A	B	C
The formal logical content	The intuitive background	The applications

Def

If sample space for an experiment is a set S with the property that each physical outcome of the experiment corresponds to exactly one element of S. An element of S is called samplepoint.

Ex

Distribution of two balls in three cells.

1. {a b | - | -}
2. {- | a b | -}
3. {- | - | a b}
4. {a | b | -}
5. {a | - | b}
6. {b | a | -}
7. {- | a | b}
8. {b | - | a}
9. $\underbrace{\{- \mid b \mid a\}}_{\text{1 of 9 sample points}}$

} Sample Space $[S]/[\Omega]$

$A = \{1,2,3\}$ – one cell is mutually occupied.

B - The first cell is not empty. => $\{1,4,5,6,8\}$

Every cell is occupied. <=> impossible event

C - "a" is in the first cell => $\{1,4,5\}$

Def

Any subset of a sample space is called an <u>event</u>. The empty set $\varnothing$ is called "the impossible set"

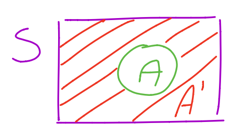
S is called the certain event.

Def

The event consisting of all points not contained in event A will be called the complementary event of A and will be denoted $[A']/[\bar{A}]$. In particular $S' \Leftrightarrow \varnothing$

Ex

B' – The first cell is empty. $\{2,3,7,9\}$

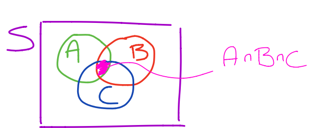


Def

The aggregate of the samplepoints which belong to all the given sets, ABC, will be denoted by $A \cap B \cap C$ and called the intersection or simultanious realization.

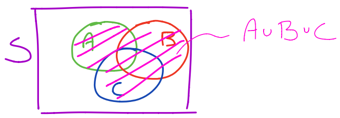
Ex

$A \cap B = \{1\}$



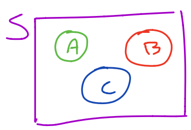
Def
The aggregate of the samplepoints which belong to at least one of the given sets, A B C, will be denoted $A \cup B \cup C$ and called the union.

Ex
$A \cup B = \{1, 2, 3, 4, 5, 6, 8\}$



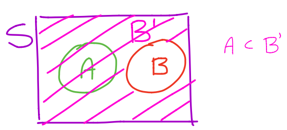
Def
The event A, B, C are mutually exclusive if no two have a point in comon.

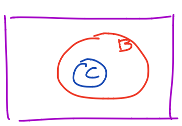
Ex
B and B' are mutually exclusive. $B \cap B' = \emptyset$



Def
The symbols $A \subset B$ and $B \supset A$ signify that every point of A is contained in B. "A implies B", "B is implied by A".

Ex





$C \subset B$
$B \not\subset C$

Ex - Coin toss, 3 random tosses.

$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

D - Two or more H.

E - Just one T.

Def
$$P(A) = \frac{\text{number of ways A can occur}}{\text{number of ways the experiment can proceed}}$$

$P(D) = \frac{4}{8} = \frac{1}{2}$

$P(E) = \frac{3}{8}$

Def
Relative frequency approximation, $P(A) = \dfrac{\text{number of times event A occured}}{\text{number of times experiment was run}}$

Axioms of probability

1. Let S denote a sample space. $P(S) = 1$
2. $P(A) \geqslant 0$, for every event A.
3. Let $A_1, A_2, A_3 \dots$ be a finite or infinite collection of mutually exclusive events. Then $P(A_1 \cup A_2 \cup A_3 \dots) = P(\bigcup_{i=1}^{\infty} A_i) = P(A_1) + P(A_2) + \dots$
$= \sum_{i=1}^{\infty} P(A_i)$

Theorems

$P(A') = 1 - P(A)$

Proof

$P(A') + P(A) = \{ax\ 3\} = P(A \cup A') = P(S) = \{ax\ 1\} = 1$

$P(\varnothing) = 0$

Proof

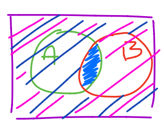
$P(S) = \{S \cup \varnothing = S\} = P(S \cup \varnothing) = \{S \cap \varnothing = 0\} + \{ax\ 3\} = P(S) + P(\varnothing)$

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

Proof

$P(A) + P(B) = \{ax\ 3\} = P(A \cap B) + P(A \cap B') + P(B \cap A) + P(B \cap A') = P(A \cup B) + P(A \cup B)$

$A = (A \cap B) \cup (A \cap B')$



A'
B'
$A \cap B$

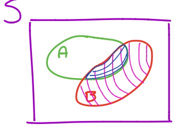
Def

Let A and B be events such that $P(B) > 0$. The conditional probability of A given B denoted by $P(A|B)$ is defined by: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Ex

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

S



The outcome will be in B.
$P(A|B)$

$\left. \begin{array}{l} P(B) = \frac{5}{9} \\ P(A \cap B) = \frac{1}{9} \end{array} \right\}$ $\frac{\frac{1}{9}}{\frac{5}{9}} = \frac{9}{9 \cdot 5} = \frac{1}{5}$

$P(A) = \frac{3}{9} = \frac{1}{9}$

Ex, prev lecture

Distribution of two balls in three cells.

1. $\{ab|-|-\}$ 4. $\{a|b|-\}$ 7. $\{-|a|b\}$
2. $\{-|ab|-\}$ 5. $\{a|-|b\}$ 8. $\{b|-|a\}$
3. $\{-|-|ab\}$ 6. $\{b|a|-\}$ 9. $\{-|b|a\}$

$\} [S]/[\Omega]$

$A = \{1,2,3\}$ - one cell is mutually occupied.
B - The first cell is not empty. $\Rightarrow \{1,4,5,6,8\}$
Every cell is occupied. $\Leftrightarrow$ impossible event
C - "a" is in the first cell $\Rightarrow \{1,4,5\}$

Def

1) Events A and B are independent iff (if and only if) $P(A\cap B) = P(A)\cdot P(B)$
2) Let A and B be events such that at least one of $P(A)$ or $P(B)$ is non-zero. A and B are independent iff $P(A|B) = P(A)$ if $P(B)\neq 0$, and $P(B|A) = P(B)$ if $P(A)\neq 0$.

Ex

$P(A) = \frac{3}{9} = \frac{1}{3}$
$P(B) = \frac{5}{9}$
$P(A\cap B) = \frac{1}{9} \neq \frac{1}{3}\frac{5}{9}$

$P(C) = \frac{3}{9} = \frac{1}{3}$
$P(A\cap C) = \frac{1}{9} = P(A)\cdot P(C)$, A and C are independent

Ex

Given that: $P(1) = \frac{1}{2}$
$P(2) = \ldots = P(9) = \frac{1}{16}$

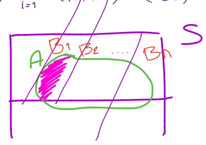
$P(A) = P(1) + P(2) + P(3) = \frac{1}{2} + \frac{1}{16} + \frac{1}{16} = \frac{5}{8}$
$P(C) = P(1) + P(4) + P(5) =$ — " — $= \frac{5}{8}$
$P(A\cap C) = P(1) = \frac{1}{2} \neq P(A)\cdot P(C)$

Law of total probability

Theorem

Let $B_1, B_2, \ldots, B_n$ be such that $B_1 \cup \ldots \cup B_n = S$ and $B_i \cap B_j \neq O$, $i\neq j$, with $P(B_i) > 0$ for every i. Then for any event A; $P(A) = \sum_{i=1}^{n} P(A|B_i)\cdot P(B_i)$

Pic



Ex

B_1 - a is in the first cell
B_2 - a — " — second cell
B_3 - a — " — third cell.

$P(B) = P(B|B_1)\cdot P(B_1) + \ldots = 1\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{1}{3} = \frac{5}{9}$

Theorem (Bayes)

Let $B_1, B_2, \ldots, B_n$ be a collection of mutually exclusive events such that $\bigcup_{i=1}^{n} B_i = S$ and $P(B_i) > 0$ for all i.
Let A be an event such that $P(A) > 0$.

$$P(B_j|A) = \frac{P(A|B_j)\cdot P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

Def. Random variables

X is the random variable if it is the function from the sample space S to the real numbers.

Def

The random number is discrete if it takes finite or at most countable number of values. Ie: its values can (in principel) be listed.

Remark I

$\mathbb{N} = \{0,1,2,3 \dots\}$

$\mathbb{Z} = \{\dots,-3,-2,-1,0,1,2 \dots\}$

$\mathbb{Q} = \{p/q : p \in \mathbb{N}, q \in \mathbb{Z}, q \neq 0\}$

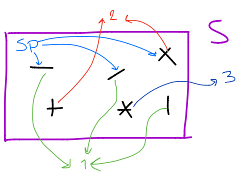
$\mathbb{R} = \mathbb{Q} \cup$ Irrational numbers.

Remark II

"Random function" is a better name for X.

The function is determenistic, but we chose events at random.

Pic



X is the random variable that takes a sample point and returns the number of intersecting lines.

$X(-) = 1$ $\quad X(\times) = 2$

$X(/) = 1$ $\quad X(\ast) = 3$

$X(|) = 1$

$X(+) = 2$

$P(-) = \dots = P(\ast) = \frac{1}{6}$

$\tilde{P}(X(S)=1) = \tilde{P}(X=1) = P(-) + P(/) + P(|) = \frac{1}{2}$

argument of "function X"

$P(X=2) = \frac{1}{3}$

$P(X=3) = \frac{1}{6}$

Def

Given any event A, define indicator random variable.

1_A equal 1 on event A and 0 otherwise.

Remark

If $A_1, A_2, \dots, A_n$ are events then $X = \sum_{i=1}^{n} 1_{A_i}$ is the number of the events that occure.

Ex

$S = \{HH, HT, TH, TT\}$, coin tosses of fair coin

A: One head in two tosses. $A = \{HT, TH\}$

$1_A(HH) = 0$, $1_A(HT) = 1, \dots$

Ex

B - The first cointoss is head.

$X = 1_A + 1_B \mid X(HH) = 1$

$X(HT) = 2$

$X(TH) = 1$

$X(TT) = 0$

Def: Discrete density

Let X be a discrete r.v. The function f given by $f(x) = P(X=x)$, for $x \in \mathbb{R}$ is called "density function" for X.

Theorem of density function:

f is a discrete density iff:

1) $\forall x: f(x) \geq 0$

2) $f(x) > 0$ only in a countable number of points

3) $\sum_{\text{all } x} f(x) = 1$

Ex

Select one of the integers 1 through 10 at random and define a r.v. X as the number of its divisors. $P(1)=\dots=P(10)$

Integer:	1	2	3	4	5	6	7	8	9	10
Divisors	1	2	2	3	2	4	2	4	3	4

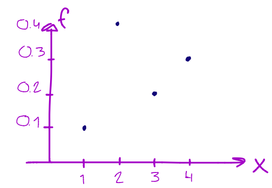
$f(x)=P(X=x)$

$f(0)=0$

$f(2)=0.1+0.1+0.1+0.1=0.4$

Number of divisors "x"	1	2	3	4
Value of $f(x)$	0.1	0.4	0.2	0.3

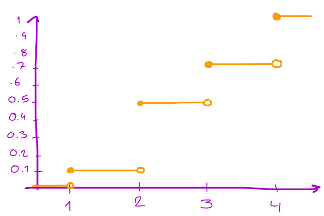
Pic



Def: cumulative distribution

Let X be a discrete r.v. with density f. The cumulative distribution function for X denoted by F_X is defined by $F(x)=P(X\leq x)$, $x\in\mathbb{R}$

Ex



$P(X\leq 1.5)=P(X=1)$

$P(X\leq 2)=P(X=1)+P(X=2)$

$$F(x)=\begin{cases}0, & x<1\\ 0.1, & 1\leq x<2\\ 0.5, & 2\leq x<3\\ 0.7, & 3\leq x<4\\ 1, & x\geq 4\end{cases}$$

Def- Expected value

Let X be a discrete r.v. with density f. The expected value (average, mean) of X is $E[X]=\sum_{\text{all } x} x f(x)$.

Ex

Integer	1	2	3	4	5	6	7	8	9	10
Number of divisors	1	2	2	3	2	4	2	4	3	4

Number of div	1	2	3	4
Value of f	0.1	0.4	0.2	0.3

$E[X]=1\cdot 0.1+2\cdot 0.4+3\cdot 0.2+4\cdot 0.3=2.7$

Def-Expected value of $H(x)$

Let X be a discrete r.v. with density f. Let H be some function. Expected value of $H(x)$ is given by $E[H(x)]=\sum_{\text{all } x} H(x) f(x)$.

Ex

$H(x)=x^2$, $E[H(x)]=E[X^2]=1^2\cdot 0.1+2^2\cdot 0.4+3^2\cdot 0.2+4^2\cdot 0.3=8.3$,

Theorem

Let X be a r.v. Then $E[aX+b]=aE[X]+b$, for any $a,b\in\mathbb{R}$.

Note

$\sum_{\text{all } x} f(x)=1$

Proof

Consider f is a density of X. $E[aX+b]=\sum_{\text{all } x}(ax+b)\cdot f(x)=\sum_{\text{all } x} axf(x)+\sum_{\text{all } x} bf(x)=aE[X]+b$

Properties of expectations

Let X and Y be r.v. $E[X], E[Y]<\infty$.

1) $E[c]=c$, c is constant
2) $E[cX]=cE[X]$
3) $E[X+Y]=E[X]+E[Y]$

Ex

$X\in\{0,1000\}$

$P(X=0)=0.99$ $\quad E[X]=0\cdot 0.99+1000\cdot 0.01=10$

$P(X=1000)=0.01$ $\quad Var[X]=9900$, $\sigma\approx 99.5$

Def- Variance

Let X be a r.v with mean $E[X]$ (μ). The variance of X, denoted by $Var[X]$ (σ_X^2, $D[X]$) is given by:

$Var[X]=E[(X-E[X])^2]$

Def-Standard deviation

Let X be a r.v. with variance σ_X^2. The std.dev. of X is given by $\sigma_X=\sqrt{\sigma_X^2}=\sqrt{Var[X]}$

Theorem

$Var[X]=E[X^2]-(E[X])^2$

Proof

Constant

$Var[X]=E[(X-E[X])^2]=E[X^2-2X E[X]+(E[X])^2]=E[X^2]-2E[X]E[X]+(E[X])^2=E[X^2]-(E[X])^2$

Ex

$Var[X]=8.3-2.7^2=1.01$, $\sigma=1.005$

Properties of variance

1) $Var[c] = 0$, c is constant
2) $Var[cX] = c^2 \cdot Var[X]$
3) $Var[X+Y] \neq Var[X] + Var[Y]$

but sometimes!

Bionomial distribution

A r.v. X is said to have binomial distribution if its density f is given by $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $0 \leq p \leq 1$, $x = 1,2,3 \ldots n$

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$

$E[X] = np$, $Var[X] = np(1-p)$

Geometric distribution

$f(x) = (1-p)^{x-1} \cdot p$, $0 \leq p \leq 1$, $x = 1,2,3,\ldots,n.$

$E[X] = \frac{1}{p}$, $Var[X] = \frac{1-p}{p^2}$

Poisson distribution

$f(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$, $x = 0,1,2 \ldots$, $\lambda > 0$

$E[X] = Var[X] = \lambda$

Def - Continuous

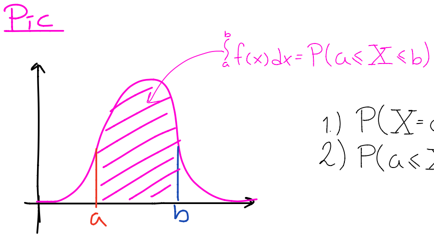
A r.v is continuous if it can assume value in some interval or intervalls of real numbers and the probability that it assumes any particular value is zero.

Def - Continuous density

Let X be a continuous r.v A function f such that:

1) $f(x) \geq 0$ for any x
2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$
3) $P(a \leq X \leq b) = \int_a^b f(x)\,dx$, this is called "density of X.

Pic



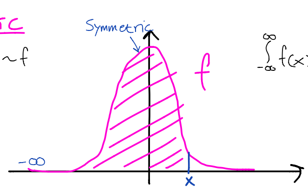
1) $P(X=c) = \int_c^c f(x)\,dx = 0$
2) $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$, if X is continuous

Def

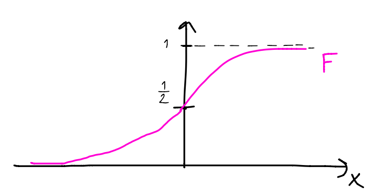
Let X be a continuous r.v with density f. The cumulative distribution function (c.d.f) is denoted by F_X, is defined by: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$, $x \in \mathbb{R}$

Pic

$X \sim f$



$\int_{-\infty}^{\infty} f(x)dx = 1$, $f(x) \geq 0$

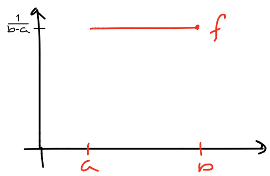


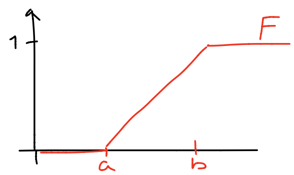
Ex

A continuous r.v. X has a uniform $(Uni[a,b])$ distribution on $[a,b]$ if X falls with the same probability in any subset (sub-interval) $(c,d) \subset (a,b)$ $a \quad c \quad d \quad b$ provided its length is constant

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b) \\ 0, & \text{otherwise} \end{cases} \qquad F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in (a,b) \\ 1, & x > b \end{cases}$$

Pic





For math

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

$$E[H(X)] = \int_{-\infty}^{\infty} H(x) f(x)\,dx$$

$$Var[X] = E[X^2] - (E[X])^2$$

Def
Let X be a continuous r.v with density f. Then expected value of X is $E[X]=\int_{-\infty}^{+\infty} x\cdot f(x)dx$

Ex
$X\sim Uni[a,b]$, $f(x)=\begin{cases}\frac{1}{b-a}, & a\leq x\leq b\\ 0, & \text{otherwise}\end{cases}$

$E[X]=\int_a^b \frac{x}{b-a}dx=\frac{x^2}{2(b-a)}\Big|_a^b=\frac{b+a}{2}$

Def
Let X be a continuous r.v with density f. Then expected value of r.v. $H(x)$, denoted by $E[H(X)]$ is given by $E[H(X)]=\int_{-\infty}^{+\infty} H(x)f(x)dx$.

Ex
$H(x)=x^2$, $X\sim Uni[a,b]$

$E[X^2]=\int_a^b x^2\cdot\frac{1}{b-a}dx=\frac{x^3}{3(b-a)}\Big|_a^b=\frac{a^2+ab+b^2}{3}$

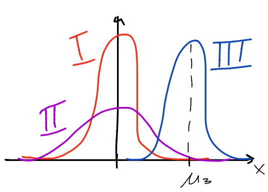
Def
Let X be a continuous r.v with density f and mean $E[X]$.
$Var[X]=E[X^2]-E^2[X]$

Ex
$Var[X]=E[X^2]-E^2[X]=\frac{a^2+ab+b^2}{3}-\left(\frac{a+b}{2}\right)^2=\frac{a^2-2ab+b^2}{12}=\frac{(a-b)^2}{12}=\frac{(b-a)^2}{12}$

Def
A r.v. X with density $f(x)=\frac{1}{\sqrt{2\pi\cdot\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, where $-\infty<\mu<+\infty$, $\sigma^2>0$ have a normal distribution with parameters μ and σ^2.

Remark
$X\sim N(\mu,\sigma^2)$
$E[X]=\mu$
$Var[X]=\sigma^2$

Pic



Expectations: I: 0 Variations: $\sigma_{II}^2>\sigma_I^2$
II: 0 $\sigma_{III}^2=\sigma_I^2$
III: $\mu_3>0$

$X_{III}=X_I+\mu_3$

Def
A r.v with parameters $\mu=0, \sigma^2=1$ is denoted by Z and is called Standard normal random variable.

Proposition
Let $X\sim N(\mu,\sigma^2)$. The variable $\frac{X-\mu}{\sqrt{\sigma^2}}$ is standard normal.

Theorem

Let $X \sim Bin(n,p)$ $\quad P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,1,\dots n$

For large values of n, X is approximately normal with mean np and variance $\underbrace{np(1-p)}_{q}$

Ex

Rolling a dice 108 times. X = number of ones after rolling.

$X \sim Bin(108, \frac{1}{6})$, $\quad E[X]=108\cdot\frac{1}{6}=18$, $\quad Var[X]=108\cdot\frac{1}{6}(1-\frac{1}{6})=15$

1. Exact calculations.
   $P(12 \leq X \leq 20)=\sum_{i=12}^{20}P(X=i)=\sum_{i=12}^{20}\binom{108}{i}\left(\frac{1}{6}\right)^i\left(\frac{5}{6}\right)^{n-i}$

2. Normal approximation, for $n>25$
   $P(12 \leq X \leq 20) \approx P(11.5 \leq X \leq 20.5)=P\left(\frac{11.5-18}{\sqrt{15}} \leq \underbrace{\frac{X-18}{\sqrt{15}}}_{\text{approx } N(0,1)} \leq \frac{20.5-18}{\sqrt{15}}\right)=F(0.645)-F(-1.678)=0.691$

   where F is a cdf of N(0,1)

Reminder

$\Phi(x)=P(Z\leq x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$

cdf for Z.

Kika i VLE-Tables

Def

Let X and Y be discrete r.v's, defined on the same sample space. The ordered pair (X,Y) is called a two-dimensional (bivaritive) discrete random variable.

A function $f_{XY}(x,y)=P(X=x, Y=y)$ is called the joint density for (X,Y).

Properties

1. $f(x,y)\geq 0$, all x,y.
2. $f(x,y)>0$ only in a countable number of points.
3. $\sum_{\text{all } x}\sum_{\text{all } y} f(x,y)=1$

EX

Three coin tossings with a fair coin, i.e. $P(\text{Heads})=P(\text{Tails})=\frac{1}{2}$.

X = {number of heads}
Y = {number of runs}

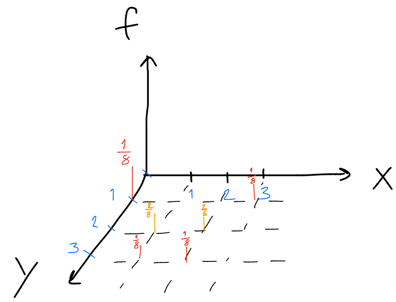
"run"
maximal number of consequtive coin flips that are the same.

Sample Point	Number of heads	Number of runs
HHH	3	1
HHT	2	2
HTH	2	3
HTT	1	2
THH	2	2
THT	1	3
TTH	1	2
TTT	0	1

$X \in \{0,1,2,3\}$
$Y \in \{1,2,3\}$

Number of runs

	Number of heads			
	0	1	2	3
1	$\frac{1}{8}$	0	0	$\frac{1}{8}$
2	0	$\frac{2}{8}$	$\frac{2}{8}$	0
3	0	$\frac{1}{8}$	$\frac{1}{8}$	0



<u>Def</u>

Let X and Y be a bivaritive r.v with joint density f_{XY}. The marginal density for X is given by $f_X(x)=\sum_{\text{all } y} f(x,y)$

Number of runs

	Number of heads			
$x\rightarrow$	0	1	2	3
1	$\frac{1}{8}$	0	0	$\frac{1}{8}$
2	0	$\frac{2}{8}$	$\frac{2}{8}$	0
3	0	$\frac{1}{8}$	$\frac{1}{8}$	0
$P(X=x$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Sample Point	Number of heads	Number of runs
HHH	3	1
HHT	2	2
HTH	2	3
HTT	1	2
THH	2	2
THT	1	3
TTH	1	2
TTT	0	1

$X \in \{0,1,2,3\}$
$Y \in \{1,2,3\}$

Number of heads

Number of runs

$y \backslash x$	0	1	2	3
1	$\frac{1}{8}$	0	0	$\frac{1}{8}$
2	0	$\frac{2}{8}$	$\frac{2}{8}$	0
3	0	$\frac{1}{8}$	$\frac{1}{8}$	0
$P(X=x$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Def

Let X and Y be r.v. with joint density f_{xy} and marginal densities f_x f_y. X and Y are independent if $f_{xy}(x,y) = f_x(x) f_y(y)$ for all x and y.

Ex

$$\left.\begin{aligned} f_{xy}(3,3) &= 0 \\ f_x(3) &= \tfrac{1}{8} \\ f_y(3) &= \tfrac{2}{8} \end{aligned}\right\} \quad f_{xy}(3,3) \neq f_x(3) f_y(3)$$

Note
If we variate the probability we can make f_{xy} independant.

Def

Let X and Y be continuous r.v. The ordered pair (X, Y) is called a two-dimensional, continuous r.v. A function f_{xy} such that:

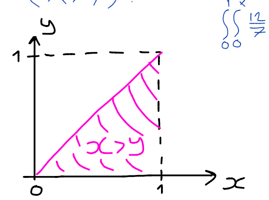
1. $f_{xy}(x,y) \geq 0$, all x, y
2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1$
3. $P(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)\,dx\,dy$

Ex

$f(x,y) = \frac{12}{7}(x^2 + xy)$, $0 \leq x$, $y \leq 1$

$P(X > Y)$?

$$\int_0^1 \int_0^x \frac{12}{7}(x^2 + xy)\,dy\,dx = \frac{9}{14}$$



Def

Let (X, Y) be a two-dimensional r.v with joint density f_{xy}. H is a function then

$$E[H(X,Y)] = \begin{cases} \sum_{\text{all } x} \sum_{\text{all } y} H(x,y) f(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y) f(x,y) & \text{continuous} \end{cases}$$

Remark

1. $H(x,y) = x \cdot y$
2. If X, Y are independent, $E[X \cdot Y] = E[X] \cdot E[Y]$

Def: Covariance

Let X and Y be r.v with means $E[X]$ and $E[Y]$. The covariance between X and Y, denoted $Cov(X,Y)$ or σ_{XY}, is given by:

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

Proposition

1. $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

Def

Let X and Y be r.v with expectations $E[X]$ and $E[Y]$, and variances $Var[X]$ and $Var[Y]$. The correlation, ρ_{xy}, between X and Y is given by:

$$-1 \leq \rho_{xy} = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} \leq 1$$

Remark

If $X = Y$ $\rho_{xy} = \frac{Cov(X,X)}{\sqrt{Var[X]Var[X]}} = \frac{Var[X]}{Var[X]} = 1$

If $X = -Y$ $\rho_{xy} = \frac{Cov(X,-X)}{\sqrt{Var[X]Var[X]}} = \frac{-Var[X]}{Var[X]} = -1$

Note

Correlation measures only linear depandence.

Prelims

1. Systems which changes states in discrete time
2. The collection of all possible states $I = \{i_0, \dots, i_n\}$ is called state space.
3. An initial distribution $\lambda = \{\lambda_{i_0}, \dots, \lambda_{i_n}\}$ defined on I, specifies the starting state.
4. The random mechanism is described by transition matrix P.
   Entry P_{ij} gives the probability that the system will change state from i to j in a unit of time.
5. Each entry in P is non-negative and not greater than one. The sum of entries on each row equals to one.
   A matrix with these properties is called stochastic.

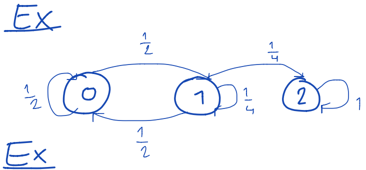
Def

A sequence of random variables X_n with values in a finite set I is a discrete-time Markov chain. (DTMC) With initial distribution λ and transition matrix P if for all $i_0, \dots, i_n \in I$, the joint probability $P(X_0 = i_0, \dots, X_n = i_n) = \lambda_{i_0} P_{i_0 i_1} \cdot \dots \cdot P_{i_{n-1} i_n}$

Def

A state i of MC is called absorbing if it is impossible to leave it. ($P_{ii} = 1$)
A MC is called absorbing if it has atleast one absorbing state and it is possible to reach that state from all other states. It is not necessary to be able to go directly there.

Ex



Theorem

1. $P(X_n = j) = (\lambda P^n)_j$, P^n – n:th power of matrix P
   jth element
2. $P_{ij}^{(n)} = P(X_{k+n} = j \mid X_k = i)$ n-step transition probability

Ex

$$P(X_{n+1} = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = \frac{P(X_0 = i_0, \dots, X_n = i, X_{n+1} = j)}{P(X_0 = i_0, \dots, X_n = i)} = \frac{\lambda_{i_0} P_{i_0 i_1} \cdot \dots \cdot P_{ij}}{\lambda_{i_0} P_{i_0 i_1} \cdot \dots \cdot P_{i_{n-1} i}} = P_{ij}$$

Def

A r.v. X that takes values in $\{0,1,2\dots\}$ is said to be a Poisson r.v with parameter $\lambda>0$, if $P(X=i)=\frac{e^{-\lambda}\lambda^i}{i!}$

Remark

$\sum_{i=0}^{\infty} P(X=i)=1$

$\sum_{i=0}^{\infty} \frac{e^{-\lambda}\cdot\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = \left\{\begin{matrix}\text{Taylor}\\ \text{series}\end{matrix}\right\} = e^{-\lambda}\cdot e^{\lambda}=1$

$X\sim Bin(n,p)$, $n\to\infty$, p is small $\Rightarrow$ $\lambda=np$ - moderate $\Rightarrow$ $p=\frac{\lambda}{n}$

Then $P(X=i)=\binom{n}{i}p^i(1-p)^{n-i}=\binom{n}{i}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}=\frac{n(n-1)\cdot\ldots\cdot(n-i+1)}{n^i}\cdot\frac{\lambda^i}{i!}\cdot\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^i}$

For n large

$\left.\begin{aligned}&\left(1-\frac{\lambda}{n}\right)^n\approx e^{-\lambda}\\ &\frac{n(n-1)\cdot\ldots\cdot(n-i+1)}{n^i}\approx 1\\ &\left(1-\frac{\lambda}{n}\right)^i\approx 1\end{aligned}\right\} \Rightarrow \approx \frac{e^{-\lambda}\lambda^i}{i!}$

Expectation & Variance

Intuition: $X\sim Bin(n,p)$, $E[X]=np$, $Var[X]=np(1-p)$

$Y\sim P_o(\lambda)$, $E[Y]\text{"="}E[X]=np=\lambda$

$Var[Y]\text{"="}Var[X]=np(1-p)\approx np=\lambda$

Proof

$$E[Y]=\sum_{i=0}^{\infty}\frac{ie^{-\lambda}\cdot\lambda^i}{i!}=\lambda\sum_{i=0}^{\infty}\frac{e^{-\lambda}\cdot\lambda^{i-1}}{(i-1)!}=\{j=i-1\}=\lambda\underbrace{\sum_{j=0}^{\infty}\frac{e^{-\lambda}\cdot\lambda^j}{j!}}_{=1}=\lambda$$

$$E[Y^2]=\sum_{i=0}^{\infty}\frac{i^2e^{-\lambda}\cdot\lambda^i}{i!}=\lambda\sum_{i=0}^{\infty}\frac{i\cdot e^{-\lambda}\cdot\lambda^{i-1}}{(i-1)!}=\{j=i-1\}=\lambda\sum_{j=0}^{\infty}\frac{(j+1)e^{-\lambda}\cdot\lambda^j}{j!}=\lambda\sum_{j=0}^{\infty}\frac{j\cdot e^{-\lambda}\cdot\lambda^j}{j!}+\lambda\sum_{j=0}^{\infty}\frac{e^{-\lambda}\cdot\lambda^j}{j!}=\lambda^2+\lambda$$

$$Var[Y]=E[Y^2]-E^2[Y]=\lambda^2+\lambda-\lambda^2=\lambda$$

Stochastic Process

Def

A stochastic process X_t is a collection of r.V.'s. That is for each t in index T, X_t is a random variable

Note

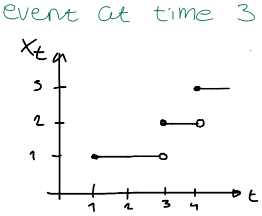
Markov Chain, $T=\{0,1,2\dots\}$

Pair of r.v's, $T=\{1,2\}$

Poisson process, $T=\mathbb{R}$

Def

Any realization of X_t is called sample path. For instance, if events are occurring randomly in time and X_t represents the number of events that occure in $[0,t]$, then the sample path of X_t, which corresponds to the initial event occuring at t_1 the next event at time 3 and the third at time 4, is given below.



Def

A stochastic process $N_t, t \geq 0$ is said to be a counting process if N_t represents the total number of "events" that have occured up to time t.

1) $N_t \geq 0$ for every t
2) N_t is integered valued.
3) $N_t - N_s$, $t>s$, represents the number of events that have happened in the time interval $[s,t]$.

Def

A counting process is said to posses independent increments if the number of events that occur in disjoint time intervals are independent.

Def

A counting process is said to posses stationary increments if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval.

Def

The counting process N_t is said to be a poisson process having rate $\lambda > 0$, if:

1) $N_0 = 0$
2) The process has independent increments.
3) The number of events in any interval of length t [time] is given by $P(N_{t+s} - N_s = i) = \frac{e^{-\lambda t} \cdot (\lambda t)^i}{i!}$

Ex

Suppose that earthquakes occur in Japan in accordance with the assumptions above, with $\lambda = 2$ and with one week as the unit of time.

Find the probability that atleast three earthquakes occur during the next two weeks.

$P(N_2 \geq 3) = 1 - P(N_2 = 0) - P(N_2 = 1) - P(N_2 = 2) = 1 - e^{-4} - \frac{e^{-4} \cdot 4^1}{1!} - \frac{e^{-4} \cdot 4^2}{2!} \approx 0.763$

Def

A continuous r.v. with density $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$, $F(x) = 1 - e^{-\lambda x}$, $E[X] = \frac{1}{\lambda}$, $Var[X] = \frac{1}{\lambda^2}$ is called exponential.

Ex

Suppose that earthquakes occur in Japan in accordance with the assumptions above, with $\lambda = 2$ and with one week as the unit of time.

Find the CDF of the time, starting from now, until the next earthquake.

Let X denote the amount of time, in weeks, 'til the next earthquake.

$P(X > t) = P(N_t = 0) = e^{-\lambda t}$

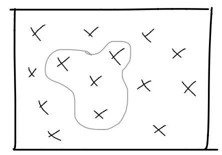
$F(x) = P(X \leq t) = 1 - P(X > t) = 1 - e^{-\lambda t}$

The waiting time between poisson events is exponential in accordance to λ.

Statistical Inference

Def
Statistics is the science of collecting, summarizing and analyzing experimental data to provide the basis for inference of decisions concerning the true nature of the population at study.

Ex
Height of a Chalmers female student.
Pick 10 students at random, measure their heights



Results: 182, 171, 177, 174, 186, 187, 193, 172, 180, 181

Def
A random sample of size n from the distribution of X is a collection of n independent r.v.s, each with the same distribution as X.

Notes
μ = the "true" average heights?
$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 179.9$

Def
A parameter of probability distribution (of population) is a number associated with the distribution (and in some way descriptive of that distribution).

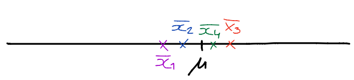
Ex
μ_X - expectation and σ_X^2 - variance

Def
A statistic is some specified numerical function of observed sample values $x_1, \dots, x_n$.

Def
A statistic used to approximate/estimate a distribution parameter θ is called point estimator for θ and is denoted $\hat{\theta}$. A numerical value obtained on a given data is called estimate.

Def
An estimator $\hat{\theta}$ is an unbiased estimator of parameter θ iff $E[\hat{\theta}] = \theta$

Picture



Def
Let $X_1 \dots X_n$ be a random sample of size n from the distribution of X. Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called sample mean.

Proposition
$\bar{X}$ is an unbiased estimator of μ_X with $\sigma_{\bar{X}}^2 = \frac{\sigma^2 X}{n}$, n = number of observations.

Proof

$E[\overline{X}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n} E[\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{=\mu_x} = \frac{1}{n}\cdot n \cdot \mu_x = \mu_x.$

$Var[\overline{X}] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2} Var[\sum_{i=1}^{n} X_i] = \{X_1 \dots X_n \text{ are independent}\} = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{Var[X_i]}_{Var[X] = \sigma_x^2} = \frac{1}{n^2}\cdot n \cdot \sigma_x^2 = \frac{\sigma_x^2}{n}$

Ex

$\overline{X}' = X_1$	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
$E[\overline{x}'] = E[X_1] = \mu_x$	$E[\overline{X}] = \mu_x$
$Var[\overline{x}'] = Var[X_1] = \sigma_x^2$	$Var[\overline{X}] = \frac{\sigma_x^2}{n}$

$E[\hat{\theta}_1] = \theta$ $\quad$ $E[\hat{\theta}_2] - \theta = b > 0$ $\quad$ $Var[\hat{\theta}_1] \gg Var[\hat{\theta}_2]$

Intuition

We need to estimate $E[X-\mu_x]^2 = \sigma_x^2$

$E[X-\overline{x}]^2$

$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ - estimator for σ_x^2

$E[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2] = E[\frac{1}{n}\sum_{i=1}^{n}((X_i-\mu_x)-(\overline{X}-\mu_x))^2] = E[\frac{1}{n}\sum_{i=1}^{n}((X_i-\mu_x)^2 - 2(X_i-\mu_x)(\overline{X}-\mu_x) + (\overline{X}-\mu_x)^2)] =$

$\frac{1}{n}E[\sum_{i=1}^{n}(X_i-\mu_x)^2 - 2(\overline{X}-\mu_x)\sum_{i=1}^{n}(X_i-\mu_x) + n(\overline{X}-\mu_x)^2] = \frac{1}{n}E[\sum_{i=1}^{n}(X_i-\mu_x)^2 - 2n(\overline{X}-\mu_x)^2 + n(\overline{X}-\mu_x)^2] =$

$\frac{1}{n}E[\sum_{i=1}^{n}(X_i-\mu_x)^2 - n(\overline{X}-\mu_x)^2] = \frac{1}{n}E[\sum_{i=1}^{n}(X_i-\mu_x)^2 - n\,E[(\overline{X}-\mu_x)^2]] = \frac{1}{n}(n\sigma_x^2 - \sigma_x^2) = \frac{n-1}{n}\cdot\sigma_x^2$

Def

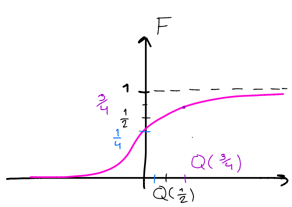
Let $X_1 \dots X_n$ be a random sample of size n from the distribution X. The statistic $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ is called a sample variance.

For computing: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2$

Def

Given a CDF F we can define the quantile function $Q(p) = \inf\{x : F(x) \geq p\}$ for $p \in (0,1)$

Pic



$P(\frac{1}{2})$ $\quad$ $Q(\frac{1}{2})$ = median

$P(\frac{3}{4})$ $\quad$ $Q(\frac{3}{4})$ = upper quartile

$P(\frac{1}{4})$ $\quad$ $Q(\frac{1}{4})$ = lower quartile

Def

A $100(1-\alpha)\%$ confidence interval (CI) for a parameter θ is a random interval $[L,R]$ such that $P(L\leqslant\theta\leqslant R)=1-\alpha$ regardless of the value of θ, where α is called confidence level.

Ex

A chocolate bar's length is denoted by $X\sim N(\mu, 0.5^2)$

Obervations: 13.07, 13.28, 12.36, 13.09, 14.12, 13.11, 12.11, 13.08 , $n=8$

Find a 95% CI for μ.

$X_i\sim N(\mu,\sigma^2)$, $\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

$P[L\leqslant\mu\leqslant R]=0.95$

$P[L-\overline{X}\leqslant\mu-\overline{X}\leqslant R-\overline{X}]=0.95$

$P[\overline{X}-R\leqslant\overline{X}-\mu\leqslant\overline{X}-L]=0.95$

$P[\frac{\overline{X}-R}{\sigma/\sqrt{n}}\leqslant\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\leqslant\frac{\overline{X}-L}{\sigma/\sqrt{n}}]=0.95$ Standard, normal, distribution

We now want to find L such that: 1. $P[Z\geqslant\frac{\overline{X}-L}{\sigma/\sqrt{n}}]=\frac{\alpha}{2}=0.025$

2. $\frac{\overline{X}-L}{\sigma/\sqrt{n}}=-\frac{\overline{X}-R}{\sigma/\sqrt{n}}$

$Q(p)=\min\{x: F(x)\geqslant p\}$ $p\in(0,1)$

$Z_{\frac{\alpha}{2}}=Z_{0.025}=\frac{\overline{X}-L}{\sigma/\sqrt{n}} \Rightarrow L=\overline{X}-Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$

$R=\overline{X}+Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$

$\overline{X}=13.02$, $\sigma=0.5$, $n=8$, $Z_{\frac{\alpha}{2}}=1.96$

CI: $(12.6, 13.37)$

Proposition

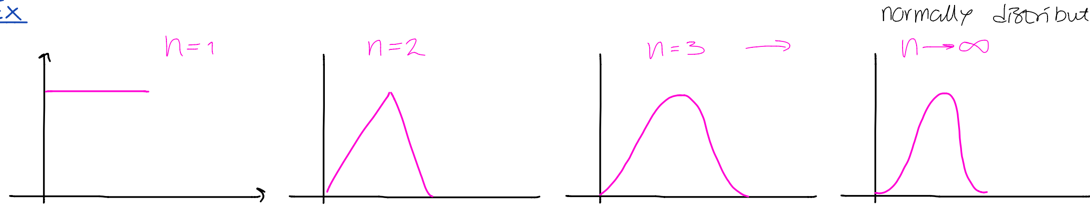
$X_1,\dots,X_n\sim N(\mu,\sigma^2)$, μ is unknown. Then $(1-\alpha)\cdot100\%$ CI is given by $(\overline{X}\pm Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}})$.

Theorem (Central limit theorem)

If $X_1\dots X_n$ are independent and identically distributed as X, where the distribution of X is "decent" ($E[X]<\infty$, $Var[X]<\infty$), then $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$. Note: $\sim$ approximately

Ex

becomes more and more normally distributed



Proposition

Let $X_1\dots X_n$ be a random sample with $\sigma^2=Var[X_1]$. Then $(1-\alpha)100\%$ CI is given by: $(\overline{X}\pm Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}})$

Find σ^2, μ

$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ is unbiased estimator of σ^2. $\sqrt{S^2}$ is a biased estimator for σ

$X_i\sim N(\mu,\sigma^2)$, n is small

Proposition

If $X_1\dots X_n$ is a sample from a $N\sim(\mu,\sigma^2)$, $\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim T_{n-1}$, called -t or student, distribution with $n-1$ degrees of freedom.

Proposition

$X_1 \ldots X_n \sim N(\mu, \sigma^2)$, μ, σ^2 - unknown, (n is small: <30). Then $(1-\alpha)100\%$ CI is given by: $\left(\overline{X} \pm t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{S^2}{n}}\right)$

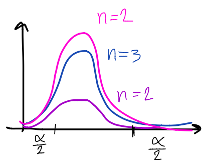
For VLE: Inverse T
n-1 df
correct tail
$\frac{\alpha}{2}$

Exact formula: $f(x) = \frac{\Gamma(n+1)}{2\sqrt{\pi n}\,\Gamma(\frac{n}{2})} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}$

Find σ^2

Proposition

If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ and are independent. $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi_{n-1}$ - Chi-squared with n-1 degrees of freedom.



$P[L \leq \sigma^2 \leq R] = 1-\alpha \longrightarrow \frac{(n-1)S^2}{\sigma^2} \longrightarrow$ Table

Proposition

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, μ, σ^2 - unknown. Then $(1-\alpha)100\%$ CI is given by: $\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right)$

Summation

$\left(\overline{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{n}\right)$. Normal, (0, 1)

$\left(\overline{X} \pm t_{\frac{\alpha}{2}, n-1} \sqrt{\frac{S^2}{n}}\right)$ t-distr; n-1 df

$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right)$ Chi-square; n-1 df

Hypothesis Testing

1. Hypothesis
   $H_0: \theta = \theta_0$ - null hypothesis
   $H_A: \theta = \theta_1$ - alternative hypothesis (simple)
   $\left.\begin{array}{l}\theta > \theta_1 \\ \theta < \theta_1\end{array}\right\}$ one sided alternatives
   $\theta \neq \theta_0$ - two sided alternatives

2. T-Statistic

3. Critical region for 2.
   Think about α

Ex

$Bin(n, \frac{1}{2})$ - $\theta_0 = \frac{1}{2}$, H_0 expectation: $\frac{n}{2}$

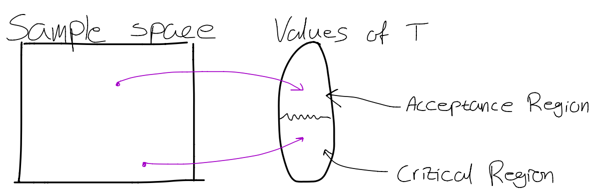
$Bin(n, \theta)$ - $\theta \neq \frac{1}{2}$, H_A expectation: $\frac{3n}{4}$

Random sample: $X_1, \ldots, X_n$

$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ - Test statistic

$\alpha = 0.05$

P-value $= 0.0001$ (unlikely)

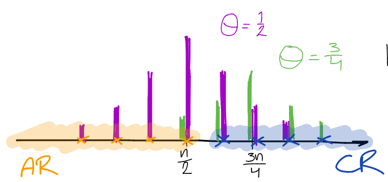


Decision \ Reality	H_0	H_A
H_0	OK	Type II ERROR
H_A	Type I ERROR	OK

Def
A type I error is an error when the null hypothesis is rejected when it is actually true.
The probability of commiting a type I error is called the level of significance of the test.
$\alpha = P[H_A \mid H_0]$

Def
A type II error is an error when the null hypothesis is not rejected when it is actually false.
$\beta = P[H_0 \mid H_A]$

Def
Power of the test is $1 - \beta$



In order to fix type I => knowing type II $\alpha = 0.05$

Def
P-value is the probability for a test statistic, given H_0 is true, to be at least as extreme as its realization obtained from data.

EX
The survey polled 1010 randomely selected employees.

Def
Consider a random sample $X_1, \ldots, X_n$ of size n drawn from a population, where
$X_i = \begin{cases} 1 - \text{if the ith member of the sample has attribute} \\ 0 - \text{otherwise} \end{cases}$
Then $X = \sum_{i=1}^{n} X_i$ gives the number of members with the attribute.

Def

A sample proportion $\hat{p} = \frac{x}{n} = \frac{\text{number in sample with attribute}}{\text{sample size}}$

Sampling distribution of $\hat{p}$:
- $E[\hat{p}] = p$ - population proportion
- $Var[\hat{p}] = \frac{p(1-p)}{n}$
- $\hat{p}$ is approximately Normal for large n (np and $n(1-p)$ are both > 5)

One proportion z-interval procedure

Assumptions:
1. Random sampling
2. The number of successes, x, and number of failures, $n-x$, are both greater than 5.

1. Find $Z_{\alpha/2}$ for a given α.
2. Calculate the end points of the CI: $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
3. Interpret the confidence interval.

Back to Ex

202 employees answered yes (they are ill when actually not)

$n = 1010$

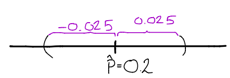
$x = 202$

$n - x = 808$

$\alpha = 0.05$

$\hat{p} = \frac{x}{n} = 0.2$

$Z_{\alpha/2} = 1.96 \Rightarrow \begin{cases} L_1 = 0.175 \\ L_2 = 0.225 \end{cases}$



Def

The margin of error, denoted by E, is given by: $E = Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

d i boken

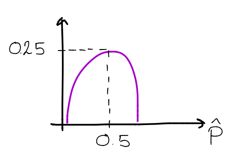
Back to Ex

$n = \hat{p}(1-\hat{p})\left(\frac{Z_{\alpha/2}}{E}\right)^2$

How to find n without $\hat{p}$

1. $0 \leq \hat{p} \leq 1$, $f(\hat{p}) = \hat{p}(1-\hat{p}) = \hat{p} - \hat{p}^2$

2. $\hat{p}_g$ - guess $\hat{p}$

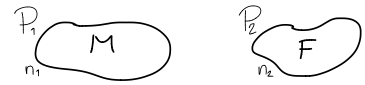


Back...

$E = 0.01, \alpha = 0.05, Z_{\frac{\alpha}{2}} = 1.96$

$n = 0.25 \cdot \left(\frac{1.96}{0.01}\right) = 9604$

Resumé



$\hat{P}_1 - \hat{P}_2$ - statistic

$E[\hat{P}_1 - \hat{P}_2] = P_1 - P_2$ - unbiased estimator

$Var[\hat{P}_1 - \hat{P}_2] = \frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}$

$\hat{P}_1 - \hat{P}_2$ - is approximately Normal for large n_1 and n_2 due to the CLT.

Assumption

- Random sampling
- Independent samples
- $x_1, n_1 - x_1, x_2, n_2 - x_2 \geq 5$ arbitrary rule of thumb

$\hat{P}_1 - \hat{P}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$, two-sided $100(1-\alpha)\%$

Ex

GP ordered a survey. 1181 adults answered the survey conducted by Svenska Veganföreningen. 747 men and 434 women answered.
276 men and 195 women said that they wanted v-food.
Find a 90% CI for $P_1 - P_2$.

$n = 1181$
$n_1 = 747$, $x_1 = 276$
$n_2 = 434$ $x_2 = 195$
$\alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow Z_{\frac{\alpha}{2}} = 1.645$
$\hat{P}_1 = \frac{276}{747} = 0.369$
$\hat{P}_2 = \frac{195}{434} = 0.449$
$0.369 - 0.449 \pm 1.645\sqrt{\frac{0.369(1-0.369)}{747} + \frac{0.449(1-0.449)}{434}} \Rightarrow$ 90% CI: $(-0.129, -0.031)$

Comparing two means

Ex

AAUP wanted to compare the mean sallery between public and private universities.
The survey included 35 faculty members from private unis and 30 from public.
The samples are given in thousands of USD/year

Sample 1 (Private)
87.3, 75.9, 108.8, 83.9, 56.6, 99.2, 54.9...

Sample 2 (Public)
49.9, 105.57, 116.1, 40.3, 123.1, 79.3...

Population 1 - All faculty in private uni.
Population 2 - All faculty in public uni.
μ_1 - Mean salery for Population 1.
μ_2 - Mean salery for Population 2.

1) $H_0: \mu_1 = \mu_2$
$H_A: \mu_1 \neq \mu_2$

2) Test statistic $\bar{X}_1 - \bar{X}_2$, $\bar{X}_x$ is the sample mean of population x.
$\bar{x}_1 = \frac{3086.8}{35} = 88.19$
$\bar{x}_2 = \frac{2195.4}{30} = 73.18$

1. $E[\overline{X}_1 - \overline{X}_2] = E[\overline{X}_1] - E[\overline{X}_2] = \mu_1 - \mu_2$
2. $Var[\overline{X}_1 - \overline{X}_2] = Var[\overline{X}_1] + Var[\overline{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ ← population
3. $\overline{X}_1 - \overline{X}_2$ is approximately normal: $N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{approx } N(0,1)$$

$$\sigma_1^2 = \sigma_2^2 \Rightarrow Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \text{approx } N(0,1)$$

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

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\underbrace{n_1+n_2-2}_{\text{makes it unbiased}}}$$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sim \text{Pooled sample deviation}$$

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \text{t-distribution, df: } n_1+n_2-2$$

Pooled t-test

Assumptions

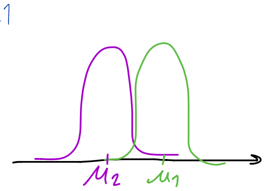
1. Random sampling
2. Independent samples
3. Normal populations, or large sample
4. Equal populations and deviations

Steps

1. Formulate H_0 and H_A.
   - $H_0: \mu_1 = \mu_2$
   - $H_A: \mu_1 \neq \mu_2$ ~ 2-sided
   - $\mu_1 \gtrless \mu_2$ ~ 1-sided
2. Decide α
3. $T = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ if $H_0: \mu_1 - \mu_2 = 0$
4. Critical values
   - $\pm t_{(\frac{\alpha}{2}, n_1+n_2-2)}$ ~ 2-sided
   - $-t_{(\frac{\alpha}{2}, n_1+n_2-2)}$ ~ 1-sided left tail
   - $+t_{(\frac{\alpha}{2}, n_1+n_2-2)}$ ~ 1-sided right tail
5. Find when t(the value of test statistics) lies
6. Interpret the result

Back to uni Ex

$S_1 = 26.21$ $S_2 = 23.95$



So far: $S_1 = 26.21$ $S_2 = 23.95$
$x_1 = 88.19$ $x_2 = 73.18$
$n_1 = 35$ $n_2 = 30$

2. $\alpha = 0.05$

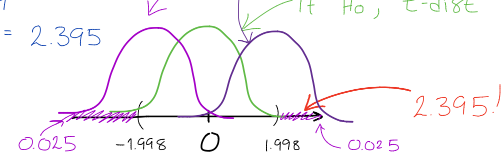
if H_A it will be moved away from 0.

3. $S_p = 25.19$

if H_0, t-dist goes towards normal around 0.

$t = \frac{88.19 - 73.18}{25.19\sqrt{\frac{1}{35} + \frac{1}{30}}} = 2.395$

4. $t_{\frac{\alpha}{2}, 63} = 1.998$



5. $2.399 \notin (-1.998, 1.998)$

6. Accept H_1 at $\alpha = 0.05$

<u>$100(1-\alpha)\%$</u>

$100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$, given $\sigma_1^2 = \sigma_2^2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \cdot S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

<u>Example, continues</u>

$$(88.19 - 73.18) \pm 1.998 \sqrt{\frac{1}{35} + \frac{1}{30}} = (2.49, 27.53) \sim 95\% \text{ CI}$$

The difference between salaries is somewhere between 2 490 and 27 530 USD.

Purpose

To perform a hypothesis test to compare two population means, μ_1 and μ_2.

Assumptions

1. Random sampling.
2. Independent samples
3. Normal population or a large enough sample.

Steps

1. The null hypothesis: $H_0: \mu_1 = \mu_2$ and the alternative
   $H_1: \mu_1 \neq \mu_2$ - 2 sided
   $H_1: \mu_1 \gtrless \mu_2$ - 1 sided
2. Significance level α.
3. Compute the value of test statistic.
   $T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$, given $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ - sample from population 1.
   $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ - sample from population 2
   $S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i^{(1)} - \overline{X}_1)$ - sample variance
4. The critical values are: $\pm t_{\frac{\alpha}{2}, \Delta}$ - 2-sided
   $-t_{\alpha, \Delta}$ - left sided ($H_1: \mu_2 < \mu_1$)
   t-student $+t_{\alpha, \Delta}$ - right sided ($H_1: \mu_2 > \mu_1$)
   $\Delta = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$ round up to closest integer

5. If value of test statistic lies within critical region reject H_0. Otherwise: not enough evidence...

Ex - ALPS vs Z-plate

Z-plate (1)

370, 360, 510, 445, 295, 315, 490, 345, 450, 505, 335, 280, 325, 500

ALPS-plate (2)

430, 445, 455, 455, 490, 535

$\bar{x}_1 = 394.6$, $n = 14$

$S_1 = 84.7$

$\bar{x}_2 = 468.3$, $n = 6$

$S_2 = 38.2$

1. $H_0: \mu_1 = \mu_2$

   $H_1: \mu_1 < \mu_2$
2. $\alpha = 0.05$
3. $t = -2.681$
4. $\Delta = 17$ should be 18.

   $-t_{0.05,17} = -1.740$
5. Rejecting H_0 in favor of H_1, given $\alpha = 0.05$

CI for difference of means

Same assumptions as above

$(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2},\Delta} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ - 2-sided

Calculating 90% CI for $\mu_1 - \mu_2$

$\alpha = 0.1$

$t_{0.05,17} = 1.740$

$(-121.5, -25.9)$ minutes

Generating Functions

Def

The ordinary generating functions (OGF) for the infinite sequence $\langle g_0, g_1, \ldots \rangle$ is the power series

$$G(x) = g_0 + g_1 x + g_2 x^2 + \ldots = \sum_{i=0}^{\infty} g_i x^i$$

Ex

$\langle 0, 0, 0, \ldots \rangle \longleftrightarrow 0 + 0x + 0x^2 + \ldots = 0$

$\langle 1, 0, 0, \ldots \rangle \longleftrightarrow 1 + 0x + 0x^2 + \ldots = 1$

$\langle 1, 2, 3, 0, 0, \ldots \rangle \longleftrightarrow 1 + 2x + 3x^2$

Recall

$1 + z + z^2 + \ldots = \frac{1}{1-z}$, $|z| < 1$

Ex

$\langle 1, 1, 1, \ldots \rangle \longleftrightarrow 1 + 1x + 1x^2 + \ldots = \frac{1}{1-x}$

$\langle 1, -1, 1, -1, \ldots \rangle \longleftrightarrow 1 - x + x^2 - x^3 + \ldots = \frac{1}{1+x}$

$\langle 1, 0, 1, 0 \ldots \rangle \longleftrightarrow 1 + x^2 + x^4 + \ldots = \frac{1}{1-x^2}$

$\langle 1, a, a^2, \ldots \rangle \longleftrightarrow 1 + ax + a^2x^2 + \ldots = \frac{1}{1-ax}$

Scaling

Ex

$\langle 1, 0, 1, 0 \ldots \rangle \longleftrightarrow \frac{1}{1-x^2}$

$\langle 2, 0, 2, 0 \ldots \rangle \longleftrightarrow \frac{2}{1-x^2}$

Theorem-ish

If $\langle g_0, g_1, g_2, \dots \rangle \longleftrightarrow G(x)$

then $\langle cg_0, cg_1, cg_2, \dots \rangle \longleftrightarrow cG(x)$

Since: $\langle cg_0, \dots \rangle \leftrightarrow cg_0 + \dots = c(g_0 + \dots) = cG(x)$

Addition

$\langle 1, 1, 1 \dots \rangle \longleftrightarrow \frac{1}{1-x}$

$\underline{\langle 1, -1, 1, -1, \dots \rangle} \overset{+}{\longleftrightarrow} \frac{1}{1+x}$

$\langle 2, 0, 2, 0 \dots \rangle \longleftrightarrow \frac{2}{1-x^2} = \frac{1}{1-x} + \frac{1}{1+x}$

Theorem-ish

$\langle g_0, g_1, g_2, \dots \rangle \longleftrightarrow G(x)$

$\underline{\langle f_0, f_1, f_2, \dots \rangle} \overset{+}{\longleftrightarrow} F(x)$

$\langle g_0 + f_0, g_1 + f_1, \dots \rangle \leftrightarrow G(x) + F(x)$

Right Shifting

$\langle 1, 1, 1, \dots \rangle \longleftrightarrow \frac{1}{1-x}$

$\langle \overbrace{0, 0, \dots, 0}^{k \text{ zeroes}}, 1 \dots \rangle \longleftrightarrow x^k + x^{k+1} + \dots = x^k \overbrace{(1 + x + x^2 \dots)}^{\frac{1}{1-x}} = \frac{x^k}{1-x}$

Theorem-ish

If $\langle g_0, g_1, g_2, \dots \rangle \longleftrightarrow G(x)$

then $\langle \overbrace{0, \dots 0}^{k \text{ zeroes}}, g_0, g_1, g_2 \dots \rangle \leftrightarrow x^k G(x)$

Differentiation

$$\frac{d}{dx}\left(1+x+x^2+\dots\right)=\frac{d}{dx}\left(\frac{1}{1-x}\right)$$

$$\iff$$

$$1+2x+3x^2+\dots=\frac{1}{(1-x)^2}$$

$$\iff$$

$$\langle 1,2,3,\dots\rangle \longleftrightarrow \frac{1}{(1-x)^2}$$

Theorem-ish

If $\langle g_0, g_1, g_2, \dots\rangle \longleftrightarrow G(x)$

then $\langle g_1, 2g_2, 3g_3, \dots\rangle \longleftrightarrow \frac{d}{dx}G(x)$

Ex

$\langle 0,1,4,9,16\dots\rangle \longleftrightarrow ?$ sequence of squares

$\langle \underset{0}{1},\underset{1}{1},\underset{2}{1},\dots\rangle \longleftrightarrow \frac{1}{1-x}$

derivative

$\langle 1,2,3,\dots\rangle \longleftrightarrow \frac{1}{(1-x)^2}$

right shift

$\langle 0,1,2,3\dots\rangle \longleftrightarrow \frac{x}{(1-x)^2}$

derivative

$\langle 1,4,9,\dots\rangle \longleftrightarrow \frac{1+x}{(1-x)^3}$

right shift

$\langle 0,1,4,9,\dots\rangle \longleftrightarrow \frac{x(1+x)}{(1-x)^3}$

Product rule

If $\langle g_0, g_1, g_2, \dots \rangle \longleftrightarrow G(x)$

and $\langle f_0, f_1, f_2, \dots \rangle \longleftrightarrow F(x)$

then $\langle c_0, c_1, c_2, \dots \rangle \longleftrightarrow G(x)\,F(x)$

$$c_n = g_0 \cdot f_n + g_1 \cdot f_{n-1} + \dots + g_{n-1} \cdot f_1 + g_n \cdot f_0$$

	g_0	$g_1 x$	$g_2 x^2$
f_0	$g_0 f_0$	$g_1 f_0 x$	$g_2 f_0 x^2$
$f_1 x$	$g_0 f_1 x$	$g_1 f_1 x^2$	$g_2 f_1 x^3$
$f_2 x^2$	$g_0 f_2 x^2$	$g_1 f_2 x^3$	$g_2 f_2 x^4$

Repetition

$F(x) = f_0 + f_1x + f_2x^2 + \dots \longleftrightarrow \langle f_0, f_1, f_2, \dots \rangle$

$F(0) = f_0$

$F'(x) = f_1 + 2f_2x + 3f_3x^2 + \dots$

$F'(0) = f_1$

$F''(x) = 2f_2 + 3\cdot 2\cdot f_3x + \dots$

$\frac{F''(0)}{2} = f_2$

$\frac{F^{(n)}(0)}{n} = f_n$

Fibonacci

$f_0 = 0$

$f_1 = 1$

$f_n = f_{n-1} + f_{n-2}$

$\langle 0, 1, 1, 2, 3, 5, 8, \dots \rangle$

$F(x) = f_0 + f_1x + f_2x^2 + \dots$

$\langle f_0, f_1, f_2, f_3, \dots \rangle$

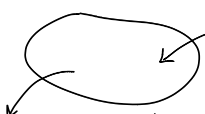
$\langle f_0, f_1, f_0+f_1, f_1+f_2, \dots \rangle$

$$\begin{array}{rl}
\langle 0, 1, 0, 0, \dots \rangle & \longleftrightarrow x \\
\langle 0, f_0, f_1, f_2, \dots \rangle & \longleftrightarrow xF(x) \\
+\ \langle 0, 0, f_0, f_1, f_2, \dots \rangle & \longleftrightarrow x\cdot x\cdot F(x) \\
\hline
\underbrace{\langle 0, f_0+1, f_0+f_1, f_1+f_2, \dots \rangle}_{\text{Fibonacci} = F(x)} & \longleftrightarrow x + x\cdot F(x) + x^2\cdot F(x)
\end{array}$$

$F(x) = x + xF(x) + x^2F(x)$

$F(x) = \frac{x}{1-x-x^2}$

Counting Problems



n elements

Choosing k elements w/o repetition

$\binom{n}{k}$ - number of ways we can choose k out of n elements

$\langle \binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \dots, \binom{n}{n}, 0, 0, \dots \rangle \longleftrightarrow 1 + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n = (1+x)^n$

a_1 $\quad \langle 1, 1, 0, 0, \dots \rangle \longleftrightarrow (1+x)$

$a_1\ a_2$ $\quad \langle 1, 2, 1, 0, 0, \dots \rangle \longleftrightarrow 1 + 2x + x^2 = (1+x)(1+x)$

a_2 $\quad \langle 1, 1, 0, 0, \dots \rangle \longleftrightarrow (1+x)$

A_1 → $F(x)$ - generating function for counting problem

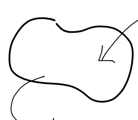
A_2 → $G(x)$ ......

$A_1 \cup A_2$

$C(x) = F(x)G(x) = c_0 + c_1x + c_2x^2 + \dots$

$c_n = f_ng_0 + f_{n-1}g_1 + \dots + f_0 \cdot g_n$

With repetition



n element set

k elements

a_1 $\quad \langle 1, 1, 1, 1, \dots \rangle \longleftrightarrow 1 + 1x + 1x^2 + \dots = \frac{1}{1-x}$

a_2 $\quad \langle 1, 1, 1, \dots \rangle \longleftrightarrow \frac{1}{1-x}$

$a_1\, a_2 \longrightarrow \frac{1}{(1-x)^2}$ $\qquad$ $a_1 \dots a_n \longrightarrow \frac{1}{(1-x)^n} \longleftrightarrow \langle ?, ?, \dots \rangle$

$$G(0) = 1$$
$$G'(x) = \left((1-x)^n\right)' = -n(1-x)^{n-1} \cdot (-1) = n(1-x)^{-(n+1)}$$
$$G''(x) = n \cdot -(n+1)(1-x)^{-n-2} \cdot (-1) = n(n-1)(1-x)^{-(n+2)}$$
$$G^{(k)}(x) = n(n+1) \cdot \ldots \cdot (n+k-1)(1-x)^{-(n+k)}$$
$$\frac{G^{(k)}(0)}{k!} = \frac{n(n+1)\cdot\ldots\cdot(n+k-1)}{k!} = \frac{(n+k-1)\cdot\ldots\cdot(n+1)\cdot n}{k!} \cdot \frac{(n-1)!}{(n-1)!} = \frac{(n+k-1)!}{k!(n-1)!} = \binom{n+k-1}{k}$$

Choosing k from n w/ repetition: $\frac{1}{(1-x)^n} \longleftrightarrow \left\langle \binom{n-1}{0}, \binom{n}{1}, \binom{n+1}{2}, \dots, \binom{n+k-1}{k}, \dots \right\rangle$

Ex

Choosing n fruits from a bag.

- The number of apples must be even.
- The numbers of bananas must be a multiple of 5.
- There can be atmost 4 oranges.
- There can be atmost 1 pear.

Consider $n=6$

Apples:	6	4	4	2	2	0	0
Bananas:	0	0	0	0	0	5	5
Oranges	0	2	1	4	3	1	0
Pears	0	0	1	0	1	0	1

7

Apples

$\langle 1,0,1,0,1,0,\dots \rangle \longleftrightarrow 1 + x^2 + x^4 + \dots = \frac{1}{1-x^2} = A(x)$

Bananas

$\langle 1,0,0,0,0,1,0,\dots \rangle \longleftrightarrow 1 + x^5 + x^{10} + \dots = \frac{1}{1-x^5} = B(x)$

5 banas

Oranges

$\langle 1,1,1,1,1,0,0,\dots \rangle \longleftrightarrow 1 + x + x^2 + x^3 + x^4 = \frac{1-x^5}{1-x} = O(x)$

Reminder

$1 + x + x^2 + \dots = \frac{1}{1-x}$

$1 + x + x^2 + \dots + x^n = \frac{1-x^{n+1}}{1-x}$

Pears

$\langle 1,1,0,0,\dots \rangle \longleftrightarrow 1 + x = P(x)$

Combined

$$A(x) \cdot B(x) \cdot O(x) \cdot P(x) = \frac{1}{1-x^2} \cdot \frac{1}{\cancel{1-x^5}} \cdot \frac{\cancel{1-x^5}}{1-x} \cdot (1+x) = \frac{\cancel{1+x}}{(1-x)\cancel{(1+x)}(1-x)} = \frac{1}{(1-x)^2} \longleftrightarrow \langle 1,2,3,4,5,\dots \rangle$$

Answer: The number of ways to form a bag of n fruits is $n+1$.

Moment generating function

Remark

$E[X^k]$: k-th rows moment
First moment : $E[X]$
Second moment: $E[X^2]$
$E([X-E[X]])^k$: k-th central moment
Second central moment: $Var[X]$

Def

The moment generating function $M(t)$ of the r.v. X is defined for all real values of t by:

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum e^{tx} \cdot p(x) & \text{if } X \text{ is discret with density } p \\ \int_{-\infty}^{\infty} e^{tx} p(x)\,dx & \text{if } X \text{ is continuous with density } p \end{cases}$$

$M'(t) = \frac{d}{dt} E[e^{tX}] = E[\frac{d}{dt} e^{tX}] = E[X \cdot e^{tX}]$
$M'(0) = E[X]$

$M''(t) = (M'(t))' = \frac{d}{dt} E[Xe^{tX}] = E[\frac{d}{dt} X \cdot e^{tX}] = E[X^2 e^{tX}]$
$M''(0) = E[X^2]$

$M^{(n)}(0) = E[X^n]$

Ex

$X \sim Bin(n, p)$
$M(t) = E[e^{tX}] = \sum_{i=0}^{n} e^{ti} \binom{n}{i} p^i (1-p)^{(n-i)} = \sum_{i=0}^{n} \binom{n}{i} (e^t p)^i (1-p)^{(n-i)} = (p \cdot e^t + 1 - p)^n$ $\qquad (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$

$M'(t) = n(p \cdot e^t + 1 - p)^{n-1} \cdot p \cdot e^t$
$M'(0) = n(p + 1 - p)^{n-1} \cdot p = np = E[X]$

Ex

$X \sim Po(\lambda)$
$M(t) = E[e^{tX}] = \sum_{k=0}^{\infty} \frac{e^{tk} \cdot e^{-\lambda} \cdot \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \cdot \lambda)^k}{k!} = \left\{ e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \right\} = e^{-\lambda} \cdot e^{e^t \cdot \lambda} = e^{\lambda(e^t - 1)}$

$M'(t) = e^{\lambda(e^t - 1)} \cdot \lambda \cdot e^t$
$M'(0) = \lambda = E[X]$

Ex

$X \sim Exp(\lambda)$, $p(x) = \lambda \cdot e^{-\lambda x}$ - density
$M(t) = E[e^{tX}] = \int_0^{\infty} e^{tx} \cdot \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda - t)x} dx = \frac{\lambda}{\lambda - t}$, $t < \lambda$

$M'(t) = \frac{\lambda}{(\lambda - t)^2}$
$M'(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} = E[X]$

Moment generating functions

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum e^{tx} p(x), & p\text{-discrete density} \\ \int_{-\infty}^{\infty} e^{tx} p(x)\,dx, & p\text{-continuous density} \end{cases}$$

$Z \sim N(0,1)$

$$M_{Z}(t) = E[e^{Zt}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \cdot e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x^2-2tx}{2}\right)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-t)^2}{2} + \frac{t^2}{2}} dx = \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-t)^2}{2}} dx =$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-t)^2}{2}} d(x-t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-t)^2}{2}} dx$$

$X \sim N(\mu, \sigma^2)$

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

$X = \sigma Z + \mu \sim N(\mu, \sigma^2)$

$$M_X(t) = E[e^{tX}] = E[e^{t(\sigma Z + \mu)}] = E[e^{(t\sigma Z)} e^{t\mu}] = e^{t\mu} \underbrace{E[e^{t\sigma Z}]}_{M_Z(t\sigma)} = e^{t\mu} M_Z(t\sigma) = e^{t\mu} \cdot e^{\frac{(t\sigma)^2}{2}} = e^{\left(\frac{\sigma^2 t^2}{2} + \mu t\right)}$$

$$\boxed{M^{(n)}(0) = E[X^n]}$$

Proposition

X, Y are independant r.v's with mgf.s $M_X(t)$ and $M_Y(t)$ respectively. Then
$M_{X+Y}(t) = M_X(t) M_Y(t)$

Proof

$M_{X+Y}(t) = E[e^{(X+Y)t}] = E[e^{Xt} \cdot e^{Yt}] = E[e^{Xt}] \cdot E[e^{Yt}] = M_X(t) \cdot M_Y(t)$

Corollary

X, Y are independent, $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{\left(\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right)} \cdot e^{\left(\frac{\sigma_2^2 \cdot t^2}{2} + \mu_2 t\right)} = e^{\frac{(\sigma_1^2 + \sigma_2^2) t}{2} + (\mu_1 + \mu_2) t}$$

$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Proposition

Mgf determines the distribution uniquely.

H/W: Check that $Y_1 + Y_2 \sim Po(\lambda_1 + \lambda_2)$ if $Y_1 \sim Po(\lambda_1)$ and $Y_2(\lambda_2)$

Misc

$Y \sim Po(\lambda)$, $Y \overset{\text{in distribution}}{=} Y_1 + Y_2$, $Y_1 \cdot Y_2 \sim Po\left(\frac{\lambda}{2}\right)$

Theorem - Markov's Inequality

If X is non-negative r.v., then for any $\varepsilon > 0$, $P(X > \varepsilon) \leq \frac{E[X]}{\varepsilon}$

Proof

O for $\varepsilon > 0$, let: $I = \begin{cases} 1 & , X \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$

Since $X \geq 0$ $I \leq \frac{X}{\varepsilon}$

Taking expectation $E[I] \leq E\left[\frac{X}{\varepsilon}\right]$, $E[I] = P(X \geq \varepsilon) \leq \frac{E[X]}{\varepsilon}$

Theorem - Chebyshev's inequality

If X is a r.v. with finite $E[X]$ and $Var[X]$, then for any $k>0$: $P(|X-\mu|\geq k)\leq \frac{Var[X]}{k^2}$

Proof

Since $(X-\mu)^2$ is a non-negative r.v. we can apply the markov inequality.

$P((X-\mu)^2\geq k^2)=\frac{E[(X-\mu)^2]}{k^2}$, $k>0$ — $Var[X]$

But since $(X-\mu)^2\geq k^2$ if and only if $|X-\mu|\geq k$

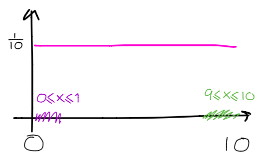
$P(|X-\mu|\geq k)\leq \frac{Var[X]}{k^2}$

$X\sim Uni(0,10)$

$E[X]=5$, $Var[X]=\frac{25}{3}$

$P(|X-5|\geq 4)\leq \frac{\frac{25}{3}}{16}\approx 0.52$ (with chubby)

$P(|X-5|\geq 4)=\frac{1}{10}+\frac{1}{10}=0.2$ (w/o chubby)



density of $X\sim Uni(0,10)$

$|X-5|\geq 4 \sim \begin{cases} 10\geq x\geq 9 \\ \text{or } 1\geq x\geq 0\end{cases}$

Theorem - Weak law of large numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d r.v., each having: $E[X_i]=\mu$. Then for any $\varepsilon>0$:

$P(|\underbrace{\frac{X_1+\ldots+X_n}{n}}_{\bar{X}_n}-\mu|\geq\varepsilon)\to 0$ as $n\to\infty$

Proof

Additional assumption: $Var[X_i]=\sigma^2$

$E[\frac{X_1+\ldots+X_n}{n}]=\mu$, $Var[\frac{X_1+\ldots+X_n}{n}]=\frac{\sigma^2}{n}$

$P(|\frac{X_1+\ldots+X_n}{n}-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{n\varepsilon^2}\longrightarrow 0$ when $n\to\infty$

constants

Theorem

Let $X_1, X_2, \ldots$ be a sequence of i.i.d r.v's, each having $E[X_i]=\mu$, $Var[X]=\sigma^2$. Then

$P(\frac{X_1+\ldots X_n-n\mu}{\sigma\sqrt{n}}\leq a)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}e^{-\frac{x^2}{2}}dx=\Phi(a)$

Remark

You can also say that: n

Proof

Proposition

$Y_1, Y_2, \ldots$ - i.i.d rv having CDF $F_{Y_i}(x)$ and mgf $M_{Y_i}(t)$ and let Y be a rv with CDF $F_Y(t)$ and mgf $M_Y(t)$.

If $M_{Y_1+\ldots+Y_n}(t)\xrightarrow[n\to\infty]{}M_Y(t)$ then $F_{Y_1+\ldots+Y_n}(x)\to F_Y(x)$ for every "suitable" x.

To be continued....

Def

Let $X_1, X_2, \dots$ be a sequence of r.v's with c.d.f's $F_1, F_2, \dots$ Let X be a r.v. with cdf F. We say that X_n converges in distribution to X if $\lim_{n\to\infty} F_n(x) = F(x)$, for every x. (Every x at which F is continuous.)

Theorem - Continuity

Let F_n be a sequence of c.d.f's with the corresponding mgf's M_n. Let F be a cdf with mgf M. If $M_n(t) \to M(t)$ when $n \to \infty$ for all t, then $F_n(x) \to F(x)$ when $n \to \infty$ for every x.

Theorem - CLT

Let $X_1, X_2, \dots$ be a sequence of i.i.d r.v's having mean 0, variance σ^2, c.d.f F and mgf M. Let $S_n = \sum_{i=1}^{n} X_i$, then $\lim_{n\to\infty} \underbrace{P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right)}_{F_{\frac{S_n}{\sigma\sqrt{n}}}(x)} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$, c.d.f of $N(0,1)$

Proof

Let $Z = \frac{S_n}{\sigma\sqrt{n}}$. We will show that $M_{Z_n}(t) \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}}$

$$M_{S_n}(t) = M_{X_1 + \dots + X_n}(t) = (M(t))^n$$
$$M_{Z_n}(t) = \{E[e^{tX}]\} = E[e^{tZ_n}] = E[e^{t\frac{S_n}{\sigma\sqrt{n}}}] = M_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(M\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

Remark - Taylor exp.

$f(s) = f(0) + f'(0)s + \frac{1}{2}f''(0)s^2 + \varepsilon$, $\frac{\varepsilon}{s^2} \to 0$ when $s \to 0$

$e^x \approx 1 + x + \frac{x^2}{2}$, when x is small.

$\sin(x) \approx x$ — " —

$\cos(x) \approx 1 - \frac{x^2}{2}$, — " —

$\ln(1-x) \approx -x - \frac{x^2}{2}$ — " —

Use Taylor expantion for mgf M.

$M(s) = M(0) + M'(0)s + \frac{1}{2}M''(0)s^2 + \varepsilon$

$M'(0) = E[X_i] = 0$

$M''(0) = E[X_i^2]$

$Var[X_i] = E[X_i^2] - E^2[X_i] \Rightarrow$

$M(t) = E[e^{tX_i}]$

$M''(0) = Var[X_i] = \sigma^2$

$$M_{Z_n}(t) = \left(M\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \underset{\text{assuming } \varepsilon \text{ small}}{\approx} \left(1 + \frac{1}{2}\sigma^2 \left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)^n = \left(1 + \frac{t^2}{2n}\right)^n$$

Remark

$\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = e^a$, where $a_n \to a$, $n \to \infty$

$$M_{Z_n}(t) \approx \left(1 + \frac{t^2}{n}\right)^n \to e^{\frac{t^2}{2}}$$

Def - Method of moments (MM)

1. Calculate low-order moments $E[X], E[X^2], E[X^3]$. Finding expressions for the moments in terms of the parameters
2. Invert the expressions from step one to find new expressions for the parameters in terms of the moments.
3. Insert the sample moments into the expressions from step 2.

Ex

$X \sim Bin(n, p)$. Find the method of moments (MM) estimator of p.

1. $E[X] = np$
2. $p = \frac{E[X]}{n}$
3. $\hat{p} = \frac{\overline{X}}{n}$

$X \sim Po(\lambda)$

1. $E[X] = \lambda$
2. $\lambda = E[X]$
3. $\lambda = \overline{X}$

$X \in \{0, 1, 2\}$

$P(X=0) = p$, $P(X=1) = P(X=2) = \frac{1-p}{2}$

1. $E[X] = 0 \cdot p + 1 \cdot \frac{1-p}{2} + 2 \cdot \frac{1-p}{2} = 3\left(\frac{1-p}{2}\right) = \frac{3-3p}{2}$
2. $1 - p = \frac{2}{3} E[X] =$

   $p = 1 - \frac{2}{3} E[X]$
3. $\hat{p} = 1 - \frac{2}{3} \overline{X}$

Def - Likelihood Function

Let $X_1, \ldots, X_n$ be i.i.d r.v's with density f. Then $Lik(\theta) = \prod_{i=1}^{n} f_\theta(X_i)$ - likelihood function.

$L(\theta) = \log(Lik(\theta)) = \sum_{i=1}^{n} \log(f_\theta(x_i))$ - log likelihood function

MLE estimator is the that maximizes likelihood function

Ex

$X \in \{0, 1, 2\}$

$P(X=0) = p$, $P(X=1) = P(X=2) = \frac{1-p}{2} \Rightarrow 0 \leq p \leq 1$

$x_1 = 0$, $x_2 = 0$, $x_3 = 1$

$lik(p) = \underbrace{p}_{\text{"}f_\theta(x_1)\text{"}} \cdot p \cdot \frac{1-p}{2}$, $\quad L(p) = \log(lik(p)) = 2\log(p) + \log\left(\frac{1-p}{2}\right)$, $\quad L'(p) = 2 \cdot \frac{1}{p} + \frac{2}{1-p} \cdot \frac{1}{2} = \frac{2}{p} - \frac{1}{1-p} = \frac{2-3p}{p(1-p)}$

$2 - 3p = 0 \Rightarrow \hat{p} = \frac{2}{3}$, mle estimate

Ex

$X_1, \ldots, X_n$ are coming from an exponential distribution: $Exp(\theta)$. Density of $Exp(\theta)$ is given by: $f_\theta(x) = \theta e^{-\theta x}$

$lik(\theta) = \theta e^{-\theta x_1} \cdot \ldots \cdot \theta e^{-\theta x_n}$

$L(\theta) = \log(\theta e^{-\theta x_1} + \ldots + \theta e^{-\theta x_n}) = \{\log\theta + \log e^{-\theta x_1} = \log\theta - \log\theta x_1\} = n\log\theta - \theta\sum_{i=1}^{n} x_i$

$L'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i = \frac{n - \theta\sum_{i=1}^{n} x_i}{\theta}$

$L'(\hat{\theta}) = 0 \Leftrightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$

mle estimate

Remark

If $X \sim Exp(\theta)$ then $E[X] = \frac{1}{\theta}$ or $\theta = \frac{1}{E[X]}$

Regarding the exam

1. Generating functions

   How to obtain GF for a given sequence.
   How to obtain a sequence given a GF.
   Counting Problems (like the example with the bag of fruits)

2. Moment Generating Functions

   Find a mgf for a given distribution. Both with given values and generally.
   Find some moments of a given distribution.

3. Limit Theorems

   Proof's of: Markov's, Chebyshev's inequalities, law of large numbers, CLT.

4. Confidence Intervals

   CI for the μ, σ^2 known
   CI for the μ, σ^2 unknown (small sample) ← t-distribution
   CI for σ^2
   CI for P - population proportion

Ex: How to construct an interval.

$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ approx $N(0,1)$ because of CLT.

$P(L_1 \leq \mu \leq L_2) = (1-\alpha)$ $\quad [L_1, L_2]$ - $100(1-\alpha)\%$ CI for μ.

$$(1-\alpha) = P\left(L'_1 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq L'_2\right) = P\left(\frac{L'_1\sigma}{\sqrt{n}} \leq \overline{X}-\mu \leq \frac{L'_2\sigma}{\sqrt{n}}\right) = P\left(\frac{L'_1\sigma}{\sqrt{n}} - \overline{X} \leq -\mu \leq \frac{L'_2\sigma}{\sqrt{n}} - \overline{X}\right) = P\left(\overline{X} - \frac{L'_2\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} - \frac{L'_1\sigma}{\sqrt{n}}\right)$$