Name	**pdf**	**Mean**	**Variance**
$Bernoulli(p)$	$P(X=k)=\begin{cases} p & k=1 \\ 1-p & k=0 \end{cases}$	p	$p(1-p)$
$Binomial(n,p)$	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,2,\dots,n$	np	$np(1-p)$
$Geometric(p)$	$P(X=k)=(1-p)^{k-1}p,\quad k=1,2,3,\dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!},\quad k=0,1,2,\dots$	λ	λ
$Uniform(a,b)$	$f_X(x)=\frac{1}{b-a},\quad a\le x\le b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Exponential(\lambda)$	$f_X(x)=\lambda e^{-\lambda x},\qquad x\ge 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$Normal(\mu,\sigma^2)$	$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\quad -\infty\le x\le\infty$	μ	σ^2

I: Moment-Generating Functions

The pdf of a random variable X contains all the information about X. There is another function – the **moment generating function** – that also provides all the information about a random variable.

The **kth moment** of a random variable X is $E(X^k)$. The kth moment is said to exist if $E(X^k)$ is finite. It can be shown that if a moment exists, then lower moments also exist. The moments contain important information about a random variable and can be used to summarize a distribution. For example, the first moment $E(X)$ is the mean. The second moment $E(X^2)$, together with the first moment, gives us the variance. Higher moments provide information such as "skewness" and "flatness". If all the moments of a random variable exist, then it may have a moment generating (mgf).

Definition: The mgf of X is $M_X(t)=E(e^{tX})$, it exists if this is finite on some open interval $(-a,a)$ containing 0. Otherwise, the mgf of X does not exist.

Example 1 Let $X\sim Geometric(p)$. Then

$$M_X(t)=E(e^{tX})=\sum_{k=1}^{\infty}e^{tk}(1-p)^{k-1}p=\frac{p}{1-p}\sum_{k=1}^{\infty}e^{tk}(1-p)^k=\frac{p}{1-p}\sum_{k=1}^{\infty}\left(e^t(1-p)\right)^k$$
$$=\frac{p}{1-p}\left(\frac{1}{1-e^t(1-p)}-1\right)=\frac{pe^t}{1-(1-p)e^t}$$

if $e^t(1-p)<1$ or equivalently $t<a=\ln 1/(1-p)$. Above we used the fact that the sum of the geometric series $\sum_{k=0}^{\infty}r^k=1/(1-r)$ if $|r|<1$. Note that $M_X(t)$ is indeed finite in the open interval $(-a,a)$ containing 0.

Example 2 Let $X\sim Exponential(\lambda)$. Then

$$M_X(t)=E(e^{tX})=\int_0^{\infty}e^{tx}\lambda e^{-\lambda x}\,dx=\lambda\int_0^{\infty}e^{-(\lambda-t)x}\,dx=\frac{\lambda}{\lambda-t}$$

if $\lambda-t>0$ or $t<\lambda$.

Example 3 Let $X\sim Normal(0,1)$. Then

$$M_X(t)=E(e^{tX})=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{tx}e^{-x^2/2}\ dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2+tx}\ dx$$

To evaluate this kind of integrals, we use a standard trick, completing the square,

$$M_X(t)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(x-t)^2/2+t^2/2}\ dx=e^{t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(x-t)^2/2}\ dx=e^{t^2/2}$$

The last integral is 1 because it is the area under a shifted normal curve.

Not all random variables have an mgf, i.e. $M_X(t)=E(e^{tX})$ is not finite on any open interval $(-a,a)$ containing 0.

Example 4 Let X be a random variable with pdf $f_X(x)=1/x^2, x\ge 1$. Then

$$M_X(t)=E(e^{tX})=\int_1^{\infty}e^{tx}\left(\frac{1}{x^2}\right)dx$$

For any $t > 0$, e^{tx} grows faster than x^2, so $\lim_{x\to\infty} e^{tx}/x^2 = \infty$ and $M_X(t) = \infty$.

In fact, the first moment of X, $E(X)$, does not exist and so X does not have any moments:

$$E(X) = \int_1^\infty x\left(\frac{1}{x^2}\right)dx = \int_1^\infty \frac{1}{x}dx = \ln x|_1^\infty = \ln\infty = \infty$$

Aside: There is, however, another function, $C_X(t) = E(e^{itX})$ (where $i = \sqrt{-1}$), called the **characteristic function**, which always exists. Note that the characteristic function is exactly the Fourier transform (and the mgf is close to the Laplace transform) if you are familiar with transforms. We will not look at characteristic functions in this course, but you may see it in a more advanced course.

As its name implies, the mgf generates moments: $E(X^n) = {M_X}^{(n)}(0)$, the nth derivative of $M_X(t)$ with respect to t evaluated at $t = 0$. We can see this by looking at two expansions of $M_X(t)$ around $t = 0$. The Taylor expansion of $M_X(t)$ is

$$M_X(t) = \sum_{n=0}^{\infty} {M_X}^{(n)}(0)\frac{t^n}{n!}$$

But we also have

$$M_X(t) = E(e^{tX}) = E\left(\sum_{n=0}^{\infty}\frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} E(X^n)\frac{t^n}{n!}$$

where we have used the Taylor expansion $e^y = \sum_{n=0}^\infty y^n/n!$ Comparing the two expressions, we have $E(X^n) = {M_X}^{(n)}(0)$.

Example 5 Let $X \sim Exponential(\lambda)$. We now use the mgf, $M_X(t) = \lambda/(\lambda - t)$ to find the mean and variance.

$$E(X) = {M_X}^{(1)}(0) = \frac{d}{dt}\bigg|_{t=0} M_X(t) = \frac{d}{dt}\bigg|_{t=0}\frac{\lambda}{\lambda - t} = \frac{\lambda}{(\lambda-t)^2}\bigg|_{t=0} = \frac{1}{\lambda}$$

$$E(X^2) = {M_X}^{(2)}(0) = \frac{d^2}{dt^2}\bigg|_{t=0}\frac{\lambda}{\lambda - t} = \frac{2\lambda}{(\lambda-t)^3}\bigg|_{t=0} = \frac{2}{\lambda^2}$$

$$Var(X) = E(X^2) - \big(E(X)\big)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

II: Properties and Applications of the Mgf

Lemma: $M_{a+bX}(t) = E\left(e^{t(a+bX)}\right) = E(e^{at}e^{btX}) = e^{at}E(e^{btX}) = e^{at}M_X(bt)$

Example 6 We have shown that for a standard normal variable Z, $M_Z(t) = e^{t^2/2}$. From Math 3081, you know that if $X \sim Normal(\mu, \sigma^2)$, then $X = \mu + \sigma Z$, where $Z \sim Normal(0,1)$. The mgf for X is

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

Lemma: If X and Y are two independent random variables, then $M_{X+Y}(t) = M_X(t)M_Y(t)$

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

Uniqueness Theorem: If two random variables have the same mgf, they must have the same distribution.

This is an extremely powerful theorem but the proof is omitted because it is too advanced for this course.

Example 7 We know from Math 3081 that the sum of normal variables is again a normal variable.

Proof: Let $Y_i \sim Normal(\mu_i, {\sigma_i}^2)$, and $X = \sum_{i=1}^n Y_i$.

$$M_X(t) = M_{Y_1}(t)M_{Y_2}(t)\cdots M_{Y_n}(t) = e^{\mu_1 t + {\sigma_1}^2 t^2/2}e^{\mu_2 t + {\sigma_2}^2 t^2/2}\ldots e^{\mu_n t + {\sigma_n}^2 t^2/2} = e^{(\mu_1+\mu_2+\cdots+\mu_n)t + ({\sigma_1}^2+{\sigma_2}^2+\cdots+{\sigma_n}^2)t^2/2}$$

But this is precisely the mgf of a normal variable with mean $\mu_1 + \mu_2 + \cdots + \mu_n$ and variance ${\sigma_1}^2 + {\sigma_2}^2 + \cdots + {\sigma_n}^2$, so according to the uniqueness of mgf,

$$X \sim Normal\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n {\sigma_i}^2\right)$$

Example 8 Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$. The Central Limit Theorem (CLT) says that for large n, $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$, or equivalently,

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \to Z$$

We will prove the CLT by showing that $\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t) = e^{t^2/2}$.

Let $S_i = (X_i - \mu)/\sigma$. Then

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} = \frac{S_1 + \cdots + S_n}{\sqrt{n}}$$

Note that $E(S_i) = 0$, and $\mathrm{Var}(S_i) = 1$. Also,

$$M_{Z_n}(t) = E(e^{tZ_n}) = E\left[e^{t\left(\frac{S_1+\cdots+S_n}{\sqrt{n}}\right)}\right] = E\left(e^{\frac{t}{\sqrt{n}}S_1} \cdots e^{\frac{t}{\sqrt{n}}S_n}\right) = E\left(e^{\frac{t}{\sqrt{n}}S_1}\right) \cdots E\left(e^{\frac{t}{\sqrt{n}}S_n}\right) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

where $M(t)$ is the mgf for S_i. Since n is large (so $t/\sqrt{n}$ is small), we can do Taylor's expansion around 0 to approximate $M\left(\frac{t}{\sqrt{n}}\right)$. Note that $M(0) = 1$ (because $M(0) = E[e^{0 \cdot S_i}] = 1$), $M'(0) = 0$ (because $M'(0) = E[S_i] = 0$), $M''(0) = 1$ (because $M''^{(0)} = E[S_i{}^2] = \mathrm{Var}[S_i] + E[S_i]^2 = 1$). So,

$$M\left(\frac{t}{\sqrt{n}}\right) \approx 1 + M'(0)\left(\frac{t}{\sqrt{n}}\right) + \frac{1}{2}M''(t)\left(\frac{t}{\sqrt{n}}\right)^2 = 1 + \frac{t^2}{2n}$$

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left(1 + \frac{t^2}{2n}\right)^n = e^{t^2/2}$$

The last equality comes from

$$\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a$$

CHAPTER 2 TRANSFORMATIONS

I. Change of Variables – One Variable (One Dimensional)

We are interested in the following question: let X be a continuous random variable with pdf $f_X(x)$, and let $Y = h(X)$ (we call such a change of variable a **transformation**), what is the pdf of Y? Later we will want the pdf of $X = Z^2$. It turns out that it is not harder to treat a more general problem.

Theorem 1 If X be a continuous random variable with pdf $f_X(x)$, then the pdf of $Y = X^2$ is

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right)$$

Proof We first find the cdf of Y

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = F_X(x) - F_X(-x)$$

The pdf of Y is (chain rule is used in the third equality)

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left(F_X(x) - F_X(-x)\right) = \frac{d}{dx}\left(F_X(x) - F_X(-x)\right)\frac{dx}{dy} = \left(f_X(x) + f_X(-x)\right)\frac{1}{2\sqrt{y}}$$

$$= \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right)$$

Example 1 Let $X \sim Normal(\mu, \sigma^2)$, what is the pdf of $Y = Z^2$?

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left(f_Z(\sqrt{y}) + f_Z(-\sqrt{y})\right) = \frac{1}{2\sqrt{y}}\left(\frac{1}{\sqrt{2\pi}}e^{-y/2} + \frac{1}{\sqrt{2\pi}}e^{-y/2}\right) = \frac{1}{\sqrt{2\pi y}}e^{-y/2}$$

Theorem 2 Let $Y = h(X)$ where h is differentiable and strictly increasing or strictly decreasing (i.e. h is a one-to-one function). Then the pdf of Y is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

If it is easier, you can find dy/dx and take its reciprocal, but $f_Y(y)$ should be a function only of y.

Proof Suppose $h(X)$ is strictly increasing, then dy/dx is positive, and the cdf of Y is

$$F_Y(y) = P(Y \le y) = P(h(X) \le y) = P\left(X \le h^{-1}(y)\right) = F_X\left(h^{-1}(y)\right) = F_X(x)$$

The pdf of Y is

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(x) = \frac{d}{dx}F_X(x)\frac{dx}{dy} = f_X(x)\frac{dx}{dy}$$

Next suppose $h(X)$ is strictly decreasing, then dy/dx is negative, and the cdf of Y is

$$F_Y(y) = P(Y \le y) = P(h(X) \le y) = P\left(X \ge h^{-1}(y)\right) = 1 - F_X\left(h^{-1}(y)\right) = 1 - F_X(x)$$

The pdf of Y is

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left(1 - F_X(x)\right) = -\frac{d}{dy}F_X(x) = -\frac{d}{dx}F_X(x)\frac{dx}{dy} = -f_X(x)\frac{dx}{dy}$$

Thus,

$$f_Y(y) = \begin{cases} f_X(x)\dfrac{dx}{dy} & when\ \dfrac{dx}{dy} > 0 \\ -f_X(x)\dfrac{dx}{dy} & when\ \dfrac{dx}{dy} < 0 \end{cases} = f_X(x)\left|\frac{dx}{dy}\right|$$

Example 2 Let $X \sim Uniform[0,1]$. Find the pdf of the following random variables.

a) $Y = -X$

The pdf for X is $f_X(x) = 1, 0 \le x \le 1$. There are two ways to calculate dx/dy:

$$y = -x \Rightarrow \frac{dy}{dx} = -1 \Rightarrow \frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1} = -1$$

$$y = -x \Rightarrow x = -y \Rightarrow \frac{dx}{dy} = -1$$

The pdf for Y is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = 1|-1| = 1, \quad -1 \le y \le 0\ (this\ domain\ is\ from\ y = -x)$$

b) $Z = e^X$

Two ways of calculating dx/dy:

$$z = e^x \Rightarrow \frac{dz}{dx} = e^x \Rightarrow \frac{dx}{dz} = e^{-x} = \frac{1}{z}$$

$$z = e^x \Rightarrow x = \ln z \Rightarrow \frac{dx}{dz} = \frac{1}{z}$$

The pdf for Z is

$$f_Z(z) = f_X(x)\left|\frac{dx}{dz}\right| = 1\left|\frac{1}{z}\right| = \frac{1}{z}, \quad 1 \le z \le e\ (this\ domain\ is\ from\ z = e^x)$$

II. Change of Variables – Multi-Variable (Multi-Dimensional)

There are situations where we are interested in 2 or more random variables simultaneously, in which case the **joint pdf** of all the variables encodes all we can know. Recall that the joint pdf for continuous random variables X and Y, denoted by $f_{X,Y}(x,y)$, satisfies $f_{X,Y}(x,y) \ge 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy = 1$. The **marginal pdf** for each of the variable is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

It is also useful to remember that if X and Y are **independent**, then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. We can easily extend to n variables: the joint pdf is a function of n variables, $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$. To obtain the marginal pdf for X_i, we integrate over all variables except x_i, i.e. the marginal pdf is a $n-1$ dimensional integral. It is convenient mathematically to represent the n variables $X_1,\ldots,X_n$ as a random vector, $\boldsymbol{X} = (X_1,\ldots,X_n)$ (we use boldface to denote vector), and we may write the joint pdf as $f_{\boldsymbol{X}}(\boldsymbol{x})$. We may just say pdf instead of joint pdf as the context should make it clear that the pdf is a joint pdf. We will usually look at the case of only two variables. However, we will start the discussion of the multi-variable transformation with n variables. The following theorem is a generalization of Theorem 2.

Theorem 3 Let $\boldsymbol{X}$ be a continuous random vector with pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$, and let $\boldsymbol{Y} = g(\boldsymbol{X})$, where g is an invertible function $\mathbb{R}^n \to \mathbb{R}^n$. Then the pdf of $\boldsymbol{Y}$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x})\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right|$$

where $\partial\boldsymbol{x}/\partial\boldsymbol{y}$ is called the **Jacobian**, which is the determinant of the $n \times n$ matrix of all partial derivatives

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} = \begin{pmatrix} \dfrac{\partial x_1}{\partial y_1} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{pmatrix}$$

As with the single variable case, conveniently $|\partial\boldsymbol{x}/\partial\boldsymbol{y}| = |\partial\boldsymbol{y}/\partial\boldsymbol{x}|^{-1}$, and you can find $|\partial\boldsymbol{y}/\partial\boldsymbol{x}|$ and take its reciprocal.

Example 3 Let $T \sim Uniform(0, 2\pi)$ and $U \sim Exponential(1)$, with T and U independent. Let $X = \sqrt{2U}\cos T$ and $Y = \sqrt{2U}\sin T$. Find the joint pdf of (X, Y). What are the marginal pdf's? Are X and Y independent?

We have $f_T(t) = 1/2\pi,\ 0 \le t \le 2\pi$, and $f_U(u) = e^{-u},\ x \ge 0$. Since T and U are independent, the joint pdf is

$$f_{T,U}(t,u) = f_T(t) f_U(u) = \frac{-e^{u}}{2\pi}\ , \qquad 0 \le t \le 2\pi, x \ge 0$$

$$\frac{\partial(x,y)}{\partial(t,u)} = \det\begin{pmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial u} \end{pmatrix} = \det\begin{pmatrix} -\sqrt{2u}\sin t & \frac{1}{\sqrt{2u}}\cos t \\ \sqrt{2u}\cos t & \frac{1}{\sqrt{2u}}\sin t \end{pmatrix} = -\sin^2 t - \cos^2 t = -1$$

$$f_{X,Y}(x,y) = f_{T,U}(t,u)|-1| = \frac{e^{-u}}{2\pi} = \frac{e^{-(x^2+y^2)}}{2\pi} = \left(\frac{1}{\sqrt{2\pi}}e^{-x^2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-y^2}\right), \qquad x, y \in \mathbb{R}$$

In the third equality, we express u in terms of x and y: $x^2 + y^2 = 2u$. To find the marginal pdf's, we integrate the joint pdf, but in this case, it is quite clear that $f_{X,Y}(x,y)$ is the product of two normal pdf's, and X and Y are independent.

In the next two sections, we will discuss two important transformations, convolutions and order statistics.

III. Convolutions

Theorem 4 Suppose X and Y are independent random variables with pdf's $f_X(x)$ and $f_Y(y)$, then the random variable $Z = X + Y$ has the pdf $f_Z(z) = \int_{-\infty}^{\infty} f_X(w)\, f_Y(z-w)dw$.

Proof One way to show this is to consider the transformation from (X,Y) to $(Z,W) = (X+Y, X)$.

$$\frac{\partial(z,w)}{\partial(x,y)} = \det\begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{pmatrix} = \det\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = -1$$

$$f_{Z,W}(z,w) = f_{X,Y}(x,y)|-1| = f_X(x) f_Y(y) = f_X(w) f_Y(z-w)$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z,w)\, dw = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\, dw$$

Example 4 Let X and Y be independent random variables with pdf $Exponential(\lambda)$. Find the pdf for $Z = X + Y$.

The only tricky part of this problem is to be careful about the integration bound. Since each of $f_X(x)$ and $f_Y(y)$ is non-zero only when $x \ge 0$ and $y \ge 0$, the integrand in the convolution integral is only non-zero when $w \ge 0$ and $z - w \ge 0$ or $w \le z$, so the integration limits are 0 and z.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w)\, f_Y(z-w)dw = \int_0^z \left(\lambda e^{-\lambda w}\right)\left(\lambda e^{-\lambda(z-w)}\right) dw = \int_0^z \lambda^2 e^{-\lambda z}\, dw = \lambda^2 e^{-\lambda z} w\Big|_0^z = \lambda^2 e^{-\lambda z} z$$

IV. Order Statistics

There is another useful transformation, that which takes n iid's $X_1, \dots, X_n$ and sorts them in order from minimum to maximum: $X_1' < X_2' < \cdots < X_n'$. These transformed variables are called the order statistics. We are interested in the pdf of the smallest and largest order statistics, $X_{\min} = X_1' = \min(X_1, \dots, X_n)$ and $X_{\max} = X_n' = \max(X_1, \dots, X_n)$.

Theorem 5 Suppose $X_1, \dots, X_n$ are iid's each with pdf $f_X(x)$ and cdf $F_X(x)$. The pdf of the extreme order statistics are

$$f_{X_{\max}}(x) = n\left(F_X(x)\right)^{n-1} f_X(x)$$

$$f_{X_{\min}}(x) = n\left(1 - F_X(x)\right)^{n-1} f_X(x)$$

Proof This transformation is **not** invertible: if we know $X_1' = \min(X_1, X_2) = 1$ and $X_2' = \max(X_1, X_2) = 2$, then we know that one of X_1 and X_2 is 1, and the other is 2, but we don't know which is which. So we cannot use Theorem 3 to prove this, but use the more basic approach of first finding the cdf and then the pdf.

$$F_{X_{\max}}(x) = P(X_{\max} < x) = P(X_1 < x, X_2 < x, \cdots, X_n < x) = P(X_1 < x)P(X_2 < x)\cdots P(X_n < x) = \left(F_X(x)\right)^n$$

$$f_{X_{\max}}(x) = \frac{d}{dx}F_{X_{\max}}(x) = \frac{d}{dx}\left(F_X(x)\right)^n = n\left(F_X(x)\right)^{n-1}\frac{d}{dx}F_X(x) = n\left(F_X(x)\right)^{n-1} f_X(x)$$

$$\begin{aligned} F_{X_{\min}}(x) &= P(X_{\min} < x) = 1 - P(X_{\min} > x) = 1 - P(X_1 > x, X_2 > x, \cdots, X_n > x) \\ &= 1 - P(X_1 > x)P(X_2 > x)\cdots P(X_n > x) = 1 - \left(1 - F_X(x)\right)^n \end{aligned}$$

$$f_{X_{\min}}(x) = \frac{d}{dx}F_{X_{\min}}(x) = \frac{d}{dx}\left(1 - \left(1 - F_X(x)\right)^n\right) = n\left(1 - F_X(x)\right)^{n-1}\frac{d}{dx}F_X(x) = n\left(1 - F_X(x)\right)^{n-1} f_X(x)$$

Example 5 Let X_1 and X_2 be independent random variables with pdf $Exponential(\lambda)$. Find the pdf for $X_{\max}$ and $X_{\min}$.

$$f_X(x) = \lambda e^{-\lambda x}$$

$$F_X(x) = \int_0^x \lambda e^{-\lambda t}\, dt = e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}$$

$$f_{X_{\max}}(x) = 2F_X(x) f_X(x) = 2\lambda\left(1 - e^{-\lambda x}\right) e^{-\lambda x}$$

$$f_{X_{\min}}(x) = 2\left(1 - F_X(x)\right) f_X(x) = 2e^{-2\lambda x}$$

Let's compare $E(X_1)$ and $E(X_{\min})$: $E(X_{\min}) = 1/2 < E(X_1) = 1$ as expected.

CHAPTER 3 SPECIAL DISTRIBUTIONS

I. Stories of Special Distributions

In Chapter 1 Section 1, we listed some common distributions. These distributions are special for a reason, which is referred to as the "story" of the distribution by Professor Blitzstein of Harvard Statistics Department. Following are the stories of the common distributions in Chapter 1 Section 1. It is very useful to remember these stories.

Name	Story
$Bernoulli(p)$	Experiment: Toss a coin with probability p of turning up heads. $X =$ Number of heads in one toss
$Binomial(n, p)$	Experiment: Toss a coin with probability p of turning up heads. $X =$ Number of heads in n tosses $Binomial(n, p)$ is the sum of n iid's of $Bernoulli(p)$
$Geometric(p)$	Experiment: Toss a coin with probability p of turning up heads. $X =$ number of tosses until the first head
$Poisson(\lambda)$	Experiment: Observe buses, which arrive with rate λ per unit time. $X =$ Number of buses that arrive per unit time
$Uniform(a, b)$	Experiment: Pick a number between a and b. $X =$ Number picked
$Exponential(\lambda)$	Experiment: Observe buses, which arrive with rate λ per unit time. $X =$ Amount of time until the arrival of first bus
$Normal(\mu, \sigma^2)$	**Story 1** Experiment: Pick an individual in a large population $X =$ Height of the individual **Story 2** Central Limit Theorem

II. Gamma and Beta Distributions

Definition The gamma distribution has two real number parameters, α, λ. The pdf of a gamma random variable $X \sim Gamma(\alpha, \lambda)$ is

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

where $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1}\, dx$ is the gamma function. In homework, you will prove the following properties:

- $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
- $\Gamma(n) = (n-1)!$ for positive integer n

So the gamma function is a generalization of the factorial function. Here is the story of gamma distribution, when α is an integer. Suppose we observe buses, which arrive with rate λ per unit time. Let T_k be the time until the kth arrival. Then $T_k \sim Gamma(k, \lambda)$, and

$$f_{T_1}(x) = \frac{\lambda^1 x^{1-1}}{\Gamma(1)} e^{-\lambda x} = \lambda e^{-\lambda x}, x > 0$$

So $Gamma(1, \lambda) = Exponential(\lambda)$, i.e. the gamma distribution is a generalization of the exponential distribution.

The expected value and variance of a gamma random variable can be calculated as follows.

$$E(X)=\int_0^\infty \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda x}\,x\,dx=\int_0^\infty \frac{\lambda^\alpha x^{\alpha}}{\Gamma(\alpha)}e^{-\lambda x}\,dx=\frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)}\int_0^\infty \frac{\lambda^{\alpha+1} x^{\alpha}}{\Gamma(\alpha+1)}e^{-\lambda x}\,dx=\frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)}=\frac{\alpha}{\lambda}$$

We have used a very cool trick in the third and fourth equalities that makes use of the fact the integral of a pdf is 1. So in the third equality we do an algebraic manipulation that results in an integral of a pdf (of $Gamma(\alpha+1,\lambda)$). Take note of this trick (called "**pdf recognition**" by some)! Using this trick you can calculate $E(X^2)$ and then $Var(X)=\alpha/\lambda^2$ (in-class exercise), and mgf $M_X(t)=E(e^{tX})=\lambda^\alpha(\lambda-t)^{-\alpha}$ (homework). Using the mgf, we can show that the sum of n independent exponential iid's with parameter λ has the distribution $Gamma(n,\lambda)$ and the sum of $Gamma(a,\lambda)$ and $Gamma(b,\lambda)$ is $Gamma(a+b,\lambda)$ (homework).

Example 1 Planes are arriving at an airport at the rate of 5 per hour. Andrew, an 8-year-old boy, is an airplane enthusiast. His father brings him to an island next to the airport where they can see all the arriving airplanes. They plan to leave when they see 6 airplanes. Let T be the time they spend at the airport. Then $T\sim Gamma(6,5)$ and $E(T)=6/5$.

Definition The beta distribution has two positive real number parameters, a,b. The pdf of a beta random variable $X\sim Beta(a,b)$ is

$$f_X(x)=\frac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1},\ \ 0<x<1$$

where $\beta(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}\,dx=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ (homework) is the beta function. When $a=b=1$, $f_X(x)=1,\ \ 0<x<1$, i.e. $Beta(1,1)=Uniform(0,1)$. So the beta distribution is a generalization of $Uniform(0,1)$.

Again, we will use the good old "pdf recognition" trick to find the expected value here.

$$E(X)=\int_0^\infty \frac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1}\,x\,dx=\int_0^\infty \frac{1}{\beta(a,b)}x^{a}(1-x)^{b-1}\,dx=\frac{\beta(a+1,b)}{\beta(a,b)}\int_0^\infty \frac{1}{\beta(a+1,b)}x^{a}(1-x)^{b-1}\,dx$$
$$=\frac{\beta(a+1,b)}{\beta(a,b)}=\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right)=\frac{a}{a+b}$$

In homework, you will show that the variance is

$$Var(X)=\frac{ab}{(a+b+1)(a+b)^2}$$

The following example reveals a connection between the gamma and beta distributions.

Example 2 Andrew, an 8-year-old boy, is an airplane enthusiast. His father brings him for airplane sighting at two airports, where planes are arriving at an airport at the rate of 5 per hour in each airport. They plan to leave the first airport when they see a airplanes there and go to the second airport. They will leave the second airport after they see b airplanes. Let X and Y be the times they spend in the first and second airports respectively. From Example 1, we know that $X\sim Gamma(a,\lambda)$ and $Y\sim Gamma(b,\lambda)$. Assume that X and Y are independent. Find the joint pdf of $T=X+Y$ (total time spent) and $W=X/(X+Y)$ (the fraction of time spent in the first airport).

$$f_X(x)=\frac{\lambda^a x^{a-1}}{\Gamma(a)}e^{-\lambda x},\ \ x>0 \qquad\qquad f_Y(y)=\frac{\lambda^b x^{b-1}}{\Gamma(b)}e^{-\lambda y},\ \ y>0$$

Since X and Y are independent, the join pdf is

$$f_{X,Y}(x,y)=f_X(x)\,f_Y(y)=\frac{\lambda^a x^{a-1}}{\Gamma(a)}e^{-\lambda x}\frac{\lambda^b y^{b-1}}{\Gamma(b)}e^{-\lambda y},\qquad x,y>0$$

To calculate the Jacobian, note that $w=x/t\ \Rightarrow x=tw$ and $y=t-x=t-tw=t(1-w)$.

$$\frac{\partial(t,w)}{\partial(x,y)}=\det\begin{pmatrix}\frac{\partial t}{\partial x} & \frac{\partial t}{\partial y}\\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y}\end{pmatrix}=\det\begin{pmatrix}1 & 1\\ \frac{y}{(x+y)^2} & -\frac{x}{(x+y)^2}\end{pmatrix}=\det\begin{pmatrix}1 & 1\\ \frac{1-w}{t} & -\frac{w}{t}\end{pmatrix}=-\frac{1}{t}$$

$$f_{T,W}(t,w)=f_{X,Y}(x,y)|-t|=\frac{\lambda^a(tw)^{a-1}}{\Gamma(a)}e^{-\lambda x}\frac{\lambda^b\big(t(1-w)\big)^{b-1}}{\Gamma(b)}e^{-\lambda y}t=\frac{1}{\Gamma(a)\Gamma(b)}w^{a-1}(1-w)^{b-1}(\lambda t)^{a+b}e^{-\lambda t}\frac{1}{t}$$
$$=\left(\frac{1}{\Gamma(a+b)}(\lambda t)^{a+b-1}e^{-\lambda t}\right)\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}w^{a-1}(1-w)^{b-1}\right)$$

To find the marginal pdf's, we integrate the joint pdf, but in this case, it is quite clear that $f_{T,W}(t,w)$ has separated into the product of a gamma pdf and a beta pdf! So T and W are independent, and $T\sim Gamma(a+b,\lambda)$ and $W\sim Beta(a,b)$.

III. Normal Distribution Theory

Because of the Central Limit Theorem, the normal distribution has a prominent role in statistics (the subject of the next 3 chapters, Chapters 4 to 6). In this section, we will look at a number of distributions related to the normal distribution.

Definition If Z_i's are independent standard normal variables, then $X = {Z_1}^2 + {Z_2}^2 + \cdots + {Z_m}^2$ has a **chi-square distribution** with degree of freedom (df) m, $X \sim \chi^2(m)$.

We have shown that the pdf for ${Z_i}^2$ is $e^{-z/2}/\sqrt{2\pi z}$, but this is the pdf for $Gamma(1/2, 1/2)$. So X is a sum of m iid's with pdf of $Gamma(1/2, 1/2)$. According to our discussion in Section II, $X \sim Gamma(m/2, 1/2)$, and we have $E(X) = m$ and $Var(X) = 2m$.

Definition Let $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$. Then the distribution of $(X/m)/(Y/n)$ is called the ***F* distribution** with degrees of freedom m and n, $F(m, n)$.

It can be shown by a change of variables from (X, Y) to (Y, W) (let $W = X$) that the pdf of $Y \sim F(m, n)$ is

$$f_Y(y) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}y\right)^{\frac{m}{2}-1}\left(1 + \frac{m}{n}y\right)^{-\frac{n+1}{2}}$$

It is a substantial exercise (it fits on one page if you write small enough) or you can also find it in many probability books.

Definition Let $Z \sim Normal(0,1)$ and $X \sim \chi^2(n)$. Then the distribution of $Z/\sqrt{X/n}$ is called the **Student *t* distribution**, or simply the t distribution, with n degrees of freedom, $t(n)$.

From the two definitions above, we see immediately that if $U \sim t(n)$ then $U^2 \sim F(1, n)$.

The pdf of $U \sim t(n)$ is given by

$$f_U(u) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{u^2}{n}\right)^{-(n+1)/2}$$

This can be shown a couple of ways but we will not go into the details. You can consult a probability book (or talk to me) if you are interested. Also, we note that $t(1)$ is known as the **Cauchy** distribution.

In Math 3081, you may have seen the strong law of large number, which states that as $n \to \infty$, the average of n iid's tends to the expected value of the random variable with probability 1. So, as $n \to \infty$, $X/n = \left({Z_1}^2 + \cdots + {Z_n}^2\right)/n \to E(Z^2) = 1$ according to the strong law of large numbers (recall that $Z^2 \sim Gamma(1/2, 1/2)$ and $E(Z^2) = 1$), i.e. $\lim_{n\to\infty} t(n) = Z$.