P1 - Probability

STAT 587 (Engineering)
Iowa State University

August 17, 2020

Probability - Interpretation

What do we mean when we say the word probability/chance/likelihood/odds? For example,

- The probability the COVID-19 outbreak is done by 2021 is 4%.
- The probability that Joe Biden will become president is 53%.
- The chance I will win a game of solitaire is 20%.

Interpretations:

- Relative frequency: Probability is the proportion of times the event occurs as the number of times the event is attempted tends to infinity.
- Personal belief: Probability is a statement about your personal belief in the event occuring.

Probability - Example

Let C be a successful connection to the internet from a laptop event.

> *From our experience with the wireless network and our internet service provider, we believe the probability we successfully connect is 90 %.*

We write $P(C) = 0.9$.

> *To be able to work with probabilities, in particular, to be able to compute probabilities of events, a mathematical foundation is necessary.*

Sets - definition

A set is a collection of things. We use the following notation

- $\omega \in A$ means ω is an element of the set A,
- $\omega \notin A$ means ω is not an element of the set A,
- $A \subseteq B$ (or $B \supseteq A$) means the set A is a subset of B (with the sets possibly being equal), and
- $A \subset B$ (or $B \supset A$) means the set A is a proper subset of B, i.e. there is at least one element in B that is not in A.

The sample space, Ω, is the set of all outcomes of an experiment.

Set - examples

The set of all possible sums of two 6-sided dice rolls is $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ and

- $2 \in \Omega$
- $1 \notin \Omega$
- $\{2, 3, 4\} \subset \Omega$

Set comparison, operations, terminology

For the following $A, B \subseteq \Omega$ where Ω is the implied universe of all elements under study,

1. Union ($\cup$): A union of events is an event consisting of all the outcomes in these events.

$$A \cup B = \{\omega \mid \omega \in A \text{ or } \omega \in B\}$$

2. Intersection ($\cap$): An intersection of events is an event consisting of the common outcomes in these events.

$$A \cap B = \{\omega \mid \omega \in A \text{ and } \omega \in B\}$$

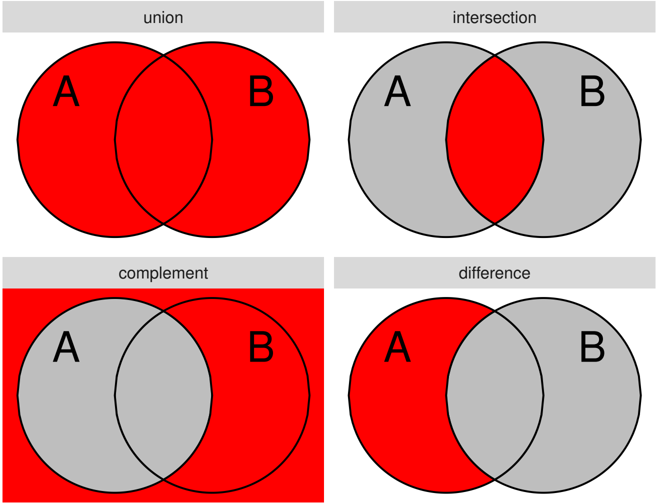
3. Complement (A^C): A complement of an event A is an event that occurs when event A does not happen.

$$A^C = \{\omega \mid \omega \notin A \text{ and } \omega \in \Omega\}$$

4. Set difference ($A \setminus B$): All elements in A that are not in B, i.e.

$$A \setminus B = \{\omega | \omega \in A \text{ and } \omega \notin B\}$$

Venn diagrams



Example

Consider the set Ω equal to all possible sum of two 6-sided die rolls i.e. $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ and two subsets

- all odd rolls: $A = \{3, 5, 7, 9, 11\}$
- all rolls below 6: $B = \{2, 3, 4, 5\}$

Then we have

- $A \cup B = \{2, 3, 4, 5, 7, 9, 11\}$
- $A \cap B = \{3, 5\}$
- $A^C = \{2, 4, 6, 8, 10, 12\}$
- $B^C = \{6, 7, 8, 9, 10, 11, 12\}$
- $A \setminus B = \{7, 9, 11\}$
- $B \setminus A = \{2, 4\}$

Set comparison, operations, terminology (cont.)

5. Empty Set $\emptyset$ is a set having no elements, i.e. $\{\}$. The empty set is a subset of every set:

$$\emptyset \subseteq A$$

6. Disjoint sets: Sets A, B are disjoint if their intersection is empty:

$$A \cap B = \emptyset$$

7. Pairwise disjoint sets: Sets $A_1, A_2, \ldots$ are pairwise disjoint if all pairs of these events are disjoint:

$$A_i \cap A_j = \emptyset \text{ for any } i \neq j$$

8. De Morgan's Laws:

$$(A \cup B)^C = A^C \cap B^C \qquad \text{and} \qquad (A \cap B)^C = A^C \cup B^C$$

Examples

Let $A = \{2, 3, 4\}, B = \{5, 6, 7\}, C = \{8, 9, 10\}, D = \{11, 12\}$. Then

- $A \cap B = \emptyset$
- A, B, C, D are pairwise disjoint
- De Morgan's:

$$\begin{aligned} (A \cup B) &= \{2, 3, 4, 5, 6, 7\} \\ (A \cup B)^C &= \{8, 9, 10, 11, 12\} \end{aligned}$$

$$\begin{aligned} A^C &= \{5, 6, 7, 8, 9, 10, 11, 12\} \\ B^C &= \{2, 3, 4, 8, 9, 10, 11, 12\} \\ A^C \cap B^C &= \{8, 9, 10, 11, 12\} \end{aligned}$$

so, by example,

$$(A \cup B)^C = A^C \cap B^C.$$

Kolmogorov's Axioms

A system of probabilities (a probability model) is an assignment of numbers $P(A)$ to events $A \subseteq \Omega$ such that

(i) $0 \leq P(A) \leq 1$ for all A

(ii) $P(\Omega) = 1$.

(iii) if $A_1, A_2, \ldots$ are pairwise disjoint events (i.e. $A_i \cap A_j = \emptyset$ for all $i \neq j$) then

$$\begin{aligned} P(A_1 \cup A_2 \cup \ldots) &= P(A_1) + P(A_2) + \ldots \\ &= \textstyle\sum_i P(A_i). \end{aligned}$$

Kolmogorov's Axioms (cont.)

These are the basic rules of operation of a probability model

- every valid model must obey these,
- any system that does, is a valid model.

Whether or not a particular model is realistic is different question.

Example: Draw a single card from a standard deck of playing cards: $\Omega = \{red, black\}$ Two different, equally valid probability models are:

Model 1	Model 2
$P(\Omega) = 1$	$P(\Omega) = 1$
$P(red) = 0.5$	$P(red) = 0.3$
$P(black) = 0.5$	$P(black) = 0.7$

Mathematically, both schemes are equally valid.
But, of course, our real world experience would prefer model 1 over model 2.

Useful Consequences of Kolmogorov's Axioms

Let $A, B \subseteq \Omega$.

- Probability of the Complementary Event: $P\left(A^C\right) = 1 - P(A)$
  Corollary: $P(\emptyset) = 0$
- Addition Rule of Probability

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

- If $A \subseteq B$, then $P(A) \leq P(B)$.

Example: Using Kolmogorov's Axioms

We attempt to access the internet from a laptop at home. We connect successfully if and only if the wireless (WiFi) network works *and* the internet service provider (ISP) network works. Assume

$$\begin{aligned} P(\text{ WiFi up }) &= .9 \\ P(\text{ ISP up }) &= .6, \text{ and} \\ P(\text{ WiFi up and ISP up }) &= .55. \end{aligned}$$

1. What is the probability that the WiFi is up or the ISP is up?
2. What is the probability that both the WiFi and the ISP are down?
3. What is the probability that we fail to connect?

Solution

Let $A \equiv$ WiFi up; $B \equiv$ ISP up

1. What is the probability that the WiFi is up or the ISP is up?

$$P(\text{ WiFi up or ISP up}) = P(A \cup B) = 0.9 + 0.6 - 0.55 = 0.95$$

2. What is the probability that both the WiFi and the ISP are down?

$$\begin{aligned} P(\text{ WiFi down and ISP down}) &= P\left(A^C \cap B^C\right) = P\left([A \cup B]^C\right) \\ &= 1 - .95 = .05 \end{aligned}$$

3. What is the probability that we fail to connect?

$$\begin{aligned} &P(\text{ WiFi down or ISP down}) \\ &\quad = P\left(A^C \cup B^C\right) = P\left(A^C\right) + P\left(B^C\right) - P\left(A^C \cap B^C\right) \\ &\quad = P\left(A^C \cup B^C\right) = (1 - .9) + (1 - .6) - .05 = .1 + .4 - .05 = .45 \end{aligned}$$

Conditional probability - Definition

The conditional probability of an event A given an event B is

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

if $P(B) > 0$.

Intuitively, the fraction of outcomes in B that are also in A.

Corrollary:

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

Random CPUs

A box has 500 CPUs with a speed of 1.8 GHz and 500 with a speed of 2.0 GHz. The numbers of good (G) and defective (D) CPUs at the two different speeds are as shown below.

	1.8 GHz	2.0 GHz	Total
G	480	490	970
D	20	10	30
Total	500	500	1000

We select a CPU at random and observe its speed. What is the probability that the CPU is defective given that its speed is 1.8 GHz?

Let

- D be the event the CPU is defective and
- S be the event the CPU speed is 1.8 GHz.

Then

- $P(S) = 500/1000 = 0.5$
- $P(S \cap D) = 20/1000 = 0.02.$
- $P(D|S) = P(S \cap D)/P(S) = 0.02/0.5 = 0.04.$

Statistical independence - Definition

Events A and B are statistically independent if

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

or, equivalently,

$$P(A|B) = P(A).$$

Intuition: the occurrence of one event does not affect the probability of the other.

Example: In two tosses of a coin, the result of the first toss does not affect the probability of the second toss being heads.

WiFi example

In trying to connect my laptop to the internet, I need

- my WiFi network to be up (event A) and
- the ISP network to be up (event B).

Assume the probability the WiFi network is up is 0.6 and the ISP network is up is 0.9. If the two events are independent, what is the probability we can connect to the internet?

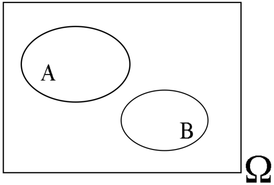
Since we have independence, we know

$$P(A \cap B) = P(A) \times P(B) = 0.6 \times 0.9 = 0.54.$$

Independence and disjoint

Warning: Independence and disjointedness are two very different concepts!

Disjoint:

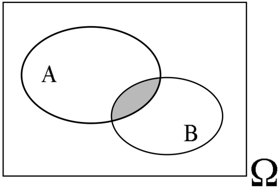


A, B are disjoint

If A and B are disjoint, their intersection is empty and therefore has probability 0:

$$P(A \cap B) = P(\emptyset) = 0.$$

Independence:



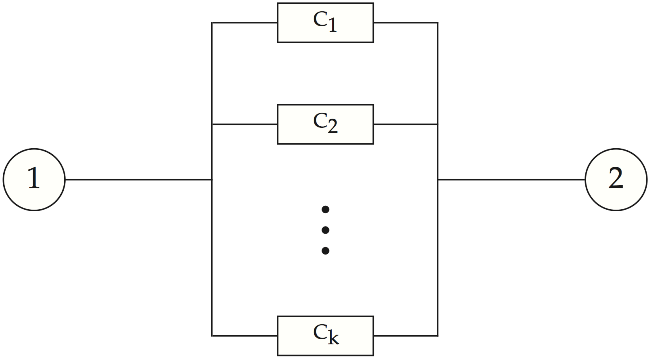
$A \cap B$

If A and B are independent events, the probability of their intersection can be computed as the product of their individual probabilities:

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

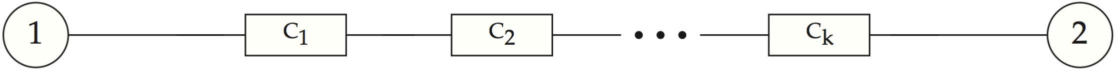
Parallel system - Definition

A parallel system consists of K components $c_1, \ldots, c_K$ arranged in such a way that the system works if **at least one** of the K components functions properly.



Serial system - Definition

A serial system consists of K components $c_1, \ldots, c_K$ arranged in such a way that the system works if and only if **all** of the components function properly.



Reliability - Definition

The reliability of a system is the probability the system works.

Example: The reliability of the WiFi-ISP network (assuming independence) is 0.54.

Reliability of parallel systems with independent components

Let $c_1, \ldots, c_K$ denote the K components in a parallel system. Assume the K components operate independently and $P(c_k \text{ works}) = p_k$. What is the reliability of the system?

$$
\begin{aligned}
P(\text{ system works }) &= P(\text{ at least one component works }) \\
&= 1 - P(\text{ all components fail }) \\
&= 1 - P(c_1 \text{ fails and } c_2 \text{ fails } \ldots \text{ and } c_k \text{ fails }) \\
&= 1 - \textstyle\prod_{k=1}^{K} P(c_k \text{ fails}) \\
&= 1 - \textstyle\prod_{k=1}^{K} (1 - p_k).
\end{aligned}
$$

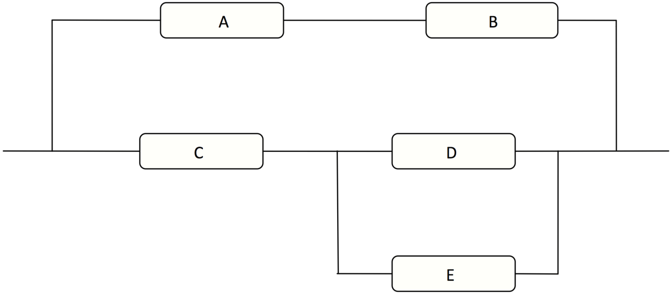
Reliability of serial systems with independent components

Let $c_1, \ldots, c_K$ denote the K components in a serial system. Assume the K components operate independently and $P(c_k \text{ works }) = p_k$. What is the reliability of the system?

$$\begin{aligned} P(\text{ system works }) &= P(\text{ all components work}) \\ &= \prod_{k=1}^{K} P(c_k \text{ works}) \\ &= \prod_{k=1}^{K} p_k. \end{aligned}$$

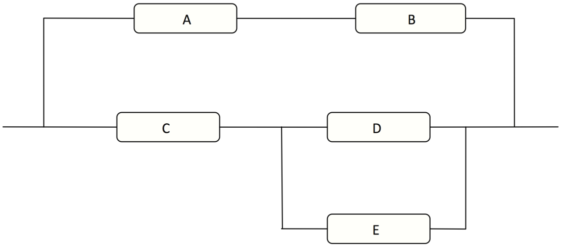
Reliability example

Each component in the system shown below is opearable with probability 0.92 independently of other components. Calculate the reliability.



Reliability example

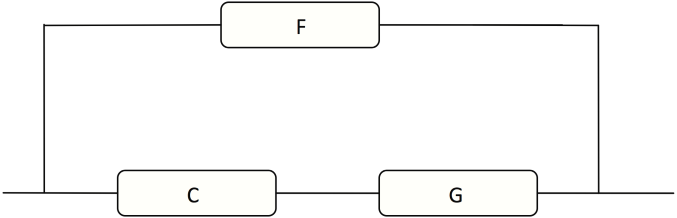
Each component in the system shown below is opearable with probability 0.92 independently of other components. Calculate the reliability.



1. Serial components A and B can be replaced by a component F that operates with probability $P(A \cap B) = (0.92)^2 = 0.8464.$
2. Parallel components D and E can be replaced by component G that operates with probability $P(D \cup E) = 1 - (1 - 0.92)^2 = 0.9936.$

Reliability example (cont.)

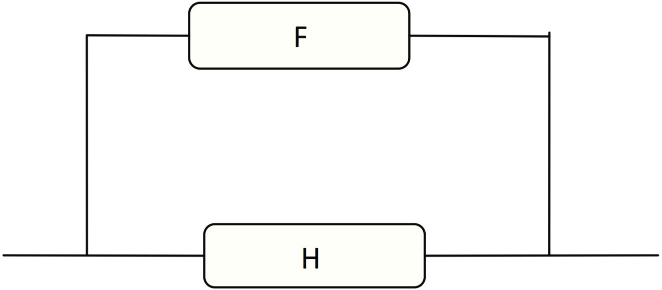
Updated circuit:



3. Serial components C and G can be replaced by a component H that operates with probability $P(C \cap G) = (0.92)(0.9936) = 0.9141$.

Reliability example (cont.)

Updated circuit:



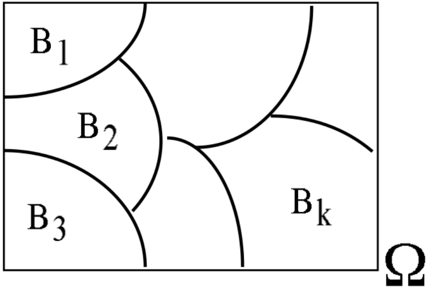
4. Parallel componenents F and H are in parallel, so the reliability of the system is $P(F \cup H) = 1 - (1 - 0.8424)(1 - 0.9141) \approx 0.99.$

Partition

Definition

A collection of events $B_1, \dots B_K$ is called a partition (or cover) of Ω if

- the events are pairwise disjoint (i.e., $B_i \cap B_j = \emptyset$ for $i \neq j$), and
- the union of the events is Ω (i.e., $\bigcup_{k=1}^{K} B_k = \Omega$).



Example

Consider the sum of two 6-sided die, i.e.

$$\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

Here are some covers:

- $\{2, 3, 4\}, \{5, 6, 7, 8, 9, 10, 11, 12\}$
- $\{2, 3, 4\}, \{5, 6, 7\}, \{8, 9, 10\}, \{11, 12\}$
- $A_2, A_3, \ldots, A_{12}$ where $A_i = \{i\}$
- any A and A^C where $A \subseteq \Omega$

Law of Total Probability

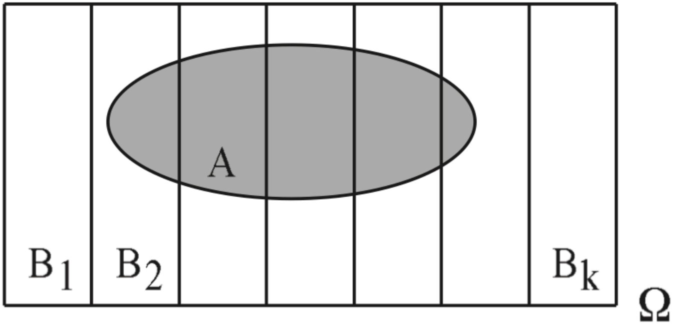
Law of Total Probability:
If the collection of events $B_1, \ldots, B_K$ is a partition of Ω, and A is an event, then

$$P(A) = \sum_{k=1}^{K} P(A|B_k)P(B_k).$$

Proof:

$$\begin{aligned} P(A) &= P\left(\bigcup_{k=1}^{K} A \cap B_k\right) && \text{partition} \\ &= \sum_{k=1}^{K} P(A \cap B_k) && \text{pairwise disjoint} \\ &= \sum_{k=1}^{K} P(A|B_k)P(B_k) && \text{conditional probability} \end{aligned}$$

Law of Total Probability - Graphically



Law of Total ProbabilityExample

In the come out roll of craps, you win if the roll is a 7 or 11. By the law of total probability, the probability you win is

$$P(\text{Win}) = \sum_{i=2}^{12} P(\text{Win}|i)P(i) = P(7) + P(11)$$

since $P(\text{Win}|i) = 1$ if $i = 7, 11$ and 0 otherwise.

Bayes' Rule

Bayes' Rule:
If $B_1, \ldots, B_K$ is a partition of Ω, and A is an event in Ω, then

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{k=1}^{K} P(A|B_k)P(B_k)}.$$

Proof:

$$\begin{aligned} P(B_k|A) &= \frac{P(A \cap B_k)}{P(A)} && \text{conditional probability} \\ &= \frac{P(A|B_k)P(B_k)}{P(A)} && \text{conditional probability} \\ &= \frac{P(A|B_k)P(B_k)}{\sum_{k=1}^{K} P(A|B_k)P(B_k)} && \text{Law of Total Probability} \end{aligned}$$

Bayes' Rule: Craps example

If you win on a come-out roll in craps, what is the probability you rolled a 7?

$$
\begin{aligned}
P(7|\text{Win}) &= \frac{P(\text{Win}|7)P(7)}{\sum_{i=2}^{12} P(\text{Win}|i)P(i)} \\
&= \frac{P(7)}{P(7)+P(11)}.
\end{aligned}
$$

Bayes' Rule: CPU testing example

A given lot of CPUs contains 2% defective CPUs. Each CPU is tested before delivery. However, the tester is not wholly reliable:

$$\begin{aligned} P(\text{ tester says CPU is good} \mid \text{CPU is good }) &= 0.95 \\ P(\text{ tester says CPU is defective} \mid \text{CPU is defective }) &= 0.94 \end{aligned}$$

If the test device says the CPU is defective, what is the probability that the CPU is actually defective?

CPU testing (cont.)

Let

- C_g (C_d) be the event the CPU is good (defective)
- T_g (T_d) be the event the tester says the CPU is good (defective)

We know

- $0.02 = P(C_d) \quad = 1 - P(C_g)$
- $0.95 = P(T_g|C_g) = 1 - P(T_d|C_g)$
- $0.94 = P(T_d|C_d) = 1 - P(T_g|C_d)$

Using Bayes' Rule, we have

$$\begin{aligned} P(C_d|T_d) &= \frac{P(T_d|C_d)P(C_d)}{P(T_d|C_d)P(C_d)+P(T_d|C_g)P(C_g)} \\ &= \frac{P(T_d|C_d)P(C_d)}{P(T_d|C_d)P(C_d)+[1-P(T_g|C_g)][1-P(C_d)]} \\ &= \frac{0.94\times 0.02}{0.94\times 0.02+[1-0.95]\times[1-0.02]} \\ &= 0.28 \end{aligned}$$

Probability Summary

- Probability Interpretation
- Sets and set operations
- Kolmogorov's Axioms
- Conditional Probability
- Independence
- Reliability
- Law of Total Probability
- Bayes' Rule

P2 - Discrete Random Variables

STAT 587 (Engineering)
Iowa State University

August 21, 2020

Random variables

If Ω is the sample space of an experiment, a random variable X is a function $X(\omega) : \Omega \mapsto \mathbb{R}$.

Idea: If the value of a numerical variable depends on the outcome of an experiment, we call the variable a *random variable*.

Examples of random variables from rolling two 6-sided dice:

- Sum of the two dice
- Indicator of the sum being greater than 5

We will use an upper case Roman letter (late in the alphabet) to indicate a random variable and a lower case Roman letter to indicate a realized value of the random variable.

8 bit example

Suppose, 8 bits are sent through a communication channel. Each bit has a certain probability to be received incorrectly. We are interested in the number of bits that are received incorrectly.

- Let X be the number of incorrect bits received.
- The possible values for X are $\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$.
- Example events:
  - No incorrect bits received: $\{X = 0\}$.
  - At least one incorrect bit received: $\{X \geq 1\}$.
  - Exactly two incorrect bits received: $\{X = 2\}$.
  - Between two and seven (inclusive) incorrect bits received: $\{2 \leq X \leq 7\}$.

Range/image of random variables

The range (or image) of a random variable X is defined as

$$Range(X) := \{x : x = X(\omega) \text{ for some } \omega \in \Omega\}$$

If the range is finite or countably infinite, we have a discrete random variable. If the range is uncountably infinite, we have a continuous random variable.
Examples:

- Put a hard drive into service, measure $Y =$ "time until the first major failure" and thus $Range(Y) = (0, \infty)$. Range of Y is an interval (uncountable range), so Y is a continuous random variable.
- Communication channel: $X =$ "# of incorrectly received bits out of 8 bits sent" with $Range(X) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$. Range of X is a finite set, so X is a discrete random variable.
- Communication channel: $Z =$ "# of incorrectly received bits in 10 minutes" with $Range(Z) = \{0, 1, \ldots\}$. Range of Z is a countably infinite set, so Z is a discrete random variable.

Distribution

The collection of all the probabilities related to X is the distribution of X.

For a discrete random variable, the function

$$p_X(x) = P(X = x)$$

is the probability mass function (pmf) and the cumulative distribution function (cdf) is

$$F_X(x) = P(X \leq x) = \sum_{y \leq x} p_X(y).$$

The set of non-zero probability values of X is called
the support of the distribution f.
This is the same as the range of X.

Examples

A probability mass function is valid if it defines a valid set of probabilities, i.e. they obey Kolmogorov's axioms.

Which of the following functions are a valid probability mass functions?

x	-3	-1	0	5	7
$p_X(x)$	0.1	0.45	0.15	0.25	0.05

y	-1	0	1.5	3	4.5
$p_Y(y)$	0.1	0.45	0.25	-0.05	0.25

z	0	1	3	5	7
$p_Z(z)$	0.22	0.18	0.24	0.17	0.18

Rolling a fair 6-sided die

Let Y be the number of pips on the upturned face of a die. The support of Y is $\{1, 2, 3, 4, 5, 6\}$. If we believe the die has equal probability for each face, then image, pmf, and cdf for Y are

y	1	2	3	4	5	6
$p_Y(y) = P(Y = y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$F_Y(y) = P(Y \leq y)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{6}{6}$

Dragonwood

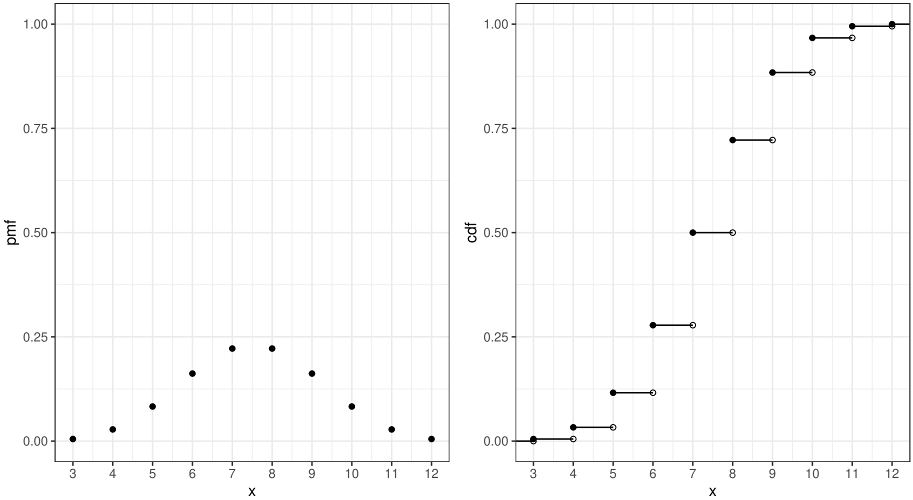
Dragonwood has 6-sided dice with the following # on the 6 sides: $\{1, 2, 2, 3, 3, 4\}$.

What is the support, pmf, and cdf for the sum of the upturned numbers when rolling 3 Dragonwood dice?

```
# Three dice
die   = c(1,2,2,3,3,4)
rolls = expand.grid(die1 = die, die2 = die, die3 = die)
sum   = rowSums(rolls); tsum = table(sum)
dragonwood3 = data.frame(x   = round(as.numeric(names(tsum)),0),
                         pmf = round(as.numeric(table(sum)/length(sum)),3)) %>%
        mutate(cdf = cumsum(pmf))

t(dragonwood3)
```

```
     [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]   [,8]   [,9]  [,10]
x   3.000 4.000 5.000 6.000 7.000 8.000 9.000 10.000 11.000 12.000
pmf 0.005 0.028 0.083 0.162 0.222 0.222 0.162  0.083  0.028  0.005
cdf 0.005 0.033 0.116 0.278 0.500 0.722 0.884  0.967  0.995  1.000
```

Dragonwood - pmf and cdf



Properties of pmf and cdf

Properties of probability mass function $p_X(x) = P(X = x)$:

- $0 \leq p_X(x) \leq 1$ for all $x \in \mathbb{R}$.
- $\sum_{x \in S} p_X(x) = 1$ where S is the support.

Properties of cumulative distribution function $F_X(x)$:

- $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$
- F_X is nondecreasing, (i.e. if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$.)
- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
- $F_X(x)$ is right continuous with respect to x

Dragonwood (cont.)

In Dragonwood, you capture monsters by rolling a sum equal to or greater than its defense. Suppose you can roll 3 dice and the following monsters are available to be captured:

- Spooky Spiders worth 1 victory point with a defense of 3.
- Hungry Bear worth 3 victory points with a defense of 7.
- Grumpy Troll worth 4 victory points with a defense of 9.

Which monster should you attack?

Dragonwood (cont.)

Calculate the probability by computing one minus the cdf evaluated at "defense minus 1". Let X be the sum of the number on 3 Dragonwood dice. Then

- $P(X \geq 3) = 1 - P(X \leq 2) = 1$
- $P(X \geq 7) = 1 - P(X \leq 6) = 0.722.$
- $P(X \geq 9) = 1 - P(X \leq 8) = 0.278.$

If we multiply the probability by the
number of victory points,
then we have the "expected points":

- $1 \times P(X \geq 3) = 1$
- $3 \times P(X \geq 7) = 2.17.$
- $4 \times P(X \geq 9) = 1.11.$

Expectation

Let X be a random variable and h be some function. The expected value of a function of a (discrete) random variable is

$$E[h(X)] = \sum_i h(x_i) \cdot p_X(x_i).$$

Intuition: Expected values are *weighted averages* of the possible values weighted by their probability.

If $h(x) = x$, then

$$E[X] = \sum_i x_i \cdot p_X(x_i)$$

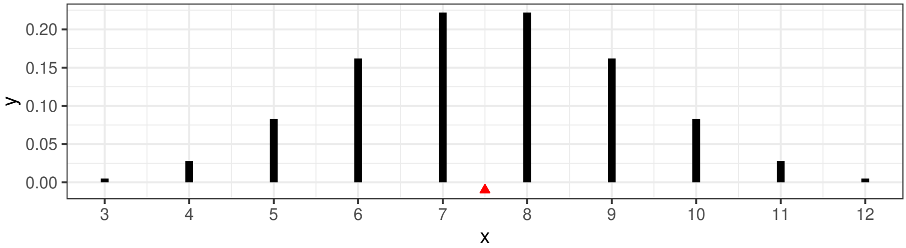
and we call this the expectation of X and commonly use the symbol μ for the expectation.

Dragonwood (cont.)

What is the expectation of the sum of 3 Dragonwood dice?

```
expectation = with(dragonwood3, sum(x*pmf))
expectation

[1] 7.5
```

The expectation can be thought of as the center of mass if we place mass $p_X(x)$ at corresponding points x.



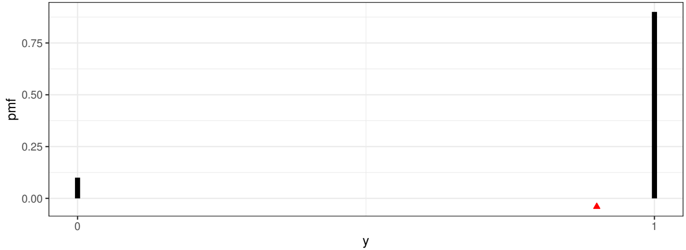
Biased coin

Suppose we have a biased coin represented by the following pmf:

y	0	1
$p_Y(y)$	$1-p$	p

What is the expected value?

If $p = 0.9$,



Properties of expectations

Let X and Y be random variables and a, b, and c be constants. Then

$$E[aX + bY + c] = aE[X] + bE[Y] + c.$$

In particular

- $E[X + Y] = E[X] + E[Y]$,
- $E[aX] = aE[X]$, and
- $E[c] = c$.

Dragonwood (cont.)

Enhancement cards in Dragonwood allow you to improve your rolls. Here are two enhancement cards:

- *Cloak of Darkness* adds 2 points to all capture attempts and
- *Friendly Bunny* allows you (once) to roll an extra die.

What is the expected attack roll total if you had 3 Dragonwood dice, the Cloak of Darkness, and are using the Friendly Bunny?

Let

- X be the sum of 3 Dragonwood dice (we know $E[X] = 7.5$),
- Y be the sum of 1 Dragonwood die which has $E[Y] = 2.5$.

Then the attack roll total is $X + Y + 2$ and the *expected* attack roll total is

$$E[X + Y + 2] = E[X] + E[Y] + 2 = 7.5 + 2.5 + 2 = 12.$$

Variance

The variance of a random variable is defined as the expected squared deviation from the mean. For discrete random variables, variance is

$$Var[X] = E[(X-\mu)^2] = \sum_i (x_i - \mu)^2 \cdot p_X(x_i)$$

where $\mu = E[X]$. The symbol σ^2 is commonly used for the variance.
The variance is analogous to moment of intertia in classical mechanics.

The standard deviation (sd) is the positive square root of the variance:

$$SD[X] = \sqrt{Var[X]}.$$

The symbol σ is commonly used for sd.

Properties of variance

Two discrete random variables X and Y are independent if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y).$$

If X and Y are independent, and a, b, and c are constants, then

$$Var[aX + bY + c] = a^2Var[X] + b^2Var[Y].$$

Special cases:

- $Var[c] = 0$
- $Var[aX] = a^2Var[X]$
- $Var[X + Y] = Var[X] + Var[Y]$
  (if X and Y are independent)

Dragonwood (cont.)

What is the variance for the sum of the 3 Dragonwood dice?

```
variance = with(dragonwood3, sum((x-expectation)^2*pmf))
variance

[1] 2.766
```

What is the standard deviation for the sum of the pips on 3 Dragonwood dice?

```
sqrt(variance)

[1] 1.66313
```

Biased coin

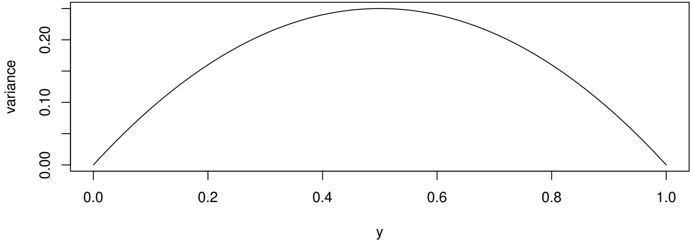
Suppose we have a biased coin represented by the following pmf:

y	0	1
$p_Y(y)$	$1-p$	p

What is the variance?

1. $E[Y] = p$
2. $Var[y] = (0-p)^2(1-p) + (1-p)^2 \times p = p - p^2 = p(1-p)$

When is this variance maximized?



Special discrete distributions

- Bernoulli
- Binomial
- Poisson

Note: The range is always finite or countable.

Bernoulli random variables

A Bernoulli experiment has only two outcomes: success/failure.

Let

- $X = 1$ represent success and
- $X = 0$ represent failure.

The probability mass function $p_X(x)$ is

$$p_X(0) = 1 - p \quad p_X(1) = p.$$

We use the notation $X \sim Ber(p)$ to denote a random variable X that follows a Bernoulli distribution with success probability p, i.e. $P(X = 1) = p$.

Bernoulli experiment examples

- Toss a coin: $\Omega = \{Heads, Tails\}$
- Throw a fair die and ask if the face value is a six:
  $\Omega = \{\text{face value is a six}, \text{face value is not a six}\}$
- Send a message through a network and record whether or not it is received:
  $\Omega = \{\text{successful transmission},\ \text{unsuccessful transmission}\}$
- Draw a part from an assembly line and record whether or not it is defective:
  $\Omega = \{\text{defective},\ \text{good}\}$
- Response to the question
  "Are you in favor of an increased in property tax
  xto pay for a new high school?":
  $\Omega = \{\text{yes}, \text{no}\}$

Bernoulli random variable (cont.)

The cdf of the Bernoulli random variable is

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

The expected value is

$$E[X] = \sum_x p_X(x) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

The variance is

$$\begin{aligned} Var[X] &= \sum_x (x - E[X])^2 p_X(x) \\ &= (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p \\ &= p(1 - p). \end{aligned}$$

Sequence of Bernoulli experiments

An experiment consisting of n independent and identically distributed Bernoulli experiments.

Examples:

- Toss a coin n times and record the nubmer of heads.
- Send 23 identical messages through the network independently and record the number successfully received.
- Draw 5 cards from a standard deck with replacement (and reshuffling) and record whether or not the card is a king.

Independent and identically distributed

Let X_i represent the i^{th} Bernoulli experiment.

Independence means

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} p_{X_i}(x_i),$$

i.e. the joint probability is the product of the individual probabilities.

Identically distributed (for Bernoulli random variables) means

$$P(X_i = 1) = p \quad \forall\, i,$$

and more generally, the distribution is the same for all the random variables.

- iid: independent and identically distributed
- ind: independent

Sequences of Bernoulli experiments

Let X_i denote the outcome of the i^{th} Bernoulli experiment. We use the notation

$$X_i \overset{iid}{\sim} Ber(p), \quad \text{for } i = 1, \ldots, n$$

to indicate a sequence of n independent and identically distributed Bernoulli experiments.

We could write this equivalently as

$$X_i \overset{ind}{\sim} Ber(p), \quad \text{for } i = 1, \ldots, n$$

but this is different than

$$X_i \overset{ind}{\sim} Ber(p_i), \quad \text{for } i = 1, \ldots, n$$

as the latter has a different success probability for each experiment.

Binomial random variable

Suppose we perform a sequence of n iid Bernoulli experiments and only record the number of successes, i.e.

$$Y = \sum_{i=1}^{n} X_i.$$

Then we use the notation $Y \sim Bin(n, p)$ to indicate a binomial random variable with

- n attempts and
- probability of success p.

Binomial probability mass function

We need to obtain

$$p_Y(y) = P(Y = y) \quad \forall\, y \in \Omega = \{0, 1, 2, \ldots, n\}.$$

The probability of obtaining a particular sequence of y success and $n - y$ failures is

$$p^y(1-p)^{n-y}$$

since the experiments are iid with success probability p. But there are

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

ways of obtaining a sequence of y success and $n - y$ failures. Thus, the binomial pmf is

$$p_Y(y) = P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}.$$

Properties of binomial random variables

The expected value is

$$E[Y] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np.$$

The variance is

$$Var[Y] = \sum_{i=1}^{n} Var[X_i] = np(1-p)$$

since the X_i are independent.

The cumulative distribution function is

$$F_Y(y) = P(Y \leq y) = \sum_{x=0}^{\lfloor y \rfloor} \binom{n}{x} p^x (1-p)^{n-x}.$$

Component failure rate

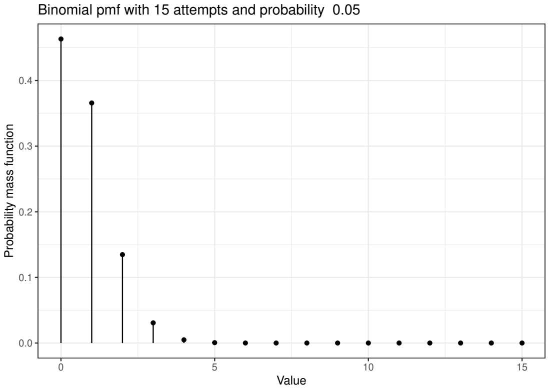
Suppose a box contains 15 components that each have a failure rate of 5%.

What is the probability that

1. exactly two out of the fifteen components are defective?
2. at most two components are defective?
3. more than three components are defective?
4. more than 1 but less than 4 are defective?

Binomial pmf

Let Y be the number of defective components and assume $Y \sim Bin(15, 0.05)$.



Component failure rate - solutions

Let Y be the number of defective components and assume $Y \sim Bin(15, 0.05)$.

$$
\begin{array}{lll}
1. & P(Y = 2) & = \binom{15}{2}(0.05)^2(1-0.05)^{15-2} \\
2. & P(Y \leq 2) & = \sum_{x=0}^{2} \binom{15}{x}(0.05)^x(1-0.05)^{15-x} \\
3. & P(Y > 3) & = 1 - P(Y \leq 3) = 1 - \sum_{x=0}^{3} \binom{15}{x}(0.05)^x(1-0.05)^{15-x} \\
4. & P(1 < Y < 4) & = \sum_{x=2}^{3} \binom{15}{x}(0.05)^x(1-0.05)^{15-x}
\end{array}
$$

Component failure rate - solutions in R

```r
n <- 15
p <- 0.05
choose(15,2)

[1] 105

dbinom(2,n,p)          # P(Y=2)

[1] 0.1347523

pbinom(2,n,p)          # P(Y<=2)

[1] 0.9637998

1-pbinom(3,n,p)        # P(Y>3)

[1] 0.005467259

sum(dbinom(c(2,3),n,p)) # P(1<Y<4) = P(Y=2)+P(Y=3)

[1] 0.1654853
```

Poisson experiments

Many experiments can be thought of as "how many *rare* events will occur in a certain amount of time or space?" For example,

- # of alpha particles emitted from a polonium bar in an 8 minute period
- # of flaws on a standard size piece of manufactured product, e.g., 100m coaxial cable, 100 sq.meter plastic sheeting
- # of hits on a web page in a 24h period

Poisson random variable

A Poisson random variable has pmf

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, \ldots$$

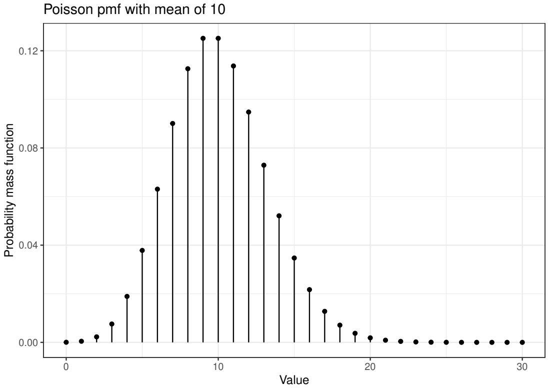
where λ is called the rate parameter.

We write $X \sim Po(\lambda)$ to represent this random variable. We can show that

$$E[X] = Var[X] = \lambda.$$

Poisson probability mass function

Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day. What is the probability that more than 8 new accounts will be initiated today?



Poisson probability

Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day. What is the probability that more than 8 new accounts will be initiated today?

Let X be the number of accounts initiated today. Assume $X \sim Po(10)$.

$$P(X > 8) = 1 - P(X \leq 8) = 1 - \sum_{x=0}^{8} \frac{\lambda^x e^{-\lambda}}{x!} \approx 1 - 0.333 = 0.667$$

In R,

```
# Using pmf
1-sum(dpois(0:8, lambda=10))

[1] 0.6671803

# Using cdf
1-ppois(8, lambda=10)

[1] 0.6671803
```

Sum of Poisson random variables

Let $X_i \overset{ind}{\sim} Po(\lambda_i)$ for $i = 1, \ldots, n$. Then

$$Y = \sum_{i=1}^{n} X_i \sim Po\left(\sum_{i=1}^{n} \lambda_i\right).$$

Let $X_i \overset{iid}{\sim} Po(\lambda)$ for $i = 1, \ldots, n$. Then

$$Y = \sum_{i=1}^{n} X_i \sim Po\left(n\lambda\right).$$

Poisson random variable - example

Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day. What is the probability that more than 16 new accounts will be initiated in the next two days?

Since the rate is 10/day, then for two days we expect, on average, to have 20. Let Y be he number initiated in a two-day period and assume $Y \sim Po(20)$. Then

$$\begin{aligned} P(Y > 16) &= 1 - P(Y \leq 16) \\ &= 1 - \textstyle\sum_{x=0}^{16} \frac{\lambda^x e^{-\lambda}}{x!} \\ &= 1 - 0.221 = 0.779. \end{aligned}$$

In R,

```r
# Using pmf
1-sum(dpois(0:16, lambda=20))

[1] 0.7789258

# Using cdf
1-ppois(16, lambda=20)

[1] 0.7789258
```

Manufacturing example

A manufacturer produces 100 chips per day and, on average, 1% of these chips are defective. What is the probability that no defectives are found in a particular day?

Let X represent the number of defectives and assume $X \sim Bin(100, 0.01)$. Then

$$P(X = 0) = \binom{100}{0}(0.01)^0(1 - 0.01)^{100} \approx 0.366.$$

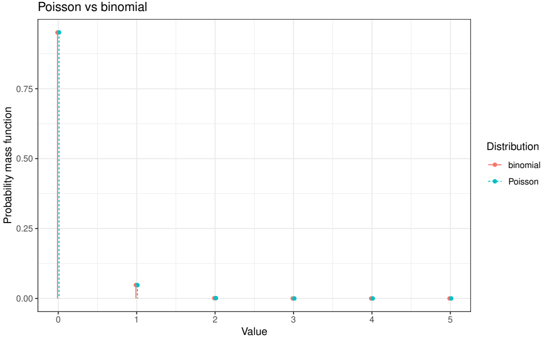
Alternatively, let Y represent the number of defectives and assume $Y \sim Po(100 \times 0.01)$. Then

$$P(Y = 0) = \frac{1^0 e^{-1}}{0!} \approx 0.368.$$

Poisson approximation to the binomial

Suppose we have $X \sim Bin(n, p)$ with n large (say ≥ 20) and p small (say ≤ 0.05). We can approximate X by $Y \sim Po(np)$ because for large n and small p

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-np} \frac{(np)^k}{k!}.$$



Example

Imagine you are supposed to proofread a paper. Let us assume that there are on average 2 typos on a page and a page has 1000 words. This gives a probability of 0.002 for each word to contain a typo. What is the probability the page has no typos?

Let X represent the number of typos on the page and assume $X \sim Bin(1000, 0.002)$. $P(X = 0)$ using R is

```
n = 1000; p = 0.002
dbinom(0, size=n, prob=p)

[1] 0.1350645
```

Alternatively, let Y represent the number of defectives and assume $Y \sim Po(1000 \times 0.002)$. $P(Y = 0)$ using R is

```
dpois(0, lambda = n*p)

[1] 0.1353353
```

Summary

- General discrete random variables
  - Probability mass function (pmf)
  - Cumulative distribution function (cdf)
  - Expected value
  - Variance
  - Standard deviation
- Specific discrete random variables
  - Bernoulli
  - Binomial
  - Poisson

P3 - Continuous random variables

STAT 587 (Engineering)
Iowa State University

August 22, 2020

Continuous vs discrete random variables

Discrete random variables have

- finite or countable support and
- pmf: $P(X = x)$.

Continuous random variables have

- uncountable support and
- $P(X = x) = 0$ for all x.

Cumulative distribution function

The cumulative distribution function for a continuous random variable is

$$F_X(x) = P(X \leq x) = P(X < x)$$

since $P(X = x) = 0$ for any x.

The cdf still has the properties

- $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$,
- F_X is monotone increasing,
  i.e. if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$, and
- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.

Probability density function

The probability density function (pdf) for a continuous random variable is

$$f_X(x) = \frac{d}{dx} F_X(x)$$

and

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Thus, the pdf has the following properties

- $f_X(x) \geq 0$ for all x and
- $\int_{-\infty}^{\infty} f(x)dx = 1.$

Example

Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

$f_X(x)$ defines a valid pdf because $f_X(x) \geq 0$ for all x and

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 3x^2 dx = x^3|_0^1 = 1.$$

The cdf is

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x^3 & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}.$$

Expected value

Let X be a continuous random variable and h be some function. The expected value of a function of a continuous random variable is

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) \cdot f_X(x) dx.$$

If $h(x) = x$, then

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

and we call this the expectation of X. We commonly use the symbol μ for this expectation.

Example (cont.)

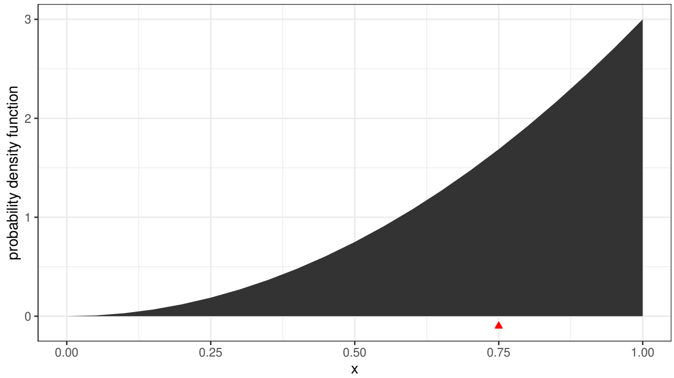
Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The expected value is

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx \\ &= \int_0^1 3x^3 dx \\ &= 3\frac{x^4}{4}\Big|_0^1 = \frac{3}{4}. \end{aligned}$$

Example - Center of mass



Variance

The variance of a random variable is defined as the expected squared deviation from the mean. For continuous random variables, variance is

$$Var[X] = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$$

where $\mu = E[X]$. The symbol σ^2 is commonly used for the variance.

The standard deviation is the positive square root of the variance

$$SD[X] = \sqrt{Var[X]}.$$

The symbol σ is commonly used for the standard deviation.

Example (cont.)

Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The variance is

$$\begin{aligned} Var[X] &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx \\ &= \int_0^1 \left(x - \tfrac{3}{4}\right)^2 3x^2 dx \\ &= \int_0^1 \left[x^2 - \tfrac{3}{2}x + \tfrac{9}{16}\right] 3x^2 dx \\ &= \int_0^1 3x^4 - \tfrac{9}{2}x^3 + \tfrac{27}{16}x^2 dx \\ &= \left[\tfrac{3}{5}x^5 - \tfrac{9}{8}x^4 + \tfrac{9}{16}x^3\right] |_0^1 dx \\ &= \frac{3}{5} - \frac{9}{8} + \frac{9}{16} \\ &= \frac{3}{80}. \end{aligned}$$

Comparison of discrete and continuous random variables

For simplicity here and later, we drop the subscript X.

	discrete	continuous
support ($\mathcal{X}$)	finite or countable	uncountable
pmf	$p(x) = P(X = x)$	
pdf		$p(x) = f(x) = F'(x)$
cdf	$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{t \leq x} p(t) \end{aligned}$	$\begin{aligned} F(x) &= P(X \leq x) = P(X < x) \\ &= \int_{-\infty}^{x} p(t)\, dt \end{aligned}$
expected value	$E[h(X)] = \sum_{x \in \mathcal{X}} h(x) p(x)$	$E[h(X)] = \int_{\mathcal{X}} h(x) p(x)\, dx$
expectation	$\mu = E[X] = \sum_{x \in \mathcal{X}} x\, p(x)$	$\mu = E[X] = \int_{\mathcal{X}} x\, p(x)\, dx$
variance	$\begin{aligned} \sigma^2 = Var[X] &= E[(X - \mu)^2] \\ &= \sum_{x \in \mathcal{X}} (x - \mu)^2\, p(x) \end{aligned}$	$\begin{aligned} \sigma^2 = Var[X] &= E[(X - \mu)^2] \\ &= \int_{\mathcal{X}} (x - \mu)^2\, p(x)\, dx \end{aligned}$

Note: we replace summations with integrals when using continuous as opposed to discrete random variables

Uniform

A uniform random variable on the interval (a, b) has equal probability for any value in that interval and we denote this $X \sim Unif(a, b)$. The pdf for a uniform random variable is

$$f(x) = \frac{1}{b-a}\mathrm{I}(a < x < b)$$

where $\mathrm{I}(A)$ is in indicator function that is 1 if A is true and 0 otherwise, i.e.

$$\mathrm{I}(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & \text{otherwise.} \end{cases}$$

The expectation is

$$E[X] = \int_a^b x \, \frac{1}{b-a} \, dx = \frac{a+b}{2}$$

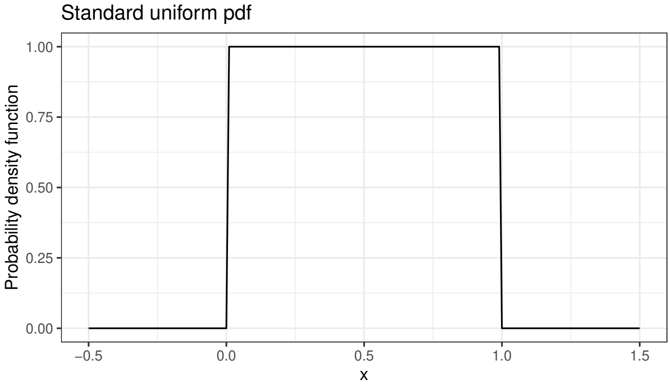
and the variance is

$$Var[X] = \int_a^b \frac{1}{b-a} \left(x - \frac{a+b}{2} \right)^2 dx = \frac{1}{12}(b-a)^2.$$

Standard uniform

A standard uniform random variable is $X \sim Unif(0,1)$. This random variable has

$$E[X] = \frac{1}{2} \qquad \text{and} \qquad Var[X] = \frac{1}{12}.$$



Example (cont.)

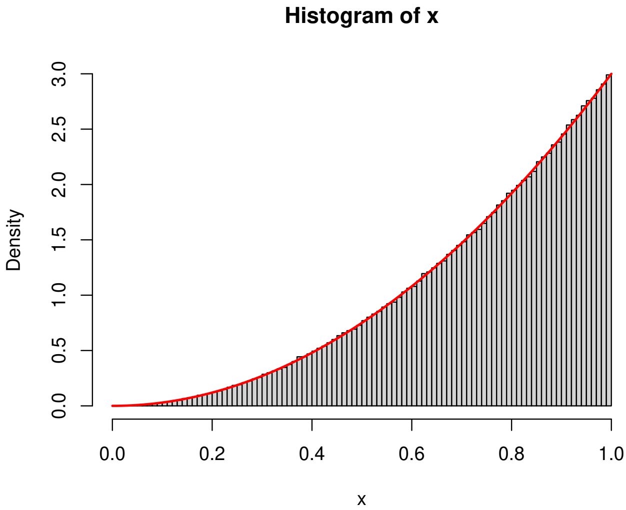
Pseudo-random number generators generate pseudo uniform values on (0,1). These values can be used in conjunction with the inverse of the cumulative distribution function to generate pseudo-random numbers from any distribution.

The inverse of the cdf $F_X(x) = x^3$ is

$$F_X^{-1}(u) = u^{1/3}.$$

A uniform random number on the interval (0,1) generated using the inverse cdf produces a random draw of X.

```
inverse_cdf = function(u) u^(1/3)
x = inverse_cdf(runif(1e6))
mean(x)

[1] 0.7502002

var(x); 3/80

[1] 0.03752111
[1] 0.0375
```


Histogram of x
Density
x
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0

Normal random variable

The normal (or Gaussian) density is a "bell-shaped" curve. The density has two parameters: mean μ and variance σ^2 and is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \qquad \text{for } -\infty < x < \infty$$

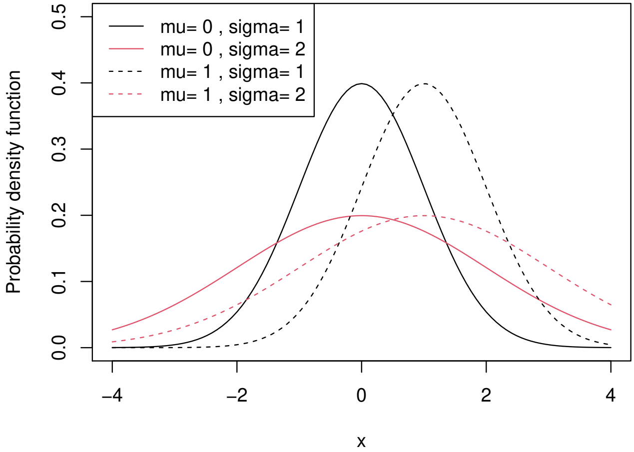
If $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x\, f(x) dx = \dots &&= \mu \\ Var[X] &= \int_{-\infty}^{\infty} (x-\mu)^2\, f(x) dx = \dots &&= \sigma^2. \end{aligned}$$

Thus, the parameters μ and σ^2 are actually the mean and the variance of the $N(\mu, \sigma^2)$ distribution.

There is no closed form cumulative distribution function for a normal random variable.

Example normal probability density functions



Properties of normal random variables

Let $Z \sim N(0,1)$, i.e. a standard normal random variable. Then for constants μ and σ

$$X = \mu + \sigma Z \sim N(\mu, \sigma^2)$$

and

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

which is called standardizing.

Let $X_i \stackrel{ind}{\sim} N(\mu_i, \sigma_i^2)$. Then

$$Z_i = \frac{X_i - \mu_i}{\sigma_i} \stackrel{iid}{\sim} N(0,1) \quad \text{for all } i$$

and

$$Y = \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Calculating the standard normal cdf

If $Z \sim N(0,1)$, what is $P(Z \leq 1.5)$? Although the cdf does not have a closed form, very good approximations exist and are available as tables or in software, e.g.

```
pnorm(1.5) # default is mean=0, sd=1

[1] 0.9331928
```

If $Z \sim N(0,1)$, then

- $P(Z \leq z) = \Phi(z)$
- $\Phi(z) = 1 - \Phi(-z)$ since a normal pdf is symmetric around its mean.

Calculating any normal cumulative distribution function

If $X \sim N(15, 4)$ what is $P(X > 18)$?

$$
\begin{aligned}
P(X > 18) \quad &= 1 - P(X \leq 18) \\
&= 1 - P\left(\frac{X-15}{2} \leq \frac{18-15}{2}\right) \\
&= 1 - P(Z \leq 1.5) \\
&\approx 1 - 0.933 = 0.067
\end{aligned}
$$

```
1-pnorm((18-15)/2)          # by standardizing

[1] 0.0668072

1-pnorm(18, mean = 15, sd = 2) # using the mean and sd arguments

[1] 0.0668072
```

Manufacturing

Suppose you are producing nails that must be within 5 and 6 centimeters in length. If the average length of nails the process produces is 5.3 cm and the standard deviation is 0.1 cm. What is the probability the next nail produced is outside of the specification?

Let $X \sim N(5.3, 0.1^2)$ be the length (cm) of the next nail produced. We need to calculate

$$P(X < 5 \text{ or } X > 6) = 1 - P(5 < X < 6).$$

```
mu = 5.3
sigma = 0.1

1-diff(pnorm(c(5,6), mean = mu, sd = sigma))

[1] 0.001349898
```

Summary

- Continuous random variables
  - Probability density function
  - Cumulative distribution function
  - Expectation
  - Variance
- Specific distributions
  - Uniform
  - Normal (or Gaussian)

P4 - Central Limit Theorem

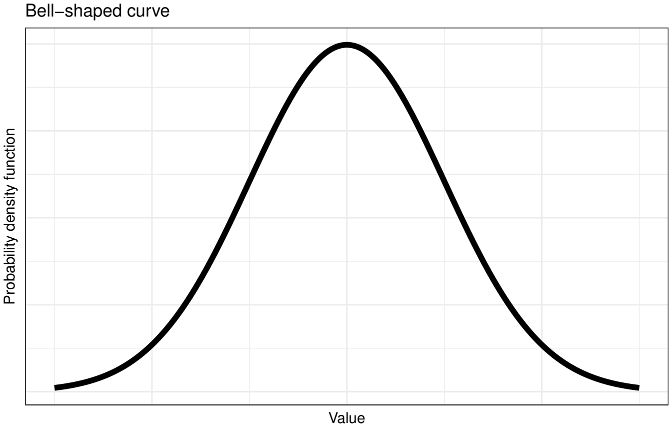
STAT 587 (Engineering)
Iowa State University

August 28, 2020

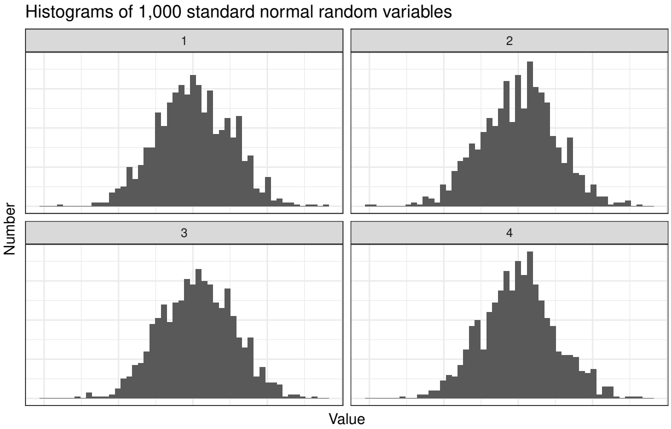
Main Idea: Sums and averages of iid random variables from any distribution have approximate normal distributions for sufficiently large sample sizes.

Bell-shaped curve

The term bell-shaped curve typically refers to the probability density function for a normal random variable:

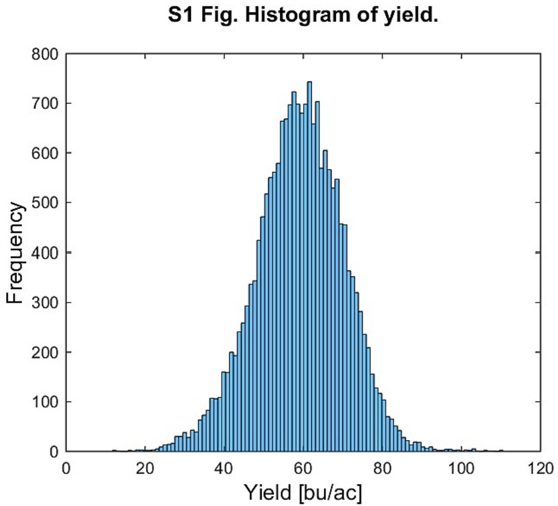


Histograms of samples from bell-shaped curves



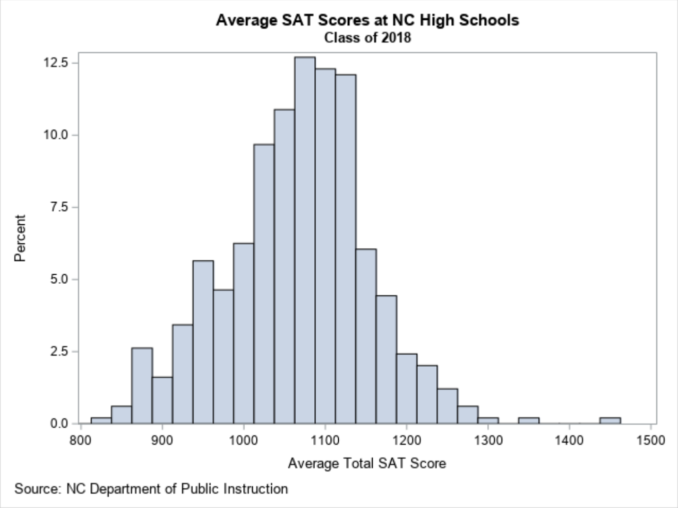
Yield

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184198

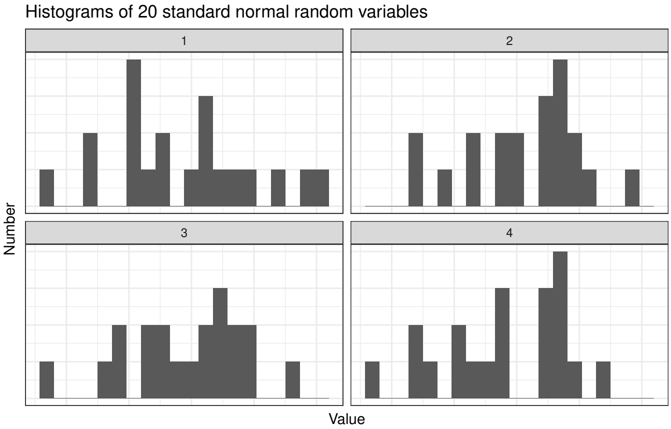


SAT scores

https://blogs.sas.com/content/iml/2019/03/04/visualize-sat-scores-nc.html

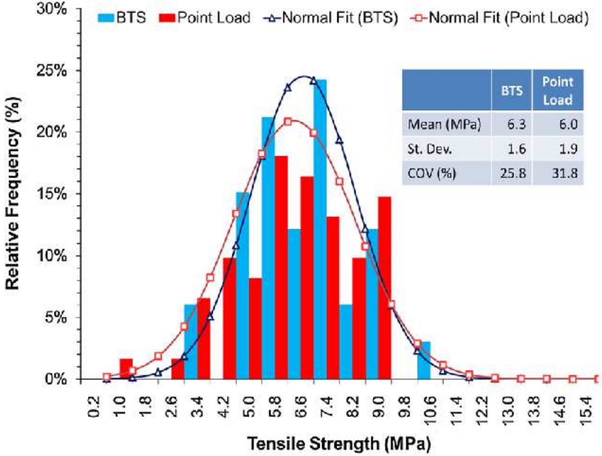


Histograms of samples from bell-shaped curves



Tensile strength

https://www.researchgate.net/figure/Comparison-of-histograms-for-BTS-and-tensile-strength-estimated-from-point-load_fig5_260617256



	BTS	Point Load
Mean (MPa)	6.3	6.0
St. Dev.	1.6	1.9
COV (%)	25.8	31.8

Sums and averages of iid random variables

Suppose $X_1, X_2, \ldots$ are iid random variables with

$$E[X_i] = \mu \quad Var[X_i] = \sigma^2.$$

Define

$$\begin{aligned} \text{Sample Sum: } S_n &= X_1 + X_2 + \cdots + X_n \\ \text{Sample Average: } \overline{X}_n &= S_n/n. \end{aligned}$$

For S_n, we know

$$E[S_n] = n\mu, \quad Var[S_n] = n\sigma^2, \quad \text{and} \quad SD[S_n] = \sqrt{n}\sigma.$$

For $\overline{X}_n$, we know

$$E[\overline{X}_n] = \mu, \quad Var[\overline{X}_n] = \sigma^2/n, \quad \text{and} \quad SD[\overline{X}_n] = \sigma/\sqrt{n}.$$

Central Limit Theorem (CLT)

Suppose $X_1, X_2, \ldots$ are iid random variables with

$$E[X_i] = \mu \quad Var[X_i] = \sigma^2.$$

Define

$$\begin{aligned} \text{Sample Sum: } S_n &= X_1 + X_2 + \cdots + X_n \\ \text{Sample Average: } \overline{X}_n &= S_n/n. \end{aligned}$$

Then the Central Limit Theorem says

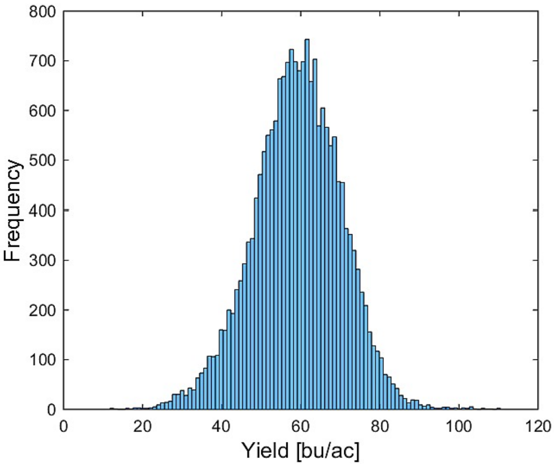
$$\lim_{n\to\infty} \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \quad \text{and} \quad \lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1).$$

Main Idea: Sums and averages of iid random variables from any distribution have approximate normal distributions for sufficiently large sample sizes.

Yield

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184198

S1 Fig. Histogram of yield.



Approximating distributions

Rather than considering the limit, I typically think of the following approximations as n gets large.

For the sample average,

$$\overline{X}_n \dot{\sim} N(\mu, \sigma^2/n).$$

where $\dot{\sim}$ indicates *approximately distributed* because

$$E\left[\overline{X}_n\right] = \mu \qquad \text{and} \qquad Var\left[\overline{X}_n\right] = \sigma^2/n.$$

For the sample sum,

$$S_n \dot{\sim} N(n\mu, n\sigma^2)$$

because

$$\begin{aligned} E[S_n] &= n\mu \\ Var[S_n] &= n\sigma^2. \end{aligned}$$

Averages and sums of uniforms

Let $X_i \stackrel{ind}{\sim} Unif(0,1)$. Then

$$\mu = E[X_i] = \frac{1}{2} \quad \text{and} \quad \sigma^2 = Var[X_i] = \frac{1}{12}.$$

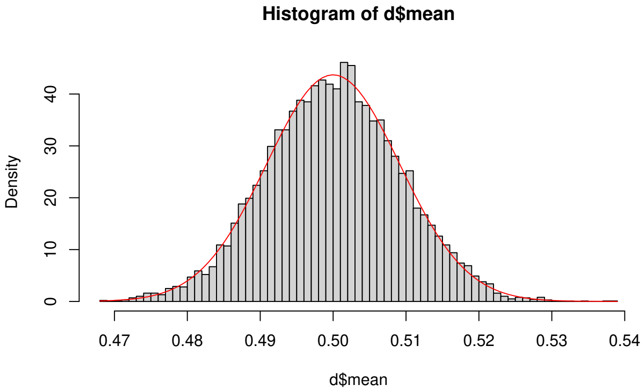
Thus

$$\overline{X}_n \dot{\sim} N\left(\frac{1}{2}, \frac{1}{12n}\right)$$

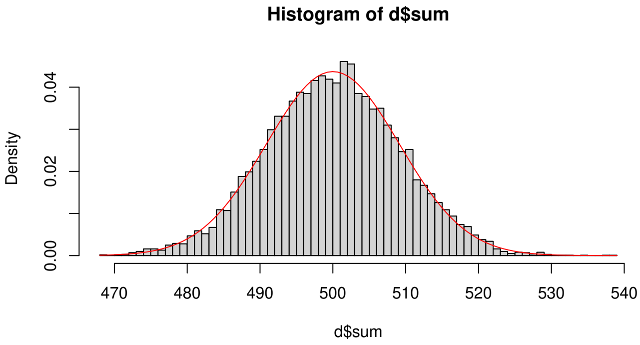
and

$$S_n \dot{\sim} N\left(\frac{n}{2}, \frac{n}{12}\right).$$

Averages of uniforms



Sums of uniforms



Normal approximation to a binomial

Recall if $Y_n = \sum_{i=1}^{n} X_i$ where $X_i \overset{ind}{\sim} Ber(p)$, then

$$Y_n \sim Bin(n, p).$$

For a binomial random variable, we have

$$E[Y_n] = np \qquad \text{and} \qquad Var[Y_n] = np(1-p).$$

By the CLT,

$$\lim_{n\to\infty} \frac{Y_n - np}{\sqrt{np(1-p)}} \to N(0,1),$$

if n is large,

$$Y_n \dot{\sim} N(np, np[1-p]).$$

Roulette example

A European roulette wheel has 39 slots: one green, 19 black, and 19 red. If I play black every time, what is the probability that I will have won more than I lost after 99 spins of the wheel?

https://isorepublic.com/photo/roulette-wheel/

Roulette example

A European roulette wheel has 39 slots: one green, 19 black, and 19 red. If I play black every time, what is the probability that I will have won more than I lost after 99 spins of the wheel?

Let Y indicate the total number of wins and assume $Y \sim Bin(n, p)$ with $n = 99$ and $p = 19/39$. The desired probability is $P(Y \geq 50)$. Then

$$P(Y \geq 50) = 1 - P(Y < 50) = 1 - P(Y \leq 49)$$

```
n = 99
p = 19/39
1-pbinom(49, n, p)

[1] 0.399048
```

Roulette example

A European roulette wheel has 39 slots: one green, 19 black, and 19 red. If I play black every time, what is the probability that I will have won more than I lost after 99 spins of the wheel?

Let Y indicate the total number of wins. We can approximate Y using $X \sim N(np, np(1-p))$.

$$P(Y \geq 50) \approx 1 - P(X < 50)$$

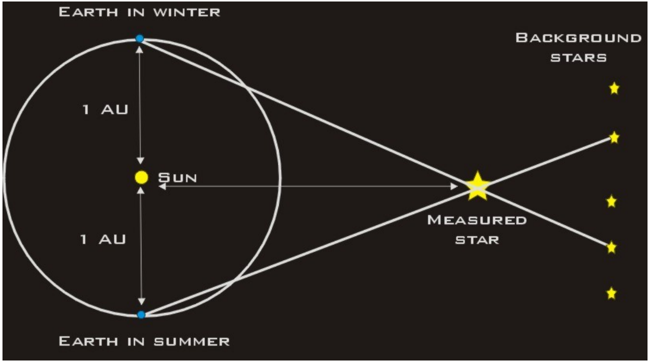
```
1-pnorm(50, n*p, sqrt(n*p*(1-p)))

[1] 0.3610155
```

A better approximation can be found using a continuity correction.

Astronomy example

An astronomer wants to measure the distance, d, from Earth to a star. Suppose the procedure has a known standard deviation of 2 parsecs. The astronomer takes 30 iid measurements and finds the average of these measurements to be 29.4 parsecs. What is the probability the average is within 0.5 parsecs?

http://planetary-science.org/astronomy/distance-and-magnitudes/



Astronomy example

Let X_i be the i^{th} measurement. The astronomer assumes that $X_1, X_2, \ldots X_n$ are iid with $E[X_i] = d$ and $Var[X_i] = \sigma^2 = 2^2$. The estimate of d is

$$\overline{X}_n = \frac{(X_1 + X_2 + \cdots + X_n)}{n} = 29.4.$$

and, by the Central Limit Theorem, $\overline{X}_n \dot{\sim} N(d, \sigma^2/n)$ where $n = 30$. We want to find

$$\begin{aligned} P\left(|\overline{X}_n - d| < 0.5\right) &= P\left(-0.5 < \overline{X}_n - d < 0.5\right) \\ &= P\left(\frac{-0.5}{2/\sqrt{30}} < \frac{\overline{X}_n - d}{\sigma/\sqrt{n}} < \frac{0.5}{2/\sqrt{30}}\right) \\ &\approx P(-1.37 < Z < 1.37) \end{aligned}$$

```
diff(pnorm(c(-1.37,1.37)))
```

```
[1] 0.8293131
```

Astronomy example - sample size

Suppose the astronomer wants to be within 0.5 parsecs with at least 95% probability. How many more samples would she need to take?

We solve

$$
\begin{aligned}
0.95 \le P\left(\left|\overline{X}_n - d\right| < .5\right) &= P\left(-0.5 < \overline{X}_n - d < 0.5\right) \\
&= P\left(\frac{-0.5}{2/\sqrt{n}} < \frac{\overline{X}_n - d}{\sigma/\sqrt{n}} < \frac{0.5}{2/\sqrt{n}}\right) \\
&= P(-z < Z < z) \qquad\qquad z = 0.5/(2/\sqrt{n}) \\
&= 1 - [P(Z < -z) + P(Z > z)] \\
&= 1 - 2P(Z < -z)
\end{aligned}
$$

where $z = 1.96$ since

```
1-2*pnorm(-1.96)

[1] 0.9500042
```

and thus $n = 61.47$ which we round up to $n = 62$ to ensure

Summary

- Central Limit Theorem
  - Sums
  - Averages
- Examples
  - Uniforms
  - Binomial
    - Roulette
- Sample size
  - Astronomy

P5 - Multiple random variables

STAT 587 (Engineering)
Iowa State University

September 4, 2020

Multiple discrete random variables

If X and Y are two discrete variables, their joint probability mass function is defined as

$$p_{X,Y}(x,y) = P(X = x \cap Y = y) = P(X = x, Y = y).$$

CPU example

A box contains 5 PowerPC G4 processors of different speeds:

#	speed
2	400 mHz
1	450 mHz
2	500 mHz

Randomly select two processors out of the box (without replacement) and let

- X be speed of the first selected processor and
- Y be speed of the second selected processor.

CPU example - outcomes

		1st processor (X)				
	Ω	400_1	400_2	450	500_1	500_2
2nd processor (Y)	400_1	-	x	x	x	x
	400_2	x	-	x	x	x
	450	x	x	-	x	x
	500_1	x	x	x	-	x
	500_2	x	x	x	x	-

Reasonable to believe each outcome is equally probable.

CPU example - joint pmf

Joint probability mass function for X and Y:

	mHz	1st processor (X) 400	450	500
2nd processor (Y)	400	2/20	2/20	4/20
	450	2/20	0/20	2/20
	500	4/20	2/20	2/20

- What is $P(X = Y)$?
- What is $P(X > Y)$?

CPU example - probabilities

What is the probability that $X = Y$?

$$
\begin{aligned}
&P(X = Y) \\
&\quad = p_{X,Y}(400, 400) + p_{X,Y}(450, 450) + p_{X,Y}(500, 500) \\
&\quad = 2/20 + 0/20 + 2/20 = 4/20 = 0.2
\end{aligned}
$$

What is the probability that $X > Y$?

$$
\begin{aligned}
&P(X > Y) \\
&\quad = p_{X,Y}(450, 400) + p_{X,Y}(500, 400) + p_{X,Y}(500, 450) \\
&\quad = 2/20 + 4/20 + 2/20 = 8/20 = 0.4
\end{aligned}
$$

Expectation

The expected value of a function $h(x, y)$ is

$$E[h(X,Y)] = \sum_{x,y} h(x,y)\, p_{X,Y}(x,y).$$

CPU example - expected absolute speed difference

What is $E[|X-Y|]$?

Here, we have the situation $E[|X-Y|] = E[h(X,Y)]$, with $h(X,Y) = |X-Y|$. Thus, we have

$$
\begin{aligned}
&E[|X-Y|] \\
&\quad = \sum_{x,y} |x-y|\, p_{X,Y}(x,y) = \\
&\quad = |400-400| \cdot 0.1 + |400-450| \cdot 0.1 + |400-500| \cdot 0.2 \\
&\qquad + |450-400| \cdot 0.1 + |450-450| \cdot 0.0 + |450-500| \cdot 0.1 \\
&\qquad + |500-400| \cdot 0.2 + |500-450| \cdot 0.1 + |500-500| \cdot 0.1 \\
&\quad = 0+5+20+5+0+5+20+5+0 = 60.
\end{aligned}
$$

Marginal distribution

For discrete random variables X and Y, the marginal probability mass functions are

$$\begin{aligned} p_X(x) &= \textstyle\sum_y p_{X,Y}(x,y) \qquad \text{and} \\ p_Y(y) &= \textstyle\sum_x p_{X,Y}(x,y) \end{aligned}$$

Marginal distribution

Joint probability mass function for X and Y:

	mHz	1st processor (X) 400	450	500
	400	2/20	2/20	4/20
2nd processor (Y)	450	2/20	0/20	2/20
	500	4/20	2/20	2/20

Summing the rows within each column provides

x	400	450	500
$p_X(x)$	0.4	0.2	0.4

Summing the columns within each row provides

y	400	450	500
$p_Y(y)$	0.4	0.2	0.4

CPU example - independence

Are X and Y independent?

X and Y are independent if $p_{x,y}(x, y) = p_X(x)p_Y(y)$ for all x and y.

Since

$$p_{X,Y}(450, 450) = 0 \neq 0.2 \cdot 0.2 = p_X(450) \cdot p_Y(450)$$

they are **not** independent.

Covariance

The covariance between two random variables X and Y is

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

where

$$\mu_X = E[X] \qquad \text{and} \qquad \mu_Y = E[X].$$

If $Y = X$ in the above definition, then $Cov[X, X] = Var[X]$.

CPU example - covariance

Use marginal pmfs to compute:

$$E[X] = E[Y] = 450 \qquad \text{and} \qquad Var[X] = Var[Y] = 2000.$$

The covariance between X and Y is:

$$\begin{aligned}
&Cov[X,Y] \\
&= \sum_{x,y}(x - E[X])(y - E[Y])p_{X,Y}(x,y) = \\
&= (400 - 450)(400 - 450) \cdot 0.1 \\
&\quad +(450 - 450)(400 - 450) \cdot 0.1 \\
&\quad +\cdots \\
&\quad +(500 - 450)(500 - 450) \cdot 0.1 \\
&= 250 + 0 - 500 + 0 + 0 + 0 - 500 + 250 + 0 \\
&= -500.
\end{aligned}$$

Correlation

The correlation between two variables X and Y is

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X] \cdot Var[Y]}} = \frac{Cov[X,Y]}{SD[X] \cdot SD[Y]}.$$

Correlation properties

- ρ is between -1 and 1
- if $\rho = 1$ or -1, Y is a linear function of X:
  - $\rho = 1 \implies Y = mX + b$ with $m > 0$,
  - $\rho = -1 \implies Y = mX + b$ with $m < 0$,
- ρ is a measure of linear association between X and Y
  - ρ near ± 1 indicates a strong linear relationship,
  - ρ near 0 indicates a lack of linear association.

CPU example - correlation

Recall

$$Cov[X, Y] = -500 \qquad \text{and} \qquad Var[X] = Var[Y] = 2000.$$

The correlation is

$$\rho[X, Y] = \frac{Cov[X, Y]}{\sqrt{Var[X] \cdot Var[Y]}} = \frac{-500}{\sqrt{2000 \cdot 2000}} = -0.25,$$

and thus there is a weak negative (linear) association.

Continuous random variables

Suppose X and Y are two continuous random variables with joint probability density function $p_{X,Y}(x, y)$. Probabilities are calculated by integrating this function. For example,

$$P(a < X < b, c < Y < d) = \int_c^d \int_a^b p_{X,Y}(x, y)\, dx\, dy.$$

Then the marginal probability density functions are

$$\begin{aligned} p_X(x) &= \int p_{X,Y}(x, y)\, dy \\ p_Y(y) &= \int p_{X,Y}(x, y)\, dx. \end{aligned}$$

Continuous random variables

Two continuous random variables are independent if

$$p_{X,Y}(x,y) = p_X(x)\, p_Y(y).$$

The expected value is

$$E[h(X,Y)] = \int \int h(x,y)\, p_{X,Y}(x,y)\, dx\, dy.$$

Properties of variances and covariances

For any random variables X, Y, W and Z,

$$
\begin{aligned}
Var[aX + bY + c] &= a^2 Var[X] + b^2 Var[Y] + 2ab Cov[X, Y] \\
\\
Cov[aX + bY, cZ + dW] &= ac Cov[X, Z] + ad Cov[X, W] \\
&\quad + bc Cov[Y, Z] + bd Cov[Y, W] \\
\\
Cov[X, Y] &= Cov[Y, X] \\
\rho[X, Y] &= \rho[Y, X]
\end{aligned}
$$

If X and Y are independent, then

$$
\begin{aligned}
Cov[X, Y] &= 0 \\
Var[aX + bY + c] &= a^2 Var[X] + b^2 Var[Y].
\end{aligned}
$$

Summary

- Multiple random variables
  - joint probability mass function
  - marginal probability mass function
  - joint probability density function
  - marginal probability density function
  - expected value
  - covariance
  - correlation

Exponential distribution

STAT 587 (Engineering)
Iowa State University

September 17, 2020

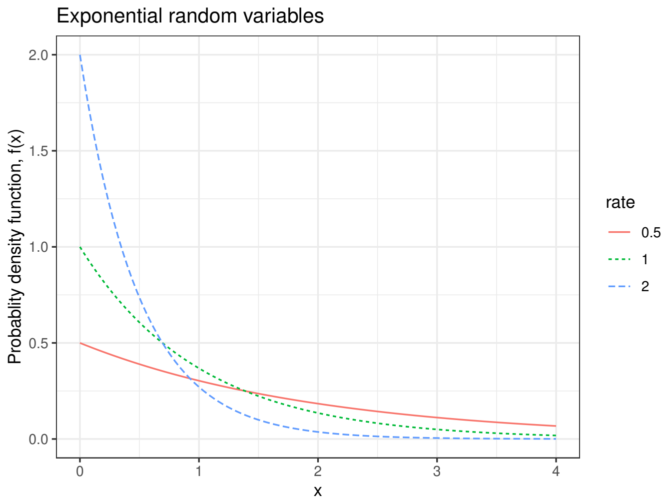
Exponential distribution

The random variable X has an exponential distribution with rate parameter $\lambda > 0$ if its probability density function is

$$p(x|\lambda) = \lambda e^{-\lambda x}\,\mathrm{I}(x > 0).$$

We write $X \sim Exp(\lambda)$.

Exponential probability density function



Exponential mean and variance

If $X \sim Exp(\lambda)$, then

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

and

$$Var[X] = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \cdots = \frac{1}{\lambda^2}.$$

Exponential cumulative distribution function

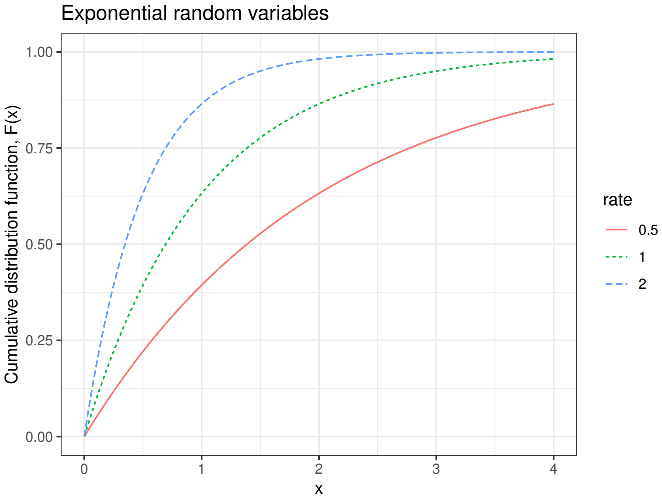
If $X \sim Exp(\lambda)$, then its cumulative distribution function is

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = \cdots = 1 - e^{-\lambda x}.$$

The inverse cumulative distribution function is

$$F^{-1}(p) = \frac{-\log(1-p)}{\lambda}.$$

Exponential cumulative distribution function - graphically



Memoryless property

Let $X \sim Exp(\lambda)$, then

$$P(X > x + c | X > c) = P(X > x).$$

Parameterization by the scale

A common alternative parameterization of the exponential distribution uses the scale $\beta = \frac{1}{\lambda}$. In this parameterization, we have

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \, \mathrm{I}(x > 0)$$

and

$$E[X] = \beta \qquad \text{and} \qquad Var[X] = \beta^2.$$

Summary

Exponential random variable

- $X \sim Exp(\lambda)$, $\lambda > 0$
- $f(x) = \lambda e^{-\lambda x}$, $x > 0$
- $F(x) = 1 - e^{-\lambda x}$
- $F^{-1}(p) = \frac{-\log(1-p)}{\lambda}$
- $E[X] = \frac{1}{\lambda}$
- $Var[X] = \frac{1}{\lambda^2}$

Gamma distribution

STAT 587 (Engineering)
Iowa State University

September 17, 2020

Gamma distribution

The random variable X has a gamma distribution with

- shape parameter $\alpha > 0$ and
- rate parameter $\lambda > 0$

if its probability density function is

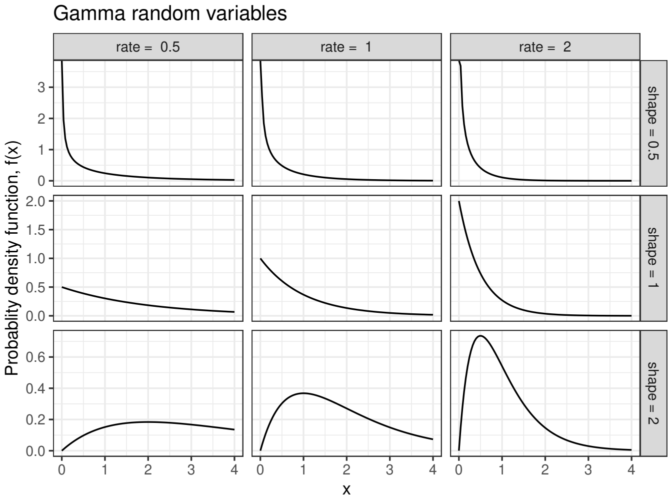
$$p(x|\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, \mathrm{I}(x > 0)$$

where $\Gamma(\alpha)$ is the gamma function,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

We write $X \sim Ga(\alpha, \lambda)$.

Gamma probability density function



Gamma mean and variance

If $X \sim Ga(\alpha, \lambda)$, then

$$E[X] = \int_0^\infty x \, \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \cdots = \frac{\alpha}{\lambda}$$

and

$$Var[X] = \int_0^\infty \left(x - \frac{\alpha}{\lambda}\right)^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \cdots = \frac{\alpha}{\lambda^2}.$$

Gamma cumulative distribution function

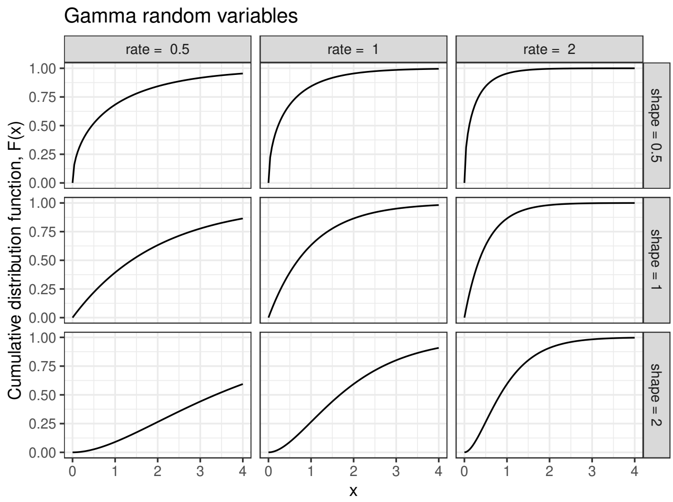
If $X \sim Ga(\alpha, \lambda)$, then its cumulative distribution function is

$$F(x) = \int_0^x \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} dt = \cdots = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

where $\gamma(\alpha, \beta x)$ is the incomplete gamma function, i.e.

$$\gamma(\alpha, \beta x) = \int_0^{\beta x} t^{\alpha-1} e^{-t} dt.$$

Gamma cumulative distribution function - graphically



Relationship to exponential distribution

If $X_i \overset{iid}{\sim} Exp(\lambda)$, then

$$Y = \sum_{i=1}^{n} X_i \sim Ga(n, \lambda).$$

Thus, $Ga(1, \lambda) \overset{d}{=} Exp(\lambda)$.

Parameterization by the scale

A common alternative parameterization of the Gamma distribution uses the scale $\theta = \frac{1}{\lambda}$. In this parameterization, we have

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}\, \mathrm{I}(x > 0)$$

and

$$E[X] = \alpha\theta \qquad \text{and} \qquad Var[X] = \alpha\theta^2.$$

Summary

Gamma random variable

- $X \sim Ga(\alpha, \lambda)$, $\alpha, \lambda > 0$
- $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, $x > 0$
- $E[X] = \frac{\alpha}{\lambda}$
- $Var[X] = \frac{\alpha}{\lambda^2}$

Inverse gamma distribution

STAT 587 (Engineering)
Iowa State University

September 17, 2020

Inverse gamma distribution

The random variable X has an inverse gamma distribution with

- shape parameter $\alpha > 0$ and
- scale parameter $\beta > 0$

if its probability density function is

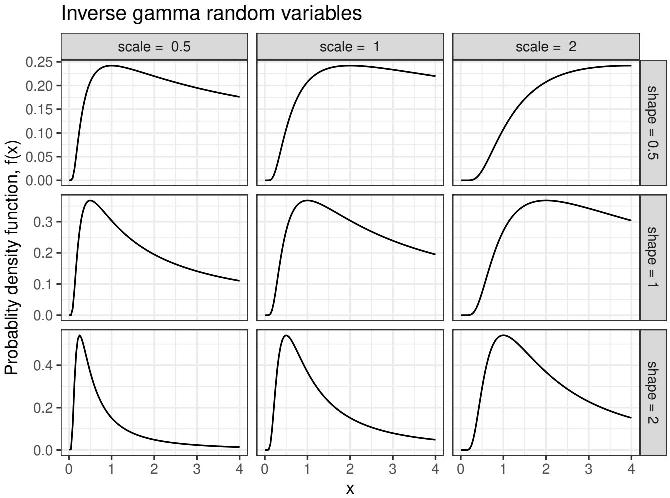
$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x} \mathrm{I}(x > 0).$$

where $\Gamma(\alpha)$ is the gamma function,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

We write $X \sim IG(\alpha, \beta)$.

Inverse gamma probability density function



Inverse gamma mean and variance

If $X \sim IG(\alpha, \beta)$, then

$$E[X] = \int_0^\infty x \, \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x} dx = \cdots = \frac{\beta}{\alpha - 1}, \quad \alpha > 1$$

and

$$\begin{aligned} Var[X] &= \int_0^\infty \left(x - \tfrac{\beta}{\alpha-1}\right)^2 \tfrac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x} dx \\ &= \cdots = \tfrac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad \alpha > 2. \end{aligned}$$

Relationship to gamma distribution

If $X \sim Ga(\alpha, \lambda)$ where λ is the rate parameter, then

$$Y = \frac{1}{X} \sim IG(\alpha, \lambda).$$

Summary

Inverse gamma random variable

- $X \sim IG(\alpha, \beta)$, $\alpha, \beta > 0$
- $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}, x > 0$
- $E[X] = \frac{\beta}{\alpha-1}$, $\alpha > 1$
- $Var[X] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$, $\alpha > 2$

Student's t-distribution

STAT 587 (Engineering)
Iowa State University

September 17, 2020

Student's t distribution

The random variable X has a Student's t distribution with degrees of freedom $\nu > 0$ if its probability density function is

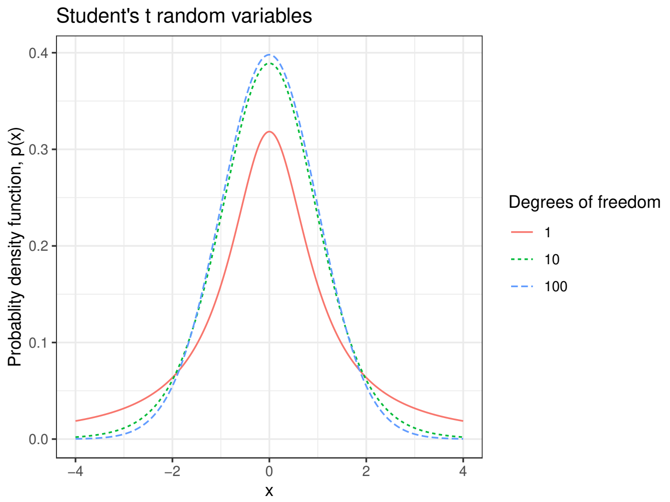
$$p(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where $\Gamma(\alpha)$ is the gamma function,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

We write $X \sim t_\nu$.

Student's t probability density function



Student's t mean and variance

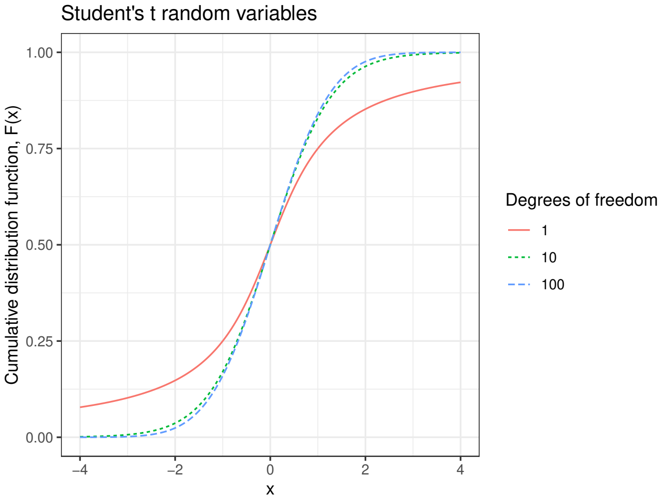
If $T \sim t_v$, then

$$E[X] = \int_{-\infty}^{\infty} x \, \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx = \cdots = 0, \quad \nu > 1$$

and

$$Var[X] = \int_{0}^{\infty} (x-0)^2 \, \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx = \cdots = \frac{\nu}{\nu-2}, \quad \nu > 2.$$

Gamma cumulative distribution function - graphically



Location-scale t distribution

If $X \sim t_\nu$, then

$$Y = \mu + \sigma X \sim t_\nu(\mu, \sigma^2)$$

for parameters:

- degrees of freedom $\nu > 0$,
- location μ and
- scale $\sigma > 0$.

By properties of expectations and variances, we can find that

$$E[Y] = \mu, \qquad \nu > 1$$

and

$$Var[Y] = \frac{\nu}{\nu - 2}\sigma^2, \qquad \nu > 2.$$

Generalized Student's t probability density function

The random variable Y has a generalized Student's t distribution with

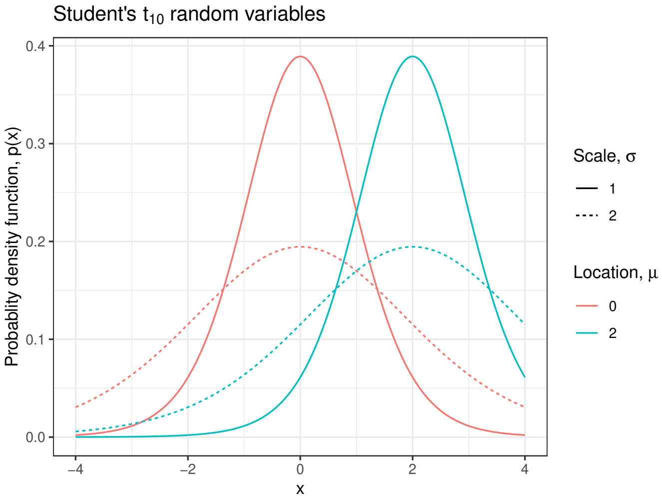
- degrees of freedom $\nu > 0$,
- location μ, and
- scale $\sigma > 0$

if its probability density function is

$$p(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left(1 + \frac{1}{\nu}\left[\frac{y-\mu}{\sigma}\right]^2\right)^{-\frac{\nu+1}{2}}$$

We write $Y \sim t_\nu(\mu, \sigma^2)$.

Generalized Student's t probability density function



t with 1 degree of freedom

If $T \sim t_1(\mu, \sigma^2)$, then T has a Cauchy distribution and we write

$$T \sim Ca(\mu, \sigma^2).$$

If $T \sim t_1(0, 1)$, then T has a standard Cauchy distribution. A Cauchy random variable has no mean or variance.

As degrees of freedom increases

If $T_\nu \sim t_\nu(\mu, \sigma^2)$, then

$$\lim_{\nu \to \infty} T_\nu \stackrel{d}{=} X \sim N(\mu, \sigma^2)$$

t distribution arising from a normal sample

Let $X_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$. We calculate the sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Then

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Inverse-gamma scale mixture of a normal

If

$$X|\sigma^2 \sim N(\mu, \sigma^2/n) \qquad \text{and} \qquad \sigma^2 \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}s^2\right)$$

then

$$X \sim t_\nu(\mu, s^2/n)$$

which is obtained by

$$p_x(x) = \int p_{x|\sigma^2}(x|\sigma^2)p_{\sigma^2}(\sigma^2)d\sigma^2$$

where

- p_x is the marginal density for x
- $p_{x|\sigma^2}$ is the conditional density for x given σ^2, and
- p_{σ^2} is the marginal density for σ^2.

Summary

Student's t random variable:

- $T \sim t_\nu(\mu, \sigma^2)$, $\nu, \sigma > 0$
- $E[X] = \mu$, $\nu > 1$
- $Var[X] = \frac{\nu}{\nu-2}\sigma^2$, $\nu > 2$
- Relationships to other distributions