Development of Some Probability Distributions

George T. Gilbert
Department of Mathematics, Texas Christian University
g.gilbert@tcu.edu

April 11, 2019

We give a detailed sketch of how the exponential, Poisson, gamma, beta, normal, χ^2, t, and F distributions arise in "natural" ways. Our only prerequisite distributions will be the binomial distribution and the uniform distribution on $[0,1]$. The details of the (more than) healthy dose of multivariable calculus along the way may certainly be slurred over by those looking for a sense of how the distributions arise.

The Exponential Distribution

Suppose we model waiting time by a "memoryless" random variable, meaning that the wait from some instant until the next occurrence does not depend upon when the last occurrence was. In mathematical terms, we mean

$$P(t > t_1 | t > t_0) = P(t > t_1 - t_0),$$

where $t_1 \geq t_0 \geq 0$. If $f(t)$ denotes the density function of such a random variable, then

$$\frac{\int_{t_1}^{\infty} f(t)\, dt}{\int_{t_0}^{\infty} f(t)\, dt} = \int_{t_1 - t_0}^{\infty} f(t)\, dt. \tag{1}$$

Let $G(t) = \int_t^{\infty} f(s)\, ds$. Then (1) implies $G(a+b) = G(a)G(b)$. With a moderate amount of work, one may prove that such a function G that is also continuous must be an exponential function. Because $G(t)$ is a decreasing function of t, we may write $G(t)$ in the form $e^{-t/\beta}$ for some positive parameter β. Noting that $G(t) = G(0) - \int_0^t f(s)\, ds$ and using the fundamental theorem of calculus, $f(t) = -G'(t) = \dfrac{1}{\beta} e^{-t/\beta}$. This is the density function for the exponential distribution.

The Poisson Distribution

Consider a binomial random variable with enormous n and miniscule p, for instance modeling the number of customers at a store or accidents at a busy intersection. Specifically, we consider the behavior of binomial random variables with fixed $\lambda = np$ as n goes to ∞. If X is the number of occurrences, then we have

$$\begin{aligned}
\lim_{n\to\infty} P(X=x) &= \lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \frac{\lambda^x}{x!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^x}{x!}\, e^{-\lambda},
\end{aligned}$$

which is the Poisson distribution. Observe that $\sum_{n=0}^{\infty} \lambda^x/x!$ is the Taylor series at 0 for e^{λ}.

If we have Poisson random variables X_1 and X_2 with means λ_1 and λ_2, then we can think of $\lambda_1 \approx n_1 p$ and $\lambda_2 \approx n_2 p$. Now the sum of binomial random variables with common p is binomial, suggesting that the sum of X_1 and X_2 should be Poisson with mean $\lambda_1 + \lambda_2$. In fact,

$$\begin{aligned}
P(X_1 + X_2 = n) &= \sum_{k=0}^{n} P(X_1 = k) P(X_2 = n-k) = \sum_{k=0}^{n} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \, \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \left(\lambda_1 + \lambda_2\right)^n .
\end{aligned}$$

Suppose we have a Poisson random variable with mean occurrence rate λ per unit time, so that the number of arrivals in time t has a Poisson distribution X_t with mean λt. Suppose we want, not the distribution of the number of occurrences over a time t, but the wait T until the next occurrence. Then the cumulative distribution function of T is given by

$$P(T \le t) = P(X_t \ge 1) = 1 - P(X_t = 0) = 1 - e^{-\lambda t}.$$

Its derivative is the density function, namely $\lambda e^{-\lambda t}$. In, other words the waiting time has an exponential distribution with $\beta = 1/\lambda$.

Conversely, given an exponential density function with parameter λ, the probability of no arrival within a time t is $e^{-\lambda t}$. We proceed by induction. Assume that the probability of exactly n arrivals is $(\lambda t)^n e^{-\lambda t}/n!$ for all $t > 0$. Then, conditioning on the first arrival being at time τ, the probability of exactly $n+1$ arrivals within a time t is

$$\int_0^t \lambda e^{-\lambda\tau} \cdot \frac{(\lambda(t-\tau))^n}{n!} e^{-\lambda(t-\tau)}\, d\tau = e^{-\lambda t} \int_0^t \frac{\lambda^{n+1}(t-\tau)^n}{n!}\, d\tau = e^{-\lambda t} \cdot \frac{\lambda^{n+1} t^{n+1}}{(n+1)!} = \frac{(\lambda t)^{n+1}}{n+1!} e^{-\lambda t}.$$

The Gamma Function

When evaluating the mean and variance for the exponential distribution, one evaluates the integrals

$$\int_0^\infty t\ \frac{1}{\beta} e^{-t/\beta}\ dt \quad \text{and} \quad \int_0^\infty t^2\ \frac{1}{\beta} e^{-t/\beta}\ dt.$$

Substituting $x = t/\beta$, one obtains the integrals

$$\beta \int_0^\infty x\ e^{-x}\ dx \quad \text{and} \quad \beta^2 \int_0^\infty x^2\ e^{-x}\ dx. \tag{2}$$

With these integrals in mind, we define the gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\ dx,$$

which converges for $\alpha > 0$. It is easy to verify that $\Gamma(1) = 1$. For $\alpha > 1$, integration by parts yields $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. Thus, $\Gamma(n) = (n-1)!$ for $n = 1, 2, \ldots$. It immediately follows from (2) that the mean and variance of the exponential distribution are β and $2\beta^2 - \beta^2 = \beta^2$, respectively.

We now show that $\Gamma(1/2) = \sqrt{\pi}$ and hence that

$$\Gamma(n + 1/2) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi} = \frac{(2n)!}{2^{2n} \cdot n!} \sqrt{\pi}.$$

The substitution $x = u^2/2$ leads to

$$\Gamma(1/2) = \int_0^\infty x^{-1/2} e^{-x}\ dx = \int_0^\infty \sqrt{2} e^{-u^2/2}\ du = \int_{-\infty}^\infty e^{-u^2/2}/\sqrt{2}\ du.$$

Thus,

$$[\Gamma(1/2)]^2 = \int_{-\infty}^\infty e^{-u^2/2}/\sqrt{2}\ du \cdot \int_{-\infty}^\infty e^{-v^2/2}/\sqrt{2}\ dv = \frac{1}{2} \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(u^2+v^2)/2}\ du\ dv.$$

This is an integral over the uv-plane, which becomes easy if we change to polar coordinates. Then

$$[\Gamma(1/2)]^2 = \frac{1}{2} \int_0^{2\pi} \int_0^\infty e^{-r^2/2} r\ dr\ d\theta = \frac{1}{2} \int_0^{2\pi} d\theta = \pi,$$

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

In a heuristic sense, the gamma function should be thought of as an extension of $(n-1)!$ to non-integral values of n.

The Gamma Distribution

If the waiting time to the next occurrence has an exponential distribution, then the waiting time until the nth occurrence is the sum of n identical, independent exponentially-distributed random variables. This distribution will turn out to be a particular case ($\alpha = n$) of the gamma distribution, which has density function

$$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta},\ x > 0,\ \alpha > 0,\ \beta > 0.$$

Upon substituting $u = x/\beta$, we note that

$$\int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}\, dx = \frac{1}{\Gamma(\alpha)} \int_0^\infty u^{\alpha-1} e^{-u}\, du = 1.$$

The exponential distribution is the special case $\alpha = 1$. We have just shown that

$$\int_0^\infty x^{\alpha-1+k} e^{-x/\beta}\, dx = \beta^{\alpha+k}\Gamma(\alpha+k),$$

so the mean $\alpha\beta$ and variance $\alpha\beta^2$ for a gamma distribution follow easily.

We now show that the sum of n independent random variables with gamma distributions with parameters α_k and β has a gamma distribution with parameters $\alpha_1 + \cdots + \alpha_n$ and β. In particular, this will prove our claim for the sum of exponential random variables. Moment generating functions provide a fairly easy way to prove this. However, we will instead take a direct approach. By iteration or induction, it suffices to prove the case $n = 2$. Let x and y denote the values from the original gamma distributions and let w denote the value of the sum. Then the cumulative distribution function for the sum is given by

$$\int_0^w \int_0^{w-x} \frac{1}{\beta^{\alpha_1}\Gamma(\alpha_1)} x^{\alpha_1-1} e^{-x/\beta}\, \frac{1}{\beta^{\alpha_2}\Gamma(\alpha_2)} y^{\alpha_2-1} e^{-y/\beta}\, dy\, dx.$$

We will change variables in the integral to $s = x+y$ and $u = x/(x+y)$. Solving for x and y, we get $x = su$, $y = s(1-u)$. Our domain $0 \le x$, $0 \le y$, $x + y \le w$ transforms to $0 \le s \le w$, $0 \le u \le 1$. The jacobian matrix ("chain rule factor") for this transformation is

$$\left| \det \begin{bmatrix} \dfrac{\partial x}{\partial s} & \dfrac{\partial x}{\partial u} \\ \dfrac{\partial y}{\partial s} & \dfrac{\partial y}{\partial u} \end{bmatrix} \right| = \left| \det \begin{bmatrix} u & s \\ 1-u & -s \end{bmatrix} \right| = s,$$

so that $dx\, dy = s\, du\, ds$. We crunch out the algebra for the substitution and obtain that the cumulative distribution function equals

$$\begin{aligned} \int_0^w \int_0^1 & \frac{1}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} s^{\alpha_1+\alpha_2-1} u^{\alpha_1-1} (1-u)^{\alpha_2-1} e^{-s/\beta}\, du\, ds \\ = & \frac{1}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^w s^{\alpha_1+\alpha_2-1} e^{-s/\beta}\, ds \cdot \int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1}\, du. \end{aligned}$$

The second integral is just a constant $c(\alpha_1, \alpha_2)$ which depends on α_1 and α_2. Therefore, the cumulative distribution function simplifies to

$$\frac{c(\alpha_1, \alpha_2)}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^w s^{\alpha_1+\alpha_2-1} e^{-s/\beta}\, ds.$$

The density function is just the derivative of the cumulative distribution function. Hence, by the fundamental theorem of calculus, the density function equals

$$\frac{c(\alpha_1, \alpha_2)}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}\, w^{\alpha_1+\alpha_2-1} e^{-w/\beta}.$$

Up to the constant, this is the gamma distribution with parameters $\alpha_1 + \alpha_2$ and β, which is enough to show it must be this gamma distribution. Since the constant is unique, we must have

$$c(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

For (unnecessary) good measure, we show this identity directly as well. We have

$$\begin{aligned} \Gamma(\alpha+\beta) \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx &= \int_0^\infty y^{\alpha+\beta-1} e^{-y}\, dy \cdot \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx \\ = \int_0^\infty \int_0^1 y^{\alpha+\beta-1} e^{-y} x^{\alpha-1}(1-x)^{\beta-1}\, dx\, dy &= \int_0^\infty \int_0^1 (xy)^{\alpha-1}((1-x)y)^{\beta-1} e^{-y} y\, dx\, dy. \end{aligned}$$

We now make the substitution $u = xy$, $v = (1-x)y$. The jacobian ("chain rule factor") for this transformation is

$$\left| \det \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \right| = \left| \det \begin{bmatrix} y & x \\ -y & 1-x \end{bmatrix} \right| = y,$$

so that $du\, dv = y\, dx\, dy$. Every pair of positive u and v corresponds to exactly one pair x and y with $0 < x < 1$ and $0 < y$. Thus,

$$\begin{aligned} \Gamma(\alpha+\beta) \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx &= \int_0^\infty \int_0^\infty u^{\alpha-1} v^{\beta-1} e^{-(u+v)}\, du\, dv \\ &= \int_0^\infty u^{\alpha-1} e^{-u}\, du \cdot \int_0^\infty v^{\beta-1} e^{-v}\, dv = \Gamma(\alpha)\Gamma(\beta). \end{aligned}$$

Now suppose we have exponential waiting time. Then the probability of exactly n occurrences in time t is the probability of at least n or occurrences minus the probability of at least $n+1$ occurrences. In terms of the gamma distribution, this is

$$\int_0^t \frac{1}{\beta^n (n-1)!} x^{n-1} e^{-x/\beta}\, dx - \int_0^t \frac{1}{\beta^{n+1} n!} x^n e^{-x/\beta}\, dx.$$

Perform one step of integration by parts on the first integral. The remaining integral equals that of the second integral. This yields that the probability of n occurrences is $\frac{1}{\beta^n n!} e^{-t/\beta}$, i.e. we have a Poisson distribution with $\lambda = 1/\beta$.

The Beta Distribution

Suppose we have $m+n+1$ independent random variables that are uniformly distributed on $(0,1)$. We wish to find the distribution of the $(m+1)$st smallest value or, equivalently, the $(n+1)$st largest value. This is an example of what is known as *order statistics*. The probability that a value from the uniform distribution on $(0,1)$ is less than x is x. Now, for a random variable X with density function $f(x)$, view $f(x)\ dx$ as the "probability" that $X=x$, which we add up, i.e. integrate, over an interval to get true probability for a continuous distribution. The "probability" that x is the $(m+1)$st smallest value is then the probability of choosing one variable to equal x, m to be smaller, and n larger,

$$((m+n+1)\ dx)\binom{m+n}{m}x^m(1-x)^n = \frac{(m+n+1)!}{m!n!}x^m(1-x)^n\ dx.$$

The generalization of this via gamma functions is the beta distribution, with density

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1},\ 0<x<1, \alpha>0,\ \beta>0.$$

Observe that this is a positive function. It will be a density function if

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\ dx = 1$$

or, equivalently,

$$\Gamma(\alpha+\beta)\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\ dx = \Gamma(\alpha)\Gamma(\beta).$$

In the last section, we saw exactly this integral in our derivation of the density for the sum of gamma distributions.

Stirling's Approximation

One of several forms of Stirling's approximation is

$$n! \sim \sqrt{2\pi}\ n^{n+1/2}e^{-n},$$

where the asymptotic symbol $\sim$ means here that the ratio has limit 1 as n goes to infinity.

One can get a slightly weaker estimate from applying the trapezoidal and midpoint estimates to two integrals of the form $\int_a^{n+b} \ln x\ dx$, (a and b constants).

We will sketch a proof using the gamma distribution. We begin more generally with the total gamma distribution

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha+1)}x^\alpha e^{-x}\ dx.$$

A quick max-min calculation shows that the maximum value of $x^\alpha e^{-x}$ occurs at $x=\alpha$. In fact, the function drops off rather sharply on both sides of $x=\alpha$. Chebyshev's theorem says that the probability of being within $k>1$ standard deviations of the mean is at least

$1 - 1/k^2$. Because $[\alpha - \alpha^{3/5}, \alpha + \alpha^{3/5}]$ represents essentially $\alpha^{1/10}$ standard deviations of $\sqrt{\alpha}$ from the mean of $\alpha + 1$, it follows that

$$\Gamma(\alpha+1) \approx \int_{\alpha-\alpha^{3/5}}^{\alpha+\alpha^{3/5}} x^{\alpha} e^{-x} \, dx,$$

where the error in the estimate divided by $\Gamma(\alpha+1)$ goes to 0 as α goes to infinity. The error terms in all of our subsequent estimates will share this property.

We now write $x^{\alpha} e^{-x}$ as $e^{\alpha \ln x - x}$. The Taylor series expansion of $\alpha \ln x - x$ about $x = \alpha$ is

$$\alpha \ln \alpha - \alpha - \frac{1}{2\alpha}(x-\alpha)^2 + \frac{1}{3\alpha^2}(x-\alpha)^3 - \frac{1}{4\alpha^3}(x-\alpha)^4 + \frac{1}{5\alpha^4}(x-\alpha)^5 - \cdots,$$

which converges to $\alpha \ln x - x$ on an interval containing $[\alpha - \alpha^{3/5}, \alpha + \alpha^{3/5}]$. Moreover, the sum of the terms of degree at least 3 goes (uniformly) to 0 on this interval as α goes to infinity. Thus,

$$\Gamma(\alpha+1) \approx \int_{\alpha-\alpha^{3/5}}^{\alpha+\alpha^{3/5}} e^{\alpha \ln \alpha - \alpha - (x-\alpha)^2/(2\alpha)} \, dx = \alpha^{\alpha} e^{-\alpha} \int_{\alpha-\alpha^{3/5}}^{\alpha+\alpha^{3/5}} e^{-(x-\alpha)^2/(2\alpha)} \, dx.$$

Upon substitution of $u = (x-\alpha)/\sqrt{\alpha}$, this becomes

$$\Gamma(\alpha+1) \approx \alpha^{\alpha+1/2} e^{-\alpha} \int_{-\alpha^{1/10}}^{\alpha^{1/10}} e^{-u^2/2} \, du.$$

This integral converges to $\displaystyle\int_{-\infty}^{\infty} e^{-u^2/2} \, du$, which we saw equals $\sqrt{2}\ \Gamma(1/2) = \sqrt{2\pi}$ in the section on the gamma function. The final estimate is Stirling's approximation,

$$\Gamma(\alpha+1) \sim \sqrt{2\pi}\ \alpha^{\alpha+1/2} e^{-\alpha}.$$

An Axiomatic Approach to The Normal Distribution

We proceed to derive the normal distribution along the lines of Gauss, as reported in [1]. Suppose the difference from the mean μ has continuously differentiable density function $f(x-\mu)$. Suppose further that f is an even function and that the sign of $f'(x-\mu)$ is opposite that of $x-\mu$ (i.e. f gets smaller as $|x-\mu|$ gets larger). Finally, assume that the maximal likelihood estimator for μ based on a random sample is the sample mean.

Consider a sample with values 0 and n values $(n+1)y$, with sample mean ny. Our assumptions imply that the maximum value of the joint density

$$f(-\mu)\,[f((n+1)y-\mu)]^n$$

occurs at $\mu = ny$. Setting the derivative at $\mu = ny$ of this joint density equal to 0, we see that

$$\frac{f'(-ny)}{f(-ny)} = -n\frac{f'(y)}{f(y)}.$$

Since f'/f is odd, we have

$$\frac{f'(ny)}{f(ny)} = n\frac{f'(y)}{f(y)}$$

for all integers n. Then

$$\frac{f'(\frac{m}{n}y)}{f(\frac{m}{n}y)} = m\frac{f'(\frac{1}{n}y)}{f(\frac{1}{n}y)} = \frac{m}{n}\frac{f'(y)}{f(y)}$$

for all integers m and $n \neq 0$. Continuity implies

$$\frac{f'(ry)}{f(ry)} = r\frac{f'(y)}{f(y)}$$

for all real r. Setting $r = 1/y$, we see that

$$\frac{f'(y)}{f(y)} = y\frac{f'(1)}{f(1)}$$

(which also holds for $y = 0$). Writing the negative constant $f'(1)/f(1)$ as $1/\sigma^2$, this differential equation has solution

$$f(y) = Ce^{-\frac{y^2}{2\sigma^2}}.$$

Since

$$\int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}}\,dy = \sqrt{2\pi}\sigma,$$

$C = 1/(\sqrt{2\pi}\sigma)$, so that the density for X is

$$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

that of the normal distribution with mean μ and variance σ^2.

Special Cases of the Central Limit Theorem

In this section, we show how the normal distribution arises naturally as limits of binomial, Poisson, and gamma distributions. These are special cases of the vitally important central limit theorem. Because the standard deviation represents a "typical" deviation from the mean, it is natural to measure the deviation of one value x from the mean in the context of this particular distribution by how many standard deviations it lies from the mean, i.e. by its z-score $z = (x - \mu)/\sigma$. The suggested importance of the z-score is, perhaps, reinforced by Chebyshev's theorem which says that the probability of being within $k > 1$ standard deviations of the mean is at least $1 - 1/k^2$.

For a binomial random variable X with parameters n and p,

$$P(X = x) = \binom{n}{x}p^x q^{n-x} = \frac{n!}{x!(n-x)!}\,p^x q^{n-x}, \quad x = 0, 1, \ldots, n,$$

where $q = 1 - p$. By Stirling's approximation,

$$P(X = x) \approx \frac{n^{n+1/2}}{\sqrt{2\pi}\; x^{x+1/2}(n-x)^{n-x+1/2}}\; p^x q^{n-x}.$$

The mean and standard deviation for the binomial distribution are np and $\sqrt{npq}$, respectively. We make the substitution $x = np + z\sqrt{npq}$ (so z is the z-score). Because we are making the transition to a continuous distribution, we need to include the chain rule factor $dx = \sqrt{npq}\; dz$ in our estimate. Substituting and simplifying, our approximation for $P(X = x)$ translates to density function

$$\frac{n^{n+1}}{\sqrt{2\pi}\; (np + z\sqrt{npq})^{np+z\sqrt{npq}+1/2}\; (nq - z\sqrt{npq})^{nq-z\sqrt{npq}+1/2}}\; p^{np+z\sqrt{npq}+1/2} q^{nq-z\sqrt{npq}+1/2}$$

$$= \frac{1}{\sqrt{2\pi}\; \left(1 + z\sqrt{\frac{q}{np}}\right)^{np+z\sqrt{npq}+1/2} \left(1 - z\sqrt{\frac{p}{nq}}\right)^{nq-z\sqrt{npq}+1/2}}.$$

The Taylor series expansion at 0 of $\ln(1 + u)$ is $u - u^2/2 + u^3/3 - \cdots$. Hence,

$$\ln \left(1 + z\sqrt{\frac{q}{np}}\right)^{np+z\sqrt{npq}+1/2}$$

$$= \quad (np + z\sqrt{npq} + 1/2) \left\{ z\sqrt{\frac{q}{np}} - \left(z\sqrt{\frac{q}{np}}\right)^2 \Big/ 2 + \left(z\sqrt{\frac{q}{np}}\right)^3 \Big/ 3 - \cdots \right\}.$$

Distributing this out yields the estimate

$$\ln \left(1 + z\sqrt{\frac{q}{np}}\right)^{np+z\sqrt{npq}+1/2} \approx npz\sqrt{\frac{q}{np}} - np\left(z\sqrt{\frac{q}{np}}\right)^2 \Big/ 2 + z\sqrt{npq}\; z\sqrt{\frac{q}{np}} = z\sqrt{npq} + z^2 q/2,$$

with the sum of all remaining terms going to 0 as n goes to infinity. Similarly (replace z with $-z$ and interchange p and q), $\ln \left(1 - z\sqrt{\frac{p}{nq}}\right)^{nq-z\sqrt{npq}+1/2}$ is approximately

$$-z\sqrt{npq} + z^2 p/2$$

as n goes to infinity. Therefore,

$$\left(1 + z\sqrt{\frac{q}{np}}\right)^{np+z\sqrt{npq}+1/2} \left(1 - z\sqrt{\frac{p}{nq}}\right)^{nq-z\sqrt{npq}+1/2}$$

$$\approx e^{z\sqrt{npq}+z^2q/2} e^{-z\sqrt{npq}+z^2p/2} = e^{z^2/2}.$$

We conclude

$$P(X = x) \approx \frac{1}{\sqrt{2\pi}}\; e^{-z^2/2}$$

as n goes to infinity, which is the density function for the standard normal distribution.

A similar derivation from the Poisson distribution is a bit easier. A Poisson distribution with parameter $n\lambda$ has mean and variance $n\lambda$ and probability function $p(x) = (n\lambda)^x e^{-n\lambda}/x!$. We transform to z-scores via $x = n\lambda + \sqrt{n\lambda}\, z$ and let $n \to \infty$. When we include the chain rule factor, the density for a continuous z is approximately

$$\frac{(n\lambda)^{n\lambda+\sqrt{n\lambda}\,z}}{(n\lambda+\sqrt{n\lambda}\,z)!}\, e^{-n\lambda}.$$

We apply Stirling's approximation to the factorial, obtaining approximate density

$$\frac{(n\lambda)^{n\lambda+\sqrt{n\lambda}\,z+1/2}}{\sqrt{2\pi}(n\lambda+\sqrt{n\lambda}\,z)^{n\lambda+\sqrt{n\lambda}\,z+1/2}e^{-(n\lambda+\sqrt{n\lambda}\,z)}}\, e^{-n\lambda} = \frac{1}{\sqrt{2\pi}}\left(1+\frac{z}{\sqrt{n\lambda}}\right)^{-(n\lambda+\sqrt{n\lambda}\,z+1/2)} e^{\sqrt{n\lambda}\,z}.$$

The Taylor series expansion of

$$\sqrt{n\lambda}\,z - (n\lambda+\sqrt{n\lambda}\,z)\ln\left(1+\frac{z}{\sqrt{n\lambda}}\right)$$

is

$$\sqrt{n\lambda}\,z - (n\lambda+\sqrt{n\lambda}\,z)\left(\frac{z}{\sqrt{n\lambda}} - \frac{z^2}{2n\lambda} + \frac{z^3}{3(n\lambda)^{3/2}} - \cdots\right) = -\frac{z^2}{2} + \frac{z^3}{6\sqrt{n\lambda}} - \frac{z^4}{12n\lambda} + \cdots .$$

Noting that $\left(1+\frac{z}{\sqrt{n\lambda}}\right)^{-1/2} \to \infty$ as $n \to \infty$, we conclude the density approaches $\frac{1}{\sqrt{2\pi}}e^{-z^2/2}$.

In the case of the gamma distribution,we will keep β fixed and let α go to infinity. As before, we can transform to z-scores for the gamma distribution, $x = \alpha\beta + \sqrt{\alpha}\,\beta z$ with $dx = \sqrt{\alpha}\,\beta\, dz$, obtaining the density of z is

$$\frac{\alpha^{1/2}}{\beta^{\alpha-1}\Gamma(\alpha)}(\alpha\beta+\sqrt{\alpha}\,\beta z)^{\alpha-1}e^{-(\alpha\beta+\sqrt{\alpha}\,\beta z)/\beta} = \frac{\alpha^{\alpha+1/2}}{\Gamma(\alpha+1)}(1+z/\sqrt{\alpha})^{\alpha-1}e^{-(\alpha+\sqrt{\alpha}\,z)}.$$

We use Stirling's approximation, $\Gamma(\alpha+1) \sim \sqrt{2\pi}\,\alpha^{\alpha+1/2}e^{-\alpha}$, to obtain an asymptotic estimate for the density

$$\frac{1}{\sqrt{2\pi}}(1+z/\sqrt{\alpha})^{\alpha-1}e^{-\sqrt{\alpha}\,z} = \frac{1}{\sqrt{2\pi}}e^{(\alpha-1)\ln(1+z/\sqrt{\alpha})-\sqrt{\alpha}\,z}.$$

Now

$$\begin{aligned}(\alpha-1)\ln(1+z/\sqrt{\alpha}) - \sqrt{\alpha}\,z &= (\alpha-1)(z/\sqrt{\alpha} - z^2/(2\alpha) + z^3/(3\alpha^{3/2}) - \cdots) - \sqrt{\alpha}\,z \\ &\sim -z^2/2.\end{aligned}$$

Thus, our asymptotic estimate for the density is $\frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. The special case of replacing α with $n\alpha$ and letting $n \to \infty$ is another case of the central limit theorem.

Moment Generating Functions and the Central Limit Theorem

The moment generating function defined to be $M_X(t) = \mathrm{E}[e^{tX}]$. Random variables with distinct cumulative distribution functions have distinct moment generating functions, though this is hard to prove. It is easily seen to satisfy $M_{aX+b}(t) = e^{bt}M_X(at)$ and, for X and Y independent $M_{X+Y}(t) = M_X(t)M_Y(t)$. If X has mean μ and variance σ^2, then $M_X(0) = 1$, $M'_X(0) = \mu$, and $M''_X(0) = \mathrm{E}[X^2] = \sigma^2 + \mu^2$. The cumulant moment generating function is $\psi_M(t) = \ln M_X(t)$. We have $\psi_M(0) = 0$, $\psi'_M(0) = \mu$, and $\psi''_M(0) = \sigma^2$.

The moment generating function for the standard normal random variable is

$$\int_{-\infty}^{\infty} e^{tz}\,\frac{1}{\sqrt{2\pi}}\,e^{-z^2/2}\,dz = e^{t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-(z-t)^2/2}\,dz = e^{t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-u^2/2}\,du = e^{t^2/2}.$$

A general normal random variable X has the form $X = \mu + \sigma Z$, where Z is standard normal. Thus, the moment generating function of a general normal distribution is $e^{\mu t + \sigma^2 t^2/2}$.

If X is normal with mean μ_X and variance σ_X^2 , and Y is normal with mean μ_Y and variance σ_Y^2, and X and Y are independent, then the moment generating function for $X+Y$ is

$$e^{\mu_X t + \sigma_X^2 t^2/2}\,e^{\mu_Y t + \sigma_Y^2 t^2/2} = e^{(\mu_X+\mu_Y)t + (\sigma_X^2+\sigma_Y^2)t^2/2},$$

i.e. $X+Y$ is normal with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$.

Now consider an arbitrary random variable X with mean 0 and moment generating function is defined to be $M_X(t)$. Let X_i be independent random variables with the same distribution as X. Then the moment generating function for $(X_1 + \cdots + X_n)/\sqrt{n}$ is

$$[M_X(t/\sqrt{n})]^n = e^{n\psi_X(t/\sqrt{n})}.$$

Applying l'Hôpital's rule,

$$\begin{aligned}\lim_{n\to\infty} n\psi_X(t/\sqrt{n}) &= \lim_{n\to\infty}\frac{\psi_X(t/\sqrt{n})}{1/n} = \lim_{n\to\infty}\frac{\psi'_X(t/\sqrt{n})(-t/(2n^{3/2}))}{-1/n^2} = \lim_{n\to\infty}\frac{t\psi'_X(t/\sqrt{n})}{2/\sqrt{n}}\\ &= \lim_{n\to\infty}\frac{\psi''_X(t/\sqrt{n})(-t^2/(2n^{3/2}))}{-1/n^{3/2}} = \lim_{n\to\infty}\frac{t^2}{2}\psi''_X(t/\sqrt{n})\frac{t^2}{2}\psi''_X(0) = \frac{\sigma^2 t^2}{2}.\end{aligned}$$

Thus,

$$\lim_{n\to\infty}[M_X(t/\sqrt{n})]^n = e^{\sigma^2 t^2/2},$$

from which we conclude the limiting distribution of $(X_1 + \cdots + X_n)/\sqrt{n}$ is normal with mean 0 and variance σ^2. When X has mean μ the limiting distribution of the sample mean $(X_1 + \cdots + X_n)/n$ is "asymptotic" to that of a normal distribution with mean μ and variance σ^2/n.

The χ^2 Distribution and Sample Variance

We explain how to derive the distribution of the sample variance s^2 when the original population has a normal distribution.

Step 1. Prove the miraculous identity

$$\begin{aligned}(n-1)s^2 &= \left(\frac{x_1-x_2}{\sqrt{2}}\right)^2+\cdots+\left(\frac{x_1+x_2+\cdots+x_{k-1}-(k-1)x_k}{\sqrt{k^2-k}}\right)^2\\ &+ \quad \cdots+\left(\frac{x_1+x_2+\cdots+x_{n-1}-(n-1)x_n}{\sqrt{n^2-n}}\right)^2. \end{aligned} \tag{3}$$

The coefficient of x_k^2 in $(n-1)s^2$ is

$$\frac{(n-1)^2}{n^2}+(n-1)\frac{1}{n^2}=\frac{n-1}{n}.$$

The coefficient of x_k^2 in the right-hand side of (3) is

$$\begin{aligned}&\frac{(k-1)^2}{k^2-k}+\frac{1}{(k+1)^2-(k+1)}+\cdots+\frac{1}{n^2-n}\\ &=\frac{k-1}{k}+\left(\frac{1}{k}-\frac{1}{k+1}\right)+\cdots+\left(\frac{1}{n-1}-\frac{1}{n}\right)=\frac{k-1}{k}+\frac{1}{k}-\frac{1}{n}=\frac{n-1}{n}.\end{aligned}$$

The coefficient of x_jx_k , $j<k$, in $(n-1)s^2$ is

$$2\,\frac{-2(n-1)}{n^2}+(n-2)\,\frac{2}{n^2}=-\frac{2}{n}.$$

The coefficient of x_jx_k , $j<k$, in the right-hand side of (3) is

$$\begin{aligned}&\frac{-2(k-1)}{k^2-k}+\frac{2}{(k+1)^2-(k+1)}+\cdots+\frac{2}{n^2-n}\\ &=\frac{-2}{k}+\left(\frac{2}{k}-\frac{2}{k+1}\right)+\cdots+\left(\frac{2}{n-1}-\frac{2}{n}\right)\\ &=\frac{-2}{k}+\frac{2}{k}-\frac{2}{n}=-\frac{2}{n}.\end{aligned}$$

This completes the proof of (3).

Step 2. Show that $\{(X_1+X_2+\cdots+X_{k-1}-(k-1)X_k)/\sqrt{k^2-k}\}_{2\le k\le n}$ have mean 0 and variance σ^2, and along with $(X_1+X_2+\cdots+X_n)/n$ are independent, normal random variables.

In general, for independent random variables $X_1,\ldots,X_n$, with means $\mu_1,\ldots,\mu_n$ and variances $\sigma_1^2,\ldots,\sigma_n^2$, the mean and variance of $a_1X_1+\cdots+a_nX_n$ are $a_1\mu_1+\cdots+a_n\mu_n$ and $a_1^2\sigma_1^2+\cdots+a_n^2\sigma_n^2$, respectively. Moreover, $a_1X_1+\cdots+a_nX_n$ is normal if $X_1,\ldots,X_n$ are. We are left to show the independence.

The joint density of $X_1,\ldots,X_n$ is

$$\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\,e^{-\dfrac{(x_k-\mu)^2}{2\sigma^2}}=\frac{1}{(2\pi)^{n/2}\sigma^n}\,e^{-\dfrac{\sum_{k=1}^{n}x_k^2-2\mu\sum_{k=1}^{n}x_k+n\mu^2}{2\sigma^2}}.$$

Let

$$Y_1 = (X_1 + X_2 + \cdots + X_n)/\sqrt{n}$$

and

$$Y_k = (X_1 + X_2 + \cdots + X_{k-1} - (k-1)X_k)/\sqrt{k^2 - k}$$

for $k = 2, \ldots, n$. Let $\mathbf{x}$ and $\mathbf{y}$ denote the column vectors of the x_k and y_k, respectively. Then $\mathbf{y} = Q\mathbf{x}$, where Q is the orthogonal matrix ($Q^{-1} = Q^t$)

$$Q = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & & \cdots & & & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & & \cdots & & & 0 \\ & & & & \cdots & & & \\ \frac{1}{\sqrt{k^2-k}} & \frac{1}{\sqrt{k^2-k}} & \cdots & \frac{1}{\sqrt{k^2-k}} & -\frac{k-1}{\sqrt{k^2-k}} & 0 & \cdots & 0 \\ & & & & \cdots & & & \\ \frac{1}{\sqrt{n^2-n}} & \frac{1}{\sqrt{n^2-n}} & & & \cdots & & \frac{1}{\sqrt{n^2-n}} & -\frac{n-1}{\sqrt{n^2-n}} \end{bmatrix}.$$

The jacobian of the transformation from $\mathbf{x}$ to $\mathbf{y}$ is $|\det Q^{-1}| = 1$. Observe that $\sum_{k=1}^{n} x_k = \sqrt{n} y_1$ and

$$\sum_{k=1}^{n} x_k^2 = \mathbf{x}^t\mathbf{x} = \mathbf{x}^t Q^t Q\mathbf{x} = \mathbf{y}^t\mathbf{y} = \sum_{k=1}^{n} y_k^2.$$

Therefore the joint density function for $\mathbf{x}$ transforms to

$$\frac{1}{(2\pi)^{n/2}\sigma^n}\, e^{-\frac{\sum_{k=1}^{n} y_k^2 - 2\sqrt{n}\mu y_1 + n\mu^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(y_1 - \sqrt{n}\mu)^2}{2\sigma^2}} \prod_{k=2}^{n} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{y_k^2}{2\sigma^2}},$$

showing that $Y_1, \ldots, Y_n$ are independent. Furthermore, transformations $g_1(Y_1), \ldots, g_n(Y_n)$ of $Y_1, \ldots, Y_n$ are independent.

Step 3. Show that for Y normal with mean 0 and variance σ^2, the distribution of Y^2/σ^2 is χ^2 with 1 degree of freedom.

Let $w = y^2/\sigma^2$ or $y = \sigma\sqrt{w}$. Then $dy = \sigma/(2\sqrt{w})dw$. Hence

$$\frac{1}{\sqrt{2\pi}\sigma}\, e^{-y^2/(2\sigma^2)} dy = \frac{1}{2\sqrt{2\pi}}\, w^{-1/2} e^{-w/2} dw = \frac{1}{2}\, \frac{1}{2^{1/2}\Gamma(1/2)}\, w^{-1/2} e^{-w/2} dw.$$

However, this transformation is 2-1, namely both y and $-y$ lead to the same value of w. In other words, two different y-intervals will lead to the same w-interval. It follows that we must double the above expression to find the density function for w,

$$\frac{1}{2^{1/2}\Gamma(1/2)}\, w^{-1/2}e^{-w/2},$$

which is the density function of the χ^2 distribution with 1 degree of freedom.

Step 4. Show that the sum of $n-1$ independent χ^2 random variables is χ^2 with $n-1$ degrees of freedom.

Observe that $(n-1)s^2/\sigma^2$ is the sum of $n-1$ independent χ^2 random variables with 1 degree of freedom. Such a χ^2 distribution is a gamma distribution with $\alpha = 1/2$, $\beta = 2$. This step is now simply a special case of our more general result on the sum of gamma distributions.

The χ^2 Distribution and Multinomial Distributions

We show that the χ^2 distribution arises as a limit of multinomial distributions. The method will be essentially a higher dimensional generalization of our derivation of the normal distribution as a limit of binomial distributions. It will require several intermediate transformations. We have $\nu+1$ mutually exclusive states with the probability state k occurs p_k, $k = 1, 2, \ldots, \nu+1$, with $\sum_{k=1}^{\nu+1} p_k = 1$. In n independent trials, the probability of x_k occurrences of state k for every k is the multinomial probability

$$\frac{n!}{x_1!\cdots x_{\nu+1}!}\, p_1^{x_1}\cdots p_{\nu+1}^{x_{\nu+1}}.$$

By Stirling's approximation, this is approximately

$$(2\pi)^{-\nu/2}\,\frac{n^{n+1/2}}{x_1^{x_1+1/2}\cdots x_{\nu+1}^{x_{\nu+1}+1/2}}\, p_1^{x_1}\cdots p_{\nu+1}^{x_{\nu+1}}$$

$$= (2\pi n)^{-\nu/2}\left(\frac{x_1}{np_1}\right)^{-x_1-1/2}\cdots\left(\frac{x_{\nu+1}}{np_{\nu+1}}\right)^{-x_{\nu+1}-1/2} p_1^{-1/2}\cdots p_{\nu+1}^{-1/2}.$$

The variance of an individual x_k is $\sqrt{np_k(1-p_k)}$. With this in mind, we transform to the individual z-score variables z_k defined by

$$x_k = np_k + z_k\sqrt{np_k(1-p_k)}.$$

(The computations would turn out to be a bit simpler by defining $x_k = np_k + z_k\sqrt{np_k}$, but there is no way to know this except in hindsight.) Note that this is only a ν-dimensional problem because $\sum_{k=1}^{\nu+1} x_k = n$, which translates to $\sum_{k=1}^{\nu+1} z_k\sqrt{p_k(1-p_k)}. = 0$. Once again, we need the jacobian of the transformation for our translation to continuous variables z_k. Here, this is just the product of

$$\frac{dx_k}{dz_k} = \sqrt{np_k(1-p_k)},\ k = 1, \ldots, \nu.$$

The approximation for multinomial probability turns into the density function

$$(2\pi)^{-\nu/2}p_{\nu+1}^{-1/2}\prod_{k=1}^{\nu+1}\left(1+z_k\sqrt{\frac{1-p_k}{np_k}}\right)^{-np_k-z_k\sqrt{np_k(1-p_k)}-1/2}\prod_{k=1}^{\nu}\sqrt{1-p_k}$$

$$= (2\pi)^{-\nu/2}p_{\nu+1}^{-1/2}\prod_{k=1}^{\nu+1}\left(1+z_k\sqrt{\frac{1-p_k}{np_k}}\right)^{-np_k}\prod_{k=1}^{\nu+1}\left(1+z_k\sqrt{\frac{1-p_k}{np_k}}\right)^{-z_k\sqrt{np_k(1-p_k)}}$$
$$\times\prod_{k=1}^{\nu+1}\left(1+z_k\sqrt{\frac{1-p_k}{np_k}}\right)^{-1/2}\prod_{k=1}^{\nu}\sqrt{1-p_k}.$$

As n goes to infinity, each term in the third product goes to 1 and each term in the second to $e^{-(1-p_k)z_k^2}$. We expand the natural logarithm of the first product using the Taylor series of $\ln(1+u) = u - u^2/2 + u^3/3 - \cdots$ to obtain

$$\sum_{k=1}^{\nu+1}\ln\left(1+z_k\sqrt{\frac{1-p_k}{np_k}}\right)^{-np_k} = \sum_{k=1}^{\nu+1}-np_k\ln\left(1+z_k\sqrt{\frac{1-p_k}{np_k}}\right)$$
$$= \sum_{k=1}^{\nu+1}-\left(z_k\sqrt{np_k(1-p_k)} - z_k^2(1-p_k)/2 + z_k^3\,\frac{(1-p_k)^{3/2}}{\sqrt{np_k}}/3 - \cdots\right).$$

The sum of the terms beyond the second goes to 0 as n goes to infinity. The first term sums to 0 by our constraint on z_k. We conclude that, as n goes to infinity, the density function has limiting value

$$(2\pi)^{-\nu/2}p_{\nu+1}^{-1/2}e^{-\sum\limits_{k=1}^{\nu+1}(1-p_k)z_k^2/2}\prod_{k=1}^{\nu}\sqrt{1-p_k}.$$

The term in the exponent

$$\sum_{k=1}^{\nu+1}(1-p_k)z_k^2 = \sum_{k=1}^{\nu+1}\frac{(x_k-np_k)^2}{np_k}$$

is the χ^2 statistic that we want; call it x. We need to combine all of the multivariate densities with a common value of x to obtain its density function. We first let $w_k = \sqrt{1-p_k}\,z_k$. Recalling that only $z_1,\ldots,z_\nu$ are our independent variables, we transform the density function to

$$(2\pi)^{-\nu/2}p_{\nu+1}^{-1/2}\;e^{-\sum\limits_{k=1}^{\nu+1}w_k^2/2}.$$

Now

$$w_{\nu+1} = \frac{-w_1\sqrt{p_1}-\cdots-w_\nu\sqrt{p_\nu}}{\sqrt{p_{\nu+1}}}.$$

Note that the length of the vector $[-\sqrt{p_1},\ldots,-\sqrt{p_\nu}]$ is $\sqrt{p_1+\cdots+p_\nu} = \sqrt{1-p_{\nu+1}}$. Let Q be a $\nu\times\nu$ orthogonal matrix with first row

$$[-\sqrt{p_1}/\sqrt{1-p_{\nu+1}},\ldots,-\sqrt{p_\nu}/\sqrt{1-p_{\nu+1}}].$$

Let $[u_1, \ldots, u_\nu]^t = Q[w_1, \ldots, w_\nu]^t$. Then, as in the last section, the jacobian of this transformation is $|\det Q^{-1}| = 1$ and

$$w_1^2 + \cdots + w_\nu^2 = u_1^2 + \cdots + u_\nu^2.$$

In terms of the new variables $u_1, \ldots, u_\nu$, the density function is

$$(2\pi)^{-\nu/2} p_{\nu+1}^{-1/2}\ e^{-\left(\sum\limits_{k=1}^{\nu} u_k^2 + u_1^2(1-p_{\nu+1})/p_{\nu+1}\right)/2} = (2\pi)^{-\nu/2} p_{\nu+1}^{-1/2}\ e^{-\left(u_1^2/p_{\nu+1} + u_2^2 + \cdots + u_\nu^2\right)/2}.$$

The penultimate transformation is to let

$$y_1 = u_1/\sqrt{p_{\nu+1}}, y_2 = u_2, \ldots, y_\nu = u_\nu,$$

which yields the density

$$(2\pi)^{-\nu/2}\ e^{-\left(y_1^2 + \cdots + y_\nu^2\right)/2}.$$

(Recall that the x we defined earlier is now $y_1^2 + \cdots + y_\nu^2$).

The values of $y_1, \ldots, y_\nu$ yielding a particular value of x is the $(\nu-1)$-dimensional "surface area" of a ν-dimensional sphere of radius $r = \sqrt{x}$. (Another viewpoint is that we are integrating out the angular variables in the higher dimensional generalizations of polar and spherical coordinates.) As in the 1, 2, and 3-dimensional cases, we obtain this by differentiating the volume. We claim the volume $V_\nu(r)$ is given by

$$V_\nu(r) = \frac{2\pi^{\nu/2}}{\nu\Gamma(\nu/2)} r^\nu.$$

The formula holds for $\nu = 1$ (and 2 and 3), so we apply induction. Assume the formula for $\nu - 1$. Then by cross sections

$$\begin{aligned} V_\nu(r) &= r^n V_\nu(1) = r^\nu \int_{-1}^{1} V_{\nu-1}(\sqrt{1-s^2})\ ds \\ &= r^\nu \int_{-1}^{1} \frac{2\pi^{(\nu-1)/2}}{(\nu-1)\Gamma((\nu-1)/2)}\ (1-s^2)^{(\nu-1)/2}\ ds \\ &= r^\nu \frac{2\pi^{(\nu-1)/2}}{(\nu-1)\Gamma((\nu-1)/2)} \int_0^1\ 2(1-s^2)^{(\nu-1)/2}\ ds. \end{aligned}$$

We may evaluate the integral via a beta distribution integral ($\alpha = 1/2$, $\beta = (\nu+1)/2$) by substituting $t = s^2$:

$$\begin{aligned} V_\nu(r) &= r^\nu \frac{2\pi^{(\nu-1)/2}}{(\nu-1)\Gamma((\nu-1)/2)} \int_0^1 t^{-1/2}(1-t)^{(\nu-1)/2} dt \\ &= r^\nu \frac{2\pi^{(\nu-1)/2}}{(\nu-1)\Gamma((\nu-1)/2)}\ \frac{\Gamma(1/2)\Gamma((\nu+1)/2)}{\Gamma((\nu+2)/2)} = \frac{2\pi^{\nu/2}}{\nu\Gamma(\nu/2)} r^\nu, \end{aligned}$$

where we have used $\Gamma(1/2)\sqrt{\pi}$ and $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ (twice). Differentiation yields

$$\frac{2\pi^{\nu/2}}{\Gamma(\nu/2)} r^{\nu-1}.$$

Hence the density function in terms of r is

$$\frac{2}{2^{\nu/2}\Gamma(\nu/2)}\, r^{\nu-1}e^{-r^2/2}.$$

(You might note that this is twice the density function for the normal distribution when $\nu = 1$, the binomial case. The factor 2 arises because we implicitly assumed $r > 0$.) Our final transformation is $x = r^2$. We obtain density function

$$\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\, x^{\nu/2-1}e^{-x/2},$$

which is the density function for the χ^2 distribution with ν degrees of freedom.

Density Functions of Quotients of Random Variables

In the next two sections, we will compute density functions for the ratios of random variables with known densities. In this section we derive an expression for such a density function. Let X be a random variable with density function $f(x)$ and Y be a positive random variable with density function $g(y)$. In terms of the variable w, the cumulative distribution function of $W = X/Y$ is

$$\int_0^\infty \int_{-\infty}^{wy} f(x)g(y)\,dx\,dy.$$

By the fundamental theorem of calculus, the density function for W is

$$\frac{d}{dw}\int_0^\infty \int_{-\infty}^{wy} f(x)g(y)\,dx\,dy = \int_0^\infty \frac{d}{dw}\int_{-\infty}^{wy} f(x)g(y)\,dx\,dy = \int_0^\infty f(wy)g(y)y\,dy.$$

The t Distribution

We begin with a normally distributed population. We write

$$t = \frac{\dfrac{\overline{x}-\mu}{\sigma/\sqrt{n}}}{s/\sigma}.$$

Then $(\overline{x}-\mu)/(\sigma/\sqrt{n})$ has the standard normal distribution. Notation will be simplified by letting $\nu = n-1$ be the degrees of freedom. By transforming the χ^2 density for $(n-1)s^2/\sigma^2$, we obtain the following for the density for $y = s/\sigma$:

$$\frac{\nu^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)}y^{\nu-1}e^{-\nu y^2/2}.$$

From the previous section on quotients, the density function for t is

$$\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2y^2/2}\frac{\nu^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)}y^{\nu-1}e^{-\nu y^2/2}y\,dy = \frac{\nu^{\nu/2}}{\sqrt{2\pi}\;2^{\nu/2-1}\Gamma(\nu/2)}\int_0^\infty y^\nu e^{-(t^2+\nu)y^2/2}\,dy.$$

Substituting $u = (t^2 + \nu)y^2/2$, we obtain

$$\frac{\nu^{\nu/2}}{\sqrt{2\pi}\ 2^{\nu/2-1}\Gamma(\nu/2)} \frac{2^{(\nu-1)/2}}{(t^2+\nu)^{(\nu+1)/2}} \int_0^\infty u^{(\nu-1)/2} e^{-u} du$$
$$= \frac{\nu^{\nu/2}}{\sqrt{2\pi}\ 2^{\nu/2-1}\Gamma(\nu/2)} \frac{2^{(\nu-1)/2}}{(t^2+\nu)^{(\nu+1)/2}} \Gamma((\nu+1)/2) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2},$$

which is the density function for the t distribution.

The F Distribution

We derive the distribution of $f = \dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$, where s_k^2 is the sample variance from a sample of size n_k from a normal distribution with variance σ_k^2. The density function for $(n_k - 1)s_k^2/\sigma_k^2$ is χ^2 with $n_k - 1$ degrees of freedom. As in the previous section, we let $\nu_k = n_k - 1$. By transformation, the density function for s_k^2/σ_k^2 is

$$\frac{\nu_k^{\nu_k/2}}{2^{\nu_k/2}\Gamma(\nu_k/2)}\ y^{\nu_k/2-1} e^{-\nu_k y/2}.$$

From the section on density functions of quotients, we see that the density function for f is given by

$$\int_0^\infty \frac{\nu_1^{\nu_1/2}}{2^{\nu_1/2}\Gamma(\nu_1/2)} (fy)^{\nu_1/2-1} e^{-\nu_1 f y/2}\ \frac{\nu_2^{\nu_2/2}}{2^{\nu_2/2}\Gamma(\nu_2/2)}\ y^{\nu_2/2-1} e^{-\nu_2 y/2} y\ dy$$
$$= \frac{\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)} f^{\nu_1/2-1} \int_0^\infty y^{(\nu_1+\nu_2)/2-1} e^{-(\nu_1 f+\nu_2)y/2}\ dy.$$

We could substitute $u = (\nu_1 f + \nu_2)\, y/2$, which yields a gamma function integral. Instead, we make use of gamma distribution integrals to obtain

$$\frac{\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)} f^{\nu_1/2-1} \left(\frac{2}{\nu_1 f+\nu_2}\right)^{(\nu_1+\nu_2)/2} \Gamma((\nu_1+\nu_2)/2)$$
$$= \frac{\Gamma((\nu_1+\nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{f^{(\nu_1-2)/2}}{(1+\nu_1 f/\nu_2)^{(\nu_1+\nu_2)/2}},$$

which is the density function for the F distribution.

References

[1] Saul Stahl, "The Evolution of the Normal Distribution," Mathematics Magazine 79:2 (April 2006), 96-113.