STAT 501 Fall 2025 Course Notes

Chapter 3: Common Families of Distributions

Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.[1] *–Box & Draper, 1987*

3.1 Introduction

Statistical distributions are used to model populations or random phenomena. Therefore, we usually deal with families of distributions, where each **family** is indexed by one or more **parameters** that allow us to vary certain characteristics of the distribution, such as shape and/or spread, while staying with one functional form.

There are many common discrete and continuous distributions, and we'll discuss a few of them, along with their interrelationships and common applications, in the following sections.

Working with Distributions in R

> **R Distribution Prefixes** The R programming language has a general convention for R functions involving distributions. Each distribution is associated with four R functions that start with `d`, `p`, `q`, or `r`:
>
> `d...`
>
> `p...`
>
> `q...`
>
> `r...`

To see a list of all the named distributions that have built-in R functions in the base library, type `?Distributions` in the R Console.

---

[1] https://medium.com/swlh/all-models-are-wrong-does-not-mean-what-you-think-it-means-610390c40c9c

Distributional Parameters and Notation

Each named distribution generally has a shorthand notation for its name and parameters. For example, we will see that the normal distribution has pdf

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right).$$

The notation $f(x|\mu, \sigma)$ implies the value of the function is dependent on the values of μ and σ. These values, required for the calculation of any probability, are known as **distributional parameters** or **parameters** of the distribution. Thus, the normal distribution is actually a **family** of distributions, parameterized by the values μ and σ. If X has a normal distribution with parameters μ and σ, we would denote this by:

3.2 Discrete Distributions

A random variable X is said to have a discrete distribution if its support, $\mathcal{X}$, is countable (or if its CDF is a step function). In most discrete distributions, the random variable has integer-valued outcomes. In this section, we'll discuss the following discrete distributions: 1) Discrete Uniform; 2) Binomial; 3) Negative Binomial; 4) Geometric; 5) Hypergeometric; and 6) Poisson.

Discrete Uniform Distribution

When Used?

This distribution puts equal mass on each of the outcomes $1, 2, 3, \ldots, N$. That is, each of the N outcomes has equal probability of being observed.

PMF

A random variable X as a **discrete uniform** $(1, N)$ distribution if

$$f_X(x|N) = P(X = x|N) = \begin{cases} \dfrac{1}{N} & x = 1, 2, \ldots, N; \quad N \in \mathbb{Z}^+ \\ \\ 0 & \text{else} \end{cases}$$

MGF

$$M_X(t) = \frac{1}{N}\sum_{k=1}^{N} e^{kt} = \frac{e^t\left(1 - e^{Nt}\right)}{N(1 - e^t)}$$

Mean and Variance

$$E(X) = \frac{N+1}{2} \qquad\qquad Var(X) = \frac{(N+1)(N-1)}{12}$$

Example

The German Tank Problem is a well-known example that uses a discrete uniform distribution to model the serial number of a tank captured during World War II in order to estimate N, the total number of tanks produced.

Fun Fact

The distribution can be generalized so that the sample space is any range of integers, $N_0, N_0 + 1, \ldots, N_1$, with pmf

$$P(X = x | N_0, N_1) = \frac{1}{N_1 - N_0 + 1} I_{\{N_0, N_0+1, \ldots, N_1\}}(x)$$

In R

The base R library doesn't have built-in functions for the discrete uniform distribution and one can use the `sample` function to simulate random draws, but there are discrete uniform R functions in the `extraDistr` library:

To find $P(X = x)$, use:	`ddunif(x, min = N0, max = N1)`
To find $P(X \leq x)$, use:	`pdunif(x, min = N0, max = N1)`
To find smallest x^* such that $P(X \leq x^*) \geq c$, use:	`qdunif(c, min = N0, max = N1)`
To simulate m random draws, use:	`rdunif(m, min = N0, max = N1)`

Bernoulli Distribution

When Used?

When an experiment has only two possible outcomes: "success" or "failure." A Bernoulli random variable X is defined in the following manner:

$$X = \begin{cases} 1 & \text{if success} \\ 0 & \text{if failure} \end{cases}$$

where $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

PMF

A random variable X has a **Bernoulli(p)** distribution if,

$$f_X(x|p) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1; \;\; 0 \le p \le 1 \\ 0 & \text{else} \end{cases}$$

MGF

$$M_X(t) = pe^t + (1 - p)$$

Mean

$$E(X) = p$$

Variance

$$Var(X) = p(1 - p)$$

Examples

- Outcome of one coin flip

- Whether a randomly selected student is infected with a disease

Fun Facts

- A Bernoulli trial is named for James Bernoulli, one of the founding fathers of probability theory.

- A Bernoulli distribution is a special case of the Binomial distribution where $n = 1$.

In R

To find $P(X = x)$, use:	`dbinom(x, n = 1, p)`
To find $P(X \le x)$, use:	`pbinom(x, n = 1, p)`
To find smallest x^* such that $P(X \le x^*) \ge c$, use:	`qbinom(c, n = 1, p)`
To simulate m random draws, use:	`rbinom(m, n = 1, p)`

Binomial Distribution

When Used?

When a random variable is a result of a sequence of independent Bernoulli trials, and we are interested in the number of successes observed during a fixed number of trials. This requires:

1. A fixed number of n identical trials.

2. Each trial results in one of two possible outcomes: "success" or "failure."

3. The probability of success (denoted by p) remains the same for each trial. The probability of failure is then denoted as $1 - p$.

4. The trials are independent.

5. The random variable, X, is defined as the number of successes observed during the n trials.

PMF

A random variable X has a **Binomial(n, p)** distribution if

$$f_X(x|n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n; \quad 0 \leq p \leq 1 \\ 0 & \text{else} \end{cases}$$

MGF

$$M_X(t) = \left[pe^t + (1-p) \right]^n$$

Mean and Variance

$$E(X) = np \qquad\qquad Var(X) = np(1-p)$$

Examples

- Number of heads observed in n coin flips

- Number of n randomly selected voters who voted for a particular candidate

- Number of n randomly selected students who are infected with a disease

Fun Fact

Recursive relationship: for $x = 1, 2, \ldots, n$,

$$P(X = x) = \left(\frac{n - x + 1}{x} \right) \left(\frac{p}{1 - p} \right) P(X = x - 1).$$

In R

To find $P(X = x)$, use:	`dbinom(x, n, p)`
To find $P(X \leq x)$, use:	`pbinom(x, n, p)`
To find smallest x^* such that $P(X \leq x^*) \geq c$, use:	`qbinom(c, n, p)`
To simulate m random draws, use:	`rbinom(m, n, p)`

Geometric Distribution

When Used?

When a random variable is a result of a sequence of independent Bernoulli trials, and we are interested in the trial on which the first success occurs. This requires:

1. Each trial results in one of two possible outcomes: "success" or "failure."

2. The probability of success (denoted by p) remains the same for each trial. The probability of failure is then denoted as $1 - p$.

3. The trials are independent.

4. The random variable, X, is defined as the trial on which the first success occurs.

PMF

A random variable X has a **Geometric(p)** distribution if

$$f_X(x|p) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, 3, \ldots; \quad 0 \leq p \leq 1 \\ 0 & \text{else} \end{cases}$$

MGF

$$M_X(t) = \frac{pe^t}{1 - (1-p)e^t}, \quad t < -\ln(1-p)$$

Mean and Variance

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$$

Examples

- Number of coin flips to observe first head flipped

- Number of students randomly chosen until one infected with a disease is found

Fun Facts

- The geometric distribution is a special case of the negative binomial distribution where $r = 1$.

- The geometric distribution has what's known as the **memoryless property** (see pg. 97).

- The geometric distribution can also be re-parameterized to represent the number of failures before the first success occurs.

Example: Suppose a baseball player has a batting average of .300. Assuming times at bat are independent, let the random variable X represent the number of times at bat since the season started when he gets his first hit. If he has had no hits in his first 20 at bats, what is the probability it takes him more than 25 at bats to get his first hit of the season?

In R

	No. of Trials	No. of Failures
To find $P(X = x)$, use:	dgeom(x-1, p)	dgeom(x, p)
To find $P(X \leq x)$, use:	pgeom(x-1, p)	pgeom(x, p)
To find smallest x^* such that $P(X \leq x^*) \geq c$, use:	qgeom(c, p) + 1	qgeom(c, p)
To simulate m random draws, use:	rgeom(m, p) + 1	rgeom(m, p)

7

Negative Binomial Distribution

When Used?

When a random variable is a result of a sequence of independent Bernoulli trials, and we are interested in the trial on which the rth success occurs $(r = 1, 2, 3, \ldots)$. This requires:

1. Each trial results in one of two possible outcomes: "success" or "failure."

2. The probability of success (denoted by p) remains the same for each trial. The probability of failure is then denoted as $1 - p$.

3. The trials are independent.

4. The random variable, X, is defined as the trial on which the rth success occurs $(r = 1, 2, 3, \ldots)$.

PMF

A random variable X has a **Negative Binomial(r, p)** distribution if

$$f_X(x|r, p) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & x = r, r+1, r+2, \ldots; \quad 0 \le p \le 1 \\ 0 & \text{else} \end{cases}$$

MGF

$$M_X(t) = \left[\frac{pe^t}{1 - (1-p)e^t} \right]^r, \quad t < -\ln(1-p)$$

Mean

$$E(X) = \frac{r}{p}$$

Variance

$$Var(X) = \frac{r(1-p)}{p^2}$$

Examples

- Number of coin flips to observe 7 heads
- Number of students randomly chosen until 100 infected with a disease are found

Fun Fact

The negative binomial distribution can be re-parameterized to represent the number of failures before the rth success occurs $(r = 1, 2, 3, \ldots)$.

In R

	No. of Trials	No. of Failures
To find $P(X = x)$, use:	`dnbinom(x-r, r, p)`	`dnbinom(x, r, p)`
To find $P(X \leq x)$, use:	`pnbinom(x-r, r, p)`	`pnbinom(x, r, p)`
To find smallest x^* such that $P(X \leq x^*) \geq c$, use:	`qnbinom(c, r, p) + r`	`qnbinom(c, r, p)`
To simulate m random draws, use:	`rnbinom(m, r, p) + r`	`rnbinom(m, r, p)`

What are the differences and similarities between Binomial and Negative Binomial random variables?

Hypergeometric Distribution

When Used?

When a random variable represents the number of items with a certain characteristic observed in a sample of size n from a finite population where there are M total items with that characteristic and $N - M$ total items without the characteristic ($M = 0, 1, 2, 3, \ldots$). The following conditions are required:

1. We are selecting a simple random sample of size n from a finite population of size N.

2. Each item in the population possesses one of two characteristics (which could be denoted as "success" or "failure"), M with the characteristic of interest and $N - M$ without that characteristic.

3. The random variable, X, is defined as the number of items with a certain characteristic observed in a sample of size n from the finite population of size N.

PMF

A random variable X has a **Hypergeometric(N, M, n)** distribution if

$$f_X(x|N, M, n) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{m}} & x = a, a+1, \ldots, b; \quad N, M, n \geq 0 \\ \\ 0 & \text{else} \end{cases}$$

where $a = \max(0, n + M - N)$ and $b = \min(n, M)$.

Note: If $n << M$ and N, then the range of $x = 0, 1, 2, \ldots, n$ will be appropriate.

MGF

Not useful.

Mean

$$E(X) = \frac{nM}{N}$$

Variance

$$Var(X) = \frac{nM}{N} \frac{(N-M)(N-n)}{N(N-1)}$$

Example

Number of defective machine parts observed in four parts randomly sampled from a shipment of 25 total parts.

Fun Fact

As $N \to \infty$ and $M \to \infty$ with $p = \frac{M}{N}$, the Hypergeometric distribution converges to a Binomial(n, p) distribution. (Why does this make intuitive sense?)

In R

To find $P(X = x)$, use:	`dhyper(x, M, N-M, n)`
To find $P(X \leq x)$, use:	`phyper(x, M, N-M, n)`
To find smallest x^* such that $P(X \leq x^*) \geq c$, use:	`qhyper(c, M, N-M, n)`
To simulate m random draws, use:	`rhyper(m, M, N-M, n)`

Example: A shipment of 50 refurbished smartphones were sent to a Bozeman distributor, and a family purchases six of them. Suppose 15 of the refurbished phones are still malfunctioning. What is the probability the family received at least one malfunctioning phone?

Poisson Distribution

When Used?

When a random variable represents a number of occurrences over a certain amount of time or space.

PMF

A random variable X has a **Poisson(λ)** distribution if

$$f_X(x|\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, 2 \ldots; \quad \lambda > 0 \\ 0 & \text{else} \end{cases}$$

MGF

$$M_X(t) = \exp(\lambda(e^t - 1))$$

Mean

$$E(X) = \lambda$$

Variance

$$Var(X) = \lambda$$

Examples

- The number of accidents at an intersection in a week

- The number of hits to a website each minute

- The number of plants of a particular species found in a $1\ m^2$ area

Fun Facts

- Recursive relationship: $P(X = x) = \frac{\lambda}{x}P(X = x - 1)$ for $x = 1, 2, \ldots$

- As $n \to \infty$ and $p \to 0$ with $np = \lambda$, the Binomial distribution converges to a Poisson distribution.

- If Y_k is a Poisson(λ_k) random variable, $k = 1, 2, \ldots, m$, then $Z = \sum_{k=1}^{m} Y_k$ is a Poisson($\sum_{k=1}^{m} \lambda_k$) random variable.

In R

To find $P(X = x)$, use:	dpois(x, λ)
To find $P(X \leq x)$, use:	ppois(x, λ)
To find smallest x^* such that $P(X \leq x^*) \geq c$, use:	qpois(c, λ)
To simulate m random draws, use:	rpois(m, λ)

Example: A certain type of tree has seedlings randomly dispersed in a large area, with the mean density of seedlings being approximately five per square yard. Let the random variable X represent the number of such seedlings in 0.25 square yards. What is the probability there are at least four seedlings in 0.25 square yards?

3.3 Continuous Distributions

In this section, we'll discuss some common families of continuous distributions: 1) Uniform; 2) Gamma; 3) Normal; 4) Beta; 5) Cauchy; 6) Lognormal; 7) Double Exponential, 8) t, and 9) F. This list of continuous distributions is by no means exhaustive.

Uniform Distribution

When Used?

When a random variable takes on any value between two limits a and b with constant probability.

PDF

A random variable X has a **Uniform(a, b)** distribution if

$$f_X(x|a,b) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{else} \end{cases}$$

MGF

$$M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}, \quad t \neq 0 \quad \text{and} \quad M_X(t) = 1, \quad t = 0$$

Mean

$$E(X) = \frac{a+b}{2}$$

Variance

$$Var(X) = \frac{(b-a)^2}{12}$$

Fun Facts

- The Uniform distribution is often used for noninformative priors in Bayesian statistical modeling.

- The Uniform distribution over the unit interval, Uniform(0, 1), is a special case of the Beta distribution.

In R

To get the density, use:	`dunif(x, a, b)`
To find $P(X \leq x)$, use:	`punif(x, a, b)`
To find x^* such that $P(X \leq x^*) = c$, use:	`qunif(c, a, b)`
To simulate m random draws, use:	`runif(m, a, b)`

Gamma Distribution

When Used?

This is a large, flexible family of distributions with many uses, such as when a random variable describes the time between events or the time to an event occurring, such as an equipment failure (reliability analysis) or death (survival analysis).

PDF

A random variable X has a **Gamma(α, β)** distribution if

$$f_X(x|\alpha, \beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & 0 \le x < \infty; \quad \alpha, \beta > 0 \\ 0 & \text{else} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ is called the **Gamma function**, and has the following properties:

- $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

- For any positive integer n, $\Gamma(n) = (n-1)!$.

- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Note: The α parameter is known as the **shape parameter**, because it most influences the peakedness of the distribution, and the β parameter is called the **scale parameter**, because most of its influence is on the spread of the distribution.

MGF

$$M_X(t) = (1 - \beta t)^{-\alpha}, \quad t < \frac{1}{\beta}$$

Mean

$$E(X) = \alpha\beta$$

Variance

$$Var(X) = \alpha\beta^2$$

Fun Fact

If $X \sim \text{Gamma}(\alpha, \beta)$, where α is an integer, then for any x, $P(X \le x) = P(Y \ge \alpha)$ where $Y \sim \text{Poisson}(x/\beta)$.

Special Cases

The Chi-squared distribution and the Exponential distribution are special cases of the Gamma distribution.

Chi-squared Distribution The Chi-squared (χ^2) distribution is a special case of the Gamma distribution where $\alpha = p/2$ and $\beta = 2$, where p is a positive integer referred to as the **degrees of freedom**.

A random variable X has a $\chi^2(p)$ distribution if

$$f_X(x|p) = \begin{cases} \frac{1}{\Gamma(\frac{p}{2})2^{p/2}}x^{\frac{p}{2}-1}e^{-x/2} & 0 \leq x < \infty; \quad p \in \mathbb{Z}^+ \\ 0 & \text{else} \end{cases}$$

The MGF, mean, and variance of a χ^2 distribution are:

$$M_X(t) = (1 - 2t)^{-p/2}, \quad t < 1/2$$
$$E(X) = p$$
$$Var(X) = 2p$$

The Chi-squared distribution plays an important role in statistical inference, especially when sampling from a normal distribution (see Chapter 5).

Exponential Distribution The Exponential distribution is a special case of the Gamma distribution where $\alpha = 1$, and it is often used to model lifetimes. The exponential distribution has the **memoryless property**.

A random variable X has an **Exponential(β)** distribution if

$$f_X(x|\beta) = \begin{cases} \frac{1}{\beta}e^{-x/\beta} & 0 \leq x < \infty; \quad \beta > 0 \\ 0 & \text{else} \end{cases}$$

The MGF, mean, and variance of an Exponential(β) distribution are:

$$M_X(t) = (1 - \beta t)^{-1}, \quad t < 1/\beta$$
$$E(X) = \beta$$
$$Var(X) = \beta^2$$

Derived Distributions

Weibull Distribution If $X \sim \text{Exponential}(\beta)$, then $Y = X^{1/\gamma}$ has a Weibull(γ, β) distribution. The Weibull distribution is important in modeling failure time data, as well as fatigue and breaking strength of materials. It is also useful in survival analysis for modeling hazard functions, which give the probability that an object survives a little past time t given that the object survives to time t.

A random variable Y has a **Weibull(γ, β)** distribution if

$$f_Y(y|\gamma, \beta) = \begin{cases} \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta} & 0 \le y < \infty; \quad \gamma, \beta > 0 \\ 0 & \text{else} \end{cases}$$

Note: The Exponential distribution is a special case of the Weibull distribution when $\gamma = 1$.

Other Distributions The following distributions are also derived from transforming a Gamma random variable (see Exercise 3.24):

- If $X \sim \text{Exponential}(\beta)$, then $Y = \left(\dfrac{2X}{\beta}\right)^{1/2}$ has a **Rayleigh** distribution.

- If $X \sim \text{Gamma}(\alpha, \beta)$, then $Y = \dfrac{1}{X}$ has an **Inverted Gamma**(α, β) distribution.

- If $X \sim \text{Gamma}\left(\frac{3}{2}, \beta\right)$, then $Y = \left(\dfrac{X}{\beta}\right)^{1/2}$ has a **Maxwell** distribution.

- If $X \sim \text{Exponential}(1)$, then $Y = \alpha - \gamma \ln(X)$ has a **Gumbel**(α, γ) distribution.

In R

	Gamma(α, β)	Exp(β)	$\chi^2(p)$
To get the density, use:	dgamma(x, α, $1/\beta$)	dexp(x, $1/\beta$)	dchisq(x, p)
To find $P(X \le x)$, use:	pgamma(x, α, $1/\beta$)	pexp(x, $1/\beta$)	pchisq(x, p)
To find x^* such that $P(X \le x^*) = c$, use:	qgamma(c, α, $1/\beta$)	qexp(c, $1/\beta$)	qchisq(c, p)
To simulate m random draws, use:	rgamma(m, α, $1/\beta$)	rexp(m, $1/\beta$)	rchisq(m, p)

Notes:

- R uses the **rate parameter** $\frac{1}{\beta}$ rather than the scale parameter β for the Gamma and Exponential distribution functions. These merely make use of another parameterization of these distributions, denoted with the following notation: $X \sim \text{Gamma}^*(\alpha, \beta)$ or $X \sim \text{Exponential}^*(\beta)$.

- The Weibull distribution also has built-in R functions (see the help file for `dweibull`). However, the Weibull parameterization used in R differs from the parameterization of Weibull(γ, β) shown above: it uses the same shape parameter (γ), but its scale parameter is equal to $\sigma = \beta^{1/\gamma}$.

Normal Distribution

When Used?

Often! The normal distribution (also called the Gaussian distribution) is the most widely used continuous probability distribution, mainly because it is tractable analytically, it follows the familiar bell shape which fits with a lot of population models, and the central limit theorem says that, with a large enough sample, the normal distribution can be used to approximate a large variety of other distributions (e.g., normal approximation to the binomial distribution).

PDF

A random variable X has a **Normal(μ, σ^2)** distribution if

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \quad -\infty < x < \infty; \quad -\infty < \mu < \infty; \quad \sigma^2 > 0$$

MGF

$$M_X(t) = \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$$

Mean and Variance

$$E(X) = \mu \qquad Var(X) = \sigma^2$$

Examples

The normal distribution is an appropriate model for a variety of real-world variables such as cognitive and psychological attributes (e.g., IQ scores, standardized test scores) and various physical and biological characteristics (e.g., height, weight, blood pressure). The **Central Limit Theorem** (Theorem 5.5.14) states that the distribution of a sum of independent and identically distributed random variables will converge to a normal distribution; thus, any phenomenon that could be conceptually thought of as a sum of errors can also be modeled with a normal distribution.

Where does the normal distribution appear in regression models?

Fun Facts

- The normal distribution was published by de Moivre in 1733!
- The chi-squared, t and F distributions can be derived from the normal distribution (see Chapter 5).

- The normal distribution follows the 68-95-99.7 Rule:

- For large n and not extreme p, the distribution of $X \sim \text{Binomial}(n, p)$ can be approximated by a $\text{Normal}(np,\, np(1-p))$ distribution. This approximation can be improved by using a "continuity correction" (see pg. 105).

Special Cases

Standard Normal Distribution   The standard normal distribution is a special case of the normal distribution where $\mu = 0$ and $\sigma^2 = 1$. If $Y \sim N(\mu, \sigma^2)$, then

$$Z = \frac{Y - \mu}{\sigma} \sim N(0, 1).$$

A random variable Z has a **standard normal distribution** $(Z \sim N(0, 1))$ if

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

The MGF, mean, and variance of a $N(0, 1)$ distribution are:

$$M_Z(t) = \exp\left(\frac{t^2}{2}\right)$$
$$E(Z) = 0$$
$$Var(Z) = 1$$

All normal probabilities may be calculated in terms of the standard normal.

In R

To get the density, use:	dnorm(x, μ, σ)
To find $P(X \le x)$, use:	pnorm(x, μ, σ)
To find x^* such that $P(X \le x^*) = c$, use:	qnorm(c, μ, σ)
To simulate m random draws, use:	rnorm(m, μ, σ)

Beta Distribution

When Used?

When a random variable is defined over the interval $0 \leq x \leq 1$; typically the beta distribution is used to model proportions, which naturally lie between 0 and 1.

PDF

A random variable X has a **Beta(α, β)** distribution if

$$f_X(x|\alpha, \beta) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 \leq x \leq 1; \quad \alpha > 0, \quad \beta > 0 \\ 0 & \text{else} \end{cases}$$

The **normalizing constant** is expressed as a function of the beta function $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

MGF

Not useful.

Mean and Variance

$$E(X) = \frac{\alpha}{\alpha + \beta} \qquad\qquad Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Fun Fact

When $\alpha = \beta = 1$, the beta distribution is the uniform distribution over the unit interval, Uniform(0, 1).

In R

To get the density, use:	dbeta(x, α, β)
To find $P(X \leq x)$, use:	pbeta(x, α, β)
To find x^* such that $P(X \leq x^*) = c$, use:	qbeta(c, α, β)
To simulate m random draws, use:	rbeta(m, α, β)

How can the Beta distribution be applied to a random variable defined over the interval $a \leq y \leq b$, where $a \neq 0$ and $b \neq 1$?

Cauchy Distribution

When Used?

Surprisingly more often than one would expect. Usually this distribution arises when comparing ratios of standard normal random variables.

PDF

A random variable X has a **Cauchy(θ, σ)** distribution if

$$f_X(x|\theta, \sigma) = \frac{1}{\pi\sigma\left[1 + \left(\dfrac{x-\theta}{\sigma}\right)^2\right]} \quad -\infty < x < \infty; \quad -\infty < \theta < \infty, \quad \sigma > 0$$

MGF

Does not exist.

Mean

Does not exist, but θ is the median of the distribution.

Variance

Does not exist.

Fun Facts

- The Cauchy distribution is a symmetric, bell-shaped distribution.

- No moments of the Cauchy distribution exist.

- The ratio of two standard normal random variables has a Cauchy distribution.

- Some texts will define the one-parameter Cauchy distribution by setting $\sigma = 1$.

In R

To get the density, use:	dcauchy(x, θ, σ)
To find $P(X \leq x)$, use:	pcauchy(x, θ, σ)
To find x^* such that $P(X \leq x^*) = c$, use:	qcauchy(c, θ, σ)
To simulate m random draws, use:	rcauchy(m, θ, σ)

Lognormal Distribution

When Used?

When the variable of interest is skewed to the right, and the natural log (log transformation) of the data is normally distributed, allowing the use of normal-theory statistical procedures. Examples include income, movement data, and electrical measurements.

PDF

If $Y \sim N(\mu, \sigma^2)$, then $X = e^Y$ follows a **Lognormal(μ, σ^2)** distribution:

$$f_X(x|\mu, \sigma^2) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2}\right) & 0 < x < \infty; \quad -\infty < \mu < \infty, \quad \sigma^2 > 0 \\ 0 & \text{else} \end{cases}$$

MGF

Does not exist, but all moments exist!

Mean

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

Variance

$$Var(X) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$$

In R

To get the density, use:	`dlnorm(x, `μ`, `σ`)`
To find $P(X \leq x)$, use:	`plnorm(x, `μ`, `σ`)`
To find x^* such that $P(X \leq x^*) = c$, use:	`qlnorm(c, `μ`, `σ`)`
To simulate m random draws, use:	`rlnorm(m, `μ`, `σ`)`

Transformations in Regression

A simple linear regression model assumes that the response variable, Y, follows a normal distribution whose mean depends on some known explanatory variable x:

Visually, this model can be represented as:

What happens if our residuals (errors around a fitted regression line) do not appear to be normally distributed? If they are right-skewed, often a log-transformation is used:

Double Exponential/Laplace Distribution

When Used?

Great question! We should investigate this more.

PDF

A random variable X has a **Double Exponential(μ, σ)** distribution if

$$f_X(x|\mu, \sigma) = \frac{1}{2\sigma} \exp\left(\frac{-|x - \mu|}{\sigma}\right) \quad -\infty < x < \infty; \quad -\infty < \mu < \infty, \quad \sigma > 0$$

MGF

$$M_X(t) = \frac{\exp(\mu t)}{1 - (\sigma t)^2}, \quad |t| > \frac{1}{\sigma}$$

Mean

$$E(X) = \mu$$

Variance

$$Var(X) = 2\sigma^2$$

Fun Facts

- The double exponential distribution is also called the Laplace distribution.

- The double exponential distribution is formed by reflecting the exponential distribution around its mean.

In R

The base R library doesn't have built-in functions for the Double Exponential distribution, but there are R functions for this distribution in the `extraDistr` library:

To get the density, use: `dlaplace(x, `μ`, `σ`)`
To find $P(X \leq x)$, use: `plaplace(x, `μ`, `σ`)`
To find x^* such that $P(X \leq x^*) = c$, use: `qlaplace(c, `μ`, `σ`)`
To simulate m random draws, use: `rlaplace(m, `μ`, `σ`)`

t-Distribution

When Used?

Extensively in statistical inference! Many **test statistics** follow a t-distribution when the null hypothesis is assumed true.

PDF

A random variable T has a $t(\nu)$ distribution if

$$f_T(t|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \qquad -\infty < t < \infty; \quad \nu \in \mathbb{Z}^+$$

The parameter ν is called the **degrees of freedom**.

MGF

Does not exist.

Mean and Variance

$$E(T) = 0, \quad \nu > 1 \qquad\qquad Var(T) = \frac{\nu}{\nu - 2}, \quad \nu > 2$$

Fun Facts

- This distribution was developed by William Gosset in 1908 while working at Guinness Brewery to address the need for statistical testing with small samples where the population variance was unknown. Since he could not publish under his own name, he published it under the pseudonym "Student", and the distribution is sometimes called the "Student's t distribution."

- The t distribution is symmetric about zero, with shape similar to the standard normal (heavier tails $\implies$ "flatter").

- If $\nu = 1$, then $T \sim \text{Cauchy}(0, 1)$.

- If $T \sim t(\nu)$, then only the first $\nu - 1$ moments exist.

- As $\nu \to \infty$, the t distribution approaches the standard normal distribution.

- If $Z \sim N(0, 1)$, $V \sim \chi^2(\nu)$, and Z and V are independent, then the random variable defined as $T = \dfrac{Z}{\sqrt{V/\nu}}$ follows a t distribution with ν degrees of freedom. (You will prove this in STAT 502! Also see Definition 5.3.4.)

In R

To get the density, use:	`dt(x, `ν`)`
To find $P(X \le x)$, use:	`pt(x, `ν`)`
To find x^* such that $P(X \le x^*) = c$, use:	`qt(c, `ν`)`
To simulate m random draws, use:	`rt(t, `ν`)`

F-Distribution

When Used?

Extensively in statistical inference! Many **test statistics** follow an F-distribution when the null hypothesis is assumed true.

PDF

A random variable X has an $F(u, v)$ distribution if

$$f_X(x|u, v) = \begin{cases} \frac{\Gamma\left(\frac{u+v}{2}\right)}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)} \left(\frac{u}{v}\right)^{u/2} x^{(u-2)/2} \left(1 + \frac{u}{v}x\right)^{-(u+v)/2} & x > 0; \quad u, v \in \mathbb{X}^+ \\ 0 & \text{else} \end{cases}$$

The parameter u is referred to as the **numerator degrees of freedom**, and the parameter v as the **denominator degrees of freedom**.

MGF

Does not exist.

Mean and Variance

$$E(X) = \frac{v}{v-2}, \quad v > 2 \qquad\qquad Var(X) = 2\left(\frac{v}{v-2}\right)^2 \left(\frac{u+v-2}{u(v-4)}\right), \quad v > 4$$

Fun Facts

- This distribution was derived by statistician George Snedecor in the mid-1930s and is sometimes called the "Snedecor's F distribution." The "F" was in honor of Sir Ronald Fisher, who developed the underlying concept of a ratio of variances (used in ANOVA).

- The F distribution is right skewed, with a shape that appears more normal as the numerator and denominator degrees of freedom increase.

- If $T \sim t(\nu)$, then $T^2 \sim F(1, \nu)$.

- If $U \sim \chi^2(u)$, $V \sim \chi^2(v)$, and U and V are independent, then the random variable defined as $X = \frac{U/u}{V/v}$ follows an F distribution with u numerator degrees of freedom and v denominator degrees of freedom. (You will prove this in STAT 502! Also see Definition 5.3.6.)

In R

To get the density, use:	`df(x, u, v)`
To find $P(X \leq x)$, use:	`pf(x, u, v)`
To find x^* such that $P(X \leq x^*) = c$, use:	`qf(c, u, v)`
To simulate m random draws, use:	`rf(m, u, v)`

3.4 Exponential Families

Families of pmfs or pdfs that belong to the exponential family have properties that make statistical inference in both the frequentist and Bayesian context simpler mathematically!

DEF 3.4.1 (Exponential Family) A family of pdfs or pmfs is called an exponential family if it can be expressed as:

$$f(x|\theta) = h(x)c(\boldsymbol{\theta})exp\left\{\sum_{i=1}^{k} t_i(x)w_i(\boldsymbol{\theta})\right\},$$

where $h(x) \geq 0$ and $t_1(x), t_2(x), ..., t_k(x)$ are real-valued functions of x alone (cannot depend on $\boldsymbol{\theta}$) and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-valued parameter $\boldsymbol{\theta}$ alone (cannot depend on x).

Notes:

- To verify that a family of pdfs or pmfs is an exponential family, we must identify the functions $h(x), c(\boldsymbol{\theta}), w_i(\boldsymbol{\theta})$, and $t_i(x)$ and show that the family has the form provided above in DEF 3.4.1.

- A **curved exponential family** is a family of densities of the form provided in DEF 3.4.1 for which the dimension of the vector $\boldsymbol{\theta}$ is $d < k$. If $d = k$ the family is a **full exponential family**.

Examples Determine whether the family of pdfs/pmfs is an exponential family...

- Let $X \sim Binomial(n, p)$. Assume $0 < p < 1$.

- Let $X \sim Normal(\mu, \sigma^2)$. Assume $\mu \in (-\infty, \infty)$ and $\sigma^2 > 0$.

- Let $X \sim Pareto(\alpha, \beta)$. Then $f_X(x|\alpha, \beta) = \dfrac{\beta\alpha^\beta}{x^{\beta+1}} I_{(\alpha,\infty)}(x)$; α, $\beta > 0$.

3.5 Location and Scale Families

We've discussed common families of continuous distributions, and in this section, we investigate different techniques for constructing families of distributions that are useful for modeling. These families include location families, scale families, and location-scale families, and the general process for constructing each of these three types of families is similar: specify a standard (or reference) pdf, and then transform that pdf in a specified way.

THM 3.5.1 (Theorem about PDFs) Let $f(\cdot)$ be any pdf and let $-\infty < \mu < \infty$ and $\sigma > 0$ be any given constants. Then the function

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

is a pdf.

Proof: Show the function $g(x|\mu, \sigma)$ satisfies the properties of a pdf for all values of $-\infty < \mu < \infty$ and $\sigma > 0$ on your own. That is, show this function is non-negative for all $x \in \mathbb{R}$ and integrates to 1 (see p. 116 in text).

Where does this pdf come from? Let Z have pdf $f(z)$. What transformation of Z, $X = h(Z)$, has the pdf defined in the Theorem above?

DEF 3.5.2 (Location Family) Let $f_X(x)$ be any pdf. Then the family of pdfs $f_X(x - \mu)$, indexed by the parameter μ, $-\infty < \mu < \infty$, is called the **location family** with the standard pdf $f_X(x)$ and μ is called the location parameter for the family. That is, the standard (or reference) pdf is where $\mu = 0$.

Notes:

- The location parameter μ shifts the graph of the pdf along the x-axis while keeping the same shape.
- The pdf $f_X(x - \mu)$ is the pdf of the transformed $Y = X - \mu$.

Examples Consider the two pdfs and the graphs of their standard pdf below. What does the parameter η represent for each distribution? What happens to the pdf if η is shifted?

- Exponential distribution with $\beta = 1$ and location parameter η:

$$f_X(x|\eta) = e^{-(x-\eta)} I_{(\eta,\infty)}(x)$$

Standard pdf:

```
curve(dexp(x, rate = 1), from = 0, to = 10, lwd = 2,
  xlim = c(-5,10), ylab = expression(f[X](x)),
  main = latex2exp::TeX(
    "$f_X(x|\\eta) = e^{-(x-\\eta)}I_{(\\eta, \\infty)}(x)$"))
```


$$f_X(x|\eta) = e^{-(x-\eta)} I_{(\eta,\,\infty)}(x)$$
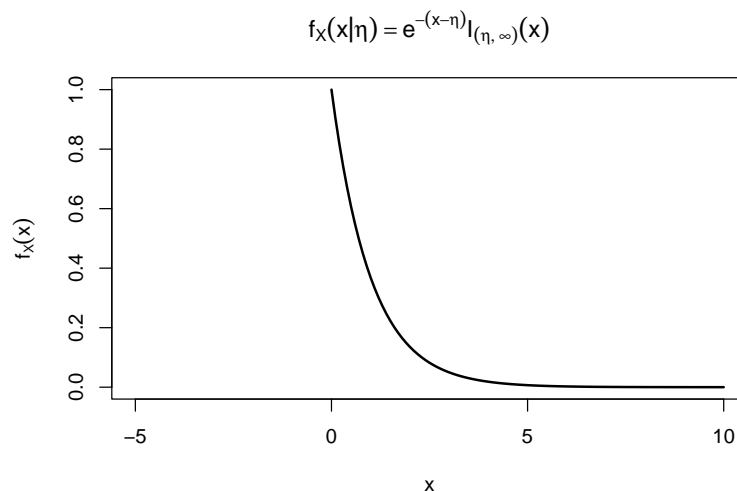
```
# Add pdfs with differing values of eta
curve(dexp(x-2, rate = 1), from = 2, to = 10,
  lty = 2, lwd = 2, add = TRUE)
curve(dexp(x+3, rate = 1), from = -3, to = 10,
  lty = 3, lwd = 2, add = TRUE)
legend("topright",
   c(expression(paste(eta," = -3")),
     expression(paste(eta," = 0")),
     expression(paste(eta," = 2"))),
   lty = c(3, 1, 2), lwd = c(1, 1, 1))
```

- Double exponential distribution with $\mu = \eta$ and $\sigma = 1$:

$$f_X(x|\eta) = \frac{1}{2}e^{-|x-\eta|}I_{(-\infty,\infty)}(x)$$

Standard pdf:

```r
curve(extraDistr::dlaplace(x), from = -5, to = 5, lwd = 2,
  ylab = expression(f[X](x)),
  main = latex2exp::TeX(
  "$f_X(x|\\eta) = \\frac{1}{2}e^{-abs(x-\\eta)}I_{(-\\infty, \\infty)}(x)$"))
```
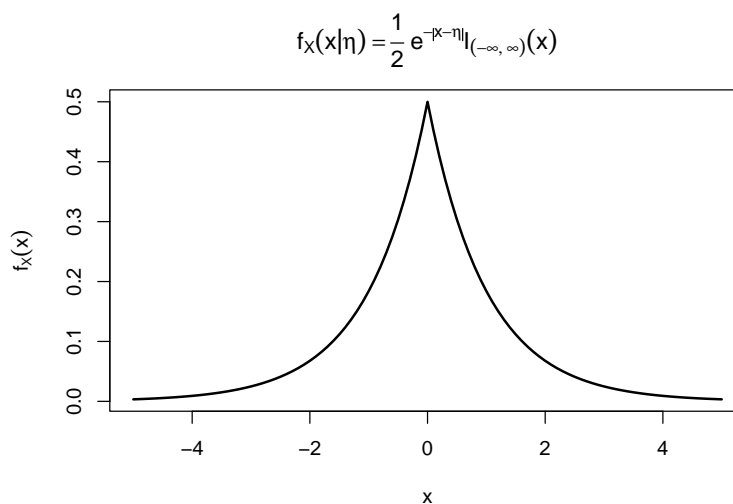


```r
# Add pdfs with differing values of eta
curve(extraDistr::dlaplace(x-2),
  lty = 2, lwd = 2, add = TRUE)
curve(extraDistr::dlaplace(x+3),
  lty = 3, lwd = 2, add = TRUE)
legend("topright",
    c(expression(paste(eta," = -3")),
      expression(paste(eta," = 0")),
      expression(paste(eta," = 2"))),
    lty = c(3, 1, 2), lwd = c(1, 1, 1))
```

What are some commonly used location families?

> **DEF 3.5.4 (Scale Family)** Let $f_X(x)$ be any pdf. Then the family of pdfs $\dfrac{1}{\sigma} f_X\left(\dfrac{x}{\sigma}\right)$, indexed by the parameter σ, $\sigma > 0$, is called the **scale family** with the standard pdf $f_X(x)$ and σ is called the scale parameter for the family. That is, the standard (or reference) pdf is where $\sigma = 1$.

Notes:

- The scale parameter σ either stretches ($\sigma > 1$) or contracts ($\sigma < 1$) the graph of $f_X(x)$, while keeping the same basic shape and the center the same.
- The pdf $\frac{1}{\sigma} f_X\left(\frac{x}{\sigma}\right)$ is the pdf of the transformed $Y = \sigma X$.

$$f_X(x|\beta) = \frac{1}{\beta} e^{-x/\beta} I_{(0,\,\infty)}(x) \qquad\qquad f_X(x|\beta) = \frac{1}{2\beta} e^{-|x|/\beta} I_{(-\infty,\,\infty)}(x)$$
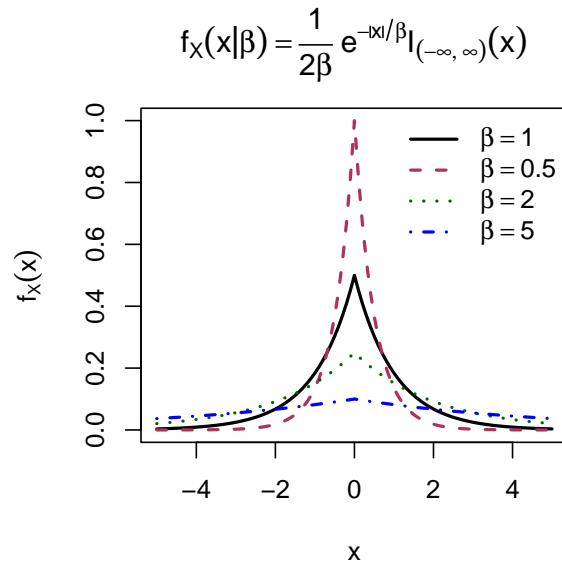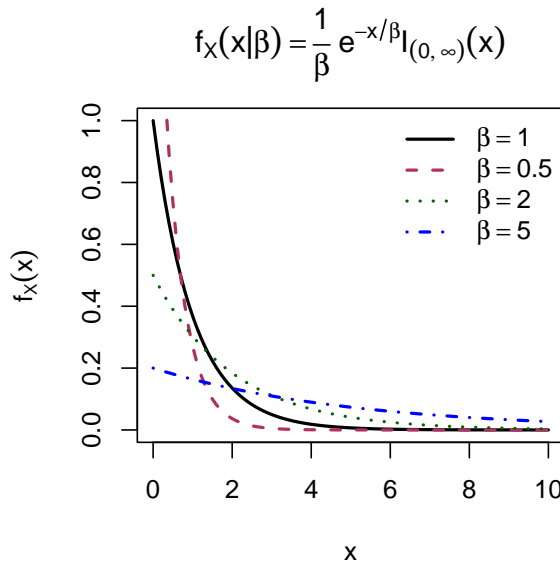
What are some examples of commonly used scale families?

DEF 3.5.5 (Location-Scale Family) Let $f_X(x)$ be any pdf. Then the family of pdfs $\frac{1}{\sigma} f_X\left(\frac{x-\mu}{\sigma}\right)$, indexed by the parameters μ and σ, where $-\infty < \mu < \infty$ and $\sigma > 0$, is called the **location-scale family** with the standard pdf $f_X(x)$; μ is called the location parameter and σ is called the scale parameter for the family. That is, the standard (or reference) pdf is where $\mu = 0$ and $\sigma = 1$.

What are some examples of commonly used location-scale families?

THM 3.5.6 (Generating location, scale, and location-scale families) Let $f(\cdot)$ be any pdf. Let μ be any real number, and let σ be any positive real number. Then, X is a random variable with pdf $f_X(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.

Proof: On your own (see p. 120 in text)

Take home message: We can start with *any* pdf $f_X(x)$ and generate a family of pdfs by introducing a location and/or scale parameter in the manner described above! Essentially, we can generate location, scale, and location-scale families by subjecting a random variable with standard pdf to a family of transformations.

3.6 Inequalities and Identities

One of the most famous inequalities is introduced at the end of this chapter, Chebychev's Inequality. We will cover several other useful inequalities involving more than one random variable at the end of Chapter 4, including the Cauchy-Schwarz Inequality and Jensen's Inequality.

THM 3.6.1 (Chebychev's Inequality) Let X be a random variable and let $g(x)$ be a non-negative function. Then for any $r > 0$,

$$P(g(X) \geq r) \leq \frac{E[g(X)]}{r}$$

Proof:

Example A special case of Chebychev's Inequality involves the following non-negative function:
$g(x) = \dfrac{(x-\mu)^2}{\sigma^2}$, where $\mu = E(X)$ and $\sigma^2 = Var(X)$.

Note: Chebychev's Theorem (particularly in its classic expression shown above) enables us to find bounds for probabilities that may be difficult or tedious to obtain. Often, we can also use it to obtain means and variances of random variables without specifying the distribution of the variable.

Example Experience has shown that the length of time X (in minutes) required to conduct a periodic maintenence check on a machine follows a gamma distribution with $\alpha = 3.1$ and $\beta = 2$. A new maintenance worker takes 22.5 minutes to check the machine. Does this length of time to perform the check seem unusually long?

Stein's Lemma (Lemma 3.6.5) Let $X \sim N(\mu, \sigma^2)$ and let g be a differentiable function satisfying $E[|g'(X)|] < \infty$. Then,
$$E[g(X)(X - \mu)] = \sigma^2 E[g'(X)].$$

- Note: This lemma is most useful for calculating higher-order moments.
- Proof: On your own (see p. 124 in text)

Example: Let $X \sim N(\mu, \sigma^2)$. Find $E(X^3)$.