Probability

Statistical Thinking (ETC2420 / ETC5242)

Week 2, Semester 2, 2025

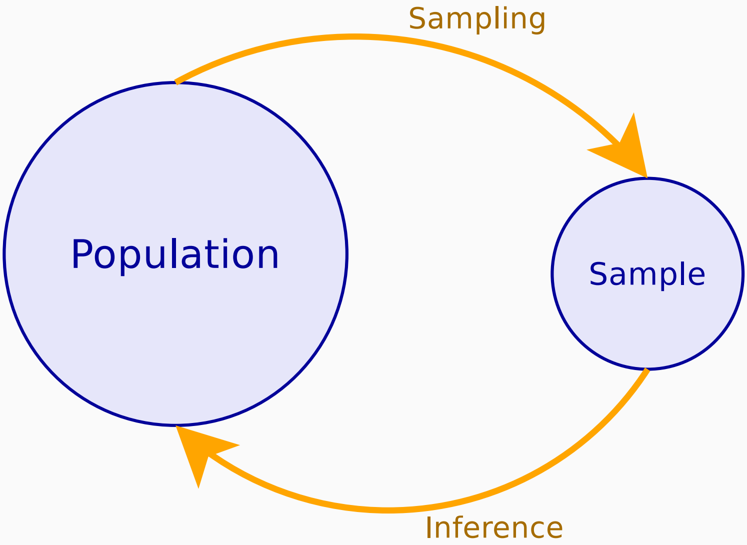
Outline

Outline

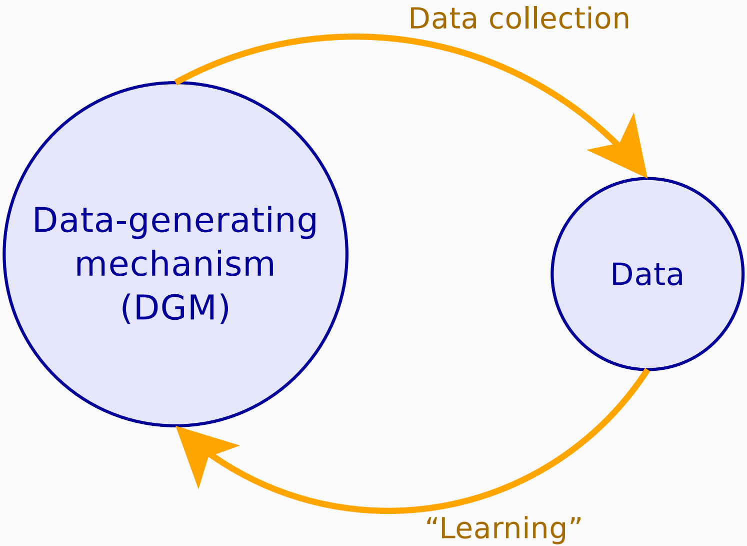
Learning goals for week 2

- Understand and apply elementary probability concepts.
- Understand and work with random variables.
- Identify common discrete and continuous univariate distributions.
- Use conditional probability and Bayes' Theorem.
- Understand the Law of Large Numbers and the Central Limit Theorem.
- Understand the vital role of randomisation in statistical inference.

Statistics: the big picture



Statistics: the big picture



Uses of probability

Abstractly:

- A tool to describe and understand apparent **randomness**
- A way to **characterise uncertainty**

Uses of probability

Abstractly:

- A tool to describe and understand apparent **randomness**
- A way to **characterise uncertainty**

More concretely:

- Model **variation** in data
- By modelling the data sampling/collection/observation process
- Think in terms of a **random process**, leading to an outcome
  - May be a single event (coin flip: H or T)
  - May be a sequence of events (HHTHTTHHTH...)

Working with probabilities

1. Mathematically
    - “Counting rules” (e.g. permutations, and other techniques)
    - These can be tricky…
2. Simulation on a computer
    - Often *much* easier!

Outline

Outcomes & sample spaces

- An **outcome** is a possible value that can be observed from some random process.
  - Examples: *H*, *T*
- The **sample space** (or **outcome space**) is the collection of all possible outcomes.
  - Can be finite or infinite.
  - Examples: $\{H, T\}$ or $\{0, 1, 2, \ldots\}$

Events

Events are subsets of outcomes. Some examples:

- Single outcomes

$$\{0\}, \{3\}$$

- Combinations of individual outcomes

$$\{0, 2, 7\}, \{3, 4\}$$

- The full sample space

$$\{H, T\}$$

- The empty set (no outcomes at all)

$$\{\}$$

Axioms of probability

The three fundamental rules of probability:

1. If $\Pr(A)$ is the probability associated with event A, then $\Pr(A) \geqslant 0$.
2. The total probability of all outcomes in the sample space is 1.
3. If $A_1, A_2, \ldots$ is a sequence of **mutually exclusive** events, then we can add up their probabilities:

$$\Pr(\text{Any of } A_1, A_2, \ldots) = \Pr(A_1 \text{ or } A_2 \text{ or } \cdots) = \Pr(A_1) + \Pr(A_2) + \cdots$$

All other properties of probability can be mathematically derived from these axioms.

For example, the fact that $\Pr(A) \leqslant 1$.

Mutually exclusive events

- **Mutually exclusive** events are also known as **disjoint** events.
- These are events that cannot happen simultaneously (their intersection is empty).
- Example:
  - A = "My birthday is on a Monday"
  - B = "My birthday is on a weekend"
- Since A and B are mutually exclusive, the third axiom implies:

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$

- If $\Pr(A) = 1/7$ and $\Pr(A) = 2/7$, then $\Pr(A \text{ or } B) = 3/7$.

Non-disjoint events (Addition rule)

- Events that "overlap" (share outcomes in common) are called **non-disjoint** events.
- We need a more general rule for working out their probabilities.
- Need to remove the "overlap" to avoid "double counting".

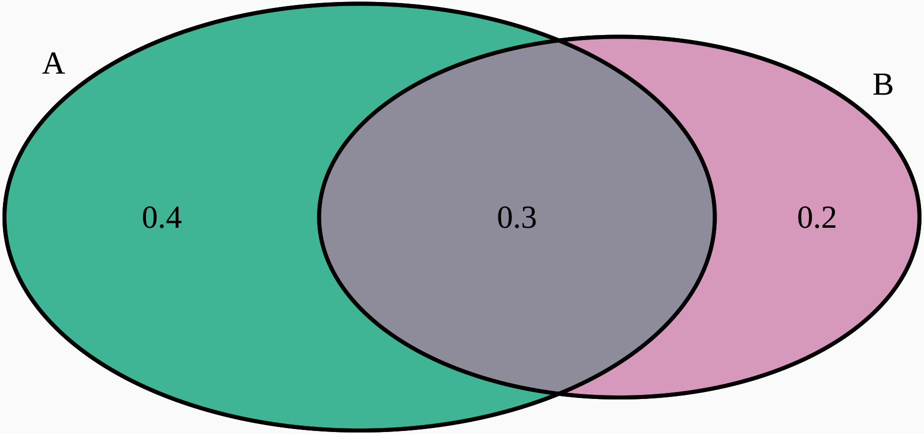
$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$

- This is known as the **addition rule**.
- Note: for disjoint events we have $\Pr(A \text{ and } B) = 0$.

Venn diagram

Suppose $\Pr(A) = 0.7$ and $\Pr(B) = 0.5$ and $\Pr(A \cap B) = 0.3$.

Then $\Pr(A \cup B) = 0.7 + 0.5 - 0.3 = 0.9$



Notation

The following are equivalent:

$$\Pr(A \text{ or } B) = \Pr(A \cup B)$$

The following are also equivalent:

$$\Pr(A \text{ and } B) = \Pr(A \cap B) = \Pr(A, B)$$

Examples

Example 1

- Let:
  - A = "My birthday is on a Friday or Saturday"
  - B = "My birthday is on a weekend"
- What is $\Pr(A \text{ or } B)$?

Example 2

- Let:
  - $A = (X > 2)$
  - $B = (X < 4)$
- What is $\Pr(A \text{ or } B)$?

Complementary events

- The **complement of event** A is the event denoted by A^c (and sometimes by $\bar{A}$).
- A^c represents **all outcomes not in** A
- Example: $A = \{H\}$ and $A^c = \{T\}$
- The probabilities for events A and A^c are related:

$$\begin{aligned} P(A) + P(A^c) &= 1 \\ P(A) &= 1 - P(A^c) \\ P(A^c) &= 1 - P(A) \end{aligned}$$

- Sometimes working with complementary events is easier.

Random variables

- A **random variable** (rv) is a quantity that can take one of many possible values (a "variable"), where each of those possibilities represents an outcome or event (with each having an associated probability).
- Typically we work with random variables that are **numerical** (integers or real numbers).
- The value of the random variable that actually occurs (e.g., is observed) is called a **realisation** of that random variable.
- Convention:
  - Random variables are denoted by uppercase letters: $X, Y, \dots$
  - Their realisations are in corresponding lowercase letters: $x, y, \dots$

Example

Flip a coin two times.
Let X = the **number of heads**.

Mapping of outcomes to values:

Outcome	X
HH	2
HT	1
TH	1
TT	0

Probability distribution of X:

x	$\Pr(X = x)$
2	0.25
1	0.5
0	0.25

Probability distributions

- A **probability distribution** describes the probabilities of events for a given random variable.
- Random variables can be **discrete** or **continuous**.
- For any random variable X, we can specify the **cumulative distribution function** (cdf): $F_X(x) = \Pr(X \leqslant x)$.
- For discrete variables, we can specify the probability for each possible value: $p_X(x) = \Pr(X = x)$. This is a **probability mass function** (pmf).
- For continuous variables, we can specify the **probability density function** (pdf): $f_X(x) = F_X'(x) = \frac{d}{dx} F_X(x)$.

Example: discrete random variable

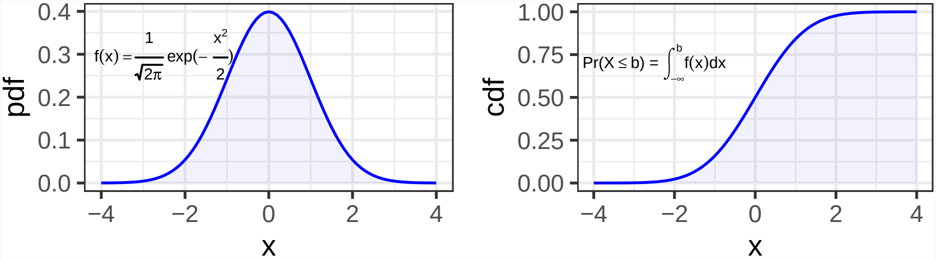
$X = x$	$\Pr(X = x)$
$X = 1$	$1/2$
$X = 2$	$1/8$
$X = 3$	$1/4$
$X = 4$	$1/8$

Find probabilities for given events:

1. $\Pr(X = 2)$
2. $\Pr(X \leqslant 2)$
3. $\Pr(X \text{ is even})$
4. $\Pr(X < 4)$
5. $\Pr(X > 2 \text{ and } X < 3)$
6. $\Pr(X > 2 \text{ or } X < 3)$

Example: continuous random variable

If $X \sim \mathrm{N}(0, 1)$



Find probabilities for given events:

1. $\Pr(X = 1)$
2. $\Pr(X < 1)$
3. $\Pr(X \text{ is even})$
4. $\Pr(X < -\frac{1}{2})$
5. $\Pr(X > 2 \text{ and } X < 3)$
6. $\Pr(X > 2 \text{ or } X < 3)$

Independence

Informal concept:

- Two random processes are **independent** if knowing the outcome of one provides no useful information about the outcome of the other.

Definitions:

- Two **events** are independent if the occurrence of one does not change the probability of occurrence of the other.
- Two **random variables** are independent if if the realisation of one does not change the probability distribution of the other.

Multiplication rule

- When we have independence, we have a convenient **multiplication rule** for calculating probabilities of intersections.
- A and B are independent events if and only if:

$$\Pr(A, B) = \Pr(A) \times \Pr(B)$$

- X and Y are independent random variables if and only if:

$$\Pr(X \leqslant x, Y \leqslant y) = \Pr(X \leqslant x) \Pr(Y \leqslant y)$$

- Can extend these definitions to more than two quantities.

Example: Left-handedness

- About 9% of people in the population are left-handed.
- Select 2 at random from the Australian population.
  - Assume population is so large that the observed handedness between the two people is independent.

1. What is the probability that both people selected are left-handed?

$$0.09 \times 0.09 = 0.0081 = 0.81\%$$

2. What is the probability that both people selected are right-handed?

$$(1 - 0.09)(1 - 0.09) = 0.91^2 = 0.8281 = 82.81\%$$

3. What is the probability that one person is left-handed and the other right-handed?

$$1 - (0.0081 + 0.8281) = 0.1638 = 16.38\%$$

Working with multiple variables

- Suppose we have more than one random variable.
- **Joint probability**
  - probability of outcomes for two or more variables.
- **Marginal probability**
  - probability of outcomes for a single variable.
- **Conditional probability**
  - probability of outcomes for a single variable *given information about a second variable.*

Example: Travel survey

Survey question: “Are you planning to travel abroad next year?”

		Age group			
		25 or less	26–40	41 or more	Total
Response	Yes	0.02	0.12	0.15	0.29
	Undecided	0.05	0.10	0.16	0.31
	No	0.10	0.15	0.15	0.40
	Total	0.17	0.37	0.46	1.0

Joint probabilities

Probability for all possible pairs

Age group / Response combination	Probability
Yes response AND (25 or less)	0.02
Yes response AND (26–40)	0.12
Yes response AND (41 or more)	0.15
Undecided response AND (25 or less)	0.05
Undecided response AND (26–40)	0.10
Undecided response AND (41 or more)	0.16
No response AND (25 or less)	0.10
No response AND (26–40)	0.15
No response AND (41 or more)	0.15
Total	1.0

Marginal probabilities

Are you planning to travel abroad next year?

Response	Probability
Yes	0.29
Undecided	0.31
No	0.40
Total	1.0

Age group

Age group	Probability
25 or less	0.17
26–40	0.37
1 or more	0.46
Total	1.0

Conditional probability

The **conditional probability** for an event A, given an event B, is defined as:

$$\Pr(A \mid B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}$$

Note that if A and B are independent, this simplifies to: $\Pr(A \mid B) = \Pr(A)$.

Examples

Example 1

Response	Pr(Response | 25 years or less)
Yes	0.02/0.17 = 0.1176
Undecided	0.05/0.17 = 0.2941
No	0.10/0.17 = 0.5882
Total	0.17/0.17 = 1.0

Example 2

Age group	Pr(Age group | Yes response)
25 years or less	0.02/0.29 = 0.0690
26–40 years	0.12/0.29 = 0.4138
40 years or more	0.15/0.29 = 0.5172
Total	0.29/0.29 = 1.0

Conditional probability distribution

A **conditional probability distribution** is a probability distribution where each probability is a conditional probability with the same condition.

Note that all of the conditional probabilities (given the same condition) will sum to 1.

(See previous slide for examples.)

General multiplication rule

Let A and B be two events. Then:

$$\begin{aligned}\Pr(A \text{ and } B) &= \Pr(A \mid B) \times \Pr(B) \\ &= \Pr(B \mid A) \times \Pr(A)\end{aligned}$$

If A and B are independent, then this simplifies to the earlier (simpler) multiplication rule.

Tree diagrams

Work out joint probabilities using the product of marginal and conditional probabilities.

Example: Travel survey (revisited)

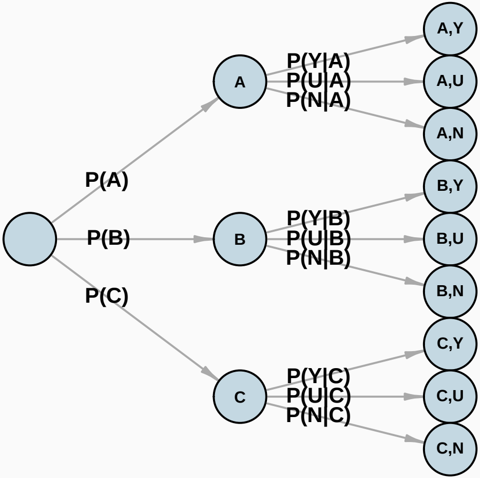
Let:

- A = 25 years or less
- B = 26–40 years
- C = 41 years or more

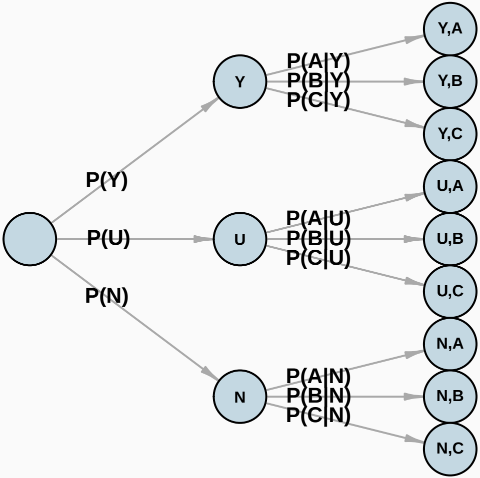
and let:

- Y = Yes response
- U = Undecided response
- N = No response

Tree diagram 1



Tree diagram 2



Bayes' theorem (inverting probabilities)

Suppose we know $\Pr(B \mid A)$ but would rather know $\Pr(A \mid B)$.

Use **Bayes' theorem**:

$$\Pr(A \mid B) = \frac{\Pr(B \text{ and } A)}{\Pr(B)}$$

Expected value (discrete)

- Let X be a discrete rv, taking values $x_1, \ldots, x_k$.
- The **mean** or **expected value** of X is:

$$\mathbb{E}(X) = x_1 \Pr(X = x_1) + \cdots + x_k \Pr(X = x_k) = \sum_{i=1}^{k} x_i \Pr(X = x_i)$$

- We often denote $\mu = \mathbb{E}(X)$
  - μ is an an attribute of the probability distribution for X
  - μ is not random

Variance (discrete)

- Let X be a discrete rv, taking values $x_1, \ldots, x_k$.
- The **variance** of X is:

$$\begin{aligned}\mathrm{var}(X) &= \mathbb{E}\left[(X-\mu)^2\right] = (x_1-\mu)^2 \Pr(X = x_1) + \cdots + (x_k-\mu)^2 \Pr(X = x_k) \\ &= \sum_{i=1}^{k} (x_i - \mu)^2 \Pr(X = x_i)\end{aligned}$$

- We often denote $\sigma^2 = \mathrm{var}(X)$
- The **standard deviation** of X is given by $\mathrm{sd}(X) = \sigma = \sqrt{\sigma^2}$.
  - σ^2 and σ are attributes of the probability distribution
  - σ^2 and σ are not random

Continuous random variables

Let X be a continuous random variable taking outcomes over the real line.

It has a probability density function (pdf) given by $f(x)$, such that:

1. $f(x) \geqslant 0$ for all $x \in \mathbb{R}$ (real numbers)
2. The probability of a given interval is given by

$$\Pr(L \leqslant X \leqslant R) = \int_L^R f(x)\,dx$$

3. The probability associated with all possible outcomes is given by

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Expected value and variance (continuous)

If X is a continuous random variable taking outcomes over the real line, having pdf $f(x)$, then the **expected value** and **variance** of X are given by:

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

and

$$\mathrm{var}(X) = \mathbb{E}\left[(X - \mu)^2\right] = \int_{-\infty}^{\infty} (x - \mu)^2\, f(x)\, dx$$

Outline

Discrete distributions

- Bernoulli
  - independent trials of a process with only two possible outcomes (Success / Failure)
- Binomial
  - Count the number of successes in n independent Bernoulli trials
- Negative Binomial
  - Count the number of failures until we reach a given number of successes

- Geometric
  - Simple version of negative binomial; number of failures until we get 1 success
- Uniform (discrete)
  - Whole numbers with equal probability of success
- Poisson

Bernoulli

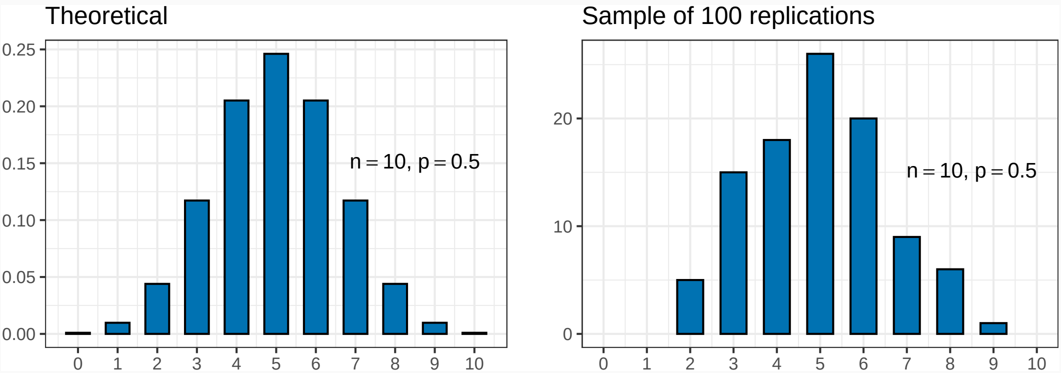
$$\Pr(X = x \mid p) = p^x(1-p)^{1-x} \quad \text{for } x \in \{0,1\}, \quad \text{given } 0 < p < 1$$

- We have seen this before
- $X = 1$ if a "heads" appears, $X = 0$ otherwise
- $X = 1$ if a person is promoted, $X = 0$ otherwise
- $\mathbb{E}[X] = p$ and $\mathrm{var}(X) = p(1-p)$

Binomial

$$\Pr(X = x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x \in \{0, 1, 2, \ldots, n\}$$

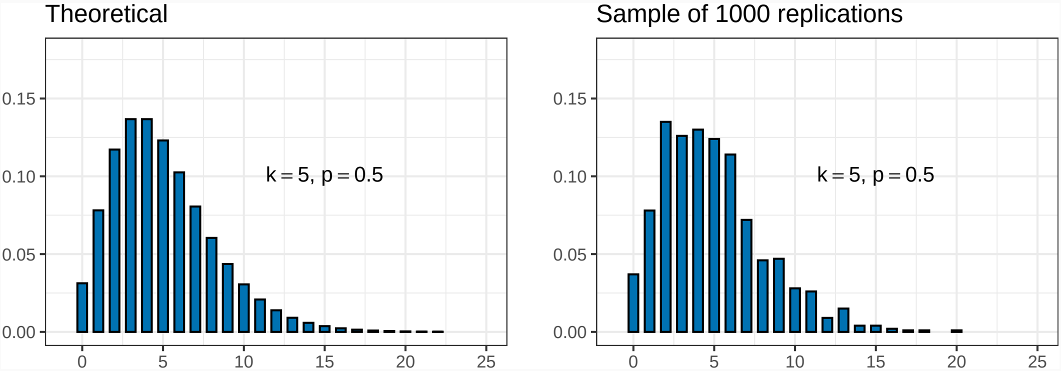
- Discrete, unimodal, right- or left-skewed or unimodal depending on p
- Arises from counting the number of successes from n independent Bernoulli trials, e.g. the number of heads in 10 coin flips
- $\mathbb{E}[X] = np$ and $\mathrm{var}(X) = np(1-p)$


Theoretical
n = 10, p = 0.5
0.25
0.20
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10
Sample of 100 replications
n = 10, p = 0.5
20
10
0
0 1 2 3 4 5 6 7 8 9 10

Negative Binomial

$$\Pr(X = x \mid k, p) = \frac{\Gamma(x+k)}{\Gamma(k)x!} p^k (1-p)^x, \quad \text{for } x \in \{0, 1, 2, \dots\}, \quad \text{given } 0 < p < 1$$

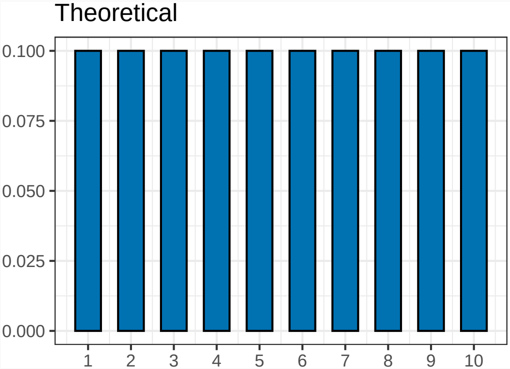
- Discrete, unimodal, right- or left-skewed or unimodal depending on p
- Arises from counting the number of 'failures' that occur in a sequence of independent Bernoulli trials until the targeted k^{th} success occurs
- $\mathbb{E}[X] = \frac{k(1-p)}{p}$ and $\mathrm{var}(X) = \frac{k(1-p)}{p^2}$
- Called the **geometric distribution** when $k = 1$.

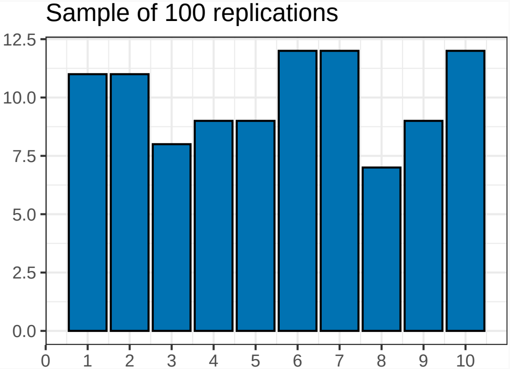

Theoretical
k = 5, p = 0.5
0.15
0.10
0.05
0.00
0
5
10
15
20
25
Sample of 1000 replications
k = 5, p = 0.5
0.15
0.10
0.05
0.00
0
5
10
15
20
25

Uniform (discrete)

$$\Pr(X = x \mid a, b) = \frac{1}{b - a + 1} \quad \text{for integers } a, b, \text{ with } x \in \{a, a+1, \ldots, b\}$$

- Discrete, symmetric, unimodal over values $\{a, a+1, \ldots, b\}$
- Arises from equally likely outcomes
- $\mathbb{E}[X] = \frac{b-a}{2}$ and $\mathrm{var}(X) = \frac{(b-a+1)^2 - 1}{12}$

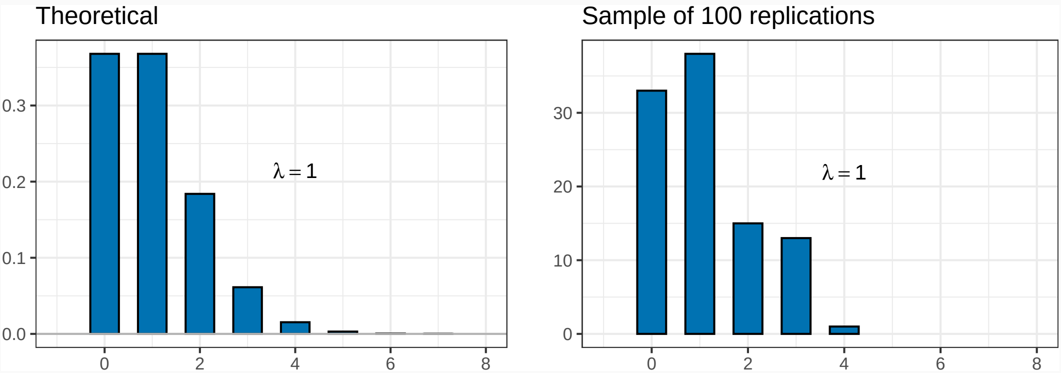

Theoretical
0.100
0.075
0.050
0.025
0.000
1 2 3 4 5 6 7 8 9 10


Sample of 100 replications
12.5
10.0
7.5
5.0
2.5
0.0
0 1 2 3 4 5 6 7 8 9 10

Poisson distribution

$$\Pr(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for } x \in \{0, 1, 2, \ldots\}, \quad \text{given } \lambda > 0$$

- Discrete, right-skewed, unimodal
- Arises when counting number of times event occurs in an interval of time, e.g. the number of patients arriving in an emergency room between 11 and 12 pm
- $\mathbb{E}[X] = \lambda$ and $\mathrm{var}(X) = \lambda$



Theoretical
0.3
0.2
0.1
0.0
λ = 1
0
2
4
6
8
Sample of 100 replications
30
20
10
0
λ = 1
0
2
4
6
8

Continuous distributions

- Uniform
- Normal
- Exponential
  - Time until first claim/accident/default
- Gamma
  - size of loan defaults, load on websites, time between accident claims
- Beta
  - models probabilities – generalisation of the uniform distribution
  - click through rate on a website
  - conversion rate of buyers

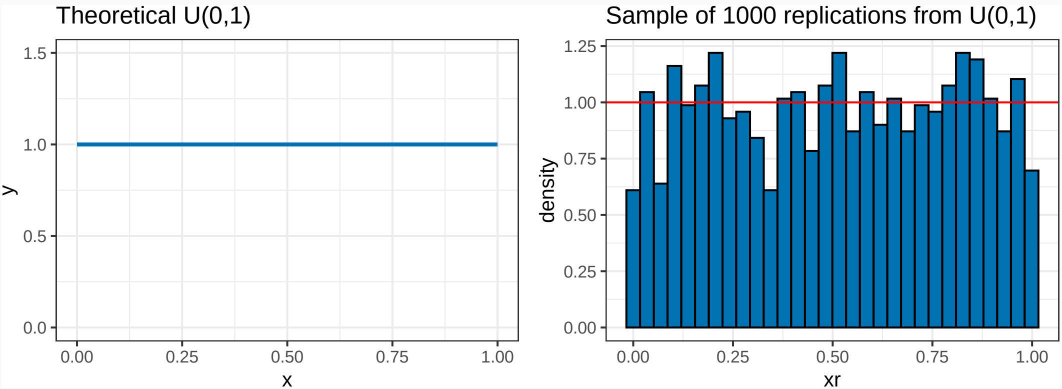
Continuous distributions

- Pareto
  - 80/20 rule
  - 20% of population controls 80% of the wealth
  - 80% of sales are from 20% of clients.
  - 80% of exam preparation is done in the last 20% of study time!
- Weibull
  - used in reliability analysis, life data analysis
  - time a user spends on a website
  - time until a system fails
- Lognormal
  - growth rates independent of size; financial data
  - Failure rates (like Weibull)

Uniform distribution (continuous)

$$f(x \mid a,\ b) = \frac{1}{(b - a)}, \quad \text{for } x \in (a, b) \text{ and } a < b$$

- continuous, symmetric, unimodal
- $\mathbb{E}[X] = \frac{a+b}{2}$ and $\mathrm{var}(X) = \frac{(b-a)^2}{12}$

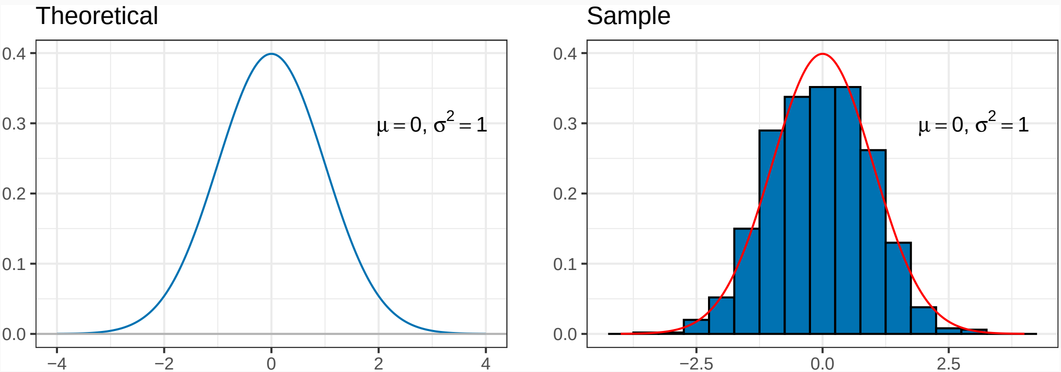


Theoretical U(0,1)
1.5
1.0
0.5
0.0
y
0.00
0.25
0.50
0.75
1.00
x
Sample of 1000 replications from U(0,1)
1.25
1.00
0.75
0.50
0.25
0.00
density
0.00
0.25
0.50
0.75
1.00
xr

Normal distribution $\mathrm{N}(\mu, \sigma^2)$

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$$

- Gaussian, bell-shaped
- symmetric, unimodal
- $\mathbb{E}[X] = \mu$ and $\mathrm{var}(X) = \sigma^2$

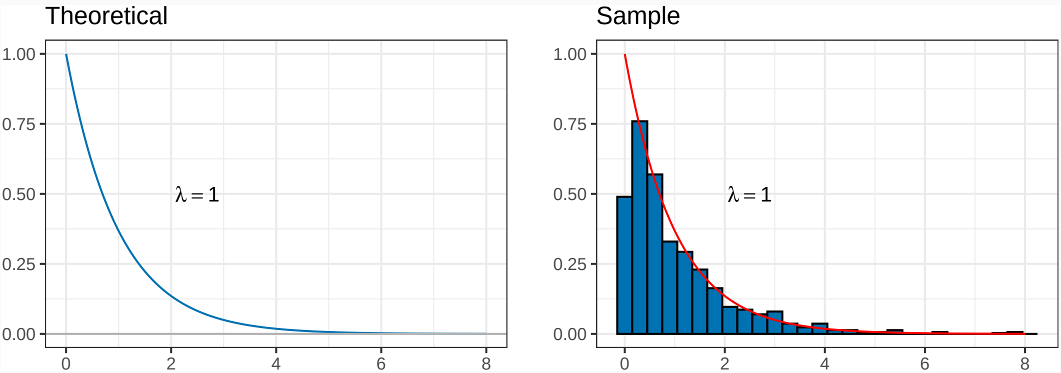


Theoretical
0.4
0.3
0.2
0.1
0.0
−4
−2
0
2
4
μ = 0, σ² = 1
Sample
0.4
0.3
0.2
0.1
0.0
−2.5
0.0
2.5
μ = 0, σ² = 1

Exponential distribution

$$f(x \mid \lambda) = \lambda e^{-\lambda x}, \quad x \geqslant 0, \lambda > 0$$

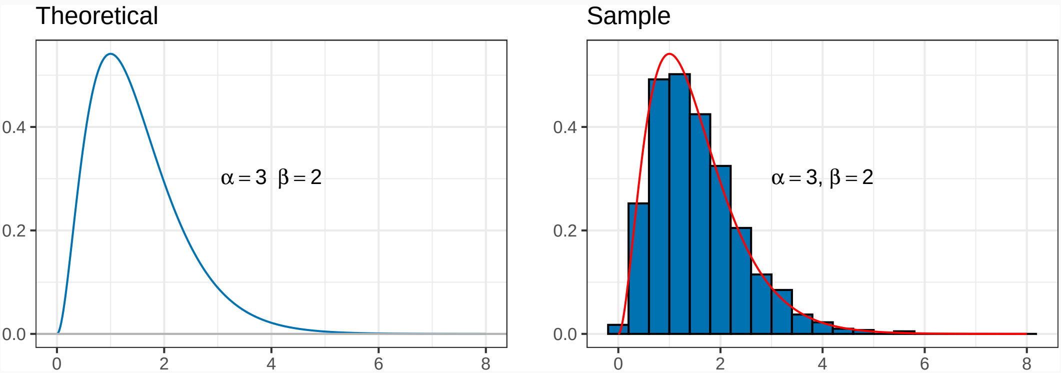
- right-skewed, unimodal
- Arises in time between or duration of events, e.g. time between successive failures of a machine, duration of a phone call to a help center
- λ is a **rate** parameter ($\beta = 1/\lambda$) is a **scale** parameter
- $\mathbb{E}[X] = \frac{1}{\lambda}$, $\mathrm{var}(X) = \frac{1}{\lambda^2} = \beta^2$


Theoretical
Sample
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
λ = 1
λ = 1

Gamma distribution

$$f(x \mid \alpha,\ \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta}, \quad x \geqslant 0, \alpha > 1, \beta > 0$$

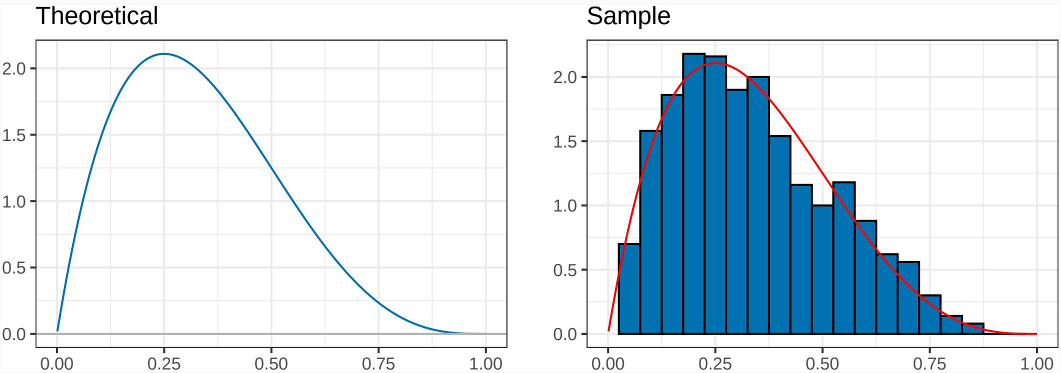
- right-skewed, unimodal
- α changes shape substantially
- β is a **rate** parameter ($b = 1/\beta$) is a **scale** parameter
- Special case is χ^2_v when $\alpha = \frac{v}{2}$ and $\beta = \frac{1}{2}$
- $\mathbb{E}[X] = \frac{\alpha}{\beta} = \alpha b$, $\mathrm{var}(X) = \frac{\alpha}{\beta^2} = \alpha b^2$


Theoretical
Sample
0.4
0.2
0.0
0
2
4
6
8
α = 3 β = 2
α = 3, β = 2

Beta

$$f(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1), \alpha > 0, \beta > 0$$

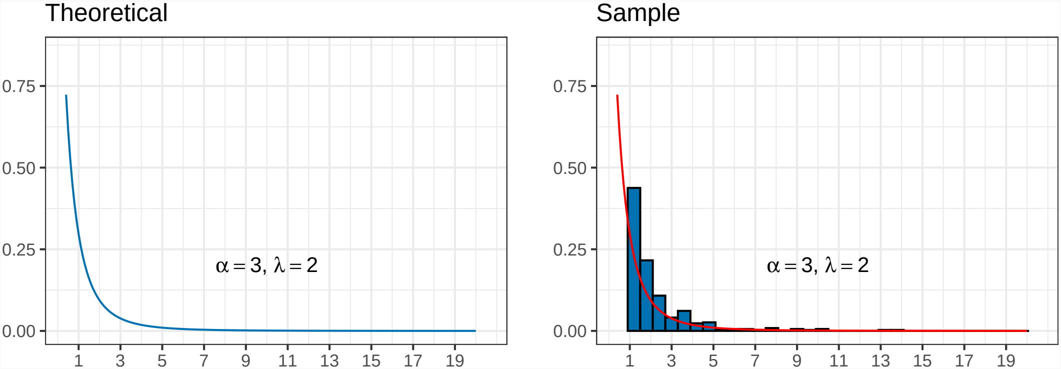
- Parameters $\alpha > 0$ and $\beta > 0$
- Generalisation of a continuous uniform on (0,1)
  - Same as a continuous uniform when $\alpha = \beta = 1$
- $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$ and $\mathrm{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$


Theoretical
Sample
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00

Pareto

$$f(x \mid \alpha, \lambda) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}, \quad x > 0, \alpha > 0, \lambda > 0$$

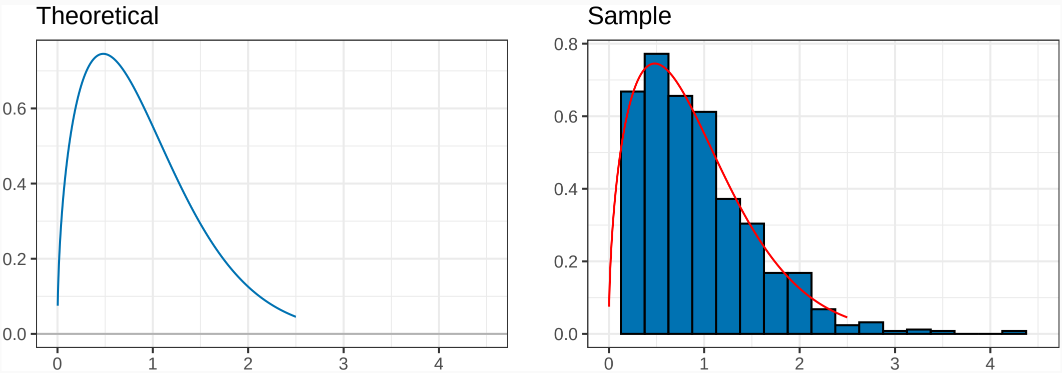
- Used to describe allocation of wealth, sizes of human settlement
- Heavier tailed than exponential distribution
- $\mathbb{E}[X] = \frac{\lambda}{\alpha-1}$, for $\alpha > 1$, and $\mathrm{var}(X) = \frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)}$, for $\alpha > 2$


Theoretical
Sample
0.75
0.50
0.25
0.00
1 3 5 7 9 11 13 15 17 19
α = 3, λ = 2
α = 3, λ = 2

Weibull

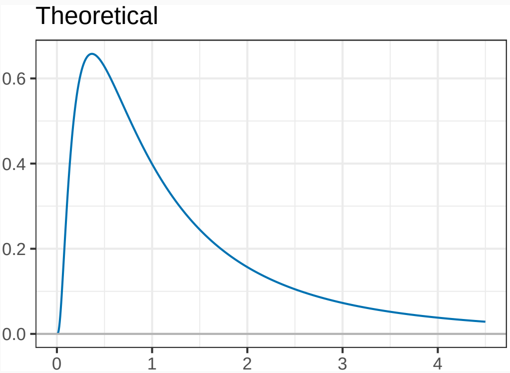
$$f(x \mid \lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad x > 0, \lambda > 0, k > 0$$

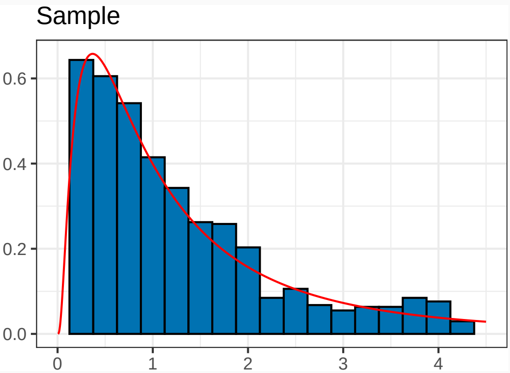
- used for particle size distribution, failure analysis, delivery time, extreme value theory
- shape changes considerably with different k
- $\mathbb{E}[X] = \lambda\Gamma\left(1 + \frac{1}{k}\right)$ and $\mathrm{var}(X) = \lambda^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2\right]$


Theoretical
Sample
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4

Lognormal

- Also called Galton's distribution
- Generated when $Y \sim \mathrm{N}(\mu, \sigma^2)$, and study $X = \exp(Y)$
- used for modeling length of comments posted in internet discussion forums, users' dwell time on the online articles, size of living tissue, highly communicable epidemics
- $\mathbb{E}[X] = \exp\{\mu + \frac{\sigma^2}{2}\}$ and $\mathrm{var}(X) = \exp\{2\mu + \sigma^2\}\left(\exp\{\sigma^2\} - 1\right)$


Theoretical
0.6
0.4
0.2
0.0
0
1
2
3
4


Sample
0.6
0.4
0.2
0.0
0
1
2
3
4

Outline

Linear combinations of random variables

- If X and Y are random variables, with mean values μ_X and μ_Y, respectively, and
- a and b are non-random constants
- then a linear combination of X and Y, denoted by Z, is given by

$$Z = aX + bY$$

- The **expected value** of Z is given by

$$\mathbb{E}[Z] = \mathbb{E}[aX + bY] = a\mu_X + b\mu_Y$$

Linear combinations of random variables

- If X and Y are random variables, with variances σ_X^2 and σ_Y^2, respectively, and
- a and b are non-random constants
- then a linear combination of X and Y, denoted by Z, is given by

$$Z = aX + bY$$

- The **variance** of Z is given by

$$\mathrm{var}(Z) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\mathrm{cov}(X, Y)$$

- The **covariance** between X and Y is given by

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X \mu_Y$$

- If X and Y are independent, then $\mathrm{cov}(X, Y) = 0$.
  - The converse is not true

Law of Large Numbers (LLN)

Let $X_1, \ldots, X_n$ be independent and identically distributed (**iid**) rvs with $\mathbb{E}(X) = \mu < \infty$.

Let,

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

With probability 1 we have:

$$\bar{X} \to \mu, \quad \text{as} \quad n \to \infty.$$

The LLN ‘guarantees’ that long-run averages behave as we expect them to:

$$\bar{X} \approx \mathbb{E}(X)$$

Central Limit Theorem (CLT)

Let $X_1, \ldots, X_n$ be iid rvs with $\mathbb{E}(X) = \mu < \infty$ and $\mathrm{var}(X) = \sigma^2 < \infty$.

Let,

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

Then

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

follows a $\mathrm{N}(0, 1)$ distribution as $n \to \infty$.

Outline

Birthday problem

- There are 25 students in the class.
- What is the probablity that at least two of them share the same birthday?

Envelope shuffling

- You have 20 letters and 20 envelopes.
- Each letter belongs to a specific envelope.
- Place the letters randomly in the envelopes (one letter per envelope).
- What is the expected number of letters in their correct envelope?

HTH vs HHT

- Keep flipping a fair coin until you see either the sequence HTH or HHT.
- Player A wins if you see HTH first.
- Player B wins if you see HHT first.
- What is the probability that player B wins?

Outline

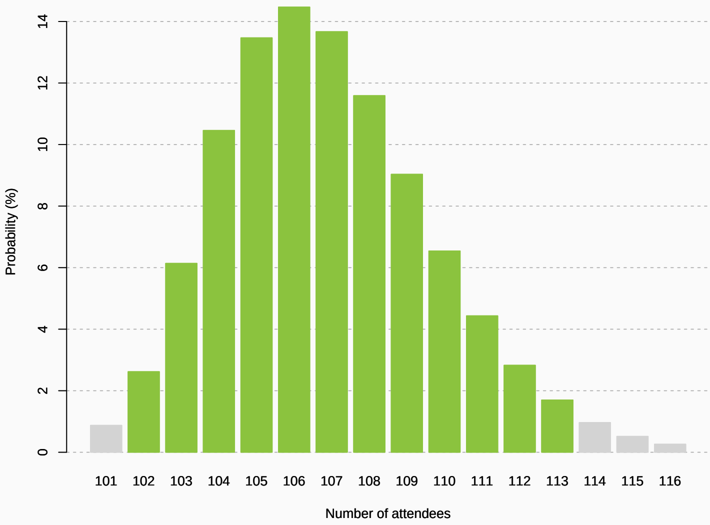
GUESTimation

Vukcevic D (2013), *Significance* 10(4):45–48.
https://doi.org/10.1111/j.1740-9713.2013.00687.x

GUESTimation

Group	Number per group	Probability	Expected attendance
Definitely	100	100 %	100
Likely	2	80 %	2
Maybe	4	50 %	2
Unlikely	33	10 %	3
—	—		—
Total	**139**		**107**

GUESTimation



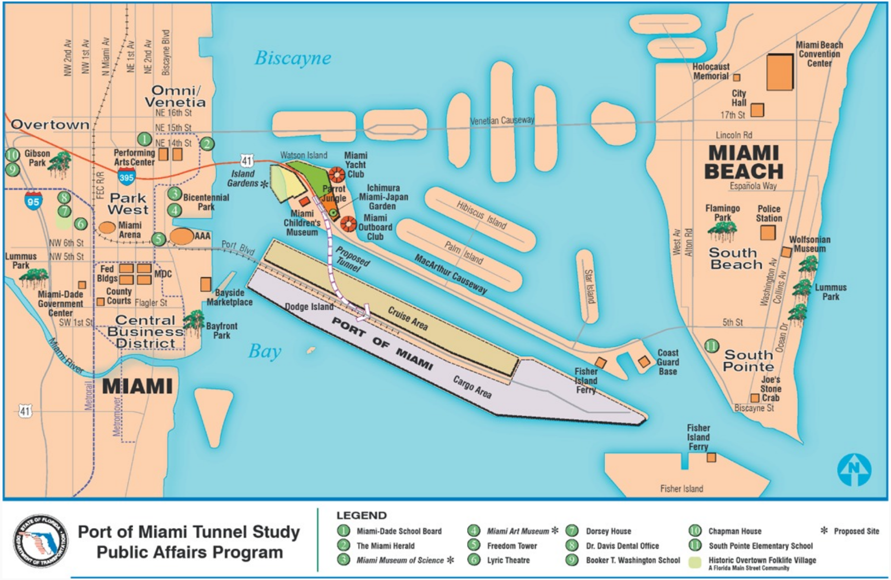
Will the big machine arrive?

Engineering the Miami Tunnel

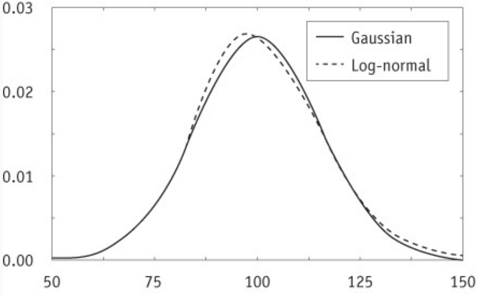
Palmer D (2011), *Significance* 8(4):148–153.
https://doi.org/10.1111/j.1740-9713.2011.00518.x

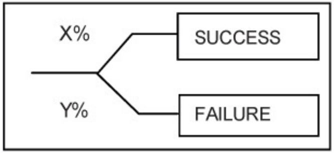
The Big Machine (TBM)

The Big Machine (TBM)

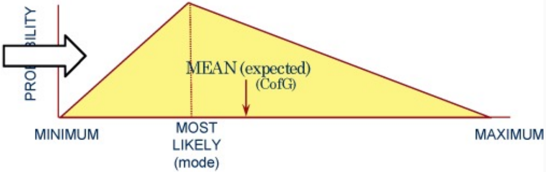


The Big Machine (TBM)



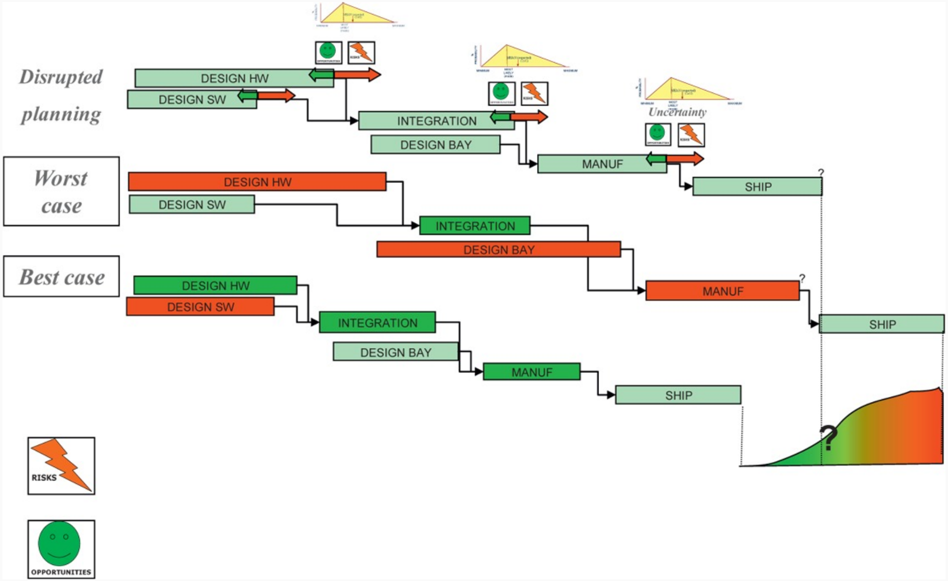


(a)

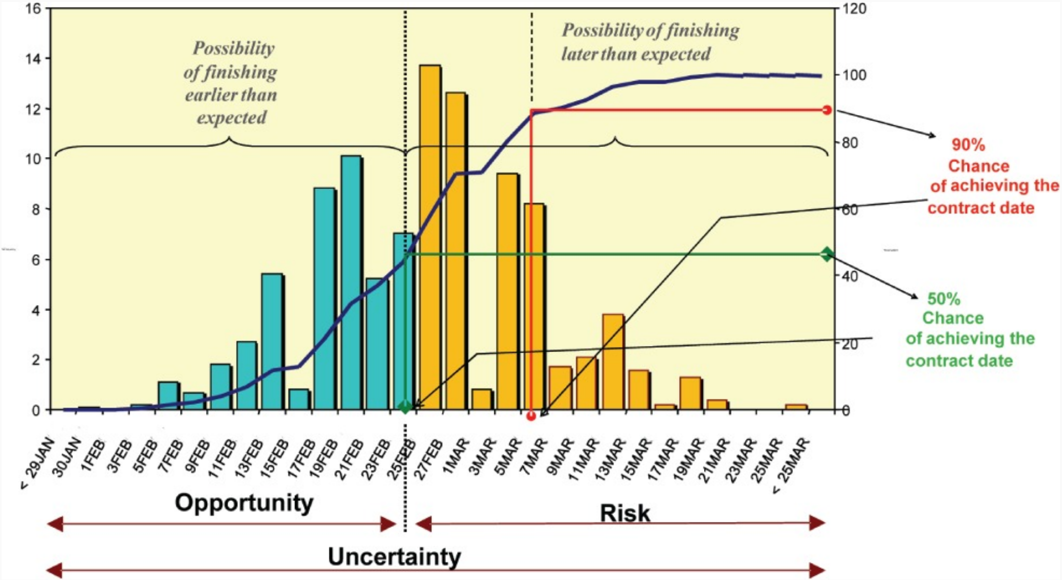


(b)

The Big Machine (TBM)



The Big Machine (TBM)



Airline overbookings

Outline

Types of data collection

Two main study types (data-generating mechanisms):

- **Observational study**. No ‘interference’ with the world.
- **Experiment**. Involves ‘manipulating’ the conditions in some way.

Sampling from a population

Compare two strategies:

- **Simple random sample** = all members of the population have an equal chance of being included in the sample.
- **Convenience sample** = members of the population that are more accessible are more likely to be included in the sample.

Sampling randomly is more likely to give a **representative** sample.

Non-response in surveys

Even with a random sample, some individuals selected to be included might choose to not respond or participate.

This can lead to **non-response bias**.

A common issue with modern surveys.

Does alcohol increase lung cancer?

- A study observes 1000 individuals
- Finding: higher alcohol intake $\Rightarrow$ greater lung cancer incidence
- Is this conclusive?

Confounding

- Actually, the underlying cause is likely to be *smoking*
- Higher smoking $\Rightarrow$ higher alcohol intake
- Higher smoking $\Rightarrow$ greater lung cancer incidence
- Once smoking is taken into account, negligible relationship between alcohol and lung cancer.
- Here, smoking is a **confounding variable**

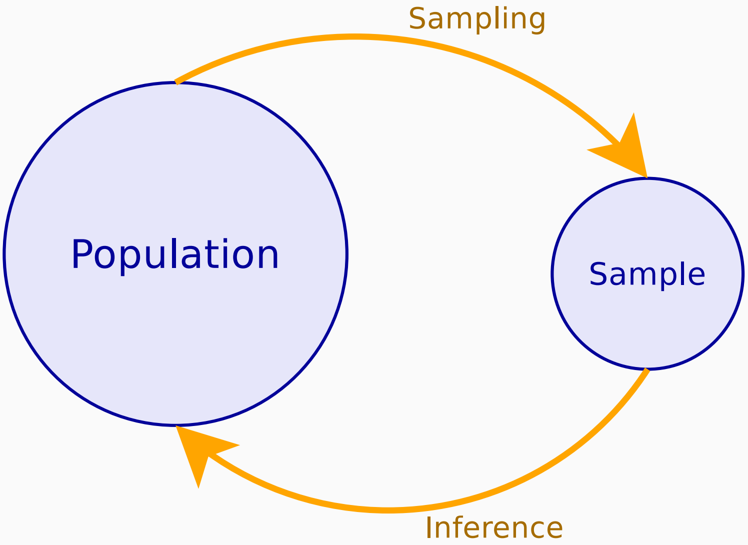
Randomly allocating treatment

- We can run experiments to eliminate confounding
- **Randomly allocate** the experimental treatment
- Also, allocate some experimental units to receive a ‘dummy’ treatment (**control**)
- This is a **randomised controlled trial** (RCT), the gold standard for quality data

Randomisation

- **Randomisation** is a key tool for statistics
  - Sampling: select random samples
  - Experiments: allocate treatments at random
- This is part of ‘thinking statistically’ even **before having any data**.

Reminder



Outline

How are we ‘thinking statistically’?

- Modelling variation
- Simulations to calculate answers to difficult problems
- Using randomisation to generate high-quality data