C1

Probability and Statistics for Computational Sciences (S. Micula)

Email: smicula@math.ubbcluj.ro

Part I. Probability Theory.

3 approaches: axiomatic, classical and geometrical.

Ch1. Probability Space

1. Experiments, Events, Sigma Fields, Probability

Def.: experiment (trial) (probă) = process or action whose outcome is unknown and random (depends on chance).

not S: - sample space = all the possible outcomes of an exp; its elements are elementary events. $S = \{e_1, e_2 \dots\}$.

- event = a collection of elementary events i.e. a subset of S; not A, B, E.

Defining events as sets, we can employ set theory.

For any exp., 2 events:

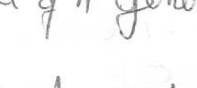
- the impossible event: not $\emptyset$.
- the sure (certain) event: not S.

- Giving an event A, the complementary event $\bar{A} = S \setminus A = \{e \in S / e \notin A\}$.
- the event A implies (induces) an event B if the occurrence of A generates the occ of B. $A \subseteq B$; $e \in A \Rightarrow e \in B$.
- two events A and B are equivalent (equal) if they imply each other. $A = B$ if $A \subseteq B$ and $B \subseteq A$.

• Usual operations:

- the union, $A \cup B$ "or" $A \cup B = \{e \in S / e \in A$ or $e \in B\}$
- intersection $A \cap B$ "and" $A \cap B = \{e \in S / e \in A$ and $e \in B\}$
- difference $A \setminus B$ "but" $A \setminus B = \{e \in S / e \in A$ but $e \notin B\}$ $(= A \cap \bar{B})$.

Def.: Two events A, B are called mutually exclusive (m.e) (disjoint, incompatible) if they cannot occur at the same time, i.e. $A \cap B = \emptyset$.

- A collection of events $\{A_i\}_{i=\overline{1,n}}$ is called m.e ~~m.e~~ if $A_i \cap A_j = \emptyset, \forall i \neq j$.
- collectively exclusive if $\bigcup_{i=1}^{n} A_i = S$.
- partition of S if $\bigcup_{i=1}^{n} A_i = S$, $A_i \cap A_j = \emptyset, \forall i \neq j$.  (sist. complet de evenimente)

Example.

1. Exp. Rolling a dice.

sample space $S = \{e_1, e_2 \dots e_6\}$. e_i: face i shows.

A: face 1 shows. $\{e_1\}$

B: face 2 shows. $\{e_2\}$

C: an even nr shows $\{e_2, e_4, e_6\}$.

D: a prime nr shows $\{e_2, e_3, e_5\}$.

E: a composite nr shows $\{e_4, e_6\}$.

$B \subseteq C$.

$C \cap D = \{e_2\}$.

$A \cap B = \emptyset$, $A \cap E = \emptyset$.

$A \cap D = \emptyset$, $D \cap E = \emptyset$.

$\{A, D, E\}$ m.e

$A \cup \Delta \cup E = S \Rightarrow \{A, \Delta, E\}$. partition of S.

<u>Theorem 1</u>. For any collection of events $\{A_i\}_{i\in I}$, De Morgan's laws <u>hold</u>:

a) $\overline{\bigcup_{i\in I} A_i} = \bigcap_{i\in I} \overline{A_i}$ b) $\overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A_i}$

Def. A coll. of events K from S is a σ-field if:

1). $K \neq \emptyset$.

2) $A \in K \Rightarrow \overline{A} \in K$

3). $A_n \in K, \forall n \in \mathbb{N} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in K$.

Ex. The easiest ex: <u>The power set</u>. $K = \mathcal{P}(S) = \{S' / S' \subseteq S\}$.

(S, K) - measurable space.

<u>Theorem 2</u> Let K be a σ-field over S. Then

a). $\emptyset, S \in K$

b). $A, B \in K \Rightarrow A \cap B, A \setminus B \in K$.

c). $A_n \in K, \forall n \in \mathbb{N} \Rightarrow \bigcap_{n\in\mathbb{N}} A_n \in K$.

Def. Let K be a σ-field over S. A mapping $P: K \to \mathbb{R}$ is called probability if.

1) $P(S) = 1$.

2) $P(A) \geq 0, \forall A \in K$.

3). $\forall \{A_n\}_{n\in\mathbb{N}} \subseteq K$ m.e. events.

$P(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$. ($P$ is σ-additive).

Then (S, K, P) is a probability space.

<u>Theorem 3</u>

Let (S, K, P) be a prob. space. Then a) <u>$P(\overline{A}) = 1 - P(A)$</u> and $0 \leq P(A) \leq 1$, $\forall A \in K$.

b) $P(\emptyset) = 0$.

c) <u>$P(A \setminus B) = P(A) - P(A \cap B)$</u>

d) if $A \subseteq B$ then $P(A) \leq P(B)$.

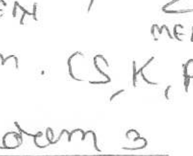
e). $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

<u>Pf</u>. a) Let $A \in K \Rightarrow \{A, \overline{A}\}$ partition of S (m.e).

$1 = P(S) = P(\underbrace{A \cup \overline{A}}_{m.e}) = P(A) + P(\overline{A}) \Rightarrow P(\overline{A}) = 1 - P(A)$.

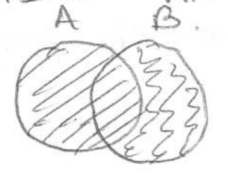
b) $P(\emptyset) = P(\overline{S}) \overset{a}{=} 1 - P(S) = 0$.

c).



$A = \underbrace{(A \cap B) \cup (A \setminus B)}_{m.e}$. $\Rightarrow P(A) = P(A \cap B) + P(A \setminus B) \Leftrightarrow P(A \setminus B) = P(A) - P(A \cap B)$.

d) $A \subseteq B \Rightarrow A \cap B = A \Rightarrow 0 \leq P(B \setminus A) \overset{c}{=} P(B) - P(A) \Rightarrow P(A) \leq P(B)$.

e) 

$$A \cup B = \underbrace{A \cup (B \setminus (A \cap B))}_{m.e.}$$

$$P(A \cup B) = P(A) + P(B \setminus (A \cap B)) \overset{c}{=} P(A) + P(B) - P(B \cap (A \cap B)) =$$
$$= P(A) + P(B) - P(A \cap B).$$

Theorem 4.

Let (S, K, P) be a prob. space. Then:

a) Poincaré's formula (Inclusion Exclu. Principle).

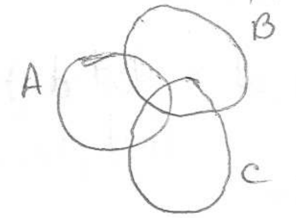
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \ldots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

Remarks

1° Notice that if $\{A_i\}_{i=\overline{1,n}}$ are m.e., then $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.

2° 3 events

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



C2 | 6.10.14. Classical Definition of Probability due to Pascal and Fermat:

Def. Consider an experiment whose outcomes are finite and equally likely. Let A be an event. Then

$$P(A) = \frac{\text{nr. of outcomes fav to the occ of } A}{\text{total nr of pos. outcomes}}$$

Ex. Two dices are thrown. Find the prob of:

A: a double appears.

B: sum of the 2 numbers is ≤ 5.

Sol. Total nr of pos. outcomes: 36.

A: nr of favorable outcomes: 6. (i,i), $i = \overline{1,6}$. $P(A) = \frac{6}{36} = \frac{1}{6}$.

B: nr. of fav. outcomes: 10 $\quad P(A) = \frac{10}{36} = \frac{5}{18}$

(1,1) (1,2) (1,3) (1,4)
(2,2) (2,3). $\quad 6 \times 2$(symmetry) $= 12 - 2$(two are sym) $= 10$.

3. Geometric Probability

a natural way of generalizing classical prob. for the case when the sample space is not finite:

"count" → "measure".

Def (Poincaré, 19th century).

Let $S \subseteq \mathbb{R}^n$ be a measurable space with measure $\mu(S) < \infty$.

Let K be σ-field over S. Let A be an event. Then the geom. prob of A.

$$P(A) = \frac{\mu(A)}{\mu(S)}.$$

The measure μ:

$n=1$, μ = distance (length).

$n=2$, μ = area

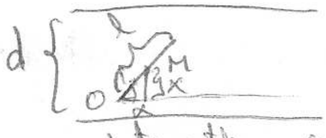
$n=3$, μ = volume.

Ex. (Buffon's Needle Problem)

In a plane, cons. a network of parallel, equidistant lines, at a dist d from each other. A needle of length l ($l<d$) is randomly placed in that plane. Find the prob. that the needle intersects some line.

Sol. The needle can interact at most 1 line (because $l<d$).

Denote by (L) the closest line to the needle. We have 2 pos. situations.



no intersection. (L).

$x \geq |MN|$.

there is intersected. $x < |MN|$.

Denote by x the distance from the midpoint of the needle to the ~~parallel line~~ to (L) ~~(passing through O)~~.

Denote by α the angle between the direction of the needle and the dir of (L) ~~N∈(L)~~). $|MN|$ = dist from M to (L).

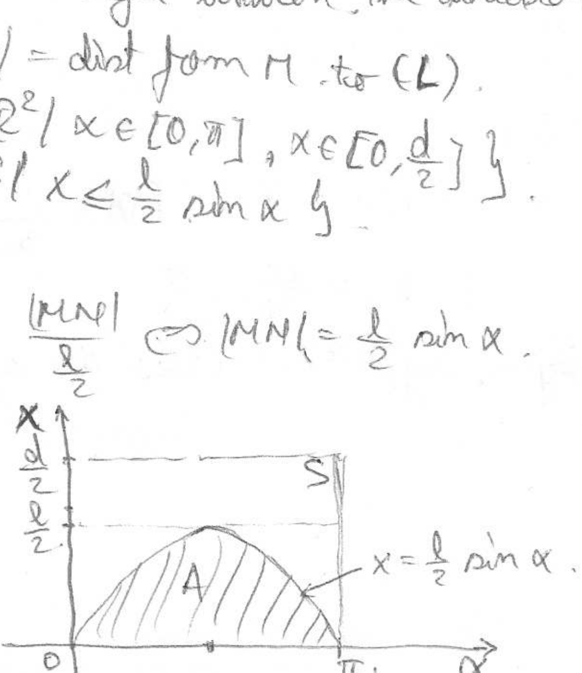
$S = \{(\alpha, x) \in \mathbb{R}^2 / \alpha \in [0,\pi], x \in [0, \frac{d}{2}]\}$.

$A = \{(\alpha, x) \in S / x \leq \frac{l}{2} \sin\alpha\}$.

from ΔOMN.

$$\sin\alpha = \frac{|MN|}{|OM|} = \frac{|MN|}{\frac{l}{2}} \Leftrightarrow |MN| = \frac{l}{2}\sin\alpha.$$

$n=2$, μ = area.



$\mu(S) = \text{area}(S) = \pi \cdot \frac{d}{2} = \frac{\pi d}{2}$.

$\mu(A) = \text{area}(A) = \int_0^{\pi} \frac{l}{2}\sin\alpha \, d\alpha =$

$= \frac{l}{2}(-\cos\alpha)\Big|_0^{\pi} = l.$

$P(A) = \frac{l}{\frac{\pi d}{2}} = \frac{2l}{\pi d}$.

4. Conditional Probability ; Independence

Def.

Let (S, K, P) be a prob space, let $B \in K$ be an event, with $P(B) \neq 0$. Then for every $A \in K$, the prob of A given B (the cond. prob of A conditioned by B) is given by:

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

Prop. 1.

$\forall A, B \in K$, with $P(A)P(B) \neq 0$, $P(A \cap B) = P(A) \cdot P(B|A) = P(B) P(A|B)$.

Prop 2 (The Multiplication Rule).

Let $\{A_i\}_{i=\overline{1,m}} \subseteq K$, with $P(\bigcap_{i=1}^{m} A_i) \neq 0$. Then

$$\underbrace{P(\bigcap_{i=1}^{m} A_i)}_{LHS} = \underbrace{P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \times \dots \times P(A_m | \bigcap_{i=1}^{m-1} A_i)}_{(RHS)}$$

Proof.

$$RHS = \cancel{P(A_1)} \cdot \frac{\cancel{P(A_1 \cap A_2)}}{\cancel{P(A_1)}} \cdot \frac{\cancel{P(A_1 \cap A_2 \cap A_3)}}{\cancel{P(A_1 \cap A_2)}} \cdots \frac{P(\bigcap_{i=1}^{m} A_i)}{\cancel{P(\bigcap_{i=1}^{m-1} A_i)}} = P(\bigcap_{i=1}^{m} A_i) = LHS.$$

Prop 3

$\forall A, B \in K$, $0 < P(A) < 1$, $P(B) = P(A) P(B|A) + P(\bar{A}) P(B|\bar{A})$.

Pf

Cons. $\{A, \bar{A}\}$ partition of S $(A \cup \bar{A} = S,\ A \cap \bar{A} \underset{m.e.}{=} \emptyset)$, (easiest ex of a partition)

$$B = B \cap S = B \cap (A \cup \bar{A}) = \underbrace{(B \cap A) \cup (B \cap \bar{A})}_{m.e.}$$

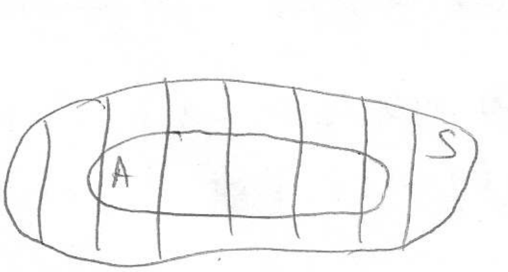
$$P(B) = P(B \cap A) + P(B \cap \bar{A}) \overset{P.1}{=} P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A}).$$

Prop 4. (The Total Probab. Rule)

Let $\{A_i\}_{i \in I} \subseteq K$ be a partition of S and let $A \in K$. Then

$$P(A) = \sum_{i \in I} P(A_i) P(A|A_i).$$

Pf. $\bigcup_{i \in I} A_i = S$, $A_i \cap A_j = \emptyset$, $\forall i \neq j$.



$$B = B \cap S = B \cap (\bigcup_{i \in I} A_i) = \underset{m.e.}{\bigcup_{i \in I} (B \cap A_i)}$$

$$P(B) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(A_i) P(B|A_i).$$

Def.
Two events $A, B \in K$ are independent if: $P(A\cap B) = P(A)\cdot P(B)$.

$\{A_i\}_{i=\overline{1,n}}$ are indep. if $P(A_{i_1}\cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$ $\forall k=\overline{2,n}$.

Remarks. 1° If A, B ind, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

2° If $A = \emptyset$ (the impos. event) or $A = S$ (the certain event), then $\forall B \in K$, A and B are indep.

Prop 5.
Let $A, B \in K$ be indep. Then $A, \overline{B}$ are also indep.

Pf.. $P(A\cap\overline{B}) = P(A\setminus B) = P(A) - P(A\cap B) = P(A) - P(A)\cdot P(B) = P(A)(1-P(B)) =$
$= P(A)\cdot P(\overline{B})$.

Remarks. 1° If A, B are indep then $(A,\overline{B}), (\overline{A}, B), (\overline{A},\overline{B})$ are also indep.

2° The same property is true for any number of events.

Chapter II. Probabilistic Models. "pattern".

→ develop a model for problems falling in the same class of prob. problems, that depends on some parameters, come up with computational formulas. (prob. model).

Sometimes, the easiest setup is to consider a box containing a nr (known or not) of balls of different colors. The exp. consists of selecting a ball and noting its color. The problem: Selecting a given nr. of balls, find the prob of having a certain color distribution.

Important distinction

sampling
- with replacement: the ball is put back; can be sel. again.
- without replacement

when nothing is specified the sampling is with repl.

1. Bernoulli Model (Binomial Model).

Bernoulli trials
- they are indep
- each trial has 2 poss. outcomes < "succes" (A). / "failure" $(\overline{A})$.
- the prob of succes, p, stays the same in every trial.

($q = 1-p$, prob. of failure)

Bernoulli Model:

Given n Bernoulli trials, with prob of succes $p \in (0,1)$, find the prob. of exactly k $(0 \leq k \leq n)$ successes occurring.

Prop 1.

That prob. $P(n;k) = C_n^k \, p^k \, q^{n-k}$.

$\boxed{C3}$ 13.10.14

1. Binomial Model

Recall

Bernoulli trials:

- indep.
- 2 poss outcomes < "success" (prob p). "failure" (prob $q = 1-p$).
- the prob of success, $p \in (0,1)$ is the same in every trial.

Model: Find the prob that in n Bernoulli trials, exactly k successes occur $(0 \leq k \leq n)$

Prop 1

That prob is $P(n;k) = C_n^k \, p^k \, q^{n-k}$

Remarks

1°. The number $P(n;k)$ is the coeff of x^k in the binomial expansion.

$(px+q)^n = \sum_{k=0}^{n} P(n;k)\, x^k$ (hence, the name).

2°. $\sum_{k=0}^{n} P(n,k) = 1$. (let $x = 1$ above).

Ex.

① A dice is rolled 5 times. Find the prob of:

a) A: getting 3 6's.

b) B: getting at least 2 even nrs.

a). success: getting a 6, $p = \frac{1}{6}$, $q = \frac{5}{6}$.

$P(A) = P(5;3) = C_5^3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2$

b). success: getting an even nr, $n = 5$.

$p = \frac{3}{6} = \frac{1}{2} = q$.

at least 2: 2 or 3 or 4 or 5. $(k \geq 2)$

complement: at most 1: 0 or 1. $(k < 2)$.

$P(B) = 1 - P(\bar{B}) = 1 - P(\underbrace{0 \overset{\cup}{\text{or}} 1}_{\text{m.e.}} \text{ successes}) = 1 - (P(0 \text{ succ.}) + P(1 \text{ succ.})) =$

$= 1 - P(5;0) - P(5;1) = 1 - C_5^0 \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^5 - C_5^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4$

2. <u>Bernoulli Model Without Repl.</u>

<u>Hypergeometric model</u>.

Model: A box contains N balls, m_1 of which are white ($m_1 < N$), the rest are black. A nr of n balls is selected, <u>without replacement</u>. Find the prob $P(n;k)$ that k balls are white ($n-k$ are black).

OR.

N-objects $<$ m_1 have a characteristic. / $N-m_1$ do not.

Choose n of them without repl. Find the prob that k have that char.

<u>Prop.</u>

$$P(n;k) = \frac{C_{m_1}^{k} \cdot C_{N-m_1}^{n-k}}{C_N^n}$$

Remark

$\sum_{k=0}^{n} P(n;k) = 1$, i.e. $\sum_{k=0}^{n} C_{m_1}^{k} C_{N-m_1}^{n-k} = C_N^n$.

Ex:

②. There are 15 boys and 20 girls in a prob class. Ten people are selected for a project. Find the prob that the group contains:

a) A: an equal nr of boys and girls,

b) B: at least one girl.

Sol.

a) success: picking a girl. $N=35$, $m_1=20$, $n=10 <$ 5 / 5

$$P(A) = P(10;5) = \frac{C_{20}^5 \cdot C_{15}^5}{C_{35}^{10}}$$

b). success: girl.

B: at least 1 success ($k \ge 1$)($k = 1, 2, \ldots 10$) (or or)

$\bar{B}$: no success: 0 succ.

$$P(B) = 1 - P(\bar{B}) = 1 - P(10;0) = 1 - \frac{C_{20}^0 C_{15}^{10}}{C_{35}^{10}}$$

3. Poisson Model.

A gen of the binomial model.

Poisson Model.

Cons. n indep. trials, each of which has 2 poss. outcomes. Let p_i denote the prob of success in trial i, $i=\overline{1,n}$. Find the prob of exactly k successes occurring in those trials.

Prop 3

$$P(n;k) = \sum_{1\le i_1<\dots<i_k\le n} p_{i_1}\cdots p_{i_k}\, q_{i_{k+1}}\cdots q_{i_n}.$$

Remarks

1°. The nr $P(n;k)$ is the coeff of x^k in the polynomial exp.

$$\boxed{(p_1x+q_1)(p_2x+q_2)\cdots(p_nx+q_n) = \sum_{k=0}^{n} P(n;k)\,x^k}$$

2°. As a conseq. $\sum_{k=0}^{n} P(n;k)=1$ (let $x=1$ in 1°)

3° If $p_i=p$ $\forall i=\overline{1,n}$ $\to$ binomial model.

Ex.

③. (Sem 2) The 3 shooters probl). The prob of 3 shooters hitting a target are $p_1=0.4$, $p_2=0.5$, $p_3=0.7$. Find the prob. of the target being hit exactly once.

Sol.

success: the target is hit, $n=3$, p_1, p_2, p_3.

Poisson Model: $k=1$; $P(3;1)=$ the coeff of x in:

$(0.4x+0.6)(0.5x+0.5)(0.7x+0.3) = \dots$

4. Pascal's Model (Negative Binomial).

Pascal Model: Cons. an infinite seq of Bernoulli trials with prob. of success $p\in(0,1)$, $(q=1-p)$. Find the prob $P(n;k)$ of the n^{th} success occurring after k failures. ($k=0,1,\dots$).

Prop. 4.

$$P(n;k) = C_{n+k-1}^{k}\, p^n q^k = C_{n+k-1}^{n-1}\, p^n q^k$$

Remark.

$$\sum_{k=0}^{\infty} P(n;k) = 1$$

5. Geometric Model.

A part. case of 4 with $n=1$.

Prop 5.

$P_k = P(1;k) = pq^k$

Rem.

$$\sum_{k=0}^{\infty} p.q^k = p \sum_{k=0}^{\infty} q^k \text{ (geom. series with ratio } q\in(0,1)) = p \cdot \frac{1}{1-q} = 1$$

Ex. 4

④. When a dice is rolled over and over, find the prob of the events:

a) A: the 1^{st} 6 appears after 5 throws.

b). B: the 3^{rd} even nr. appears after 5 thrown.

Sol.

a). success: getting a 6 $p=\frac{1}{6}, q=\frac{5}{6}$ (Geometric model).

$P(A) = p_5 = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^5$.

b). success: getting an even nr. $p=q=\frac{1}{2}$

B: 3^{rd} success after 5 trials : 3^{rd} success after 3 failures

Pascal's Model: $p=\frac{1}{2}$ $n=3, k=3$.

$P(B) = P(3,3) = C_5^3 \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^3$.

Chapter III. Random Variables and Random Vectors

a quantitative description of random phenomena.

→ random var.s < discrete / continuous.

1. Discrete R. Var.s and Probab. Distribution Function (pdf).

Let (S, K, P) be a prob. space.

Def. A discrete rand. var. (d.r.v) is a function $X: S \to R$ satisf 2 conditions: (i). the set of values that it takes $(X(S))$ is at most countable (discrete). (ii) $X^{-1}(x) = \{e \in S / X(e) = x\} = (X<x) \in K$ $\forall x \in R$.

Ex.

1. Cons. the exp of rolling a dice. $S = \{e_1, \dots e_6\}$. e_i: face i shows

Let $K = P(S) = \{S' / S' \subseteq S\}$. Let $X: S \to R$ be def. by $X(e_i) = i$.

(i) $X(S) = \{1, \dots 6\}$. ✓.

(ii). $\forall x = i \in \{1, \dots 6\}$. $X^{-1}(x) = X^{-1}(i) = \{e_i\} \in K$

$\forall x \in \mathbb{R} \setminus \{1, \dots, 6\}$

$X^{-1}(x) = \emptyset \in K$

②. The indicator of an event. Let $A \in K$, let $X_A : S \to \mathbb{R}$; $X_A(e) = \begin{cases} 1, & e \in A \\ 0, & e \notin A \end{cases}$

(i) $X_A(S) = \{0, 1\}$.

(ii) $X^{-1}(1) = A \in K$; $X^{-1}(0) = \bar{A} \in K$

$\forall x \in \mathbb{R} \setminus \{0,1\}$, $X^{-1}(x) = \emptyset \in K$.

Random Variables.

Discrete R.Vs.

$X : S \to \mathbb{R}$.

i) $X(S)$ is at most countable (discrete).

ii). $X^{-1}(x) = \{e \in S \mid X(e) = x\} = (X = x) \in K$

inv. image,

Ex. ①. Rolling a dice $X : S \to \mathbb{R}$ $X(e_i) = i$ $i = \overline{1,6}$.

② The indicator of an ev. $A \in K$

$X_A : S \to \mathbb{R}$ $X_A(e) = \begin{cases} 0, & e \notin A \ (e \in \bar{A}) \\ 1, & e \in A \end{cases}$

③ $\{A_i\}_{i \in I}$ partition of S. I at most countable.

$\{x_i\}_{i \in I} \subseteq \mathbb{R}$.

$X : S \to \mathbb{R}$.

$X(e) = \sum_{i \in I} x_i X_{A_i}(e)$. $\quad e \in A_i \Leftrightarrow X(e) = x_i$.

$\boxed{C4}$ With what prob does a discrete r.v take each of its values?

Def: Let $X : S \to \mathbb{R}$ be a d.r.v. The prob. distribution func (pdf in MATLAB) is an array of the form $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$. $p_i = P(X = x_i)$ $(*)$.

Remarks

1° All values in $(*)$ are distinct.

2°. All prob p_i are $\neq 0$.

3°. $\sum_{i \in I} p_i = 1$.

Ex:

④ for ①.

$$X\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \qquad f:\mathbb{R}\to\mathbb{R} \quad f(x)=\begin{cases} \frac{1}{6}, & \text{if } x\in\{1,\dots,6\} \\ 0, & \text{otherwise} \end{cases}$$

⑤ for ②.

$$X\begin{pmatrix} 0 & 1 \\ P(\bar{A}) & P(A) \end{pmatrix} \qquad p=P(A).$$

$1-P(A)$.

$$f(x)=\begin{cases} p^{x}(1-p)^{1-x}, & x=0,1. \\ 0, & \text{otherwise}. \end{cases}$$

2. Cumulative Distr. Function. (Funcţie de repartiţie)

Def: Let X be a r.v. The cdf (ratted). $F_X:\mathbb{R}\to\mathbb{R}$, $F(x)=P(X<x)$.

Remark

$$F(x)=P(\{e\in S \mid X(e)<x\})=P(X<x).$$

Ex:

⑥. The cdf ~~for~~ of the indicator r.v.

$$X_A\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix} \qquad p=P(A). \qquad F_A:\mathbb{R}\to\mathbb{R}.$$

$x=-1$.

$F(-1)=P(X<-1)=0$.

$F(0)=P(X<0)=0$. because $X=0$ or 1.

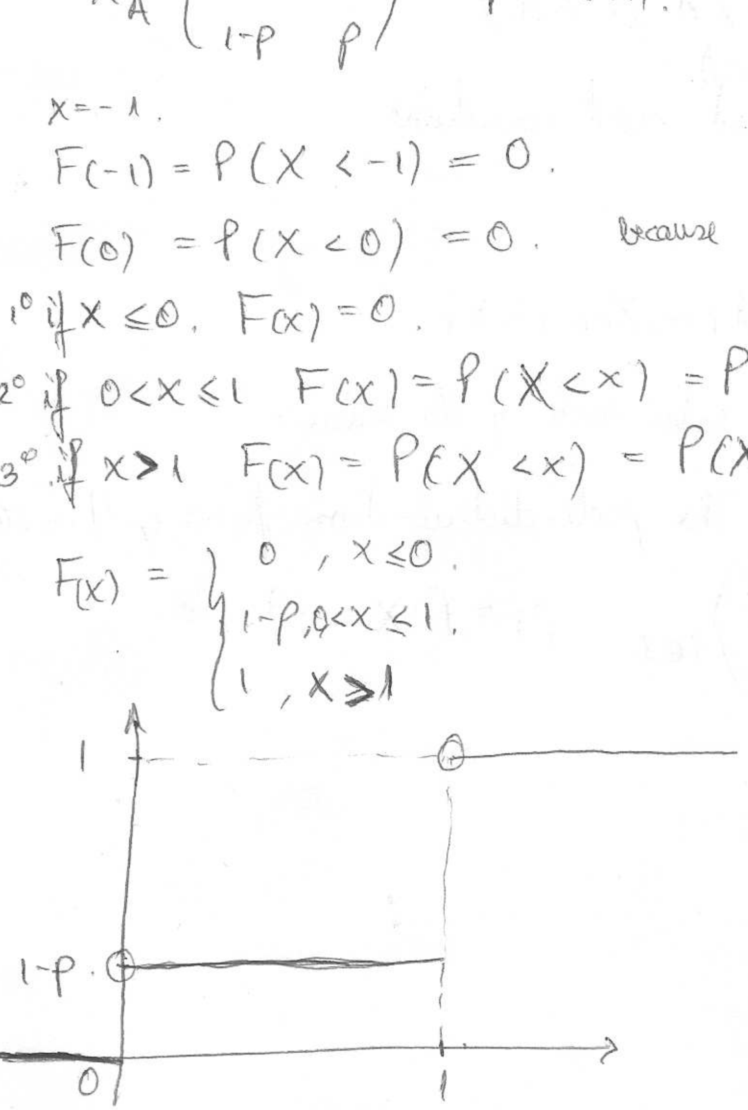
1° if $x\le 0$, $F(x)=0$.

2° if $0<x\le 1$ $F(x)=P(X<x)=P(X=0)=1-p$

3° if $x>1$ $F(x)=P(X<x)=P(X=0 \overset{\cup}{\text{or}} X=1)=P(X=0)+P(X=1)=1-p+p=1$.

m.e.

$$F(x)=\begin{cases} 0, & x\le 0. \\ 1-p, & 0<x\le 1. \\ 1, & x>1 \end{cases}$$



<u>Theorem 1</u> (Properties of c.d.f.).

Let X be a r.v. and let F be its c.d.f. Then F satisfies:

a). $a<b \Rightarrow P(a \le X < b) = F(b) - F(a)$.

b) F is monoton. increasing.

c). F is left continuous. $F(x-0) = F(x)$

$$\lim_{y \nearrow x} F(y) = F(x)$$

d) end behaviour. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

e). $P(X \le x) = F(x+0) (= \lim_{y \searrow x} F(y))$,

$P(X = x) = F(x+0) - F(x)$.

Pf of a).

a) $a \le X < b \Leftrightarrow X < b$, but not $X < a$.

$\Leftrightarrow X < b$ and $X < a$

$\Leftrightarrow (X<b) \setminus (X<a)$.

$$P(a \le X < b) = P((X<b) \setminus (X<a)) = P(X<b) - P((X<b) \cap (X<a)) =$$

$$= P(X<b) - P(X<a) = F(b) - F(a).$$

<u>Remark</u>.

If X is a ~~disc~~ discr. r.v., then $F(x) = \sum_{x_i < x} p_i$.

③. Discrete Distributions. (Prob. Laws)

① <u>Bernoulli Distr.</u>

$X \in Bern(p)$, $p \in (0,1)$ if its pdf is $X \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$, $q = 1-p$.

② <u>Binomial Distr.</u> (bino in Matlab).

$X \in B(n,p)$, $n \in \mathbb{N}$, $p \in (0,1)$. $X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$.

X represents the nr of successes in a Binomial model.

<u>Remark</u>

$Bern(p) = B(1,p)$.

③ <u>Discrete Uniform Distribution</u> (unid in Matlab).

$X \in U(m)$, $m \in \mathbb{N}^*$

pdf.

$$X \begin{pmatrix} k \\ \frac{1}{m} \end{pmatrix}_{k=\overline{1,m}}.$$

Rolling a dice $X \in U(6)$.

④ Hypogeometric Distr. (hyge).

$X \in H(N, n_1, n)$. $N, n_1, n \in \mathbb{N}$, $n, n_1 \leq N$.

$$X \begin{pmatrix} k \\ \frac{C_{n_1}^{k} \cdot C_{N-n_1}^{n-k}}{C_N^n} \end{pmatrix}_{k=\overline{0,n}}.$$

X repr. the nr. of successes in a hypogeometric model.

⑤. Poisson Distr (poiss).

$X \in P(\lambda)$, $\lambda > 0$. $X \begin{pmatrix} k \\ \frac{\lambda^k}{k!} e^{-\lambda} \end{pmatrix}_{k=0,1,\dots}$

X arises in a Poisson process: observing a discr. event in a cont medium.

eg.: nr of earthquakes in a year.

nr of white blood cells in a drop of blood.

The param. λ repr. the average nr of occurences.

Remarks

1°. Poisson law is also known as "The law of Rare Events" because.

$$\lim_{k\to\infty} P(X=k) = \lim_{k\to\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 0.$$

$$2°.\ \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{\text{Taylor Series for } e^{\lambda}} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

6. Pascal (Neg. Binomial) Distr (nbin)

$X \in NB(n, p)$, $n \in \mathbb{N}$, $p \in (0,1)$, its pdf.

$$X \begin{pmatrix} k \\ C_{n+1-k}^{k} p^n q^k \end{pmatrix} \quad q = 1-p. \quad k = 0, 1, \dots$$

X repr. the nr. of failures after which the n^{th} success occurs in a Pascal Model.

7. Geometrical Distr (geo).

$$X \in G(p) = NB(1,p) \quad X \begin{pmatrix} k \\ p \cdot q^k \end{pmatrix}_{k=0,1,\dots}$$
$p \in (0,1)$.

4. Discrete Random vectors and Joint Prob. Distr. Funct. (2-dim)

Def: Let (S, K, P) be a prob. space, A discr (2-dim) random vector. is a function $(X,Y): S \to \mathbb{R}^2$ satisf.:

1). $(X,Y)(S)$ is at most countable.

2). $(X,Y)^{-1}(x,y) = \{e \in S \mid (X,Y)(e) = (x,y)\} = ((X,Y) = (x,y)) =$
$= (X = x, Y = y) \in K$.

The joint pdf

$Y \backslash X$	...	x_i
⋮		⋮
y_j	...	p_{ij}

$$p_{ij} = P((X,Y) = (x_i, y_j)) =$$
$$= P(X = x_i, Y = y_j) \quad (i,j) \in I \times J.$$

Remark.

$\sum_{j \in J} p_{ij} = p_i = P(X = x_i)$. i.e from the joint pdf of the vector (X,Y) we can get the pdf's of X, Y, resp. called marginal distributions

$\sum_{i \in I} p_{ij} = q_j = P(Y = y_j)$.

Def. Two d. r. v's $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ and $Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}$ are called independent if:

$$p_{ij} = P((X,Y) = (x_i, x_j)) = P(X = x_i) P(Y = y_j) = p_i q_j$$
$\forall (i,j) \in I \times J$.

Recall ev A, B indep if $P(A \cap B) = P(A) \cdot P(B)$.

5. Operations with Discr. R.V.'s.

$$X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I},\quad Y\begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j\in J}$$

1) The sum of X, Y.

$$X+Y\begin{pmatrix} x_i+y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}$$

$p_{ij} = P(X=x_i, Y=y_j)$. if X, Y ind then $p_{ij} = p_i q_j$.

2) <u>Scalar Multiple</u>

$\alpha\in\mathbb{R}$. $\alpha X\begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i\in I}$.

3) <u>Difference</u>. $X-Y = X+(-1)Y$.

$$X-Y = \begin{pmatrix} x_i-y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}.$$

4) <u>Product</u>

$$XY\begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}.$$

5) <u>Quotient</u> if $y_j \neq 0$, $\forall j\in J$.

$$X/Y\begin{pmatrix} x_i/y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}.$$

$\boxed{C5}$

6. Continuous Random Variables and Probability Density Function.

<u>Def.</u> Let (S, K, P) be a prob space. A random variable is a function $X: S\to\mathbb{R}$ ~~st~~ satisf. the inverse image

$$X^{-1}((-\infty, x)) = \{e\in S / X(e) < x\} = (X<x)\in K,\ \forall x\in\mathbb{R}.$$

<u>Def</u> Let $F:\mathbb{R}\to\mathbb{R}$ be def by <u>$F(x) = P(X<x)$</u> the cdf of X.

Def. Let X be a r.v. with cdf F. We say that X is a cont r.v if F is abs. cont, i.e $\exists f: \mathbb{R} \to \mathbb{R}$ s.t.

$$\boxed{F(x) = \int_{-\infty}^{x} f(t)\,dt} \quad (*)$$

The funct f from $(*)$ is called prob. density function (pdf).

Theorem 1. (Properties of a pdf).
Let X be a cont. r.v with cdf F and pdf f. Then:

a). $F'(x) = f(x)$, $\forall x \in \mathbb{R}$.

b) $f(x) \geq 0$, $\forall x \in \mathbb{R}$.

c) $\int_{\mathbb{R}} f(x)\,dx = 1$

d) $P(X=x) = 0$.

$P(a \underset{<}{\leq} X \underset{\leq}{<} b) = \int_a^b f(x)\,dx$.

Pf. a). Differentiate $(*)$. $F'(x) = f(x)$

b) Recall that F is mon. increasing $\Rightarrow F'(x) \geq 0 \Leftrightarrow f(x) \geq 0$.

c) Recall $F(\infty) = 1$. $(= \lim\limits_{x\to\infty} F(x))$.

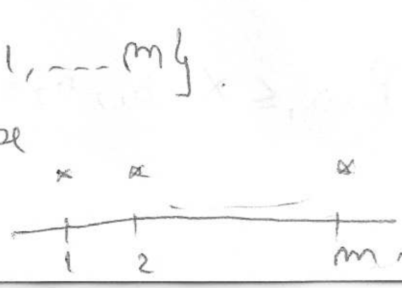
$\int_{\mathbb{R}} f(t)\,dt = \int_{-\infty}^{\infty} f(t)\,dt = \lim\limits_{x\to\infty} \int_{-\infty}^{x} f(t)\,dt = \lim\limits_{x\to\infty} F(x) = F(\infty) = 1$.

d) Recall that $P(X=x) = F(x+0) - F(x)$. But F is abs. cont, so cont.
$\Rightarrow F(x+0) = F(x)$ $(= F(x-0))$. $\Rightarrow P(X=x) = 0$.

Recall $P(a \leq X < b) = F(b) - F(a) \overset{(*)}{=} \int_a^b f(x)\,dx$.

I. Continuous Distribution (Prob. Laws).

① Uniform (unif). $X \in U(a,b)$, $a<b$, pdf $f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$

> discr. $X \in U(m)$.
> pdf $f(x) = \begin{cases} \frac{1}{m}, & x \in \{1, \ldots m\} \\ 0, & \text{otherwise} \end{cases}$
>
> pdf $\begin{pmatrix} k \\ \frac{1}{m} \end{pmatrix}_{k=\overline{1,m}}$.
>
> 

②. Normal Distr. (norm).

$X \in N(\mu, \sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$

pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

$f'(x) = 0 \Leftrightarrow x = \mu.$

$f''(x) = 0 \Leftrightarrow x = \mu \pm \sigma.$

③ Standard (Reduced) Normal Distr.

$X \in N(0,1)$. pdf $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$.

Statistics.

- Normal $N(0,1)$, $N(\mu, \sigma)$.
- Student $T(n)$, $n \in \mathbb{N}$ (t).
- (chi2) $\chi^2(n)$, $n \in \mathbb{N}$.
- Fischer. $F(m,n)$ $m, n \in \mathbb{N}$ (f).

8. Continuous. R. Vectors. and Joint CDF.

only 2-dim case.

Def Let (S, K, P) be a prob space. A (2-dim). r. vector is a funct $(X,Y): S \to \mathbb{R}^2$ satisf. $(X,Y)^{-1}((-\infty,x) \times (-\infty,y)) = \{e \in S / X(e) < x, Y(e) < y\} = (X<x, Y<y) \in K$, $\forall (x,y) \in \mathbb{R}^2$.

Def. Let (X,Y) be a r. vector. The function $F: \mathbb{R}^2 \to \mathbb{R}$. $F(x,y) = P(X<x, Y<y)$ is called the (joint) cdf of (X,Y).

Theorem 1 (Properties of the joint cdf).

Let (X,Y) be a r. vector with joint cdf $F_{(X,Y)}: \mathbb{R}^2 \to \mathbb{R}$ and, let $F_X, F_Y: \mathbb{R} \to \mathbb{R}$ be the cdf's of X, Y.

Then.

a) If $a_k < b_k$, $k = \overline{1,2}$ $P(a_1 \le X < b_1, a_2 \le X < b_2) = F(b_1, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, a_2)$.

b) F is mon. incr. in each var..

c) F is left cont. in each var.

d) $F(\infty,\infty)=1$ e) $F(x,-\infty) = F(-\infty,y)=0$.

$F(x,\infty)= F_X(x)$.

$F(\infty,y) = \cancel{F_{(y)}}\, F_Y(y)$.

9 - Joint Density Function (joint pdf), Marginal Densities, Indep. R. Variables.

~~Def~~. Let (X,Y) be a r. vector with cdf $F:\mathbb{R}^2\to\mathbb{R}$. We say that (X,Y) is a cont. r. vector if $\exists f:\mathbb{R}^2\to\mathbb{R}$ s.t. $\boxed{F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv}$

Then f is called the (joint) pdf of the cont r vector (X,Y).

Theorem 2 (Prop. of a joint pdf). Let (X,Y) be a cont r. vector with joint cdf $F:\mathbb{R}^2\to\mathbb{R}$, joint pdf $f:\mathbb{R}^2\to\mathbb{R}$ and $f_X, f_Y:\mathbb{R}\to\mathbb{R}$ the pdf's of X, Y, resp. Then

a) $\dfrac{\partial^2 F(x,y)}{\partial x\,\partial y} = f(x,y), \quad \forall (x,y)\in\mathbb{R}^2$.

b) $f(x,y)\geq 0, \quad \forall (x,y)\in\mathbb{R}^2$.

c) $\iint_{\mathbb{R}^2} f(u,v)\,du\,dv = 1$.

d) $\forall D\subseteq\mathbb{R}^2,\ P((X,Y)\in D) = \iint_D f(x,y)\,dx\,dy$.

e) $f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy, \quad f_Y(y)=\int_{\mathbb{R}} f(x,y)\,dx$.

~~Def~~ Two r. variables X,Y are independent if

$\cancel{F}\ F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y) \quad (\forall (x,y)\in\mathbb{R}^2)$.

Remark 1°. In the discr. case (1) $\Leftrightarrow p_{ij} = P(X=x_i, Y=y_j) = P(X=x_i)P(Y=y_j) = p_i q_j \quad \forall i,j$.

2° In the cont case (1) $\Leftrightarrow f_{(X,Y)}(x,y) = f_X(x)\, f_Y(y) \quad \forall x,y\in\mathbb{R}$.

10. Functions of cont. R. Var's.

Let (X,Y) be a cont. r. vector with pdf f, f_X, f_Y pdf's of X, Y.

Prop 1.

a) The pdf of the sum $Z = X+Y$.

$$f_Z(z) = \int_{\mathbb{R}} f(u, z-u)\, du, \quad \forall z \in \mathbb{R}.$$

b) product $Z = X \cdot Y$.

$$f_Z(z) = \int_{\mathbb{R}} f(u, \tfrac{z}{u}) \cdot \frac{1}{|u|}\, du.$$

c) quotient $Z = X/Y$.

$$f_Z(z) = \int_{\mathbb{R}} f(zu, u) \cdot |u|\, du.$$

Prop 2

Let $g: \mathbb{R} \to \mathbb{R}$ be a strictly mon. and differentiable function, with $g'(x) \neq 0$, $\forall x \in \mathbb{R}$. Let X be a cont. r. v. with pdf f_X and let $Y = g(X)$. Then

$$f_Y(y) = \begin{cases} \dfrac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}, & y \in g(\mathbb{R}) \\ 0, & y \notin g(\mathbb{R}). \end{cases}$$

Pf. Note that under our assumptions, the inverse function $g^{-1}: g(\mathbb{R}) \to \mathbb{R}$ exists.

a). g is strictly increasing. Start with the cdf $F_Y(y) = P(Y < y) =$
$= P(g(X) < y)$

- if $y < \min g(x)$, $P(g(X) < y) = 0$.
- if $y > \max g(x)$, $P(g(X) < y) = 1$.

if $y \in g(\mathbb{R})$.

$g(X) < y \Leftrightarrow X < g^{-1}(y)$.

$F_Y(y) = P(g(X) < y) = P(X < g^{-1}(y)) = F_X(g^{-1}(y))$

$$F_Y(y) = \begin{cases} 0, & y < \min g \\ F_X(g^{-1}(y)), & y \in g(\mathbb{R}) \\ 1, & y > \max g \end{cases}$$

Take the derivative

$$(F_X(g^{-1}(y)))' = F_X'(g^{-1}(y)) \cdot (g^{-1}(y))' = f_X(g^{-1}(y)) \cdot \frac{1}{g'(g^{-1}(y))}.$$

$$f_Y(y) = \begin{cases} \dfrac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}, & y \in g(\mathbb{R}) \\ 0, & y \notin g(\mathbb{R}) \end{cases}$$

Prop. 3.

Let $X_1, \dots, X_m \in N(0,1)$ indep. Then $Y_m = \sum_{i=1}^{m} X_i^2 \in \chi^2(m)$.

Prop 4.

Let $X \in N(0,1)$, $Y \in \chi^2(m)$ ind. Then $Z = \frac{X}{\sqrt{\frac{1}{m} Y}} \in T(m)$.

Prop 5. $X \in \chi^2(m)$, $Y \in \chi^2(m)$ ind. Then $Z = \frac{Y_m X}{Y_m Y} \in F(m, m)$.

6.6 cred.

Ch IV. Numerical Characteristics of Random Variables.

1. Expectation.

Def.

- If X is a discr. r.v., pdf $\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$, $p_i = P(X = x_i)$ the expected value (expectation, mean value) of X. $E(X) = \sum_{i \in I} x_i p_i = \sum_{i \in I} x_i P(X = x_i)$, if it exists ($< \infty$).
- If X is a cont r.v. with pdf $f: \mathbb{R} \to \mathbb{R}$, then $E(X) = \int_{\mathbb{R}} x f(x)\, dx$ if it exists.

Remark

If $h: \mathbb{R} \to \mathbb{R}$, X a r.v. then if $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}$, then $E(h(X)) = \sum_{i \in I} h(x_i) p_i$.

~~X f:~~ - X $f: \mathbb{R} \to \mathbb{R}$ then $E(h(X)) = \int_{\mathbb{R}} h(x) f(x)\, dx$

Ex:

1. $X \in Bern(p)$, $p \in (0,1)$ pdf $\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$. $E(X) = 0(1-p) + 1p = p$.
2. $X \in U(a,b)$, $a, b \in \mathbb{R}$, $a < b$, pdf $f: \mathbb{R} \to \mathbb{R}$ $f(x) = \frac{1}{b-a}$, $x \in [a,b]$.

$$E(X) = \int_{\mathbb{R}} x f(x)\, dx \left(= \int_{-\infty}^{a} + \int_{a}^{b} + \int_{b}^{\infty} \right)$$

$$= \frac{1}{b-a} \int_a^b x\, dx = \frac{1}{b-a} \frac{1}{2} x^2 \Big|_a^b = \frac{1}{\cancel{b-a}} \cdot \frac{1}{2} (\underbrace{b^2 - a^2}_{a+b}) = \frac{a+b}{2}.$$

Theorem 1 (Properties of $E(X)$)

If X, Y are both discr. r.v.'s or both cont r.v.'s, $a, b \in \mathbb{R}$, then

a). $E(aX+b) = a \cdot E(X) + b$.

b) $E(X+Y) = E(X) + E(Y)$.

c) if X, Y indep. $E(XY) = E(X)E(Y)$.

d) if $X \le Y$, then $E(X) \le E(Y)$.

Pf. (selected) only X, Y discr.

a) $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I} \Rightarrow aX+b \begin{pmatrix} ax_i+b \\ p_i \end{pmatrix}_{i\in I}$.

$$E(aX+b) = \sum_{i\in I}(ax_i+b)p_i = a\underbrace{\sum_i x_i p_i}_{=E(X)} + b\underbrace{\sum_i p_i}_{=1} = aE(X)+b.$$

c). $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$, $Y\begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j\in J} \Rightarrow XY\begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}$

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

$$E(XY) = \sum_{(i,j)} x_i y_j p_{ij} = \sum_i \sum_j x_i y_j p_i q_j = \left(\sum_i x_i p_i\right)\left(\sum_j y_j q_j\right) = E(X)E(Y)$$

d) "$X \le Y$" means $X(e) \le Y(e)$, $\forall e \in S$.

We prove $Z \ge 0 \Rightarrow E(Z) \ge 0$ $\quad Z\begin{pmatrix} z_i \\ p_i \end{pmatrix}_{i\in I}$.

$Z \ge 0 \Rightarrow z_i \ge 0, \forall i \in I$.

$E(Z) = \sum_i \underbrace{z_i}_{\ge 0}\underbrace{p_i}_{\ge 0} \ge 0$. Apply this to $Z = Y - X$.

$X \le Y \Rightarrow Z \ge 0 \Rightarrow E(Z) \ge 0$

$E(Z) = E(Y) - E(X)$ so $E(Y) \ge E(X)$.

Remark

1° b) in T.1. can be extended to any nr of r.v.'s.

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i).$$

2°. Same for c), if $\{X_i\}_{i=\overline{1,n}}$ indep. r.v.'s. $E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$

Ex

③ $X \in B(n,p)$, $n \in \mathbb{N}$, $p \in (0,1)$.

Fact (Sem 4).

If $X_1, \dots X_n \in Bern(p)$, indep., then $X = \sum_{i=1}^{n} X_i \in B(n,p)$.

$X_i \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$ $q = 1-p$ $E(X_i) = p$.

By T.1 b) $\Rightarrow E(X) = \sum_{i=1}^{n} \underbrace{E(X_i)}_{p} \Rightarrow E(X) = np$.

2. Variance, Standard Deviations, Moments.

Def.

Let X be a r.v. The variance (dispersion) of X is the nr.

$V(X) = E[(X-E(X))^2]$, (if it exists) and

$\sigma(X) = \sqrt{V(X)}$ is called the standard derivation of X (abatere medie patratica)

Theorem 2 (Properties of $V(X)$). Let X, Y be r.v.'s. Then.

a) $V(X) = E(X^2) - (E(X))^2$.

b) $V(aX+b) = a^2 V(X)$

c) if X, Y indep $\Rightarrow V(X+Y) = V(X) + V(Y)$.

d) if X, Y indep then. $V(XY) = E(X^2)E(Y^2) - (E(X))^2 (E(Y))^2$.

Pf (selected).

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

$= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (E(X))^2$

b). $V(aX+b) = E[(aX+b-aE(X)-b)^2] = E[a^2(X-E(X))^2] = a^2V(X)$

Remark.

1°. a). – X discr. pdf $\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ $V(X) = \sum_i x_i^2 p_i - \left(\sum_i x_i p_i\right)^2$

– X cont pdf $f: \mathbb{R} \to \mathbb{R}$ $V(X) = \int_{\mathbb{R}} x^2 f(x)dx - \left(\int_{\mathbb{R}} x f(x)\, dx\right)^2$.

2°. c) can be generalized : $\{X_i\}_{i=\overline{1,n}}$ indep. $V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i)$.

3'. By T1. a) (we know $V(X) \geq 0$) $\to E(X^2) \geq (E(X))^2$.

$|E(X)| \leq \sqrt{E(X^2)}$

Ex:

①. $X \in Bern(p)$ $X\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$.

$E(X) = p$.

$X^2\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$

$E(X^2) = p$.

$V(X) = p - p^2 = p(1-p) = pq$.

2. $X \in B(n,p)$ $(X = \sum_{i=1}^{n} X_i, X_i \in Bern(p)$ ind$) \to V(X) = \sum_{i=1}^{n} V(X_i) = npq$

3. (Sem 6). $X \in N(\mu, \sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

$E(X) = \mu$

$V(X) = \sigma^2$ (st. dev $\sigma(X) = \sqrt{V(X)} = \sigma$).

Def.

Let X be a r.v, let $k \in \mathbb{N}$

- the (initial) moment of order k of X.

$\nu_k = E(X^k)$.

- the absolute moment of order k of X.

$\bar{\nu}_k = E(|X|^k)$

- the central (centered) mom of ord. k.

$\mu_k = E[(X - E(X))^k]$ ~~centered at a.~~

Remarks

1°. $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ discr. $\nu_k = \sum_{i \in I} x_i^k p_i = \sum_i |x_i|^k p_i$.

$\mu_k = \sum_i ((x_i - E(X))^k)$.

2°. X cont pdf $f: \mathbb{R} \to \mathbb{R}$. $\nu_k = \int_{\mathbb{R}} x^k f(x)\,dx = \int_{\mathbb{R}} |x|^k f(x)\,dx$.

$\mu_k = \int_{\mathbb{R}} (x - E(X))^k f(x)\,dx$.

2°. $E(X) = \nu_1$.

$V(X) = \mu_2 = \nu_2 - \nu_1^2$.

$\mu_1 = 0$.

$\mu_3 = \nu_3 - 3\nu_1\nu_2 - 2\nu_1^3$.

3°. If ν_n exists for some $n \in \mathbb{N}$, then ν_k and $\nu_k(\mu_k)$ exist $\forall k = \overline{1,n}$.

3. Covariance, Correlation Coefficient, Quantiles.

Def.

Let X, Y be r.v.'s.

- the covariance of X, Y (cof (X,Y)) is the nr: $cov(X,Y) = E[(X-E(X))(Y-E(Y))]$
- the corr.coeff. of X, Y is $\rho(X,Y) = \dfrac{cov(X,Y)}{\sqrt{V(X)V(Y)}} = \dfrac{cov(X,Y)}{\sigma(X)\sigma(Y)}$.

Remarks

1° Obviously, $cov(X,Y) = cov(Y,X)$.

$\rho(X,Y) = \rho(Y,X)$.

2° $cov(X,Y) > 0$ means X, Y grow at the same time.

3° - $cov(X,Y) < 0$ means as one increases, the other decr.

Theorem 3.

Let X, Y, Z be r.v.'s. Then a) $cov(X,X) = V(X)$

b) $cov(X,Y) = E(XY) - E(X)E(Y)$.

c) if X, Y indep $\Rightarrow cov(X,Y) = \rho(X,Y) = 0$. ($X, Y$ are uncorrelated).

d) $V(aX+bY) = a^2 V(X) + b^2 V(Y) + 2ab\, cov(X,Y)$.

e). $cov(X+Y, Z) = cov(X,Z) + cov(Y,Z)$.

Remarks.

1° c). X, Y ind $\underset{\not\Leftarrow}{\Rightarrow}$ X, Y uncorrel.

2° d). can be generalized.

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{i<j} a_i a_j\, cov(X_i, X_j).$$

Theorem 4.

Let X, Y be r.v.'s. Then a) $-1 \le \rho(X,Y) \le 1$.

$|\rho(X,Y)| \le 1$ $((\rho(X,Y))^2 \le 1)$.

b) $|\rho(X,Y)| = 1 \Leftrightarrow \exists\, a, b \in \mathbb{R}, a \ne 0$, s.t. $Y = aX + b$.

Remark

$\rho(X,Y)$ measures the linear trend between X, Y. The closer $\rho(X,Y)$ is to ± 1, the "more linear" the relationship between X and Y is.

$\rho(X,Y) = 1 \Rightarrow$ perfect pos. correlation.
$\rho(X,Y) = -1 \Rightarrow$ perfect neg correlation.

Def

Let X be a r.v. with cdf $F:\mathbb{R}\to\mathbb{R}$ and let $\alpha \in (0,1)$. A quantile ( ) of order α is a nr q_α satisf.

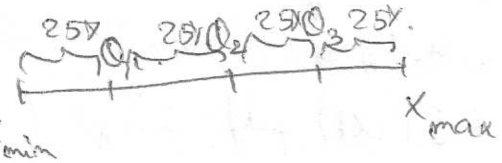
$P(X < q_\alpha) \le \alpha \le P(X \le q_\alpha)$, or equiv. $F(q_\alpha) \le \alpha \le F(q_\alpha + 0)$.

Of these,

- the median $m = q_{1/2}$.
- the quartiles. $Q_1 = q_{\frac{1}{4}}$

$Q_2 = m = q_{1/2}$

$Q_3 = q_{3/4}$.



Remark.

If X is a discr r.v, some quantiles may not be unique.

If X is a cont r.v, then F is cont, so $F(q_\alpha) = F(q_\alpha + 0)$ so $(*)$ is equiv. to

$F(q_\alpha) = \alpha$. $q_\alpha = F^{-1}(\alpha)$. So q_α is uniquely det, $\forall \alpha \in (0,1)$.

4. Inequalities.

- may help approx. probabilities or num. char.'s associated with r.v.'s.

① Hölder's Ineq.

X, Y r.v's, $p, q > 1$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$. Then $E(|XY|) \le (E(|X|^p))^{1/p} \cdot (E(|Y|^q))^{\frac{1}{q}}$

② Schwartz's Ineq.

Let $p = q = 2$ in ①.

$E(|XY|) \le \sqrt{E(X^2)}\sqrt{E(Y^2)}$.

③ Cauchy - Buneakowsky Ineq.

Let $Y \equiv 1$ in ②. $(Y(!))$.

$E(|X|) \le \sqrt{E(X^2)}$

④ Minkowsky's Ineq.

X, Y r.v's and $p \ge 1$.

$(E(|X+Y|^p))^{1/p} \le (E(|X|^p))^{\frac{1}{p}} + (E(|Y|^p))^{\frac{1}{p}}$.

5. Markov's Ineq.

Let X be a r.v, let $a > 0$. Then $P(|X| \ge a) \le \frac{1}{a} E(|X|)$.

$\left(P(|X| < a) \ge 1 - \frac{1}{a} E(|X|) \right)$.

Cons. the event $A = \{e \in S \mid |X(e)| \geq a\}$.

Cons. its indicator r.v. $I_A(e) = \begin{cases} 0, & |X(e)| < a \\ 1, & |X(e)| \geq a \end{cases}$

$a I_A(e) = \begin{cases} 0, & |X(e)| < a \\ a, & |X(e)| \geq a \end{cases}$

We want to compare $a I_A$ and $|X|$ (as r.v.'s). Let $e \in S$: if $|X(e)| < a$, then $a I_A(e) = 0 \leq |X|(e)$.

if $|X(e)| \geq a$ then $a I_A(e) = a \leq |X|(e)$.

So, either way, $a I_{A}(e) \leq |X|(e)$, $\forall e \in S$, so $a I_A \leq |X|$.

Then $E(a I_A) \leq E(|X|)$, pdf $a I_A = \begin{pmatrix} 0 & a \\ 1-P(A) & P(A) \end{pmatrix}$.

$E(a I_A) = a P(A)$. So $a P(|X| \geq a) \leq E(|X|)$.

$P(|X| \geq a) \leq \frac{1}{a} E(|X|)$.

6. Chebyshev's Ineq.

Let X be a r.v. and let $\varepsilon > 0$. Then $P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} V(X)$.

$\left(\qquad < \qquad \geq 1 - \frac{1}{\varepsilon^2} V(X) \right)$

Pf.

⑤ for $(X - E(X))^2$ and $a = \varepsilon^2 \Rightarrow P(|X - E(X)| \geq \varepsilon) = P((X - E(X))^2 \geq \varepsilon^2) \leq$

$\leq \frac{1}{\varepsilon^2} \underbrace{E((X - E(X))^2)}_{V(X)}$.

Ch V. Sequences of Random Variables;

Laws of Large Numbers and Limit Theorems.

Convergence of Sequences of R.V.'s.

Let $\{X_n\}_{n \in \mathbb{N}}$ of r.v.'s with cdfs F_n and X a r.v. with cdf F.

Def:

① $\{X_n\}_n$ conv. in probability to X, $X_n \xrightarrow{P} X$, if $\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1, \forall \varepsilon > 0$.

$\left(\qquad \geq \varepsilon) = 0 \right)$.

② $\{X_n\}_n$ conv. almost surely to X, ($X_n \xrightarrow{a.s.} X$) if $P(\lim_{n\to\infty} X_n = X) = 1$.

$\left(P(\qquad \neq X) = 0 \right)$

③ $\{X_n\}_n$ conv. in distribution to X, $X_n \xrightarrow{d} X$ ("în repartiţie", $X_n \xrightarrow{r} X$), if $\lim_{n\to\infty} F_n(x) = F(x)$, $\forall x$ a point of continuity of F.

④ $\{X_n\}_n$ conv in mean of order r to X, $X_n \xrightarrow{L^r} X$ if $\lim_{n\to\infty} E(|X_n - X|^r) = 0$. $(0 < r < \infty)$.

Rem. Special case for ④, $r=2$. mean square conv.

Theorem 1

Let $\{X_n\}_n$ be a seq. of r.v.'s and X a r.v. Then the following are equiv:

a) $X_n \xrightarrow{a.s} X$. (almost surely = a.s)

b) $\lim_{n\to\infty} P(\sup_{k\geq n} |X_k - X| < \varepsilon) = 1$, $\forall \varepsilon > 0$.

($P($ $\geq \varepsilon) = 0$)

Theorem 2.

Same setup. Then a) $X_n \xrightarrow{a.s} X. \Rightarrow X_n \xrightarrow{P} X$.

b). $X_n \xrightarrow{L^r} \Rightarrow X_n \xrightarrow{P} X$.

c) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$

Pf (selected)

a) $X_n \xrightarrow{a.s} X$, i.e (by T1). $\lim_{n\to\infty} (\sup_{k\geq n} |X_k - X| < \varepsilon) = 1$, $\forall \varepsilon > 0$.

Want. $X_n \xrightarrow{P} X$, i.e. $\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$, $\forall \varepsilon > 0$. (sup - supremum the largest val)

Fix. $\varepsilon > 0$. If $\sup_{k\geq n} |X_k - X| < \varepsilon$, then $|X_n - X| < \varepsilon$. As events

$\{\sup_{k\geq n} |X_k - X| < \varepsilon\} \subseteq \{|X_n - X| < \varepsilon\}$.

Remark

1°. ② $\Rightarrow$ ① $\Rightarrow$ ③.

④ $\nearrow$.

2°. In general, the reverse implications are not true.

3°. $X_n \xrightarrow{d} \cancel{X} a$. (a const) $\Rightarrow X_n \xrightarrow{P} a$.

$X_n \xrightarrow{P} X \Rightarrow \exists \{X_{n_k}\}_k \subseteq \{X_n\}_n$ s.t. $X_{n_k} \xrightarrow{a.s} X$.

Def. Two r.v.'s X, Y are almost surely equal. $X \overset{a.s}{=} Y$ if $P(X = Y) = 1$. $(P(X \neq Y) = 0)$.

Theorem 3

Let $\{X_n\}_n$ be a seq of r.v.'s, let X, Y be r.v.'s s.t. $X_n \xrightarrow{P} X$, $X_n \xrightarrow{P} Y$. Then. $X \overset{a.s}{=} Y$.

Pf

$|X - Y| = |X - X_n + X_n - Y| \leq |X_n - X| + |X_n - Y|$.

Fix. $\varepsilon > 0$.

If $|X_n - X| < \frac{\varepsilon}{2}$, $|X_n - Y| < \frac{\varepsilon}{2}$, then $|X - Y| < \varepsilon$.

As events $\{|X_n - X| < \frac{\varepsilon}{2}\} \cap \{|X_n - Y| < \frac{\varepsilon}{2}\} \subseteq \{|X - Y| < \varepsilon\}$.
Take $C\, A \subseteq B$, $\forall$. ...

Remark
T.3 holds for $\xrightarrow{a.s}$, $\xrightarrow{d}$, $\xrightarrow{L^r}$.

2. Laws of Large Numbers and Limit Theorems.
We are interested in the stability of the average of "large" nr. of observations.
$\frac{1}{n}\sum_{k=1}^{n} |X_k - E(X_k)| \xrightarrow{?} 0$

$\xrightarrow{P} 0$ $\{X_n\}_n$ follows the WLLN (weak law of large numbers.

$\xrightarrow{a.s} 0$ $\{X_n\}_n$ follows the SLLN (strong ...)

Let $\{X_n\}_n$ be a seq. of indep r.v.'s. Notations :- $m_k = E(X_k)$ ($\forall k$).
- $\sigma_k^2 = V(X_k)$ (σ_k stand. dev.).
- $\sigma_{(n)}^2 = \sum_{k=1}^{n} \sigma_k^2$

Define: $X_{(n)} = \frac{1}{\sigma_{(n)}} \sum_{k=1}^{n} (X_k - m_k)$ (reduced variable. $\frac{X - E(X)}{\sigma(X)}$)).

Notice
$E(X_{(n)}) = \frac{1}{\sigma_{(n)}} \sum_{k=1}^{n} (E(X_k) - m_k) = 0$. $V(X_{(n)}) = \frac{1}{\sigma_{(n)}^2} \sum_{k=1}^{n} V(X_k) = 1$.
Limit Problem to find X s.t. $X_{(n)} \xrightarrow{d} X$. the answer a limit theorem. If X has a normal distr $\Rightarrow$ central limit theorem.

Theorem (CLT).
Let $\{X_k\}_n$ be a seq of indep and identically distributed r.v.s with $\underset{=m}{E(X_k)} < \infty$, $V(X_k) < \infty$. Then, $X_{(n)} \xrightarrow{d} X \in N(0,1)$

C8 Ch.7. Estimation

7.1 Sample Theory
Notations: Population: characteristic X, $\overset{\text{mean}}{E(X) = \mu}$, $\overset{\text{var.}}{V(X) = \sigma^2}$
Sample of size n.: n. random var's $X_1 \ldots X_n$ indep and having the same pdf as X.
- sample fns $h(X_1, \ldots X_n)$ are random variables.

- sample mean, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- sample moments: $\overline{\nu}_k = \frac{1}{N}\sum_{i=1}^{n} X_i^k$,
- sample central moments: $\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k$

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

Notations.

	Pop	Sample
Mean	$\mu = E(X)$	$\overline{X}$
Var	$\sigma^2 = V(X)$	S^2
Std dev	σ	S
Moment	$\nu_k = E(X^k)$	$\overline{\nu}_k$
C. Moment	$\mu_k = E(X - E(X^k))$	$\overline{\mu}_k$

Prop 1 Let X be with $E(X) = \mu$, $V(X) = \sigma^2$. Then $E(\overline{X}) = \mu$, $V(\overline{X}) = \frac{\sigma^2}{n}$

Pf.

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E(X_i)}_{E(X) = \mu} = \frac{1}{n}\cdot n\mu = \mu.$$

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{V(X_i)}_{V(X) = \sigma^2} = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

Prop 2. $E(\overline{\nu}_k) = \nu_k$ ...

Prop 3. $E(S^2) = \sigma^2$.

- X has pdf depending on some parameter θ, unknown (to be estimated) pdf $f(x;\theta)$

θ - target parameter. sample of size n: $X_1 \dots X_n$ sample variables, indep and having the same pdf as X ($f(x;\theta)$).

2 types of est:

- point estimate

$\overline{\theta} = \overline{\theta}(X_1 \dots X_n)$.

- interval estimate

$\theta \in [\overline{\theta}_L, \overline{\theta}_U]$. $\overline{\theta}_L(X_1 \dots X_n)$

$\overline{\theta}_U(X_1 \dots X_n)$

Point Estimators.

Def: An estimator $\overline{\theta}(X_1 \dots X_n)$ is called unbiased for (the estimation of) θ if

$E(\overline{\theta}) = \theta$

the bias of $\overline{\theta}$ is $B(\overline{\theta}) = |\theta - E(\overline{\theta})|$.

Recall:

- $\overline{X}$ is unbiased estimator for μ.
- $\overline{\nu}_k$ is unbiased estimator for ν_k.
- S^2 is an unbiased estimator for σ^2

Another trait for an unbiased est that its values do not vary too much from its mean.

2. Estimation; Confidence.
Intervals – General Frameworks.

pop. char. X whose pdf
$f(X;\theta)$ θ – target parameter
Sample of size n: n sample var's $X_1 \ldots X_n$ indep. and have the same pdf as X, $f(X;\theta)$

Point estimator $\bar{\theta} = \bar{\theta}(X_1, \ldots X_n)$. $\bar{\theta}$ has a mean, $E(\bar{\theta})$ and a variance $\sigma^2_{\bar{\theta}}$ ($\sigma_{\bar{\theta}}$ std. deviation).

- $\bar{\theta}$ unbiased if $E(\bar{\theta}) = \theta$
- the standard error $\sigma_{\bar{\theta}}$ to be small, i.e. $\sigma_{\bar{\theta}}$ becomes smaller as n becomes larger.

Find an interval estimate: find $\bar{\theta}_L, \bar{\theta}_U$ s.t.

$$P(\theta \in (\bar{\theta}_L, \bar{\theta}_U)) = \boxed{1-\alpha} \leftarrow \text{given} \quad (*)$$

- $(\bar{\theta}_L, \bar{\theta}_U)$ - conf interval. $100(1-\alpha)\%$ c.i.
- $\bar{\theta}_L, \bar{\theta}_U$ - conf. limits.
- $1-\alpha$ - conf coeff (level).
- α significance level.

Remarks

1° Relation $(*)$ does not uniquely determine the c.i.

2° This is a two-sided c.i. sometimes it is enough to find a one-sided ci $(-\infty, \bar{\theta}_U)$ $(\bar{\theta}_L, \infty)$.

General Framework

We use the pivotal method:
We have a pivot, W, (a sample function) satisf:
1) W depends on $X_1, \ldots X_n$, other known quantities and θ (the only unknown is θ)
2) its pdf (and, hence, cdf) is known and does not dep. on θ.

For ex, for $\theta = \mu$ (pop. mean), if the sample is large enough ($n > 30$) or comes from an approx normal pop and, σ (pop. std. dev) is known then.

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \in N(0,1).$$

Let us assume that for a target param. θ and an est $\bar{\theta}$ with mean $E(\bar{\theta}) = \theta$, and std. error $\sigma_{\bar{\theta}}$, the r.v.

$$Z = \frac{\bar{\theta} - \theta}{\sigma_{\bar{\theta}}} \in N(0,1).$$

We use Z as a pivot start with 2 values. z_L, z_U s.t. $P(z_L < Z < z_U) = 1-\alpha$

We choose these 2 values z_L, z_U using the symmetry of $N(0,1)$.

(Recall quantiles of order $\alpha \in (0,1)$: q_α, $P(X < q_\alpha) = \alpha$.)

$z_L = z_{\frac{\alpha}{2}}$ and $z_U = z_{1-\frac{\alpha}{2}}$

$$1-\alpha = P(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) = P\left(z_{\frac{\alpha}{2}} < \frac{\bar{\theta}-\theta}{\sigma_{\bar{\theta}}} < z_{1-\frac{\alpha}{2}}\right) =$$

$$= P(\sigma_{\bar{\theta}} \cdot z_{\frac{\alpha}{2}} < \bar{\theta} - \theta < \sigma_{\bar{\theta}} \cdot z_{1-\frac{\alpha}{2}}) = P(\bar{\theta} - \sigma_{\bar{\theta}} \cdot z_{1-\frac{\alpha}{2}} < \theta < \bar{\theta} - \sigma_{\bar{\theta}} \cdot z_{\frac{\alpha}{2}}).$$

<u>Remark</u>

$z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$.

So the $100(1-\alpha)\%$ c.i.

$$\theta \in (\bar{\theta} - \sigma_{\bar{\theta}} \cdot z_{1-\frac{\alpha}{2}}, \bar{\theta} - \sigma_{\bar{\theta}} \cdot z_{\frac{\alpha}{2}}).$$

~~So~~ $\bar{\theta} \pm \sigma_{\bar{\theta}} z_{1-\frac{\alpha}{2}}$ only for sym. distr.'s. $N(0,1)$, $T(n-1)$.

3. C.I's for the Parameters of One Pop.

1. C.I for the pop. mean, $\theta = \mu$.

If either the sample is large enough ($n > 30$) or drawn from an approx. normal pop.

a) <u>σ known</u>

pivot $Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \in N(0,1)$. | c.i. $\mu \in (\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}})$ (sym)

b) <u>σ unknown</u>

pivot $T = \dfrac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \in T(n-1)$ | c.i $\mu \in (\bar{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}, \bar{X} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}})$ (sym.)

2. <u>c.i for the pop. variance, $\theta = \sigma^2$</u>

If the sample is drawn from the normal pop.

pivot $\chi^2 = \dfrac{(n-1) s^2}{\sigma^2} \in \chi^2(n-1)$. | c.i $\sigma^2 \in \left(\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}, \dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right)$ (not sym).

$(\sigma \in (\sqrt{}, \sqrt{}))$.

<u>Ex</u> ① The shopping times of 64 rand. selected costumers at a local supermarket were recorded. These were found to have an average of 33 mins. and a variance of 256 mins. Find a 95% c.i for the time average shopping time per costumer.

Sol $\theta = \mu$, $n = 64$, $\bar{X} = 33$, $s^2 = 256$.

$100(1-\alpha)\% = 95\% \Leftrightarrow 1-\alpha = 0.95 \Rightarrow \alpha = 0.05$.

$t_{1-\frac{\alpha}{2}} = t_{0.975} = 1.9983$ (tinv (0.975, 64)).

$t_{\frac{\alpha}{2}} = t_{0.025} = -1.9983$.

$\mu \in (29.0033, 36.9967)$. So, on average, a costumer spends between 29.0033 and 36.9967 mins in the supermarket and we are 95% sure of it.

4. C.I.'s for Comparing Two Populations.

Two characteristics

$X_{(1)}$, with mean μ_1, var. σ_1^2

$X_{(2)}$, with mean μ_2, var. σ_2^2

Two indep selections

$X_{11}, \ldots, X_{1n_1}$ from the 1st pop. size n_1, $\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}$.

$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2$.

same for 2nd pop

In addition: pooled variance (var. centralizata).

$$s_p^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \bar{X}_2)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

1 C.I. for pop. means, $\theta = \mu_1 - \mu_2$.

- either large samples ($n_1 + n_2 > 40$) or normal underlying pop ($N(\mu_1, \sigma_1)$, $N(\mu_2, \sigma_2)$).

a) σ_1, σ_2 known.

pivot $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \in N(0,1).$$ $\boxed{\text{sym}}$

b) $\sigma_1 = \sigma_2$ unknown

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \in T(n_1 + n_2 - 2).$$ $\boxed{\text{sym}}$

c) $\sigma_1 \neq \sigma_2$ unknown.

$$T^* = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \in T(n)$$ $\boxed{\text{sym}}$

2 C.I. for pop. variances, $\theta = \frac{\sigma_1^2}{\sigma_2^2}$. normal pop.

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \in F(n_1 - 1, n_2 - 1)$$ $\boxed{\text{not sym}}$.

Ch VIII Hypothesis Testing

1. Basic concepts

So far, we estimated pop parameters, based solely on sample data, without any preconceived notion about those target parameters.

In contrast, here, we start with such a preconceived notion.

- a stat. hyp.: an assumption about some pop. char. X.
- a hyp test: the method(s) used to decide if a hyp is true or not.

2 hyp

- H_0: the null hyp: what is believed to be known or true.
- H_1: the alternative (research) hyp: something new, proposed by the experiment. (H_a)

The goal: to decide to reject H_0 (in favor of H_1) or not.

- test statistic (TS): a stat used to determine the value of truth of a hyp.
- rejection region (RR) or critical region: will consist of the values of TS for which H_0 is rejected.

* If the test refers to a pop. parameter $\Rightarrow$ par. test, otherwise nonparam. test.

For a param. test, $\theta \in A = A_0 \cup A_1$, $A_0 \cap A_1 = \emptyset$, $H_0: \theta \in A_0$,
$H_1: \theta \in A_1$

If $A_0 = \{\theta_0\} \Rightarrow$ simple hyp. We only consider simple null hyp.

So: $H_0: \theta = \theta_0$, with one of the alt.:

$$H_1: \begin{cases} \theta < \theta_0 & \text{left-tailed test} \\ \theta > \theta_0 & \text{right-tailed test} \\ \theta \neq \theta_0 & \text{two-tailed test} \end{cases}$$

For any hyp test, we have the foll. possibilities.

Decision \ Actual sit.	H_0 true	H_1 true
Reject H_0	Type I error α	correct decision
Do not Reject H_0	correct decision	Type II error β

$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \mid H_0) =$
$= P(TS \in RR \mid H_0)$

α - significance level $\theta = \theta_0$.

$\beta = P(\text{type II error}) = P(\text{not reject } H_0 \mid H_1) =$
$= P(TS \notin RR \mid H_1)$.

As one decreases, the other increases, but both decrease as n gets larger.

In general, $\alpha \in (0,1)$ is given and RR is determined to make β as small as possible.

1. Cont.

θ target parameter

null H_0 $\theta = \theta_0$ (simple hyp)

alt H_1 one of $\begin{cases} \theta < \theta_0 \text{ (left tailed)} \\ \theta > \theta_0 \text{ (right tailed)} \\ \theta \neq \theta_0 \text{ (two-tailed)} \end{cases}$

GOAL: to decide if - we reject H_0 (in fav of H_1).
- do not reject H_0.

Given $\alpha \in (0,1)$ - significance level.

$\alpha = P$ (type I error) $= P$ (reject H_0 / H_0).

$\beta = P$ (type II error) $= P$(do not reject H_0 / H_1).

- test statistic, TS. (same as the pivot).
- rejection region, RR (critical).

$\alpha = P(TS \in RR \mid \theta = \theta_0)$.

Suppose that for the t. par. θ, we have an unbiased point estimator $\bar{\theta}(X_1, \ldots X_n)$. $(E(\bar{\theta}) = \theta)$ with standard error $\sigma_{\bar{\theta}}$ s.t. $Z = \dfrac{\bar{\theta} - \theta}{\sigma_{\bar{\theta}}} \in N(0,1)$.

We choose: $TS = Z$.

- observed value of the TS $\quad \overset{Z_0:}{TS_0} = TS(\theta = \theta_0)$.

We want $\alpha = P(\text{reject } H_0 \mid H_0) = P(Z_0 \in RR \mid \theta = \theta_0) = P(Z \in N(0,1))$.

<u>For a left-tailed test</u>

$H_0 : \theta = \theta_0$.

$H_1 : \theta < \theta_0$.

We should reject the H_0 if the values of θ are "far away" from θ_0 (in the sense of a left-tailed test).

$\alpha = P(Z \in RR \mid \theta = \theta_0)$.

We choose a RR of the form $RR = \{Z < k_{left}\} = (\ldots, k_{left})$.

$\alpha = P(Z_0 < k_{left})$.

$RR_{left} = \{Z_0 < z_\alpha\}$,

<u>Right-tailed Test</u>

$H_0 : \theta = \theta_0$

$H_1 : \theta > \theta_0$.

$RR = (k_{right}, \infty)$.

$\alpha = P(Z_0 > k_{right})$.

<u>Two-Tailed Test</u>

$\begin{cases} H_0 : \theta = \theta_0. \\ H_1 : \theta \neq \theta_0 \end{cases}$

$\alpha = P(Z_0 < k_1 \text{ or } Z_0 > k_2)$.

$1 - \alpha = P(k_1 < Z < k_2)$. $\quad$ sym. $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$.

$RR = (-\infty, k_{\alpha/2}) \cup (z_{1-\frac{\alpha}{2}}, \infty)$. $\quad RR = \{|Z_0| > z_{1-\frac{\alpha}{2}}\}$.

2. Hyp. Tests for the Parameters of One Pop.

X pop. char. with $E(X) = \mu$.

$V(X) = \sigma^2$.

$X_1, \dots X_n$ sample var's.

[1] Tests for the mean $\theta = \mu$.

$H_0: \mu = \mu_0$.

$$H_1: \begin{cases} \mu < \mu_0. \\ \mu > \mu_0. \\ \mu \neq \mu_0 \end{cases}$$

If the sample is large ($n > 30$) or from a normal pop.

a) σ known. (z test)

$$TS = Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \in N(0,1).$$

$$TS_0 = Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$RR: \begin{cases} z_0 < z_\alpha & \text{left} \\ z_0 > z_{1-\alpha} & \text{right.} \\ |z_0| > z_{1-\frac{\alpha}{2}} & \text{two} \end{cases}$$

b) σ unknown. (t test).

$$TS = T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \in T(n-1).$$

$$TS_0 = T_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}.$$

$$RR: \begin{cases} T_0 < t_\alpha, & \text{left} \\ T_0 > t_{1-\alpha}, & \text{right.} \\ |T_0| > t_{1-\frac{\alpha}{2}}, & \text{two} \end{cases}$$

[2] Tests for the pop. variance $\theta = \sigma^2$ (var test)

If the sampl comes from a normal distr.

$H_0: \sigma^2 = \sigma_0^2$. $\qquad \sigma = \sigma_0$.

$$H_1: \begin{cases} \sigma^2 < \sigma_0^2 & \sigma < \sigma_0 \\ \sigma^2 > \sigma_0^2 & \sigma > \sigma_0 \\ \sigma^2 \neq \sigma_0^2 & \sigma \neq \sigma_0. \end{cases}$$

$$RR: \begin{cases} \chi_0^2 < \chi^2_\alpha. \\ \chi_0^2 > \chi^2_{1-\alpha}. \\ \chi_0^2 < \chi^2_{\frac{\alpha}{2}} \text{ or } \chi_0^2 > \chi^2_{1-\frac{\alpha}{2}} \end{cases}$$

$$TS = \chi^2 = \frac{(n-1)S^2}{\sigma^2} \in \chi^2(n-1).$$

$$TS_0 = \chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \in \boxed{\text{not sym } \chi^2(n-1).}$$

EX:

① The number of monthly sales at a large corporation is known to have an average of 20 and a std dev of 4 (and all salary, tax, bonus figures are based on these). However, in times of econ recession, a VP fears that his people do not meet this standard (while his employees think otherwise). A nr of 36 salespeople are selected randomly and for those, it was found that they averaged 19 sales in one month.
At the 5% sign. level, does the data confirm a contradict the V.P.'s suspicion?

$\begin{cases} H_0 : \mu = 20 \\ H_1 : \mu < 20. \end{cases}$

$\sigma = 4$

$n = 36, \; \overline{x} = 19.$

$\alpha = 0,05.$

$z_{0.05} = -1.645$ (= norminv (0.05, 0, 1)).

$RR = (-\infty, -1.645).$

$$Z_0 = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{19-20}{\frac{4}{6}} = -\frac{3}{2} = -1.5$$

$Z_0 \notin RR \Rightarrow$ do not reject H_0

i.e. the data seems to contradict the V.P's suspicion.

3. Significance testing, P-values.

Situation: we set up a test, det a RR and we find that the obs. value of the test stat., TS_0 does not belong to RR, hence we can not reject H_0. However, when we compute the prob of TS getting that value or more extreme, under the assumption that H_0 is true, we find it very small.

To avoid such situations we compute the prob. of TS taking values at least as "extreme" as TS_0.

If P is small, we reject H_0. $\begin{cases} P \le \alpha \Rightarrow \text{reject } H_0 \\ P > \alpha \Rightarrow \text{do not reject.} \end{cases}$

Remark

1° supp. the TS has cdf F.

- left-tailed

$P = P(TS < TS_0) = F(TS_0).$

- right-tailed

$P = P(TS > TS_0) = 1 - F(TS_0).$

- two-tailed

$P = 2 * \min\{P(TS < TS_0), P(TS > TS_0)\} = 2 \min\{F(TS_0), 1 - F(TS_0)\}.$

2°. P can be thought of as the minimum rejection sign. level.

EX:

② the previous example.

$Z_0 = -1.5$

$\alpha = 0.05.$

$P = P(Z < -1.5)$ (= normcdf (-1.5, 0, 1)) = 0.0668.

$P > \alpha \Rightarrow$ do not reject H_0.

4. Tests for Comparing 2 Populations

1. $\theta = \mu_1 - \mu_2$

$H_0: \mu_1 - \mu_2 = 0. \quad \mu_1 = \mu_2.$

$$H_1: \begin{cases} < 0, & < \\ > 0 & > \\ \neq 0 & \neq. \end{cases}$$

a) σ_1, σ_2 known. $N(0,1)$

b) $\sigma_1 = \sigma_2$ unknown. $T(n_1 + n_2 - 2)$ (ttest 2).

c) $\sigma_1 \neq \sigma_2$ unknown $T(n)$ (ttest 2 "unequal").

2. $\theta = \frac{\sigma_1^2}{\sigma_2^2}$. (vartest 2)

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1, \quad \sigma_1^2 = \sigma_2^2. \quad \sigma_1 = \sigma_2$

$$H_1: \begin{cases} < & < & < \\ > & > & > \\ \neq & \neq. & \neq. \end{cases}$$

$F(n_1 - 1, n_2 - 1)$.

COURSE 11 - P&S

Back to Chapter VII

5. Properties of Point Estimators

5.1. Minimum Variance and Efficient Estimators

Recall: X population characteristic whose pdf $f(X,\theta)$ depends on a parameter θ (target parameter)

X has mean $\mu = E(X)$

variance $\sigma^2 = V(X)$

std. deviation $\sigma = \sqrt{V(X)}$

Sample of size n: $X_1, \dots, X_n$ sample variables, independent and identically distributed i.e. with the same pdf as X, $f(x,\theta)$,

estimator: $\overline{\theta} = \overline{\theta}(X_1, \dots, X_n) \Rightarrow \overline{\theta}$ has a mean, $E(\overline{\theta})$, a variance $V(\overline{\theta})$

std. deviation $\sigma_{\overline{\theta}} = \sqrt{V(\overline{\theta})}$ (standard error of $\overline{\theta}$)

- unbiased estimator if $E(\overline{\theta}) = \theta$

Another desirable trait of an unbiased estimator is that its values do not vary too much from its mean value, i.e., to have a small variance.

Def: An unbiased estimator is called minimum variance unbiased estimator (MVUE) (estimator optimal) if $V(\overline{\theta}) \le V(\tilde{\theta})$, $\forall \tilde{\theta}$ unbiased estimator.

Remark: It can be shown that if an unbiased estimator exists, then an MVUE also exists and it is unique.

Def:

5.2. Efficient Estimators

__Def__: The __likelihood function__ (funcția de verosimilitate) of a sample $X_1, \dots, X_n$ is the joint pdf of the vector $(X_1, \dots, X_n)$, i.e. $L(X_1, \dots, X_n; \theta) = f(X_1; \theta) \dots f(X_n; \theta) = \prod_{i=1}^{n} f(X_i, \theta)$ with value $\prod_{i=1}^{n} f(x_i; \theta)$

__Def__: For a sample of size n, the (Fisher) information of the sample relative to a target parameter θ is the quantity $I_n(\theta) = E\left[\left(\frac{\partial \ln L(X_1, \dots X_n; \theta)}{\partial \theta}\right)^2\right]$

$$I_n(\theta) = E\left[\left(\frac{\partial \ln L(X_1, \dots, X_n; \theta)}{\partial \theta}\right)^2\right]$$

__Prop 1__: If the range of X does __not__ depend on θ and L is twice differentiable with respect to θ, then:

$$I_n(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

__Corolary 2__: If the range of X does not depend on θ, then

$$I_n(\theta) = n I_1(\theta)$$

__Proof__: $L(X_1, \dots, X_n, \theta) = \prod_{i=1}^{n} f(X_i; \theta)$

$$\ln L = \sum_{i=1}^{n} \ln f(X_i, \theta)$$

$$\frac{\partial \ln L}{\partial \theta} = \sum_{1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \sum_{1}^{n} \frac{\partial^2 \ln f(X_i, \theta)}{\partial \theta^2}$$

$$-E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right] = \sum_{1}^{n} \underbrace{\left(-E\left(\frac{\partial^2 \ln f(X_i, \theta)}{\partial \theta^2}\right)\right)}_{I_1(\theta)}$$

$$I_n(\theta) = n I_1(\theta)$$

<u>Def</u>: An estimator $\bar{\theta} = \bar{\theta}(X_1, \dots, X_n)$ is called <u>absolutely correct</u> if (i) $E(\bar{\theta}) = \theta$ (unbiased estimator)

(ii) $\lim\limits_{n \to \infty} V(\bar{\theta}) = 0$

<u>Theorem 3</u> (Crammer-Rao Inequality)

Under our assumptions, if $\bar{\theta}$ is an absolutely correct estimator, then $V(\bar{\theta}) \geq \dfrac{1}{I_n(\theta)}$

<u>Def</u>: The <u>efficiency</u> of an absolutely correct estimator $\bar{\theta}$,

$$(\text{eff}(\bar{\theta}))\ e(\bar{\theta}) = \frac{1}{V(\bar{\theta})\, I_n(\theta)}$$

<u>Remarks</u>:

1° By theorem 3, we have $e(\bar{\theta}) \leq 1$, so an efficient estimator has the maximum efficiency possible.

2° An efficient estimator may not always exist, but when it does exist, it is unique and it is also a MVUE (the MVUE is not necessarily efficient)

<u>Ex</u>: ① Let X be a characteristic where pdf is given by $f(x_i, \theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}}$ $x > 0$, where $\theta > 0$ is unknown. For a sample $X_1, \dots, X_n$ consider the estimator $\bar{\theta} = \frac{1}{2}\bar{X} \left(= \frac{1}{2} \cdot \frac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i \right)$ show that:

a) θ is absolutely correct

b) find its efficiency, $e(\bar{\theta})$

<u>Sol</u>:

<u>Preliminaries</u>: [Sem 1] Euler's Gamma Function

$$\Gamma : (0, \infty) \to (0, \infty)$$

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

Prop: $\Gamma(1) = 1$

$\Gamma(n+1) = n!,\ \forall n \in \mathbb{N}$

Let's see that $f(x;\theta)$ is indeed a pdf: $\int_{\mathbb{R}} f(x;\theta)\,dx = 1$

$\frac{1}{\theta^2}\int_0^\infty x\cdot e^{-\frac{x}{\theta}}\,dx$ → change of variable:

$$t = \frac{1}{\theta}x \Rightarrow x = \theta t$$
$$dx = \theta\,dt$$
$$(0,\infty) \to (0,\infty)$$

$$\frac{1}{\cancel{\theta^2}}\int_0^\infty (\theta t)e^{-t}(\cancel{\theta}\,dt) = \int_0^\infty t e^{-t}\,dt = \Gamma(2) = 1 \checkmark$$

a) We need $E(X)$, $V(X)$

$$E(X) = \int_{\mathbb{R}} x f(x)\,dx = \frac{1}{\theta^2}\int_0^\infty x^2 e^{-\frac{x}{\theta}}\,dx \quad \text{same change of variable}$$

$$E(X) = \theta\int_0^\infty t^2 e^{-t}\,dt = \theta\,\Gamma(3) = 2\theta$$

$$V(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = \int_{\mathbb{R}} x^2 f(x)\,dx = \frac{1}{\theta^2}\int_0^\infty x^3 e^{-\frac{x}{\theta}}\,dt \quad \text{same change of variable}$$

$$E(X^2) = \theta^2\int_0^\infty t^3 e^{-t}\,dt = \theta^2\Gamma(4) = 6\theta^2$$

$$V(X) = 6\theta^2 - (2\theta)^2 = 6\theta^2 - 4\theta^2 = 2\theta^2$$

$$\boxed{\begin{array}{l} E(X) = 2\theta \\ V(X) = 2\theta^2 \end{array}}$$

absolutely correct estimator: (i) $E(\bar{\theta}) = \theta$

(ii) $\lim_{n\to\infty} V(\bar{\theta}) = 0$

i) $E(\bar{\theta}) = E\left(\frac{1}{2}\bar{X}\right) = \frac{1}{2}E(\bar{X}) = \frac{1}{2}E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{2n}\sum_{i=1}^{n}\underbrace{E(X_i)}_{=E(X)} =$

$= \frac{1}{\cancel{2n}}\cdot\cancel{n}\cdot\cancel{2}\theta = \theta \checkmark$

ii) $V(\bar{\theta}) = V\left(\frac{1}{2}\bar{X}\right) = \frac{1}{4}V(\bar{X}) = \frac{1}{4}\frac{V(X)}{n} = \frac{1}{\cancel{4}_2 n}\cdot\cancel{2}\theta^2 = \frac{\theta^2}{2n} \xrightarrow{n\to\infty} 0$

b) The range of X does not depend on θ, so $e(\bar{\theta}) = \dfrac{1}{I_n(\theta)\cdot V(\bar{\theta})}$

$$\boxed{\begin{array}{l} \ln(ab) = \ln a + \ln b \\ \ln\left(\frac{a}{b}\right) = \ln a - \ln b \\ \ln a^n = n \ln a \\ (\ln(g(x)))' = \frac{g'(x)}{g(x)} \\ \hline \ln(e^A) = A \end{array}}$$

$I_n(\theta) = n I_1(\theta)$

$I_1(\theta) = E\left(\dfrac{\partial \ln f(x_i,\theta)}{\partial \theta^2}\right)$

$f(x_i;\theta) = \frac{1}{\theta^2} x_i e^{-\frac{1}{\theta}x_i}$

$\ln f = -2\ln\theta + \ln X_i - \frac{1}{\theta}X_i$

$\dfrac{\partial \ln f}{\partial \theta} = -\dfrac{2}{\theta} + \dfrac{1}{\theta^2}X_i$

$\dfrac{\partial^2 \ln f}{\partial \theta^2} = \dfrac{2}{\theta^2} - \dfrac{2}{\theta^3}X_i$

$$I_1(\theta) = -E\left(\frac{2}{\theta^2} - \frac{2}{\theta^3}X_i\right) = -\frac{2}{\theta^2} + \frac{2}{\theta^3}\underbrace{E(X_i)}_{E(X)=2\theta} = -\frac{2}{\theta^2} + \frac{2}{\theta^3}\cdot 2\theta = \frac{2}{\theta^2}$$

$I_n(\theta) = \dfrac{2n}{\theta^2}$

$e(\bar{\theta}) = \dfrac{1}{\frac{2n}{\theta^2}\cdot\frac{\theta^2}{2n}} = 1 \Rightarrow \bar{\theta} = \frac{1}{2}\bar{X}$ is an efficient estimator, so also a MVUE

5.3. Methods of Pointwise Estimation

$\boxed{1}$ Method of Moments (K. Pearson ≈ 1800)

Idea: sample moments $(\bar{\nu}_k)$ are unbiased estimators for population moments (ν_k). The equate them: $\nu_k = E(X^k)$

$\bar{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

$\boxed{\nu_k = \bar{\nu}_k}$ $k = 1, 2, \dots$ as many as needed (nr. of unknowns)

$\boxed{2}$ Method of Maximum Likelihood

Intuitive ex: Suppose there are 3 balls in a box, black and white, the nr of each unknown. We randomly select 2, without replacement and get both white. What would be a good estimate, $\bar{w}$, for the nr. of white balls w?

Obviously, $\bar{w} < \frac{2}{3}$

We choose the estimator that makes our sample the most likely.

If $w = 2$, the probability of getting 2 white balls without replacement (hypergeometric model)

$$\frac{C_2^2 \, C_1^0}{C_3^2} = \frac{1}{3}$$

If $w = 3$, the probability is 1

So, we choose $\bar{w} = 3$

Formally, we choose parameters that maximize the likelihood function, i.e.

$\frac{\partial L}{\partial \theta_k} = 0$, $k = 1, 2, \ldots$ (nr. of parameters)

or equivalently, $\frac{\partial \ln L}{\partial \theta_k} = 0 \longrightarrow$ MLE (Maximum Likelihood Estimator)

<u>Ex</u>: Previous ex: $f(x_i;\theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}}$, $x > 0$, $\theta > 0$

$$E(X) = 2\theta \ , \quad V(X) = 2\theta^2$$

a) the method of moments estimator, $\bar{\theta}$

b) the MLE, $\hat{\theta}$

<u>Sol</u>:

a) $\nu_1 = \bar{\nu}_1$

$\nu_1 = E(X) = 2\theta$

$\bar{\nu}_1 = \bar{X}$

$2\bar{\theta} = \bar{X}$

$\bar{\theta} = \frac{1}{2}\bar{X}$

b) $L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta)$

$= \prod_{i=1}^{n} \left(\frac{1}{\theta^2} x_i e^{-\frac{1}{\theta} x_i} \right)$

$= \frac{1}{\theta^{2n}} \underbrace{\left(\prod_{i=1}^{n} x_i \right)}_{k \text{ (const. w.r.t. } \theta)} e^{-\frac{1}{\theta} \underbrace{\sum_{i=1}^{n} x_i}_{= n\bar{X}}}$

$= \frac{1}{\theta^{2n}} k e^{-\frac{n}{\theta}\bar{X}}$

$\ln L = \ln k - 2n \ln\theta - n\bar{X} \cdot \frac{1}{\theta}$

$\left.\begin{array}{l} \frac{\partial \ln L}{\partial \theta} = -2n \frac{1}{\theta} + \frac{n\bar{X}}{\theta^2} \\ \frac{\partial \ln L}{\partial \theta} = 0 \end{array}\right\} \Rightarrow \frac{2n}{\theta} = \frac{n\bar{X}}{\theta^2}$

$\hat{\theta} = \frac{1}{2}\bar{X}$

> <u>Remark</u>: Many times, the 2 methods yield the same results, if not, the method of maximum likelihood is more reliable.

Back to Chapter VIII

5. Type II Errors; Power of a Test; Neyman-Pearson Lemma:

$H_0: \theta = \theta_0$.

$H_1: \theta < \theta_0$ left

$\theta > \theta_0$ right.

$\theta \neq \theta_0$ both.

$\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0) = P(TS \in RR | \theta = \theta_0)$ - preset.

$\beta = P(\text{type II error}) = P(\text{do not reject } H_0 | H_1) = P(TS \notin RR | H_1)$.

We can only determine β if the alt H_1 is also a simple hyp.

$H_0: \theta = \theta_0$.

$H_1: \theta = \theta_1$ $\begin{pmatrix} < \theta_0 \\ > \theta_0 \\ \neq \theta_0 \end{pmatrix}$.

$\beta(\theta_1)$.

Ex.

① Recall the Vice President problem,

$H_0: \mu = 20 (= \mu_0)$. 5% sign. level

$H_1: \mu < 20$ $\alpha = 0.05$.

$n = 36$, $\bar{X} = 19$, $\sigma = 4$.

$TS = Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \in N(0,1)$.

observed value $TS_0 = Z_0 = Z(\mu = 20)$.

We found,

$RR = \{ Z_0 < \underset{\text{quantile}}{Z_{0.05}} \} = \left\{ \dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -1.645 \right\} = \left\{ \dfrac{\bar{X} - 20}{\frac{4}{6}} < -1.645 \right\}$

$= \{ \bar{X} < 20 - \frac{2}{3} \cdot 1.645 \} = \{ \bar{X} < 18.9 \}$

Conclusion: do not reject H_0. $(\bar{X} = 19)$.

Suppose:

$H_0: \mu = 20$ $(= \mu_0)$

$H_1: \mu = 18$ $(= \mu_1)$.

$\beta(\mu_1) = P(\text{do not reject } H_0 | H_1) = P(TS \notin RR | H_1) = P(\bar{X} > 18.9 | \mu = 18)$.

$= P\left(\dfrac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} > \dfrac{18.9 - \mu_1}{\frac{\sigma}{\sqrt{n}}} \right)$

Let $Z_1 = \dfrac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}}$.

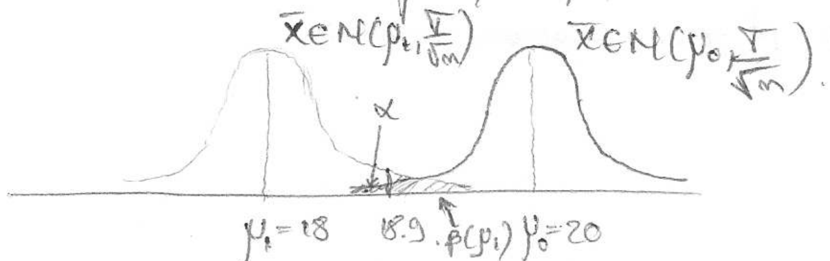
$\Rightarrow \beta(\mu_1) = P(Z_1 > 1.35 | \mu = \mu_1 = 18) =$

$= P(Z_1 > 1.35 | Z_1 \in N(0,1)) =$

$= 1 - P(Z_1 < 1.35) = 1 - \text{normcdf}(1.35, 0, 1)$

$= 0.0885$

Remark We used the fact that, under these conditions, $\overline{X} \in N(\mu, \frac{\sigma}{\sqrt{n}})$.



Def The power of a test for a param θ, unknown, is the prob of rejecting H_0, when the true value is $\theta = \theta^*$

$\pi(\theta^*) = P(\text{reject } H_0 \mid \theta = \theta^*) = P(TS \in RR \mid \theta = \theta^*)$.

$H_0: \theta = \theta_0$.

$H_1: \{ \ne, <, > \}$

$\theta = \theta_0$.

$\pi(\theta_0) = P(\text{reject } H_0 \mid H_0) = \alpha$.

For any other value $\theta_1 \ (\ne \theta_0)$.

$\pi(\theta_1) = P(\text{reject } H_0 \mid \theta = \theta_1) = 1 - \underbrace{P(\text{do not reject } H_0 \mid H_1)}_{\beta(\theta_1)} = 1 - \beta(\theta_1)$

So, making β small $\Leftrightarrow$ making π large.

$H_0: \theta = \theta_0$.

$H_1: \theta = \theta_1$.

We preset $\alpha = \pi(\theta_0)$ and determine RR so that $\pi(\theta_1)$ is maximum possible. $\rightarrow$ (a) most powerful test.

Theorem: (The Neyman - Pearson Lemma, NPL).

Let X be a char. with pdf $f(x, \theta)$, θ unknown. Suppose we test 2 simple hyp: $H_0: \theta = \theta_0$

$H_1: \theta = \theta_1$.

Let $L(x_1, \ldots x_n; \theta)$ be the likelihood function of the sample. Then for a fixed $\alpha \in (0, 1)$ a most powerful test is the one with RR

$$RR = \left\{ \frac{L(\theta_1)}{L(\theta_0)} \geq k_\alpha > 0 \right\}.$$

Ex

Ex. ②. Supp. X_1 represents a single observation from a distr. with pdf.

$$f(x;\theta)=\begin{cases}\theta\, x^{\theta-1} & , x\in(0,1).\\ 0 & ,\text{otherwise}\end{cases}$$

Find the NPL most powerful test that at 5% sign. level tests

$H_0: \theta=1 \quad (=\theta_0)$

$H_1: \theta=30 \quad (=\theta_1)$. Also find $\beta(\theta_1)$

(a sample of size $n=1$)

X_1.

Sol: $L(X_1,\dots X_n;\theta)=\prod_{i=1}^{n} f(X_i;\theta)$

In our case $n=1$, X_1.

$$\frac{L(\theta_1)}{L(\theta_0)}=\frac{f(X_1;\theta=\theta_1)}{f(X_1;\theta=\theta_0)}=\frac{30\cdot X_1^{29}}{1}\geq k_\alpha>0.$$

$$RR=\{30X_1^{29}\geq k_\alpha\}=\{X_1\geq \underbrace{(\tfrac{1}{30}k_\alpha)^{\frac{1}{29}}}_{K_\alpha}\}=\{X_1\geq K_\alpha\}$$

$\alpha=P(\text{reject } H_0\mid H_0)=P(X_1\geq K_\alpha\mid\theta=1)$ (X_1 same pdf with X.)

$$=\int_{K_\alpha}^{\infty} f_{X_1}(x;\underset{1}{\theta})\,dx=\int_{K_\alpha}^{1} 1\,dx=1-K_\alpha.\Rightarrow K_\alpha=0.95$$

$RR=\{X_1\geq 0.95\}$.

$$\beta(\theta_1)=P(\text{do not reject } H_0\mid H_1)=P(X_1<0.95\mid\theta=30)=\int_{\infty}^{0.95} f(x;\theta_1)\,dx=$$

$$=\int_{0}^{0.95}30\cdot x^{29}\,dx=x^{30}\Big|_0^{0.95}=(0.95)^{30}\cong 0.166.$$

$\Pi(\theta_1)=0.834$.

Remark. In general, Π, β, RR depend on θ_1 (from H_1). However, sometimes they are the same for all values θ_1 in an alt H_1.

$\to$ a uniformly most powerful test.

Ex.

③. Let $X_1 \dots X_n$ be a r. sample drawn from a $N(\mu,\sigma)$ distr. with $\mu\in\mathbb{R}$ unknown and $\sigma>0$ known. At the $\alpha\in(0,1)$ sign. level, find a (unif). most powerful test for the right-tailed alternative, i.e.

$H_0: \mu=\mu_0$.

$H_1: \mu>\mu_0$

Sol: $H_0: \mu=\mu_0$.

$H_1: \mu=\mu_1\ (>\mu_0)$

$f(x;\mu)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x\in\mathbb{R}$.

$L(x_1\ldots x_n;\mu)=\prod_{i=1}^{n} f(x_i;\mu)=\prod_{i=1}^{n}\left(\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)=$

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

$\frac{L(\mu_1)}{L(\mu_0)}=e^{\frac{1}{2\sigma^2}\left[\sum(x_i-\mu_0)^2-\sum(x_i-\mu_1)^2\right]}\geq k_\alpha>0.$

Take ln (increasing function). $\qquad \bar{x}=\frac{1}{n}\sum x_i$

$\frac{1}{2\sigma^2}\left[\cancel{\sum x_i^2}-2\mu_0\underbrace{\sum x_i}_{n\bar{x}}+n\mu_0^2-\cancel{\sum x_i^2}+2\mu_1\underbrace{\sum x_i}_{n\bar{x}}-n\mu_1^2\right]\geq \ln k_\alpha.$

$n(\mu_0^2-\mu_1^2)+2(\mu_1-\mu_0)\cdot n\bar{x}\geq 2\sigma^2\ln k_\alpha$

$2\underbrace{(\mu_1-\mu_0)}_{>0.}n\bar{x}\geq 2\sigma^2\ln k_\alpha+n(\mu_1^2-\mu_0^2).$

$\bar{x}\geq\underbrace{\frac{\sigma^2\ln k_\alpha}{n(\mu_1-\mu_0)}+\frac{\mu_1+\mu_0}{2}}_{K_\alpha.}$

RR: $\{\bar{x}\geq K_\alpha\}$.

$\alpha=P(\text{reject } H_0|H_0)=P(\bar{X}\geq K_\alpha\mid\mu=\mu_0).$

$X_1\ldots X_n\in N(\mu,\sigma).$

$\bar{X}\in N\left(\mu,\frac{\sigma}{\sqrt{n}}\right)\Rightarrow\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\in N(0,1).$

$\alpha=P\left(Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\geq\frac{K_\alpha-\mu}{\frac{\sigma}{\sqrt{n}}}\mid\mu=\mu_0\right)=P\left(Z_0\geq\frac{K_\alpha-\mu_0}{\frac{\sigma}{\sqrt{n}}}\mid Z_0\in N(0,1)\right\}$

$Z\in N(0,1)$ $\qquad =P(Z_0\geq z_{1-\alpha})$

quantile

$\Rightarrow\frac{K_\alpha-\mu_0}{\frac{\sigma}{\sqrt{n}}}=z_{1-\alpha}.\quad\left|K_\alpha=\mu_0+\frac{\sigma}{\sqrt{n}}z_{1-\alpha}\right|$ indep of μ_1.

$\Rightarrow$ unif most powerful test

Exam. 3 lists: - formulas_prob_stat.pdf
- continuous_distr.pdf
- means and variances (Lab 5).

problems prob and stat

seminar
dar mai ușoare