimum value of n in terms of σ^2 and c. (It is independent of μ.)
2. Let $X_1, \dots, X_{n_1}, Y_1, \dots Y_{n_2}$ be independent random variables, with the X_i normal (μ_1, σ_1^2) and the Y_i normal (μ_2, σ_2^2). If $\overline{X}$ is the sample mean of the X_i and $\overline{Y}$ is the sample mean of the Y_i, explain how to compute the probability that $\overline{X} > \overline{Y}$.
3. Let $X_1, \dots, X_n$ be iid, each normal (μ, σ^2), and let S^2 be the sample variance. Explain how to compute $P\{a < S^2 < b\}$.
4. Let S^2 be the sample variance of iid normal (μ, σ^2) random variables $X_i, i = 1 \dots, n$. Calculate the moment-generating function of S^2 and from this, deduce that S^2 has a gamma distribution.

Lecture 5. The T and F Distributions

5.1 Definition and Discussion

The *T distribution* is defined as follows. Let X_1 and X_2 be independent, with X_1 normal (0,1) and X_2 chi-square with r degrees of freedom. The random variable $Y_1 = \sqrt{r}X_1/\sqrt{X_2}$ has the T distribution with r degrees of freedom.

To find the density of Y_1, let $Y_2 = X_2$. Then $X_1 = Y_1\sqrt{Y}_2/\sqrt{r}$ and $X_2 = Y_2$. The transformation is one-to-one with $-\infty < X_1 < \infty, X_2 > 0 \iff -\infty < Y_1 < \infty, Y_2 > 0$. The Jacobian is given by

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \sqrt{y_2/r} & y_1/(2\sqrt{ry_2}) \\ 0 & 1 \end{vmatrix} = \sqrt{y_2/r}.$$

Thus $f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(x_1, x_2)\sqrt{y_2/r}$, which upon substitution for x_1 and x_2 becomes

$$\frac{1}{\sqrt{2\pi}} \exp[-y_1^2 y_2/2r] \frac{1}{\Gamma(r/2)2^{r/2}} y_2^{(r/2)-1} e^{-y_2/2} \sqrt{y_2/r}.$$

The density of Y_1 is

$$\frac{1}{\sqrt{2\pi}\Gamma(r/2)2^{r/2}} \int_0^\infty y_2^{[(r+1)/2]-1} \exp[-(1 + (y_1^2/r))y_2/2]\, dy_2/\sqrt{r}.$$

With $z = (1 + (y_1^2/r))y_2/2$ and the observation that all factors of 2 cancel, this becomes (with y_1 replaced by t)

$$\frac{\Gamma((r+1)/2)}{\sqrt{r\pi}\Gamma(r/2)} \frac{1}{(1 + (t^2/r))^{(r+1)/2}}, -\infty < t < \infty,$$

the *T density* with r degrees of freedom.

In sampling from a normal population, $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ is normal (0,1), and nS^2/σ^2 is $\chi^2(n-1)$. Thus

$$\sqrt{n-1}\frac{(\overline{X} - \mu)}{\sigma/\sqrt{n}} \quad \text{divided by} \quad \sqrt{n}S/\sigma \quad \text{is} \quad T(n-1).$$

Since σ and $\sqrt{n}$ disappear after cancellation, we have

$$\boxed{\frac{\overline{X} - \mu}{S/\sqrt{n-1}} \quad \text{is} \quad T(n-1)}$$

Advocates of defining the sample variance with $n - 1$ in the denominator point out that one can simply replace σ by S in $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ to get the T statistic.

Intuitively, we expect that for large n, $(\overline{X} - \mu)/(S/\sqrt{n-1})$ has approximately the same distribution as $(\overline{X} - \mu)/(\sigma/\sqrt{n})$, i.e., normal (0,1). This is in fact true, as suggested by the following computation:

$$\left(1 + \frac{t^2}{r}\right)^{(r+1)/2} = \sqrt{\left(1 + \frac{t^2}{r}\right)^r} \left(1 + \frac{t^2}{r}\right)^{1/2} \to \sqrt{e^{t^2}} \times 1 = e^{t^2/2}$$

as $r \to \infty$.

5.2 A Preliminary Calculation

Before turning to the F distribution, we calculate the density of $U = X_1/X_2$ where X_1 and X_2 are independent, positive random variables. Let $Y = X_2$, so that $X_1 = UY, X_2 = Y$ (X_1, X_2, U, Y) are all greater than zero). The Jacobian is

$$\frac{\partial(x_1, x_2)}{\partial(u, y)} = \begin{vmatrix} y & u \\ 0 & 1 \end{vmatrix} = y.$$

Thus $f_{UY}(u, y) = f_{X_1X_2}(x_1, x_2)y = yf_{X_1}(uy)f_{X_2}(y)$, and the density of U is

$$h(u) = \int_0^\infty yf_{X_1}(uy)f_{X_2}(y)\,dy.$$

Now we take X_1 to be $\chi^2(m)$, and X_2 to be $\chi^2(n)$. The density of X_1/X_2 is

$$h(u) = \frac{1}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)}u^{(m/2)-1}\int_0^\infty y^{[(m+n)/2]-1}e^{-y(1+u)/2}\,dy.$$

The substitution $z = y(1+u)/2$ gives

$$h(u) = \frac{1}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)}u^{(m/2)-1}\int_0^\infty \frac{z^{[(m+n)/2]-1}}{[(1+u)/2]^{[(m+n)/2]-1}}e^{-z}\frac{2}{1+u}\,dz.$$

We abbreviate $\Gamma(a)\Gamma(b)/\Gamma(a+b)$ by $\beta(a, b)$. (We will have much more to say about this when we discuss the beta distribution later in the lecture.) The above formula simplifies to

$$h(u) = \frac{1}{\beta(m/2, n/2)}\,\frac{u^{(m/2)-1}}{(1+u)^{(m+n)/2}}, \quad u \ge 0.$$

5.3 Definition and Discussion

The F *density* is defined as follows. Let X_1 and X_2 be independent, with $X_1 = \chi^2(m)$ and $X_2 = \chi^2(n)$. With U as in (5.2), let

$$W = \frac{X_1/m}{X_2/n} = \frac{n}{m}U$$

so that

$$f_W(w) = f_U(u)\left|\frac{du}{dw}\right| = \frac{m}{n}f_U\left(\frac{m}{n}w\right).$$

Thus W has density

$$\frac{(m/n)^{m/2}}{\beta(m/2, n/2)}\,\frac{w^{(m/2)-1}}{[1 + (m/n)w]^{(m+n)/2}}, \quad w \ge 0,$$

the F *density* with m and n degrees of freedom.

5.4 Definitions and Calculations

The *beta function* is given by

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx, \quad a,b>0.$$

We will show that

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

which is consistent with our use of $\beta(a,b)$ as an abbreviation in (5.2). We make the change of variable $t=x^2$ to get

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\,dt = 2\int_0^\infty x^{2a-1}e^{-x^2}\,dx.$$

We now use the familiar trick of writing $\Gamma(a)\Gamma(b)$ as a double integral and switching to polar coordinates. Thus

$$\begin{aligned}\Gamma(a)\Gamma(b) &= 4\int_0^\infty\int_0^\infty x^{2a-1}y^{2b-1}e^{-(x^2+y^2)}\,dx\,dy\\ &= 4\int_0^{\pi/2} d\theta \int_0^\infty (\cos\theta)^{2a-1}(\sin\theta)^{2b-1}e^{-r^2}r^{2a+2b-1}\,dr.\end{aligned}$$

The change of variable $u=r^2$ yields

$$\int_0^\infty r^{2a+2b-1}e^{-r^2}\,dr = (1/2)\int_0^\infty u^{a+b-1}e^{-u}\,du = \Gamma(a+b)/2.$$

Thus

$$\frac{\Gamma(a)\Gamma(b)}{2\Gamma(a+b)} = \int_0^{\pi/2}(\cos\theta)^{2a-1}(\sin\theta)^{2b-1}\,d\theta.$$

Let $z=\cos^2\theta, 1-z=\sin^2\theta, dz=-2\cos\theta\sin\theta\,d\theta = -2z^{1/2}(1-z)^{1/2}\,d\theta$. The above integral becomes

$$-\frac{1}{2}\int_1^0 z^{a-1}(1-z)^{b-1}\,dz = \frac{1}{2}\int_0^1 z^{a-1}(1-z)^{b-1}\,dz = \frac{1}{2}\beta(a,b)$$

as claimed. The *beta density* is

$$f(x) = \frac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1}, \quad 0\le x\le 1 \quad (a,b>0).$$

Problems

1. Let X have the beta distribution with parameters a and b. Find the mean and variance of X.
2. Let T have the T distribution with 15 degrees of freedom. Find the value of c which makes $P\{-c \le T \le c\} = .95$.
3. Let W have the F distribution with m and n degrees of freedom (abbreviated $W = F(m,n)$). Find the distribution of $1/W$.
4. A typical table of the F distribution gives values of $P\{W \le c\}$ for $c = .9, .95, .975$ and $.99$. Explain how to find $P\{W \le c\}$ for $c = .1, .05, .025$ and $.01$. (Use the result of Problem 3.)
5. Let X have the T distribution with n degrees of freedom (abbreviated $X = T(n)$). Show that $T^2(n) = F(1,n)$, in other words, T^2 has an F distribution with 1 and n degrees of freedom.
6. If X has the exponential density $e^{-x}, x \ge 0$, show that $2X$ is $\chi^2(2)$. Deduce that the quotient of two exponential random variables is $F(2,2)$.