

01. PROBABILITY

- probability** of an event \rightarrow the limiting relative frequency of its occurrence as the experiment is repeated many times
- the **realisation** x is a constant, and X is a generator
 - running r experiments gives us r realisations x_1, \dots, x_r

Expectation

discrete: (mass function)	continuous: (density function)
$E(X) := \sum_{i=1}^n x_i p_i$	$E(X) := \int_{-\infty}^{\infty} x f(x) dx$

expectation of a function $h(X)$

$$E\{h(X)\} = \begin{cases} \sum_{i=1}^n h(x_i) p_i & X \text{ is discrete} \\ \int_{-\infty}^{\infty} h(x) f(x) dx & X \text{ is continuous} \end{cases}$$

Variance

$$\begin{aligned} \text{variance, } \text{var}(X) &:= E\{(X - \mu)^2\} \\ \text{standard deviation, } SD(X) &:= \sqrt{\text{var}(X)} \\ \text{var}(X) &= E(X^2) - E(X)^2 \\ E(X - \mu) &= 0 \end{aligned}$$

Law of Large Numbers

mean and variance of r realisations:

$$\bar{x} := \frac{1}{r} \sum_{i=1}^r x_i \quad v := \frac{1}{r} \sum_{i=1}^r (x_i - \bar{x})^2$$

LLN: for a function h , as $r \rightarrow \infty$,

$$\begin{aligned} \frac{1}{r} \sum_{i=1}^r h(x_i) &\rightarrow E\{h(X)\} \\ \bar{x} \rightarrow E(X), \quad v &\rightarrow \text{var}(X) \end{aligned}$$

Monte Carlo approximation

simulate x_1, \dots, x_r from X . by LLN, as $r \rightarrow \infty$, the approximation becomes exact

$$E\{h(X)\} \approx \frac{1}{r} \sum_{i=1}^r h(x_i)$$

Joint Distribution

(discrete) mass function:

$$P(X = x_i, Y = y_j) = p_{ij}$$

(continuous) density function:

$$f : \mathbb{R}^2 \rightarrow [0, \infty), \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

(expectation) for $h : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$\begin{cases} \sum_{i=1}^I \sum_{j=1}^J h(x_i, y_j) p_{ij} & X \text{ is discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy & Y \text{ is continuous} \end{cases}$$

Algebra of RV's

let X, Y be RVs and a, b, c be constants

- $Z = aX + bY + c$ is also an RV
 - $z = ax + by + c$ is a realisation of Z
- linearity of expectation: $E(Z) = aE(X) + bE(Y) + c$
- any theorem about a RV is true about a constant

Covariance

let $\mu_X = E(X)$, $\mu_Y = E(Y)$.

$$\begin{aligned} \text{covariance, } \text{cov}(X, Y) &= E\{(X - \mu_X)(Y - \mu_Y)\} \\ \text{cov}(X, Y) &= E(XY) - \mu_X \mu_Y \\ \text{cov}(X, Y) &= \text{cov}(Y, X) \\ \text{cov}(X, X) &= \text{var}(X) \\ \text{cov}(W, aX + bY + c) &= a \text{cov}(W, X) + b \text{cov}(W, Y) \\ \text{var}(aX + bY + c) &= a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab \text{cov}(X, Y) \\ \text{var}(\sum_{i=1}^N a_i X_i) &= \sum_{i=1}^N a_i^2 \text{var}(X_i) + 2 \sum_{1 \leq i < j \leq N} a_i a_j \text{cov}(X_i, X_j) \end{aligned}$$

joint = marginal \times conditional distributions

$$\begin{aligned} f(x, y) &= f_X(x) f_Y(y|x) \\ &= f_Y(y) f_X(x|y), \quad x, y \in \mathbb{R} \end{aligned}$$

- $f(x, y)$ is the *joint density*
- $f_X(x)$, $f_Y(y)$ are the *marginal densities*
- $f_Y(\cdot|x)$ is the **conditional** density of Y given $X = x$
- $f_X(\cdot|y)$ is the **conditional** density of X given $Y = y$
- for discrete case, *density* \equiv *probability*, $x \equiv x_i$, $y \equiv y_j$

Independence

- X, Y are independent $\iff \forall x, y \in \mathbb{R}$,
 - $f(x, y) = f_X(x) f_Y(y)$
 - $f_Y(y|x) = f_Y(y)$
 - $f_X(x|y) = f_X(x)$
- X, Y are independent \Rightarrow
 - $E(XY) = E(X)E(Y)$
 - $\text{cov}(X, Y) = 0$
 (the converse does not hold)

Conditional expectation

discrete case

let $f_Y(\cdot|x_i)$ be the conditional pmf of Y given $X = x_i$.

$$\begin{aligned} E[Y|x_i] &:= \sum_{j=1}^J y_j f_Y(y_j|x_i) \\ \text{var}[Y|x_i] &:= \sum_{j=1}^J (y_j - E[Y|x_i])^2 f_Y(y_j|x_i) \end{aligned}$$

$E[Y|x_i]$ is like $E(Y)$, with conditional distribution replacing marginal distribution $f_Y(\cdot)$. likewise, $\text{var}[Y|x_i]$ like $\text{var}(Y)$.

continuous case

$$\begin{aligned} E[Y|x] &:= \int_{-\infty}^{\infty} y f_Y(y|x) dy \\ \text{var}[Y|x] &:= \int_{-\infty}^{\infty} (y - E[Y|x])^2 f_Y(y|x) dy \\ &= E(Y^2|x) - \{E(Y|x)\}^2 \end{aligned}$$

Distributions

if X is iid with expectation μ , SD σ and $S_n = \sum_{i=1}^n X_i$,

- $E(S_n) = n\mu$
- $SD(S_n) = \sqrt{n}\sigma$
- variance of sum = sum of variances
- $\text{var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{var}(x_i)$

bernoulli

$X \sim \text{Bernoulli}(p) \Rightarrow$ coin flip with probability p

$$\begin{aligned} E(X_i) &= p & \text{var}(X_i) &= p(1 - p) \\ E(S_n) &= np & \text{var}(S_n) &= np(1 - p) \end{aligned}$$

binomial

$X \sim \text{Bin}(n, p) \Rightarrow X_i \stackrel{i.i.d.}{\sim} \text{Bernoulli}(p)$

$$E(X) = np, \quad \text{var}(X) = np(1 - p)$$

$$E(X) = \sum_{k=1}^n k \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\text{cov}(X, n - X) = -\text{var}(X)$$

multinomial

$X \sim \text{Multinomial}(n, \mathbf{p})$

- for k outcomes E_1, \dots, E_k , $\Pr(E_i) = p_i$. For some $1 \leq i \leq k$, E_i occurs X_i times in n runs.

(X_1, \dots, X_k) has the **multinomial distribution**:

$$\Pr(X_1 = x_1, \dots, X_k = x_k) = \binom{n}{x_1, \dots, x_k} \prod_{i=1}^k p_i^{x_i}$$

- where $\binom{n}{x_1, \dots, x_k} = \frac{n!}{x_1! x_2! \dots x_k!}$
 - combinatorially, # of arrangements of x_1, \dots, x_k
 - $\sum_{i=1}^n x_i = n$, $x_i \geq 0$

$$E(X) = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix}, \quad \text{var}(X_i) = np_i(1 - p_i)$$

$\text{var}(X) = \text{covariance matrix } M$ with

$$m_{ij} = \begin{cases} \text{var}(X_i) & \text{if } i = j \\ \text{cov}(X_i, X_j) & \text{if } i \neq j \end{cases}$$

- $\text{cov}(X_i, X_j) < 0$
- $X_i \sim \text{Bin}(n, p_i)$
- $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$

02. PROBABILITY (2)

Mean Square Error (MSE)

$$MSE = E\{(Y - c)^2\}$$

- predicting Y :
 - $MSE = \text{var}(Y) + \{E(Y) - c\}^2$
 - $\min MSE = \text{var}(Y)$ when $c = E(Y)$
- Y and X are correlated:
 - $MSE = \text{var}[Y|x] + \{E[Y|x] - c\}^2$
 - $MSE = E[(Y - c)^2|x] = E[\{Y - E(Y)\}^2|x]$
 - $\min MSE = \text{var}(Y|x)$ when $c = E[Y|x]$
 - if $c = E(Y)$ instead of $E(Y|x) \Rightarrow$ the MSE increases by $(E(Y|x) - E(Y))^2$

mean MSE

$$\frac{1}{n} \sum_{i=1}^n \text{var}[Y|x_i] \approx E\{\text{var}[Y|X]\}$$

random conditional expectations

let X, Y be r.v.s.

- $E[Y|X]$ is a r.v. which takes value $E[Y|x]$ with probability/density $f_X(x)$
- $\text{var}[Y|X]$ is a r.v. which takes value $\text{var}[Y|x]$ with probability/density $f_X(x)$

$$\begin{aligned} E(E[X_2|X_1]) &= E(X_2) \\ \text{var}(E[X_2|X_1]) + E(\text{var}[X_2|X_1]) &= \text{var}(X_2) \end{aligned}$$

CDF (cumulative distribution function)

for r.v. X , let $F(x) = P(X \leq x)$

- domain: \mathbb{R} ; codomain: $[0, 1]$

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

Standard Normal Distribution

$Z \sim N(0, 1)$ has density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, \quad -\infty < z < \infty$$

$$E(Z) = 0, \quad \text{var}(Z) = 1$$

$$\text{CDF, } \Phi(x) = P(Z \leq x) = \int_{-\infty}^x \phi(z) dz$$

- $E(Z) = \int_{-\infty}^{\infty} z \phi(z) dz = 0$
 - $E(Z^2) = \int_{-\infty}^{\infty} z^2 \phi(z) dz = 1$
 - $E(Z^{2k+1}) = 0 \quad \forall k \in \mathbb{Z}_{\geq 0}$

general normal distribution

let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$

$$\text{standardisation: } \frac{X - \mu}{\sigma} \sim N(0, 1)$$

- summations:
 - for constants $a, b \neq 0$,
 - $a + bX \sim N(a + b\mu, b^2\sigma^2)$
 - $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2 + 2\text{cov}(X, Y))$
 - $\text{cov}(X, Y) = 0, \Rightarrow X \perp Y$
 - $X \perp Y \Rightarrow X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$
- for $W = a + bX$,
 - density, $f_W(w) = \frac{d}{dw} F_W(w)$
 - CDF, $F_W(w) = P(X \leq \frac{w-a}{b}) = \Phi(\frac{w-a}{b})$

Central Limit Theorem

let X_1, \dots, X_n be iid rv's with expectation μ and SD σ , with $S_n = \sum_{i=1}^n X_i$

CLT

as $n \rightarrow \infty$, the distribution of the standardised $S_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ converges to $N(0, 1)$

- $E(S_n) = n\mu$, $\text{var}(S_n) = n\sigma^2$
- for large n , approximately $S_n \sim N(n\mu, n\sigma^2)$

bernoulli

let $X_i \sim \text{Bernoulli}(p)$. then $S_n \sim \text{Binom}(n, p)$

- for large n , $S_n = N(np, np(1 - p))$
- CLT: standardised $\frac{S_n - np}{\sqrt{n}\sqrt{p(1-p)}} \rightarrow N(0, 1)$ as $n \rightarrow \infty$

Distributions

chi-square (χ²)

- let $Z \sim N(0, 1)$. \Rightarrow then $Z^2 \sim \chi_1^2$
- Z^2 has χ^2 distribution with 1 degree of freedom
- degrees of freedom = number of RVs in the sum

$$\begin{aligned} E(Z^2) &= 1, & E(Z^4) &= 3 \\ \text{var}(Z^2) &= E(Z^4) - \{E(Z^2)\}^2 = 2 \end{aligned}$$

let V_1, \dots, V_n be iid χ_1^2 RVs and $V = \sum_{i=1}^n V_i$. then

$$\begin{aligned} V &\sim \chi_n^2 \\ E(V) &= n & \text{var}(V) &= 2n \end{aligned}$$

gamma

let $\alpha, \lambda > 0$. The *Gamma*(α, λ) density is

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

where $\Gamma(\alpha)$ is a number that makes density integrate to 1

- χ_n^2 RV $\sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$
 - χ_n^2 is a special case of Gamma!
 - density of χ_1^2 RV = $\frac{1}{\sqrt{2\pi}} v^{-1/2} e^{-v/2}$, $v > 0$
 $= \text{Gamma}(\frac{1}{2}, \frac{1}{2})$
- if $X_1 \sim \text{Gamma}(\alpha_1, \lambda)$ and $X_2 \sim \text{Gamma}(\alpha_2, \lambda)$ are independent, then $X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \lambda)$

t distribution

let $Z \sim N(0, 1)$ and $V \sim \chi_n^2$ be independent.

$$\frac{Z}{\sqrt{V/n}} \sim t_n$$

has a *t* distribution with *n* degrees of freedom.

- t* distribution is symmetric around 0
- $t_n \rightarrow Z$ as $n \rightarrow \infty$ (because $\frac{V}{n} \rightarrow 1$)

F distribution

let $V \sim \chi_m^2$ and $W \sim \chi_n^2$ be independent.

$$\frac{V/m}{W/n} \sim F_{m,n}$$

has an *F* distribution with (*m*, *n*) degrees of freedom.

- even if $m = n$, still two RVs V, W as they are independent
- for $T \sim t_n$, $T^2 = \frac{Z^2}{V/n} \sim F_{1,n}$

IID Random Variables

let X_1, \dots, X_n be iid RVs with mean \bar{X} .

$$\begin{aligned} \text{sample variance, } S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ S &\text{ is an estimate of } \sigma \end{aligned}$$

let X_1, \dots, X_n be iid $N(\mu, \sigma^2)$. $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

$$\begin{aligned} \bar{X} &\sim N(\mu, \frac{\sigma^2}{n}) \\ E(\bar{X}) &= \mu, & \text{var}(\bar{X}) &= \frac{\sigma^2}{n} \end{aligned}$$

more distributions:

$$\begin{aligned} \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} &\sim N(0, 1) \\ \frac{(n-1)S^2}{\sigma^2} &\sim \chi_{n-1}^2 \\ \frac{\bar{X} - \mu}{S/\sqrt{n}} &\sim t_{n-1} \end{aligned}$$

- \bar{X} and S^2 are independent

Multivariate Normal Distribution

let μ be a $k \times 1$ vector and Σ be a *positive-definite* symmetric $k \times k$ matrix.

the random vector $\mathbf{X} = (X_1, \dots, X_k)'$ has a multivariate normal distribution $N(\mu, \Sigma)$ if its density function is

$$\frac{1}{(2\pi)^{k/2} \sqrt{\det \Sigma}} \exp \left(-\frac{(\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)}{2} \right)$$

- $E(\mathbf{X}) = \mu$, $\text{var}(\mathbf{X}) = \Sigma$
- for any non-zero $k \times 1$ vector \mathbf{a} ,

$$\mathbf{a}' \mathbf{X} \sim N(\mathbf{a}' \mu, \mathbf{a}' \Sigma \mathbf{a})$$

- $\mathbf{a}' \Sigma \mathbf{a} > 0$ because Σ is positive-definite
- the product $\mathbf{a}' \mathbf{X}$ is a scalar (same for $\mathbf{a}' \mu, \mathbf{a}' \Sigma \mathbf{a}$)
- two multinomial normal random vectors \mathbf{X}_1 and \mathbf{X}_2 , sizes *h* and *k*, are independent if $\text{cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}_{h \times k}$
 - $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ has a multivariate normal distribution; the covariance between \bar{X} and $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ is 0, thus they are independent

03. POINT ESTIMATION

for a variable *v* in population *N*,

$$\mu = \frac{1}{N} \sum_{i=1}^N v_i \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^N (v_i - \mu)^2$$

- μ, σ^2 are **parameters** (unknown constants)
- a **simple random sample** is used to estimate parameters: individuals drawn from the population at random without replacement

binary variable

for variable *v* with proportion *p* in the population,

$$\mu = p, \quad \sigma^2 = p(1 - p)$$

single random draw

for variable *v* (population of size *N*, mean μ , variance σ^2), let *X* be the chosen *v*-value.

$$E(X) = \mu, \quad \text{var}(X) = \sigma^2$$

draws with replacement

let X_1, \dots, X_n be random draws with replacement from a population of mean μ and variance σ^2 .

$$\text{random sample mean, } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\begin{aligned} X_1, \dots, X_n &\text{ are iid with } E(X_i) = \mu, \text{var}(X_i) = \sigma^2 \\ E(\bar{X}) &= \mu, \text{var}(\bar{X}) = \frac{\sigma^2}{n} \end{aligned}$$

let x_1, \dots, x_n be realisations of *n* random draws with replacement from the population.

$$\text{sample mean, } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- as $n \rightarrow \infty$, $\bar{x} \rightarrow \mu$ (LLN)
- sample distribution, x_i has the same distribution as X_i and the population distribution

representativeness

- X_1, \dots, X_n is **representative** of the population
 - as *n* gets larger, \bar{X} gets closer to μ
- x_1, \dots, x_n are *likely* representative of the population

estimating mean

given data x_1, \dots, x_n ,

- sample mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ is an **estimate** of μ
- the error in \bar{x} is $\mu - \bar{x}$; it cannot be estimated
- \bar{x} is a realisation of the **estimator** \bar{X}
 - this realisation is used to estimate μ

standard error

the size of error in estimate \bar{x} is roughly $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

the **standard error** (SE) in \bar{x} is $\frac{\sigma}{\sqrt{n}}$

- SE is a constant by definition: $SE = SD(\hat{X}) = \frac{\sigma}{\sqrt{n}}$

estimating σ

intuitive estimate of σ^2 , $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

$$\begin{aligned} \text{sample variance, } s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ E(s^2) &= \sigma^2 \end{aligned}$$

Point estimation of mean

a population (size *N*) has unknown mean μ , variance σ^2 .

for random draws (without replacement) x_1, \dots, x_n :

\bar{x} is a realisation of \bar{X} , with $E(\bar{X}) = \mu$, $\text{var}(\bar{X}) = \frac{\sigma^2}{n}$

- μ is estimated as $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- error in \bar{x} is measured by the SE: $\frac{\sigma}{\sqrt{n}} = SD(\bar{X})$

- SE is estimated as $\frac{s}{\sqrt{n}}$

$\Rightarrow \mu$ is around \bar{x} , give or take $\frac{s}{\sqrt{n}}$

unbiased estimation

- since $E(\bar{X}) = \mu$, \bar{X} is an **unbiased** estimator of μ . \bar{x} is an unbiased estimate.
- S^2 is unbiased for σ^2 : $E(S^2) = \sigma^2$
- S* is *not* unbiased for σ : $E(S) < \sigma$

Simple random sampling (SRS)

n random draws *without replacement* from a population of mean μ and variance σ^2 .

- for $i = 1, \dots, n$, $E(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2$

- for $i \neq j$, $\text{cov}(X_i, X_j) = -\frac{\sigma^2}{N-1}$

- if n/N is relatively large,

- multiply SE by correction factor $\sqrt{\frac{N-n}{N-1}}$

- standard error = $\frac{N-n}{N-1} \frac{\sigma}{\sqrt{n}}$

$$E(\bar{X}) = \mu, \quad \text{var}(\bar{X}) = \frac{N-n}{N-1} \frac{\sigma^2}{n}$$

- if $n \ll N$, then SRS is like sampling *with replacement* (treat the data as if they come from IID RVs X_1, \dots, X_n)

$$E(\bar{X}) = \mu, \quad \text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

estimating proportion *p*

- in a 0-1 population, $\mu = p$, $\sigma^2 = p(1 - p)$
 - p* is estimated as \bar{x} (sample proportion of 1's)

- $SE = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = SD(\hat{p})$
 - estimated by replacing *p* with \hat{x}
- unbiased estimator \hat{p}

- $E(\hat{p}) = p$, $\text{var}(\hat{p}) = \frac{p(1-p)}{n}$, $SD(\hat{p}) = SE$
- the estimate of σ is $\hat{\sigma}$, not *s*
- e.g. if a SRS of size 100 has 78 white balls,
 $p \approx 0.78 \pm \frac{\sqrt{0.78 \times 0.22}}{\sqrt{100}}$

Gauss Model

Let x_i be a realisation of X_i . X_1, \dots, X_{100} are random draws with replacement from an imaginary population with mean *w* and variance σ^2 . *w* and σ^2 are parameters (unknown constants).

- $E(X_i) = w$, $\text{var } X_i = \sigma^2$ (since X_i is just 1 draw)
- $E(\bar{X}) = w$, $\text{var } \bar{X} = \frac{\sigma^2}{100}$

04. ESTIMATION (SE, bias, MSE)

let x_1, \dots, x_n be from random draws X_1, \dots, X_n with replacement from a population of mean μ and variance σ^2 .

sample mean \bar{x} is an *unbiased estimate* of μ

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

SE = $\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$ tells us roughly how far \bar{x} is from μ

sample variance, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

MSE and bias

suppose measurements were from a population with mean *w* + *b* where *b* is a constant: $x_i = w + b + \epsilon_i$

- $E(\bar{X}) = w + b$
- $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
 - $SE = \frac{\sigma}{\sqrt{n}}$ measures how far \bar{x} is from *w* + *b*, not *w*

- if *b* \neq 0, then \bar{x} is a biased estimate for *w*

$$\begin{aligned} MSE &= E\{(\bar{X} - w)^2\} = \frac{\sigma^2}{n} + b^2 \\ MSE &= SE^2 + bias^2 \end{aligned}$$

as $n \rightarrow \infty$, $MSE \rightarrow b^2$

conclusion

let θ be a parameter (constant) and $\hat{\theta}$ be an estimator (RV).

$$SE = SD(\hat{\theta}), \text{ bias} = E(\hat{\theta}) - \theta,$$

$$MSE = E\{(\hat{\theta} - \theta)^2\} = SE^2 + bias^2\}$$

05. INTERVAL ESTIMATION

let x_1, \dots, x_n be realisations of IID RVs X_1, \dots, X_n with unknown $\mu = E(X_i)$ and $\sigma^2 = \text{var}(X_i)$.

sample mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

sample variance, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

standard error, $SE = \frac{s}{\sqrt{n}}$

point estimation: $\mu \approx \bar{x}$, give or take $\frac{s}{\sqrt{n}}$

interval estimation: interval contains μ with some confidence level

interval estimation works well if

- X_i has a normal distribution, for any $n > 1$
- X_i has any other distribution but *n* is large

normal "upper-tail quantile" z_p

let $Z \sim N(0, 1)$. for $0 < p < 1$, let z_p be such that $p = \text{Pr}(Z > z_p)$

- e.g. $z_{0.5} = 0$
- $z_p = (1 - p)$ -quantile of *Z*
- for $0 < p < 0.5$, $\text{Pr}(-z_p \leq Z \leq z_p) = 1 - 2p$

(case 1) normal distribution with known σ^2

assume X_1, \dots, X_n are IID $\sim N(0, 1)$ with known σ^2 .
for $0 < \alpha < 1$, $\Pr(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$

confidence interval for μ : the random interval
$$\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

contains μ with probability $1 - \alpha$,
and produces the realisation $(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$

- $1 - \alpha$ is the **confidence level**
- Proof.* since $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$,
 - $\Pr(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$
 - $\Pr(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$

(case 2) normal distribution with unknown σ^2

assume X_1, \dots, X_n are IID $\sim N(\mu, \sigma^2)$ with unknown σ^2 .
replace σ with S :

- for $0 < p < 1$, let $t_{p,n}$ be such that $\Pr(t_n > t_{p,n}) = p$
- $t_{p,n}$ is the *upper* p quartile of the t distribution with n degrees of freedom
 - e.g. $t_{0.1,5} = 1.48$ (using $qt(0.9, 5)$)
- as $n \rightarrow \infty$, $t_{n,p} \rightarrow z_p$

- $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$
- $\Pr(\bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}})$
the random interval
$$\left(\bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \right)$$

contains μ with probability $1 - \alpha$.

- data x_1, \dots, x_n give realisations \bar{x} of \bar{X} and s of S , thus the random interval gives a $(1 - \alpha)$ -CI for μ :
$$\left(\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \right)$$

(case 3) general distribution with unknown σ^2

IID X_1, \dots, X_n with $E(X_i) = \mu$, $\text{var}(X_i) = \sigma^2$ unknown

- for large n , approximately $\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$
- since $\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0, 1) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$,
 - $\Pr(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}}) \approx 1 - \alpha$
 - $\Pr(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) \approx 1 - \alpha$for large n , the random interval
$$\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

contains μ with probability $\approx 1 - \alpha$

- data x_1, \dots, x_n give realisations \bar{x} of \bar{X} and s of S .
- $(\bar{x} - z_{\frac{\alpha}{2}} SE, \bar{x} + z_{\frac{\alpha}{2}} SE)$
is an *approximate* $(1 - \alpha)$ -CI for μ .
 - $SE = \frac{s}{\sqrt{n}}$
 - for SRS, multiply SE by correction factor $\sqrt{\frac{N-n}{N-1}}$

- contains μ with probability $< 1 - \alpha$
- probability $\rightarrow 1 - \alpha$ as $n \rightarrow \infty$
- exception:** for Bernoulli, $\sigma = \sqrt{p(1-p)}$ is not estimated by s , but by replacing p with the sample proportion

06. METHOD OF MOMENTS

modified notation of mass/density functions:

- bernoulli:** $f(x|p) = p^x(1-p)^{1-x}$, $x = 0, 1$
 - parameter space is $(0, 1)$
- poisson:** $f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$, $x = 0, 1, \dots$
 - parameter space is \mathbb{R}_+

parameter estimation

assuming data x_1, \dots, x_n are realisations of IID RVs X_1, \dots, X_n with mass/density function $f(x|\theta)$, where θ is unknown in parameter space Θ .

- 2 methods to estimate θ :
 - method of moments (MOM)
 - method of maximum likelihood (MLE)
- for both:
 - the estimate of θ is a realisation of an estimator $\hat{\theta}$
 - SE is $SD(\hat{\theta})$
 - bias is $E(\hat{\theta}) - \theta$
- parameter space Θ : set of values that can be used to estimate the real parameter value θ

Moments of an RV

the k -th moment of an RV X is
$$\mu_k = E(X^k), \quad k = 1, 2, \dots$$

estimating moments

let X_1, \dots, X_n be IID with the same distribution as X .
the k -th sample moment is
$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

- $E(\hat{\mu}_k) = \mu_k \Rightarrow$ unbiased estimator!
- $\hat{\mu}_k$ is an estimator of μ_k . For realisations x_1, \dots, x_n , the realisation $\frac{1}{n} \sum_{i=1}^n x_i^k$ is an *unbiased* estimate of μ_k .
- hat (^) means estimator (random variable)
 - note that this violates the uppercase=RV, lowercase=(fixed)realisation notation
- $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

MOM: Poisson

assume x_1, \dots, x_n are realisations of IID *Poisson*(λ) RVs X_1, \dots, X_n . Let λ be the mean number of emissions per 10 seconds (λ is a parameter).

- let $X \sim \text{Poisson}(\lambda)$. $\mu_1 = \lambda$. Estimate λ by estimating μ_1 using sample mean \bar{x} , which is an estimator of \bar{X} .
- the MOM estimator is $\hat{\lambda} = \hat{\mu}_1 = \bar{X}$
 - the random sample mean
- $\text{var}(X) = \lambda$, $\text{var}(\bar{X}) = \frac{\lambda}{n}$, SE = SD of estimator = $\sqrt{\frac{\lambda}{n}}$
$$\lambda \approx \bar{x} \pm \sqrt{\frac{\lambda}{n}}$$

MOM: Bernoulli

Assume X_1, \dots, X_n are iid *Bernoulli*(p) RVs.
Finding MOM estimator of p :

- let $X \sim \text{Bernoulli}(p)$. $\Rightarrow \mu_1 = p$
- MOM estimator, $\hat{p} = \hat{\mu}_1 = \bar{X}$
 - random sample proportion of 1's
- SE = SD of estimator = $\sqrt{\text{var}(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}$

MOM: Normal

let X_1, \dots, X_n be iid $N(\mu, \sigma^2)$ with parameters μ, σ^2
for $X \sim N(\mu, \sigma^2)$: parameter space, $\Theta = \mathbb{R} \times \mathbb{R}_+$

- $\mu_1 = \mu$, $\mu_2 = \sigma^2 + \mu^2$
- express $\mu = \mu_1$; $\sigma^2 = \mu_2 - \mu_1^2$; then add hats
- MOM estimators:
$$\hat{\mu} = \bar{X} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

(to construct CI for σ^2 : use $S^2 \Rightarrow$ since $E(S^2) = \sigma^2$)

MOM: Geometric

let x_1, \dots, x_n be realisations of IID *Geometric*(p) RVs X_1, \dots, X_n with expectation $1/p$.

- for $X \sim \text{Geometric}(p) \Rightarrow E(X) = \frac{1}{p}$
 - $\Pr(X = i) = p(1-p)^{i-1}$ for $i = 1, 2, \dots$
 - $E(X) = \sum_{i=1}^{\infty} i p(1-p)^{i-1} = \frac{1}{p}$
- $\mu_1 = \frac{1}{p} \Rightarrow p = \frac{1}{\mu_1} \Rightarrow \hat{p} = \frac{1}{\bar{X}}$
- MOM estimator, $\hat{p} = \frac{1}{\bar{X}}$
 - then MOM estimate = $\frac{1}{\bar{x}}$
- SE = $SD(1/\bar{X}) \Rightarrow$ use monte carlo to approximate

MOM: Gamma

let X_1, \dots, X_n be iid *Gamma*(α, λ) RVs with shape parameter $\alpha > 0$, rate parameter $\lambda > 0$

- $X \sim \text{Gamma}(\alpha, \lambda)$, $E(X) = \frac{\alpha}{\lambda}$, $E(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}$
- express parameters in terms of moments:
$$\mu_1 = \frac{\alpha}{\lambda}, \mu_2 - \mu_1^2 = \frac{\alpha}{\lambda^2} \Rightarrow \lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}, \alpha = \lambda \mu_1$$
- MOM estimators: $\hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2}$, $\hat{\lambda} = \frac{\bar{X}}{\hat{\sigma}^2}$

MOM estimators are consistent

let X_1, \dots, X_n be iid with mass/density $f(x|\theta)$, where $\theta \in \Theta \subset \mathbb{R}$.
Suppose $\theta = g(\mu_1)$ for some *continuous* function g .
Then the MOM estimator is **consistent** (approaches θ with more data)

- the MOM estimator is $\hat{\theta} = g(\hat{\mu}_1)$. as $n \rightarrow \infty$, $\hat{\mu}_1 \rightarrow \mu_1$
- since g is continuous, $\hat{\theta} \rightarrow g(\mu_1) = \theta$
 - asymptotic unbiasedness:** $E(\hat{\theta}) \rightarrow \theta$

07. MLE

MOM: works through estimating moments - if no formula is available for $SD(\hat{\theta})$ or $E(\hat{\theta})$, monte carlo can be used
MLE: another estimation method

Likelihood function

let x_1, \dots, x_n be realisations of iid rvs X_1, \dots, X_n with density $f(x|\theta)$, $\theta \in \Theta \subset \mathbb{R}^k$.
likelihood function $L: \Theta \rightarrow \mathbb{R}_+$ is
$$L(\theta) = f(x_1|\theta) \times \dots \times f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$$
loglikelihood function $\ell: \Theta \rightarrow \mathbb{R}$ is
$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i|\theta)$$

Maximum Likelihood Estimation (MLE)

$$L(\theta) = \prod_{i=1}^n f(x_i|\theta)$$

- maximiser** of $L \rightarrow$ the maximum likelihood estimate of θ (a realisation of the MLEstimator $\hat{\theta}$)

poisson (log)likelihood/MLE

Poisson(λ): $f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$, $x = 0, 1, 2, \dots$

- let x_1, \dots, x_n be realisations of iid *Poisson*(λ) RVs X_1, \dots, X_n . the joint probability of data is
$$f(x_1|\lambda) \times \dots \times f(x_n|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{x_1! \dots x_n!}$$
- likelihood:** probability as a function of only λ
$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{x_1! \dots x_n!}$$
- loglikelihood:**
$$\ell(\lambda) = (\sum_{i=1}^n x_i) \log \lambda - n\lambda - \sum_{i=1}^n \log(x_i!)$$
- MLE** of $\lambda = \bar{x}$ (maximiser of $L(\lambda)$)
 - differentiate $\ell(\lambda)$: $\ell'(\lambda) = \frac{\sum_{i=1}^n x_i}{\lambda} - n$
 - $\ell'(\lambda) = 0 \Rightarrow \lambda = \bar{x}$
 - $\ell''(\lambda) < 0$ (thus max point)

normal (log)likelihood/MLE

$N(\mu, \sigma^2)$: for $x \in \mathbb{R}$,
$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = (2\pi)^{\frac{1}{2}} \sigma^{-1} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- let x_1, \dots, x_n be realisations of iid $N(\mu, \sigma)$ RVs X_1, \dots, X_n . the joint probability of data is
$$f(x_1|\lambda) \times \dots \times f(x_n|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{x_1! \dots x_n!}$$
- likelihood** function: joint density as a function of (μ, σ)
$$L(\mu, \sigma) = f(x_1|\mu, \sigma) \times \dots \times f(x_n|\mu, \sigma)$$

$$= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

- loglikelihood:**
$$\ell(\mu, \sigma) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$
- MLE:**
 - MLE of $\mu = \bar{x}$
 - MLE of $\sigma = \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$