PROBABILITY: AN INTERACTIVE REVIEW, II

CS1951A INTRO TO DATA SCIENCE

DAN POTTER

teaching datascience .org

CONTENT CREDIT

Many of of these slides include content produced by

Jeremy Orloff

Jonathan Bloom

MIT Course Number
18.05
As Taught In Spring 2014

OUTLINE

- Warm-Up
  - Statistical Data Mining
  - Probability Trees
- Bayes Wrap-Up
- Random Variables
- LoLN and CLT

STATISTICAL DATA MINING

- Assume a statistical inference has a chance of 1% to be wrong
- How many test can you run before the likelihood of being wrong at least once is 50% or more?

STATISTICAL DATA MINING

- Assume a statistical inference has a chance of 1% to be wrong
- How many test can you run before the likelihood of being wrong at least once is 50% or more?

Let P(F)=P("test incorrect")=.1, P(T)=P("test correct") =.99

P("wrong at least once") = ?

STATISTICAL DATA MINING

- Assume a statistical inference has a chance of 1% to be wrong
- How many test can you run before the likelihood of being wrong at least once is 50% or more?

Let P(F)=P("test incorrect")=.1, P(T)=P("test correct") =.99

$$P(T_1) = 1 - P(F_1) = 0.99$$

$$P\left(T_1 T_2\right) = 0.99 \times 0.99 \approx 0.98$$

$$P\left(F_1 F_2, F_1 T_2, T_1 F_2\right) = 1 - P\left(T_1 T_2\right) \approx 0.02$$

$$P\left(\text{Being Wrong At Least Once}\right) =$$

$$1 - P\left(T_1 T_2 \ldots T_{n-1} T_n\right) = 1 - P\left(T\right)^n$$

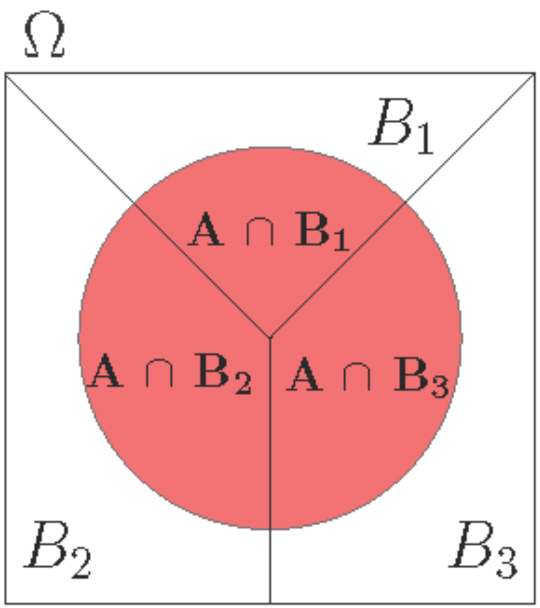
$$n = \log(0.5)/\log(0.99) \approx 68.96 \approx 69$$

Multiplication Rule, Law of Total Probability

MR: $P(A \cap B) = P(A|B) \cdot P(B)$.

LoTP: If B_1, B_2, B_3 *partition* Ω then

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \end{aligned}$$

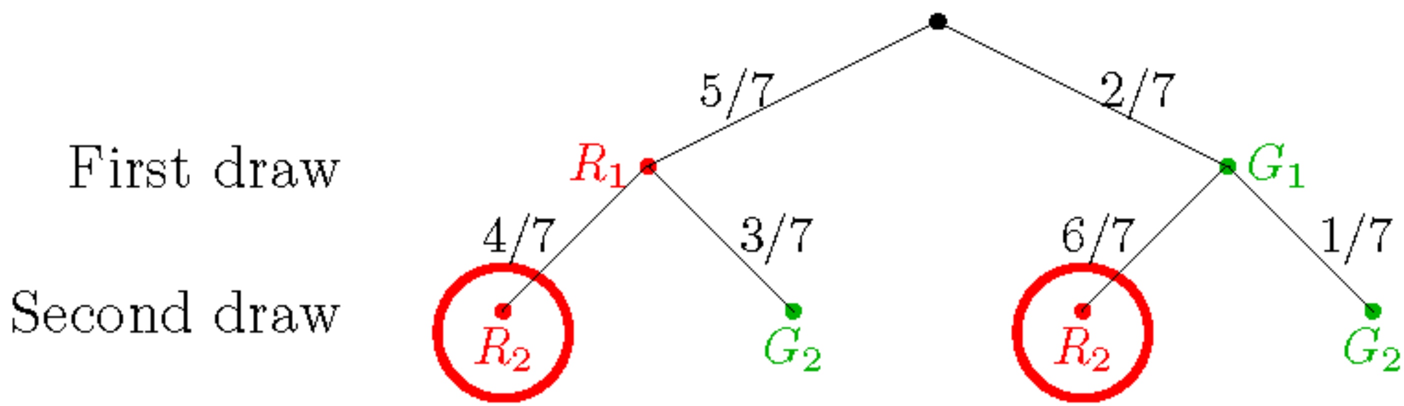


Trees

- Organize computations
- Compute total probability
- Compute Bayes formula

Example. : Game: 5 red and 2 green balls in an urn. A random ball is selected and replaced by a ball of the other color.

1. What is the probability the second ball is red?
2. What is the probability the first ball was red given the second ball was red?

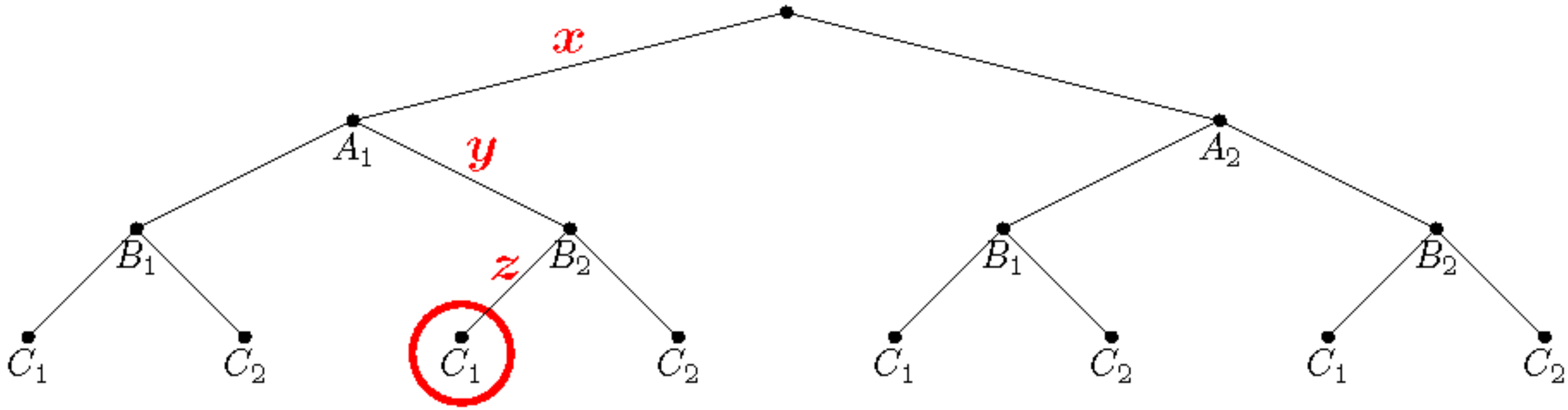


Solution

1. The law of total probability gives $P(R_2) = \frac{5}{7} \cdot \frac{4}{7} + \frac{2}{7} \cdot \frac{6}{7} = \frac{32}{49}$

2. Bayes rule gives $P(R_1|R_2) = \frac{P(R_1 \cap R_2)}{P(R_2)} = \frac{20/49}{32/49} = \frac{20}{32}$

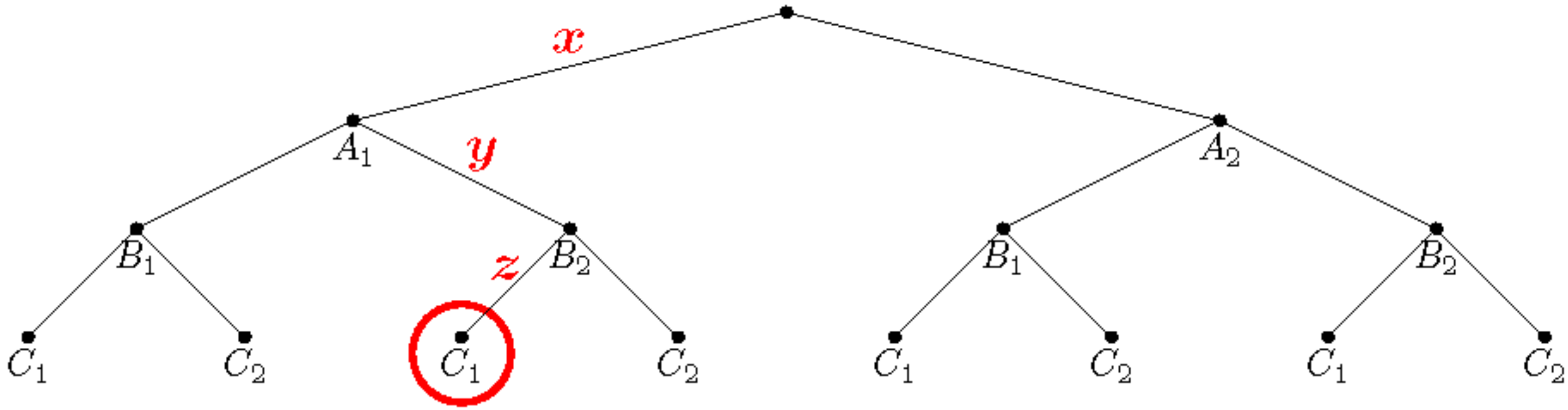
Concept Question: Trees 1



1. The probability x represents

a) $P(A_1)$
b) $P(A_1|B_2)$
c) $P(B_2|A_1)$
d) $P(C_1|B_2 \cap A_1)$.

Concept Question: Trees 1

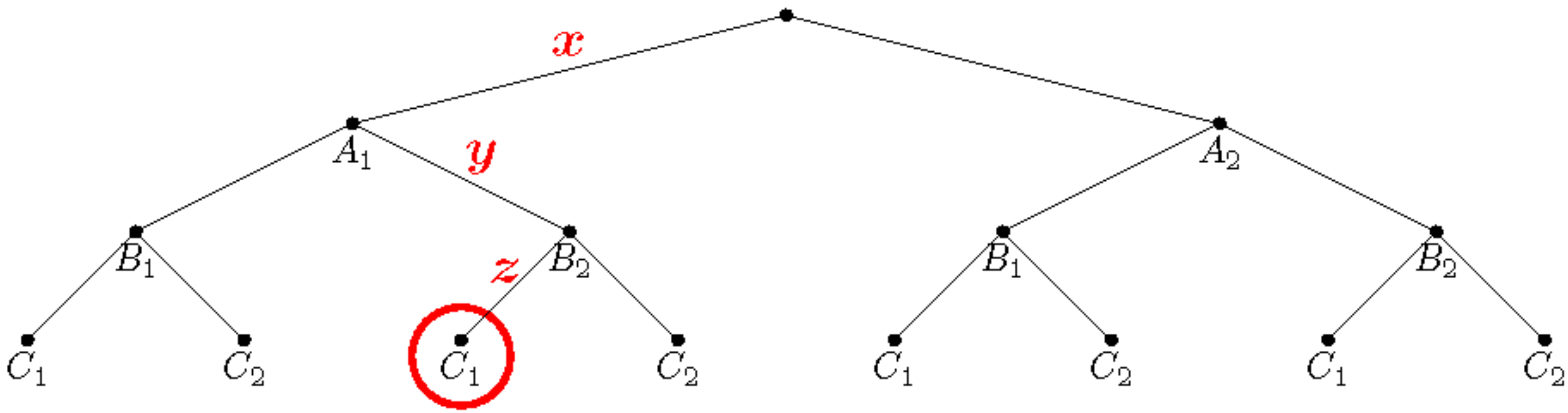


1. The probability x represents

a) $P(A_1)$
b) $P(A_1|B_2)$
c) $P(B_2|A_1)$
d) $P(C_1|B_2 \cap A_1)$.

<u>**answer:**</u> (a) $P(A_1)$.

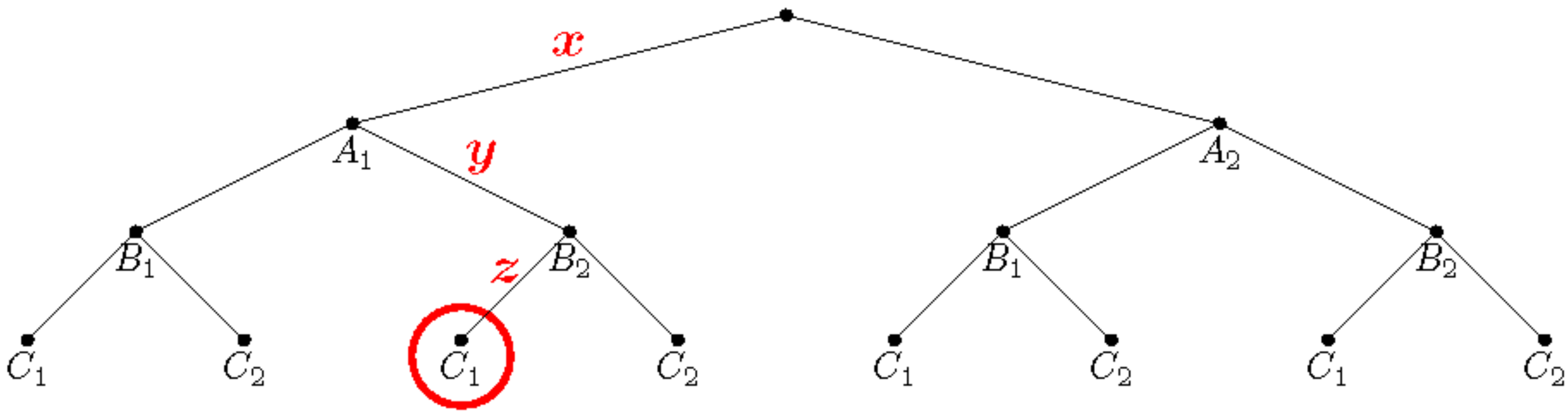
Concept Question: Trees 2



2. The probability y represents

a) $P(B_2)$
b) $P(A_1|B_2)$
c) $P(B_2|A_1)$
d) $P(C_1|B_2 \cap A_1)$.

Concept Question: Trees 2

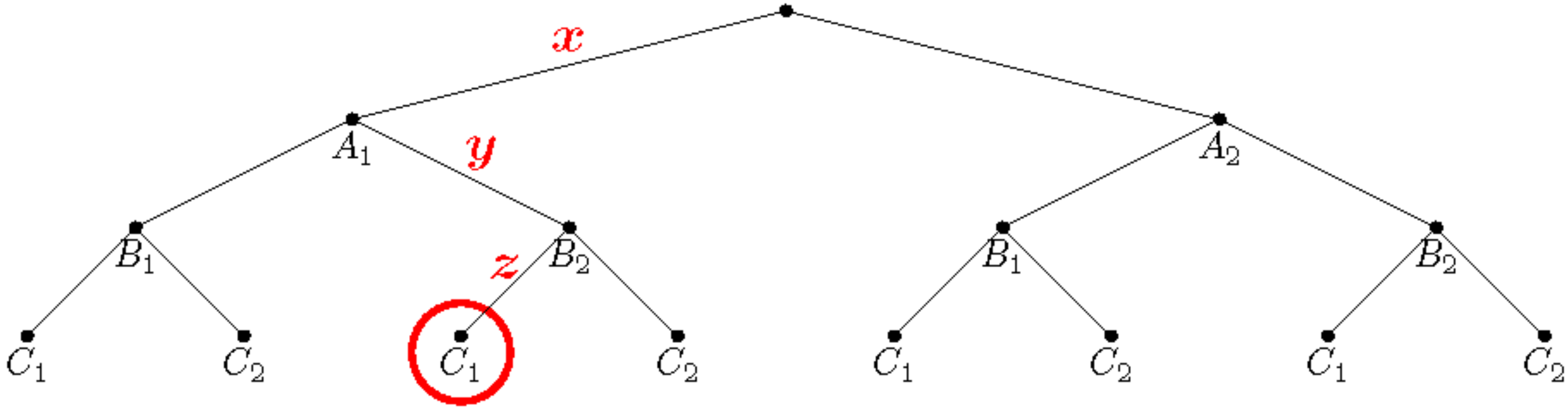


2. The probability y represents

a) $P(B_2)$
b) $P(A_1|B_2)$
c) $P(B_2|A_1)$
d) $P(C_1|B_2 \cap A_1)$.

answer: (c) $P(B_2|A_1)$.

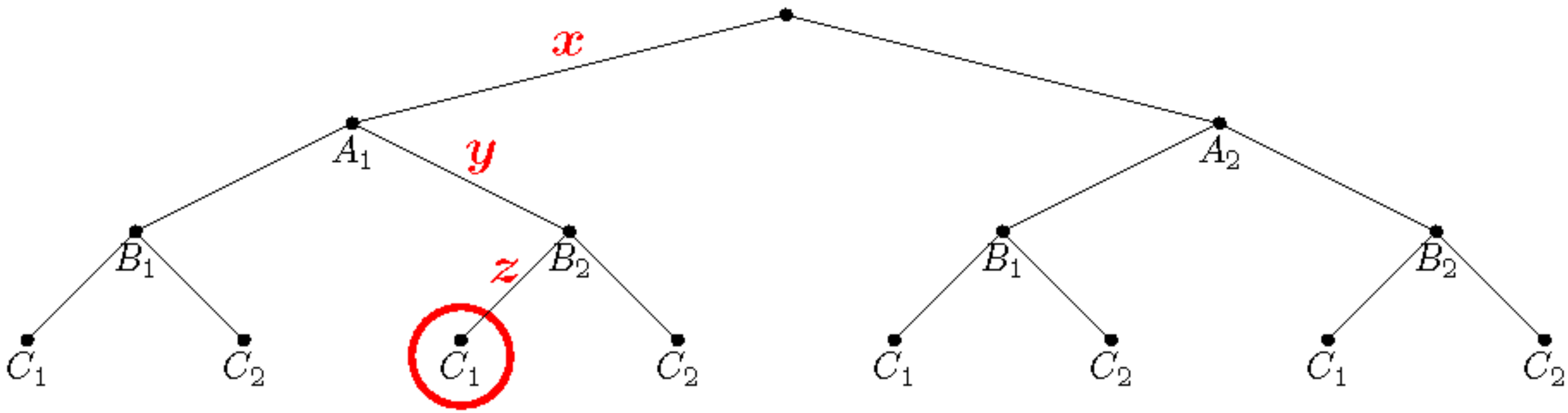
Concept Question: Trees 3



3. The probability z represents

a) $P(C_1)$
b) $P(B_2|C_1)$
c) $P(C_1|B_2)$
d) $P(C_1|B_2 \cap A_1)$.

Concept Question: Trees 3

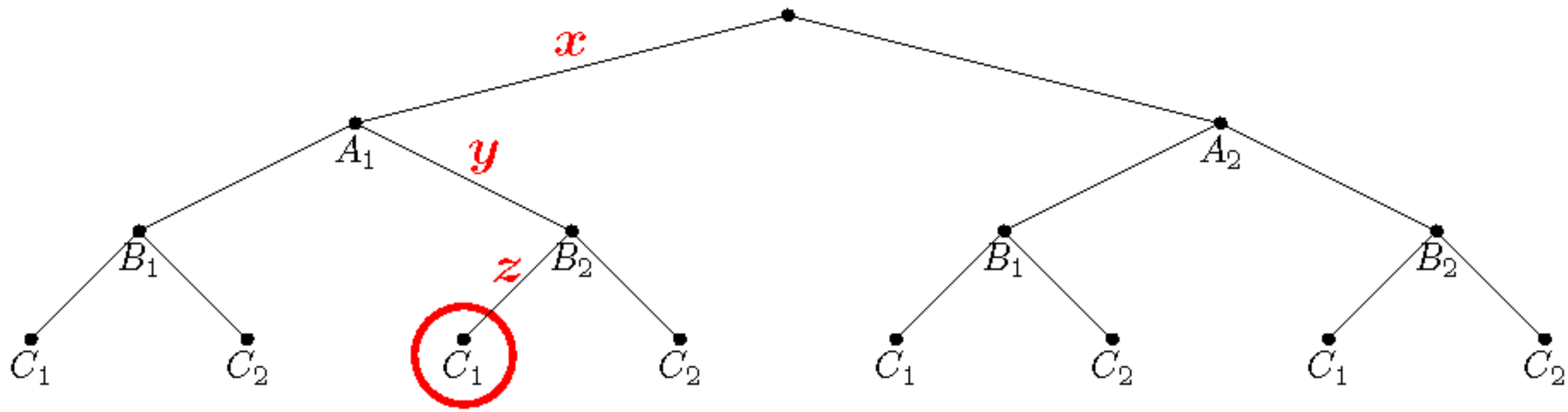


3. The probability z represents

a) $P(C_1)$
b) $P(B_2|C_1)$
c) $P(C_1|B_2)$
d) $P(C_1|B_2 \cap A_1)$.

answer: (d) $P(C_1|B_2 \cap A_1)$.

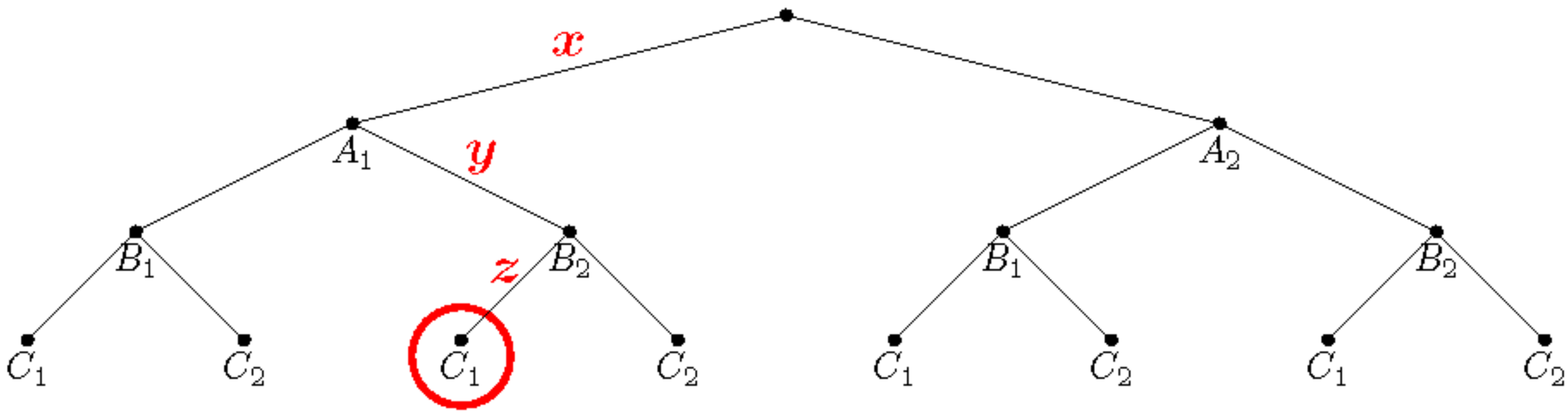
Concept Question: Trees 4



4. The circled node represents the event

a) C_1
b) $B_2 \cap C_1$
c) $A_1 \cap B_2 \cap C_1$
d) $C_1 | B_2 \cap A_1$.

Concept Question: Trees 4



4. The circled node represents the event

a) C_1
b) $B_2 \cap C_1$
c) $A_1 \cap B_2 \cap C_1$
d) $C_1 | B_2 \cap A_1$.

answer: (c) $A_1 \cap B_2 \cap C_1$.

MONTY HALL PROBLEM

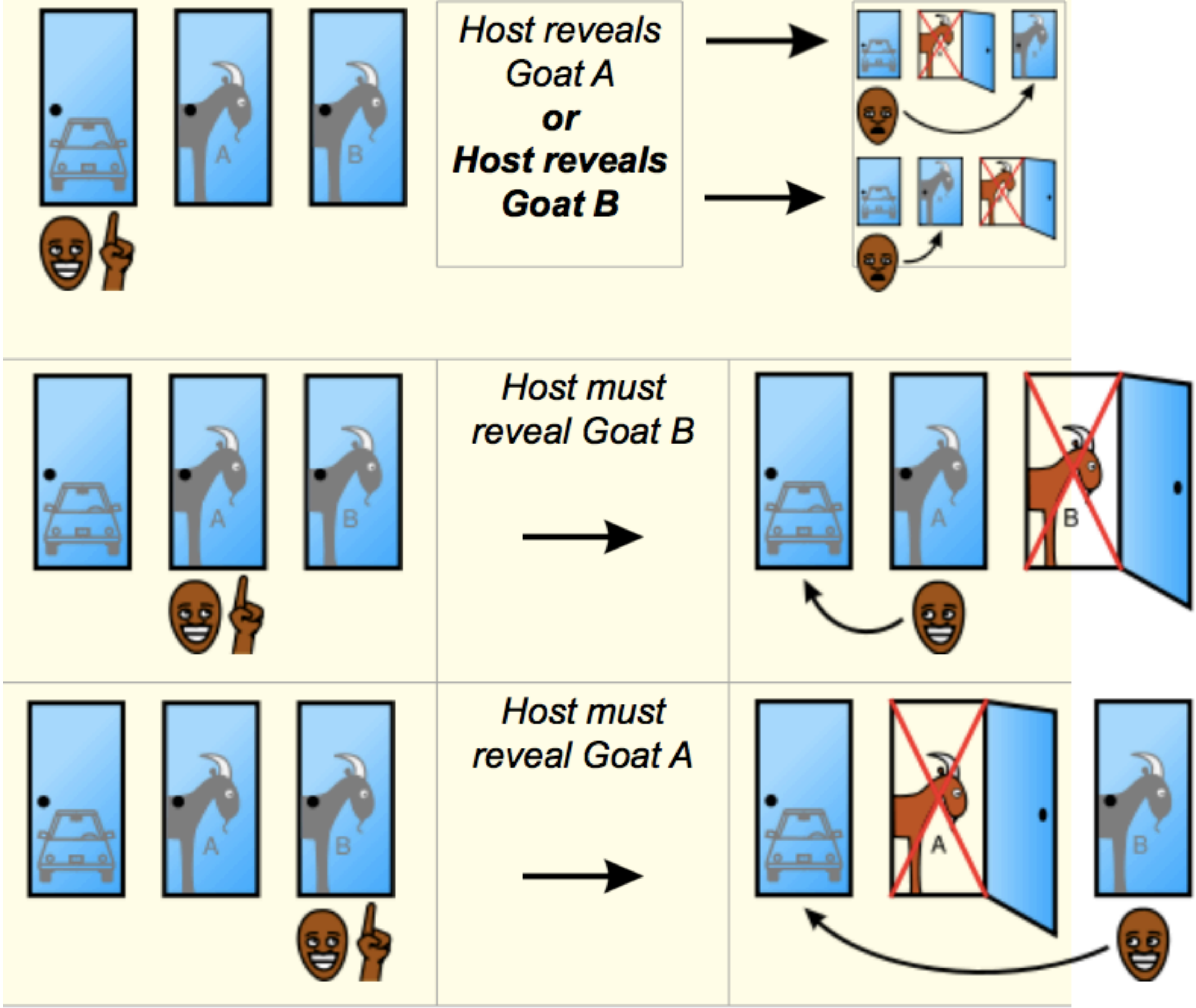
Let's Make a Deal

1. One door hides a car, two hide goats.
2. You choose door.
3. Monty (who knows where the car is) opens a different door with a goat.
4. You can switch doors or keep your original choice.

What is the best strategy for winning a car?

A) Switch
B) Don't switch
C) It doesn't matter

You can play it here:
http://math.ucsd.edu/~crypto/Monty/monty.html



A
B
Host reveals Goat A
or
Host reveals Goat B
Host must reveal Goat B
Host must reveal Goat A

Concept question: Monty Hall

Let's Make a Deal

1. One door hides a car, two hide goats.
2. Contestant chooses door.
3. Monty (who knows where the car is) opens a different door with a goat.
4. Contestant can switch doors or keep her original choice.

What is the best strategy for winning a car?

a) Switch b) Don't switch c) It doesn't matter

Board question: Monty Hall

Organize the Monty Hall problem into a tree and compute the probability of winning if you always switch.

Hint first break the game into a sequence of actions.

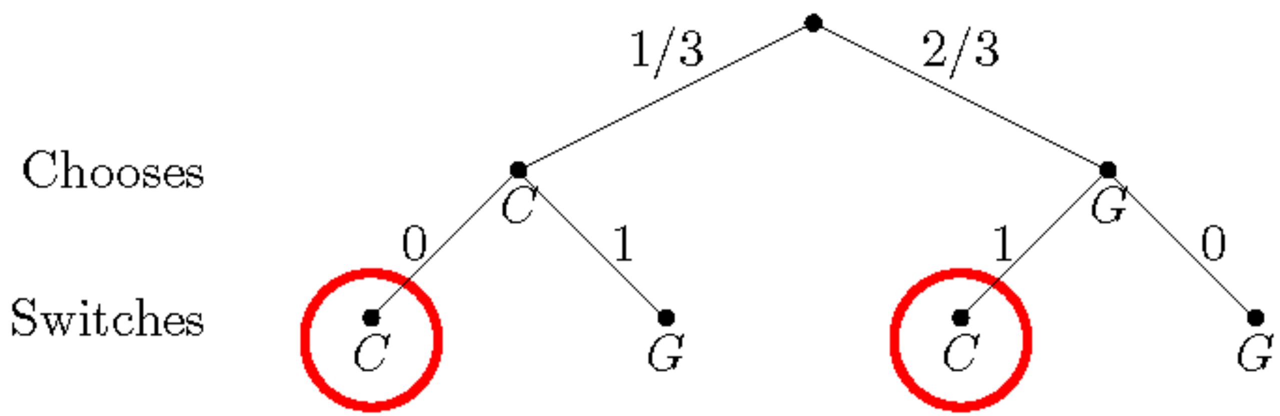
Board question: Monty Hall

Organize the Monty Hall problem into a tree and compute the probability of winning if you always switch.

Hint first break the game into a sequence of actions.

answer: Switch. $P(C|switch) = 2/3$

It's easiest to show this with a tree representing the process: contestant chooses (Monty shows a goat), contestant switches.



The (total) probability of C is $P(C|switch) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$.

BAYESIAN STATISTICS

teaching
datascience
.org

Bayes Theorem

Also called Bayes Rule and Bayes Formula.

Allows you to find $P(A|B)$ from $P(B|A)$, i.e. to 'invert' conditional probabilities.

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

BAYES' LAW

P(D|H): **Likelihood**
Probability of collecting this data when our hypothesis is true

P(H): **Prior**
The probability of the hypothesis being true before collecting data

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

P(H|D): **Posterior**
The probability of our hypothesis being true given the data collected

P(D): **Marginal**
What is the probability of collecting this data under all possible hypotheses?

MONTY HALL PROBLEM: BAYES RULE

You pick door A.
Data D = Monty opened door B

Hypothesis space H:

h_1 = Car is behind door A
h_2 = Car is behind door C
h_3 = Car is behind door B

What is $P(h_1 \mid D)$?
What is $P(h_2 \mid D)$?
What is $P(h_3 \mid D)$?

Prior probability:

$P(h_1) = 1/3$ $P(h_2) = 1/3$ $P(h_3) = 1/3$

Likelihood:

$P(D \mid h_1) = 1/2$
$P(D \mid h_2) = 1$
$P(D \mid h_3) = 0$

$$P(D) = p(D|h_1)p(h_1) + p(D|h_2)p(h_2) + p(D|h_3)p(h_3) = 1/6 + 1/3 + 0 = 1/2$$

By Bayes rule:

$P(h1|D) = P(D|h1)p(h1) / P(D)$
$= ½\ 1/3 / ½ = 1/3$

$P(h2|D) = P(D|h2)p(h2) / P(D)$
$= 1\ 1/3 / ½ = 2/3$

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

BAYES THEOREM
(ALTERNATE FORM)

Let $E_1, E_2, \ldots E_n$ be **mutually disjoint** events in a sample space Ω, and $\bigcup_{i=1}^{n} E_i = \Omega$

Then:

$$\boxed{\Pr(E_j \mid B) = \frac{\Pr(B \mid E_j)\Pr(E_j)}{\sum_{i=1}^{n} \Pr(B \mid E_i)\Pr(E_i)}}$$

$$\text{Conditional Probability: } \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\text{Law of Total Probability: } \Pr(B) = \sum_{j=1}^{n} \Pr(B \mid E_j)\Pr(E_j)$$

APPLICATION: FINDING A BIASED COIN

- We are given three coins. 2 coins are fair, and the 3rd is biased (landing heads with probability 2/3)
- **We need to identify the the biased coin**
- We flip each of the coins. The first and second come up heads, and the third comes up tails
- **What is the probability that the first coin was the biased one?**

APPLICATION: FINDING A BIASED COIN

Let E_i be the event that the *ith* coin flip is the biased one and let B be the event that the three coin flips came up HEADS, HEADS, and TAILS. Before we flip the coins we have $\Pr(E_i) = 1/3$ for $i=1,\ldots,3$, thus

$$\mathrm{P}(B|E_1) = \mathrm{P}(B|E_2) = \frac{2}{3}\times\frac{1}{2}\times\frac{1}{2} = \frac{1}{6}$$

and

$$\Pr(B|E_3) = \frac{1}{2}\times\frac{1}{2}\times\frac{1}{3} = \frac{1}{12}$$

Applying Bayes we have

$$\Pr(E_1|B) = \frac{\mathrm{P}(B|E_1)\mathrm{P}(E_1)}{\sum_{i=1}^{3}\Pr(B|E_i)\Pr(E_i)} = \frac{2}{5}$$

The outcome HHT increases the probability that the first coin is the biased one from 1/3 to 2/5.

EXAMPLE REVISITED

Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

$$BB = \text{Two Boys}, \text{GG} = \text{Two girls}$$

$$B = \text{At least one kid is a boy}$$

$$P(BB) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

$$P(B) = 1 - P(GG) = 1 - \frac{1}{2} * \frac{1}{2} = \frac{3}{4}$$

$$P(BB|B) = \frac{P(BB \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Direct Calculation

$$\boxed{P(BB|B) = \frac{P(B|BB) * P(BB)}{P(B)} = \frac{1 * 1/4}{3/4} = \frac{1}{3}}$$

Bayes Calculation

IN CLASS: DRUG TEST

- 0.4% of Rhode Island PhDs use Marijuana*
- Drug Test: The test will produce 99% true positive results for drug users and 99% true negative results for non-drug users.
- If a randomly selected individual is tested positive, what is the probability he or she is a user?

$$
\begin{aligned}
P(User|+) &= \frac{P(+|User)P(User)}{P(+)} \\
&= \frac{P(+|User)P(User)}{P(+|User)P(User)+P(+|!User)P(!User)} \\
&= \frac{0.99\times 0.004}{0.99\times 0.004+0.01\times 0.996} \\
&= 28.4\%
\end{aligned}
$$

SPAM FILTERING WITH NAÏVE BAYES

$$P(spam|words) = \frac{P(spam)P(words|spam)}{P(words)}$$

$$P(spam|viagra, rich, ..., friend) = \frac{P(spam)P(viagra, rich, ..., friend|spam)}{P(viagra, rich, ..., friend)}$$

$$P(spam|words) \approx \frac{P(spam)P(viagra|spam)P(rich|spam)\ldots P(friend|spam)}{P(viagra, rich, ..., friend)}$$

Washington Post

Annual physical exam is probably unnecessary if you're generally healthy

> *For patients, the negatives include time away from work and possibly unnecessary tests. "Getting a simple urinalysis could lead to a false positive, which could trigger a cascade of even more tests, only to discover in the end that you had nothing wrong with you." Mehrotra says.*

http://www.washingtonpost.com/national/health-science/annual-physical-exam-is-probably-unnecessary-if-youregenerally-healthy/2013/02/08/2c1e326a-5f2b-11e2-a389 ee565c81c565_story.html

Is Officer Drew a Male or Female?

Officer Drew

Luckily, we have a small database with names and sex.

We can use it to apply Bayes rule…

Name	Sex
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

A) P(male|drew) = 0.25 / (3/8)
   P(female|drew) = 0.25 / (3/8)
B) P(male|drew) = 0.125 / (3/8)
   P(female|drew) = 0.250 / (3/8)
C) P(male|drew) = 0.5
   P(female|drew) = 0.5
D) P(male|drew) = 1
   P(female|drew) = 0.5

Officer Drew

$$p(c_j \mid d) = \frac{p(d \mid c_j)\, p(c_j)}{p(d)}$$

Name	Sex
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

$$p(\text{male} \mid drew) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8}$$

$$p(\text{female} \mid drew) = \frac{2/5 * 5/8}{3/8} = \frac{0.250}{3/8}$$

Officer Drew is more likely to be a Female.

Officer Drew

Officer Drew IS a female!

$$p(\text{male} \mid drew) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8}$$

$$p(\text{female} \mid drew) = \frac{2/5 * 5/8}{3/8} = \frac{0.250}{3/8}$$

Officer Drew

Is Officer Drew a Male or Female?

Luckily, we have a small database with names and sex.

We can use it to apply Bayes rule…

Name	Sex
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

A) P(male|drew) = 0.25 / (3/8)
   P(female|drew) = 0.25 / (3/8)
B) P(male|drew) = 0.125 / (3/8)
   P(female|drew) = 0.250 / (3/8)
C) P(male|drew) = 0.5
   P(female|drew) = 0.5
D) P(male|drew) = 1
   P(female|drew) = 0.5

NEXT UP

- Random Variables
- LoLN, CLT

Random Variables

Random variable X assigns a number to each outcome:

$$X : \Omega \to \mathbf{R}$$

"$X = a$" denotes the event $\{\omega \,|\, X(\omega) = a\}$.

Probability mass function (pmf) of X is given by

$$p(a) = P(X = a).$$

Cumulative distribution function (cdf) of X is given by

$$F(a) = P(X \leq a).$$

Example

Suppose X is a random variable with the following table.

values of X:	-2	-1	0	1	2
pmf $p(a)$:	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$
cdf $F(a)$:	$1/5$	$2/5$	$3/5$	$4/5$	$5/5$

The cdf is the probability 'accumulated' from the left.

Examples. $F(-1) = 2/5$, $F(1) = 4/5$, $F(1.5) = 4/5$, $F(-5) = 0$, $F(5) = 1$.

Properties of $F(a)$:

1. Nondecreasing
2. Way to the left, i.e. as $a \to -\infty$), F is 0
3. Way to the right, i.e. as $a \to \infty$, F is 1.

Concept Question: cdf and pmf

X a random variable.

values of X:	1	3	5	7
cdf $F(a)$:	.5	.75	.9	1

1. What is $P(X \leq 3)$?

a) .15 b) .25 c) .5 d) .75

2. What is $P(X = 3)$

a) .15 b) .25 c) .5 d) .75

Concept Question: cdf and pmf

X a random variable.

values of X:	1	3	5	7
cdf $F(a)$:	.5	.75	.9	1

1. What is $P(X \leq 3)$?

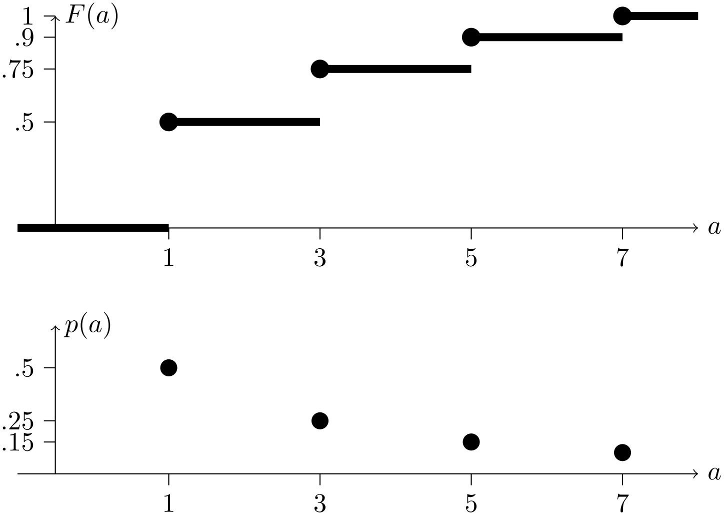
a) .15 b) .25 c) .5 d) .75

2. What is $P(X = 3)$

a) .15 b) .25 c) .5 d) .75

1. **answer:** (d) .75. $P(X \leq 3) = F(3) = .75$.
2. **answer:** (b) $P(X = 3) = .75 - .5 = .25$.

CDF and PMF



Deluge of discrete distributions

$$\text{Bernoulli}(p) = \begin{array}{l} 1 \text{ (success) with probability } p, \\ 0 \text{ (failure) with probability } 1-p. \end{array}$$

In more neutral language:

$$\text{Bernoulli}(p) = \begin{array}{l} 1 \text{ (heads) with probability } p, \\ 0 \text{ (tails) with probability } 1-p. \end{array}$$

$$\text{Binomial}(n,p) = \# \text{ of successes in } n \text{ independent Bernoulli}(p) \text{ trials.}$$

$$\text{Geometric}(p) = \# \text{ of tails before first head in a sequence of indep. Bernoulli}(p) \text{ trials.}$$

(Neutral language avoids confusing whether we want the number of successes before the first failure or vice versa.)

Concept Question

1. Let $X \sim \text{binom}(n, p)$ and $Y \sim \text{binom}(m, p)$ be independent. Then $X + Y$ follows:

a) $\text{binom}(n + m, p)$ b) $\text{binom}(nm, p)$
c) $\text{binom}(n + m, 2p)$ d) other

2. Let $X \sim \text{binom}(n, p)$ and $Z \sim \text{binom}(n, q)$ be independent. Then $X + Z$ follows:

a) $\text{binom}(n, p + q)$ b) $\text{binom}(n, pq)$
c) $\text{binom}(2n, p + q)$ d) other

Concept Question

1. Let $X \sim \text{binom}(n, p)$ and $Y \sim \text{binom}(m, p)$ be independent. Then $X + Y$ follows:

a) $\text{binom}(n + m, p)$ b) $\text{binom}(nm, p)$

c) $\text{binom}(n + m, 2p)$ d) other

2. Let $X \sim \text{binom}(n, p)$ and $Z \sim \text{binom}(n, q)$ be independent. Then $X + Z$ follows:

a) $\text{binom}(n, p + q)$ b) $\text{binom}(n, pq)$

c) $\text{binom}(2n, p + q)$ d) other

1. **answer:** (a). Each binomial random variable is a sum of independent Bernoulli(p random variables, so their sum is also a sum of Bernoulli(p) r.v.'s.

2. **answer:** (d) This is different from problem 1 because we are combining Bernoulli(p) r.v.'s with Bernoulli(q) r.v.'s. This is not one of the named random variables we know about.

Computing expected value

Definition: $$E(X) = \sum_i x_i \, p(x_i)$$

1. $E(aX + b) = aE(X) + b$
2. $E(X + Y) = E(X) + E(Y)$
3. $E(h(X)) = \sum_i h(x_i) \, p(x_i)$

Computing Expectations

1.	X:	-2	-1	0	1	2
2.	pmf:	1/5	1/5	1/5	1/5	1/5

3. $E(X) =$ -2/5 - $1/5 + 0/5 + 1/5 + 2/5 = 0$

4.	X^2:	4	1	0	1	4

5. $E(X^2) = 4/5 + 1/5 + 0/5 + 1/5 + 4/5 = 2$

Line 3 computes $E(X)$ by multiplying the probabilities in line 2 by the values in line 1 and summing.

Line 4 gives the values of X^2.

Line 5 computes $E(X^2)$ by multiplying the probabilities in line 2 by the values in line 4 and summing. This illustrates the use of the formula

$$E(h(X)) = \sum_i h(x_i)\, p(x_i).$$

Continued on the next slide.

Class example continued

Notice that in the table on the previous slide, some values for X^2 are repeated. For example the value 4 appears twice. Summing all the probabilities where $X^2 = 4$ gives $P(X^2 = 4) = 2/5$. Here's the full table for X^2

1.	X^2:		4		1		0		
2.	pmf:		$2/5$		$2/5$		$1/5$		
3.	$E(X^2)$	$=$	$8/5$	$+$	$2/5$	$+$	$0/5$	$=$	2

Here we used the definition of expected value to compute $E(X^2)$. Of course, we got the same expected value $E(X^2) = 2$ as we did earlier.

Board Question: Interpreting Expectation

a) Would you accept a gamble that offers a 10% chance to win \$95 and a 90% chance of losing \$5?

b) Would you pay \$5 to participate in a lottery that offers a 10% percent chance to win \$100 and a 90% chance to win nothing?

• Find the expected value of your change in assets in each case?

Board Question: Interpreting Expectation

a) Would you accept a gamble that offers a 10% chance to win \$95 and a 90% chance of losing \$5?

b) Would you pay \$5 to participate in a lottery that offers a 10% percent chance to win \$100 and a 90% chance to win nothing?

- Find the expected value of your change in assets in each case?

a) E(payoff) = 95*.10-5*.9 = \$5

b) E(payoff) = 100*.10-5 = \$5

Discussion

Framing bias / cost versus loss. The two situations are identical, with an expected value of gaining \$5. In a study, 132 undergrads were given these questions (in different orders) separated by a short filler problem. 55 gave different preferences to the two events. Of these, 42 rejected (a) but accepted (b). One interpretation is that we are far more willing to pay a cost up front than risk a loss. (See *Judgment under uncertainty: heuristics and biases* by Tversky and Kahneman.)

Loss aversion and cost versus loss sustain the insurance industry: people pay more in premiums than they get back in claims on average (otherwise the industry wouldn't be sustainable), but they buy insurance anyway to protect themselves against substantial losses. Think of it as paying \$1 each year to protect yourself against a 1 in 1000 chance of losing \$100 that year. By buying insurance, the expected value of the change in your assets in one year (ignoring other income and spending) goes from negative 10 cents to negative 1 dollar. But whereas without insurance you might lose \$100, with insurance you always lose exactly \$1.

Board Question (skip)

Suppose (hypothetically!) that everyone at your table got up, ran around the room, and sat back down randomly (i.e., all seating arrangements are equally likely).

What is the expected value of the number of people sitting in their original seat?

Neat fact: A permutation in which nobody returns to their original seat is called a derangement. The number of derangements turns out to be the nearest integer to $n!/e$. Since there are $n!$ total permutations, we have:

$$P(\text{everyone in a different seat}) \approx \frac{n!/e}{n!} = 1/e \approx 0.3679.$$

It's surprising that the probability is about 37% regardless of n, and that it converges to $1/e$ as n goes to infinity

Solution (skip)

Number the people from 1 to n. Let X_i be the Bernoulli random variable with value 1 if person i returns to their original seat and value 0 otherwise. Since person i is equally likely to sit back down in any of the n seats, the probability that person i returns to their original seat is $1/n$. Therefore $X_i \sim Bernoulli(1/n)$ and $E(X_i) = 1/n$. Let X be the number of people sitting in their original seat following the rearrangement. Then

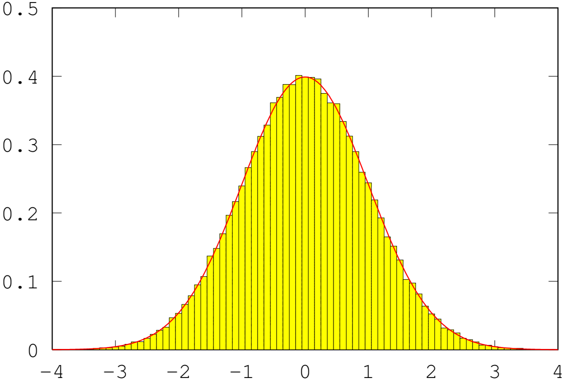
$$X = X_1 + X_2 + \cdots + X_n.$$

By linearity of expected values, we have

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} 1/n = 1.$$

- It's neat that the expected value is 1 for any n.
- If $n = 2$, then both people either retain their seats or exchange seats. So $P(X = 0) = 1/2$ and $P(X = 2) = 1/2$. In this case, X never equals $E(X)$.
- The X_i are not independent (e.g. for $n = 2$, $X_1 = 1$ implies $X_2 = 1$).
- Expectation behaves linearly even when the variables are dependent.

Continuous Expectation and Variance, the Law of Large Numbers, and the Central Limit Theorem \



Expected value

Expected value: measure of location, central tendency

X continuous with range $[a, b]$ and pdf $f(x)$:

$$E(X) = \int_a^b x f(x)\, dx.$$

X discrete with values $x_1, \ldots, x_n$ and pmf $p(x_i)$:

$$E(X) = \sum_{i=1}^{n} x_i p(x_i).$$

View these as essentially the same formulas.

Variance and standard deviation

Standard deviation: measure of spread, scale

For *any* random variable X with mean μ

$$\mathrm{Var}(X) = E((X-\mu)^2), \qquad \sigma = \sqrt{\mathrm{Var}(X)}$$

X continuous with range $[a, b]$ and pdf $f(x)$:

$$\mathrm{Var}(X) = \int_a^b (x-\mu)^2 f(x)\, dx.$$

X discrete with values $x_1, \ldots, x_n$ and pmf $p(x_i)$:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i).$$

View these as essentially the same formulas.

Properties

Properties:

1. $E(X + Y) = E(X) + E(Y)$.
2. $E(aX + b) = aE(X) + b$.

1. If X and Y are independent then
   $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.
2. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
3. $\mathrm{Var}(X) = E(X^2) - E(X)^2$.

Board question

The random variable X has range [0,1] and pdf cx^2.

a) Find c.

b) Find the mean, variance and standard deviation of X.

c) Find the median value of X.

d) Suppose $X_1, \ldots X_{16}$ are independent identically-distributed copies of X. Let $\overline{X}$ be their average. What is the standard deviation of $\overline{X}$?

e) Suppose $Y = X^4$. Find the pdf of Y.

answer: *See next slides.*

Solution

a) Total probability is 1: $\displaystyle\int_0^1 cx^2\,dx = 1 \;\Rightarrow\; \boxed{c = 3}$.

b) $\mu = \int_0^1 3x^3\,dx = 3/4$.
$\sigma^2 = (\int_0^1 (x-3/4)^2\,3x^2\,dx) = \frac{3}{5} - \frac{9}{8} + \frac{9}{16} = \frac{3}{80}$.
$\sigma = \sqrt{3/80} = \frac{1}{4}\sqrt{3/5} \approx .194$

c) Set $F(q_{.5}) = .5$, solve for $q_{.5}$: $F(x) = \displaystyle\int_0^x 3u^2\,du = x^3$. Therefore,
$F(q_{.5}) = q_{.5}^3 = .5$. We get, $\boxed{q_{.5} = (.5)^{1/3}}$.

d) Because they are independent
$\mathrm{Var}(X_1 + \ldots + X_{16}) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_{16}) = 16\mathrm{Var}(X)$.
Thus, $\mathrm{Var}(\overline{X}) = \frac{16\mathrm{Var}(X)}{16^2} = \frac{\mathrm{Var}(X)}{16}$. Finally, $\sigma_{\overline{X}} = \boxed{\dfrac{\sigma_X}{4} = .194/4}$.

Solution continued

e) **Method 1 use the cdf:**
$F_Y(y) = P(X^4 < y) = P(X < y^{\frac{1}{4}}) = F_X(y^{\frac{1}{4}}) = y^{\frac{3}{4}}$.
Now differentiate. $f_Y(y) = F'_Y(y) = \boxed{\frac{3}{4} y^{-\frac{1}{4}}}$.

Method 2 use the pdf: We have

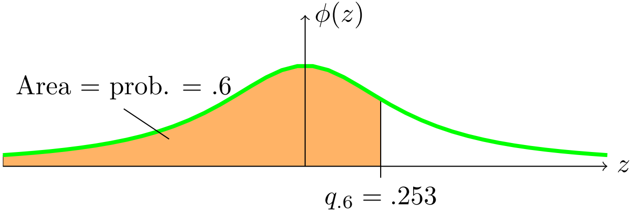
$$y = x^4 \;\Rightarrow\; dy = 4x^3\,dx \;\Rightarrow\; \frac{dy}{4y^{3/4}} = dx$$

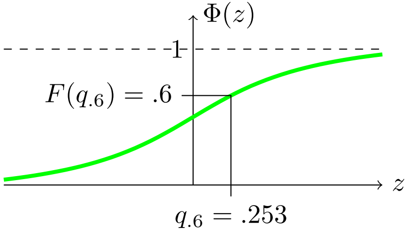
This implies $$f_X(x)\,dx = f_X(y^{1/4})\frac{dy}{4y^{3/4}} = \frac{3y^{2/4}\,dy}{4y^{3/4}} = \frac{3}{4y^{1/4}}\,dy$$

Therefore $$f_Y(y) = \frac{1}{4y^{1/4}}$$

Quantiles

Quantiles give a measure of **location.**



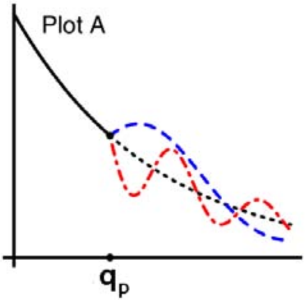


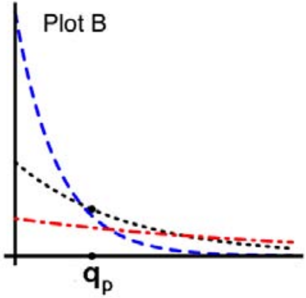
$q_{.6}$: *left tail area* $= .6 \;\Leftrightarrow\; F(q_{.6}) = .6$

Concept question

In each of the plots some densities are shown. The median of the black plot is always at q_p. In each plot, which density has the greatest median?

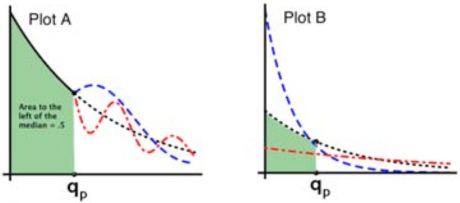
1. Black
2. Red
3. Blue
4. All the same
5. Impossible to tell





answer: *See next frame.*

Solution



Plot A: $\boxed{\text{4. All three medians are the same.}}$ Remember that probability is computed as the area under the curve. By definition the median q_p is the point where the shaded area in Plot A .5. Since all three curves coincide up to q_p. That is, the shaded area in the figure is represents a probability of .5 for all three densities.

Plot B: $\boxed{\text{2. The red density has the greatest median.}}$ Since q_p is the median for the black density, the shaded area in Plot B is .5. Therefor the area under the blue curve (up to q_p) is greater than .5 and that under the red curve is less than .5. This means the median of the blue density is to the left of q_p (you need less area) and the median of the red density is to the right of q_p (you need more area).

Law of Large Numbers (LoLN)

Informally: An average of many measurements is more accurate than a single measurement.

Formally: Let X_1, X_2, ... be i.i.d. random variables all with mean μ and standard deviation σ.
Let

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then for any (small number) a, we have

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < a) = 1.$$

Concept Question: Desperation

- You have \$100. You need \$1000 by tomorrow morning.
- Your only way to get it is to gamble.
- If you bet \$k, you either win \$k with probability p or lose \$k with probability $1 - p$.

Maximal strategy: Bet as much as you can, up to what you need, each time.
Minimal strategy: Make a small bet, say \$5, each time.

1. If $p = .45$, which is the better strategy?

A. Maximal B. Minimal

Concept Question: Desperation

- You have \$100. You need \$1000 by tomorrow morning.
- Your only way to get it is to gamble.
- If you bet \$k, you either win \$k with probability p or lose \$k with probability $1 - p$.

Maximal strategy: Bet as much as you can, up to what you need, each time.
Minimal strategy: Make a small bet, say \$5, each time.

1. If $p = .45$, which is the better strategy?
   A. Maximal B. Minimal

2. If $p = .8$, which is the better strategy?
   A. Maximal B. Minimal

answer: *On next slide*

Solution to previous two problems

answer: $p = .45$ use maximal strategy; $p = .8$ use minimal strategy.
If you use the minimal strategy the law of large numbers says your average winnings per bet will almost certainly be the expected winnings of one bet. The two tables represent $p = .45$ and $p = .8$ respectively.

Win	-10	10
p	.55	.45

Win	-10	10
p	.2	.8

The expected value of a \$5 bet when $p = .45$ is -\$0.50 Since on average you will lose \$0.50 per bet you want to avoid making a lot of bets. You go for broke and hope to win big a few times in a row. It's not very likely, but the maximal strategy is your best bet.

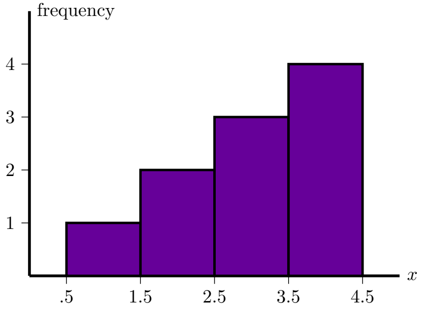
The expected value when $p = .8$ is \$3. Since this is positive you'd like to make a lot of bets and let the law of large numbers (practically) guarantee you will win an average of \$6 per bet. So you use the minimal strategy.

Histograms

Made by 'binning' data.

Frequency: *height* of bar over bin = number of data points in bin.

Density: *area* of bar is the fraction of all data points that lie in the bin. So, total area is 1.



Board question

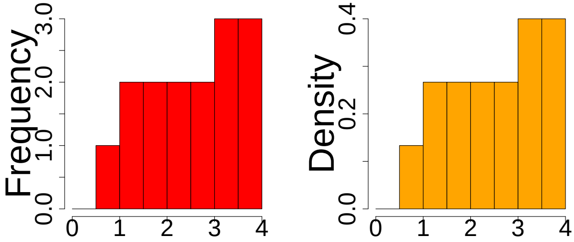
1. Make a both a frequency and density histogram from the data below.

Use bins of width 0.5 starting at 0. The bins should be right closed.

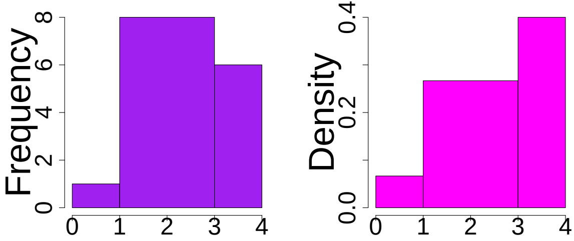
1	1.2	1.3	1.6	1.6
2.1	2.2	2.6	2.7	3.1
3.2	3.4	3.8	3.9	3.9

2. Same question using unequal width bins with edges 0, 1, 3, 4.

Solution



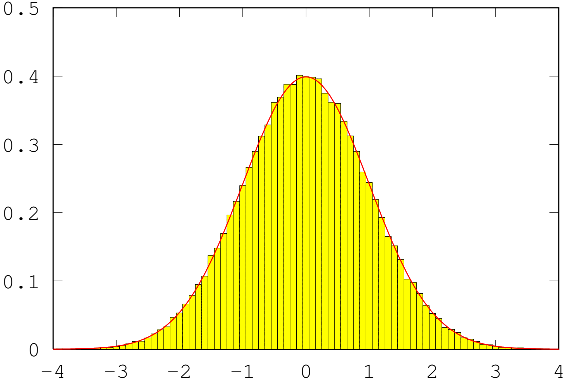
Histograms with equal width bins



Histograms with unequal width bins

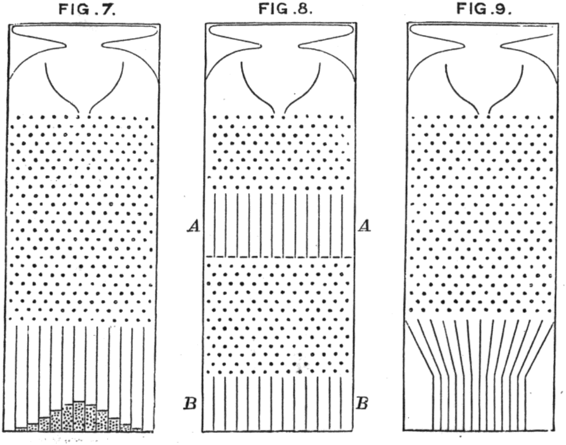
LoLN and histograms

LoLN implies density histogram converges to pdf:



Histogram with bin width .1 showing 100000 draws from a standard normal distribution. Standard normal pdf is overlaid in red.

LoLN & CLT in Action



http://www.mathsisfun.com/data/quincunx.html

http://www.youtube.com/watch?v=9xUBhhM4vbM

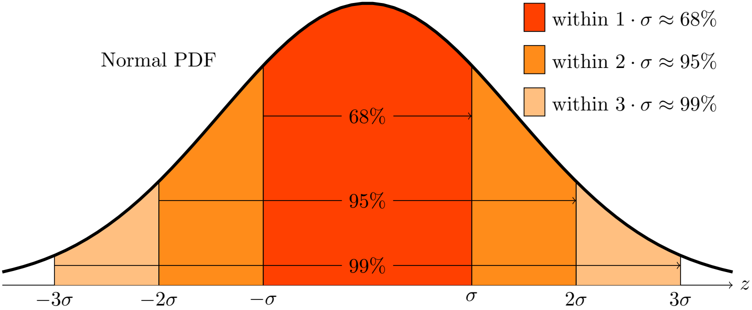
Standardization

Random variable X with mean μ and standard deviation σ.

Standardization: $\quad Z = \dfrac{X - \mu}{\sigma}.$

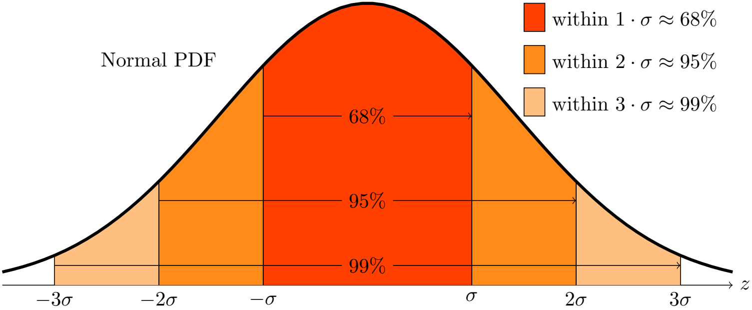
- Z has mean 0 and standard deviation 1.
- Standardizing any normal random variable produces the standard normal.
- If $X \approx$ normal then standardized $X \approx$ stand. normal.

Concept Question: Standard Normal



1. $P(-1 < Z < 1)$ is
   a) .025 b) .16 c) .68 d) .84 e) .95
2. $P(Z > 2)$
   a) .025 b) .16 c) .68 d) .84 e) .95

Concept Question: Standard Normal



1. $P(-1 < Z < 1)$ is
   a) .025 b) .16 c) .68 d) .84 e) .95
2. $P(Z > 2)$
   a) .025 b) .16 c) .68 d) .84 e) .95

answer: 1c, 2a: (1-P(-2<Z<2))/2

Central Limit Theorem

Setting: $X_1, X_2, \ldots$ i.i.d. with mean μ and standard dev. σ.

For each n:

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$
$$S_n = X_1 + X_2 + \ldots + X_n.$$

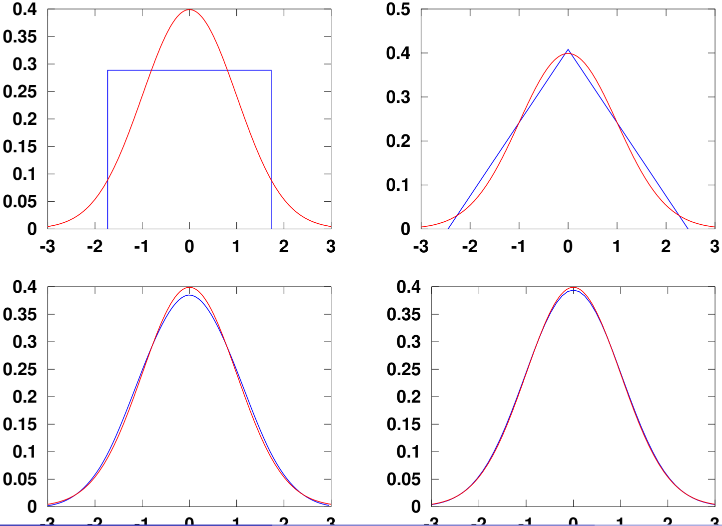
Conclusion: For large n:

$$\overline{X}_n \approx \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$
$$S_n \approx \mathrm{N}\left(n\mu, n\sigma^2\right)$$

Standardized S_n or $\overline{X}_n \approx \mathrm{N}(0,1)$

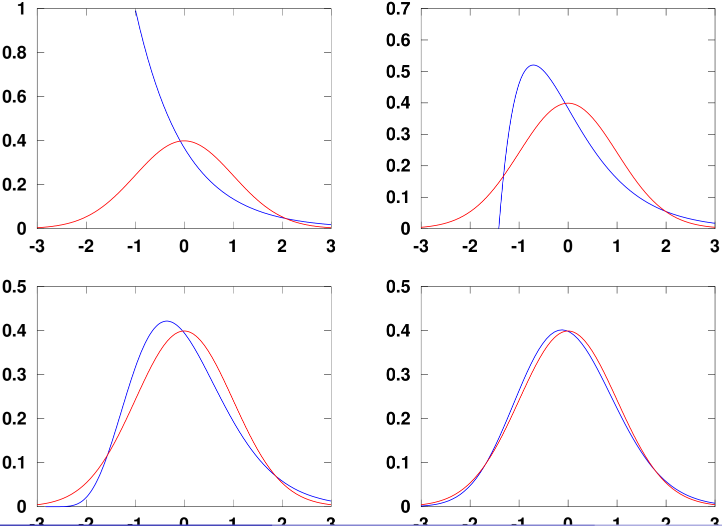
CLT: pictures

Standardized average of n i.i.d. uniform random variables with $n = 1, 2, 4, 12$.



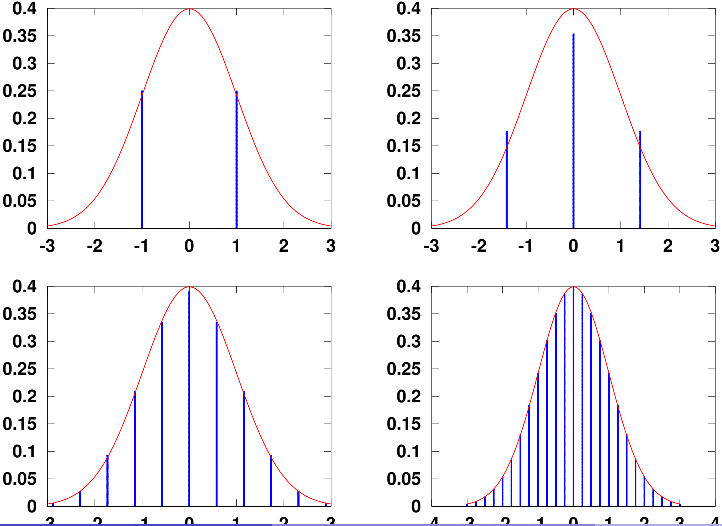
CLT: pictures 2

The standardized average of n i.i.d. exponential random variables with $n = 1, 2, 8, 64$.



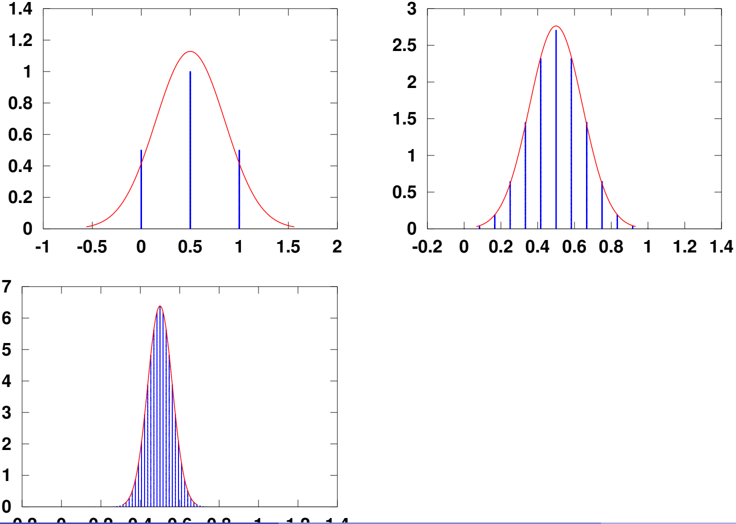
CLT: pictures 3

The standardized average of n i.i.d. Bernoulli(.5) random variables with $n = 1, 2, 12, 64$.



CLT: pictures 4

The (non-standardized) average of n Bernoulli(.5) random variables, with $n = 4, 12, 64$. (Spikier.)



Board Question: CLT

1. Carefully write the statement of the central limit theorem.

2. To head the newly formed US Dept. of Statistics, suppose that 50% of the population supports Erika, 25% supports Ruthi, and the remaining 25% is split evenly between Peter, Jon and Jerry.

A poll asks 400 random people who they support. What is the probability that at least 55% of those polled prefer Erika?

3. What is the probability that less than 20% of those polled prefer Ruthi?

answer: On next slide.

Solution

answer: 2. Let $\mathcal{E}$ be the number polled who support Erika.
The question asks for the probability $\mathcal{E} > .55 \cdot 400 = 220$.

$E(\mathcal{E}) = 400(.5) = 200$ and $\sigma^2_{\mathcal{E}} = 400(.5)(1 - .5) = 100 \Rightarrow \sigma_{\mathcal{E}} = 10$.

Because $\mathcal{E}$ is the sum of 400 Bernoulli(.5) variables the CLT says it is approximately normal and standardizing gives

$$\frac{\mathcal{E} - 200}{10} \approx Z$$

and

$$P(\mathcal{E} > 220) \approx P(Z > 2) \approx .025$$

3. Let $\mathcal{R}$ be the number polled who support Ruthi.
The question asks for the probability the $\mathcal{R} < 0.2 \cdot 400 = 80$.

$E(\mathcal{R}) = 400(.25) = 100$ and $\sigma^2_{\mathcal{R}} = 400(.25)(.75) = 75 \Rightarrow \sigma_{\mathcal{R}} = \sqrt{75}$.

So $(\mathcal{R} - 100)/\sqrt{75} \approx Z$ and

$$P(\mathcal{R} < 80) \approx P(Z < -20/\sqrt{75}) \approx 0.0105$$

Clicker:

You know from a previous study, that the customers have a normally distributed income with mean μ = $50k and σ=$10k

If you randomly select 100 customers, how many do you expect to have have an income of over $70k?

a) ≤ 0.02
b) ≈ 1.00
c) ≈ 2.28
d) ≈ 15.87

Ignore that there are no negative incomes

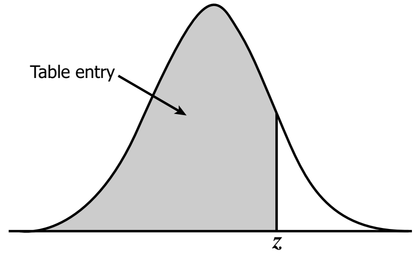


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

You know from a previous study, that the customers have a normally distributed income with mean μ = $50k and σ=$10k

If you randomly select 100 customers, how many do you expect to have have an income of over $70k?

$$Z = \frac{X-\mu}{\sigma} = \frac{70k-50k}{10k} = 2$$

$$P(X \geq 70k) \times 100$$

$$= 100 \times (1-0.9772) = 2.28$$

Ignore that there are no negative incomes

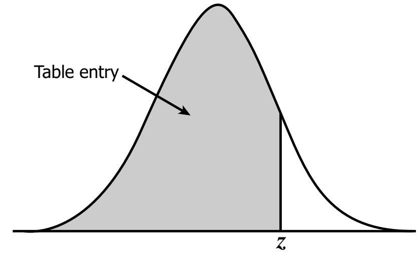


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Clicker:

You know from a previous study, that the customers have a normally distributed income with mean μ = \$50k and σ=\$10k

If you randomly select 100 customers, how many do you expect to have have an income of over \$70k?

a) ≤ 0.02
b) ≈ 1.00
c) ≈ 2.28
d) ≈ 15.87

Ignore that there are no negative incomes

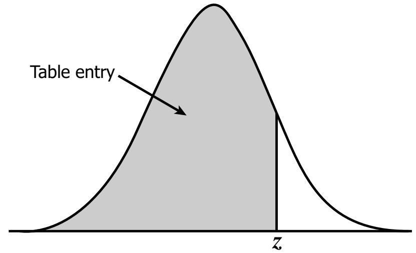


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Clicker:

For a normal distributed random variable X with

$$\mu = 5$$

$$\sigma = 10$$

What is the likelihood to see a value less than 10.5

P(X < 10.5)

a) 0.5
b) 0.6915
c) 0.7088
d) 0.9987

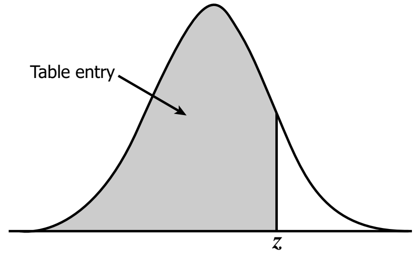


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Clicker Solution:

For a normal distributed random variable X with

$$\mu = 5$$

$$\sigma = 10$$

What is the likelihood that to see a value less than 10.5

$$Z = \frac{X-\mu}{\sigma}$$

$$= \frac{10.5-5}{10} = 0.55$$

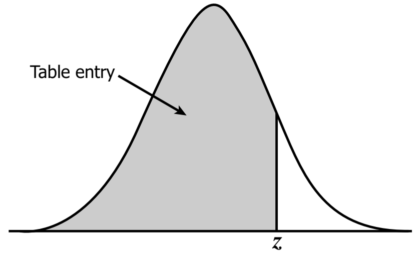


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Clicker:

For a normal distributed random variable X with

$$\mu = 5$$

$$\sigma = 10$$

What is the likelihood to see a value less than 10.5

P(X < 10.5)

a) 0.5
b) 0.6915
c) 0.7088
d) 0.9987

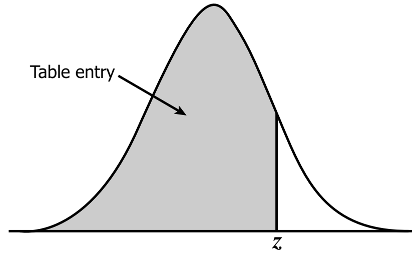


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Clicker:

For a normal distributed random variable X with

$$\mu = 5$$

$$\sigma = 10$$

What is the value of d so that

$$P(Z < d) \leq 0.05$$

a) -21.5
b) -11.5
c) 5
d) 21.5

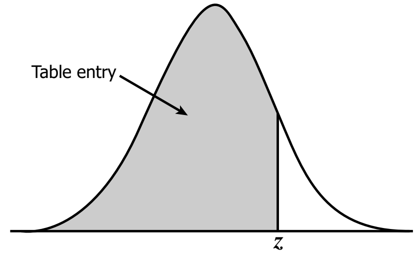


Table entry for z is the area under the standard normal curve to the left of z.

z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Clicker:

For a normal distributed random variable X with

$$\mu = 5$$

$$\sigma = 10$$

What is the value of d so that

$$P(Z < d) \leq 0.05$$

$$-1.65 = \frac{x-\mu}{\sigma} = \frac{d-5}{10}$$

$$d = -1.65 \times 10 + 5 = -11.5$$

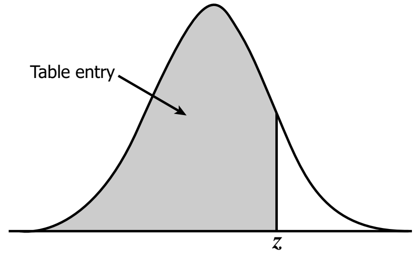


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Clicker:

For a normal distributed random variable X with

$$\mu = 5$$

$$\sigma = 10$$

What is the value of d so that

$$P(Z < d) \leq 0.05$$

a) -21.5
b) -11.5
c) 5
d) 21.5

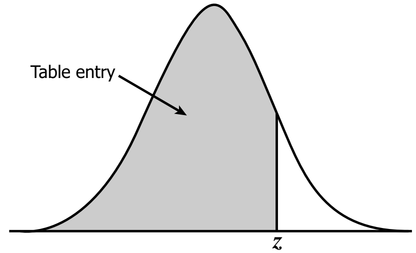


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998