Probability ← Assumes we understand rules of the game

Rules of the game → Outcome

← We know the outcome

Statistics

$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probability Question

If I flip two coins, what is the probability that they are both heads?

$$P(E) = \lim_{n \to \infty} \frac{\text{num of occurrences in } n \text{ trials}}{n}$$

Question of perspective where the coins are already flipped, but we just don't know if the are heads or tails.

Frequentist: Long-run probabilities

Bayesian: Probability changes with new events. We take in these events to find probability

disjoint

why $P(x<a)=P(x\leq a$

sample space

Counting and Sets

What is the probability of flipping exactly one heads out of 3 coin flips?

$$P(E) = \frac{\text{Num of ways } E \text{ can occur}}{\text{Total num of things that can happen}} = \frac{|E|}{|\Omega|}$$

* If all outcomes are equally likely

Magnitude (num of elements in set).

Event i.e. "exactly one heads"

or S, Sample Space

$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ ➔ $\frac{|E|}{|\Omega|} = \boxed{\frac{3}{8}}$

$E = \{HTT, THT, TTH\}$

Sets

"A set is a collection of elements", unordered"

$A = \{1, 2, 3\}$ $B = \{\text{red}, \text{apple}, \frac{1}{2}, x\}$

$C = \{1, \{2\}, \{3\}\}$ $D = \{1, 2\}$

Notation:

$\in$ - "is an element of" $1 \in A$, $\{2\} \in C$, $2 \notin C$ ← Also, $\subset$

$\subseteq$ - "is a subset of" $D \subseteq A$, $A \subseteq A$

$\subsetneq$ - "is a proper subset of" $D \subsetneq A$, $A \not\subsetneq A$ Everything in $D \subseteq A$ but $D \neq A$

$\emptyset$ - "empty set" $\{\}$ $\emptyset \subseteq A$ for any A

Note: $A = B \Leftrightarrow A \subseteq B$ and $B \subseteq A$

$\mathbb{N} = \{1, 2, 3, \dots\}$ natural numbers

$\mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}$ Integers

Set Operations

Union:

"or" $A \cup B = \{\text{elements in } A \text{ or } B\}$

Intersection:

"and" $A \cap B = \{\text{elements in both } A \text{ and } B\}$

Complement:

A^c A' $= \{\text{elements not in } A, \text{but in } U\}$

$\bar{A}$ $\neg A$ or $\setminus A$

Universal Set: U

Power Set:

$P(A)$ = set of all subsets of A

Ex: $S = \{1, 2, 3\}$ $\quad U = \mathbb{N}$

$T = \{3, 4\}$

$S \cup T = \{1, 2, 3, 4\}$

$S \cap T = \{3\}$

$S' = \{4, 5, 6, 7, \ldots\}$

$T = \{3, 4\}$

$P(T) = \{\{3\}, \{4\}, \{3, 4\}, \emptyset\}$

Counting

For finite sets A, B, with $A \cap B \neq 0$

$|A \cup B| = |A| + |B|$

$|A \times B| = |A| \times |B|$

← If no elements are the same in A and B then...

Ex. $A = \{1, 2\}$ $B = \{a, b, c\}$

$A \cup B = \{1, 2, a, b, c\}$ $\quad |A \cup B| = 5$

$A \times B = \{(1,a), (1,b), (1,c), (2,a), (2,b), (2,c)\}$ $\quad |A \times B| = 6$

Ex. 2 A restaurant has 6 appetizers and 8 main courses.

$|A| = 6$ $\quad |M| = 8$

$|A \cup M| = 6 + 8 = 14$

There are 14 items on the menu.

How many different combinations of one appeticer and one main are possible?

$|A \times M| = 6 \cdot 8 = 48$

Monka

Inclusion – Exclusion Principle

$$|A \cup B| = |A| + |B| - \underbrace{|A \cap B|}$$

← Think about 3-set inclusion-exclusion $|A \cup B \cup C|$

So we don't double count same variables in two different sets

Use often when dependence is unknown.

Two set Inclusion-Exclusion Principle

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Idempotence: $A \cap A = A$

De Morgan's Laws

For sets A and B

$(A \cap B)' = A' \cup B'$

$(A \cup B)' = A' \cap B'$

Combinatorics

Ex: Alice, Bob, and Charlie run a race, how many results are possible?

ABC BAC CAB ACB BCA CBA

$\underline{3}\cdot\underline{2}\cdot\underline{1}=6$ ← Just a factorial (3!)

3 options for who can win race

2 others can get second

Someone gets last

Ex. 2: 8 people run a race, how many distinct ways can 1st, 2nd, and 3rd be awarded?

$8\cdot 7\cdot 6=504$ ← This is a permutation where $n=8$, $k=3$

Permutations ← Order matters

$$P(n,k) \text{ or } {}_nP_k = \frac{n!}{(n-k)!}$$

Π ← sigma

Same as Σ

$$P(8,3)=\frac{8!}{(8-3)!}=\frac{8!}{5!}=\frac{8\cdot7\cdot6\cdot\cancel{5}\cdot\cancel{4}\cdot\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}{\cancel{5}\cdot\cancel{4}\cdot\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}=8\cdot7\cdot6=504$$

Ex. 3: 8 people run a qualifying race where, the top 3 move on to the next stage. How many distinct groups of 3 qualifiers are possible?

$8\cdot7\cdot6$ ← We don't care which place everyone is in the top 3. This is double counting since ABC is the same for us as BAC. We double count by 3!.

To counter this we can write as:

$$\frac{8\cdot7\cdot6}{3\cdot2\cdot1}=54$$ ← This is a combination

Combinations ← Order doesn't matter

$$\left.\begin{matrix} C(n,k) \\ {}_nC_k \\ \binom{n}{k} \end{matrix}\right\} = \frac{\overbrace{n\cdot(n-1)\cdot(n-2)\ldots}^{k}}{k!}=\frac{n!}{(n-k)!\,k!}$$

$$C(8,3)=\frac{8\cdot7\cdot6\cdot\cancel{5\cdot4\cdot3\cdot2\cdot1}}{(\cancel{5\cdot4\cdot3\cdot2\cdot1})(3\cdot2\cdot1)}=\frac{8\cdot7\cdot6}{3\cdot2\cdot1}=54$$

Exponentiation:
- Order matters
- Repitition allowed

Permutations:
- Order matters
- No repitition

Combinations:
- Order does NOT matter
- No repitition

Note: replacement means repitition

Probability Rules

- $P(A') = 1 - P(A)$

Ex: A = I make the free throw

A' = I don't make the free throw

$P(A) = 0.3$ $P(A') = 0.7$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Could be dependent

Ex: A = The train is late $P(A) = 0.3$

B = It is raining $P(B) = 0.2$

- If A and B are independent,

$P(A \cap B) = P(A)P(B)$

Ex: A = I roll a 4 (On a 6-sided dice)

B = I get "tails" (On a coin flip)

- If all outcomes are equally likely,

$$P(A) = \frac{|A|}{|S|}$$

Ex: I roll 2 dice, let A be the event where the sum is 7.

$A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

$|A| = 6$

$|S| = 36$

$P(A) = \frac{6}{36} = \frac{1}{6}$

Conditional Probability

- Updating probability given new information

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

"|" - given

Ex: I flip 2 coins

B = "Both are heads" A = "Left coin is heads"

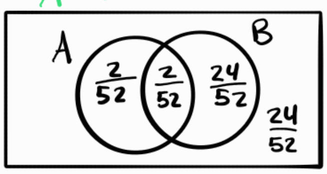
$P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(B|A) = 1 \cdot \frac{1}{2} = \frac{1}{2}$

B given A

Ex 2: I draw a card

A = The card is an Ace B = The card is Red



$P(A) = \frac{1}{13}$ $P(B) = \frac{1}{2}$

$$P(B|A) = \frac{2}{4} = \frac{1}{2} = \frac{P(A \text{ and } B)}{P(A)} = \frac{2/52}{4/52} = \frac{1}{2}$$

- Two events A, B are independent if, $P(B|A) = P(B)$ or $P(A \cap B) = P(A)P(B)$

- Two events A, B are disjoint if, $P(A \text{ and } B) = 0$

disjoint

← Can't possibly both happen

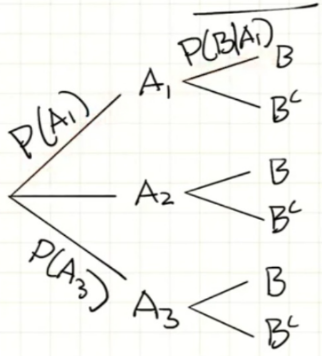
Generalized Multiplication Rule

$$P(A \text{ and } B) = P(A)P(B|A)$$

if A and B are independent,

$P(A \text{ and } B) = P(A)P(B)$

Tree Diagrams



$P(A_1 \text{ and } B) = P(A_1)$

Bayes Theorem

$$P(H|E) = \boxed{\frac{P(H)P(E|H)}{P(H)P(E|H) + P(\neg H)P(E|\neg H)}} = \boxed{\frac{P(H)P(E|H)}{P(E)}}$$

Posterior: Hypothesis after seeing evidence

When to use:

You have a hypothesis and You've observed some evidence so You want $P(H|E)$

Hypothesis given evidence

$P(H)$ ← "prior"

$P(E|H)$ ← Evidence given H is true
"Likelihood"

$P(E|\neg H)$ ← Evidence given H is NOT true

Sensitivity and Specificity

Ex: "When it snows, how often does Bellevue cancel school?

$P(c|s)$ probability school is canceled given snow.

"When Bellevue cancels school, how often is it because it snowed?"

$P(S|C)$ probability of snow given school canceled

Ex: If you have HIV, how likely are you to test positive? $P(+|H)$

If you tested positive for HIV, how likely is it you actually have it $P(H|+)$

H = actually has HIV, + = tests positive

Suppose:

$P(H) = 0.01$ ← prevalence

$P(+|H) = 0.99$ ← sensitivity

$P(-|H') = 0.98$ ← Specificity

What is $P(H|+)$? ← Positive predictive value

Test \ Reality	HIV	No HIV
+	True positive	False positive
−	False Negative	True Negative

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$PPV = \frac{TP}{TP+FP}$$

1,000,000 population

	Reality: HIV	No HIV	
Test +	9,900	19,800	29,700
Test −	100	970,200	970,300
	10,000	990,000	

$$PPV = P(H \mid +) = \frac{9{,}900}{29{,}700} = \boxed{0.33}$$

Discrete Probability Distributions

Random variables

- A numerical value X determined by the results of an experiment
  - Formerly $X : \Omega \to \mathbb{R}$

Ex: Roll two 6-sided dice, let X be the sum.

$$P(X=7) = \frac{6}{36} = \frac{1}{6}$$

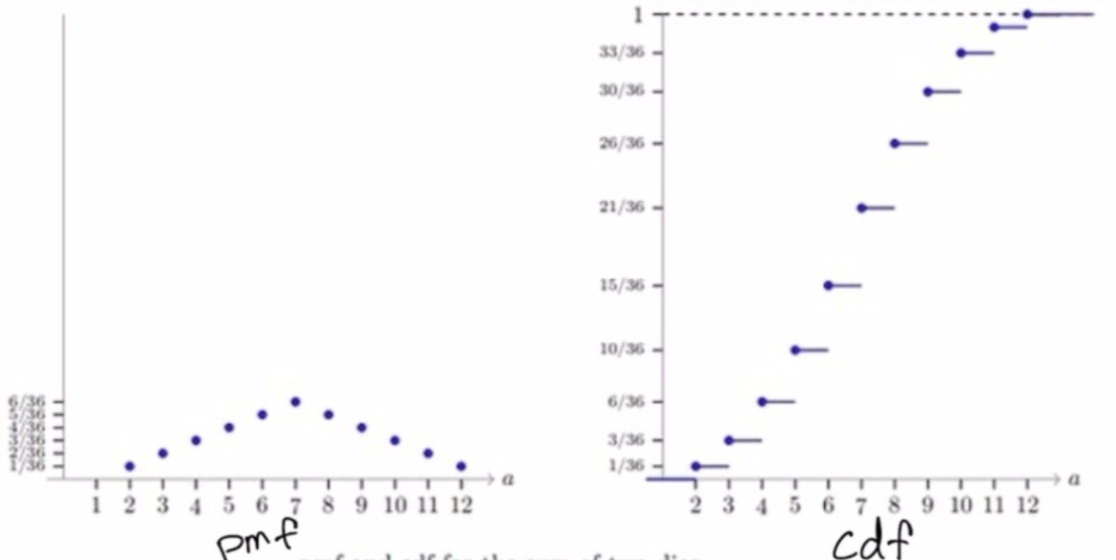
$$P(X=2) = \frac{1}{36}$$

Instead of writing each out we can use...

Probability Distribution

(aka "probability mass function" (PMF))

$$P(a) = P(X=a)$$



pmf and cdf for the sum of two dice

Cumulative Distribution Function (cdf):

$$F(a) = P(X \le a)$$

Expected Value

If you roll a 6-sided die what is the average result?

$E[x] = \frac{1}{6}1 + \frac{1}{6}2 + \frac{1}{6}3 + \frac{1}{6}4 + \frac{1}{6}5 + \frac{1}{6}6 = 3.5$

The expected value of x

$\mu = E[x] = \sum p_i x_i = p_1 x_1 + p_2 x_2 + \ldots$

Ex: I propose a game:
We'll roll 2 dice,
- If the sum is 2, you get \$100
- If not you lose \$10

pmf
$P(x=-10) = \frac{35}{36}$
$p(x=100) = \frac{1}{36}$

$E(x) = (-10)\frac{35}{36} + (100)\frac{1}{36} \approx -69.4$

Algebra of $E(x)$

$E(cx) = cE(x)$

$E(x+y) = E(x) + E(y)$

Variance (and Standard Deviation)

Two data sets:

$\{79, 80, 81\}$ $\{60, 80, 100\}$

$\mu = 80$ $\mu = 80$

what's the difference?

Variance: $\frac{(79-80)^2 + (80-80)^2 + (81-80)^2}{3}$

$= \boxed{\frac{2}{3}}$

variance: $\frac{(60-80)^2 + (80-80)^2 + (100-80)^2}{3}$

$= \frac{800}{3}$

For Discrete Probability Distributions

$E(x) = \sum_i x_i p_i$

$Var(x) = \sum_i (x_i - E(x))^2 p_i$

$= \sum_i (x_i - \mu)^2 p_i$

$= E[(x-\mu)^2]$

$var(x+y) = var(x) + var(y)$

Standard Deviation $\sigma = \sqrt{var(x)}$

Probability Models

Bernoulli Trials

- Two possible outcomes (success or failures)
- Fixed probability p = prob of success $q = 1-p$
- Independent Trials

can be estimated w/ normal dist

	Geometric Model	Binomial Model
X	How many attempts until first success?	How many successes in a fixed number (n) of attempts?
	$P(X=k) = q^{k-1}p$ $E(x) = \frac{1}{p}$ $SD(x) = \sqrt{\frac{p}{q^2}}$	$P(x=k) = \binom{n}{k} p^k q^{n-k}$ $E(x) = np$ $SD(x) = \sqrt{npq}$

Geometric Model:

$P(X=k)$ is the probability that the first success is in the k^{th} element

Ex: $p=0.4$ (i.e. shooting free throws)

$P(X=3) = \underline{0.6}\ \underline{0.6}\ \underline{0.4} = 0.144$

3 steps

(14.4% that 1st free throw is made.

$P(X=3) = q\ q\ p = q^2 p$

$q = p'$

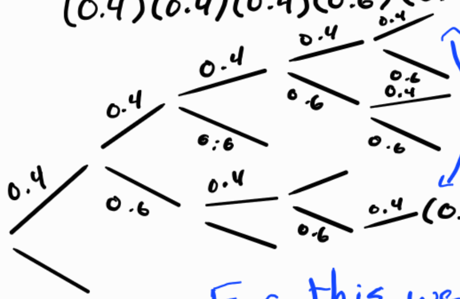
Binomial Model (n attempts):

$P(x=k)$ = probability of exactly k successes

Ex: $P=0.4$ (i.e. shooting free throws)
I attempt $n=5$ shots, what is the probability I make exactly 3?

$(0.4)(0.4)(0.4)(0.6)(0.6)$

Issue is with the order of these things occuring



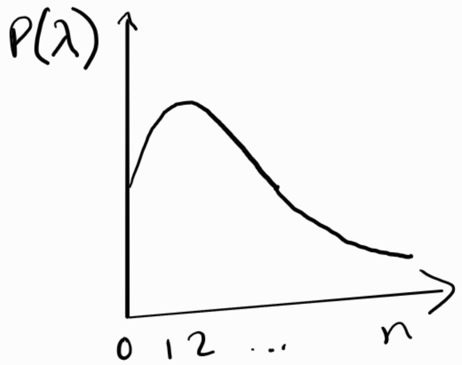
$(0.4)(0.6)(0.4)(0.6)(0.4)$

All nodes on same level will add up to 1

For this we use combinations and permutations

Poisson Distribution

- The frequency an event happens



$$P(y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

$$E(y) = y_0 \frac{\lambda^{y_0} e^{-\lambda}}{y_0!} + y_1 \frac{\lambda^{y_1} e^{-\lambda}}{y_1!} + \dots$$

$$\mu = \sigma^2 = \lambda$$

Mean = variance = λ

Ex. Usually we get 4 questions per day but yesterday we got 7

$\lambda = 4$
Interval of one day $\Big\}$ Probability function for $P_o(4)$

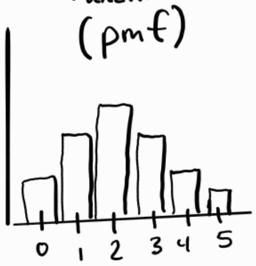
$y = 7$

$$P(7) = \frac{4^7 e^{-4}}{7!} \approx 0.06$$

Continuous Random Variables

Discrete

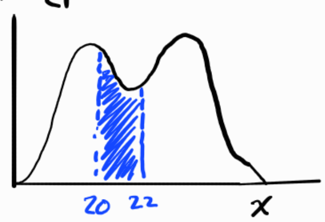
Probability Mass Function (pmf)



Continuous

Probability Density Function

$P(X=x)$ (pdf)

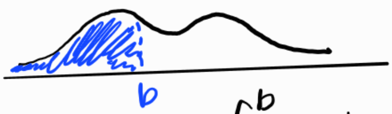


Properties of pdf:

Let X be a continuous random variable and $f(x)$ be its pdf.

- $f(x) \geq 0$
- $P(c \leq X \leq d) = \int_c^d f(x)\,dx$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Cumulative Distribution Function



$$\text{cdf} = F(b) = \int_{-\infty}^{b} f(x)\,dx$$

Properties:

- $0 \leq F(x) \leq 1$
- $F(x)$ is nondecreasing

$P(a \leq X \leq b) = F(b) - F(a)$

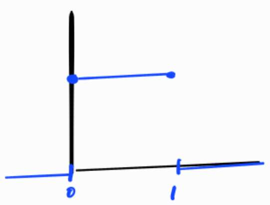
Continuous Random Variables

Suppose random variable X has pdf $f(x)$

Expected value

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

Example: Suppose $f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$



$$\int_0^1 x \cdot 1\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$$

$f(x) = 1$

$X \sim U(a,b)$

$U(a,b)$ has density $f(x) = \frac{1}{b-a}$ on $[a,b]$

Variance:

$\mu = E[x]$

$Var(x) = E[(x-m)^2]$

$var(x) = E[x^2] - E[x]^2$

$= E[x^2] - \mu^2$

$Var(X+Y) = Var(X) + Var(Y)$

$Std\ Dev(X) = \sqrt{var(X)}$

Linearity of Expectation: ← Mean of Y

$E[aX+b] = aE[X] + b$

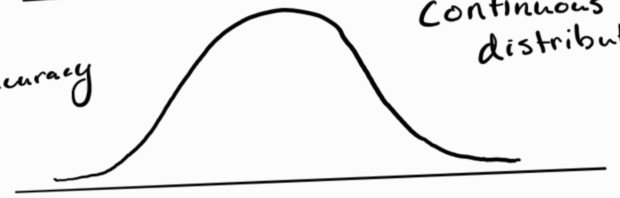
Variance Transformation Rule: ← variance of Y

$Var(aX+b) = a^2 Var(X)$

Normal Distributions

Higher sample means higher accuracy

continuous distribution



$\text{Z-score} = \frac{x-\mu}{\sigma}$

68-95-99.7 Rule

$\sigma = 68\%$

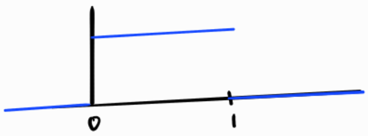
$2\sigma = 95\%$

$3\sigma = 99.7\%$

Continuous Random Variables Examples

Constant or Uniform:

Let $X \sim \text{uniform}(0,1)$



$E[X] = \int_0^1 1 \cdot x\, dx$

$= \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$

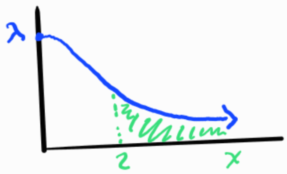
$Var(X) = \int_0^1 (x-\frac{1}{2})^2 dx = \frac{1}{12}$

Exponential:

Parameter λ

Notation $exp(\lambda)$

pdf: $f(x) = \lambda e^{-\lambda x}$ $x \geq 0$



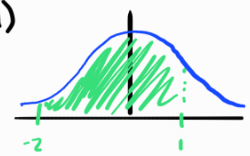
$\int_0^\infty \lambda e^{-\lambda x} dx = e^{-\lambda x}\Big|_0^\infty = 1$

$E[X] = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$

Normal:

$X = N(0, 1) = \text{norm}(\overset{\mu}{0}, \overset{\sigma}{1})$

$\text{pdf} = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$



Joint Discrete Distributions

Joint Probability Table:

Joint Probability Table

		True Status		
		Positive	Negative	
Test Result	+	0.04	0.02	0.06
	-	0.01	0.93	0.94
		0.05	0.95	1

Ex: I roll two 4-sided dice

$P(4,1) = 1/16$

Y	1	2	3	4	
1	1/16	1/16	1/16	1/16	1/4
2	1/16	1/16	1/16	1/16	1/4
3	1/16	1/16	1/16	1/16	1/4
4	1/16	1/16	1/16	1/16	1/4
					1

Event: $E = \{X + Y \geq 7\}$

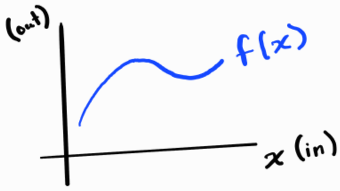
$\sum P(X+Y \geq 7) = 3/16$

We need

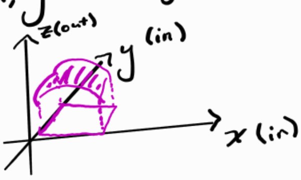
$$\sum_{i=1} \sum_{j=1} P(x_i, y_i) = 1$$

Multivariable Calculus

Single Variable Calculus	Multivariable Calculus
Function: $x \to f(x)$	$x, y \to f(x,y)$



$\int_a^b f(x)\,dx$



$\int_a^b \int_c^d f(x,y)\,dy\,dx$

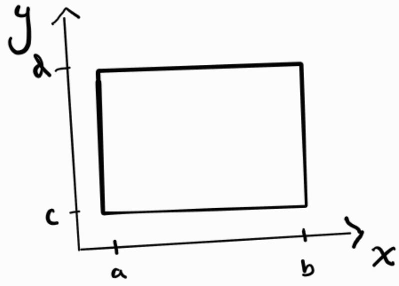
For region R

$\iint_R f(x,y)\,dy\,dx$

Joint Continuous Distributions

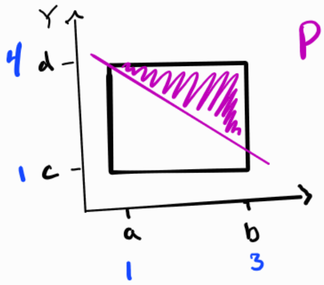
A joint pdf must satisfy:

- $f(x,y) \geq 0$
- $\int_c^d \int_a^b f(x,y)\,dx\,dy = 1$



Ex: uniform $(\overset{\int^a}{1}, \overset{\int^b}{3}, \overset{\int^c}{1}, \overset{\int^d}{4})$

has pdf $f(x,y) = \begin{cases} 1/6 & \text{for } 1 \leq x \leq 3 \text{ and } 1 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$



$P(X+Y \geq 5)$

$\iint_R f(x,y)\,dx\,dy = 2 \cdot 1/6 = \frac{1}{3}$

Area $R = \frac{1}{2} \cdot 2 \cdot 2 = 2$

Independence of Joint Distributions

Recall: $P(A \cap B) = P(A)P(B|A)$

Two events are independent if

$P(A \cap B) = P(A)P(B)$

aka if $P(B) = P(B|A)$ ← Knowing outcome of A doesn't affect prob. of B occuring

Ex. I roll two 4-sided die, X is the result of the first, Y is the maximum

Y \ X	1	2	3	4	
1	1/16	0	0	0	1/16
2	1/16	2/16	0	0	3/16
3	1/16	1/16	3/16	0	5/16
4	1/16	1/16	1/16	4/16	7/16
	1/4	1/4	1/4	1/4	

Marginal PMF

The bottom and right margins of our joint pmf tables

$P_x(x_i) = \sum_j P(x_i, x_j)$

Joint Discrete Variables are independent if $P(x_i, y_i) = P_x(x_i)P_Y(y_i)$

Y \ X	1	2	3	4	
1	1/16	1/16	1/16	1/16	1/4
2	1/16	1/16	1/16	1/16	1/4
3	1/16	1/16	1/16	1/16	1/4
4	1/16	1/16	1/16	1/16	1/4
	1/4	1/4	1/4	1/4	

These two multiplied by each othere equals (1,1)

Covariance and Correlation

Measures of the linear relationship between X and Y

$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$

Ex: I flip 3 coins,
X is number of heads in first two flips
Y is number of heads in second two flips

$\mu_X = \frac{1}{4}\cdot 0 + \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 = 1 = \mu_Y$

Y \ X	0	1	2	
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
	1/4	1/2	1/4	

$Cov(X,Y) = E[(X-1)(Y-1)] = \underbrace{\frac{1}{8}(0-1)(0-1)}_{(0,0)} + \underbrace{\frac{1}{8}(1-1)(0-1)}_{(1,0)} + \frac{1}{8}(1)(1) = \frac{1}{4}$

can skip when =1

Properties of Covariance:

2) $Cov(X_1+X_2, Y = Cov(X_1, Y) + Cov(X_2+Y)$

3) $Cov(X,X) = Var(X)$

4) $Cov(X,Y) = E[XY] - \mu_X\mu_Y$

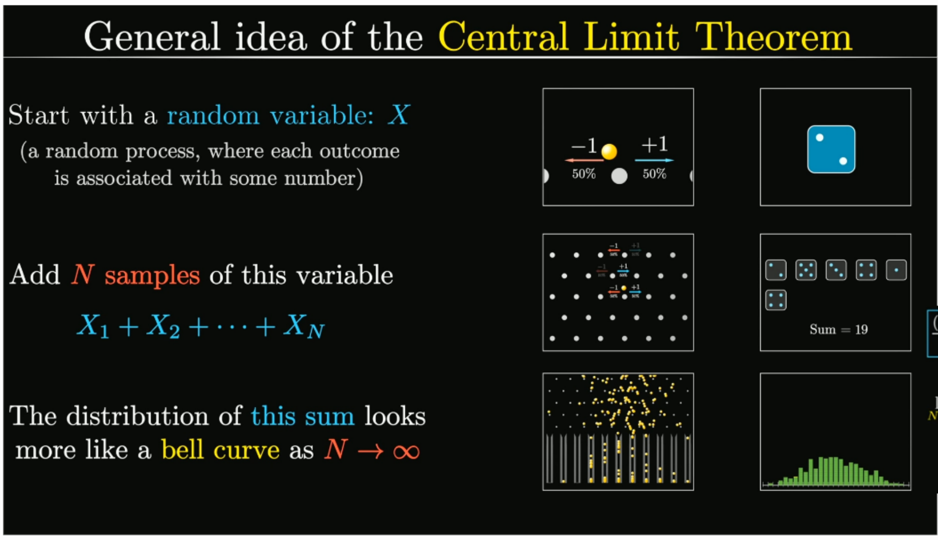
6) If X, Y are independent, $Cov(X,Y) = 0$

Central Limit Theorem

Let X be any random variable with $E(X) = \mu$ and $var(X) = \sigma^2$

Let Y_n be the sum of n samples from X

As n grows, Y_n approaches a normal distribution with $E(Y_n) n\mu$, $Var(Y_n) = n\sigma^2$

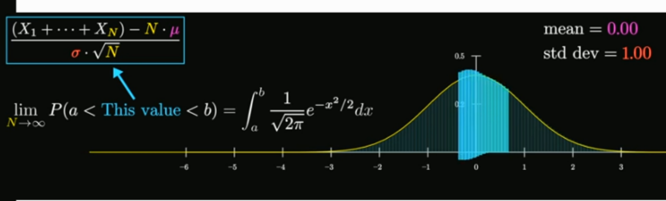


3 Assumptions

1) All X_i's are independent from each other

2) Each X_i is drawn from the same distribution

3) $0 < Var(X_i) < \infty$



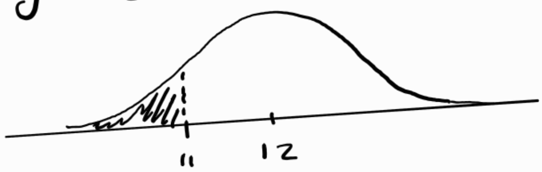
Ex. Salmon weigh 12 lbs on average with a standard deviation of 2 lbs.

If I catch 10 salmon, what is the probability that the average weight is less than 11 lbs?

Let Y be the avg. weight of the 10.

$E(Y) = 12$

$\sigma_Y = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{10}}$



$\text{cdf}(\text{Normal}(12, 2/\sqrt{10}), 11) \approx 0.057$

Representative Samples

What can we say about the real world? We'd like to ask everyone, but thats not generally possible. Instead we ask for a sample.

Famous Example: Literary Digest Poll

- 1936 Presidential Election: FDR vs. Robert Landon
- Asked 10 Million readers who they'll vote for.
  2.4 Million Responded:
  - FDR: 43%
  - Landon: 57%

  Reality:
  - FDR: 62%
  - Landon: 37%

Sampling Bias

Often to combat this statitians use random sampling

Types of Random Samples

- Simple: Randomly select from a population

Stratified:
- split population into groups based on a property
- ensure sample has appropriate representation of each group

Cluster:
- Take a random sample from non-random subgroups

Convinience/straw poll:
- Easiest
- Most open to bias

Sampling Distributions

Probability →

Population Parameters ⇄ Sample Statistics

← Inference

Quantatative $\left[\begin{array}{l}\mu \\ \sigma\end{array}\right.$ $\quad$ $\left.\begin{array}{l}\bar{x} \\ s\end{array}\right]$

Categorical $\left[\ p\right.$ $\quad$ $\left.\hat{p}\ \right]$

Ex. Assume that vanilla is the favorite ice cream flavor for 30% of Americans.

$p = 0.3$

If we poll 100 people, how many would report vanilla as their favorite ice cream flavor?

$E(X) = np = 100(0.3) = 30$

$E(\hat{p}) = p = 0.3$

$SD(X) = \sqrt{npq} = 4.5$

$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = 0.45$

Approximating a Sample Distribution with a normal model

Assumptions:
- Independence
- Randomization
- 10% condition ← sample size < 10% of population
- Success / Failure condition ← $np > 10$, $nq > 10$

Ex: Let p be the probability that the bracket lands flat.
What can we say about p?

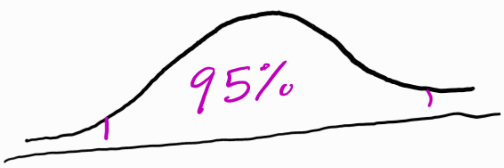
Flipped 10 times, it landed flat 6 times

$\hat{p} = 0.6$ we don't know p and q

$SD(\hat{p}) = \sqrt{\frac{pq}{n}}$ so we use standard error which is an approximation

$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Critical Values

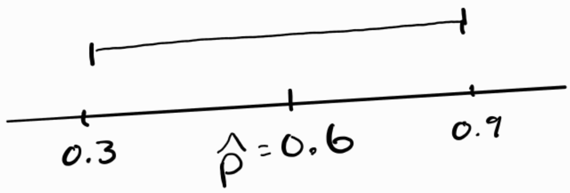


z^*	Confidence Level
1.64	90%
1.96	95%
2.58	99%

Most common

$SE(\hat{p}) = 0.155$ 95% confidence interval

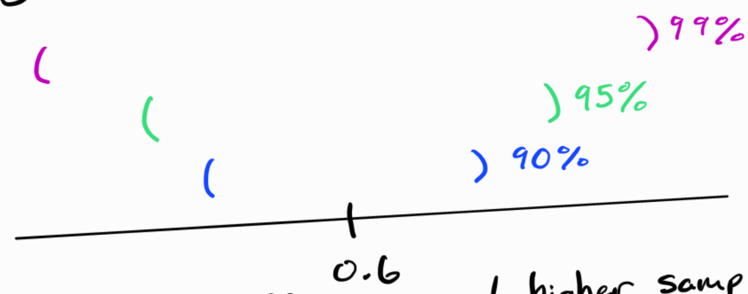
$0.6 \pm 1.96(0.155) = (0.3, 0.9)$



What does a confidence interval mean?

We are 95% confident that this interval contains the true population (p).

Certainty vs. Precision



To minimize trade-off we need higher sample size

Margin of Error $= 0.05 = 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$

Confidence Interval for proportions: p

$\hat{p} \pm z^* SE(\hat{p})$

$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$SD(\hat{p}) = \sqrt{\frac{pq}{n}}$

$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Confidence Interval for means: μ

~~$\bar{x} \pm z^* SE(\bar{x})$~~

$\bar{x} \pm t^* SE(\bar{x})$

$SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

$SE(\bar{x}) = \frac{s}{\sqrt{n}}$

s not great estimate for σ so we don't use first equation

Ex. Find the 90% confidence interval for the length of the twizzlers if:

$n = 30$ $\quad \bar{x} = 8.1$ in. $\quad s = 0.1$ in.

$8.1 \pm t^* SE(\bar{x})$

$8.1 \pm 1.699 \left(\frac{0.1}{\sqrt{30}}\right)$

$df = n - 1$

degrees of freedom

90% $\quad t^*_{29} = 1.699$

$SE(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{0.1}{\sqrt{30}}$

$CI = (8.069 \text{ in}, 8.131 \text{ in})$

Hypothesis Testing

Are the results of an experiment or survey sufficient to change baseline assumptions?

Sample vs. population

Hypothesis Testing:

1) Hypothesis
2) Model
3) Mechanics
4) Conclusion

Definition: P-value
Given the null hypothesis, what is the probability that the observed result happens?

Conclusion:
"cut-off" p-value; convention: 0.05

$\boxed{\text{Failed to reject the null hypothesis.}}$

Ex. 30% of M&M's are supposed to be green, but your bag of 140 only contains 37 green.
Is this evidence that the 30% is wrong?

· Hypothesis:

$H_0: p = 0.3$ ← Null Hypothesis

$H_A: p \neq 0.3$ ← Alternate Hypothesis

Two-sided

· Model: One proportion z-test

Check Assumptions:

- Randomization Condition
- 10% Condition
- Success/Failure $\begin{pmatrix} np > 10 \\ nq > 10 \end{pmatrix}$

· Mechanics: Goal: Calculate the P-value

Assume: $H_0: p = 0.3$

$E(\hat{p}) = 0.3$ ← Assuming we know $p = 0.3$

$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.3)(0.7)}{140}} \approx 0.039$

$z\text{-score} = \frac{0.264 - 0.3}{0.039} \approx -0.92$

P-value: $P(z \le -0.92) + P(z > 0.92)$
$= 0.178 + 0.178$
$= \boxed{0.36}$

Quantitative Hypothesis Testing

A coffee machine dispenses coffee into paper cups. You're supposed to get 10 ounces of coffee, but the amount varies slightly from cup to cup. Here are the amounts measured in a random sample of 20 cups. Is there evidence that the machine is shortchanging customers?

9.9	9.7	10.0	10.1
9.9	9.6	9.8	9.8
10.0	9.5	9.7	10.1
9.9	9.6	10.2	9.8
10.0	9.9	9.5	9.9

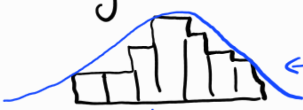
Hypothesis: $H_0: \mu = 10\text{oz}$

$H_A: \mu < 10\text{oz}$ ← One-sided

Model: One-sample t-test

Assumptions:

- Random, Independent Sample
  - ↑ Maybe not random
- Nearly normal condition



Mechanics: Goal: Find P-value

$\bar{x} = 9.845$

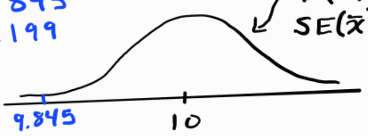
Assuming $H_0: \mu = 10\text{ oz.}$

$\bar{x} = 9.845$

$s = 0.199$

$E(\bar{x}) = 10\text{ oz}$

$SE(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{0.199}{\sqrt{20}} \approx 0.0445$



$t\text{-score} = \frac{\bar{x} - \mu}{SE(\bar{x})} = \frac{9.845 - 10}{0.0445} = -3.49$

$df = n - 1 = 19$

$pt(-3.49, df = 19) = \boxed{0.0012}$

Conclusion:

Reject the null hypothesis

More on P-values

$$P\text{-value} = P\left(\begin{matrix}\text{Observed Statistic} \\ \text{(or more extreme)}\end{matrix} \middle| H_0\right)$$

α: The cutoff value for judging H_0

if p-value $< \alpha$, reject H_0

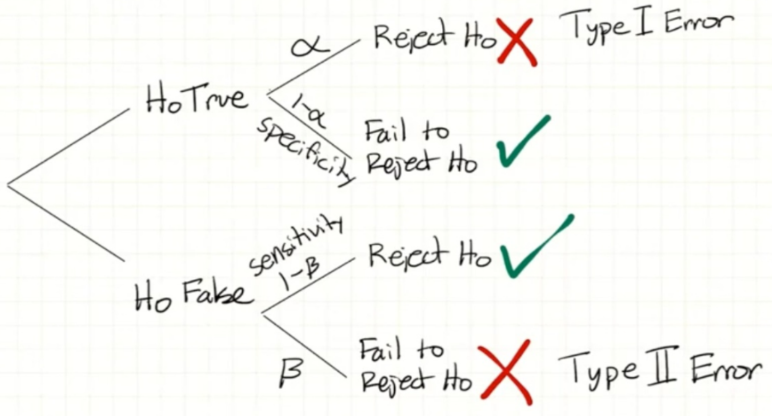
if p-value $> \alpha$, fail to reject H_0

Ex.

Drug X	Drug Y
H_0: Drug X is the same as a placebo. H_A: Drug X is better than a placebo. $\alpha = 0.05$	H_0: Drug Y is the same as a placebo. H_A: Drug Y is better than a placebo. $\alpha = 0.05$
$P = 0.01 < 0.05$; reject H_0	$P = 0.0001 < 0.05$; reject H_0

The p-value only says if we should reject null hypothesis.
We CANNOT say drug Y works better than drug X with only the p-value.

Error – Chapter 19



Decrease α $\Rightarrow$ Increase β

Increase n $\Rightarrow$ Decrease β,
No Effect on α

Comparing Groups

Ex: p_1 = The percentage of M&M's that are green

p_2 = The percentage of Skittles that are green

Comparing Means

	Categorical	Quantitative
Single Population	1 proportion z-test $H_0: p = 0.4$	1 sample t-test $H_0: \mu = 7.9$ inches
Two populations	2 proportion z-test $H_0: p_1 - p_2 = 0$	2 sample (unpaired) t-test $H_0: \mu_1 - \mu_2 = 0$ Paired t-test

Unpaired vs. Paired

Ex: Fuel Economy

Cars	Trucks
40	15
35	20
96	30
60	45

Ex: Fuel Economy

	w/ chains	w/o chains
car 1	30	34
car 2	50	51
truck 1	15	17
car 3	45	41
⋮		

(Unpaired) 2-sample t-test

Ex: Do cars have better fuel economy (mpg) than trucks?

μ_1 = Average mpg for cars

μ_2 = Average mpg for trucks

Sample:

$n_1 = 10$ $\quad \bar{x} = 35$ $\quad s_1 = 15$

$n_2 = 8$ $\quad \bar{x}_2 = 25$ $\quad s_2 = 10$

$H_0: \mu_1 - \mu_2 = 0$

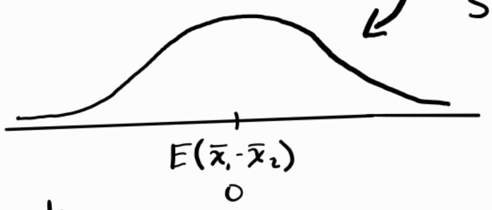
$H_A: \mu_1 - \mu_2 > 0$

Assume H_0, what do we expect of $\bar{x}_1 - \bar{x}_2$

$$SD(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

df gets weird here



2-Sample t-test Assumptions

- Independence Assumption (within groups)
- Nearly Normal Condition
- Randomization Condition
- Independent Group Assumption

Paired t-test

	w/ chains	w/o chains	Difference
car 1	30	34	+4
car 2	50	51	+1
truck 1	15	17	+2
car 3	45	45	+0
⋮			

$\bar{d}$ = mean of difference

s_d = Standard deviation of difference

$H_0: \mu_d = 0$

$H_A: \mu_d > 0$

Paired t-test assumptions

- Independence Assumption (within groups)
- Nearly Normal Condition
- Randomization Condition
- Paired Data Assumption

Confidence Intervals (95%)

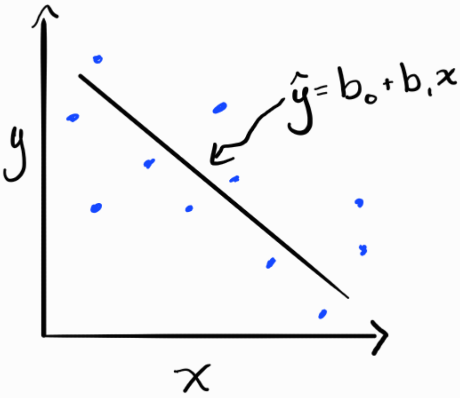
Sample: $n = 10$ $\quad \bar{d} = 1.3$ $\quad s_d = 0.4$

$1.3 \pm t^* SE$

$1.3 \pm (1.812)\left(\frac{0.4}{\sqrt{10}}\right) = (1.071,\ 1.529)$

Inference for Regression

Q: Is our sample evidence of a trend?



Notation:

population	sample
μ	$\bar{x}$
σ	s
p	$\hat{p}$
$\boxed{\beta_1}$	$\boxed{b_1}$

Trend

$H_0: \beta_1 = 0$ ← There is no trend

$H_A: \beta_1 \neq 0$ ← There is a trend

Mechanics:

$$t = \frac{b_1 - \beta_1}{SE(b_1)}$$

$$SE(b_1) = \frac{S_e}{\sqrt{n-1}\; S_x}$$

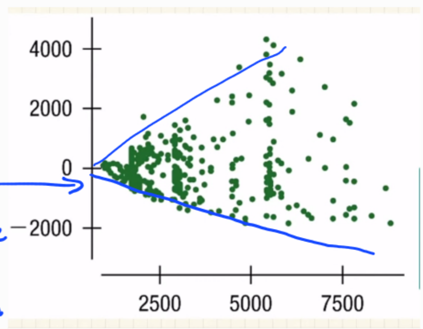
S_e ← standard deviations of difference in each data point (hard to do by hand)

S_x ← standard deviation of just x-values

$$df = n - 2$$

Assumptions

- Linearity Assumption
- Independence Assumption
- Equal Variance Assumption ← Equal variance between data sets
- Normal Population Assumption



Not equal var because spread of y is different with different x-values

instead of

or or =

Comparing Counts: χ^2 Test

Ex:

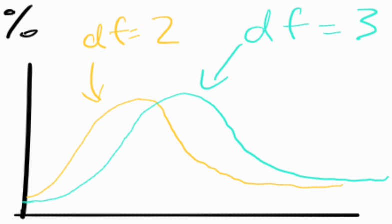
	Birth Season Winter	Spring	Summer	Fall	$n=31$
Observed:	3	8	10	10	
Expected:	7.75	7.75	7.75	7.75	

"Goodness of Fit" Test

H_0: "Even Distribution" aka 25%, 25%, 25%, 25%

$$\boxed{\chi^2 = \sum_{\text{all cells}} \frac{(\text{Obs}-\text{Exp})^2}{\text{Exp}} \qquad df = (\#\text{ of cells} - 1)}$$

Only w/ one row

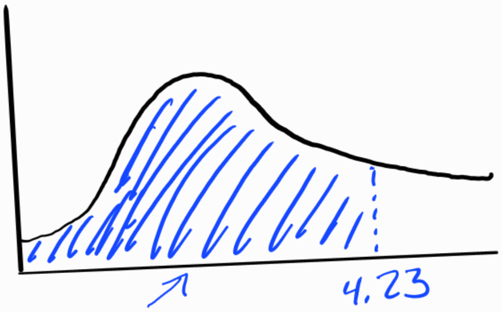


$$\chi^2 = \frac{(3-7.75)^2}{7.75} + \frac{(8-7.75)^2}{7.75} + \frac{(10-7.75)^2}{7.75} + \frac{(10-7.75)^2}{7.75} = \boxed{4.23}$$

$df = 3$

Python: from scipy.stats import chi2

chi2.cdf(x, df=3)



p-value: 1-chi2.cdf(4.23, df=3) = 0.24 > 0.05

Fail to reject null hypothesis

Assumptions:
- Working w/ counts (Not proportions)
- Each cell's expected count ≥ 5

χ^2 Test for independence/Homogeneity

Ex:

$\text{Exp} = 42\left(\frac{19}{114}\right)$

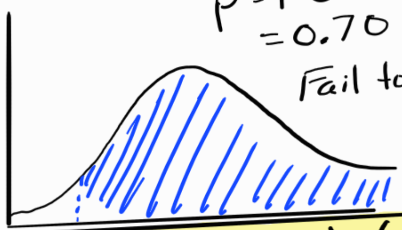
Eye color	Handedness Left	Right	Total
Brown	6 7	36 35	42
Blue	7 5.5	26 27.5	33
Other	6 6.5	33 32.5	39
Total	19	95	114

H_0: Handedness and Eye Color are independent

$$\chi^2 = \sum_{\text{cells}} \frac{(\text{Obs}-\text{Exp})^2}{\text{Exp}} = \frac{(7-6)^2}{6} + \ldots = 0.71$$

$p = 1 - \text{chi2.cdf}(0.71, df=2)$
$= 0.70$

Fail to reject H_0



$$\boxed{df = \left(\#\text{ of rows} - 1\right)\left(\#\text{ of columns} - 1\right)}$$

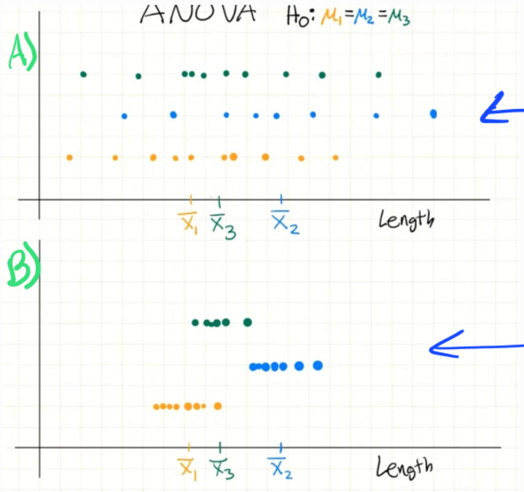
If multiple rows

Big χ^2 ⇒ small p-value ⇒ Reject H_0

Small χ^2 ⇒ big p-value ⇒ Retain H_0

ANOVA (Analysis Of Variance)

	Categorical	Quantitative
Single Population	1 proportion z-test $H_0: p = 0.4$	1 sample t-test $H_0: \mu = 7.9$ inches
Two populations	2 proportion z-test $H_0: p_1 - p_2 = 0$	2 sample (unpaired) t-test $H_0: \mu_1 - \mu_2 = 0$ Paired t-test $H_0: \mu_d = 0$
Many Populations	χ^2 Warning: Counts not proportions	ANOVA $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots$



More overlap

Less overlap (x_1 and x_2 have none)

Comparing where the variance comes from

Measurement:

$$\text{F-statistic} = \frac{MS_T}{MS_E}$$

MS_T: Variance between groups

MS_E: Variance within groups

A has smaller F-stat; large p-value

B has bigger F-stat; small p-value

Assumptions:

- Independence Assumptions
- Equal Variance Assumption
- Nearly Normal Assumption

May need to change scale of datasets (n^2, log n, etc.)

Bayesian Statistics

Frequentist: The population parameter (θ) (aka the true value) is fixed but unknown.

Probability is used to describe the likelihood of observing certain data in a sample given.

$$P(\text{Data} \mid \theta)$$

Bayesian: Probability is used to represent our "degree of belief" about the value of a parameter. Since θ is unknown, we assign a probability distribution to it.

$$P(\theta \mid \text{Data}) = \frac{P(\text{Data} \mid \theta)\, P(\theta)}{P(\text{Data})}$$

Prior

Bayes' Rule
(Probability)

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Likelihood

Bayesian Statistics

$$P(\theta \mid \text{Data}) = \frac{P(\text{Data} \mid \theta)\, P(\theta)}{P(\text{Data})}$$

Probability distribution function

Posterior

Ex: A = Bus running late
B = Raining

Bayesian Statistics with Discrete Priors

Ex. 5 coins in a drawer:

Two "Type A": Probability of heads = 0.5
Two "Type B": Probability of heads = 0.6
One "Type C": Probability of heads = 0.9

I select one at random, flip it, get "heads"
Which was it?

D = Data = "get heads"

$P(A \mid D)$, $P(B \mid D)$, $P(C \mid D)$ ⟵ Posteriors

Prior: $P(A) = 0.4$
$P(B) = 0.4$
$P(C) = 0.2$

Likelihood: $P(D \mid A) = 0.5$
$P(D \mid B) = 0.6$
$P(D \mid C) = 0.9$

$$P(D) = P(A)P(D \mid A) + P(B)P(D \mid B) + P(C)P(D \mid C)$$
$$= (0.4)(0.5) + (0.4)(0.6) + (0.2)(0.9) = 0.62$$

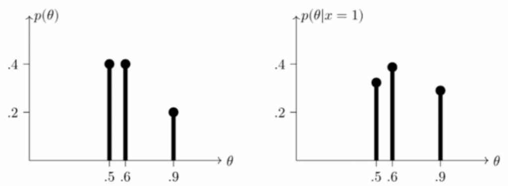
$$P(A \mid D) = \frac{P(D \mid A)\, P(A)}{P(D)} = \frac{(0.5)(0.4)}{0.62} \approx 0.3226$$

$P(B \mid D) = \ldots \approx 0.3871$

$P(C \mid D) = \ldots \approx 0.2903$

Hypothesis	θ	prior pmf $p(\theta)$	posterior pmf $p(\theta \mid x = 1)$
A	0.5	$P(A) = p(0.5) = 0.4$	$P(A \mid \mathcal{D}) = p(0.5 \mid x = 1) = 0.3226$
B	0.6	$P(B) = p(0.6) = 0.4$	$P(B \mid \mathcal{D}) = p(0.6 \mid x = 1) = 0.3871$
C	0.9	$P(C) = p(0.9) = 0.2$	$P(C \mid \mathcal{D}) = p(0.9 \mid x = 1) = 0.2903$

Here are plots of the prior and posterior pmf's from the example.



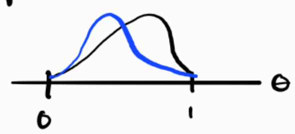
Prior pmf $p(\theta)$ and posterior pmf $p(\theta \mid x = 1)$ for Example 1

Bayesian Statistics w/ Continuous Priors

Ex. We have a coin with an unknown probability of heads θ.

What do we think this θ is?

Recall: Probability Density Functions:

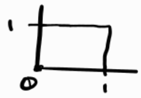


Bayes Theorem

pdf's

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{\int_a^b P(D|\theta)P(\theta)\,d\theta}$$

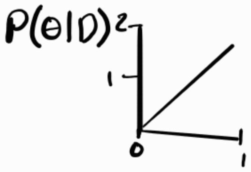
Ex. "Flat prior": all possibilities are equally likely



$P(\theta)=1 \quad 0 \le x \le 1$

I flip it and get "heads"

$$P(\theta|D) = \frac{\theta \cdot 1}{\int_0^1 \theta\,d\theta} = \frac{\theta}{\frac{\theta^2}{2}\Big|_0^1} = \frac{\theta}{1/2} = 2\theta$$



$$P(\theta > 0.5) = 0.5$$

$$P(\theta > 0.5|D) = \int_{0.5}^1 2\theta\,d\theta = \frac{3}{4}$$

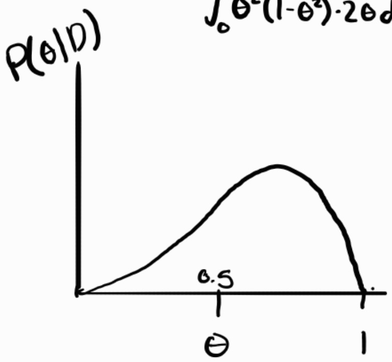
← More likely we are on higher end from our perspective

Ex. Suppose the prior is $P(\theta) = 2\theta$, ← Baseline assumption

and you flip 3 coins and get HHTT

Binomial dist. $P(D|\theta)$

$$P(\theta|D) = \frac{\theta^2(1-\theta^2)\cdot 2\theta}{\int_0^1 \theta^2(1-\theta^2)\cdot 2\theta\,d\theta} = 60\theta^3(1-\theta)^2$$



$$P(\theta > 0.05|D) = \int_{0.5}^1 60\theta^3(1-\theta)^2\,d\theta$$

$$= 0.65625$$

Beta(4, 3)

$\alpha = \beta + 1$

Beta Function

$$\text{Beta}(\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx}$$

Law of total probability: $P(B) = \sum_i P(B|A) \cdot P(A_i)$

B = data

Prior odds: $O_{prior} = \frac{P(A)}{P(\neg A)}$

Posterior odds: $O_{posterior} = \frac{P(A|B)}{P(\neg A|B)}$

Bayes Factor: $\frac{P(B|A)}{P(B|\neg A)}$ ← Now, when do we use this?

Odds update rule: $O_{posterior} = O_{prior} \cdot$ Bayes Factor