Review Handout 2: Bernoulli, Normal and t- Distributions

Edward Vytlacil, Yale University

June 24, 2025

This handout reviews the Bernoulli distribution, and connects it to indicator/dummy variables. It also reviews the normal distribution, the distribution of the absolute value of a normal random variable, and the the t-distribution. We will use these distributions extensively for inference.

Contents

Bernoulli Distribution

Example 1 of Handout 1 considered the random variable for a weighted coin toss, with $X = 1$ if heads and $X = 0$ if tails, and $\Pr[X = 1] = p$, $\Pr[X = 0] = 1 - p$. Such a random variable is called a Bernoulli random variable, and its distribution is called the Bernoulli distribution:

Definition 1: Bernoulli Random Variable

A random variable X is distributed *Bernoulli(p)*, $X \sim \text{Bernoulli}(p)$, if:

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p. \end{cases}$$

For a Bernoulli random variable, we call $X = 1$ a "*success*" and $X = 0$ a "*failure*".

The terminology of "success" and "failure" need not correspond to good and bad outcomes. For example, $X = 1$ might indicate that a patient dies and $X = 0$ that the patient lives, without implying that death is a good outcome. Likewise, a machine failing or a dam failing might be a "success" despite the tension in the terminology.

While we motivated the Bernoulli distribution using a coin toss, it is the distribution for any *binary variable* taking the values 0 and 1, and is how we encode as a variable yes-no and true-false answers. It is often productive in probability theory (as in economics, as in coding) to abstract from details. Thus, instead of trying to separately consider random variables for whether an individual is employed or not, or a person tests positive for a virus or not, or a machine fails or not, or a dam bursts or not, we instead study Bernoulli random variables and think of hypothetical experiments involving weighted coin flips with our abstract model of weighted coin flips describing any yes-no, true-false type variable, just with different probabilities of "success" or "failure."

In Handout 1, Example 1, we already derived the expectation and variance of a Bernoulli random variable, which we restate here in the following theorem:

Theorem 1: Expected Value and Variance of a Bernoulli r.v.

Suppose $X \sim \text{Bernoulli}(p)$. Then

$$\mathbb{E}[X] = p,$$
$$\text{Var}[X] = p \cdot (1 - p).$$

Logical Indicator Function and Indicator Variables

WE WILL OFTEN refer to variables representing whether some event is true or not as indicator variables, where the variable equals 1 if the event is true, and equals 0 otherwise. Indicator variables are also called dummy variables. We will often construct indicator variables using logical indicator functions, defined as follows:

Definition 2: Logical Indicator Function

Let $\mathbb{1}$ denote the *logical indicator function*, defined for any statement (event) A as

$$\mathbb{1}[A] = \begin{cases} 1 & \text{if } A \text{ is true,} \\ 0 & \text{if } A \text{ is false.} \end{cases}$$

We define indicator variables using logical indicator functions:

Definition 3: Indicator Variables

X is an *indicator variable*, also called a *dummy variable*, if, for some given event A, $X = \mathbb{1}[A]$.

If X is an indicator variable for some event A, $X = \mathbb{1}[A]$, then X equals 1 when A is true and equals 0 when A is false. Thus, the indicator variable $X = \mathbb{1}[A]$ is a Bernoulli(p) random variable with $p = \Pr[A]$. For example, we might define $X = \mathbb{1}[\text{employed}]$ as a indicator variable for the event that an individual is employed, which is distributed Bernoulli(p) where p equals probability of being employed. From Theorem 1, we see that the expected value of a indicator variable is the probability that the event occurs, and that the variance of an indicator variable is relatively large when the probability of the event is close to a half and is small when the probability of the event is close to 0 or 1.

We will also often use logical indicator functions to construct indicator variables from some other non-binary random variable, as illustrated by the following examples.

Example 1 (Positive Returns). *Let r_A denote the return on an asset, and define X by $X = \mathbb{1}[r_A > 0]$. Then X is a indicator variable for the asset having a positive return and $\mathbb{E}[X] = \Pr[r_A > 0]$.*

Example 2 (Schooling). *Let S denote an individual's years of schooling. Then $X = \mathbb{1}[S \geq 12]$ is a indicator variable for having completed high school, and $\mathbb{E}[X] = \Pr[S \geq 12]$.*

Dummy variables are pervasive in econometrics, representing, for example, an individual working or not working, having graduated from college or not, being retired or not, being married or not, having children or not, having positive savings or not, having investments in the stock market or not, and so forth.

Normal Distribution

We now consider the normal distribution, which is different in that we rarely believe that random variables in practice are exactly normally distributed. However, we often believe that random variables are approximately normally distributed, and normal approximations plays a critical role in large samples due to the Central Limit Theorem.

Definition 4: Normal Distribution

A random variable X is *Normally Distributed* with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$, if the probability density function of X is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Particularly important for hypothesis testing will be the normal distribution with $\mu = 0$ and $\sigma^2 = 1$, called the standard normal distribution.

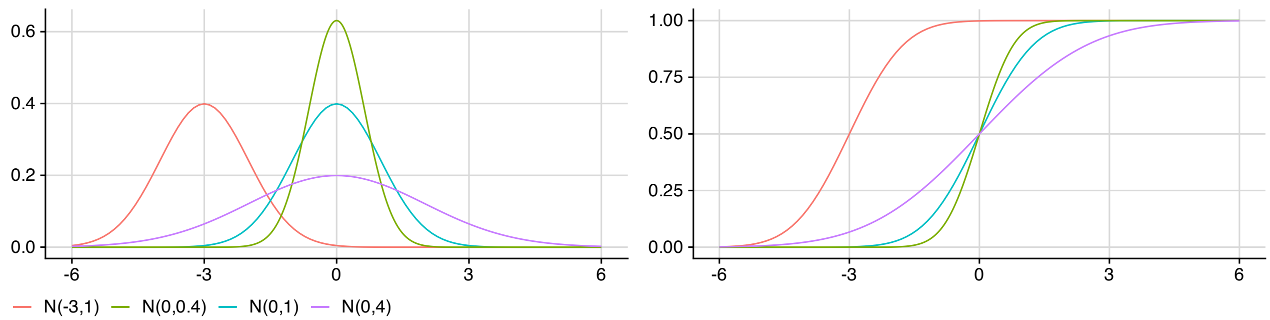


Figure 1: Plotting PDF (left) and CDF (right) of $N(\mu, \sigma^2)$. Note that μ determines where the density is centered and the density is symmetric around μ, while σ^2 determines the spread of the density. Note that — $N(0, 1)$ is standard normal.

Definition 5: Standard Normal Distribution

A random variable X is distributed *Standard Normal* if $X \sim N(0, 1)$. We denote the pdf of a standard normal distribution by $\phi(x)$, so that

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

and the CDF of a standard normal by $\Phi(x)$, so that

$$\Phi(t) = \Pr[X \leq t] = \int_{-\infty}^{t} \phi(x) dx.$$

Remark 1 (Symmetry of Std. Normal). *A $N(\mu, \sigma^2)$ density is symmetric around μ, so that the standard normal pdf, $\phi(\cdot)$, is symmetric around 0 . The symmetry of $\phi(\cdot)$ around 0 has the following implications, which will be useful for inference:*

1. *$\phi(t) = \phi(-t)$ for all t;*
2. *$\Phi(t) = 1 - \Phi(-t)$ for all t;*
3. *$X \sim N(0, 1)$ implies $-X \sim N(0, 1)$.*
4. *$q_\alpha = -q_{1-\alpha}$ for all α, where q_α is the α quantile of a standard normal.*

Quantiles of $N(0, 1)$

α	q_α
0.01	-2.33
0.025	-1.96
0.05	-1.64
0.95	1.64
0.975	1.96
0.99	2.33

We will often work with linear functions of normal random variables.

Theorem 2: Linear Function of Normal

Suppose $X \sim N(\mu, \sigma^2)$. Let a and b denote constants with $b \neq 0$. Then $a + bX \sim N(a + b\mu, b^2\sigma^2)$.

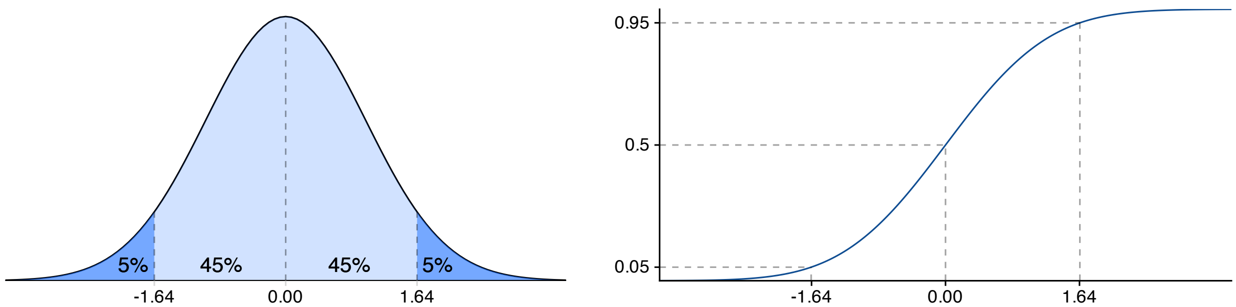


Figure 2: Standard Normal PDF (left) and CDF (right).

Remark 2 (Converting $N(\mu, \sigma^2)$ r.v. to Std. Normal)**.** *Suppose* $X \sim N(\mu, \sigma^2)$. *Then, by Theorem 2*

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

Converting a $N(\mu, \sigma^2)$ *r.v. to a standard normal r.v. allows us to use* Φ *to find* $\Pr[X \leq t]$ *for* $X \sim N(\mu, \sigma^2)$. *In particular,*

$$\begin{aligned} \Pr[X \leq t] =& \Pr\left[\frac{X - \mu}{\sigma} \leq \frac{t - \mu}{\sigma}\right] \\ =& \Phi\left(\frac{t - \mu}{\sigma}\right). \end{aligned}$$

For example, if $X \sim N(1, 4)$, *then* $\Pr[X \leq 2] = \Phi(\frac{2-1}{2}) = \Phi(\frac{1}{2})$.

Consider the sum of independent normal random variables.

Theorem 3: Sum of Independent Normal r.v.'s

Suppose that $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ and X and Y are independent. Let a and b denote constants with $a \neq 0$ or $b \neq 0$. Then

$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2).$$

Iterating on Theorem 3 for i.i.d. random variables leads to the following corollary:

Corollary 4: Mean of i.i.d. Normal Random Variables

Suppose $X_1, X_2, ..., X_N$ are i.i.d. with $X_i \sim N(\mu, \sigma^2)$. Then

$$\bar{X}_N \sim N(\mu, \sigma^2/N).$$

Following Remark 2 and applying Corollary 4, we have that, if $X_1, X_2, ..., X_N$ are i.i.d. with $X_i \sim N(\mu, \sigma^2)$, then

$$\frac{\bar{X}_N - \mu}{\sigma / \sqrt{N}} \sim N(0,1),$$

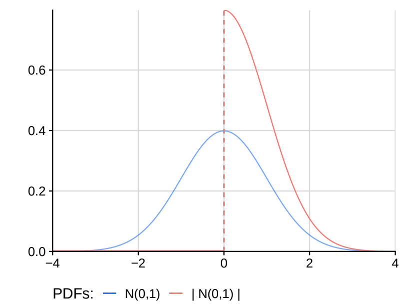
so that

$$\Pr[\bar{X}_N \leq t] = \Phi\left(\frac{t-\mu}{\sigma/\sqrt{N}}\right).$$

Remark 3 (Absolute Value of Std. Normal)**.** *The distribution of the absolute value of a standard normal r.v. is important for inference. Suppose $X \sim N(0,1)$ so that $|X| \sim |N(0,1)|$. Then, using the symmetry of $\phi(\cdot)$,*

- *for $x \geq 0$, the density of $|X|$ at x is $2 \cdot \phi(x)$;*
- *for $x \geq 0$, $\Pr[\ |X| > x\ ] = 2 \cdot \Pr[X > x] = 2\ (1 - \Phi(x))$;*
- *the $1-\alpha$ quantile of $|X|$ equals the $1-\alpha/2$ quantile of X.*

For example, if $X \sim N(0,1)$, then $\Pr[X > 1.64] = 1 - \Phi(1.64) = 0.05$, $\Pr[|X| > 1.64] = 2 \cdot (1 - \Phi(1.64)) = 0.10$, and thus 1.64 is the 0.95 quantile of $X \sim N(0,1)$ which is the 0.90 quantile of $|X|$.



$|N(0,\sigma^2)|$ is called the *half-normal* distribution, so that $|N(0,1)|$ is an example of a half-normal distribution .

*Normal Distribution in **R***

WE WILL TYPICALLY use **R** to calculate probabilities for the normal distribution.

R Functions for Normal Distribution

Function	Returns
`dnorm(x, m, s)`	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-m}{s})^2}$, the $N(m,s^2)$ density evaluated at x,
`pnorm(x, m, s)`	$\Pr[X \leq x]$ for $X \sim N(m,s^2)$.
`qnorm(q, m, s)`	qth quantile of $N(m,s^2)$,
`rnorm(n, m, p)`	n random draw of X, $X \sim N(m,s^2)$.

These functions set $m = 0$ and $s = 1$ by default if their values are not specified. Thus, `dnorm(0.5)` returns the same value as `dnorm(0.5, 0, 1)` .

```
> # 0.05 Quantile of N(0,1)
> qnorm(0.05,0,1)
[1] -1.644854
> # Std. Norm Density at -1.645
> dnorm(-1.6449,0,1)
[1] 0.1031278
> # Prob Std. Norm. Less than -1.645
> pnorm(-1.645,0,1)
[1] 0.04998491
> # Prob Std. Norm. Greater than -1.645
> 1-pnorm(-1.645,0,1)
[1] 0.9500151
> # Default mu=0, sigma=1
> pnorm(-1.645)
[1] 0.04998491
```

```
> # .95 Quantile of N(0,1)
> qnorm(0.95,0,1)
[1] 1.644854
> # Std. Norm Density at 1.645
> dnorm(1.6449,0,1)
[1] 0.1031278
> # Prob Std. Norm. Less than 1.645
> pnorm(1.645,0,1)
[1] 0.9500151
> # Prob Std. Norm. Greater than 1.645
> 1-pnorm(1.645,0,1)
[1] 0.04998491
> # Prob -1.64<Std. Norm.< 1.64
> pnorm(1.645)-pnorm(-1.645)
[1] 0.9000302
```

```
> # Simulating 1 draw from N(0,1)
> rnorm(1,0,1)
[1] 1.207962
> # Another draw, using defaults
> rnorm(1)
[1] -0.4028848
 > # Simulating 3 draws from N(0,1)
> rnorm(3,0,1)
[1]  0.55391765 -0.06191171 -0.30596266
 > # Another 3 draws  from N(0,1)
> rnorm(3)
[1] -0.46665535  0.77996512 -0.08336907
```

```
> # Consider X ~ N(1,4)
> # Pr[X<=2]
> pnorm(2,1,2)
[1] 0.6914625
> # Using Remark 3
> # Pr[X<=2]
> pnorm(0.5,0,1)
[1] 0.6914625
> # Simulating 3 draws from N(1,4)
> # Using Theorem 3
> 1 + 2 * rnorm(3)
[1]  0.7841654 -1.3307927  5.5379119
```

t-Distribution

A FAMILY OF DISTRIBUTIONS CLOSELY RELATED to the standard normal distribution is the t-distribution, which will play a key role in inference.

Definition 6: t-distribution

A random variable X has a *t-distribution* with ν degrees of freedom, written $X \sim t_\nu$, if its probability density function is given by

$$f_X(x) = \frac{\Gamma(\nu+1)/2}{\Gamma(\nu/2)\sqrt{\nu\pi}}(1+\frac{x^2}{\nu})^{-(\nu+1)/2}, \qquad (1)$$

where ν is a positive integer and Γ is the gamma function.

In this course, you do not need to remember the formula for the normal or t-densities, and we will never work with them directly but rather use **R** when we need to evaluate them.

The t-distribution is a family of distributions indexed by the parameter ν, called the degrees of freedom. Like the standard normal distribution, the t-distribution is symmetric around zero and bell-shaped. However, the tails of the t-distribution are heavier than those of the normal distribution, with how much heavier depending on the parameter ν. When $\nu = 1$, the distribution is called the *cauchy distribution* and has much heavier tails and very different properties than a standard normal distribution. The larger is ν, the thinner the tails, and the closer the t_ν distribution is to a standard normal distribution. As ν goes to infinity, the t_ν distribution approaches t_∞, which is the $N(0,1)$ distribution.

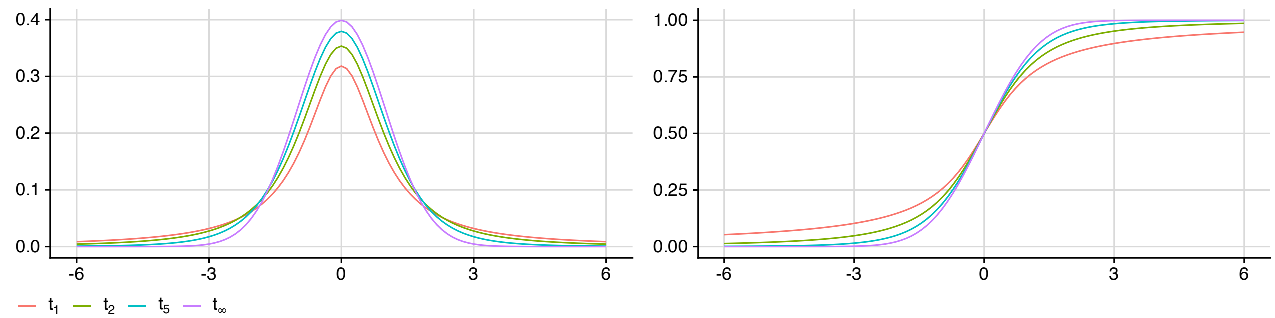


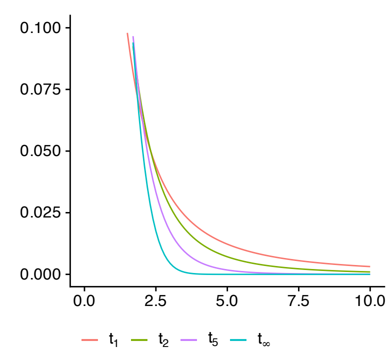
Figure 3: Plotting the t_ν PDF (above left), CDF (above right), and right tail of PDF (below). Note that — t_1 is the Cauchy density, and — t_∞ is the $N(0,1)$ density.

The heavier tails of a t-distribution results in extreme values being more likely for a t-distribution than for a standard normal distribution, especially for ν small. In the extreme case of $\nu = 1$, i.e., a

Examples of t_ν distribution

ν	$\mathbb{E}[X]$	$\text{Var}(X)$	$\Pr[\|X\| > 3]$	$\Pr[\|X\| > 5]$	name
1	Does not exist	Does not exist	0.205	0.126	Cauchy
2	0	Does not exist	0.095	0.038	
3	0	3	0.058	0.015	
4	0	2	0.040	0.007	
5	0	$1\frac{2}{3}$	0.030	0.004	
∞	0	1	0.003	0.000	Std. Normal

cauchy distribution, the tails are so heavy that that $\mathbb{E}[X]$ and $\text{Var}(X)$ do not exist. For a t_2 distribution, $\mathbb{E}[X]$ does exists and equals 0, but $\text{Var}(X)$ does not exist. For a t_ν distribution with $\nu \geq 3$, then $\mathbb{E}[X] = 0$ and $\text{Var}(X) = \frac{\nu}{\nu-2}$. For a t_∞ distribution, i.e., a $N(0,1)$ distribution, $\mathbb{E}[X] = 0$ and $\text{Var}(X) = 1$.



A normal distribution can take any value on the whole real line, and thus can take extreme values. However, the tails of a normal distribution go to zero so quickly (i.e., are so thin) that one can essentially ignore the possibility of extreme values with a normal distribution. For example, if $X \sim N(0,1)$, then $\Pr[|X| > 5] = 0.0000006$. In contrast, for a t-distribution with small ν, one cannot ignore the extreme values. For example, if $X \sim t_1$ (Cauchy), $\Pr[|X| > 5] = 0.126$ and X takes extreme values so often that $\mathbb{E}[X]$ and $\text{Var}(X)$ don't exist. Researchers sometimes use the t-distribution, especially Cauchy, when modeling variables where it is important to account for the variables taking extreme values. It is often used in physics, but is also sometimes used to model asset returns and study financial risk in the context of returns that take extreme values with too high of a probability to be ignored.

However, the most common use of a t-distribution in statistics is to model the distribution of the *studentized* mean (also called the *t-statistic* or *t-ratio*) when sampling from a normal distribution. Suppose $X_1, X_2, ..., X_n$ are i.i.d. with $X_i \sim N(\mu, \sigma^2)$. Then it follows from Corollary 4 that $\bar{X}_n \sim N(\mu, \sigma^2/n)$ so that $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$, where it is important to note that the expressions are dividing by the true, population σ which is generally unknown. Suppose we had a value of μ that we hypothesized as the true value, but don't know the value of σ^2, so that we need to estimate it. Define the studentized mean as

$$T_n = \frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \tag{2}$$

where

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2. \tag{3}$$

Then $T_n \sim t_{n-1}$, as stated in the following theorem.

Theorem 5: Studentized Mean of i.i.d. Normal r.v.s

Suppose $X_1, X_2, ..., X_N$ are i.i.d. with $X_i \sim N(\mu, \sigma^2)$. Let T_n and s_n^2 be defined by equations (2) and (3). Then $T_n \sim t_{n-1}$.

An implication of Theorem 5 is that when n is small, the Studentized statistic $\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}$ will be far from normal and will take extreme values relatively frequently because we are dividing by s_n instead of σ. On the other hand, when n is large, the distribution of $\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}}$ will be close to normal and that we are substituting s_n for σ will be unimportant.

*Student-t Distribution in **R***

We will typically use **R** to calculate probabilities for the Student-t distribution.

R Functions for Student-t Distribution

Function	Returns
`dt(x, v)`	the PDF of a t_v density evaluated at x,
`pt(x, v)`	$\Pr[X \leq x]$ for $X \sim t_v$.
`qnorm(q, v)`	qth quantile of t_v,
`rt(n, v)`	n random draw of X, $X \sim t_v$.

```
> # 0.05 Quantile of Cauchy
> qt(0.05,1)
[1] -6.313752
> # 0.05 Quantile of t_5
> qt(0.05,5)
[1] -2.015048
> # 0.05 Quantile of t_infinity
> qnorm(0.05)
[1] -1.644854
> # density of Cauchy at -5
> dt(-5,1)
[1] 0.01224269
> # density of t_5 at -5
> dt(-5,5)
[1] 0.001757438
> # density ot t_infinity at -5
> dnorm(-5)
[1] 0.00000148672
```

```
> # Prob t_1 between -5 and 5
> pt(5,1)-pt(-5,1)
[1] 0.8743341
> # Prob  | t_1 | >5
> 1 - (pt(5,1)-pt(-5,1))
[1] 0.1256659
> # Prob t_5 between -5 and 5
> pt(5,5)-pt(-5,5)
[1] 0.9958953>
# Prob  | t_5 | >5
> 1 - (pt(5,5)-pt(-5,5))
[1] 0.004104716
> # Prob t_infinity between -5 and 5
> pnorm(5)-pnorm(-5)
[1] 0.9999994
> # Prob  | t_infinity | >5  3
> 1 - (pnorm(5)-pnorm(-5))
[1] 0.0000005733031
```

```
> # Simulating 6 draw from Cauchy
 > rt(6,1)
[1] -0.1055119 -3.5727334 -0.1666175
    14.4183067  2.8204201 -0.1955553
> # Another 6 draws from Cauchy
> rt(6,1)
  [1] -0.4865291 -0.4264724  1.2611932
       1.0142861 -0.6185245 -0.4247870
 > # Simulating 6 draws from t_5
> rt(3,5)
[1] -0.7905094 -0.2588564 -0.3107277
     1.3989390  0.7775500 -0.4130591
```

```
> # another 6 draws from t_5
> rt(3,5)
[1]   -2.0170308  1.0184921  1.7427500
     0.3218333  1.1276689 -0.1284361
> # Simulating 6 draws  from t_infinity
> rnorm(6)
[1] -0.6989059  0.4868481 -0.8626596
     1.5333999  0.4064118  0.1865867
> # Another 6 draws  from t_infinity
> rnorm(6)
[1] -1.5115780  2.0045310 -1.7778798
    -0.8635079 -0.1826787  0.2622432
```

Note that many draws from a Cauchy look very similar to draws from t_5 or from $N(0,1)$, though once in a while the Cauchy takes extreme values.

Summary

Important Definitions

Def 1: $X \sim$ **Bernoulli(p)** if $\Pr[X = 1] = p$, $\Pr[X = 0] = 1 - p$. We call $X = 1$ a "success" and $X = 0$ a "failure".

Def 2: $\mathbb{1}$ denotes the **logical indicator** function where, for any event A, $\mathbb{1}[A] = 1$ if A is true and $\mathbb{1}[A] = 0$ if A is false.

Def 3: X is an **indicator variable**, also called a **dummy variable**, if, for some given event A, $X = \mathbb{1}[A]$.

Def 4: X is **Normally Distributed** with mean μ and variance σ^2, written $\boldsymbol{X \sim N(\mu, \sigma^2)}$, if the pdf of X is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Def 5: X is distributed **Standard Normal** if $X \sim N(0, 1)$. We denote the standard normal pdf by. $\boldsymbol{\phi}(\cdot)$ and the standard normal CDF by $\boldsymbol{\Phi}(\cdot)$.

Def 6: X is distributed according to a **t-distribution** with ν degrees of freedom, written $X \sim t_\nu$, if it's pdf is given by (1).

- When $\nu = 1$, the distribution is called the **Cauchy** distribution.
- As $\nu \to \infty$, the distribution approaches t_∞, the standard normal distribution.

Important Results

Thm 1: If $X \sim$ Bernoulli(p) then $\mathbb{E}[X] = p$, $\text{Var}(X) = p \cdot (1 - p)$.

Thm 2: If $X \sim N(\mu, \sigma^2)$, and a and b denote constants with $b \neq 0$, then $a + bX \sim N(a + b\mu, b^2\sigma^2)$.

Thm 3: If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, X and Y are independent, and a and b denote constants with $a \neq 0$ or $b \neq 0$, then $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$.

Cor 4: Suppose $X_1, X_2, ..., X_N$ are i.i.d. with $X_i \sim N(\mu, \sigma^2)$. Then $\bar{X}_N \sim N(\mu, \sigma^2/N)$ so that $\frac{\bar{X}_N - \mu}{\sigma/\sqrt{N}} \sim N(0, 1)$.

Thm 5: Suppose $X_1, X_2, ..., X_N$ are i.i.d. with $X_i \sim N(\mu, \sigma^2)$. Then $\frac{\bar{X}_N - \mu}{s_n/\sqrt{N}} \sim t_{n-1}$, where $s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$.