Stat 5101 Lecture Slides: Deck 3

Probability and Expectation on Infinite Sample Spaces, Poisson, Geometric, Negative Binomial, Continuous Uniform, Exponential, Gamma, Beta, Normal, and Chi-Square Distributions

Charles J. Geyer
School of Statistics
University of Minnesota

Deja Vu

Now we go back to the beginning and do everything again.

Probability Mass Functions

A *probability mass function* (PMF) is a function

$$S \xrightarrow{f} \mathbb{R}$$

whose domain S, which can be any nonempty set, is called the *sample space*, whose codomain is the real numbers, and which satisfies the following conditions: its values are nonnegative

$$f(x) \geq 0, \qquad x \in S$$

and sum to one

$$\sum_{x \in S} f(x) = 1.$$

(Exactly the same as slide 20, deck 1.)

Infinite Sample Spaces

This time we allow infinite sample spaces. That means the sum

$$\sum_{x \in S} f(x) = 1$$

is an infinite series. So we are now using calculus.

Bernoulli Process

A *Bernoulli process* is an infinite sequence of random variables X_1, X_2, ... (a stochastic process), that are IID Ber(p).

Geometric Distribution

The number of zeros (failures) before the first one (success) in a Bernoulli process is a random variable Y that has the *geometric distribution* with success probability p, denoted Geo(p) for short.

Clearly, Y takes values in $\mathbb{N} = \{ 0, 1, 2, \dots \}$. Its PMF is given by

$$f_p(y) = \Pr(Y = y)$$

because that is the formula for *any* PMF.

Geometric Distribution (cont.)

If $Y = y$, then we know that the first y variables in the Bernoulli process have the value zero and that $X_{y+1} = 1$, and we don't know anything else about the rest of the infinite sequence X_1, X_2, ...

The probability of observing y failures and one success *in that order* is $(1-p)^y p$. There is no binomial coefficient, because there is only one order considered.

Hence the PMF of the Geo(p) distribution is

$$f_p(y) = p(1-p)^y, \qquad y = 0, 1, 2, \ldots.$$

Geometric Distribution (cont.)

With every brand name distribution comes a theorem that says the probabilities sum to one. For the geometric distribution, this theorem is

$$\sum_{y=0}^{\infty} p(1-p)^y = 1.$$

This is a special case of the geometric series (deck 2, slides 127 ff.)

$$\sum_{n=0}^{\infty} s^n = \frac{1}{1-s}$$

whenever $-1 < s < 1$.

Here $s = 1 - p$.

Geometric Distribution (cont.)

The geometric series only converges when $-1 < s < 1$, which is $-1 < 1 - p < 1$, which is $0 < p < 2$. Of course, we know $p \leq 1$ because p is a probability. Thus the parameter space of the geometric family of distributions is

$$\{ p \in \mathbb{R} : 0 < p \leq 1 \}$$

unlike the Bernoulli and binomial distributions $p = 0$ is not allowed.

What goes wrong is that when we try to sum the infinite series

$$\sum_{y=0}^{\infty} (1 - p)^y = 1 + 1 + 1 + \cdots$$

it does not converge.

Geometric Distribution (cont.)

So we had to be careful. The phrase "number of failures before the first success in a Bernoulli process" does not define a random variable when the success probability is $p = 0$ because the first success never happens!

Geometric Distribution (cont.)

There is also something different about the case $p = 1$. Then we do have a well defined random variable. The special case of

$$f_p(y) = p(1 - p)^y, \qquad y = 0, 1, 2, \ldots$$

when $p = 1$ gives

$$f_1(y) = \begin{cases} 1, & y = 0 \\ 0, & y > 0 \end{cases}$$

so the support is different. A Geo(p) random variable with $p = 1$ is concentrated at zero. It is a constant random variable.

Expectation

Just like the case where the sample space S is finite, in the case where the sample space is infinite the *expectation* of a random variable X is defined by

$$E(X) = \sum_{s \in S} X(s) f(s)$$

when this expression makes sense. But it doesn't always make sense.

Rearrangement of Series

One of the reasons for using the notation

$$E(X) = \sum_{s \in S} X(s) f(s)$$

instead of defining $S = \{s_1, s_2, \ldots\}$ and writing

$$E(X) = \sum_{i=1}^{\infty} X(s_i) f(s_i)$$

is that S is just a set. Its elements need not have a natural order. The sum should not depend on a particular enumeration $\{s_1, s_2, \ldots\}$.

But the sum of an infinite series can depend on the order of summation.

Rearrangement of Series (cont.)

$$(1-1)+\left(\frac{1}{2}-\frac{1}{2}\right)+\left(\frac{1}{3}-\frac{1}{3}\right)+\left(\frac{1}{4}-\frac{1}{4}\right)+\cdots$$

converges to zero, but

$$\left(1+\frac{1}{2}-1\right)+\left(\frac{1}{3}+\frac{1}{4}-\frac{1}{2}\right)+\left(\frac{1}{5}+\frac{1}{6}-\frac{1}{3}\right)+\cdots$$
$$=\frac{1}{2}+\frac{1}{12}+\frac{1}{30}+\cdots$$

converges to something positive, and both series have the same terms: one is the other rearranged.

Rearrangement of Series (cont.)

A theorem from advanced calculus (rearrangement of series).

A series of nonnegative terms has the same sum however rearranged. If the series diverges (in which case we say the sum is ∞), then it diverges however rearranged.

If a series

$$\sum_{i=1}^{\infty} a_i$$

converges absolutely, which means

$$\sum_{i=1}^{\infty} |a_i| < \infty,$$

then it converges to the same sum however rearranged.

Expectation (cont.)

In expectation theory, we don't mess around with series that do not converge absolutely.

The *expectation* of a random variable X is defined by

$$E(X) = \sum_{s \in S} X(s) f(s)$$

when the series on the right had side converges absolutely. Otherwise, we say the expectation does not exist.

The expectation operator E is a map $L^1(E) \to \mathbb{R}$, where $L^1(E)$ is the set of all random variables that do have expectation in this probability model.

Infinite Expectation

When X is a nonnegative random variable, we write $E(X) = \infty$ to indicate that the expectation of X does not exist and write $E(X) < \infty$ to indicate that the expectation of X does exist.

More generally, if $X = Y - Z$, where Y and Z are both nonnegative, we write

$$\begin{array}{llll} E(X) = +\infty & \text{if} \quad E(Y) = \infty & \text{and} & E(Z) < \infty \\ E(X) = -\infty & \text{if} \quad E(Y) < \infty & \text{and} & E(Z) = \infty \end{array}$$

There is nothing we can write to indicate the case $E(Y) = \infty$ and $E(Z) = \infty$, because $\infty - \infty$ has no sensible definition.

Expectation (cont.)

If we were going to pursue this subject in a logical manner, we would now develop tools to tell which expectations exist. But we defer that (to deck 6).

For now we just calculate a few expectations that do exist.

Unfortunately, the only distribution with infinite sample space we have already learned about has expectations that are tricky to calculate. Hence we first learn a new tool for calculating expectations.

Moment Generating Functions

Define $\varphi : \mathbb{R} \to \mathbb{R}$ by

$$\varphi(t) = E(e^{tX}), \qquad t \in \mathbb{R},$$

using the convention $\varphi(t) = \infty$ when the expectation does not exist.

If the function so defined is finite on a neighborhood of zero, that is, there exists an $\varepsilon > 0$ such that

$$\varphi(t) < \infty, \qquad -\varepsilon < t < \varepsilon,$$

then we call φ the *moment generating function* (MGF) of the random variable X. Otherwise we say X does not have an MGF.

Moment Generating Functions (cont.)

A theorem from advanced probability theory.

If a random variable X has an MGF φ, then

$$E(X^k) = \left. \frac{d^k \varphi(t)}{dt^k} \right|_{t=0}$$

Moment Generating Functions (cont.)

The idea of the proof is simple. If it is valid to differentiate the series

$$\varphi(t) = \sum_{s \in S} e^{tX(s)} f(s)$$

term by term (interchange the order of summation and differentiation), then we have

$$\varphi'(t) = \sum_{s \in S} X(s) e^{tX(s)} f(s)$$

$$\varphi''(t) = \sum_{s \in S} X(s)^2 e^{tX(s)} f(s)$$

and so forth.

Moment Generating Functions (cont.)

Hence

$$
\begin{aligned}
\varphi'(0) &= \sum_{s \in S} X(s) f(s) \\
\varphi''(0) &= \sum_{s \in S} X(s)^2 f(s)
\end{aligned}
$$

and so forth.

But we will not develop the tools for when it is valid to differentiate a series term by term in this course. You just have to take our word for it that this operation is always valid for differentiating an MGF at the point zero.

MGF of the Geometric Distribution

The MGF of a Geo(p) random variable is

$$\begin{aligned}
\varphi(t) &= \sum_{x=0}^{\infty} p(1-p)^x e^{tx} \\
&= \sum_{x=0}^{\infty} p \left[(1-p)e^t \right]^x \\
&= \frac{p}{1-(1-p)e^t}
\end{aligned}$$

by the geometric series theorem whenever

$$-1 < (1-p)e^t < 1$$

MGF of the Geometric Distribution (cont.)

Recall that the parameter space of the geometric family of distributions is $0 < p \leq 1$. It is important that $p = 0$ is not allowed.

In case $0 < p < 1$ we have

$$-1 < (1 - p)e^t < 1 \qquad (*)$$

whenever

$$t < \log\left(\frac{1}{1 - p}\right)$$

and the log of a number greater than one is greater than zero, so Geo(p) random variables have MGF for all such p.

In case $p = 1$, then we clearly have $(*)$ for all t, so Geo(1) random variables also have MGF.

Geometric Distribution (cont.)

If X has the Geo(p) distribution, then

$$\begin{aligned}
\varphi(t) &= \frac{p}{1-(1-p)e^t} \\
\varphi'(t) &= -\frac{p}{\left[1-(1-p)e^t\right]^2} \cdot \left[-(1-p)e^t\right] \\
&= \frac{p(1-p)e^t}{\left[1-(1-p)e^t\right]^2} \\
\varphi'(0) &= \frac{p(1-p)}{[1-(1-p)]^2} \\
&= \frac{1-p}{p}
\end{aligned}$$

Geometric Distribution (cont.)

If X has the Geo(p) distribution, then

$$
\begin{aligned}
\varphi'(t) &= \frac{p(1-p)e^t}{\left[1-(1-p)e^t\right]^2} \\
\varphi''(t) &= \frac{p(1-p)e^t}{\left[1-(1-p)e^t\right]^2} \\
&\qquad - 2\frac{p(1-p)e^t}{\left[1-(1-p)e^t\right]^3} \cdot \left[-(1-p)e^t\right] \\
&= \frac{p(1-p)e^t}{\left[1-(1-p)e^t\right]^2} + \frac{2p(1-p)^2e^{2t}}{\left[1-(1-p)e^t\right]^3}
\end{aligned}
$$

Geometric Distribution (cont.)

$$
\begin{aligned}
\varphi''(t) &= \frac{p(1-p)e^t}{\left[1-(1-p)e^t\right]^2} + \frac{2p(1-p)^2e^{2t}}{\left[1-(1-p)e^t\right]^3} \\
\varphi''(0) &= \frac{p(1-p)}{[1-(1-p)]^2} + \frac{2p(1-p)^2}{[1-(1-p)]^3} \\
&= \frac{p(1-p)\,[p+2(1-p)]}{p^3} \\
&= \frac{(1-p)(2-p)}{p^2}
\end{aligned}
$$

Geometric Distribution (cont.)

$$
\begin{aligned}
E(X) &= \frac{1-p}{p} \\
E(X^2) &= \frac{(1-p)(2-p)}{p^2} \\
\text{var}(X) &= \frac{(1-p)(2-p)}{p^2} - \left(\frac{1-p}{p}\right)^2 \\
&= \frac{(1-p)(2-p-1+p)}{p^2} \\
&= \frac{1-p}{p^2}
\end{aligned}
$$

Geometric Distribution (cont.)

What a struggle! But now we know.

If X has the Geo(p) distribution, then

$$
\begin{aligned}
E(X) &= \frac{1-p}{p} \\
\operatorname{var}(X) &= \frac{1-p}{p^2}
\end{aligned}
$$

Poisson Distribution

It's not about fish. It's named after a man named Poisson.

A random variable X has the *Poisson distribution* with parameter $\mu \geq 0$, abbreviated Poi(μ), if it has PMF

$$f_\mu(x) = \frac{\mu^x}{x!} e^{-\mu}, \qquad x = 0, 1, 2, \ldots.$$

Poisson Distribution (cont.)

As always, there is a theorem that the probabilities sum to one

$$\sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} = 1$$

which is equivalent to

$$\sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{\mu}$$

which is the Maclaurin series for the exponential function.

Poisson Distribution (cont.)

The Poisson distribution has an MGF, but we won't use it. We calculate the mean and variance using the theorem, just like we did for the binomial distribution.

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \cdot \frac{\mu^x}{x!} e^{-\mu} \\
&= \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} e^{-\mu} \\
&= \mu \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} e^{-\mu} \\
&= \mu \sum_{y=0}^{\infty} \frac{\mu^y}{y!} e^{-\mu} \\
&= \mu
\end{aligned}
$$

Poisson Distribution (cont.)

$$
\begin{aligned}
E\{X(X-1)\} &= \sum_{x=0}^{\infty} x(x-1) \cdot \frac{\mu^x}{x!} e^{-\mu} \\
&= \sum_{x=2}^{\infty} \frac{\mu^x}{(x-2)!} e^{-\mu} \\
&= \mu^2 \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!} e^{-\mu} \\
&= \mu^2 \sum_{y=0}^{\infty} \frac{\mu^y}{y!} e^{-\mu} \\
&= \mu^2
\end{aligned}
$$

Poisson Distribution (cont.)

$$
\begin{aligned}
\text{var}(X) &= E(X^2) - E(X)^2 \\
&= E\{X(X-1)\} + E(X) - E(X)^2 \\
&= \mu^2 + \mu - \mu^2 \\
&= \mu
\end{aligned}
$$

Poisson Distribution (cont.)

In summary, if X has the Poi(μ) distribution, then

$$\begin{aligned} E(X) &= \mu \\ \operatorname{var}(X) &= \mu \end{aligned}$$

Poisson Approximation to the Binomial Distribution

So far we have given no rationale for the Poisson distribution. What kind of random variable would have that?

It is an approximation to the Bin(n, p) distribution when p is very small, n is very large, and $np = \mu$ is moderate.

Poisson Approximation to the Binomial Distribution (cont.)

$$
\begin{aligned}
\binom{n}{x} p^x (1-p)^{n-x} &= \frac{n!}{x!\,(n-x)!} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} \\
&= \frac{\mu^x}{x!} \cdot \frac{n(n-1)\cdots(n-x+1)}{n^x} \left(1 - \frac{\mu}{n}\right)^{n-x} \\
&= \frac{\mu^x}{x!} \left[\prod_{k=0}^{x-1} \left(1 - \frac{k}{n}\right)\right] \left(1 - \frac{\mu}{n}\right)^{n-x}
\end{aligned}
$$

Now take the limit as $n \to \infty$. Clearly $1 - k/n \to 1$, so the term in square brackets converges to one. Hence, in order for this to converge to the PMF of the Poisson distribution, all we need is the validity of

$$
\lim_{n \to \infty} \left(1 - \frac{\mu}{n}\right)^{n-x} = e^{-\mu}
$$

Poisson Approximation to the Binomial Distribution (cont.)

To show the latter, take logs

$$\log\left(1 - \frac{\mu}{n}\right)^{n-x} = (n - x)\log\left(1 - \frac{\mu}{n}\right)$$

and use the definition of derivative

$$\lim_{h \to 0} \frac{\log(1 - h\mu) - \log(1)}{h} = \left.\frac{d \log(1 - \mu x)}{dx}\right|_{x=0} = -\mu$$

Hence

$$\begin{aligned}\lim_{n \to \infty} (n - x)\log\left(1 - \frac{\mu}{n}\right) &= \left[\lim_{n \to \infty} \frac{n - x}{n}\right]\left[\lim_{n \to \infty} n \log\left(1 - \frac{\mu}{n}\right)\right] \\ &= 1 \cdot (-\mu)\end{aligned}$$

using the theorem that the limit of a product is the product of the limits. Continuity of the exponential function finishes the proof.

Poisson Process

Imagine a bunch of IID Ber(p) random variables that represent presence or absence of a point in a region of space. Denote them X_t, $t \in T$, where the elements of t are the regions. We assume the elements of T are disjoint sets and each contains at most one point.

Let $\mathcal{A}$ denote the family of all unions of elements of T, including unions of just one element or no elements, and for each $A \in \mathcal{A}$, let X_A denote the number of points in A. This does not conflict with our earlier notation because each $t \in T$ is also an element of $\mathcal{A}$.

Let $n(A)$ denote the number of elements of T contained in A. Then X_A has the binomial distribution with sample size $n(A)$ and success probability p.

Poisson Process (cont.)

Now suppose p is very very small, so

$$E(X_A) = n(A)p$$

is also very very small unless $n(A)$ is very very large, in which case the distribution of X_A is approximately Poisson.

This gives rise to the following idea.

Poisson Process (cont.)

A random pattern of points in space is called a *spatial point process*, and such a process is called a *Poisson process* if the number of points X_A in region A has the following properties.

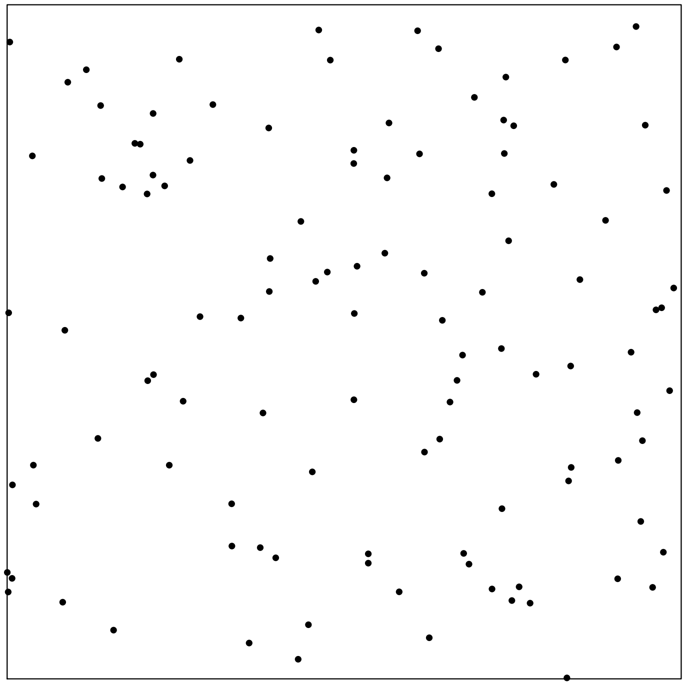
(i) If $A_1, \ldots, A_k$ are disjoint regions, then $X_{A_1}, \ldots, X_{A_k}$ are independent random variables.

(ii) For any region A, the random variable X_A has the Poisson distribution.

A Poisson process is *homogeneous* if $E(X_A)$ is proportional to the size of the region A.

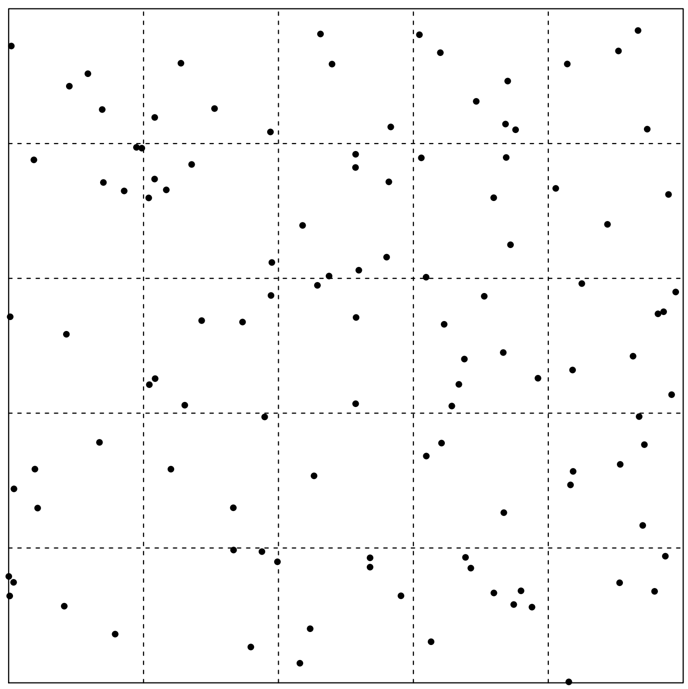
Poisson Process (cont.)

Here is an example.



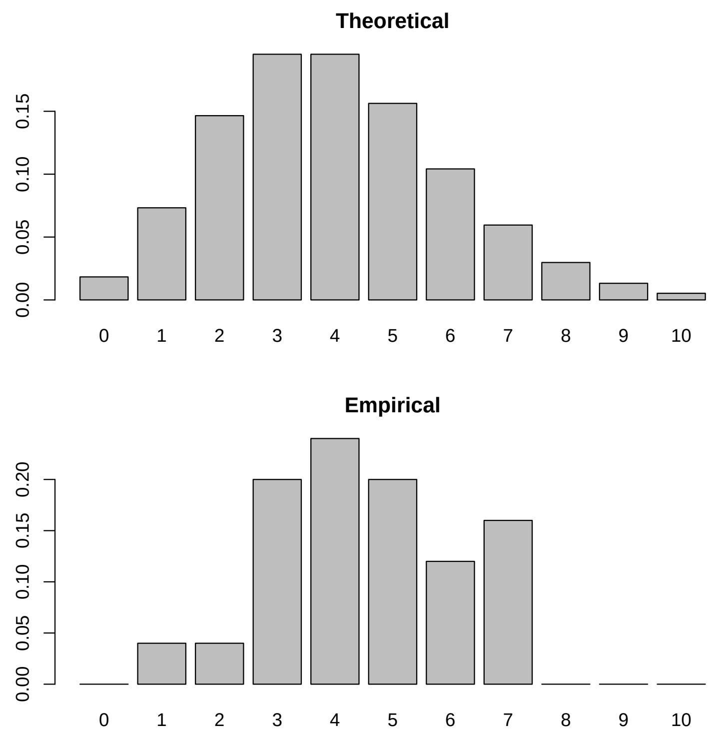
Poisson Process (cont.)

Suppose we divide the whole region into disjoint subregions and count the points in each.



Poisson Process (cont.)

Above, the PMF of the relevant Poisson distribution. Below, the "empirical" PMF, the histogram of counts in subregions.



Poisson Process (cont.)

The Poisson process is considered a reasonable model for any pattern of points in space, where space can be any dimension.

One dimension, the times of calls arriving at a call center, the times of radioactive decays.

Two dimensions, the pattern of anthills on a plain, or prairie dog holes, or trees in a forest.

Three dimensions, the pattern of raisins in a carrot cake.

Poisson Process (cont.)

What is the distribution of the number of raisins in a box of raisin bran?

Poisson (approximately) with parameter that is the mean number of raisins in a box.

The Addition Rule for Geometric

Suppose X_1, ..., X_n are IID Geo(p) random variables? What is the distribution of $Y = X_1 + \cdots + X_n$?

Each X_1 can be thought of as the number of zeros between ones in a Bernoulli process. Then Y is the number of zeros before the n-th one.

The probability of a particular pattern of zeros and ones that has n ones and y zeros is $p^n (1 - p)^y$.

The number of such patterns that end with a one is $\binom{n+y-1}{y}$.

The Negative Binomial Distribution

The *negative binomial distribution* with *shape parameter* n and *success probability* p has PMF

$$f_p(y) = \binom{n + y - 1}{y} p^n (1 - p)^y, \qquad y = 0, 1, 2, \ldots.$$

We abbreviate this distribution NegBin(n, p).

The Addition Rule for Geometric (cont.)

If X_1, ..., X_n are IID random variables having the Geo(p) distribution, then $Y = X_1 + \ldots + X_n$ has the NegBin(n, p) distribution.

The Addition Rule for Negative Binomial

If X_1, ..., X_n are independent (but not necessarily identically distributed) random variables, X_i having the NegBin(r_i, p) distribution, then $Y = X_1 + \ldots + X_n$ has the NegBin($r_1 + \cdots + r_n, p$) distribution.

Mean and Variance for Negative Binomial

If X has the NegBin(n, p) distribution, then

$$E(X) = n \cdot \frac{1-p}{p}$$

$$\text{var}(X) = n \cdot \frac{1-p}{p^2}$$

Convolution Formula

The rather odd name we will not try to explain. It gives the answer to the question: if X and Y are independent random variables with PMF f and g, respectively, then what is the PMF of $Z = X + Y$?

The PMF of the random vector (X, Y) is the product

$$h(x, y) = f(x)g(y)$$

by independence.

The map $(x, y) \mapsto (x, z)$ is invertible, hence one-to-one. Thus the PMF of the vector (X, Z) is

$$j(x, z) = f(x)g(z - x)$$

In order for this to make sense, we may have to define $g(y) = 0$ for values y not in the support of Y.

Convolution Formula (cont.)

To find the PMF of Z, we calculate

$$\Pr(Z = z) = \sum_x j(x, z) = \sum_x f(x)g(z - x)$$

where the sum runs over the support of x.

The Addition Rule for Poisson

If X and Y are independent Poisson random variables having means μ and ν, then what is the PMF of $Z = X + Y$?

$$
\begin{aligned}
h(z) &= \sum_x f(x) g(z - x) \\
&= \sum_{x=0}^{z} \frac{\mu^x}{x!} \cdot \frac{\nu^{z-x}}{(z-x)!} e^{-\mu-\nu}
\end{aligned}
$$

The sum stops at z because if $x > z$ then $y = z - x$ would be negative, which is impossible for a Poisson random variable.

The Addition Rule for Poisson (cont.)

$$
\begin{aligned}
h(z) &= \sum_{x=0}^{z} \frac{\mu^x}{x!} \cdot \frac{\nu^{z-x}}{(z-x)!} e^{-\mu-\nu} \\
&= \frac{(\mu+\nu)^z}{z!} e^{-\mu-\nu} \sum_{x=0}^{z} \frac{z!}{x!\,(z-x)!} \left(\frac{\mu}{\mu+\nu}\right)^x \left(\frac{\nu}{\mu+\nu}\right)^{z-x} \\
&= \frac{(\mu+\nu)^z}{z!} e^{-\mu-\nu} \sum_{x=0}^{z} \binom{z}{x} \left(\frac{\mu}{\mu+\nu}\right)^x \left(\frac{\nu}{\mu+\nu}\right)^{z-x} \\
&= \frac{(\mu+\nu)^z}{z!} e^{-\mu-\nu}
\end{aligned}
$$

which is the PMF of the Poi$(\mu+\nu)$ distribution.

The Addition Rule for Poisson (cont.)

If X_1, $\ldots$, X_n are independent (but not necessarily identically distributed) random variables, X_i having the Poi(μ_i) distribution, then $Y = X_1 + \ldots + X_n$ has the Poi($\mu_1 + \cdots + \mu_n$) distribution.

And now for something completely different ...

Defining Probabilities with Integrals

Integrals are limits of sums. It stands to reason that we can not only approximate probabilities with infinite sums but also with integrals.

Probability Density Functions

A real-valued function f defined on an interval (a, b) of the real numbers is called a *probability density function* (PDF) if

$$f(x) \geq 0, \qquad a < x < b$$

and

$$\int_a^b f(x)\,dx = 1.$$

The values $a = -\infty$ or $b = +\infty$ are allowed for endpoints of the interval.

A PDF is just like a PMF except that we integrate rather than sum.

Probability Density Functions (cont.)

A real-valued function f defined on a region S of $\mathbb{R}^2$ is also called a PDF if

$$f(x_1, x_2) \geq 0, \qquad (x_1, x_2) \in S$$

and

$$\iint_S f(x_1, x_2) \, dx_1 \, dx_2 = 1.$$

Probability Density Functions (cont.)

A real-valued function f defined on a region S of $\mathbb{R}^n$ is also called a PDF if

$$f(\mathbf{x}) \geq 0, \qquad \mathbf{x} \in S$$

and

$$\int_S f(\mathbf{x})\, d\mathbf{x} = 1.$$

Here only the boldface indicates that $\mathbf{x}$ is a vector and hence we are dealing with a multiple integral (n-dimensional).

Discrete and Continuous

If X is a random variable or $\mathbf{X}$ is a random vector whose distribution is described by a PMF, we say the distribution or the random variable or vector is *discrete*.

If X is a random variable or $\mathbf{X}$ is a random vector whose distribution is described by a PDF, we say the distribution or the random variable or vector is *continuous*.

Continuous Uniform Distribution

We say a continuous random variable or random vector is *uniform* if its PDF is a constant function. Different domains of definition give different random variables or random vectors.

In one dimension, the continuous uniform distribution on the interval (a, b) has the PDF

$$f(x) = \frac{1}{b - a}, \qquad a < x < b.$$

This distribution is abbreviated Unif(a, b).

That this constant is correct is obvious from an integral being the area under the "curve" (which in this case is flat). The area is that of a rectangle with base $b - a$ and height $1/(b - a)$.

Continuous Uniform Distribution (cont.)

In two dimensions, the continuous uniform distribution on the triangle

$$\{ (x, y) \in \mathbb{R}^2 : 0 < x < y < 1 \}$$

has the PDF

$$f(x, y) = 2, \qquad 0 < x < y < 1.$$

That this constant is correct is obvious from an integral being the volume under the "surface" (which in this case is flat). The volume is that of a parallelepiped having height 2 and triangular base having area $1/2$.

Exponential Distribution

The positive, continuous random variable having PDF

$$f_\lambda(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

is said to have the *exponential distribution* with *rate parameter* $\lambda > 0$. This is abbreviated $\mathrm{Exp}(\lambda)$.

Exponential Distribution (cont.)

Let us check that the PDF of the exponential distribution does integrate to one

$$
\begin{aligned}
\int_0^\infty \lambda e^{-\lambda x}\,dx &= -e^{-\lambda x}\Big|_0^\infty \\
&= \left[\lim_{x\to\infty}\left(-e^{-\lambda x}\right)\right] - \left[-e^{-\lambda 0}\right] \\
&= 0 - (-1) \\
&= 1
\end{aligned}
$$

Expectation

If $\mathbf{X}$ is a continuous random vector with PDF $f : S \to \mathbb{R}$, then

$$E\{g(\mathbf{X})\} = \int_S g(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}$$

if

$$\int_S |g(\mathbf{x})| f(\mathbf{x})\, d\mathbf{x} < \infty.$$

Otherwise, we say the expectation of $g(\mathbf{X})$ does not exist.

Again, this is just like the discrete case. In the discrete case, we are only interested in absolute summability. Here we are only interested in absolute integrability.

In both cases, $g(\mathbf{X})$ has expectation if and only if $|g(\mathbf{X})|$ has expectation.

Axioms for Expectation

The axioms we used before

$$E(X+Y) = E(X) + E(Y) \tag{1}$$
$$E(X) \geq 0, \qquad \text{when } X \geq 0 \tag{2}$$
$$E(aX) = aE(X) \tag{3}$$
$$E(1) = 1 \tag{4}$$

hold for expectation defined in terms of PMF and infinite sums or in terms of PDF and integrations just as they did for expectation defined in terms of PMF and finite sums, *when all of the expectations exist*.

Consequently, every property of expectation we derived from these axioms (all of deck 2) hold for these new kinds of expectation, just as they did for the old, again *when all of the expectations exist*.

Axioms for Expectation (cont.)

The proof that these axioms hold for expectation defined in terms of PDF, is very similar to homework problem 3-1. Just use

$$\int_S [g(\mathbf{x}) + h(\mathbf{x})] \, d\mathbf{x} = \int_S g(\mathbf{x}) \, d\mathbf{x} + \int_S h(\mathbf{x}) \, d\mathbf{x}$$

$$\int_S a g(\mathbf{x}) \, d\mathbf{x} = a \int_S g(\mathbf{x}) \, d\mathbf{x}$$

in place of the analogous properties of summation.

Continuous Uniform Distribution (cont.)

Suppose X has the Unif(a, b) distribution. Then

$$
\begin{aligned}
E(X) = \int_a^b x f(x)\, dx &= \frac{1}{b-a} \int_a^b x\, dx \\
&= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\
&= \frac{1}{b-a} \left[\frac{b^2}{2} - \frac{a^2}{2} \right] \\
&= \frac{(b^2 - a^2)}{2(b-a)} \\
&= \frac{(b-a)(b+a)}{2(b-a)} \\
&= \frac{b+a}{2}
\end{aligned}
$$

Continuous Uniform Distribution (cont.)

And

$$
\begin{aligned}
E(X^2) = \int_a^b x^2 f(x)\,dx &= \frac{1}{b-a}\int_a^b x^2\,dx \\
&= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\
&= \frac{1}{b-a}\left[\frac{b^3}{3} - \frac{a^3}{3}\right] \\
&= \frac{(b^3 - a^3)}{3(b-a)} \\
&= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} \\
&= \frac{b^2 + ab + a^2}{3}
\end{aligned}
$$

Continuous Uniform Distribution (cont.)

And

$$\begin{aligned}
\text{var}(X) &= E(X^2) - E(X)^2 \\
&= \frac{b^2 + ab + a^2}{3} - \left(\frac{b + a}{2}\right)^2 \\
&= \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} \\
&= \frac{4(b^2 + ab + a^2) - 3(b^2 + 2ab + a^2)}{12} \\
&= \frac{b^2 - 2ab + a^2}{12} \\
&= \frac{(b - a)^2}{12}
\end{aligned}$$

Continuous Uniform Distribution (cont.)

In summary, if X is a Unif(a, b) random variable, then

$$E(X) = \frac{a + b}{2}$$
$$\text{var}(X) = \frac{(b - a)^2}{12}$$

Continuous Distributions Approximate Discrete

Let X have the discrete uniform distribution on $\{1, \ldots n\}$, then the random variable $Y = X/n$ should be well approximated by U having the continuous uniform distribution on the interval $(0, 1)$ when n is large.

Compare mean and variance, for discrete

$$
\begin{aligned}
E(X) &= \frac{n+1}{2} \\
\operatorname{var}(X) &= \frac{(n+1)(n-1)}{12} \\
E(Y) &= \frac{n+1}{2n} = \frac{1}{2}\left(1 + \frac{1}{n}\right) \\
\operatorname{var}(Y) &= \frac{(n+1)(n-1)}{12n^2} = \frac{1}{12}\left(1 + \frac{1}{n}\right)\left(1 - \frac{1}{n}\right)
\end{aligned}
$$

Continuous Distributions Approximate Discrete (cont.)

$$
\begin{aligned}
E(Y) &= \frac{1}{2}\left(1 + \frac{1}{n}\right) \\
\operatorname{var}(Y) &= \frac{1}{12}\left(1 + \frac{1}{n}\right)\left(1 - \frac{1}{n}\right) \\
E(U) &= \frac{1}{2} \\
\operatorname{var}(U) &= \frac{1}{12}
\end{aligned}
$$

almost the same for large n.

Of course, this doesn't prove that Y and U have nearly the same distribution, since very different distributions can have the same mean and variance. More on this later.

Exponential Distribution (cont.)

If X has the exponential distribution with rate parameter λ, then

$$\begin{aligned} E(X) &= \int_0^\infty x f(x)\, dx \\ &= \int_0^\infty x \lambda e^{-\lambda x}\, dx \end{aligned}$$

We do this by integration by parts

$$\int u\, dv = uv - \int v\, du$$

with $u = x$ and $dv = \lambda e^{-\lambda x}\, dx$.

Exponential Distribution (cont.)

$$\begin{aligned}
E(X) &= \int_0^{\infty} x\lambda e^{-\lambda x}\,dx \\
&= -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx \\
&= \int_0^{\infty} e^{-\lambda x}\,dx \\
&= -\frac{1}{\lambda}e^{-\lambda x}\Big|_0^{\infty} \\
&= \frac{1}{\lambda}
\end{aligned}$$

The Gamma Function

Useful in calculating expectations with respect to the exponential distribution is a special function you may not have heard of but which is just as important as the logarithm, exponential, sine, or cosine functions.

The *gamma function* is defined for all positive real numbers α by

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

It is part of the definition that this integral exists for all $\alpha > 0$ (we won't verify that until we get to the unit on when infinite sums and integrals exist).

The Gamma Function (cont.)

We use the same integration by parts argument we used to calculate $E(X)$ for the exponential distribution with $u = x^\alpha$ and $dv = e^{-x}\, dx$.

$$\begin{aligned}
\Gamma(\alpha + 1) &= \int_0^\infty x^\alpha e^{-x}\, dx \\
&= -x^\alpha e^{-x} \Big|_0^\infty + \alpha \int_0^\infty x^{\alpha - 1} e^{-x}\, dx \\
&= \alpha \Gamma(\alpha)
\end{aligned}$$

This

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \qquad \alpha > 0$$

is the very important *gamma function recursion formula*.

The Gamma Function (cont.)

We know from the fact that the Exp(1) distribution has PDF that integrates to one

$$\int_0^\infty e^{-x}\,dx = 1$$

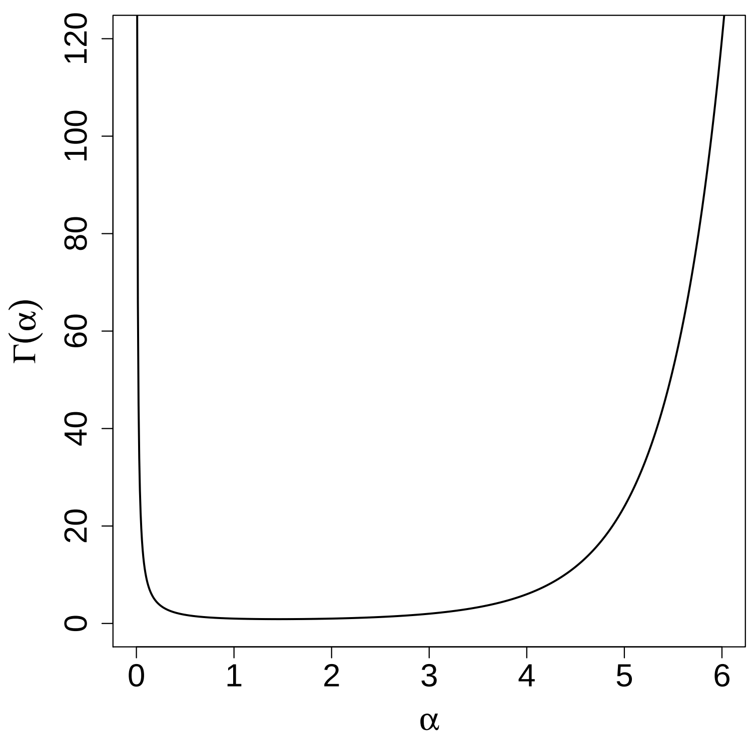
that $\Gamma(1) = 1$. Hence

$$\begin{aligned}
\Gamma(2) &= 1 \cdot \Gamma(1) = 1 \\
\Gamma(3) &= 2 \cdot \Gamma(2) = 2 \\
\Gamma(4) &= 3 \cdot \Gamma(3) = 3 \cdot 2 \\
\Gamma(5) &= 4 \cdot \Gamma(4) = 4 \cdot 3 \cdot 2 \\
&\vdots \\
\Gamma(n+1) &= n!
\end{aligned}$$

The gamma function "interpolates the factorials".

The Gamma Function (cont.)

The function $\alpha \mapsto \Gamma(\alpha)$ is a smooth function that goes to infinity as $\alpha \to 0$ and as $\alpha \to \infty$. Here is part of its graph.



Exponential Distribution (cont.)

Using the gamma function, we can find $E(X^\beta)$ for any $\beta > -1$ when X has the Exp(λ) distribution

$$
\begin{aligned}
E(X^\beta) &= \int_0^\infty x^\beta \cdot \lambda e^{-\lambda x} \, dx \\
&= \frac{1}{\lambda^\beta} \int_0^\infty y^\beta e^{-y} \, dy \\
&= \frac{\Gamma(\beta + 1)}{\lambda^\beta}
\end{aligned}
$$

Exponential Distribution (cont.)

As particular cases of

$$E(X^\beta) = \frac{\Gamma(\beta + 1)}{\lambda^\beta}$$

we have

$$\begin{aligned} E(X) &= \frac{\Gamma(2)}{\lambda} = \frac{1}{\lambda} \\ E(X^2) &= \frac{\Gamma(3)}{\lambda^2} = \frac{2}{\lambda^2} \end{aligned}$$

so

$$\begin{aligned} \mathrm{var}(X) &= E(X^2) - E(X)^2 \\ &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \end{aligned}$$

Exponential Distribution (cont.)

In summary, if X has the $\text{Exp}(\lambda)$ distribution, then

$$
\begin{aligned}
E(X) &= \frac{1}{\lambda} \\
\text{var}(X) &= \frac{1}{\lambda^2}
\end{aligned}
$$

Probabilities and PDF

As always, probability is just expectation of indicator functions.

If X is a continuous random variable with PDF f, then

$$\Pr(X \in A) = \int I_A(x) f(x)\, dx = \int_A f(x)\, dx$$

And similarly for random vectors (same equation but with boldface).

Probabilities and PDF (cont.)

Suppose X has the Exp(λ) distribution and $0 \le a < b < \infty$, then

$$\begin{aligned} \Pr(a \le X \le b) &= \int_a^b \lambda e^{-\lambda x}\, dx \\ &= -e^{-\lambda x}\Big|_a^b \\ &= e^{-\lambda a} - e^{-\lambda b} \end{aligned}$$

Probabilities and PDF (cont.)

Suppose (X, Y) has PDF

$$f(x, y) = x + y, \qquad 0 < x < 1,\ 0 < y < 1$$

and $0 < a < 1$.

Probabilities and PDF (cont.)

Then

$$
\begin{aligned}
\Pr(X \le a) &= \int_0^a \int_0^1 (x+y)\,dy\,dx \\
&= \int_0^a dx \left[xy + \frac{y^2}{2} \right]_0^1 \\
&= \int_0^a \left(x + \frac{1}{2} \right) dx \\
&= \left. \frac{x^2}{2} + \frac{x}{2} \right|_0^a \\
&= \frac{a^2 + a}{2}
\end{aligned}
$$

Neither Discrete Nor Continuous

It is easy to think of random variables and random vectors that are neither discrete nor continuous.

Detection Limit Model

Here X models a measurement, which is a real number (say weight), but there is a detection limit ε, which is the lowest value the measurement device can read. For values above ε the distribution is continuous. For the value ε, the distribution is discrete. We can write

$$E\{g(X)\} = pg(\varepsilon) + (1 - p) \int_{\varepsilon}^{\infty} g(x) f(x) \, dx$$

where $p = \Pr(X = \varepsilon)$ and f is a PDF giving the part of the distribution when $X > \varepsilon$.

Neither Discrete Nor Continuous (cont.)

Some Components Discrete and Some Continuous

If X and Y are independent random vectors, X is Geo(p) and Y is Exp(λ), then the random vector (X, Y) is neither discrete nor continuous. We can write

$$E\{g(X,Y)\} = \sum_{x=0}^{\infty} \int_0^{\infty} g(x,y) p(1-p)^x \lambda e^{-\lambda y} \, dy$$

There is no problem with expectations, we integrate over the continuous variable and sum over the discrete one. We could also define a model where the components are not independent and one is discrete and the other continuous.

Neither Discrete Nor Continuous (cont.)

Degenerate Random Vectors

Suppose X has the Unif$(0, 1)$ distribution. Then the random vector $\mathbf{Y} = (X, X)$ does not have a PDF. Nor does it have a PMF.

We sometimes say it has a *degenerate* continuous distribution. Although it is a two-dimensional random vector, it is really one-dimensional, since it is a function of the one-dimensional variable X.

We can write

$$E\{g(Y_1, Y_2)\} = E\{g(X, X)\} = \int_0^1 g(x, x)\, dx$$

Neither Discrete Nor Continuous (cont.)

We can handle some models that are neither discrete nor continuous, but we won't discuss them much, nor provide general methods for handling them, except for the next method.

Distribution Functions

Our last method of specifying a probability model!

The *distribution function* (DF) of a random variable X is the function $\mathbb{R} \to \mathbb{R}$ defined by

$$F(x) = \Pr(X \leq x), \qquad x \in \mathbb{R}$$

Note that the domain is always the *whole real line* no matter what the support of X may be.

Also called *cumulative distribution function* (CDF), but not in theory courses.

Distribution Functions (cont.)

If X is Exp(λ), we have calculated

$$\Pr(a \leq X \leq b) = e^{-\lambda a} - e^{-\lambda b}, \qquad 0 \leq a < b < \infty$$

We also know $\Pr(X \leq a) = 0$ for negative a because X is a nonnegative random variable. Thus X has DF

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

Distribution Functions (cont.)

We can generalize the argument about the support. If X has support $[a, b]$, then we know the DF has the form

$$F(x) = \begin{cases} 0, & x < a \\ \text{something}, & a \le x < b \\ 1, & x \ge b \end{cases}$$

Distribution Functions (cont.)

If X has the $\mathsf{Unif}(a, b)$ distribution, then for $a \le x < b$ we have

$$F(x) = \Pr(X \le x) = \int_a^x \frac{1}{b-a}\,ds = \frac{s}{b-a}\bigg|_a^x = \frac{x-a}{b-a}$$

so

$$F(x) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a \le x < b \\ 1, & x \ge b \end{cases}$$

PDF are Different

So far PDF are much the same as PMF. You just integrate instead of sum. But something is a bit strange about PDF.

If X has the Unif$(0, 1)$ distribution, what are $\Pr(X \leq 1/2)$ and $\Pr(X < 1/2)$?

Same integral

$$\int_0^{1/2} dx$$

for both! Hence

$$\Pr(X = 1/2) = \Pr(X \leq 1/2) - \Pr(X < 1/2) = 0$$

because $X < 1/2$ and $X = 1/2$ are mutually exclusive events.

PDF are Different (cont.)

Generalizing this argument. For any continuous random variable X and any constant a we have $\Pr(X = a) = 0$.

This seems paradoxical. If every point in the sample space has probability zero, where is the probability?

It also seems weird. But it is a price we pay for the simplicity of calculation that comes with continuous random variables (integration is easier than summation).

Continuous random variables don't really exist, because no random phenomenon is measured or recorded to an infinite number of decimal places. Nor, since the universe is really discrete (atoms, quanta, etc.) would it make sense to do so even if we could.

PDF are Different (cont.)

Continuous random variables are an idealization. They approximate discrete random variables with a very large support having very small spacing — measured to a large, but not infinite number, of decimal places.

For example, the discrete model having the uniform distribution on the set

$$\left\{ \frac{1}{n}, \frac{2}{n}, \ldots, 1 \right\}$$

is well approximated by the Unif$(0, 1)$ distribution when n is large.

In a discrete model well approximated by a continuous one, the probability of any point is very small. In the continuous approximation, the probability of any point is zero. Not so weird when thought about this way.

PDF are Different (cont.)

Because points have probability zero, a PDF can be arbitrarily redefined at any point, or any finite set of points, without changing probabilities or expectations. Suppose we wish to define the Unif(a, b) distribution on the whole real line rather than just on the interval (a, b). How do we define the PDF at a and b?

It doesn't matter. We can define

$$f(x) = \begin{cases} 1/(b-a), & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

or

$$f(x) = \begin{cases} 1/(b-a), & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

PDF are Different (cont.)

or

$$f(x) = \begin{cases} 1/(b-a), & a < x < b \\ 42, & x = a \text{ or } x = b \\ 0, & \text{otherwise} \end{cases}$$

Probabilities and expectations are not affected by these changes.

PDF are Different (cont.)

Because points have probability zero, there is no difference between

$$
\begin{aligned}
&\Pr(a < X < b) \\
&\Pr(a < X \le b) \\
&\Pr(a \le X < b) \\
&\Pr(a \le X \le b)
\end{aligned}
$$

when X is continuous.

(When X is discrete, there can be a big difference!)

PDF are Different (cont.)

The situation is worse for continuous random vectors. What is the PDF of the continuous uniform distribution on the square $(0,1)^2$?

If we want to define the PDF on all of $\mathbb{R}^2$, then it doesn't matter how we define the PDF on the boundary of the support, or on any one-dimensional line or curve.

$$f(x,y) = \begin{cases} 1, & 0 < x < 1 \text{ and } 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

or

$$f(x,y) = \begin{cases} 1, & 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

PDF are Different (cont.)

or

$$f(x,y) = \begin{cases} 1, & 0 < x < 1 \text{ and } 0 < y < 1 \text{ and } x \neq y \\ 0, & \text{otherwise} \end{cases}$$

all define the same probabilities and expectations, so we say they define the same probability distributions.

DF and PDF

We already know how to go from PDF f to DF F

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(s)\,ds$$

Note that we either have to change the dummy variable of integration from x to s (or anything other than x) or we have to change the free variable

$$F(s) = \Pr(X \leq s) = \int_{-\infty}^{s} f(x)\,dx$$

Both define exactly the same function (mathematics is invariant under changes of notation).

DF and PDF (cont.)

We go the other way using the fundamental theorem of calculus.

$$f(x) = F'(x), \qquad \text{when } F \text{ is differentiable at } x.$$

Typically, F is differentiable except at a finite set of points, and *it doesn't matter* how f is defined at those points.

DF and PDF (cont.)

For the Exp(λ) distribution we found

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

If we apply $f = F'$ to this, we obtain

$$f(x) = \begin{cases} 0, & x < 0 \\ \text{whatever}, & x = 0, \\ \lambda e^{-\lambda x}, & x > 0 \end{cases}$$

which is a valid PDF for this distribution.

DF and PDF (cont.)

For the Unif(a, b) distribution we found

$$F(x) = \begin{cases} 0, & x < a \\ (x - a)/(b - a), & a \le x < b \\ 1, & x \ge b \end{cases}$$

If we apply $f = F'$ to this, we obtain

$$f(x) = \begin{cases} 0, & x < a \\ \text{whatever}, & x = a, \\ 1/(b - a), & a < x < b \\ \text{whatever}, & x = b, \\ 0, & x > b \end{cases}$$

which is a valid PDF for this distribution.

DF (cont.)

DF exist for any random variable, and are defined by the same general formula

$$F(x) = \mathrm{Pr}(X \leq x), \qquad x \in \mathbb{R}$$

DF and PMF

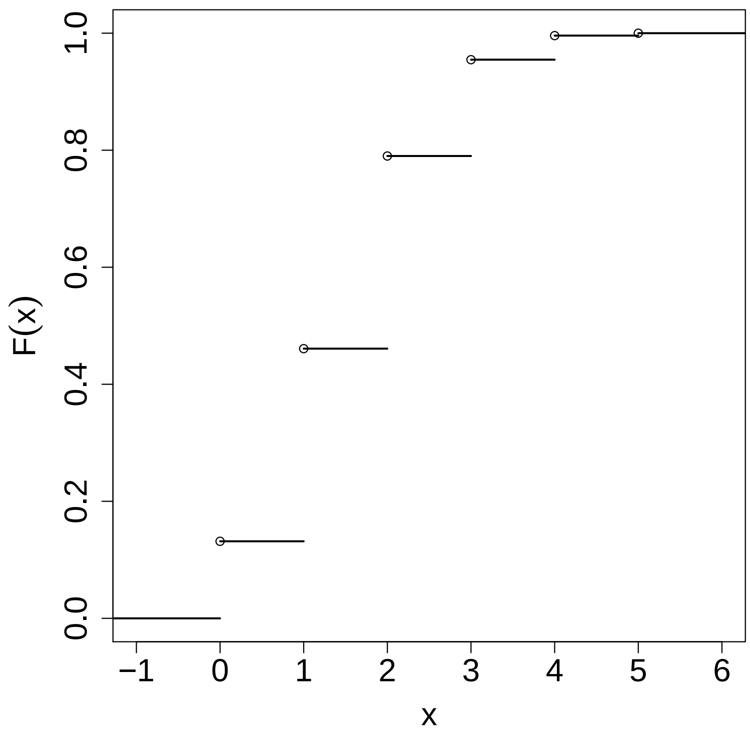
The DF of a discrete random variable X having PMF $f : S \to \mathbb{R}$ is

$$
\begin{aligned}
F(x) &= \Pr(X \le x) \\
&= \sum_{\substack{s \in S \\ s \le x}} f(s)
\end{aligned}
$$

The DF is flat between elements of the support of X, and has a jump $f(x)$ at each point x in the support.

DF and PMF (cont.)

Here is an example, the DF of the $\text{Bin}(5, 1/3)$ distribution



DF for Neither Continuous Nor Discrete

Recall the expectation operator for the detection limit model

$$E\{g(X)\} = pg(\varepsilon) + (1 - p) \int_{\varepsilon}^{\infty} g(x) f(x) \, dx$$

We have

$$\begin{aligned}
F(x) &= \Pr(X \le x) = 0, & x < \varepsilon \\
F(x) &= \Pr(X \le x) = p, & x = \varepsilon \\
F(x) &= \Pr(X \le x) = p + \int_{\varepsilon}^{x} f(s) \, ds, & x > \varepsilon
\end{aligned}$$

Properties of DF

Since $F(x)$ is a probability $\Pr(X \leq x)$ we have

$$0 \leq F(x) \leq 1, \qquad \text{for all } x.$$

By monotonicity of probability, F is nondecreasing

$$x_1 \leq x_2 \quad \text{implies} \quad F(x_1) \leq F(x_2)$$

If X is a continuous random variable, then F is a continuous function. If X is a discrete random variable, then F is a step function, with jumps at the elements of the support of X.

$$\begin{aligned} F(x) &= 0, \qquad && x \text{ below the support of } X \\ F(x) &= 1, \qquad && x \text{ above the support of } X \end{aligned}$$

Properties of DF (cont.)

A property we cannot yet prove in general, but is easily seen to be true for DF of either discrete or continuous is that DF are right continuous

$$F(x) = \Pr(X \le x) = \lim_{y \downarrow x} F(y)$$

and have left limits

$$F_{-}(x) = \Pr(X < x) = \lim_{y \uparrow x} F(y)$$

and

$$\lim_{y \downarrow -\infty} F(y) = 0$$
$$\lim_{y \uparrow +\infty} F(y) = 1$$

Properties of DF (cont.)

As far as applications are concerned, the continuity and limit properties of DF are useful only as "sanity checks".

The only effect in practice is that when you make a plot of a discontinuous DF, you should indicate that it is right continuous as we did in our plot.

Formulas will automatically indicate right continuity (if they are correct!)

PDF and Independence

Verifying independence from PDF almost the same as with PMF.

- First we check that the support (the set where the PDF is nonzero) is a Cartesian product.

- Second we check that the PDF is a product of functions of each variable

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} h_i(x_i)$$

We have independence if these two checks hold.

PDF and Independence (cont.)

The only difference between the check with PDF and the check with PMF is that the conditions were "if and only if" with PMF and now are only "if" with PDF.

The issue is that PDF can be redefined on sets of probability zero without changing the distribution and such redefinition can make the check fail but does not change the distribution (so the random variables may still be still independent).

If the random variables are independent, then there is some definition of the PDF such that the check works, but it may not be the given definition.

PDF and Independence (cont.)

For the examples on slide 102 the check works. For the example on slide 103 the check does not work. But these are all PDF's for the same distribution.

Usually this is not an issue to worry about. For most nontricky definitions of PDF the check will work if the random variables are independent.

Change of Variable

The term *change of variable* refers to the process of determining the distribution of $Y = g(X)$ when the distribution of X is given.

We already know how to do this for discrete random variables and random vectors (slides 81–89, deck 1). Now we do for continuous random variables and random vectors.

Invertible Functions

A function $g : S \to T$ is *invertible* if there is a function $h : T \to S$ such that

$$\begin{aligned} h(g(x)) &= x, & \qquad x \in S \\ g(h(y)) &= y, & \qquad y \in T \end{aligned}$$

The domain of g is the codomain of h and vice versa.

h is said to be the *inverse* of g and vice versa. This relationship is sometimes denoted $h = g^{-1}$.

Invertible Functions (cont.)

One-dimensional continuous and invertible functions are strictly monotone (strictly increasing or strictly decreasing). Examples of inverse pairs

$$\begin{aligned} \exp &: \mathbb{R} \to (0, \infty) \\ \log &: (0, \infty) \to \mathbb{R} \end{aligned}$$

and

$$\begin{aligned} x \mapsto x^2 &\qquad (0, \infty) \to (0, \infty) \\ y \mapsto \sqrt{y} &\qquad (0, \infty) \to (0, \infty) \end{aligned}$$

Note that $x \mapsto x^2$ considered as a function $\mathbb{R} \to [0, \infty]$ is not invertible because the equation $y = x^2$ has two solutions for x.

Change of Variable (cont.)

The change-of-variable process very different for PDF and PMF. Suppose $g : S \to T$ is an invertible function, where S and T are open subsets of $\mathbb{R}^n$ and h is the inverse of g. Then for any function w

$$\int_S w(\mathbf{x})\, d\mathbf{x} = \int_T w[h(\mathbf{y})] |\det J(\mathbf{y})|\, d\mathbf{y}$$

(if the integrals exist and h is differentiable). This is the multivariate change-of-variable formula for integration.

Here $J(\mathbf{y})$ is the Jacobian matrix for the change-of-variable h, the $n \times n$ matrix whose i, j component is $\partial h_i(\mathbf{y}) / \partial y_j$, sometimes written $\partial x_i / \partial y_j$.

Change of Variable (cont.)

If we apply this to integrals that are expectations, we get

$$E\{r(\mathbf{Y})\} = \int_S r[g(\mathbf{x})] f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = \int_T r(\mathbf{y}) f_{\mathbf{X}}[h(\mathbf{y})] |\det J(\mathbf{y})|\, d\mathbf{y}$$

for any function r so long as the integrals exist, where f_X is the PDF of $\mathbf{X}$.

From this we see that

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}[h(\mathbf{y})] \cdot |\det J(\mathbf{y})|, \qquad \mathbf{y} \in T$$

serves as a PDF of $\mathbf{Y}$. This is the multivariate change-of-variable theorem for PDF.

Change of Variable (cont.)

Before we get to examples we specialize to the univariate case. Here

$$f_Y(y) = f_X[h(y)] \cdot |h'(y)|, \qquad y \in T$$

is the PDF of the random variable $Y = g(X)$, where h is the inverse of g and is differentiable.

Change of Variable (cont.)

Suppose X has the Exp(λ) distribution and $Y = 1/X$. What is the PDF of Y?

Here

$$
\begin{aligned}
g(x) &= 1/x \\
h(y) &= 1/y \\
h'(y) &= -1/y^2 \\
f_X(x) &= \lambda e^{-\lambda x}
\end{aligned}
$$

so

$$
\begin{aligned}
f_Y(y) &= f_X[h(y)] |h'(y)| \\
&= \lambda e^{-\lambda(1/y)} |-1/y^2| \\
&= \lambda e^{-\lambda/y} / y^2
\end{aligned}
$$

Change of Variable (cont.)

The only thing left to do is add the domain of definition. The map $x \mapsto 1/x$ maps the domain $(0, \infty)$ of X to $(0, \infty)$. Hence the PDF of Y is

$$f(y) = \frac{\lambda e^{-\lambda/y}}{y^2}, \qquad y > 0$$

Here we have dropped the subscript on f_Y now that it is no longer needed to avoid confusion.

The Gamma Distribution

The function

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \qquad x > 0$$

is the PDF of a random variable when $\alpha > 0$ and $\lambda > 0$. This distribution is called the *gamma distribution* with shape parameter α and rate parameter λ. It is abbreviated Gam(α, λ).

That this density integrates to one, is shown by the substitution $y = \lambda x$

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x} \, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\lambda} \right)^{\alpha-1} e^{-y} \frac{dy}{\lambda} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{\lambda^\alpha} = 1$$

The Gamma Distribution (cont.)

The special case where $\alpha = 1$ is the exponential distribution.

$\text{Gam}(1, \lambda) = \text{Exp}(\lambda)$.

Change of Variable (cont.)

Theorem. Suppose X and Y are independent gamma random variables

$$
\begin{aligned}
X &\sim \text{Gam}(\alpha_1, \lambda) \\
Y &\sim \text{Gam}(\alpha_2, \lambda)
\end{aligned}
$$

then

$$
\begin{aligned}
U &= X + Y \\
V &= \frac{X}{X + Y}
\end{aligned}
$$

are independent random variables and

$$
U \sim \text{Gam}(\alpha_1 + \alpha_2, \lambda)
$$

Change of Variable (cont.)

Two things are important about this theorem.

First, it contains the addition rule for gamma random variables. If X_1, $\ldots$, X_n are gamma random variables all with the same rate parameter, X_i has the $\text{Gam}(\alpha_i, \lambda)$ distribution, then $X_1 + \cdots + X_n$ has the $\text{Gam}(\alpha_1 + \cdots + \alpha_n, \lambda)$ distribution. (Apply mathematical induction to the theorem.)

Second, the distribution of V is also a brand name distribution, but we haven't named it yet. (We will get to it.)

Change of Variable (cont.)

The first step in applying the change-of-variable theorem is to find the inverse transformation.

$$
\begin{aligned}
u &= x + y \\
v &= \frac{x}{x + y}
\end{aligned}
$$

implies $v = x/u$ so $x = uv$. Then $y = u - x = u - uv = u(1 - v)$. In summary

$$
\begin{aligned}
x &= uv \\
y &= u(1 - v)
\end{aligned}
$$

Change of Variable (cont.)

For the transformation

$$
\begin{aligned}
x &= uv \\
y &= u(1 - v)
\end{aligned}
$$

the Jacobian matrix is

$$
\begin{aligned}
J(u, v) &= \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \\
&= \begin{pmatrix} v & u \\ 1 - v & -u \end{pmatrix}
\end{aligned}
$$

Change of Variable (cont.)

And the Jacobian determinant is

$$\begin{aligned}
\det J(u, v) &= \begin{vmatrix} v & u \\ 1 - v & -u \end{vmatrix} \\
&= -uv - u(1 - v) \\
&= -u
\end{aligned}$$

Change of Variable (cont.)

By independence the PDF of the random vector (X, Y) is the product $f_X(x) f_Y(y)$, hence the PDF of (U, V) is

$$
\begin{aligned}
f_{U,V}(u, v) &= f_X(uv) f_Y[u(1-v)] |\det J(u, v)| \\
&= \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} (uv)^{\alpha_1 - 1} e^{-\lambda uv} \cdot \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} [u(1-v)]^{\alpha_2 - 1} e^{-\lambda u(1-v)} \cdot u \\
&= \frac{\lambda^{\alpha_1} \lambda^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \cdot u^{\alpha_1 + \alpha_2 - 1} e^{-\lambda u} \cdot v^{\alpha_1 - 1} (1-v)^{\alpha_2 - 1}
\end{aligned}
$$

Since this is a function of u times a function of v, the random variables U and V are independent if the support is a Cartesian product.

Change of Variable (cont.)

$u = x + y$ satisfies $0 < u < \infty$ and $v = x/(x + y)$ satisfies $0 < v < 1$.

Conversely, for any (u, v) in the set

$$\{ (u, v) \in \mathbb{R}^2 : 0 < u < \infty \text{ and } 0 < v < 1 \} \tag{$**$}$$

we have $x = uv$ satisfies $0 < x < \infty$ and and $y = u(1 - v)$ satisfies $0 < y < \infty$.

Hence $(**)$ is the support of the random vector (U, V). Since $(**)$ is a Cartesian product, we have finished checking (satisfactorily) that U and V are independent.

Change of Variable (cont.)

So what is the distribution of $U = X + Y$? We know

$$\begin{aligned} f_{U,V}(u, v) &= f_U(u) f_V(v) \\ &= \frac{\lambda^{\alpha_1} \lambda^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \cdot u^{\alpha_1+\alpha_2-1} e^{-\lambda u} \cdot v^{\alpha_1-1}(1-v)^{\alpha_2-1} \end{aligned}$$

Clearly

$$f_U(u) \propto u^{\alpha_1+\alpha_2-1} e^{-\lambda u}$$

where $\propto$ means "proportional to". Since the $\text{Gam}(\alpha_1 + \alpha_2, \lambda)$ distribution has a PDF of this form, that is the distribution of U

$$f_U(u) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} u^{\alpha_1+\alpha_2-1} e^{-\lambda u}, \qquad u > 0$$

Change of Variable (cont.)

Plugging

$$f_U(u) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} u^{\alpha_1+\alpha_2-1} e^{-\lambda u}, \qquad u > 0$$

in to

$$f_U(u) f_V(v) = \frac{\lambda^{\alpha_1}\lambda^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \cdot u^{\alpha_1+\alpha_2-1} e^{-\lambda u} \cdot v^{\alpha_1-1}(1-v)^{\alpha_2-1}$$

we get

$$f_V(v) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1-1}(1-v)^{\alpha_2-1}, \qquad 0 < v < 1$$

Beta Distribution

Conclusion:

$$f(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1}, \qquad 0 < x < 1$$

is a PDF. This distribution has two parameters α_1 and α_2, which can be any positive numbers. It is abbreviated Beta(α_1, α_2).

The reason for the name is that

$$B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} = \int_0^1 x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1}\, dx$$

is called the *beta function*. For us, the definition of the beta function is just the fact that the PDF of the beta distribution integrates to one (which we derived using the change-of-variable theorem).

Gamma Distribution (cont.)

If X has the $\text{Gam}(\alpha, \lambda)$ distribution, then

$$\begin{aligned} E(X^\beta) &= \int_0^\infty x^\beta f(x)\, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\beta \cdot x^{\alpha - 1} e^{-\lambda x}\, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha + \beta)}{\lambda^{\alpha + \beta}} \end{aligned}$$

because the integrand is, except for constants, the PDF of the $\text{Gam}(\alpha + \beta, \lambda)$ distribution. This also tells us that the integral exists if and only if $\alpha + \beta > 0$. Hence the formula above is valid for negative β, so long as $\beta > -\alpha$.

Gamma Distribution (cont.)

If β is a positive integer, then gamma functions can be eliminated from

$$E(X^\beta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha + \beta)}{\lambda^{\alpha+\beta}}$$

using the gamma function recursion formula. For example,

$$E(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha + 1)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda}$$

Linear Change of Variable

Suppose X is a continuous random variable having PDF f_X which is defined on the whole real line. Then $Y = \mu + \sigma X$ has PDF

$$f_Y(y) = \frac{1}{|\sigma|} f_X\left(\frac{y - \mu}{\sigma}\right)$$

if $\sigma \neq 0$ (otherwise Y is the constant random variable always having the value μ and does not have a PDF).

Proof: Solving $y = \mu + \sigma x$ for x we obtain

$$h(y) = \frac{y - \mu}{\sigma}$$

for the inverse transformation and this has "Jacobian" $\frac{1}{\sigma}$. Now apply the change-of-variable formula.

Location-Scale Families

Clearly as x goes from $-\infty$ to ∞ so does y, and vice versa. Hence the range of Y is the whole real line.

The parametric family of distributions having PDF of the form

$$f_{\mu,\sigma}(y) = \frac{1}{\sigma} f\left(\frac{y - \mu}{\sigma}\right)$$

where μ and σ are parameters, μ called the *location parameter* and σ called the *scale parameter*, and where $\sigma > 0$ and μ can be any real number, is called the *location-scale* family with *standard distribution* having PDF $f = f_{0,1}$, which must be a PDF defined on the whole real line.

Location-Scale Families (cont.)

The location-scale family with standard PDF $f = f_{0,1}$ is the set of all distributions of random variables $Y = \mu + \sigma X$, where X has PDF f. We know

$$\begin{aligned} E(Y) &= \mu + \sigma E(X) \\ \operatorname{var}(Y) &= \sigma^2 \operatorname{var}(X) \\ \operatorname{sd}(Y) &= \sigma \operatorname{sd}(X) \end{aligned}$$

Hence, *if we choose* f *so that* $E(X) = 0$ *and* $\operatorname{var}(X) = 1$, then μ is the mean and σ is the standard deviation of Y.

Otherwise, μ and σ cannot be the mean and standard deviation.

Location-Scale Families (cont.)

The only location-scale family we already know is the Unif(a, b) family. However, a and b are not a location-scale pair of parameters. We can take a to be the location parameter and $\sigma = b - a$ to be the scale parameter. Then the *standard continuous uniform distribution* is the one with $a = 0$ and $b - a = 1$, so $b = 1$, that is, the Unif$(0, 1)$ distribution is the standard one.

Then

$$
\begin{aligned}
f_{a,\sigma}(x) &= \frac{1}{\sigma} \cdot f_{0,1}\left(\frac{x - a}{\sigma}\right) \\
&= \frac{1}{b - a} \cdot I_{(0,1)}\left(\frac{x - a}{b - a}\right) \\
&= \frac{1}{b - a} \cdot I_{(a,b)}(x)
\end{aligned}
$$

Standard Normal Distribution

The function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty$$

is a PDF.

Standard Normal Distribution (cont.)

Once when lecturing in class He [Lord Kelvin] used the word 'mathematician' and then interrupting himself asked his class: 'Do you know what a mathematician is?' Stepping to his blackboard he wrote upon it:

$$\int_{-\infty}^{+\infty} e^{-x^2}\, dx = \sqrt{\pi}.$$

Then putting his finger on what he had written, he turned to his class and said, 'a mathematician is one to whom that is as obvious as that twice two makes four is to you.'

S. P. Thompson, *Life of Lord Kelvin*.

Standard Normal Distribution (cont.)

Proof that the standard normal PDF integrates to one. Let

$$c = \int_{-\infty}^{\infty} e^{-x^2/2} \, dx,$$

then

$$\begin{aligned} c^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2/2 - y^2/2} \, dx \, dy \\ &= \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} \, r dr \, d\theta \\ &= 2\pi \int_0^{\infty} e^{-r^2/2} \, r dr \\ &= 2\pi \Big[-e^{-r^2/2} \Big]_0^{\infty} \\ &= 2\pi \end{aligned}$$

General Normal Distributions

The location-scale family whose standard PDF is the standard normal PDF, is called the family of *normal distributions*. The normal distribution with location parameter μ and scale parameter σ is abbreviated $\mathcal{N}(\mu, \sigma^2)$. It has PDF

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

Note the oddity that in writing $\mathcal{N}(\mu, \sigma^2)$ it is not the scale parameter but its square that goes in the second slot.

Symmetry

We say a random variable X is *symmetric about zero* if $-X$ has the same distribution as X.

We say a random variable X is *symmetric about the point* a if $X - a$ is symmetric about zero, that is, if $-(X - a)$ has the same distribution as $X - a$.

In this case we say a is the *center of symmetry* of the distribution of X.

Symmetry (cont.)

If the distribution of X is specified by a PMF or PDF $f : S \to \mathbb{R}$, then the distribution is symmetric about zero if

$$f(-x) = f(x), \qquad x \in S$$

(this implicitly requires that $-x \in S$ whenever $x \in S$).

For PDF this check can fail due to arbitrary redefinition at a finite set of points. One must have a "nice" definition of the PDF.

One should not think of this as the *definition* of symmetry about zero. The real definition, that X and $-X$ have the same distribution is (1) not dependent on how the PDF is defined and (2) much simpler to use.

Symmetry (cont.)

We don't yet know many symmetric distributions.

The discrete uniform distribution on $\{1, \ldots, n\}$ is symmetric about $(n + 1)/2$.

The Bin(n, p) distribution is symmetric about $n/2$ if $p = 1/2$.

The Unif(a, b) distribution is symmetric about $(a + b)/2$.

The $\mathcal{N}(\mu, \sigma^2)$ distribution is symmetric about μ.

The Beta(α_1, α_2) distribution is symmetric about $1/2$ if $\alpha_1 = \alpha_2$.

Moments

For any random variable X, the numbers

$$\alpha_k = E(X^k), \qquad k = 1, 2, \ldots$$

are called the *ordinary moments* of X and α_k is called the k-th ordinary moment or the ordinary moment of order k.

Of course, the moments need not all exist. This is what they are called if they exist. If X has a moment generating function, then moments of all orders exist.

The first ordinary moment is also called the expectation of X or the mean of X (we already knew that).

The "ordinary" in "ordinary moment" is our private terminology. Most probabilists would just say "moment".

Central Moments

For any random variable X with $E(X) = \mu$, the numbers

$$\mu_k = E\{(X - \mu)^k\}, \qquad k = 1, 2, \ldots$$

are called the *central moments* of X and μ_k is called the k-th central moment or the central moment of order k.

Of course, the moments need not all exist. This is what they are called if they exist.

The first central moment is necessarily zero if it exists, because $E(X - \mu) = E(X) - \mu = 0$.

The second central moment is also called the variance of X (we already knew that).

Ordinary and Central Moments

$$
\begin{aligned}
\mu_k &= E\{(X-\mu)^k\} \\
&= E\left\{\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} X^j \mu^{k-j}\right\} \\
&= \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \mu^{k-j} E(X^j) \\
&= \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \alpha_j \mu^{k-j}
\end{aligned}
$$

where $\alpha_0 = E(X^0) = E(1) = 1$ and $\alpha_1 = \mu$.

Of course, this only holds if the moments in the formula exist.

Change of Variable, Method 2

We can also use DF's to calculate change-of-variable. If X has PDF f_X and $Y = g(X)$, then the DF of Y is

$$F(y) = \Pr(Y \le y) = \Pr\{g(X) \le y\}$$

Then we can find the PDF of Y by differentiation.

Change of Variable, Method 2 (cont.)

Suppose $Y = X^2$ and X has PDF f_X. What is the PDF of Y?

Since $x \mapsto x^2$ is not invertible if X takes both positive and negative values, the "Jacobian method" is not usable. We use "method 2".

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr(X^2 \le y) \\ &= \Pr(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Change of Variable, Method 2 (cont.)

Differentiating

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

with respect to y we get

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\Big[F_X(\sqrt{y}) - F_X(-\sqrt{y})\Big] \\ &= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} \end{aligned}$$

If X is symmetric about zero, then

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{\sqrt{y}}, \qquad y > 0$$

The Chi-Square Distribution

If X is standard normal, then the distribution of $Y = X^2$ is called the *chi-squared distribution for one degree of freedom*. It has PDF

$$
\begin{aligned}
f_Y(y) &= f_X(\sqrt{y}) \frac{1}{\sqrt{y}} \\
&= \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}
\end{aligned}
$$

From the form of the PDF we see that this is another name for the $\text{Gam}(1/2, 1/2)$ distribution.

Chi-Square Distribution and Gamma Function

From

$$\frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)}y^{-1/2}e^{-y/2}$$

we obtain

$$\frac{1}{\sqrt{2\pi}} = \frac{(1/2)^{1/2}}{\Gamma(1/2)}$$

hence

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

Chi-Square Distribution and Gamma Function (cont.)

Hence using the recursion relation

$$
\begin{aligned}
\Gamma(1/2) &= \sqrt{\pi} \\
\Gamma(3/2) &= \frac{1}{2} \cdot \sqrt{\pi} \\
\Gamma(5/2) &= \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi} \\
\Gamma(7/2) &= \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi}
\end{aligned}
$$

and so forth.

Now we know algebraic values for the gamma function at integer and half integer arguments, but nowhere else.

The Chi-Square Distribution (cont.)

If X_1, $\ldots$, X_n are IID standard normal random variables, then the distribution of $Y = X_1^2 + \cdots + X_n^2$ is called the *chi-squared distribution for n degrees of freedom*. This distribution is abbreviated chi$^2(n)$.

Since each X_i^2 has the Gam$(1/2, 1/2)$ distribution, and since we know the addition rule for gamma random variables, we know that chi$^2(n)$ is another name for the Gam$(n/2, 1/2)$ distribution.

Also from the addition rule for gamma random variables, we get the addition rule for chi-squared random variables. If Y_1, $\ldots$, Y_n are independent chi-squared random variables, Y_i having degrees of freedom k_i, then $Y_1 + \cdots + Y_n$ has the chi-squared distribution for $k_1 + \cdots + k_n$ degrees of freedom.

The Chi-Square Distribution (cont.)

We know the mean and variance of the $\text{Gam}(\alpha, \lambda)$ distribution are

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} \\
\text{var}(X) &= \frac{\alpha}{\lambda^2}
\end{aligned}
$$

(the latter from a homework problem). Specializing to $\alpha = n/2$ and $\lambda = 1/2$ gives

$$
\begin{aligned}
E(Y) &= n \\
\text{var}(Y) &= 2n
\end{aligned}
$$

for the mean and variance of the $\text{chi}^2(n)$ distribution.

The Chi-Square Distribution (cont.)

The latter could also have been calculated directly from the definition of Y as the sum of n IID squared standard normal random variables. From the rules for mean and variance of the sum of IID, these are n times the mean and variance for one, and if $n = 1$, we have

$$E(Y) = \text{var}(X) = E(X^2) = 1$$

where X is standard normal, and

$$\text{var}(Y) = E(Y^2) - E(Y)^2 = E(X^4) - E(X^2)^2 = 2$$

using the values for $E(X^4)$ and $E(X^2)$ calculated in homework.