Precept 4: Hypothesis Testing

Soc 500: Applied Social Statistics

Ian Lundberg

Princeton University

October 6, 2016

Learning Objectives

1. Introduce vectorized R code
2. Review homework and talk about RMarkdown
3. Review conceptual ideas of hypothesis testing
4. Practice for next homework
5. If time allows...random variables!

Expected value, bias, consistency (relevant to homework)

Suppose we flip a fair coin n times and estimate the proportion of heads as

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

Derive the expected value of this estimator

$$\begin{aligned}
E(\hat{\pi}) &= E\left(\frac{H_1 + H_2 + ... + H_n + 1}{n}\right) \\
&= \frac{1}{n}E(H_1 + H_2 + ... + H_n + 1) \\
&= \frac{1}{n}(E[H_1] + E[H_2] + ... + E[H_n] + E[1]) \\
&= \frac{1}{n}(0.5 + 0.5 + ... + 0.5 + 1) \\
&= \frac{1}{n}(n(0.5) + 1) \\
&= 0.5 + \frac{1}{n}
\end{aligned}$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

$$E(\hat{\pi}) = 0.5 + \frac{1}{n}$$

Is this estimator biased? What is the bias?

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

$$E(\hat{\pi}) = 0.5 + \frac{1}{n}$$

Is this estimator biased? What is the bias?

Bias = Expected value - Truth

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

$$E(\hat{\pi}) = 0.5 + \frac{1}{n}$$

Is this estimator biased? What is the bias?

Bias = Expected value - Truth

$$= E(\hat{\pi}) - \pi$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

$$E(\hat{\pi}) = 0.5 + \frac{1}{n}$$

Is this estimator biased? What is the bias?

$$\text{Bias} = \text{Expected value - Truth}$$

$$= E(\hat{\pi}) - \pi$$

$$= \left(0.5 + \frac{1}{n}\right) - 0.5 = \frac{1}{n}$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

Derive the variance of this estimator. Remember, $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$ and $Var(aX) = a^2 Var(X)$!

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

Derive the variance of this estimator. Remember, $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$ and $Var(aX) = a^2 Var(X)$!

$$V(\hat{\pi}) = V\left(\frac{H_1 + H_2 + ... + H_n + 1}{n}\right)$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

Derive the variance of this estimator. Remember, $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$ and $Var(aX) = a^2 Var(X)$!

$$V(\hat{\pi}) = V\left(\frac{H_1 + H_2 + ... + H_n + 1}{n}\right)$$

$$= \frac{1}{n^2}[V(H_1 + H_2 + ... + H_n + 1)]$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

Derive the variance of this estimator. Remember, $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$ and $Var(aX) = a^2 Var(X)$!

$$V(\hat{\pi}) = V\left(\frac{H_1 + H_2 + ... + H_n + 1}{n}\right)$$

$$= \frac{1}{n^2}[V(H_1 + H_2 + ... + H_n + 1)]$$

$$= \frac{1}{n^2}[V(H_1) + V(H_2) + ... + V(H_n) + V(1)] \text{ (since independent)}$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

Derive the variance of this estimator. Remember, $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$ and $Var(aX) = a^2 Var(X)$!

$$V(\hat{\pi}) = V\left(\frac{H_1 + H_2 + ... + H_n + 1}{n}\right)$$

$$= \frac{1}{n^2}[V(H_1 + H_2 + ... + H_n + 1)]$$

$$= \frac{1}{n^2}[V(H_1) + V(H_2) + ... + V(H_n) + V(1)] \text{ (since independent)}$$

$$= \frac{1}{n^2}[.5(.5) + .5(.5) + ... + .5(.5)] \text{ (since variance of Bernoulli is } p(1-p))$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

Derive the variance of this estimator. Remember, $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$ and $Var(aX) = a^2 Var(X)$!

$$V(\hat{\pi}) = V\left(\frac{H_1 + H_2 + ... + H_n + 1}{n}\right)$$

$$= \frac{1}{n^2}[V(H_1 + H_2 + ... + H_n + 1)]$$

$$= \frac{1}{n^2}[V(H_1) + V(H_2) + ... + V(H_n) + V(1)] \text{ (since independent)}$$

$$= \frac{1}{n^2}[.5(.5) + .5(.5) + ... + .5(.5)] \text{ (since variance of Bernoulli is } p(1-p))$$

$$= \frac{1}{n^2}[n(.5).5] = \frac{.25}{n}$$

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

$$E(\hat{\pi}) = 0.5 + \frac{1}{n}$$

$$V(\hat{\pi}) = \frac{.25}{n}$$

Is this estimator consistent?

Expected value, bias, consistency (relevant to homework)

$$\hat{\pi} = \frac{H_1 + H_2 + ... + H_n + 1}{n}$$

$$E(\hat{\pi}) = 0.5 + \frac{1}{n}$$

$$V(\hat{\pi}) = \frac{.25}{n}$$

Is this estimator consistent? Yes. In large samples $E(\hat{\pi}) \to \pi$ and $V(\hat{\pi}) \to 0$, so we will get **arbitrarily close to the truth** as the sample size grows.

Hypothesis Testing: Setup

Copied from Brandon Stewart's lecture slides

Goal: test a **hypothesis** about the value of a parameter.

Statistical **decision theory** underlies such hypothesis testing.

Hypothesis Testing: Setup

Copied from Brandon Stewart's lecture slides

Goal: test a **hypothesis** about the value of a parameter.

Statistical **decision theory** underlies such hypothesis testing.

Trial Example:

Suppose we must decide whether to convict or acquit a defendant based on evidence presented at a trial. There are four possible outcomes:

		Defendant	
		Guilty	Innocent
Decision	Convict		
	Acquit		

Hypothesis Testing: Setup

Copied from Brandon Stewart's lecture slides

Goal: test a **hypothesis** about the value of a parameter.

Statistical **decision theory** underlies such hypothesis testing.

Trial Example:

Suppose we must decide whether to convict or acquit a defendant based on evidence presented at a trial. There are four possible outcomes:

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	
	Acquit		Correct

We could make two types of errors:

Hypothesis Testing: Setup

Copied from Brandon Stewart's lecture slides

Goal: test a **hypothesis** about the value of a parameter.

Statistical **decision theory** underlies such hypothesis testing.

Trial Example:

Suppose we must decide whether to convict or acquit a defendant based on evidence presented at a trial. There are four possible outcomes:

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	Type I Error
	Acquit		Correct

We could make two types of errors:

- Convict an innocent defendant (**type-I error**)

Hypothesis Testing: Setup

Copied from Brandon Stewart's lecture slides

Goal: test a **hypothesis** about the value of a parameter.

Statistical **decision theory** underlies such hypothesis testing.

Trial Example:

Suppose we must decide whether to convict or acquit a defendant based on evidence presented at a trial. There are four possible outcomes:

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	Type I Error
	Acquit	Type II Error	Correct

We could make two types of errors:

- Convict an innocent defendant (**type-I error**)
- Acquit a guilty defendant (**type-II error**)

Hypothesis Testing: Setup

Copied from Brandon Stewart's lecture slides

Goal: test a **hypothesis** about the value of a parameter.

Statistical **decision theory** underlies such hypothesis testing.

Trial Example:

Suppose we must decide whether to convict or acquit a defendant based on evidence presented at a trial. There are four possible outcomes:

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	Type I Error
	Acquit	Type II Error	Correct

We could make two types of errors:

- Convict an innocent defendant (**type-I error**)
- Acquit a guilty defendant (**type-II error**)

Our goal is to limit the probability of making these types of errors.

Hypothesis Testing: Setup

Copied from Brandon Stewart's lecture slides

Goal: test a **hypothesis** about the value of a parameter.

Statistical **decision theory** underlies such hypothesis testing.

Trial Example:

Suppose we must decide whether to convict or acquit a defendant based on evidence presented at a trial. There are four possible outcomes:

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	Type I Error
	Acquit	Type II Error	Correct

We could make two types of errors:

- Convict an innocent defendant (**type-I error**)
- Acquit a guilty defendant (**type-II error**)

Our goal is to limit the probability of making these types of errors.

However, creating a decision rule which minimizes both types of errors at the same time is impossible. We therefore need to balance them.

Hypothesis Testing: Error Types

Copied from Brandon Stewart's lecture slides

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	Type-I error
	Acquit	Type-II error	Correct

Now, suppose that we have a statistical model for the probability of convicting and acquitting, conditional on whether the defendant is actually guilty or innocent.

Hypothesis Testing: Error Types

Copied from Brandon Stewart's lecture slides

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	α
	Acquit	Type-II error	Correct

Now, suppose that we have a statistical model for the probability of convicting and acquitting, conditional on whether the defendant is actually guilty or innocent.

Then, our decision-making rule can be characterized by two probabilities:

- $\alpha \;=\; \Pr(\text{type-I error}) \;=\; \Pr(\text{convict} \mid \text{innocent})$

Hypothesis Testing: Error Types

Copied from Brandon Stewart's lecture slides

		Defendant	
		Guilty	Innocent
Decision	Convict	Correct	α
	Acquit	β	Correct

Now, suppose that we have a statistical model for the probability of convicting and acquitting, conditional on whether the defendant is actually guilty or innocent.

Then, our decision-making rule can be characterized by two probabilities:

- $\alpha = \Pr(\text{type-I error}) = \Pr(\text{convict} \mid \text{innocent})$
- $\beta = \Pr(\text{type-II error}) = \Pr(\text{acquit} \mid \text{guilty})$

Hypothesis Testing: Error Types

Copied from Brandon Stewart's lecture slides

		Defendant	
		Guilty	Innocent
Decision	Convict	$1-\beta$	α
	Acquit	β	$1-\alpha$

Now, suppose that we have a statistical model for the probability of convicting and acquitting, conditional on whether the defendant is actually guilty or innocent.

Then, our decision-making rule can be characterized by two probabilities:

- α = Pr(type-I error) = Pr(convict | innocent)
- β = Pr(type-II error) = Pr(acquit | guilty)

The probability of making a correct decision is therefore $1-\alpha$ (if innocent) and $1-\beta$ (if guilty).

Hypothesis Testing: Error Types

Copied from Brandon Stewart's lecture slides

		Defendant	
		Guilty	Innocent
Decision	Convict	$1-\beta$	α
	Acquit	β	$1-\alpha$

Now, suppose that we have a statistical model for the probability of convicting and acquitting, conditional on whether the defendant is actually guilty or innocent.

Then, our decision-making rule can be characterized by two probabilities:

- α = Pr(type-I error) = Pr(convict | innocent)
- β = Pr(type-II error) = Pr(acquit | guilty)

The probability of making a correct decision is therefore $1-\alpha$ (if innocent) and $1-\beta$ (if guilty).

Hypothesis testing follows an analogous logic, where we want to decide whether to **reject** (= convict) or **fail to reject** (= acquit) a **null hypothesis** (= defendant) using sample data.

Hypothesis Testing: Steps

Copied from Brandon Stewart's lecture slides

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1-\beta$	α
	Fail to Reject	β	$1-\alpha$

1. Specify a **null hypothesis** H_0 (e.g. the defendant = innocent)

Hypothesis Testing: Steps

Copied from Brandon Stewart's lecture slides

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1-\beta$	α
	Fail to Reject	β	$1-\alpha$

1. Specify a **null hypothesis** H_0 (e.g. the defendant = innocent)

2. Pick a value of $\alpha = \Pr(\text{reject } H_0 \mid H_0)$ (e.g. 0.05). This is the maximum probability of making

Hypothesis Testing: Steps

Copied from Brandon Stewart's lecture slides

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1-\beta$	α
	Fail to Reject	β	$1-\alpha$

1. Specify a **null hypothesis** H_0 (e.g. the defendant = innocent)

2. Pick a value of $\alpha = \Pr(\text{reject } H_0 \mid H_0)$ (e.g. 0.05). This is the maximum probability of making a type-I error we decide to tolerate, and called the **significance level** of the test.

Hypothesis Testing: Steps

Copied from Brandon Stewart's lecture slides

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1-\beta$	α
	Fail to Reject	β	$1-\alpha$

1. Specify a **null hypothesis** H_0 (e.g. the defendant = innocent)
2. Pick a value of $\alpha = \Pr(\text{reject } H_0 \mid H_0)$ (e.g. 0.05). This is the maximum probability of making a type-I error we decide to tolerate, and called the **significance level** of the test.
3. Choose a **test statistic** T, which is a function of sample data and related to H_0 (e.g. the count of testimonies against the defendant)

Hypothesis Testing: Steps

Copied from Brandon Stewart's lecture slides

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1-\beta$	α
	Fail to Reject	β	$1-\alpha$

1. Specify a **null hypothesis** H_0 (e.g. the defendant = innocent)
2. Pick a value of $\alpha = \Pr(\text{reject } H_0 \mid H_0)$ (e.g. 0.05). This is the maximum probability of making a type-I error we decide to tolerate, and called the **significance level** of the test.
3. Choose a **test statistic** T, which is a function of sample data and related to H_0 (e.g. the count of testimonies against the defendant)
4. Assuming H_0 is true, derive the **null distribution** of T (e.g. standard normal)

Hypothesis Testing: Steps

Copied from Brandon Stewart's lecture slides

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1-\beta$	α
	Fail to Reject	β	$1-\alpha$

5. Using the **critical values** from a statistical table, evaluate how unusual the observed value of T is under the null hypothesis:
   - If the probability of drawing a T **at least as extreme** as the observed T is less than α, we reject H_0.
     (e.g. there are too many testimonies against the defendant for her to be innocent, so reject the hypothesis that she was innocent.)
   - Otherwise, we fail to reject H_0.
     (e.g. there is not enough evidence against the defendant, so give her the benefit of the doubt.)

Hypothesis testing example: Our ages

Suppose we are interested in whether the average age among Princeton students in graduate courses is 25. We assume our class is a representative sample (though this is a HUGE assumption). Go to `http://tiny.cc/ygplfy` and enter your age in years.

Hypothesis testing example: Our ages

What is the null?

Hypothesis testing example: Our ages

What is the null? $H_0 : \mu = 25$
What is the alternative?

Hypothesis testing example: Our ages

What is the null? $H_0 : \mu = 25$
What is the alternative? $H_a : \mu \neq 25$
What is the significance level?

Hypothesis testing example: Our ages

What is the null? $H_0 : \mu = 25$
What is the alternative? $H_a : \mu \neq 25$
What is the significance level? $\alpha = .05$
What is the test statistic?

Hypothesis testing example: Our ages

What is the null?$H_0 : \mu = 25$
What is the alternative?$H_a : \mu \neq 25$
What is the significance level?$\alpha = .05$
What is the test statistic?$Z = \frac{\bar{X}-25}{\sigma/\sqrt{n}}$
What is the null distribution?

Hypothesis testing example: Our ages

What is the null?$H_0 : \mu = 25$
What is the alternative?$H_a : \mu \neq 25$
What is the significance level?$\alpha = .05$
What is the test statistic?$Z = \frac{\bar{X}-25}{\sigma/\sqrt{n}}$
What is the null distribution? $Z \sim N(0,1)$
What is are the critical values?

Hypothesis testing example: Our ages

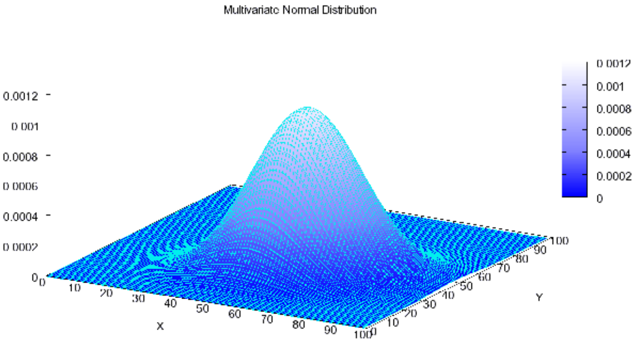
What is the null?$H_0 : \mu = 25$
What is the alternative?$H_a : \mu \neq 25$
What is the significance level?$\alpha = .05$
What is the test statistic?$Z = \frac{\bar{X}-25}{\sigma/\sqrt{n}}$
What is the null distribution? $Z \sim N(0,1)$
What is are the critical values? -1.96 and 1.96
What is the rejection region?

Hypothesis testing example: Our ages

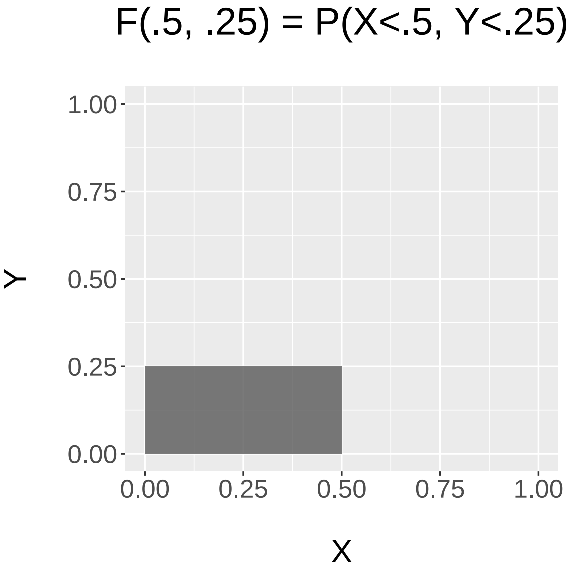
What is the null? $H_0 : \mu = 25$
What is the alternative? $H_a : \mu \neq 25$
What is the significance level? $\alpha = .05$
What is the test statistic? $Z = \frac{\bar{X} - 25}{\sigma / \sqrt{n}}$
What is the null distribution? $Z \sim N(0, 1)$
What is are the critical values? -1.96 and 1.96
What is the rejection region? We reject if $Z < -1.96$ or $Z > 1.96$.

Joint Distributions

The joint distribution of X and Y is defined by a **joint PDF** $f(x, y)$, or equivalently by a **joint CDF** $F(x, y)$.



Join CDF visualization



CDF practice problem 1

Modified from Blitzstein and Morris

Suppose $a < b$, where a and b are constants (for concreteness, you could imagine $a = 3$ and $b = 5$). For some distribution with PDF f and CDF F, which of the following must be true?

1. $f(a) < f(b)$
2. $F(a) < F(b)$
3. $F(a) \leq F(b)$

Joint CDF practice problem 2

Modified from Blitzstein and Morris

Suppose $a_1 < b_1$ and $a_2 < b_2$. Show that
$F(b_1, b_2) - F(a_1, b_2) + F(b_1, a_2) - F(a_1, a_2) \geq 0$.

Joint CDF practice problem 2

Modified from Blitzstein and Morris

Suppose $a_1 < b_1$ and $a_2 < b_2$. Show that
$F(b_1, b_2) - F(a_1, b_2) + F(b_1, a_2) - F(a_1, a_2) \geq 0$.

$$= [F(b_1, b_2) - F(a_1, b_2)] + [F(b_1, a_2) - F(a_1, a_2)]$$

Joint CDF practice problem 2

Modified from Blitzstein and Morris

Suppose $a_1 < b_1$ and $a_2 < b_2$. Show that
$F(b_1, b_2) - F(a_1, b_2) + F(b_1, a_2) - F(a_1, a_2) \geq 0$.

$$= [F(b_1, b_2) - F(a_1, b_2)] + [F(b_1, a_2) - F(a_1, a_2)]$$

$$= (\text{something } \geq 0) + (\text{something } \geq 0)$$

Joint CDF practice problem 2

Modified from Blitzstein and Morris

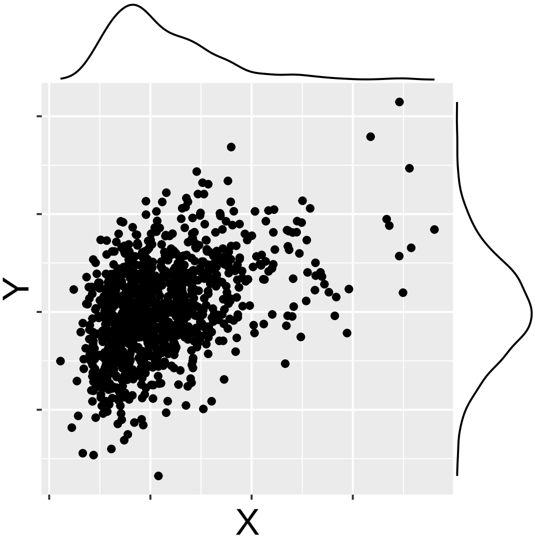
Suppose $a_1 < b_1$ and $a_2 < b_2$. Show that $F(b_1, b_2) - F(a_1, b_2) + F(b_1, a_2) - F(a_1, a_2) \geq 0$.

$$= [F(b_1, b_2) - F(a_1, b_2)] + [F(b_1, a_2) - F(a_1, a_2)]$$

$$= (\text{something } \geq 0) + (\text{something } \geq 0)$$

$$\geq 0$$

Marginalizing



Correlation between sum and difference of dice

Suppose you roll two dice and get numbers X and Y. What is $Cov(X+Y, X-Y)$?
Let's solve by simulation!

Correlation between sum and difference of dice

Correlation between sum and difference of dice

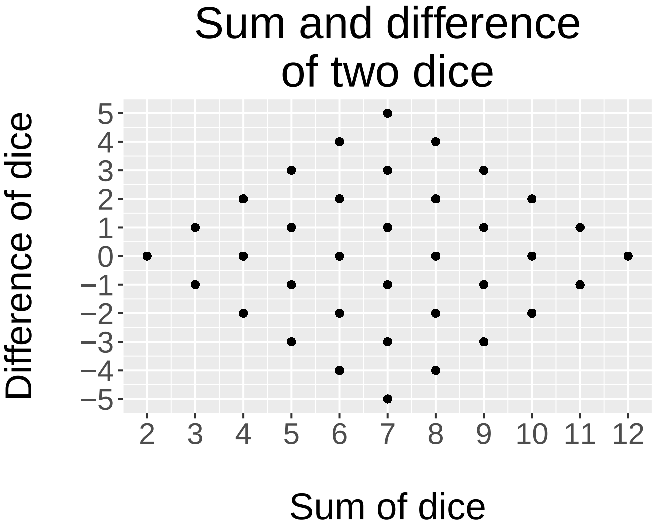
```
draw.sum.diff <- function() {
  x <- sample(1:6,1)
  y <- sample(1:6,1)
  return(c(x+y,x-y))
}
samples <- matrix(nrow=5000,ncol=2)
colnames(samples) <- c("x","y")
set.seed(08544)
for (i in 1:5000) {
  samples[i,] <- draw.sum.diff()
}
```

Sum and difference of dice: Plot

Sum and difference of dice: Plot

```
ggplot(data.frame(samples), aes(x=x,y=y)) +
  geom_point() +
  scale_x_continuous(breaks=c(2:12),
                     name="\nSum of dice") +
  scale_y_continuous(breaks=c(-5:5),
                     name="Difference of dice\n") +
  ggtitle("Sum and difference\nof two dice") +
  theme(text=element_text(size=20)) +
  ggsave("SumDiff.pdf",
         height=4, width=5)
```

Sum and difference of dice: Plot



Examples to clarify independence

Blitzstein and Hwang, Ch. 3 Exercises

1. Give an example of dependent r.v.s X and Y such that $P(X < Y) = 1$.
2. Can we have independent r.v.s X and Y such that $P(X < Y) = 1$?
3. If X and Y are independent and Y and Z are independent, does this imply X and Z are independent?

Examples to clarify independence

Blitzstein and Hwang, Ch. 3 Exercises

1. Give an example of dependent r.v.s X and Y such that $P(X < Y) = 1$.
   - $Y \sim N(0,1)$, $X = Y - 1$
2. Can we have independent r.v.s X and Y such that $P(X < Y) = 1$?
   - $X \sim Uniform(0,1)$, $Y \sim Uniform(2,3)$
3. If X and Y are independent and Y and Z are independent, does this imply X and Z are independent?
   - No. Consider $X \sim N(0,1)$, $Y \sim N(0,1)$, with X and Y independent, and $Z = X$.

The power of conditioning: Coins

Suppose your friend has two coins, one which is fair and one which has a probability of heads of $3/4$. Your friend picks a coin randomly and flips it. What is the probability of heads?

The power of conditioning: Coins

Suppose your friend has two coins, one which is fair and one which has a probability of heads of $3/4$. Your friend picks a coin randomly and flips it. What is the probability of heads?

$$P(H) =$$

The power of conditioning: Coins

Suppose your friend has two coins, one which is fair and one which has a probability of heads of 3/4. Your friend picks a coin randomly and flips it. What is the probability of heads?

$$P(H) = P(H \mid F)P(F) + P(H \mid F^C)P(F^C)$$

The power of conditioning: Coins

Suppose your friend has two coins, one which is fair and one which has a probability of heads of 3/4. Your friend picks a coin randomly and flips it. What is the probability of heads?

$$P(H) = P(H \mid F)P(F) + P(H \mid F^C)P(F^C)$$

$$= 0.5(.05) + .75(.05) = 0.625$$

The power of conditioning: Coins

Suppose your friend has two coins, one which is fair and one which has a probability of heads of 3/4. Your friend picks a coin randomly and flips it. What is the probability of heads?

$$P(H) = P(H \mid F)P(F) + P(H \mid F^C)P(F^C)$$

$$= 0.5(.05) + .75(.05) = 0.625$$

Suppose the flip was heads. What is the probability that the coin chosen was fair?

$$P(F \mid H) =$$

The power of conditioning: Coins

Suppose your friend has two coins, one which is fair and one which has a probability of heads of 3/4. Your friend picks a coin randomly and flips it. What is the probability of heads?

$$P(H) = P(H \mid F)P(F) + P(H \mid F^C)P(F^C)$$

$$= 0.5(.05) + .75(.05) = 0.625$$

Suppose the flip was heads. What is the probability that the coin chosen was fair?

$$P(F \mid H) = \frac{P(H \mid F)P(F)}{P(H)}$$

The power of conditioning: Coins

Suppose your friend has two coins, one which is fair and one which has a probability of heads of 3/4. Your friend picks a coin randomly and flips it. What is the probability of heads?

$$P(H) = P(H \mid F)P(F) + P(H \mid F^C)P(F^C)$$

$$= 0.5(.05) + .75(.05) = 0.625$$

Suppose the flip was heads. What is the probability that the coin chosen was fair?

$$P(F \mid H) = \frac{P(H \mid F)P(F)}{P(H)}$$

$$= \frac{0.5(0.5)}{0.625)} = 0.4$$

Exponential: $-\log$ Uniform

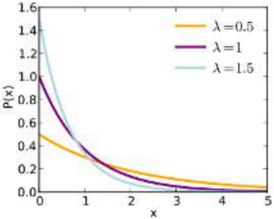
Figure credit: Wikipedia

The Uniform distribution is defined on the interval $(0, 1)$. Suppose we wanted a distribution defined on all positive numbers.

Definition

X follows an **exponential** distribution with rate parameter λ if

$$X \sim -\frac{1}{\lambda} log(U)$$



Exponential: $-\log$ Uniform

The exponential is often used for wait times. For instance, if you're waiting for shooting stars, the time until a star comes might be exponentially distributed.

Key properties:

- Memorylessness: Expected remaining wait time does not depend on the time that has passed
- $E(X) = \frac{1}{\lambda}$
- $V(X) = \frac{1}{\lambda^2}$

Exponential-uniform connection

Suppose $X_1, X_2 \overset{iid}{\sim} Expo(\lambda)$. What is the distribution of $\frac{X_1}{X_1+X_2}$?

Exponential-uniform connection

Suppose $X_1, X_2 \overset{iid}{\sim} Expo(\lambda)$. What is the distribution of $\frac{X_1}{X_1+X_2}$? The proportion of the wait time that is represented by X_1 is uniformly distributed over the interval, so

$$\frac{X_1}{X_1 + X_2} \sim Uniform(0, 1)$$

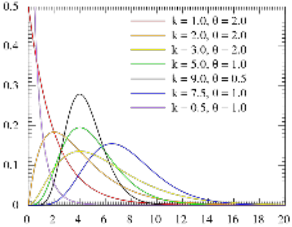
Gamma: Sum of independent Exponentials

Figure credit: Wikipedia

Definition

Suppose we are waiting for a shooting stars, with the time between stars $X_1, \ldots, X_a \overset{iid}{\sim} Expo(\lambda)$. The distribution of time until the ath shooting star is

$$G \sim \sum_{i=1}^{a} X_i \sim Gamma(a, \lambda)$$



Gamma: Properties

Properties of the *Gamma*(a, λ) distribution include:

- $E(G) = \frac{a}{\lambda}$
- $V(G) = \frac{a}{\lambda^2}$

Beta: Uniform order statistics

Suppose we draw $U_1, \ldots, U_k \sim \mathit{Uniform}(0, 1)$, and we want to know the distribution of the jth *order statistic*, $U_{(j)}$. Using the Uniform-Exponential connection, we could also think of these $U_{(j)}$ as being the location of the jth Exponential in a series of $k+1$ Exponentials. Thus,

$$U_{(j)} \sim \frac{\sum_{i=1}^{j} X_i}{\sum_{i=1}^{j} X_i + \sum_{i=j+1}^{k+1} X_i} \sim \mathit{Beta}(j, k-j+1)$$

This defines the **Beta distribution**.
Can we name the distribution at the top of the fraction?

Beta: Uniform order statistics

Suppose we draw $U_1, \ldots, U_k \sim \mathit{Uniform}(0,1)$, and we want to know the distribution of the jth *order statistic*, $U_{(j)}$. Using the Uniform-Exponential connection, we could also think of these $U_{(j)}$ as being the location of the jth Exponential in a series of $k+1$ Exponentials. Thus,

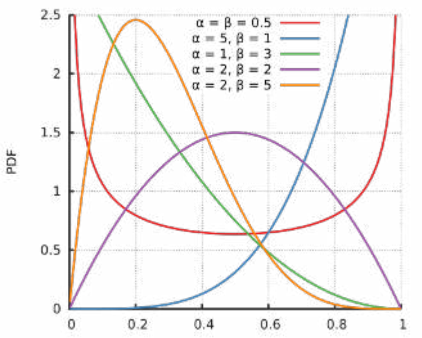
$$U_{(j)} \sim \frac{\sum_{i=1}^{j} X_i}{\sum_{i=1}^{j} X_i + \sum_{i=j+1}^{k+1} X_i} \sim \mathit{Beta}(j, k-j+1)$$

This defines the **Beta distribution**.
Can we name the distribution at the top of the fraction?

$$\sim \frac{G_j}{G_j + G_{k-j+1}}$$

What do Betas look like?

Figure credit: Wikipedia



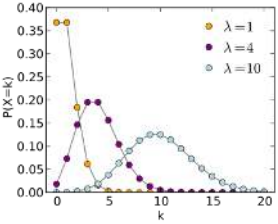
Poisson: Number of Exponential events in a time interval

Figure credit: Wikipedia

Definition

Suppose the time between shooting stars is distributed $X \sim Expo(\lambda)$. Then, the number of shooting stars in an interval of time t is distributed

$$Y_t \sim Poisson(\lambda t)$$



Poisson: Number of Exponential events in a time interval

Properties of the Poisson:

- If $Y \sim Pois(\lambda t)$, then $V(Y) = E(Y) = \lambda t$
- Number of events in disjoint intervals are independent

χ_n^2: A particular Gamma

Definition

We define the **chi-squared distribution** with n degrees of freedom as

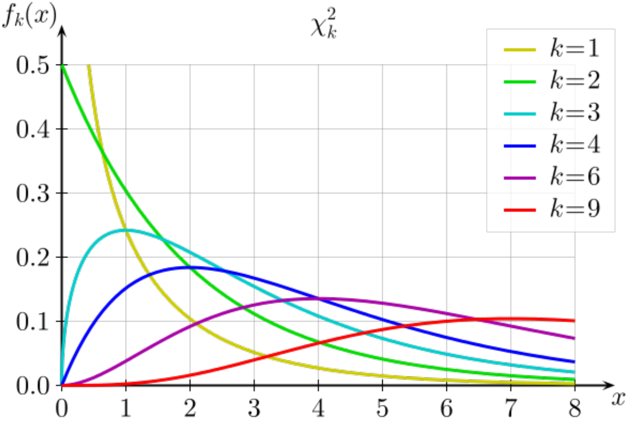
$$\chi_n^2 \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

More commonly, we think of it as the sum of a series of independent squared Normals, $Z_1, \ldots, Z_n \overset{iid}{\sim} Normal(0,1)$:

$$\chi_n^2 \sim \sum_{i=1}^{n} Z_i^2$$

χ^2_n: A particular Gamma

Figure credit: Wikipedia

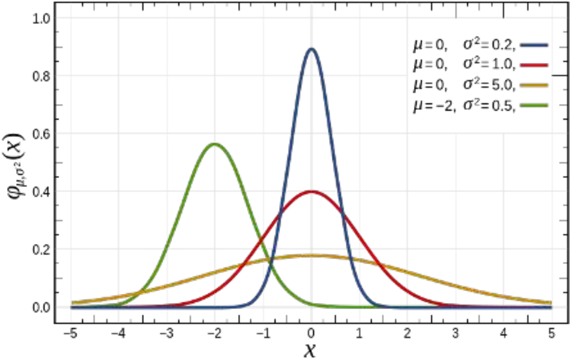


Normal: Square root of χ_1^2 with a random sign

Figure credit: Wikipedia

Definition

Z follows a **Normal** distribution if $Z \sim S\sqrt{\chi_1^2}$, where S is a random sign with equal probability of being 1 or -1.



Normal: An alternate construction

Note: This is far above and beyond what you need to understand for the course!

Box-Muller Representation of the Normal

Let $U_1, U_2 \overset{iid}{\sim} \textit{Uniform}$. Then

$$Z_1 \equiv \sqrt{-2\log U_2}\cos(2\pi U_1)$$

$$Z_2 \equiv \sqrt{-2\log U_2}\sin(2\pi U_1)$$

so that $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

What is this?

- **Inside the square root:** $-2\log U_2 \sim \textit{Expo}\left(\frac{1}{2}\right) \sim \textit{Gamma}\left(\frac{2}{2}, \frac{1}{2}\right) \sim \chi^2_2$
- **Inside the cosine and sine:** $2\pi U_1$ is a uniformly distributed angle in polar coordinates, and the cos and sin convert this to the cartesian x and y components, respectively.
- **Altogether in polar coordinates:** The square root part is like a radius distributed $\chi \sim |Z|$, and the second part is an angle.