Mathematical Notation

$\equiv$ identity (always reversible)
$\Rightarrow$ is implied by
$\Leftrightarrow$ if and only if (iff)
$\checkmark$ checked
$\therefore$ therefore
$\because$ because
$\square$ Q.E.D.
$:=$ is defined by
$\rightarrow$ tends to
$\forall$ for all
$f: x \mapsto y$ function x maps to y
$\dot{x}, \ddot{x}, \dddot{x}$... nth derivative w.r.t. x
$P(A|B)$ probability of A given B
ϕ probability density function of standardized $N(0,1)$
Φ corresponding cumulative distribution function

$\mathbb{N}$ natural numbers
$\mathbb{Z}$ integers (e.g. $\mathbb{Z}_0^+$)
$\mathbb{Q}$ rational numbers $\{\frac{a}{b}, a \in \mathbb{Z}, b \in \mathbb{N}\}$
$\mathbb{R}$ real numbers
$\mathbb{I}$ imaginary numbers
$\mathbb{C}$ complex numbers
$\mathbb{P}$ prime numbers
$\in \notin$ is (not) an element of
$\subseteq \subset$ is a (proper) subset of
$\{x: ...\}$ set of all x such that ...
$n(A)$ # of elements in set A
$\varnothing$ empty set
ξ 'universal set'
A' complement of set A
$\cup$ union (OR)
$\cap$ intersection (AND)

w.l.o.g. without loss of generality
w.r.t. with respect to
w.t.p. want to prove
s.t. such that
c.v. critical value
p.m.c.c. Pearson's product moment correlation coefficient
α significance level
ν degrees of freedom
r.v. random variable
$E(X)$ expectation of r.v. X
$Var(X)$ variance of r.v. X
$p(x), P(X=x)$ probability function of discrete r.v. X, and its value for x
$X \sim B(n,p)$ X has binomial distr. w/ 'n' # of trials and 'p' probability of success in trial
$X \sim N(\mu, \sigma^2)$ X has normal distr. w/ mean μ and variance σ^2

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\bar{X}$ is the sample mean distribution with mean μ and s.d. $\sigma/\sqrt{n}$
$Z \sim N(0,1)$ standard Normal distribution
$X \sim U(n)$ X has Uniform distribution over $1, 2, ..., n$
$X \sim Geo(p)$ X has Geometric distribution with probability 'p' per outcome

μ population mean — $\bar{x}$ sample mean
σ^2 population variance — s^2 sample variance
σ population s.d. — s sample s.d.
ρ population p.m.c.c. — r sample p.m.c.c

H_0 Null Hypothesis — H_1 Alternative Hypothesis

Deriving Quadratic Formula

stretch factor, line of symmetry $x = h$, vertex point (h, k); domain: $\mathbb{R}$; range: $[b, \infty)$ and $(-\infty, k]$

$$f(x) = ax^2 + bx + c \Rightarrow a(x-h)^2 + k$$

$$= a\left[\left(x^2 + \frac{b}{a}x\right) + \frac{c}{a}\right] = a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a^2} + \frac{c}{a}\right]$$

$$= a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a} \qquad \boxed{\text{vertex point: } \left(-\frac{b}{2a}, \frac{4ac - b^2}{4a}\right)}$$

$$\Rightarrow \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}, \text{ so when } f(x) = 0, \quad \boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

Finding Inverse Functions

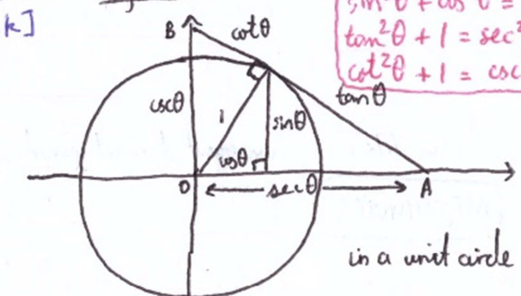
Given $f(x) = 3x^2 - 4$, find $f^{-1}(x)$ where $\{x \in \mathbb{R} : x \geq 0\}$

$\rightarrow f(x) = 3x^2 - 4$, let $y = f^{-1}(x)$ $\therefore x = f(y)$

$\rightarrow x = 3y^2 - 4 \Rightarrow y^2 = \frac{x+4}{3} \quad \therefore y = \sqrt{\frac{x+4}{3}}$ (as $y \geq 0$)

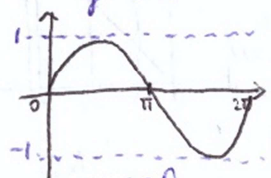
Definitions

$$\boxed{\begin{aligned} \sin^2\theta + \cos^2\theta &= 1 \\ \tan^2\theta + 1 &= \sec^2\theta \\ \cot^2\theta + 1 &= \csc^2\theta \end{aligned}}$$



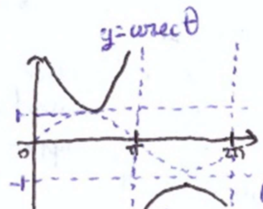
in a unit circle

$y = \sin\theta$



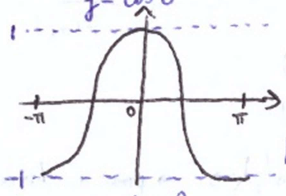
domain $\theta \in \mathbb{R}$, range $[-1, 1]$

$y = \csc\theta$



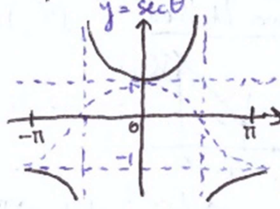
domain $\theta \in \mathbb{R}$, $\theta \neq n\pi$, range $(-\infty, -1] \cup [1, \infty)$

$y = \cos\theta$



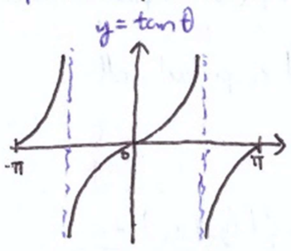
domain $\theta \in \mathbb{R}$, range $[-1, 1]$

$y = \sec\theta$



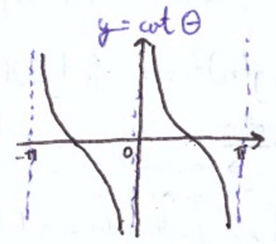
domain $\theta \in \mathbb{R}$, $\theta \neq \frac{2n-1}{2}\pi$, range $(-\infty, -1] \cup [1, \infty)$

$y = \tan\theta$



domain $\theta \in \mathbb{R}$, $\theta \neq \frac{2n-1}{2}\pi$, range $y \in \mathbb{R}$

$y = \cot\theta$

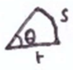


domain $\theta \in \mathbb{R}$, $\theta \neq n\pi$, range $y \in \mathbb{R}$

Trigonometry

$\boxed{Area = \frac{1}{2}ab\sin C}$ $\boxed{\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}}$ $\boxed{\cos A = \frac{b^2 + c^2 - a^2}{2bc}}$

$\boxed{S = r\theta}$ $\boxed{A = \frac{1}{2}r^2\theta}$ $\boxed{\theta = \text{degrees} \times \frac{\pi}{180}}$



CO-FUNCTION

$$\boxed{\begin{aligned} \sin\theta &= \cos(\pi/2 - \theta) \\ \tan\theta &= \cot(\pi/2 - \theta) \\ \sec\theta &= \csc(\pi/2 - \theta) \end{aligned}} \quad \boxed{\begin{aligned} \cos\theta &= \sin(\pi/2 - \theta) \\ \cot\theta &= \tan(\pi/2 - \theta) \\ \csc\theta &= \sec(\pi/2 - \theta) \end{aligned}}$$

SUM/DIFFERENCE

$$\boxed{\begin{aligned} \sin(u \pm v) &= \sin u \cos v \pm \cos u \sin v \\ \cos(u \pm v) &= \cos u \cos v \mp \sin u \sin v \\ \tan(u \pm v) &= \frac{\tan u \pm \tan v}{1 \mp \tan u \tan v} \end{aligned}}$$

⚠ always draw the graph for questions

DOUBLE

$$\boxed{\begin{aligned} \sin(2u) &= 2\sin u \cos u \\ \cos(2u) &= \cos^2 u - \sin^2 u = 2\cos^2 u - 1 = 1 - 2\sin^2 u \\ \tan(2u) &= \frac{2\tan u}{1 - \tan^2 u} \end{aligned}}$$

Inverse Trigonometric Functions

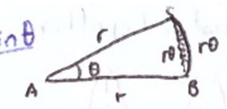
	domain	range
$\arcsin\theta$	$[-1, 1]$	$[-\pi/2, \pi/2]$
$\arccos\theta$	$[-1, 1]$	$[0, \pi]$
$\arctan\theta$	$(-\infty, \infty)$	$(-\pi/2, \pi/2)$
$\text{arccot}\,\theta$	$(-\infty, \infty)$	$(0, \pi)$

HALF-ANGLE

$$\boxed{\sin\frac{A}{2} = \pm\sqrt{\frac{1 - \cos A}{2}}} \quad \boxed{\cos\frac{A}{2} = \pm\sqrt{\frac{1 + \cos A}{2}}} \quad \boxed{\tan\frac{A}{2} = \pm\sqrt{\frac{1 - \cos A}{1 + \cos A}}}$$

Small Angle Approximations

$\sin\theta$



$$Area(segment) = \frac{1}{2}r^2\theta - \frac{1}{2} \cdot r \cdot r \cdot \sin\theta = \frac{1}{2}r^2(\theta - \sin\theta)$$

For small angles, $\theta - \sin\theta \simeq 0$ (as area of segment is negligible)

$\therefore \boxed{\sin\theta \simeq \theta}$ (overestimate)

$\cos\theta$ Given that $a^2 = b^2 + c^2 - 2bc\cos A$

$\therefore (r\theta)^2 \simeq r^2 + r^2 - 2r^2\cos\theta \Rightarrow \cancel{r^2}\theta^2 \simeq 2\cancel{r^2}(1 - \cos\theta)$

$\Rightarrow \frac{1}{2}\theta^2 \simeq 1 - \cos\theta \quad \therefore \boxed{\cos\theta \simeq 1 - \frac{1}{2}\theta^2}$

$\tan\theta$ $\tan\theta = \frac{\sin\theta}{\cos\theta} \simeq \frac{\theta}{1 - \frac{\theta^2}{2}} \simeq \frac{\theta(1 + \frac{\theta^2}{2})}{1 - \frac{1}{4}\theta^4} \simeq \frac{\theta}{1 - \frac{1}{4}\theta^4} + \cancel{\frac{\frac{1}{2}\theta^3}{1 - \frac{1}{4}\theta^4}}$

$\therefore \boxed{\tan\theta \simeq \theta}$ (underestimate)

Also given $\triangle ABC$, $\tan\theta \simeq \frac{BC}{AB} \simeq \frac{r\theta}{r} = \theta$

other $\frac{1 - \cos\theta}{\theta} = \frac{1 - \cos\theta}{\theta} \cdot \frac{1 + \cos\theta}{1 + \cos\theta} = \frac{1 - \cos^2\theta}{\theta(1 + \cos\theta)} = \frac{\sin\theta}{\theta} \cdot \frac{\sin\theta}{1 + \cos\theta}$

$\lim_{\theta \to 0} \frac{\sin\theta}{\theta} \cdot \frac{\sin\theta}{1 + \cos\theta} = 1 \cdot \frac{\sin 0}{1 + \cos 0} = 0 \quad \boxed{\lim_{\theta \to 0} \frac{1 - \cos\theta}{\theta} = 0}$

Derivative of Trigonometric Functions

$\sin x$ $\frac{dy}{dx} = \lim_{p,q \to x} \frac{\sin(q) - \sin(p)}{q - p} = \frac{\sin(x+h) - \sin(x-h)}{2h} = \frac{\cancel{\sin x \cos h} + \cos x \sin h - (\cancel{\sin x \cos h} - \cos x \sin h)}{2h}$

$= \frac{2\cos x \sin h}{2h} = \frac{\cos x \sin h}{h} = \cos x \cdot \frac{\sin h}{h}$. $\frac{dy}{dx} = \lim_{h \to 0} \cos x \cdot \frac{\cancel{h}}{\cancel{h}} = \cos x \quad \therefore \boxed{\frac{dy}{dx} = \cos x}$

$\cos x$ $\frac{dy}{dx} = \lim_{h \to 0} \frac{\cos(x+h) - \cos(x-h)}{2h} = \frac{(\cancel{\cos x \cos h} - \sin x \sin h) - (\cancel{\cos x \cos h} + \sin x \sin h)}{2h}$

$= \frac{-2\sin x \sin h}{2h} = \lim_{h \to 0} -\sin x \cdot \frac{\sin h}{h} = -\sin x \quad \therefore \boxed{\frac{dy}{dx} = -\sin x}$

$\tan x$ $\frac{dy}{dx} = \left(\frac{\sin x}{\cos x}\right)' = \frac{\cos x \cdot \cos x - \sin x(-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x \quad \therefore \boxed{\frac{dy}{dx} = \sec^2 x}$

$\sec x$ $\frac{dy}{dx}$: $\tan^2 x + 1 = \sec^2 x \Rightarrow 2\tan x \sec^2 x = 2\sec x(\sec x)' \Rightarrow (\sec x)' = \frac{2\tan x \sec^2 x}{2\sec x} \quad \therefore \boxed{\frac{dy}{dx} = \tan x \sec x}$

$\cot x$ $(\cot x)' = \left(\frac{\cos x}{\sin x}\right)' = \frac{\sin x(-\sin x) - \cos x(\cos x)}{\sin^2 x} = \frac{-(\sin^2 x + \cos^2 x)}{\sin^2 x} = \frac{-1}{\sin^2 x} \quad \therefore \boxed{\frac{dy}{dx} = -\csc^2 x}$

$\csc x$ $\cot^2 x + 1 = \csc^2 x \Rightarrow 2\cot x \cdot -\csc^2 x = 2\csc x(\csc x)' \Rightarrow (\csc x)' = \frac{-\cancel{2}\cot x \csc^{\cancel{2}} x}{\cancel{2\csc x}} \quad \therefore \boxed{\frac{dy}{dx} = -\cot x \csc x}$

Differentiation from First Principles

$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$ (right derivative) or $f'(x) = \lim_{p,q \to x} \frac{f(q) - f(p)}{q - p}$ (left/right derivative)

$f(x) = x^3$

$f'(x) = \frac{(x+h)^3 - x^3}{h} = \frac{\cancel{x^3} + 3x^2h + 3xh^2 + h^3 - \cancel{x^3}}{h} = \lim_{h \to 0} 3x^2 + 3xh + h^2 = 3x^2$

$f(x) = x^6$

$f'(x) = \frac{q^6 - p^6}{q - p} = \frac{\cancel{(q-p)}(q^5 + q^4p + q^3p^2 + q^2p^3 + qp^4 + p^5)}{\cancel{q - p}}$

$= \lim_{p,q \to x} q^5 + q^4p + q^3p^2 + q^2p^3 + qp^4 + p^5$

$= 6x^5$

Exponentials

$(a^x)' = (e^{x\ln a})' = \ln a \cdot e^{x\ln a} \quad \therefore \boxed{(a^x)' = a^x \ln a}$

$$\boxed{\int a^x\,dx = \frac{a^x}{\ln a} + c}$$

Numerical Methods

BINOMIAL EXPANSION

$\binom{n}{0} = \binom{n}{n} = 1$; $\binom{n}{r} = \binom{n}{n-r}$; $\binom{k}{r-1} + \binom{k}{r} = \binom{k+1}{r}$

$$\boxed{(1+a)^n = 1 + an + \frac{a^2 n(n-1)}{2!} + \frac{a^3 n(n-1)(n-2)}{3!} + \dots}$$

Graphical Method – split into linear, quadratic (etc) graphs and find intersections

e.g. $e^x + x - 3 = 0 \Rightarrow e^x = 3 - x$ ∴ 1 root; $\ln x - x^2 + 4 \Rightarrow \ln x = x^2 - 4$ ∴ 2 roots

Change of Sign – to locate roots where $f(x) = 0$ when $f(x)$ is sufficiently well-behaved. find solution in interval (a, b) so work out $f(a)$ and $f(b)$

> "there is a root in the interval (a, b) as there is a change in sign and $f(x)$ is continuous in this interval"

Problems: → curve touches x-axis (repeated root) → several roots close together → discontinuity in $f(x)$ (vertical asymptotes, jump discont.)

Fixed Point Iteration

e.g. $y = x^3 - 3x + 1$, draw a sketch, pick x_0

$\Rightarrow 3x = x^3 + 1$ ∴ $x = \frac{1}{3}(x^3 + 1)$

$x_{n+1} = \frac{1}{3}(x_n^3 + 1)$

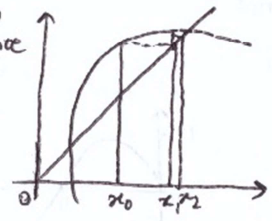
$\Rightarrow x^3 = 3x - 1$ ∴ $x = \sqrt[3]{3x-1}$

$x_{n+1} = \sqrt[3]{3x_n - 1}$

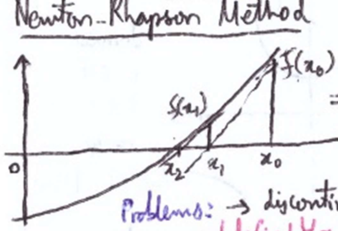
Staircase & Cobweb Diagrams

Problems → diverges from x_0 → wrong root convergence

Prove root to 2.d.p use change in sign w/ upper and lower bound



Newton-Raphson Method



∴ $f'(x_0) = \frac{f(x_0)}{x_0 - x_1} \Rightarrow x_0 - x_1 = \frac{f(x_0)}{f'(x_0)}$

$\Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$ ∴ $\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}}$

Problems: → discontinuous (not defined $\forall x \in \mathbb{R}$) → gradient (early) $f'(x) = 0$ ↔ → poor x_0 choice (approximates wrong root)

Converges of $x_{n+1} = g(x_n)$ to $x = a$ iff $|g'(a)| < 1$ and x_0 picked well.

Numerical Integration

⚠ use rectangles to work out upper and lower bounds for integrals.

Trapezium Rule → concave ∩ underestimate → convex ∪ overestimate

Area $= \frac{1}{2} \times h \times \sum$ parallel sides $= \frac{1}{2}h(y_1 + y_2) + \frac{1}{2}h(y_2 + y_3) + \dots$

$\boxed{= \frac{1}{2} \times \text{strip width} \times (\text{first} + \text{last} + 2 \cdot \text{middles})}$

Logarithm Laws $\boxed{\log_a(a^x) = x}$; $\boxed{a^{\log_a x} = x}$ Given $\boxed{a = b^c \iff c = \log_b a}$ $\boxed{\log_a 1 = 0}$ $\boxed{\log_a a = 1}$

$\boxed{\log_a(MN) = \log_a M + \log_a N}$ $\boxed{\log_a\left(\frac{M}{N}\right) = \log_a M - \log_a N}$ $\boxed{\log_a(M^p) = p\log_a M}$

$\boxed{\log_a(b) = \frac{\log_c(b)}{\log_c(a)}}$ $\boxed{\log_a(b) = \frac{1}{\log_b(a)}}$ $\boxed{\log_{(a^b)} c = \frac{\log_a c}{b}}$

Differentiation Laws $\boxed{(uv)' = u'v + uv'}$ $\boxed{\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}}$ $\boxed{(m^n)' = nm^{n-1}}$ $\boxed{\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}}$

Integration Laws $\boxed{\int \frac{g'(x)}{g(x)}\,dx = \ln|g(x)| + c}$ $\boxed{\int x^r\,dx = \frac{x^{r+1}}{r+1} + c}$ $\boxed{\int u'(u)^r\,dx = \frac{u^{r+1}}{r+1} + c}$

$\boxed{\int \sin x\,dx = -\cos x + c}$ $\boxed{\int \cos x\,dx = \sin x + c}$ $\boxed{\int \sec^2 x\,dx = \tan x + c}$ $\boxed{\int \tan x \sec x\,dx = \sec x + c}$

$\boxed{\int \csc^2 x\,dx = -\cot x + c}$ $\boxed{\int \cot x \csc x = -\csc x + c}$ $\boxed{\int_a^b uv'\,dx = uv\Big|_a^b - \int_a^b u'v\,dx}$

Sequences & Series

A.P $\boxed{S_n = \frac{n}{2}(a + l) = \frac{n}{2}(a + [a + (n-1)d])}$

G.P $\boxed{a_n = a_1 r^{n-1}}$ $S_n = a + ar + ar^2 + \dots + ar^{n-1}$

$rS_n = ar + ar^2 + ar^3 + \dots + ar^n$

$\boxed{S_\infty = \frac{a}{1-r}}$ when $|r| < 1$

$S_n(1-r) = a - ar^n = a(1 - r^n)$

$\boxed{S_n = \frac{a(1 - r^n)}{1-r}}$

(convergent, divergent, periodic, oscillating)

Convergent Values

Given a recurrence relation, e.g. $u_1 = 1$, $u_{n+1} = \frac{10 - 3u_n}{u_n}$

∵ $u_n = u_{n+1}$, $n \to \infty$

∴ $u = \frac{10 - 3u}{u}$

$\Rightarrow u^2 + 3u - 10 = 0$

$(u+5)(u-2) = 0$

∴ $u = -5$

Increasing & Decreasing

e.g. $u_n = 2n + 1$

$u_{n+1} = 2(n+1) + 1 = 2n + 3$

$u_{n+1} - u_n = 2 > 0$

∴ increasing sequence $\forall n \in \mathbb{N}$

MECHANICS

$\boxed{v = u + at}$ $\boxed{s = vt - \frac{1}{2}at^2 = ut + \frac{1}{2}at^2 = \frac{1}{2}(u+v)t}$ $\boxed{v^2 = u^2 + 2as}$

coalesce – join together

$\boxed{F = ma}$ $\boxed{F \le \mu R}$ $\boxed{\mu = \tan\theta}$ $\boxed{W = mg}$ $\boxed{\text{moment} = F \cdot d}$

$\boxed{\text{time of flight } T = \frac{2u\sin\theta}{g}}$ $\boxed{\text{horizontal range } R = \frac{u^2\sin 2\theta}{g}}$ $\boxed{\text{maximum height } h_{max} = \frac{u^2\sin^2\theta}{2g}}$ $\boxed{\text{trajectory equation } y = x\tan\theta - \frac{x^2 g}{2u^2\cos^2\theta}}$

Assumptions
→ no A.R.
→ negligible effect of curvature / Earth rotation

STATISTICS

Measures of Location – gives idea of size of numbers (median, mean)
Measures of Spread – idea of variability (s.d., range, variance, semi-interquartile range)

Sampling Techniques

- systematic sampling – items selected at regular intervals
  + representative + unlikely to be biased
  – data must be listed – practical consideration (data clumping)
- simple random sampling – every sample of a given size is equally likely to be selected
- random sampling – every item in a population has a non-zero probability of selection
- stratified sampling – population split into clear groups / strata. Random samples taken within, which may be proportional.
- cluster sampling – population is in clusters; which are selected (e.g. choosing several towns to represent one region)
- quota sampling – in social surveys, quotas on diff. types of people to interview
  + easy to collect – likely biased
- opportunity sampling – takes on easily available sample (e.g. delegates at a conference)
  + easy to collect – not representative, usually unreliable – only those who bothered filled in
- self-selecting sampling – participants volunteer to take part in survey

Data Processing · interpolate in data range though not explicitly data · extrapolate outside data range

- cleaning the data → exclude missing values / outliers
- use appropriate sampling technique
- represent in suitable diagrams
- calculate summary measures

Binomial Distribution

→ fixed number of independent trials → two outcomes $(p + q = 1)$ → constant p (equal likelihood of success)

If $X \sim B(n, p)$, then $P(X = x) = \boxed{\binom{n}{x} p^x (1-p)^{n-x}}$, $E(X) = \boxed{np}$; $Var(X) = \boxed{np(1-p)}$

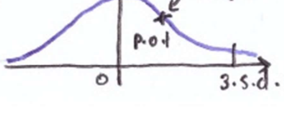
Normal Distribution

→ continuous w/ symmetry → discrete w/ small intervals vs. σ → continuous corrections → $P(x) = 0$ (no area)

If $Y \sim N(\mu, \sigma^2)$, then $P(Y = y) = \boxed{f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)}$, $E(X) = \boxed{\mu}$, $Var(X) = \boxed{\sigma^2}$

When standardised, $\left(\frac{x-\mu}{\sigma}\right)^2 = x^2$

PROBLEMS:
→ lack of symmetry: skewed, truncated, lack of tails
→ 3.s.d is unrealistic:



KEY VALUES

1.s.d. $P(-1 < Z < 1) = 68\%$
2.s.d. $P(-2 < Z < 2) = 95\%$ = outliers
3.s.d. $P(-3 < Z < 3) = 99.7\%$

Test Statistic: a statistic that is calculated from sample data in order to test a hypothesis about a population

Bivariate Data: data covers two variables in scatter graph

Random Variable: sum of probabilities of all possibilities = 1
discrete: can only take certain values
continuous: can take any value in a given range

Box Plot Advice ⚠
- 1.5×IQR outside it = outlier
- use a scale
- compare IQR, median, outliers

Linear Transformations

$\to aX + b = Y \Rightarrow \boxed{\bar{y} = a\bar{x} + b}$ and $\boxed{\sigma_y = a\sigma_x}$

For standardisation, $X \sim N(\mu, \sigma^2) \Rightarrow X \sim N(0, 1)$:

$a = \frac{1}{\sigma}$, $b = -\frac{\mu}{\sigma} \Rightarrow z = \frac{1}{\sigma}\bar{x} - \frac{\mu}{\sigma} \Rightarrow \boxed{z = \frac{x - \mu}{\sigma}}$

$\boxed{Var(X) = \frac{\sum(x - \mu)^2}{n}}$ $\boxed{\sigma = \sqrt{Var(X)} = \sqrt{\frac{\sum x^2}{n} - \mu^2}}$

Approximating the Binomial Distribution w/ Normal $\boxed{X \sim N(np, np(1-p))}$ ⚠ Use continuity correction beware < and ≤!

→ n is large $(n > 50)$ → p is close to $\frac{1}{2}$: $(0.1 < p < 0.9)$

Reversing Normal Calculations

$\boxed{x = \text{InvNorm}(p, \mu, \sigma)}$

Use Probabilities to find μ and σ

$P(X < x) = 0.8 \Rightarrow P(Z < z) = 0.8$

∴ InvNorm$(0.8) \Rightarrow z = 0.8416$

∴ use $z = \frac{x - \mu}{\sigma}$ to find missing μ or σ.

or use simultaneous eq. to find μ and σ for 2 $P(X = x)$'s.

Testing a Mean Using Test Statistic

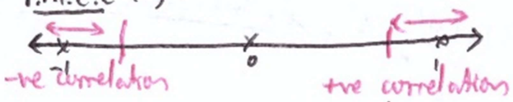
$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$ $\boxed{C.V = \text{InvNorm}(\alpha)}$

Standardised: $\boxed{z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}}$

Testing a Mean using Probability

w.t.f. $P(X = x)$, so use $P(X > x)$ or $P(X < x)$

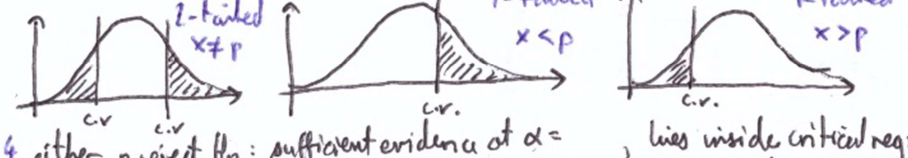
P.M.C.C (r)



$H_0: r = 0$, $H_1: r \ne 0$ so two tailed or $r > 0$ or $r < 0$ (one-tailed)

HYPOTHESIS TESTING

1. define H_0 and H_1 (e.g. $x \ne 0$ or $x > 0$ or $x < 0$), and let X be the r.v "..."
2. model X as a distribution, state sig. level ($\alpha = ..$) and one / two – tailed
3. calculate relevant values, compared to c.v.



4. either
   - reject H_0: sufficient evidence at α = , lies inside critical region
   - accept H_0: insufficient evidence at α = ; lies outside critical region

FURTHER STATISTICS

Probability

$^nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$ $\quad {}^nP_r = \frac{n!}{(n-r)!}$

Discrete Random Variables

X is the r.v. taking values x_i in a discrete distribution w/ $P(X = x_i) = p_i$

$\boxed{\mu = E(X) = \sum x_i p_i}$ $\boxed{\sigma^2 = Var(X) = \sum (x_i - \mu)^2 p_i = \sum x_i^2 p_i - \mu^2}$

w.t.p. $\sum (x_i - \mu)^2 = \sum x_i^2 p_i - \mu^2$ N.B: μ is a constant $\sum p_i = 1$

L.H.S $= \sum (x_i^2 - 2\mu x_i + \mu^2) p_i = \sum x_i^2 p_i - 2\mu \sum x_i p_i + \mu^2 \sum p_i$

$= \sum x_i^2 p_i - 2\mu^2 + \mu^2 = \sum x_i^2 p_i - \mu^2$ □

Linear Transformations

square of the s.d.

$\boxed{E(aX+b) = aE(X) + b}$ $\boxed{Var(aX+b) = a^2 Var(X)}$

Binomial Distribution where $X \sim B(n,p)$

→ fixed # of independent trials
→ constant p of success
→ two outcomes $(p+q=1)$

$\boxed{E(X) = np}$ $\boxed{Var(X) = np(1-p)}$ $\boxed{P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}}$

w.t.p. $E(X) = np$

$E(X) = \sum x_i p_i = 0 \cdot \binom{n}{0} \cdot p^0 (1-p)^n + 1 \cdot \binom{n}{1} p^1 (1-p)^{n-1} + 2 \cdot \binom{n}{2} p^2 (1-p)^{n-2} + \dots$

$\dots + (n-1) \cdot \binom{n}{n-1} p^{n-1} (1-p) + n \cdot \binom{n}{n} p^n (1-p)^0$

$= 0 + np(1-p)^{n-1} + \frac{2n(n-1)}{2!} p^2 (1-p)^{n-2} + \frac{3n(n-1)(n-2)}{3!} + \dots$

$\dots + (n-1)np^{n-1}(1-p) + np^n$

$= np\left[(1-p)^{n-1} + (n-1)p(1-p)^{n-2} + \frac{(n-1)(n-2)}{2!} p^2 (1-p)^{n-3} + \dots \right.$

$\left. \dots + (n-1)p^{n-2} + p^{n-1}\right]$

$= [(1-p) + p]^{n-1} \times np = np$ □

Discrete Uniform Distribution where $X \sim U(n)$ over $1, 2, \dots, n$

$\boxed{E(X) = \frac{1}{2}(n+1)}$ $\boxed{Var(X) = \frac{1}{12}(n^2-1)}$ $\boxed{P(X=x) = \frac{1}{n}}$

w.t.p. $E(X) = \frac{1}{2}(n+1)$

$E(X) = \sum x_i p_i = 1 \times \frac{1}{n} + 2 \times \frac{1}{n} + \dots + n \times \frac{1}{n} = \sum r \times \frac{1}{n} = \left[\frac{n(n+1)}{2}\right] \times \frac{1}{n} = \frac{1}{2}(n+1)$ □

w.t.p. $Var(X) = \frac{1}{12}(n^2-1)$

$Var(X) = E(X^2) - [E(X)]^2 = \sum x_i^2 p_i - \left(\frac{n+1}{2}\right)^2 = \frac{1}{n}(1^2 + 2^2 + \dots + n^2) - \frac{1}{4}(n+1)^2$

$= \frac{1}{n} \times \frac{n}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 = \frac{n+1}{12}[2(2n+1) - 3(n+1)] = \frac{(n+1)(n-1)}{12} = \frac{1}{12}(n^2-1)$ □

Geometric Distribution where $X \sim Geo(p)$

→ no upper limit to # of trials
→ constant p of success for each trial
→ independent trials, 2 outcomes

$\boxed{E(X) = \frac{1}{p}}$ $\boxed{Var(X) = \frac{1-p}{p^2}}$ $\boxed{P(X=x) = p(1-p)^{x-1}}$

w.t.p. $E(X) = \frac{1}{p}$

$E(X) = 1 \times p + 2 \times (1-p)p + 3 \times (1-p)^2 p + \dots$

$(1-p)[E(X)] = \quad 1 \times (1-p)p + 2 \times (1-p)^2 p + \dots$

$E(X)[1-(1-p)] = p + p(1-p) + p(1-p)^2 + \dots$

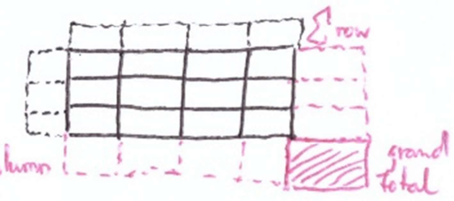
$\therefore S_\infty = \frac{a}{1-r} = \frac{p}{1-(1-p)} = 1 \Rightarrow pE(X) = 1 \quad \therefore E(X) = \frac{1}{p}$ □

Chi-Squared Tests (χ^2)

Expected Frequency $= \boxed{E_i = \frac{\text{row total} \times \text{column total}}{\text{grand total}}}$

Contingency Tables

with m rows and n columns
⇒ m×n contingency tables



Add E_i into cell i in subscript. Then:

$\boxed{\chi^2_{calc} = \sum \frac{(O_i - E_i)^2}{E_i} = \sum \frac{O_i^2}{E_i} - N}$ grand total

Degrees of Freedom (ν)

how many values required, given $\sum$s, to fill out the rest of the table

$\boxed{\nu = (n-1)(m-1)}$

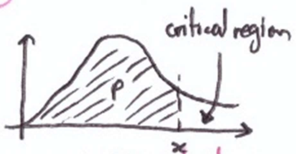
when $m = 1$, $\nu = n - 1$

Yates' Correction when $\nu = 1$ in contingency tables

$\boxed{\chi^2_{Yates} = \sum \frac{(|O_i - E_i| - 0.5)^2}{E_i}}$

We need all $E_i \geq 5$, so combine rows/columns accordingly.

Chi-Squared Hypothesis Testing



1 H_0: X is independent of Y
H_1: X is not independent of Y

2 Work out χ^2_{calc} (or χ^2_{Yates}).

← independent 10% sig. level dependent →

p	0.01	0.025	0.90	0.95	...
$\nu = 1$					
2					
3					
...					

3 Find χ_ν^2 from c.v. table.

4 if $\chi^2_{calc} > \chi_\nu^2$, reject H_0, sufficient evidence of association
$\chi^2_{calc} < \chi_\nu^2$, accept H_0, insufficient evidence

Testing Suitability of a Model

1 H_0: a $B(n,p)$ distribution is suitable
H_1: distribution is not suitable

N = ☐ α = ☐

2

x					...
$p(x)$					...
E_i					...

3 $\chi^2_{calc} = \sum \frac{O_i^2}{E_i} - N$ or equivalent

4 Conclude whether model is suitable

Estimating Parameters – removes a degree of freedom

e.g. **BINOMIAL**

$E(X) = np = \bar{x} = \frac{\sum (x \cdot f)}{N}$ value is O_i N ← total freq.

$\Rightarrow p = \frac{\sum x \times f}{Nn}$

GEOMETRIC

$E(X) = \frac{1}{p} = \bar{x} \Rightarrow p = \frac{1}{\bar{x}}$

Pearson's Product Moment Correlation Coefficient (P.M.C.C)

Correlation does not imply causation.

$\boxed{S_{xx} = \sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n}}$ $\boxed{S_{xy} = \sum (x-\bar{x})(y-\bar{y}) = \sum xy - \frac{\sum x \sum y}{n}}$ $\boxed{S_{yy} = \sum (y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n}}$

$\boxed{r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{xy}}} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}}$

Effect of Coding on P.M.C.C:

None, as it should be. $r = \frac{a \times b}{\sqrt{a^2 \times b^2}} = 1$