# ST2132
AY23/24 SEM 1
github/jovyntls

# 01. PROBABILITY

- **probability** of an event → 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)
$$E(X) := \sum_{i=1}^{n} x_i p_i$$

**continuous**: (density function)
$$E(X) := \int_{-\infty}^{\infty} x f(x)\, dx$$

### 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

**variance**, $\text{var}(X) := E\{(X-\mu)^2\}$
**standard deviation**, $SD(X) := \sqrt{\text{var}(X)}$

- $\text{var}(X) = E(X^2) - E(X)^2$
- $E(X - \mu) = 0$

## Law of Large Numbers

mean and variance of $r$ realisations:
$$\bar{x} := \frac{1}{r}\sum_{i=1}^{r} x_i \qquad v := \frac{1}{r}\sum_{i=1}^{r}(x_i - \bar{x})^2$$

**LLN:** for a function $h$, as $r \to \infty$,
$$\frac{1}{r}\sum_{i=1}^{r} h(x_i) \to E\{h(X)\}$$
$$\bar{x} \to E(X), \quad v \to \text{var}(X)$$

### Monte Carlo approximation

simulate $x_1, \dots, x_r$ from $X$. by LLN, as $r \to \infty$, 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:
$$f : \mathbb{R}^2 \to [0, \infty), \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$$
(**expectation**) for $h : \mathbb{R}^2 \to \mathbb{R}$,
$$E\{h(X,Y)\} = \begin{cases} \sum_{i=1}^{I}\sum_{j=1}^{J} h(x_i, y_j) p_{ij} & X \text{ is discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\, dx\, dy & Y \text{ is continuous} \end{cases}$$

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

## joint = marginal × conditional distributions

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

- $f(x,y)$ 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* ≡ *probability*, $x \equiv x_i$, $y \equiv y_j$

## Independence

- $X, Y$ are independent $\iff \forall x, y \in \mathbb{R}$,
  1. $f(x,y) = f_X(x) f_Y(y)$
  2. $f_Y(y|x) = f_Y(y)$
  3. $f_X(x|y) = f_Y(x)$
- $X, Y$ are independent ⇒
  - $E(XY) = E(X)E(Y)$
  - $\text{cov}(X,Y) = 0$

  (the converse does not hold)

## Conditional expectation

### discrete case

let $f_Y(\cdot|x_i)$ be the conditional pmf of $Y$ given $X = x_i$.
$$E[Y|x_i] := \sum_{j=1}^{J} y_j f_Y(y_j|x_i)$$
$$\text{var}[Y|x_i] := \sum_{j=1}^{J}(y_j - E[Y|x_i])^2 f_Y(y_j|x_i)$$

$E[Y|x_i]$ 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

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

## Distributions

if $X$ is iid with expectation $\mu$, SD $\sigma$ 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

$X \sim Bernoulli(p)$ ⇒ coin flip with probability $p$
$$E(X_i) = p \qquad \text{var}(X_i) = p(1-p)$$
$$E(S_n) = np \qquad \text{var}(S_n) = np(1-p)$$

### binomial

$X \sim Bin(n,p) \quad \Rightarrow X_i \overset{i.i.d.}{\sim} Bernoulli(p)$
$$E(X) = np, \quad \text{var}(X) = np(1-p)$$
$$E(X) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$
$$\text{cov}(X, n - X) = -\text{var}(X)$$

### multinomial

$X \sim Multinomial(n, \mathbf{p})$
- for $k$ outcomes $E_1, \dots, E_k$, $Pr(E_i) = p_i$. For some $1 \le i \le k$, $E_i$ occurs $X_i$ times in $n$ runs.

$(X_1, \dots, X_k)$ has the **multinomial distribution:**
$$Pr(X_1 = x_1, \dots, X_k = x_k) = \binom{n}{x_1, \dots, x_k}\Pi_{i=1}^{k} p_i^{x_i}$$
- where $\binom{n}{x_1, \dots, x_k} = \frac{n!}{x_1! x_2! \dots x_k!}$
  - combinatorially, # of arrangements of $x_1, \dots, x_k$
  - $\sum_{i=1}^{n} x_i = n, \quad x_i \ge 0$

$$E(X) = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix}, \quad \text{var}(X_i) = np_i(1-p_i)$$
$\text{var}(X)$ = *covariance matrix* $M$ with
$$m_{ij} = \begin{cases} \text{var}(X_i) & \text{if } i = j \\ \text{cov}(X_i, X_j) & \text{if } i \ne j \end{cases}$$
- $\text{cov}(X_i, X_j) < 0$
- $X_i \sim Bin(n, p_i)$
- $X_i + X_j \sim Bin(n, p_i + p_j)$

# 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 $E(Y|x)$ ⇒ 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 $f_X(x)$

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

## CDF (cumulative distribution function)

for r.v. $X$, let $F(x) = P(X \le x)$
- domain: $\mathbb{R}$; codomain: $[0,1]$
$$F(x) = \int_{-\infty}^{\infty} f(x)\, dx$$

## Standard Normal Distribution

$Z \sim N(0,1)$ has density function
$$\phi(z) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{z^2}{2}\}, \quad -\infty < z < \infty$$
$$E(Z) = 0, \quad \text{var}(Z) = 1$$
**CDF**, $\Phi(x) = P(Z \le x) = \int_{-\infty}^{x}\phi(z)\, dz$
- $E(Z) = \int_{-\infty}^{\infty} z\phi(z)\, dz = 0$
  - $E(Z^2) = \int_{-\infty}^{\infty} z^2\phi(z)\, dz = 1$
  - $E(Z^{2k+1}) = 0 \quad \forall k \in \mathbb{Z}_{\ge 0}$

### general normal distribution

let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$
**standardisation:** $\frac{X - \mu}{\sigma} \sim N(0,1)$
- summations:
  - for constants $a, b \ne 0$,
    $a + bX \sim N(a + b\mu, \ b^2\sigma^2)$
  - $X + Y \sim N(\mu + \nu, \ \sigma^2 + \tau^2 + 2\,\text{cov}(X,Y))$
    - $\text{cov}(X,Y) = 0, \quad \Rightarrow \quad X \perp Y$
    - $X \perp Y \quad \Rightarrow \quad X + Y \sim N(\mu + \nu, \ \sigma^2 + \tau^2)$
- for $W = a + bX$,
  - density, $f_W(w) = \frac{d}{dw}F_W(w)$
  - CDF, $F_W(w) = P(X \le \frac{w-a}{b}) = \Phi(\frac{w-a}{b})$

## Central Limit Theorem

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

**CLT**
as $n \to \infty$, 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 $X_i \sim Bernoulli(p)$. then $S_n \sim Binom(n,p)$
- for large $n$, $S_n = N(np, np(1-p))$
- CLT: standardised $\frac{S_n - np}{\sqrt{n}\sqrt{p(1-p)}} \to N(0,1)$ as $n \to \infty$

# Distributions

## chi-square ($\chi^2$)

let $Z \sim N(0,1)$. $\Rightarrow$ then $Z^2 \sim \chi^2_1$

- $Z^2$ has $\chi^2$ distribution with 1 degree of freedom
- degrees of freedom = number of RVs in the sum

$$E(Z^2) = 1, \quad E(Z^4) = 3$$
$$\text{var}(Z^2) = E(Z^4) - \{E(Z^2)\}^2 = 2$$

let $V_1, \ldots, V_n$ be iid $\chi^2_1$ RVs and $V = \sum_{i=1}^n V_i$. then

$$V \sim \chi^2_n$$
$$E(V) = n \quad \text{var}(V) = 2n$$

## gamma

let $\alpha, \lambda > 0$. The $Gamma(\alpha, \lambda)$ 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

- $\chi^2_n$ RV $\sim Gamma(\frac{n}{2}, \frac{1}{2})$
  - $\chi^2_n$ is a special case of Gamma!
  - density of $\chi^2_1$ RV $= \frac{1}{\sqrt{2\pi}} v^{-1/2} e^{-v/2}, \quad v > 0$
    $= Gamma(\frac{1}{2}, \frac{1}{2})$
- if $X_1 \sim Gamma(\alpha_1, \lambda)$ and $X_2 \sim Gamma(\alpha_2, \lambda)$ are independent, then $X_1 + X_2 \sim Gamma(\alpha_1 + \alpha_2, \lambda)$

## t distribution

let $Z \sim N(0,1)$ and $V \sim \chi^2_n$ 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
- $t_n \to Z$ as $n \to \infty$ (because $\frac{V}{n} \to 1$)

## *F* distribution

let $V \sim \chi^2_m$ and $W \sim \chi^2_n$ 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 $X_1, \ldots, X_n$ be iid RVs with mean $\bar{X}$.

**sample variance**, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

$S$ is an estimate of $\sigma$

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

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

more distributions:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$
$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

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

# Multivariate Normal Distribution

let $\boldsymbol{\mu}$ be a $k \times 1$ vector and $\boldsymbol{\Sigma}$ be a *positive-definite* symmetric $k \times k$ matrix.

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

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

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

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

  - $\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} > 0$ because $\boldsymbol{\Sigma}$ is positive-definite
  - the product $\boldsymbol{a}'\boldsymbol{X}$ is a scalar (same for $\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}$)
- two multinomial normal random vectors $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$, sizes $h$ and $k$, are independent if $\text{cov}(\boldsymbol{X_1}, \boldsymbol{X_2}) = \boldsymbol{0}_{h \times k}$
  - $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ has a multivariate normal distribution; the covariance between $\bar{X}$ and $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ is $0$, thus they are independent

# 03. POINT ESTIMATION

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

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

- $\mu$, $\sigma^2$ 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,

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

## single random draw

for variable $v$ (population of size $N$, mean $\mu$, variance $\sigma^2$), let $X$ be the chosen $v$-value.

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

## draws with replacement

let $X_1, \ldots, X_n$ be random draws with replacement from a population of mean $\mu$ and variance $\sigma^2$.

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

$X_1, \ldots, X_n$ are iid with $E(X_i) = \mu$, $\text{var}(X_i) = \sigma^2$

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

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

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

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

## representativeness

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

## estimating mean

given data $x_1, \ldots, x_n$,

- sample mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ is an **estimate** of $\mu$
- 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 $\mu$

## 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 $\sigma$

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

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

$$E(s^2) = \sigma^2$$

## Point estimation of mean

a population (size $N$) has unknown mean $\mu$, variance $\sigma^2$. for random draws (without replacement) $x_1, \ldots, x_n$:

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

- $\mu$ 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 $\mu$. $\bar{x}$ is an unbiased estimate.
- $S^2$ is unbiased for $\sigma^2$: $E(S^2) = \sigma^2$
- $S$ is *not* unbiased for $\sigma$: $E(S) < \sigma$

## Simple random sampling (SRS)

$n$ random draws *without replacement* from a population of mean $\mu$ and variance $\sigma^2$.

- for $i = 1, \ldots, n$, $E(X_i) = \mu$ 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 = $\frac{N-n}{N-1}$

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

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

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

## estimating proportion $p$

- in a 0-1 population, $\mu = p, \quad \sigma^2 = p(1-p)$
  - $p$ 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 $\bar{x}$
- unbiased estimator $\hat{p}$
  - $E(\hat{p}) = p, \quad \text{var}(\hat{p}) = \frac{p(1-p)}{n}, \quad SD(\hat{p}) = SE$
- the estimate of $\sigma$ 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 $X_i$. $X_1, \ldots, X_{100}$ are random draws with replacement from an imaginary population with mean $w$ and variance $\sigma^2$. $w$ and $\sigma^2$ are parameters (unknown constants).

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

# 04. ESTIMATION (SE, bias, MSE)

let $x_1, \ldots, x_n$ be from random draws $X_1, \ldots, X_n$ with replacement from a population of mean $\mu$ and variance $\sigma^2$.

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

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

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

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: $\quad x_i = w + b + \epsilon_i$

- $E(\bar{X}) = w + b$
- $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
  - $SE = \frac{\sigma}{\sqrt{n}}$ measures how far $\bar{x}$ is from $w + b$, not $w$
- if $b \neq 0$, then $\bar{x}$ is a biased estimate for $w$

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

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

## conclusion

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

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

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

# 05. INTERVAL ESTIMATION

let $x_1, \ldots, x_n$ be realisations of IID RVs $X_1, \ldots, X_n$ 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 $\mu$ 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

$$p = \Pr(Z > z_p)$$

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

**(case 1) normal distribution with known $\sigma^2$**

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

**confidence interval for $\mu$:** 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 $\mu$ with probability $1-\alpha$,

and produces the realisation $\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

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

**(case 2) normal distribution with unknown $\sigma^2$**

assume $X_1, \ldots, X_n$ are IID $\sim N(\mu, \sigma^2)$ with unknown $\sigma^2$.
replace $\sigma$ 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 $n \to \infty$, $t_{n,p} \to z_p$
- $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
- $\Pr(\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}})$

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 $\mu$ with probability $1-\alpha$.

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

**(case 3) general distribution with unknown $\sigma^2$**

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

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

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 $\mu$ with probability $\approx 1-\alpha$

- data $x_1, \ldots, x_n$ give realisations $\bar{x}$ of $\bar{X}$ and $s$ of $S$.
- $\left(\bar{x} - z_{\frac{\alpha}{2}}SE, \bar{x} + z_{\frac{\alpha}{2}}SE\right)$
  is an *approximate* $(1-\alpha)$-CI for $\mu$.
  - $SE = \frac{s}{\sqrt{n}}$
  - for SRS, multiply $SE$ by correction factor $\sqrt{\frac{N-n}{N-1}}$
- contains $\mu$ with probability $< 1-\alpha$
- probability $\to 1-\alpha$ as $n \to \infty$
- **exception**: for Bernoulli, $\sigma = \sqrt{p(1-p)}$ is not estimated by $s$, but by replacing $p$ with the sample proportion