Sets

- De Morgan's Law
  - $(A \cup B)^C = A^C \cap B^C$
- Inclusion-Exclusion
  - $|A \cup B| = |A| + |B| - |A \cap B|$
- sample space
  - set of all possible outcomes that could happen
- event
  - event A: $A \subseteq S$

Combinatorics

- Combinations of possible outcomes
  - $|A| \cdot |B| \cdot |C| = 2 \cdot 2 \cdot 3 = 12$
- $n = |S| =$ number of elements in samplespace
- $k =$ number of picks

Ordered

Sampling with replacement

- number of possible ordered outcomes:
  - $|S| = n \cdot n \cdot \ldots \cdot n = n^k$
  - rolling dice

Sampling without replacement

- number of possible ordered outcomes:
  - $|S| = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)$
  - picking a card

Unordered / Permutations

- number of permutations
  - $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$
- Binomial coefficient
  - $\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}$
  - total number of unique (unordered) subsets that you could pick
  - $\binom{52}{5} = \frac{52!}{47!5!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{311,875,200}{120} = 2,598,960$
  - $P(A) = \frac{\text{Number of outcomes favourable to A}}{\text{Total number of outcomes in S}} = \frac{|A|}{|S|}$
  - given $A \subseteq S$
  - Rolling a die (getting at least 5):
    $|S| = 6$
    $A = \{x : x \in S, x \leq 5\} \to |A| = 2$
    $P(A) = \frac{|A|}{|S|} = \frac{2}{6} = 33\%$
- Probability function
  - Maps events to numbers between 0 and 1
- Probability space
  - Consists of sample space (S) and probability function (P)
- $\text{Odds} = \frac{P(A)}{P(A^C)}$
- $P(A^C) = 1 - P(A)$
- If $A \subseteq B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{|A| + |B| + |A \cap B|}{|S|}$

Conditional Probabilities

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  - "Probability of A, given B"
  - $P(A)$: Prior probability, $P(B)$: Posterior probability
- $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$
  - $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1, A_2)$
  - $P(A_1 \cap A_2 \cap A_3) = P(A) \cdot P(B) \cdot P(C)$
- $P(A \cap B|E) = P(A|E) \cdot P(B|E)$
- Bayes rule
  - $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$
- Law of Total Probability (LoTP)
  - $P(B) = P(B|A) \cdot P(A) + P(B|A^C) \cdot P(A^C)$
  - $P(B) = \Sigma_i^n = 1 P(B|A_i) \cdot P(A_i)$
- Independence
  - A and B are independent, if $P(A \cap B) = P(A) \cdot P(B)$
  - If A and B are independent $\Longrightarrow A$ and B^C, A^C and B, A^C and B^C are independent

Random Variables

- Definition
  - A random variable is a function mapping the sample space to the real line.
    $X(s)$: X for outcome s
  - A Indicator random variable indicates if something happens (either 1 or 0)
  - Discrete or Continuous
- Properties
  - Sample space S, random variable X, and a function $g: \mathbb{R} \to \mathbb{R}$
    - $Y = g(X)$ is the random variable that maps s to $g(X(S))$ for all $s \in S$
      - TLDR: X^2 is a random variable, if X is a random variable
    - $g(X, Y)$ can also be a random variable wich maps S to $g(X(s), Y(s))$
- Independence
  - X and Y are independent, if
    - $P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$
    - discrete case:
      $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$
      $, \forall\ x \in \text{supp}(X), y \in \text{supp}(Y)$
- Independence and Identically distributed
  - random variables are i.i.d., if...
    1. they are mutually independent and
    2. have the same probability distribution
- Conditional Independence
  - random variables X and Y are conditionally independent given a random variable $Z, \forall\ x, y \in \mathbb{R} \wedge z \in \text{supp}(Z)$, if
    $P(X \leq x, Y \leq y \mid Z = z) = P(X \leq x \mid Z = z) \cdot P(Y \leq y \mid Z = z)$

Discrete Random Variables

- $\text{supp}(X)$
  - distinct set of values x that X could give
  - $p_X(x \in \text{supp}(X)) > 0$
  - $p_X(x \notin \text{supp}(X)) = 0$

Random Indicator Variable

- $I \in 1, 0$
- Properties
  - $(I_A)^k = I_A, \forall k > 0$
  - $I_{Acomp} = 1 - I_A$
  - $I_{Acomp} = 1 - I_A$
  - $I_A \cap B = I_A \cdot I_B$
  - $I_A \cup I_B = I_A + I_B - I_A \cdot I_B$

Continuous Random Variable

- $P(X = x) = 0$ for any real number
  - Infinitly thin slice (width = 0)
  - If $P(X = x) > 0$ ==> sum would be infinit
- $\text{supp}(X) = \{f(x) > 0 \mid \forall x \in X\}$
- Non-negative: $f(x) \geq 0$
- Area under the curve is equal to 1

Probability Functions

Discrete

Probability Mass Function (PMF)

- $P(X = x) = p_X(x)$
  - Probability of the random variable X being a value of x
- Properties
  - $\sum_{x_j \in \text{supp}(X)} p_X(x_j) = 1$
  - $p_X(x) > 0$ when $x \in \text{supp}(X)$
  - $p_X(x) = 0$ otherwise $x \notin \text{supp}(X)$
  - $p_X(x) \geq 0\ \forall\ \{x \in \mathbb{R} | x \geq 0\}$

Cumulative Distribution Function (CDF)

- $F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_X(x_i)$
  - show if a random variable X will have a value equal or less than a specific value x
- Properties
  - Increasing
    - $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_1)$
  - "Right-continuous"
    - "jumps" happen from left to right
  - Convergence to 0 and 1 in the limits
    - Towards negative infinity for x, $F(x) \to 0$;
    - towards positive infinity for x, $F(x) \to 1$

Continuous

Probability Density Function (PDF)

- $f(x) = F'(x) = \frac{dF(x)}{dx}$

Cumulative Density Function (CDF)

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

$$P(a < X \leq b) = P(X \leq b) - P(X < a) = \int_a^b f(x)dx = F(b) - F(a)$$

Probability Distributions

Discrete Distributions

Bernoulli Distribution

- Situation (Bernoulli trial/experiment)
  - single event/experiment
  - can either result in "success" (1) or failure (0)
- $X \sim \text{Bern}(p)$
- PMF
  - $P(X = 1) = p$
  - $P(X = 0) = 1 - p$

Binominal Distribution

- Situation
  - repeated Bernoulli trials
  - sampling with replacement
  - $X \geq 0$: All trials could be "failures" ($X = 0$)
  - $X \leq n$: All trials (n) could be "successes" ($X = n$)
- $X \sim \text{Bin}(n, p)$
  - n = Number of trials
  - p = probability for success each trial
- PMF
  - $P(X = k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$
    - probability of getting k "successes" after repeating a Bernoulli trial with probability p for n times
- Expected Value
  - $E(X) = n \cdot p$
- Variance
  - $\text{Var}(X_{\text{Bin}(n,p)}) = n \cdot p \cdot (1-p)$

Hypergeometric Distribution

- Situation
  - repeated Bernoulli trials
  - sampling without replacement
- $X \sim \text{HGeom}(w, b, n)$
  - w = "successes"
  - b = "failures"
- PMF
  - $P(X = k) = \frac{\binom{w}{k} \cdot \binom{b}{n-k}}{\binom{w+b}{n}}$
    - w: successes
    - b: fails
    - n: number of picks
    - k: getting k number of successes
  - Expected Value
    - $E(X) = n \cdot \frac{w}{w+b}$
- Properties
  - For $w + b$ is large relative to n $\text{HGeom}(w, b, n)$ can be approximated with $\text{Bin}(n, \frac{w}{w+b})$

Geometric Distribution

- Situation
  - repeated Bernoulli trials
  - Number of *failures* before success
- $X \sim \text{Geom}(p)$
- PMF
  - $P(X = k) = (1-p)^k \cdot p$
    - $k = 0, 1, 2, \ldots$
- Expected Value
  - $E(X) = \frac{1-p}{p} = \frac{q}{p}$
- Variance
  - $\text{Var}(X_{\text{Geom}(p)}) = \frac{(1-p)}{p^2}$
- Properties
  - $X \sim \text{Geom}(p) \Longleftrightarrow X + 1 \sim \text{FS}(p)$
  - Memoryless
    - We are already waiting for 4 days and nothing happened ==> it doesn't affect the probability you will be waiting an additional day, also doesnt change the expected value
    - $P(X \leq t + s \mid X \leq s) = P(X \leq t)$

First Success / Shifted Geometric Distribution

- Situation
  - repeated Bernoulli trials
  - Number of *attempts* (failures + the final success) until a success
- $X \sim \text{FS}(p)$
- PMF
  - $P(Y = k) = (1-p)^{k-1} \cdot p$, for $k = 0, 1, 2, \ldots$
- Expected Value
  - $E(X) = E(X+1) = E(X) + 1 = \frac{1-p}{p} = \frac{1}{p}$
- Properties
  - $X \sim \text{FS}(p) \Longleftrightarrow X - 1 \sim \text{Geom}(p)$
  - Memoryless
    - We are already waiting for 4 days and nothing happened ==> it doesn't affect the probability you will be waiting an additional day, also doesnt change the expected value
    - $P(X \leq t + s \mid X \leq s) = P(X \leq t)$

Discrete Uniform Distribution

- Situation
  - picking an element of a set C, where all elements have the same probability of getting picked. X is the chosen element
- $X \sim \text{DUnif}(C)$
  - C: finite, non-empty set of numbers
- PMF
  - $P(X = x) = \frac{1}{|C|}$
    - ... for $x \in C$
  - For $x \notin C : P(X = x) = 0$

Poissen Distribution

- Situation
  - Counting number of successes in a particular **region** or **interval of time**
  - Large number of "sort-of-blurry" trials and small probabilities of success
  - no/difficult to percieve discrete trials
  - popular for counting data
- $X \sim \text{Pois}(\lambda)$
  - λ: Average number of events per interval. Could indeed be <1
- PMF
  - $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$
    - $k = 0, 1, 2, \ldots$
- Expected Value
  - $E(X) = \lambda$
- Variance
  - $Var(X) = \lambda$
- Properties
  - For $X \sim \text{Pois}(\lambda_1)$ and $Y \sim \text{Pois}(\lambda_2) \Longrightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
  - If $X \sim \text{Pois}(\lambda)$ on 1 unit scale (time interval), then $Y\,\text{Pois}(c \cdot \lambda)$ on c unit interval (scalable)
    - C *Pois*(\lambda) =/= *Pois*(*C*\lambda)

Continuous

Continuous Uniform Distribution

- $X \sim \text{Unif}(a, b)$
- PDF
  - $f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$
- CDF
  - $F(x) = \begin{cases} 0, & \text{if } x \leq a \\ \frac{x-a}{b-a}, & \text{if } a < x < b \\ 1, & \text{if } x \geq b \end{cases}$
- Expected Value
  - $E(X) = \frac{a+b}{2}$
- Variance
  - $Var(X) = \frac{(b-a)^2}{12}$
- Properties
  - Transformations
    - $Y = \sigma \cdot X + \mu$
    - σ, μ being constants ($\sigma > 0$)
    - μ: controls location (up down the number line)
    - σ: controls scale (squeezed, elongated)

Normal Distribution

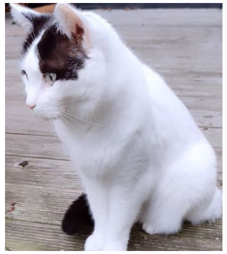
- $X \sim N(\mu, \sigma^2)$
- Expected Value
  - $E(X) = \mu$
- Variance
  - $Var(X) = \sigma^2$
- PDF
  - $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}, \quad -\infty < z < \infty$
- CDF
  - $\Phi(z) = \int_{-\infty}^{z} \phi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$
- Properties
  - Symetry: $\phi(z) = \phi(-z)$
  - Symetry of tail areas: $P(Z < -z) = \Phi(-z) = 1 - \Phi(z) = P(Z > z), \forall z$
  - $\frac{X - \mu}{\sigma} \sim N(0, 1)$
  - $X = Z \cdot \sigma + \mu$
  - Standard normal distribution: N(0,1)

Exponential distribution

- Situation
  - Number of failures before first success (First success: number of trials until first success)
  - "waiting time": number of discrete failres that happens before the success
  - continuous counterpart to the geometric
- $X \sim \text{Expo}(\lambda)$
  - λ: rate of success (per unit of time), average number of successes in a time interval of length t is λ
- PDF
  - $f(x) = \lambda e^{-\lambda x}, x > 0$
  - dexp(x, rate=lambda)
- CDF
  - $F(x) = 1 - e^{-\lambda x}, x > 0$
  - dexp(x, rate=lambda)
- Expected Value
  - $E(X) = \frac{1}{\lambda}$
- Variance
  - $Var(X) = \frac{1}{\lambda^2}$
- Properties
  - Scaling
    - $X \sim \text{Expo}(1), Y = \frac{X}{\lambda} \implies Y \sim \text{Expo}(\lambda)$
    - $Y \sim \text{Expo}(\lambda) \implies \lambda \cdot Y \sim \text{Expo}(1)$
  - Memoryless
    - We are already waiting for 4 days and nothing happened ==> it doesn't affect the probability you will be waiting an additional day, also doesnt change the expected value
    - $P(X \leq t + s \mid X \leq s) = P(X \leq t)$

Student t distribution

- Situation
  - Same situation as normal distribution but for samplecounts of n<30
  Degrees of freedom
- $\text{df} = \nu = n - 1$
- Expected value
  - $E(T) = 0$
- Variance
  - $Var(T) = \frac{\nu}{\nu - 2}$
- PDF: `dt(x, df = df)`
- CDF: `pt(x, df=df)`
- CDF^-1 : `qt(p, df=df)`
- Properties
  - for $\nu \to \infty$ the t distribution $\to N(0, 1)$



- Law of large numbers
  - As sample size n grows, the sample mean $\bar{X}_n$ converges to the true population mean μ.
- Central limit theorem
  - As sample size n increases, the distribution of the sample mean $\bar{X}_n$ approaches a normal distribution. For standardized X a standard normal distribution is approached

The Fundamental Bridge

- $E(I_A) = P(A)$
  - $I_A \sim \text{Bern}(p)$ and $p = P(A)$
- $E(X_i) = P(X_i = 1)$
  - because $E(X_i) = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1)$

Law of the unconscious statistician

- Discrete:
  - $E(g(X)) = \sum_{x \in \text{supp}(X)} g(x) \cdot P(X = x)$
    $\Longrightarrow E(X^3) = \sum_{x \in \text{supp}(X)} x^3 \cdot P(X = x)$
- Continuous:
  - $E(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx$

Confidence Interval

$$P(\hat{\theta}_l \le \theta \le \hat{\theta}_h) \le 1-\alpha$$

"Probability that the (unknown) population mean is somewhere in the interval $\hat{\mu}_l - \hat{\mu}_h$ is 0.95"

CI for Normaldistribution

$$CI = \bar{X} \pm Z \cdot \frac{s}{\sqrt{n}}$$

z-tables or qnorm()

CI for Student t distribution

Degrees of freedom:

$$\text{df} = \nu = n-1$$

$$CI = \bar{X} \pm t_{\frac{\alpha}{2},\nu} \cdot \frac{s}{\sqrt{n}}$$

CI for Proportions

If $n\hat{p} > 10$ and $n(1-\hat{p}) > 10$:

$$CI = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Hypothesis testing

two sided test:

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

- Errors
  - Type 1
    - reject H_0 when H_0 is true
    - P(t1 error) >= alpha
  - Type 2
    - support H_0 when H_1 is true
- Teststatistic
  - $T = \frac{\bar{X}-\mu_0}{\frac{s}{\sqrt{n}}}$
  - $H_0 = \text{true if } |T| \le t_{\frac{\alpha}{2};n-1}$
- p-value
  - $p = 2 \cdot \min(P(T \le t), P(T \ge t))$

Proportions

- check
  - $n \cdot p_0 > 10$
  - $n \cdot (1-p_0) > 10$
- Teststatistic
  - $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$
- p-value
  - $P = 2 \cdot P(Z < \text{textstatistic Z})$

Expectation

- measures center of distribution
- Discrete Case
  - $E(X) = \sum_{x \in supp(X)} x \cdot P(X=x)$
- Properties
  - Equivalence of expectations
    - If $P(X=x) = P(Y=y)$ then $E(X) = E(Y)$
  - Linearity of expectations
    - $E(X+Y) = E(X) + E(Y)$
    - $E(X-Y) = E(X) - E(Y)$
    - $E(c \cdot X) = c \cdot E(X)$, c: constant
  - Monotonicity of expectations
    - $X \ge Y \Longrightarrow E(X) \ge E(Y)$
- Continuous Random Variables:
  - $E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$
  - $P(E(X)) = 0$

Variance

- $\text{Var}(X) = E((X-EX)^2) = E(X^2) - (EX)^2$
- $\text{SD}(X) = \sqrt{\text{Var}(X)}$
- Properties
  - $\text{Var}(X+c) = \text{Var}(X)$
  - $\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X)$
  - $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
  - $\text{Var}(X) \ge 0$, only equal to 0 if X is constant

Inferential statistics

- measures of location
  - mean
  - median
    - 1. order values, 2. get the middle one (if 2 middlevalues then take mean)
    - 50% | median | 50%
    - robust measure: outliers have less impact
  - mode
    - most common value
- standard error of sample mean
  - $\sqrt{Var(\bar{X}_n)} = \sqrt{\frac{s^2}{n}} = \frac{\sigma}{\sqrt{n}}$
  - $\lim_{n\to\infty} \frac{\sigma}{\sqrt{n}} \longrightarrow 0, \frac{\sigma}{\sqrt{n}} > 0$
- Sample Variance
  - $S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$

Location scale transformation:

Distribution	Transformation
Uniform	$Y \sim Unif(c*a+d, c*b+d)$ where c>0
Normal	$Z \sim \frac{X-\mu}{\sigma}$

Conditional probabilities of continuous r.v.s.:

$P(X \ge s+t \mid X \ge s) = \frac{P(X \ge s+t)}{P(X \ge s)} =$

$if\ memoryless = P(X \ge t)$

Sample statistics/Point estimates:

Statistic	Calculation
Sample mean	$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$
Sample median	c where $P(X \le c) \ge \frac{1}{2}$ and $P(X \ge c) \ge \frac{1}{2}$
Sample mode	$P(X=c) \ge (P\ X = x), for\ x \in supp(X)$ $f(c) \ge f(x), for\ all\ x$
Sample variance	$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$
Standard error	$\sqrt{Var(\bar{X}_n)} = \frac{\sigma}{\sqrt{n}}$ *The variance of the overall sample mean is the variance of the population/n*

Law of Large Numbers: As sample size n grows, the sample mean $\bar{X}_n$ converges to the true (population) mean μ!

***Central Limit Theorem:** For large n (→∞; with replacement)/repeated sampling, the distribution of $\bar{X}_n$ /sample means approaches a (standard) Normal distribution (when X is standardized)

Symbol	Name	Explanation
$\in$ or $\notin$	Membership	Is Set/… in another Set/…?
$\subseteq$	Subset of	Is Event/… a subset of Set/…?
A^C	Complement	Everything in *U* without *A*
$A \setminus B$	Set difference	Elements from *A* minus Elements from *B*
$A \cup B$	Union	Elements in *A* or *B*
$A \cap B$	Intersection	Elements in *A* and *B*; if $A \cap B = \emptyset$ they are disjoint

Rules for events and their probabilities:

$$P(A^C) = 1 - P(A)$$

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

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

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

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

$$P(A \mid B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|A^C) * P(A^C)}$$

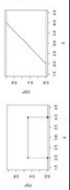
Independence of two events:

$Independent\ if\ P(A \cap B) = P(A) * P(B)$

Algorithms for sampling and combinatorics:

Algorithm	Use case
n^k	Sampling with replacement / order matters
$n*(n-1)*\dots*(n-k+1)$	Sampling without replacement / order matters
$\frac{n!}{(n-k)!*k!} = \binom{n}{k}$	Sampling without replacement / order doesn't matter
$\frac{n+k-1}{k}$	Sampling with replacement / order doesn't matter

Distributions of continuous random variables:

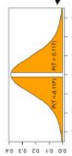


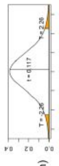
Name	Short	CDF	Use case	E(X) and Var(X)
Uniform	$X \sim Unif(a,b)$	$F(x) = \frac{x-a}{b-a}$ If a < x < b	a = min; b = max (of interval) Success happens in the given interval (b-a)	$E(X) = \frac{a+b}{2}$ $Var(X) = \frac{(b-a)^2}{12}$
(Standard) Normal	$X \sim N(\mu,\sigma^2)$ $Z \sim N(0,1)$	$\phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$	μ = mean, σ = sd z-transformation to use ϕ; R: pnorm()	$E(X) = \mu$ $Var(X) = \sigma^2$
Exponential	$X \sim Expo(\lambda)$	$F(x) = 1 - e^{-\lambda x}$, x > 0	λ = rate of success Waiting time until success; helpful to think of E() when determining λ; R: pexp()	$E(X) = \frac{1}{\lambda}$ $Var(X) = \frac{1}{\lambda^2}$

Confidence intervals:

Formula	Context
$CI_\mu = \left(\bar{X} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right)$ or $\left(\bar{X} \pm z_{\alpha/2} * \frac{s}{\sqrt{n}}\right)$ (when σ unknown and n > 30°)	For normal distributed population parameters θ α = confidence level; z = z-value (R: qnorm(p)) σ = population sd; s = sample sd; n = sample size
$CI_\mu = \left(\bar{X} \pm t_{\frac{\alpha}{2}n-1} * \frac{s}{\sqrt{n}}\right)$ (when σ unknown and n < 30)	For t-distributed θ, i.e., for small n's and unknown σ α = confidence level; t = t-value (R: qt(p)) s = sample sd; n = sample size
$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	For sample proportion of success $\hat{p}$: Estimate for p in an indicator r.v. when we don't know the true p α = confidence level; z = z-value; n = sample size $\hat{p}$ = estimate for success probability

Two-sided hypothesis testing (for the mean):





For comparing means assumption is always: $H_0: \mu = \mu_0,\ H_1: \mu \neq \mu_0$

Test statistic T is: $T = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ (we divide variation explained by the model ($\bar{X} - \mu_0$) by unsystematic variance)

We accept H_0 if: $|T| \le t_{\frac{\alpha}{2}n-1}$ (acceptance of H_0 when T is in the confidence interval of the t-distribution)

P-value for T: $p = 2 * \min(P(T \le t), P(T \ge t))$ (we're basically interested in the probability to get the test statistic t or something greater than t by chance)

If p is small (typically $p < 0.05$) it is very unlikely that effect occurs by chance. Thus, data contradicts H_0.

(T follows itself a t-distribution, so we put the value of T in the CDF of the t-distribution to get the probability for its occurrence (R: pt). We use min() to get one of the tails. We multiply with 2, because test is two-sided)

Test statistic for the proportion:

Test statistic Z is: $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ (when $np_0 > 10$ and $n(1-p) > 10$) (For calculating the p-value, we use $\sim N(0,1)$ (R: pnorm))

Name	Short	PMF	Use case	E(X) and Var(X)
Bernoulli	$X \sim Bern(p)$	$P(X=1) = p$ $P(X=0) = 1-p$	Indicator variable can be 0 (failure) or 1 (success)	$E(X) = p$ $Var(X) = p*(1-p)$
Binomial	$X \sim Bin(n,p)$	$P(X=k) = \binom{n}{k} * p^k * (1-p)^{n-k}$	n = number of trials; k = successes; p = success probability Helpful where identifiable Bernoulli trials exist. Success probability doesn't change through experiment. R: dbinom(k, n, p).	$E(X) = n*p$ $Var(X) = n*p*(1-p)$
Hyper-geometric	$X \sim HGeom(w,b,n)$	$P(X=k) = \frac{\binom{w}{k}*\binom{b}{n-k}}{\binom{w+b}{n}}$	w = white balls; b = black balls; k = successes; n = number of picks Helpful when probabilities change from pick to pick – so when "sampling" without replacement. Use this when having information on groups of different sizes. R: dhyper(k, w, b, n)	$E(X) = n * \frac{w}{w+b}$
Uniform	$X \sim DUnif(C)$	$P(X=x) = 1/\|C\|$	C = sum of all x All outcomes of X are equally likely.	-
Geometric	$X \sim Geom(p)$	$P(X=k) = (1-p)^k * p$	k = number of failures; p = success probability Helpful whenever probability of number of failures is searched for. R: dgeom(k, p) $X \sim Geom = X + 1 \sim FS(p)$	$E(X) = \frac{q}{p}$ $Var(X) = (1-p)/p^2$
First-success	$X \sim FS(p)$	$P(X=k) = (1-p)^{k-1} * p$	k = point at which first success occurs p = success probability	$E(X) = \frac{1}{p}$
Poisson	$X \sim Pois(\lambda)$	$P(X=k) = \frac{e^{-\lambda} * \lambda^k}{k!}$	λ = average/expected number of events per interval; k = successes Helpful to count number of successes when number of trials is not specified. E.g., for certain time periods. R: dpois(k, λ)	$E(X) = Var(X) = \lambda$

Formula of the PMF of discrete r.v.s.:

$$\sum_{x_j \in supp(X)} p_X(x_j) = 1$$

Basics on Expectation:

$$E(X) = \sum_{x \in supp(X)} x * P(X = x)$$

$E(X+Y) = E(X) + E(Y)$ *(linearity)*

$E(c*X) = c*E(X)$ *(linearity)*

Basics on Variance:

$$Var(X) = E(X^2) - (EX)^2$$

$SD(X) = \sqrt{Var(X)}$

$Var(X+c) = Var(X)$

$Var(c*X) = c^2 * Var(X)$

$Var(X+Y) = Var(X) + Var(Y)$ *(only if independent)*

Law of the unconscious statistician (LOTUS):

$E(g(X)) = \sum_{x \in supp(X)} g(x) * P(X=x)$