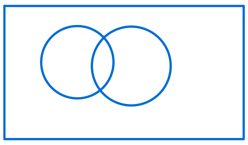
1. probability

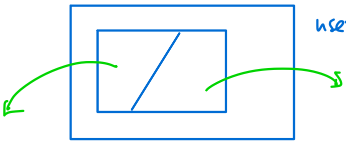
① properties of operations of event

1. $A \cap A' = \phi \qquad A \cup A' = S$
2. $(A')' = A$
3. $A \cup \phi = A \qquad A \cap \phi = \phi$
4. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
5. $A \cup B = A \cup (B \cap A')$
6. $A = (A \cap B) \cup (A \cap B')$
7. $(A_1 \cup A_2 \cup A_3 \ldots)' = A_1' \cap A_2' \ldots$
   $(A_1 \cap A_2 \cap A_3 \ldots)' = A_1' \cup A_2' \ldots$

② Basic properties of probability

1) $P(\phi) = 0$
2) $P(A') = 1 - P(A)$
3) $P(A) = P(A \cap B) + P(A \cap B')$





useful application to sample points and counting
⇒ split sample space and use considerations to count

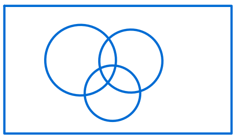
when even no. of events, you add too many $P(A_i)$ ⇒ double counted, need to remove eg. 2 event case

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

5) $P(A_1 \cup A_2 \ldots) = \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P(A_1 \cap \ldots A_n)$

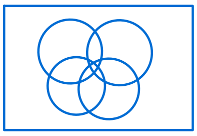
↳ inclusion – exclusion principle

visualise as for loops to achieve all possible combination



odd

vs



even

add individual – pairwise (but now middle gone) + add back middle

add individual – pairwise + add back missing triple overlaps – remove additional 4 overlap

axioms

1. $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. if events are mutually exclusive, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

5 conditional probability

↳ the probability of an event occuring on the condition that another has occured.

⇒ since event B has occured, instead of considering the set of all possible outcomes we consider the set of sample points for event B

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ assuming } P(B) \neq 0$$

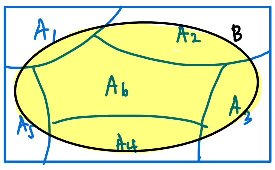
1) multiplication rule of probability

$P(A \cap B) = P(A) \cdot P(B/A)$ or $P(B) \cdot P(A|B)$ but not both (double counting)

$$\Rightarrow P(A_1 \cap A_2 \dots A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap \dots A_{n-1})$$

↳ useful if given a lot of probabilities

2) law of total probability



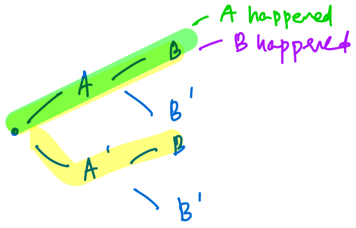
$A_1 \dots A_n$ are mutually exclusive

$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)$ ← notice you can just sub for $P(A_i \cap B)$

intuition: what is the total probability of B occuring, if there are all the ways $A_1 \dots A_n$ with which it could happen?

3) bayes' theorem

↳ intuition: if B happened, what are the ways in which A could have happened?



$$P(A|B) = \frac{\overbrace{P(A)\ P(B|A)}^{\text{A happening given B happened}}}{P(B)}$$

notice can sub for $P(A \cap B)$

---

6 independent events

cannot be checked by venn diagram, because venn only shows if $\neq 0$ or $= 0$, not magnitude

↳ whether the knowledge that an event has occured changes your knowledge about an event (its probability)

Two events are statistically independent iff $P(A \cap B) = P(A) \times P(B)$

1) properties

1. conditional probability is same $P(A) = P(A|B)$
2. Not mutually exclusive. $P(A \cap B) = P(A) \times P(B) > 0 \Rightarrow P(A), P(B) \neq 0$
3. S and ϕ are independent of any event, because $P(S) = 1$ and $P(\phi) = 0$
4. if $A \subset B$, $A \cap B = A$. so A and B are dependent unless $B = S$.

$P(A \cap B) = P(A) \neq P(A) \times P(B)$

2) mutual exclusivity and independence

mutually exclusive $\longrightarrow$ not independent $\equiv$ independent $\longrightarrow$ not mutually exclusive

not mutually exclusive $\longrightarrow$ may or may not be independent

3) pairwise and mutual independence

pairwise independence

a set of events $A_1 \dots A_n$ are pairwise independent iff $P(A_i \cap A_j) = P(A_i) \times P(A_j)$ for $i \neq j$ and $i, j = 1, 2 \dots n$

$\Rightarrow$ each pair of events has an intersection

- ↳ events are individually independent
- ↳ events are individually not mutually exclusive
- ↳ for any pair of events, multiplication rule holds

mutual independence

a set of events $A_1 \dots A_n$ are mutually independent iff $P(A \cap \dots A_n)$

$= P(A_1) \times P(A_2) \dots P(A_n)$

$\Rightarrow$ there is an intersection of all events

- ↳ events are collectively independent
- ↳ events are collectively not mutually exclusive
- ↳ for any n events, multiplication rule holds

mutual independence $\longrightarrow$ pairwise independence

pairwise independence $\not\longrightarrow$ mutual independence

---

2. Counting

permutations

$${}^nP_n = n(n-1)\dots = n!$$

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

in a circle (or where no reference) $N = (n-1)!$

$$N = \frac{n!}{n_1!\, n_2! \dots}$$

no. of identical objects

combinations

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

1. ${}^nC_r = {}^nC_{n-r}$; $\binom{n}{r} = \binom{n}{n-r}$
2. ${}^nC_r = {}^{n-1}C_r + {}^{n-1}C_{r-1}$ for $1 \leq r \leq n$
3. ${}^nC_0 = {}^nC_n = 1$
4. ${}^nC_{n-1} = {}^nC_1 = n$

---

multisets

$$N = {}^{n+r-1}C_r$$

3. Random variables

① Discrete and continuous

	Discrete random variables	Continuous random variables
distribution	$f(x) = F(n) - F(n-1)$ $F(x) = \sum_{t=0}^{x} f(t)$	$f(x) = \frac{d}{dx} F(x)$ $F(x) = \int_0^{x} f(t)\, dt$ take note of diff symbols
probability	$P(A) = P(X_1) + P(X_2) + P(X_3)$ $P(a \leq X \leq b) = P(X \leq b) - P(X < a)$ $= F(b) - F(a-1)$	$P(A) = \int_b^a f(x)\, dx$ $P(a \leq X \leq b) = P(X \leq b) - P(X < a)$ $= F(b) - F(a)$
Expectation	$E(X) = \sum_x x f(x)$ $E(g(x)) = \sum_x g(x) f(x)$	$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$ $E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$ value probability
variance	$E(X^2) = \sum_x x^2 \cdot f(x)$ $V(X) = \sum_x (x-\mu_x)^2 f(x)$ $= E(X^2) - E(X)^2$	$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx$ $V(X) = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\, dx$ $= E(X^2) - E(X)^2$

② E(X), V(X), n moment

Expectation

$$E(aX+b) = aE(X) + b$$

cumulative

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) \quad \text{for } x \in \mathbb{Z}_{\geq 1}$$

n moment

$g(x) = x^k$, $E(g(x))$ is called the kth moment of X

Variance

$$Var(X) = \sigma_X^2 = E((X-\mu_X)^2)$$
$$= \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\, dx$$
$$= \sum_x (x-\mu_x)^2 f(x)$$
$$= E(X^2) - [E(X)]^2$$

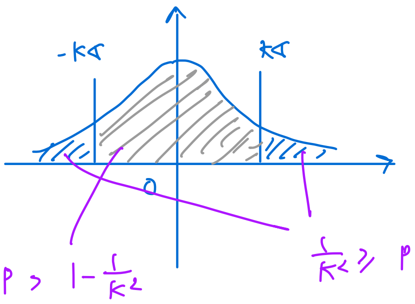
piecewise if necessary in computation to get value

③ chebyshev inequality

remember that this is two sided.

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P(|X-\mu| < k\sigma) \geq 1-\frac{1}{k^2}$$



---

4. 2 dim Random Variables

probability

1. $f_{x,y}(x,y) \geq 0$ for all x,y & total probability $=1$

2. $P(X,Y) = \sum\sum_{(x,y)\in A} f_{x,y} = \iint_A f_{x,y}\, dA$

can derive if discrete

eg. $f_{xy} = \frac{{}^4C_x \cdot {}^3C_y \cdot {}^2C_{3-x-y}}{{}^9C_3}$

ways of interest

total no. of ways

marginal distribution

$$f_X(x) = \sum_y f_{x,y} = \int_{-\infty}^{\infty} f_{x,y}\, dy$$

x	y			$f_X(x)$
	1	3	5	
2	0.1	0.2	0.1	0.4
4	0.15	0.3	0.15	0.6
$f_Y(y)$	0.25	0.5	0.25	1

conditional

marginal

conditional distribution

$$f_{Y|X} = \frac{f(x,y)}{f(x)} \qquad f(x,y) = f_{Y|X} \cdot f(x)$$

1. find $f(x)$ → discrete (manual) / → continuous (integrate)

2. find $\frac{f(x,y)}{f(x)}$ at each interval → discrete (manual decision) / → continuous (manipulation)

independence

RVs are independent iff $f(x,y) = f(x) \cdot f(y)$

≥ 0 ≥ 0 ≥ 0

biconditional; so knowledge allow you to construct joint from marginal expectation

can think of it still as weighted sum of some function involving both variables, effectively same, just more options / fidelity, more need to collapse area

eg. $E(X) = \iint_A x f(x,y)\, dx = \int x \int f(x,y)\, dy$

$= \int x f(x)\, dx \Rightarrow$ same

$E(g(X,Y)) = \iint_A g(x,y) f(x,y)\, dA$

covariance of X and Y

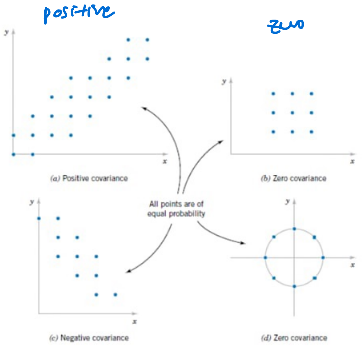
$Cov(X,Y) = E((X-\mu_x)(Y-\mu_y))$

$= \iint \underbrace{\underbrace{(x-\mu_x)}_{\text{magnitude \& direction of deviation}}\underbrace{(y-\mu_y)}_{\text{magnitude \& direction of deviation}}}_{\text{magnitude \& direction of how they vary (from global mean) w.r.t. each other}} f(x,y)\, dA$

weighted sum

positive

zero



negative

zero

1. $Cov(X,Y) = E(XY) - \mu_x \mu_y$
2. $Cov(aX+b, cY+d) = ac\, Cov(X,Y)$
3. if X and Y are independent $\to Cov(X,Y) = 0$

not converse.

variance of aX + bY

$E(aX+bY) = aE(X) + bE(Y) \ldots$

$V(aX_1 + bX_2 + \ldots)$

$= \underbrace{a^2 V(X_1)}_{} + \underbrace{b^2 V(X_2)}_{} \ldots + \underbrace{2ab\, Cov(X_1, X_2)}_{} \ldots$

individual contributor to variance of final RV

"mutual" interactions of variances

correlation

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

absolute variation from mean of x and y

absolute variations from individual mean

variance from mean together as a ratio to that of individual total deviation

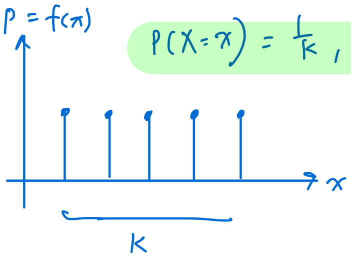
5. Special probability distributions

(Parameter & family of distributions) : if $f(x)$ depends on parameters, then the collection of all probability distributions for all possible parameter values is called the family of distributions.

① Discrete distributions

(Uniform distribution)

$$P(X=x) = \frac{1}{k}, \quad x = x_1, x_2 \ldots$$

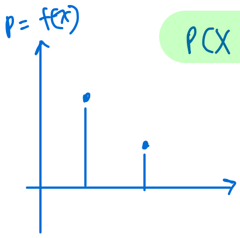


$$E(X) = \frac{1}{k}\sum_{i=1}^{k} x_i$$

$$V(X) = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu^2$$

(Bernoulli distribution)

↳ X : a binary random variable, only can take success / failure ; 1/0



$$P(X=x) = p^x(1-p)^{1-x}, \quad x = 1, 0$$

↳ so $f(x) = p$ or $1-p$

$$E(X) = p$$

$$V(X) = p(1-p)$$

(Binomial distribution)

↳ X : the number of successes in n independent Bernoulli trials. $X = Y_1 + Y_2 \ldots$ (binomial; n bernoulli trials)

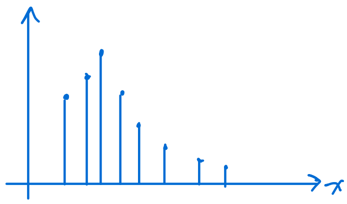
↳ conditions : n identical, independent Bernoulli trials (identical → constant p; → binary outcome)

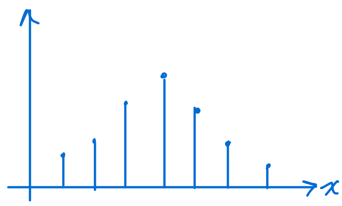
$$P(X=x) = {}^nC_x\, p^x(1-p)^{n-x}$$

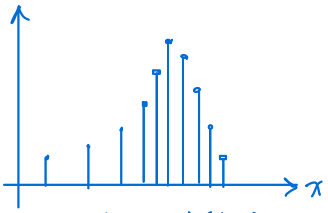
$$E(X) = np$$

$$V(X) = np(1-p)$$

$$X \sim B(n,p)$$







↳ (normal approximation to binomial) & continuity correction

$X \sim N(np, npq)$ approximately. As $n \to \infty$ and $p \to \frac{1}{2}$, binomial goes to normal.

$np > 5$ and $n(1-p) > 5$.

(Negative binomial distribution)

↳ X : the number of trials to get the k^{th} success in a set of independent, identical (same p) Bernoulli trials

$$X \sim NB(k,p)$$

$$E(X) = \frac{k}{p}$$

$$V(X) = \frac{(1-p)k}{p^2}$$

$$P(X=x) = {}^{x-1}C_{k-1}\, p^k (1-p)^{x-k}$$

(${}^{x-1}C_{k-1}$: choosing for non-final; p^k: successes; $(1-p)^{x-k}$: failures)

↳ (geometric distribution) $X \sim NB(1,p)$: no. of trials to get the first success

Poisson distribution

↳ X : the number of successes occurring over a continuous domain of ∞ bernoulli trials

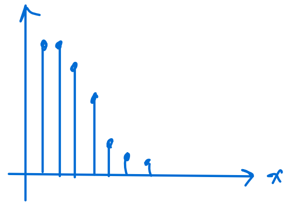
↳ characteristics :
1. no. of success in a cts. interval is independent of that of a disjoint interval
2. probability of a single success in an interval is proportional to the size and is independent of events outside the interval
3. probability of >1 success in a very small interval is negligible

$X \sim P(\lambda)$ $\quad P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2 \ldots$

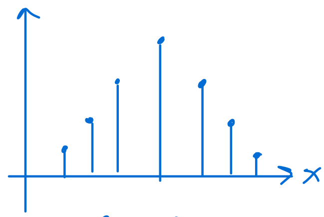
$E(X) = \lambda$

$V(X) = \lambda$

↳ λ is the average no. of success over the given interval



λ is small
⇒ right skew



λ very large
⇒ symmetric

↳ derivation :

assume probability of a bernoulli experiment is constant. → want to find no. of success over a specified continuous domain → infinite number of Bernoulli trials needed $n \to \infty$ → but in increasing resolution, average / expectation should not change. $p \to 0$

Binomial $\bar{c}$ $n \to \infty$, $p \to 0$

1. Define average rate over some interval

$$\lambda = np \Rightarrow p = \frac{\lambda}{n}.$$

$$\lim_{n\to\infty} P(X=x) = \lim_{n\to\infty} \frac{n!}{x!\,(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{1-x}$$

2. By simplifying & using $e = \lim_{\lambda\to\infty}\left(1+\frac{1}{\lambda}\right)^\lambda$ and taking the limit, we get :

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

→ we essentially increase the no. of Bernoulli trials infinitely but encapsulate its behavior in average success, λ

poisson approximation to binomial

↳ poisson distribution is the asymptotic distribution of binomial as $n \to \infty$ and $p \to 0$ s.t. np is constant

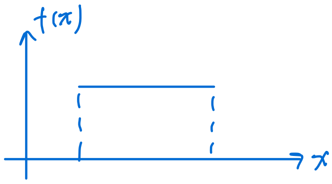
② Continuous distributions

continuous uniform distribution

↳ rectangular distribution

$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$

$E(X) = \frac{a+b}{2}$ $\quad V(X) = \frac{(b-a)^2}{12}$



exponential distribution

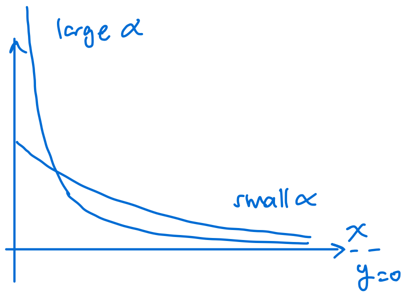
$f(x) = \alpha e^{-\alpha x}, \quad x > 0.$

$E(X) = \frac{1}{\alpha}$ $\quad \alpha > 0$

$V(X) = \frac{1}{\alpha^2}$ $\quad f(x) > 0$

$CDF = P(X \leq x) = 1 - e^{-\alpha x}$

$P(X > x) = e^{-\alpha x}$



↳ no memory property :

$P(X > s+t \mid X > s) = P(X > t)$

probability that it will last addt'l t minutes given already last s, same as probability of a new piece lasting s

normal distribution

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

$E(X) = \mu$

$V(X) = \sigma^2$

$X \sim N(\mu, \sigma^2)$ $\quad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

chi-square χ^2 distribution

Z^2

degrees of freedom

$f(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} \cdot e^{-y/2}$

gamma function, $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$

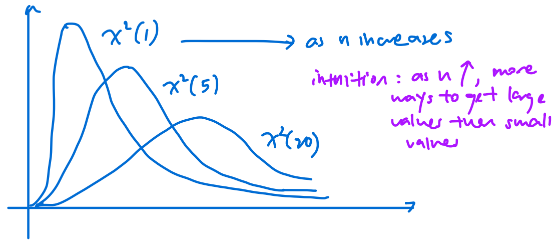
$= (n-1)!$

$E(\chi^2) = n$ $\quad V(\chi^2) = 2n$

$\chi^2(n) = Z_1^2 + Z_2^2 \ldots Z_n^2, \quad Z \sim N(0,1)$

↳ χ^2 distribution is the distribution of sum of n independent, identical normal random variable.

$n \to \infty, \quad \chi^2(n) \sim N(n, 2n)$ approx.



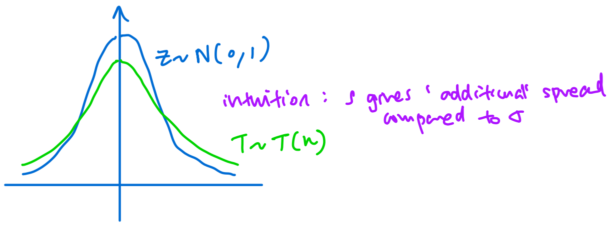
t distribution

$T = \frac{Z}{\frac{\chi^2(n)}{n}} \sim T(n) \Rightarrow T \text{ w/ } n \text{ DOF}$

standard normal

chi-square w/ n DOF

$n \to \infty, \quad T \sim N(0,1)$ approx.



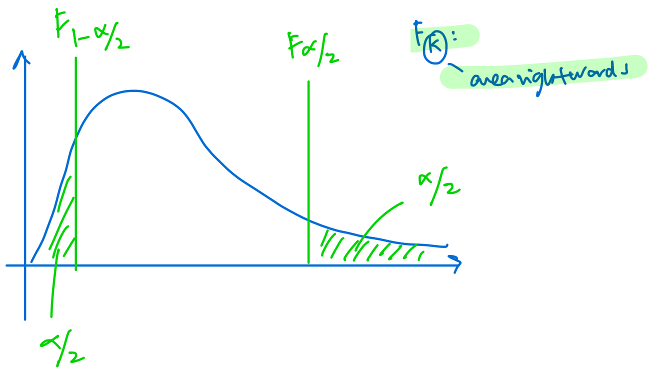
F distribution

T^2

$F = \frac{\chi_1^2(n_1)}{n_1} \Big/ \frac{\chi_2^2(n_2)}{n_2} \sim F(n_1, n_2)$

$F \sim F(n_1, n_2), \quad \frac{1}{F} \sim F(n_2, n_1)$

n_1 & $n_2 \to \infty, \quad F \sim N$ approximately



6. sampling and testing

① population & sample

a collection of units of interest (not necessarily their values)
eg. all customers who use credit cards

1) finite and infinite populations

finite no. of units

countably / uncountably infinite no. of units

2) probability sampling

from a finite population

without replacement
↳ there are ${}^N C_n$ samples of size n. Each has $\frac{1}{{}^N C_n}$ chance of being selected

with replacement
↳ there are N^n samples of size n, each with $\frac{1}{N^n}$ chance of selection

also equivalent to sampling from a (diff.) infinite population (all possible obs. ever)

from an infinite population

↳ take n independent observations, then random sample is: $(X_1, X_2 \ldots X_n)$ with joint distribution $f_{X_1, X_2 \ldots X_n}(x_1, x_2 \ldots x_n)$

use for

② Statistics, estimators and estimates

1) Statistics: a function of a random sample which does not depend on any unknown parameters

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

$W = \frac{1}{n}\sum_{i=1}^{n} X_i - \mu$ ✗

2) estimators and estimates

statistic, non-concrete, to estimate population parameters

concrete estimations in realized values

3) estimator bias & variance for some estimator $\hat{\theta}$

accuracy of estimator

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

precision of estimator

$var(\hat{\theta}) = E(\hat{\theta}^2) - E(\theta)^2$

MSE = bias² + estimator variance

4) point and interval estimates

$\hat{\theta}$, some statistic that estimates θ

$\hat{\theta}_L < \theta < \hat{\theta}_R$, two 'bounds' estimators that estimate interval in which θ lies

③ Estimating w confidence → probability interval (analytical) / confidence interval (constructed)

↳ the calculated interval, calculated from this method, will contain the true population parameter C% of the time. ie. if repeated large no. of times and each time C% CI constructed, C% of them will have θ

↳ Not that the calculated interval has a C% of containing interval, because it either does or does not

---

④ common statistics

sample means of normal (approx.) population

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

1) law of large numbers:

$$P(|\bar{X} - \mu| > \varepsilon) \to 0 \text{ as } n \to \infty$$

2) central limit theorem:

$X_1, X_2 \dots X_n$ is n independent observations.

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ approx. when $n > 30$.

↳ if X is normal, then $\bar{X}$ will be exactly normal.

3) distribution $P(X - m < \mu < X + m)$

known variance:

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

$$m = \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

unknown variance:

$$T_{n-1} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$m = \pm t_{n-1, \alpha/2} \left(\frac{s}{\sqrt{n}}\right)$$

if $n > 30$

Estimator of variance of a normal population

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

↳ distribution

known μ:

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

$$\sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_{n;\alpha/2}}} < \sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\chi^2_{n;1-\alpha/2}}}$$

unknown μ:

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}} < \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}}$$

α if single tail

ratio of sample variance two variances of normal population w unknown mean

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, n-2)$$

$H_0: \sigma_1^2 = \sigma_2^2$ $\quad F = \frac{S_1^2}{S_2^2}$

$\frac{\sigma_1^2}{\sigma_2^2}$ → >1, $\neq 1$, <1

$$\sqrt{\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1, n_2-1, \alpha/2}}} < \frac{\sigma_1}{\sigma_2} < \sqrt{\frac{S_1^2}{S_2^2} F_{n_2-1, n_1-1, \alpha/2}}$$

change to α for two sided test

↳ $= F_{n_2-1, n_1-1, 1-\alpha/2}$

---

difference in sample means Assume underlying normal

1. σ_1, σ_2 are known and not equal, population (approximately) normal, independent

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

mean; → sd

$$m = \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

2. σ_1, σ_2 are unknown but large sample. just replace σ_1, σ_2 w S_1 and S_2. $n_1, n_2 > 30$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

$$m = \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

3. σ_1, σ_2 are known and equal, distribution (approximately) normal

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$m = \pm z_{\alpha/2}\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

4. σ_1, σ_2 are unknown but equal, population normal but $n_1, n_2 \le 30$

$$T_{n_1+n_2-2} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$m = \pm t_{n_1+n_2-2;\alpha/2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$, so $\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$

5. paired (dependent) data that is normal distributed, μ_D and unknown σ_D^2

↳ do as if mean ω unknown σ, calculate each data point as $x_i - y_i$

$$\bar{D} = X_{after} - Y_{before} \sim N\left(\mu_D, \frac{\sigma_D^2}{n}\right)$$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \qquad T_{n-1} = \frac{\bar{d} - \mu_D}{\frac{s_d}{\sqrt{n}}} \qquad m = \pm t_{n-1, \alpha/2}\left(\frac{s_D}{\sqrt{n}}\right)$$

$$s_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

↳ if $n > 30$, use $z_{\alpha/2}$

⑤ sample size for estimation

$$P(|\bar{x} - \mu| \leq error) \geq 1 - \alpha$$

↳ we want an error smaller ω a probability of at least $1-\alpha$ 　 $error \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ 　 $error \geq m$

$$P\left(|\bar{X} - \mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

⑥ Hypothesis testing

Significance tests
A formal procedure for comparing observed data with a hypothesis whose truth we want to assess. The results of the test are expressed in terms of a probability that measures how well the data and the null hypothesis (what you want to disprove) agree. Intuition Finding probability of event occurring by random chance if certain premises are true, from which you can falsify the premise (or not) – because of its low probability of occurrence. Hypothesis testing is a method of quantifying likelihood of an event to falsify something you want to falsify. It cannot prove something. Checks if the observed effect is *not* compatible with H_0, or compatibility with H_1, since you can only reject H_0 or fail to reject it. Key terms • **Null hypothesis H_0:** proposes that there is no difference between certain characteristics of a population, eg. μ = some k. Often the value you want to disprove. • **Alternative hypothesis H_1:** proposes that true population parameter is not k (>, <, ≠), determines contextual interpretation of p-value. • **Test statistic:** estimate parameter that appears in hypothesis under the assumption that H_0 is true (ie. using its numbers for computation). • **P-value:** probability of getting an outcome (test statistic) as extreme or more extreme than the observed outcome, in the context of the null hypothesis. Ie. quantifying the likelihood of getting that observation of the *random* variable or something more extreme than it by chance variation given that H_0 is true • **Level of significance:** probability of rejecting H_0 wrongly given that H_0 is true. Probability of type I error. • **Statistical significance:** if p-value is smaller than level of significance, we conclude that the result is statistically significant and reject H_0, as we believe that the likelihood of getting this observation by chance is too small. Procedure 1. State H_0 and H_1 2. Calculate standardised test statistic 3. Find p-value and compare with level of significance

Two-sided significance tests and confidence intervals
A level α two-sided significance test rejects a null hypothesis exactly when the sample mean falls outside a level 1- α confidence interval and population is normally distributed. Else approximately. Mathematically they mean the same thing, so you can use CI to do hypothesis testing more directly.

Cautions in hypothesis testing
1. Statistical significance is an opinion – there is no definitive line between significant and insignificant – only increasingly strong evidence as p-value decreases. 2. Statistical significance does not mean meaningful significance: e.g. there is strong evidence of some association (e.g. r ≠ 0), but r = 0.01, so the association exists, but may not meaningfully have any impact on what you do with that information. 3. Lack of statistical significance does not mean that evidence that H_0 (no effect) is true: when a p-value falls within the confidence interval, it suggests that there is not enough evidence to *reject* H_0, but it also does not mean there is evidence to *accept* H_0. Further experiments should be conducted, often with larger samples. E.g. incident rate ratio between control and experimental groups is 1.00, confidence interval is 0.63 to 1.58 – true value might be anywhere between 0.63 to 1.58 or not at all, rather than 1.00 (no effect at all). 4. Meaningful but statistically insignificant results may be used as pilot studies. E.g. p-value = 0.1, but correlation = 0.7 5. Formal statistical inference is not valid if sampling is biased or haphazard – e.g. presence of confounding, biased samples etc.

② Power of tests

Type I and Type II errors

Type I: wrongly rejecting H0 when H0 is true. Hypothesis testing controls for this error through the level of significance – which is the probability of wrongly rejecting H0 when H0 is true, since the model is built assuming H0 is true. Level of significance controls for this.

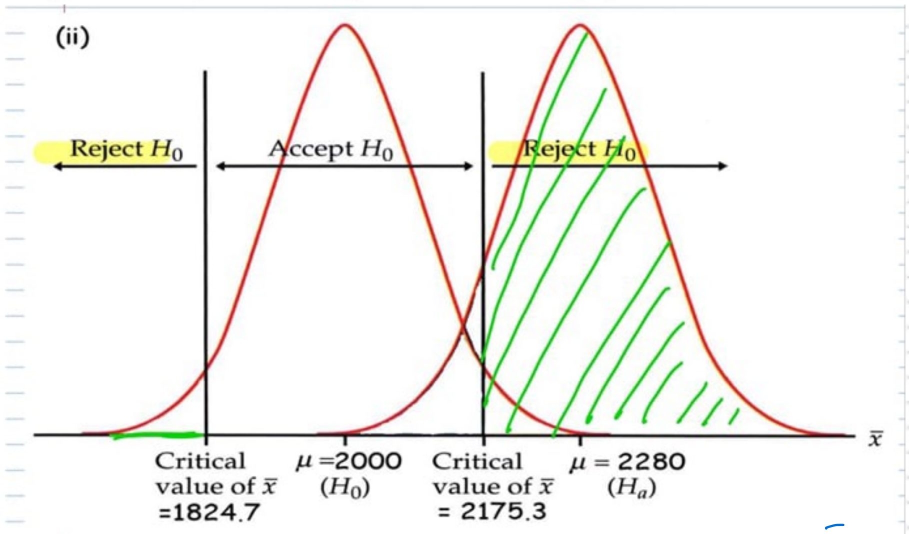
Type II: wrongly accepting H0 when H1 is true. If this occurs, it means that the 'true value' is close to the null hypothesis, even if it is inaccurate. Power of test controls for this.

Power of tests

The probability that a fixed α level significance test will reject H_0 when an alternative value is true is called the *power* of the test to detect that alternative. Intuitively, it is the region of overlap between the rejection zones and true population distribution.

Power of tests

The probability that a fixed α level significance test will reject H_0 when an alternative value is true is called the *power* of the test to detect that alternative. Intuitively, it is the region of overlap between the rejection zones and true population distribution.



$$\text{power} = P(\text{reject } H_0 \mid H_1 \text{ is true}) = \frac{P(\text{reject } H_0 \wedge H_1 \text{ is true})}{P(H_1 \text{ is true})}$$

$\bar{X} \sim N(\mu_a, \frac{\sigma^2}{n})$

$\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$

reject when $P(|Z| > \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}})$

1. Find rejection zone test statistic given H_0
2. Find probability of true value (given H_1) in those rejection zones

Practically speaking, this is not likely to be a significant deviation – so you control for type II error (though power) by setting the maximum deviation of true mean from H_0 that you can accept (counterpart to level of significance). You then test for it by finding an appropriate sample size that can allow you detect the deviation through hypothesis testing.

Increasing the power

- Increase α – higher chance of rejecting H_0 if H_1 is true since evidence required for rejection is less
- Increase sample size – shrinks wrong distribution assuming H_0 is true, thus increasing likelihood of rejecting H_0 if H_1 is true and thus power, since degree of overlap between H_1 distribution and rejection zones will be larger