

Part I. Probability Theory:

3 approaches: axiomatic, classical and geometrical.

1. 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.

- Given 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 one 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 $B \subseteq A$.

2. Usual operations:

- the union, $A \cup B$. "or" $A \cup B = \{e \in S | e \in A \text{ or } e \in B\}$.

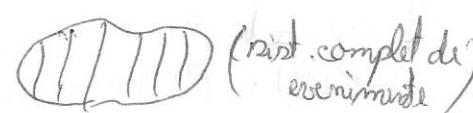
- intersection, $A \cap B$ "and" $A \cap B = \{e \in S | e \in A \text{ and } e \in B\}$.

- difference $A \setminus B$ "but" $A \setminus B = \{e \in S | e \in A \text{ but } e \notin B\} (= A \cap \bar{B})$.

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 is called 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 $\bigcup_{i=1}^n A_i = S, A_i \cap A_j = \emptyset, \forall i \neq j$. 

Example:

- Exp. Rolling a dice.

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

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

$B \subseteq C$.

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

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

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

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

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

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

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

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

AUDUE = $\Sigma \Rightarrow \{A, \bar{A}, E\}$, partition of S .

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

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

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

$$1). K \neq \emptyset.$$

$$2). A \in K \Rightarrow \bar{A} \in K$$

$$3). A_m \in K, \forall m \in \mathbb{N} \Rightarrow \bigcup_{m \in \mathbb{N}} A_m \in K.$$

Ex. The easiest ex: the power set. $K = P(S) = \{S' | S' \subseteq S\}$.
 (S, K) - measurable space.

Theorem 2. Let K be a σ -field over S . Then,

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

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

$$c). A_m \in K, \forall m \in \mathbb{N} \Rightarrow \bigcap_{m \in \mathbb{N}} A_m \in K.$$

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

$$i) P(S) = 1.$$

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

$$iii). \bigcup_{m \in \mathbb{N}} A_m \subseteq K \text{ m.e. events.}$$

$$P\left(\bigcup_{m \in \mathbb{N}} A_m\right) = \sum_{m \in \mathbb{N}} P(A_m). \quad (P \text{ is } \sigma\text{-additive}).$$

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

Theorem 3.

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

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

$$c) P(A \setminus B) = P(A) - P(A \cap B)$$

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

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

Pf. a) Let $A \in K \Rightarrow \{A, \bar{A}\}$ partition of S (m.e.).

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A}). \Rightarrow P(\bar{A}) = 1 - P(A).$$

$$B) P(\emptyset) = P(S) \stackrel{\text{m.e.}}{=} 1 - P(S) = 0.$$



$$C). A = (A \cap B) \cup (A \setminus B), \Rightarrow P(A) = P(A \cap B) + P(A \setminus B). \Rightarrow P(A \setminus B) = P(A) - P(A \cap B).$$

$$d) A \subseteq B \Rightarrow A \cap B = A \Rightarrow 0 \leq P(B|A) \stackrel{c}{=} P(B) - P(A) \Rightarrow P(A) \leq P(B)$$

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

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \setminus (A \cap B)) \stackrel{c}{=} P(A) + P(B) - P(B \cap (A \cap B)) = \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Theorem 4.

Let (S, \mathcal{K}, P) be a prob. space. Then:

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

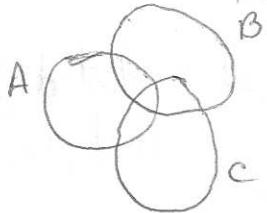
$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \dots + (-1)^{n+1} P(\bigcap_{i=1}^n A_i).$$

Remarks

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

2° 3 events

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



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.

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

Ex. Two dice 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. $(i, i) \forall i = 1, 6$. $P(A) = \frac{6}{36} = \frac{1}{6}$.

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

$(1,1)(1,2)(1,3)(1,4)(2,2)(2,3)$ 6×2 (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" \rightarrow "measure".

Def (Poincaré, 19th century).

Let $S \subseteq \mathbb{R}^m$ 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 :

$m=1$, $\mu = \text{distance (length)}$.

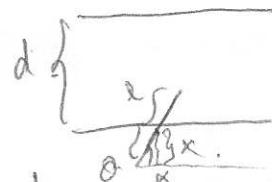
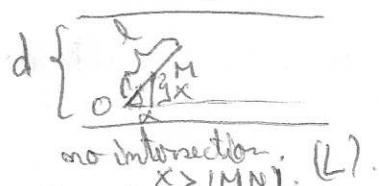
$m=2$, $\mu = \text{area}$

$m=3$, $\mu = \text{volume}$.

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. intersect at most 1 line (because $l < d$). Denote by (L) the closest line to the needle. We have 2 pos. situations.



there is intersection: $x < 1/2 d$.

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

Denote by α the angle between the direction of the needle and the dir of (L) . $(MN) = \text{dist from } M \text{ to } (L)$.

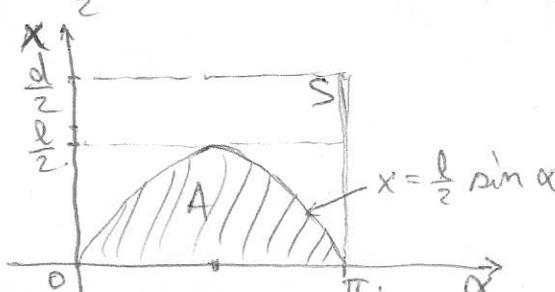
$$S = \{(\alpha, x) \in \mathbb{R}^2 / \alpha \in [0, \pi], x \in [0, \frac{d}{2}] \}$$

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

from $\triangle OMN$:

$$\sin \alpha = \frac{|MN|}{|OM|} = \frac{|MN|}{\frac{d}{2}} \Rightarrow |MN| = \frac{d}{2} \sin \alpha.$$

$m=2$, $\mu = \text{area}$



$$\begin{aligned} \mu(S) &= \text{area}(S) = \pi \cdot \frac{d}{2} = \frac{\pi d}{2}, \\ \mu(A) &= \text{area}(A) = \int_{-\pi/2}^{\pi/2} \frac{d}{2} \sin \alpha \, d\alpha = \\ &= \frac{d}{2} (-\cos \alpha) \Big|_0^{\pi} = d. \end{aligned}$$

$$P(A) = \frac{l}{\frac{1}{2}d} = \frac{2l}{d}.$$

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

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

(RHS)

LHS

Proof.

$$\text{RHS} = P(A_1) \cdot \frac{P(A_2 | A_1)}{P(A_1)} \cdot \frac{P(A_3 | A_1 \cap A_2)}{P(A_2 | A_1)} \cdots \frac{P(A_m | \bigcap_{i=1}^{m-1} A_i)}{P(A_{m-1} | \bigcap_{i=1}^{m-2} A_i)} = P(\bigcap_{i=1}^m A_i) = \text{LHS}$$

Prop 3

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

If cons. $\{A, \bar{A}\}$ partition of S . ($A \cup \bar{A} = S$, $A \cap \bar{A} = \emptyset$), (earliest ex of a partition)

$$B = B \cap S = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A}).$$

m.e.

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

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

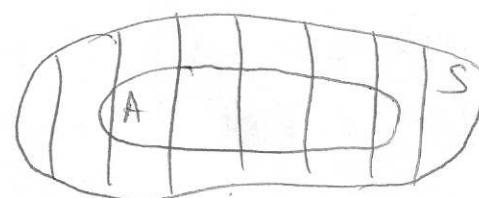
$$P(A) = \sum_{i \in I} P(A_i)P(A|A_i).$$

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

$$B = B \cap S = B \cap \left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} (B \cap A_i)$$

m.e.

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



Def.

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

$\{A_i\}_{i=1,m}$ are indep. if $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$ for $k=2, \dots, m$.

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, \bar{B} are also indep.

$$\begin{aligned} \text{L. } P(A \cap \bar{B}) &= P(A|\bar{B}) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A)/(1 - P(B)) = \\ &= P(A) \cdot P(\bar{B}) \end{aligned}$$

Remarks. 3° If A, B are indep. then $(A, \bar{B}), (\bar{A}, B), (\bar{A}, \bar{B})$ 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
1 - with replacement: the ball is put back; can be sel. again.

1 - 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 \langle "succes" (A) $,$ "failure" (\bar{A}) \rangle

- the prob of success, p , stays the same in every trial.

$(q = 1 - p, \text{prob. of failure})$

Bernoulli Model:

L7

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

Prop 1.

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

C 3 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.

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

Remarks,

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

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

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

Ex.

(i) A dice is rolled 5 times. Find the prob of :-

a) A : getting 3 6's.

b) B : getting at least 2 even nos.

c) 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 no. $n=5$.

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

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

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

$$P(B) = 1 - P(\bar{B}) = 1 - P(\text{0 or } \checkmark \text{ successes}) = 1 - (P(0 \text{ succ.}) + P(1 \text{ succ.})) =$$

m.e.

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

2. Bernoulli Model Without Repl.

Hypergeometric model.

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

OR:

N -objects $\begin{cases} m_1 \text{ have a characteristic} \\ N-m_1 \text{ do not} \end{cases}$

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

Prop 2

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

Remark

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

Ex:

Q. There are 15 boys and 20 girls in a prob class. ~~10~~ 10 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. $N=35, m_1=20, n=10, \begin{cases} 5 \\ 5 \end{cases}$

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

b). success: girl.

B: at least 1 success ($k \geq 1$) ($k = 1, 2, \dots, 10$)

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

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

3. Poisson Model.

A gen. of the binomial model.

Poisson Model.

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

Prop 3

$$P(m; k) = \sum_{i_1+i_2+\dots+i_m=k} p_{i_1} q_{i_2} \dots q_{i_m}$$

Remarks

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

$$(p_1 x + q_1)(p_2 x + q_2) \dots (p_m x + q_m) = \sum_{k=0}^m P(m; k) x^k$$

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

3°. If $p_i=p$ & $i=1, m$. \rightarrow binomial model.

Ex.

③. (Sem 2) The 3 shooters prob). The prob. of 3 shooters hitting a target or $P_1=0.4, P_2=0.5, P_3=0.7$. Find the prob. of the target being hit exactly once.

Sol.

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

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

$$(0.4x + 0.6)(0.5x + 0.5)(0.7x + 0.3) = - - -$$

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. $P(m; k)$ of the m^{th} success occurring after k failures. ($k=0, 1, \dots$).

Prop. 4.

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

Remark.

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

5. Geometric Model.

L10

A part. case of $\text{geom. with } n=1$.

Prop 5.

$$P_k = P(C_1; k) = pq^k$$

Rem.

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

Ex. 4.

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

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

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

Sol.

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

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

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

B: 3rd success after 5 trials. $\therefore 3^{\text{rd}}$ success after 5 failures,

$$\text{Pascal's Model: } P = \frac{1}{2} \quad n=3, k=3$$

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

Chapter III. Random Variables and Random Vectors

a quantitative description of random phenomena.

\rightarrow random vars $\begin{cases} \text{discrete} \\ \text{continuous} \end{cases}$

i. Discret 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 $X: S \rightarrow \mathbb{R}$, satisfying 2 conditions: (i). the set of values that it takes. $(X(S))$ is at most countable (discrete). (ii) $X^{-1}(x) := \{e \in S \mid X(e) = x\} = (X < x), e \in K \quad \forall x \in \mathbb{R}$.

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

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

(i) $X(S) = \{1, \dots, 6\}$, v.

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

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

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

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

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

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

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

Random Variables.

Discrete R.V.s.

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

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

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

Ex. ①. Rolling a dice $X: S \rightarrow \mathbb{R} \quad X(e_i) = i \quad i=1, 2, \dots, 6$.

② The indicator of an ev. $A \in K$

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

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

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

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

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

C4 With what prob does a discrete r.v. take each of its values?

Def: Let $X: S \rightarrow \mathbb{R}$ be a d.r.v. The prob. distribution func (pdf) in NOT

is an array of the form $X \left(\begin{matrix} x_i \\ p_i \end{matrix} \right)_{i \in I}$. $p_i = P(X=x_i)$ \otimes .

Remarks

1^o All values in $(*)$ are distinct.

2^o All prob p_i are $\neq 0$.

3^o $\sum_{i \in I} p_i = 1$.

Ex:

④ for ①.

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

⑤ for ②.

$$X \begin{pmatrix} 0 & 1 \\ P(\bar{A}) & P(A) \\ 1-P(A) \end{pmatrix} \quad p = P(A) \quad f(x) = \begin{cases} p^x (1-p)^{1-x}, & x=0, 1 \\ 0, & \text{otherwise} \end{cases}$$

2. Cumulative Distr. Function. (Fonction de répartition)

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

$$F(x) = P(\{x_i \in S | X(x_i) < x\}) = P(X < x).$$

Ex:⑥. The cdf ~~f~~ of the indicator n.r.

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

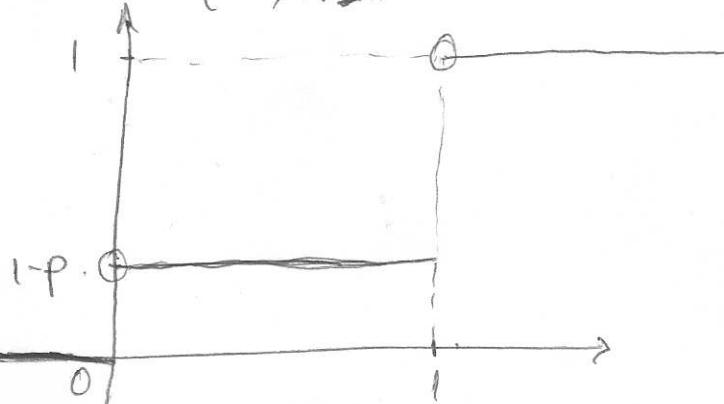
$$x = -1,$$

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

$$F(0) = P(X < 0) = 0. \quad \text{because } X = 0 \text{ or } 1.$$

1° if $x \leq 0$, $F(x) = 0$.2° if $0 < x \leq 1$ $F(x) = P(X < x) = P(X=0) = 1-p$.3° if $x > 1$ $F(x) = P(X < x) = P(X=0 \text{ or } X=1) = P(X=0) + P(X=1) = 1-p+p = 1$.

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



0.7
0.3
0.2
0.1

Theorem 1 (Properties of cdf).

Let X be a r.v. and let F be its cdf. Then F satisfies:

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

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. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$.

e). $P(X \leq x) = F(x+0) (= \lim_{y \nearrow x} F(y))$.

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

pf of a).

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

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

$\Leftrightarrow (X < b) \cap (X \geq a)$.

$$P(a \leq X < b) = P((X < b) \setminus (X \geq a)) = P(X < b) - P((X < b) \cap (X \geq a)) = P(X < b) - P(X \geq a) = F(b) - F(a).$$

Remark:

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

④ Discrete Distributions (Prob. Laws)

① Bernoulli Distr.

$X \in \text{Bern}(p)$, $p \in (0, 1)$. If the pdf is $X \binom{0}{q} p^0 q^1$, $q = 1-p$.

② Binomial Distr. (bino in Matlab).

$X \in \text{B}(n, p)$, $n \in \mathbb{N}$, $p \in (0, 1)$. $X \binom{k}{n} p^k q^{n-k}$

X represents the nr of successes in a binomial model.

Remark

$$\text{Bern}(p) = B(1, p)$$

③ Discrete Uniform Distribution (unid in Matlab)

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

pdf

$$X \binom{k}{m} \quad k = \overline{1, m}$$

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

④ Hypogeometric Distr. (hyge).

$X \in H(N, m_1, n)$, $N, m_1, n \in \mathbb{N}$, $n, m_1 \leq N$.

$$X \left(\frac{\binom{k}{m_1} \cdot \binom{n-k}{N-m_1}}{\binom{n}{N}} \right)_{k=0, n}$$

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

⑤ Poisson Distr (poiss).

$$X \in P(\lambda), \lambda \geq 0, X \left(\frac{\lambda^k}{k!} e^{-\lambda} \right)_{k=0, 1, \dots}$$

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 occurrences.

Remarks

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

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

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

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

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

$$X \left(\binom{k}{m+k} p^m q^k \right)_{k=0, 1, \dots} \quad q = 1-p$$

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

7. Geometrical Distr (geo).

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

$p \in (0,1)$.

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 $(X, Y) : S \rightarrow \mathbb{R}^2$ s.t.

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

$$\begin{aligned} 2). (X, Y)^{-1}(x, y) &= \{e \in S | (X, Y)(e) = (x, y)\} = \\ &= \{(X=x, Y=y) \in K\}. \end{aligned}$$

The joint pdf

$$\begin{array}{ccc} x_i & \cdots & x_i \\ | & & | \\ y_j & \cdots & y_j \end{array} \quad p_{ij} = P((X, Y) = (x_i, y_j)) = \\ = P(X=x_i, Y=y_j) \quad (i, j) \in I \times J.$$

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

$\sum_{i \in I} p_{ij} = q_j = P(Y=y_j)$. of X, Y , resp. called marginal distributions.

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

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

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

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

$$X \left(\begin{matrix} x_i \\ p_i \end{matrix} \right)_{i \in I}, Y \left(\begin{matrix} y_j \\ q_j \end{matrix} \right)_{j \in J}$$

1) The sum of X, Y .

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

$$P_{ij} = P(X=x_i, Y=y_j). \text{ If } X, Y \text{ ind then } p_{ij} = p_i q_j;$$

2) Scalar Multiple

$$\alpha \in \mathbb{R}, \alpha X \left(\begin{matrix} \alpha x_i \\ p_i \end{matrix} \right)_{i \in I}.$$

3) Difference. $X - Y = X + (-1)Y$.

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

4) Product

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

5) Quotient if $y_j \neq 0, \forall j \in J$.

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

C5

6. Continuous Random Variables and Probability Density Function.

Def. Let (S, \mathcal{K}, P) be a prob space. A random variable is a function $X: S \rightarrow \mathbb{R}$ s.t. satisif. the inverse image

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

Def. Let $F: \mathbb{R} \rightarrow \mathbb{R}$ be def by $F(x) = P(X < x)$ 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. $\exists f: \mathbb{R} \rightarrow \mathbb{R}$ st.

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

The funct f from $\textcircled{*}$ 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) $P(X=x) = 0$.

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

Pf. a). Differentiate $\textcircled{*}$: $F'(x) = f(x)$

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

c) Recall $F(\infty) = 1$. ($= \lim_{x \rightarrow \infty} F(x)$)

$$\int_{\mathbb{R}} f(t) dt = \int_{-\infty}^{\infty} f(t) dt = \lim_{x \rightarrow \infty} \int_{-\infty}^x f(t) dt = \lim_{x \rightarrow \infty} F(x) = F(\infty) = 1.$$

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

2. Continuous Distribution (Prob. Laws).

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

discrete $X \in U(m)$.

pdf $f(x) = \begin{cases} \frac{1}{m}, & x \in \{1, \dots, m\} \\ 0, & \text{otherwise} \end{cases}$



② Normal Distr. (mom).

$X \in N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.
pdf

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

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

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

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 $N(0, 1)$, $N(\mu, \sigma^2)$.
- Student T_m , $m \in \mathbb{N} \setminus \{1\}$.

(Chi) χ^2_m , $m \in \mathbb{N}$.

- Fischer, $F(m, n)$ $m, n \in \mathbb{N} \setminus \{1\}$.

8. Continuous R. Vectors. and Joint CDF.

only 2-dim case.

Def. Let (S, \mathcal{K}, P) be a prob space. A (2-dim). rv. vector is a function $(X, Y): S \rightarrow \mathbb{R}^2$ satif. $(X, Y): ((-\infty, x) \times (-\infty, y)) \mapsto \{e \in S | X(e) < x, Y(e) < y\} = (X < x, Y < y) \in \mathcal{K}$, $\forall (x, y) \in \mathbb{R}^2$.

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

Theorem 1 (Properties of the joint cdf).

Let (X, Y) be a rv. vector with joint cdf $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ and let

$F_x, F_y: \mathbb{R} \rightarrow \mathbb{R}$ be the cdf's of X, Y .

Then.

- If $a_k < b_k$, $k=1, 2$ $P(a_1 \leq X < b_1, a_2 \leq Y < b_2) = F(b_1, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, a_2)$.

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

c) F is left cont. in each var.

- d) $F(\infty, \infty) = 1$ e) $F(x, -\infty) = F(-\infty, y) = 0$.
 $F(x, \infty) = F_x(x)$.
 $F(\infty, y) = \underset{\cancel{F}}{F_y(y)}$.

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

Def. Let (X, Y) be a r. vector with cdf $F: \mathbb{R}^2 \rightarrow \mathbb{R}$. We say that (X, Y) is a cont. r. vector if $\exists f: \mathbb{R}^2 \rightarrow \mathbb{R}$ 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 $F: \mathbb{R}^2 \rightarrow \mathbb{R}$, joint pdf $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ and $f_x, f_y: \mathbb{R} \rightarrow \mathbb{R}$ the pdf's of X, Y , resp. Then

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

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

e) $f_x(x) = \int_R f(x, y) dy, \quad f_y(y) = \int_R f(x, y) dx$.

Def Two r. variables X, Y are independent if

~~$F_{(X,Y)}(x, y) = F_X(x) + F_Y(y)$~~ $\Leftrightarrow (x, y) \in \mathbb{R}^2$.

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

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

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

Let (X, Y) be a cont. r. vector. with pdf $f_{X,Y}$. pdf's f_X, f_Y .

Prop 1.

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

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

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

$$f_Z(z) = \int_R f(u, \frac{z}{u}) \cdot \frac{1}{|u|} du.$$

c) quotient $Z = X/Y$.

$$f_Z(z) = \int_R f(zu, u) \cdot |u| du.$$

Prop 2

Let $g: \mathbb{R} \rightarrow \mathbb{R}$ 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.

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

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

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

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

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

$$\begin{cases} 1, & y > \max g \end{cases}$$

Take the derivative

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

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

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

Ex 1 sol.

Ch IV. Numerical characteristics of Random Variables.

1. Expectation.

Def.

- If X is a discr. r.v., pdf. $(\frac{x_i}{p_i})_{i \in I}$, $p_i = P(X=x_i)$ 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 $f: \mathbb{R} \rightarrow \mathbb{R}$, then $E(X) = \int_{\mathbb{R}} x f(x) dx$ if it exists.

Remark

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

\rightarrow if:

\rightarrow X $f: \mathbb{R} \rightarrow \mathbb{R}$ then $E(h(X)) = \int_{\mathbb{R}} h(x) f(x) dx$

Ex:

①. $X \in \text{Bern}(p)$, $p \in (0,1)$ 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 $f: \mathbb{R} \rightarrow \mathbb{R}$. $f(x) = \frac{1}{b-a}$, $x \in [a, b]$.

$$E(X) = \int x f(x) dx. (= \int_a^b + \int_b^\infty)$$

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

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

22

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

- $E(aX + b) = a \cdot E(X) + b$.
- $E(X+Y) = E(X) + E(Y)$.
- If X, Y indep. $E(XY) = E(X)E(Y)$.
- If $X \leq Y$, then $E(X) \leq E(Y)$.

Pf. (select & only X, Y discr.)

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

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

c) $X \left(\begin{matrix} x_i \\ p_i \end{matrix} \right)_{i \in I}, Y \left(\begin{matrix} y_j \\ q_j \end{matrix} \right)_{j \in J} \Rightarrow XY \left(\begin{matrix} x_i y_j \\ p_i q_j \end{matrix} \right)_{(i,j) \in I \times J}$

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

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

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

We prove $Z \geq 0 \Rightarrow E(Z) \geq 0$ $Z \left(\begin{matrix} z_i \\ p_i \end{matrix} \right)_{i \in I}$.

$$Z \geq 0 \Rightarrow z_i \geq 0, \forall i \in I$$

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

$$X \leq Y \Rightarrow Z \geq 0 \Rightarrow E(Z) \geq 0$$

$$E(Z) = E(Y) - E(X) \text{ so } E(Y) \geq E(X).$$

Remark

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

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

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

Ex

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

Fact (Sem 4).

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

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

By T. 1 b) $\Rightarrow E(X) = \sum_{i=1}^m \frac{E(X_i)}{p} \Rightarrow E(X) = mp$.

2. Variance, Standard Deviations, Moments.

Def.

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

$$V(X) = \mathbb{E}\left[(X - E(X))^2\right], \text{ (if it exists) and }$$

$\sigma(X) = \sqrt{V(X)}$ is called the standard deviation of X (aboluta media patrationis).

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

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

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

$$c) \text{ If } X, Y \text{ indep } \Rightarrow V(X+Y) = V(X) + V(Y).$$

$$d) \text{ If } X, Y \text{ indep then. } V(XY) = E(X^2)E(Y^2) - (E(X))^2(E(Y))^2.$$

Pf (sketched).

$$a) V(X) = \mathbb{E}\left[(X - E(X))^2\right] = \mathbb{E}[X^2 - 2E(X)X + (E(X))^2] = \\ = E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (E(X))^2$$

$$b) V(aX + b) = \mathbb{E}\left[(aX + b - aE(X) - b)^2\right] = \mathbb{E}[a^2(X - E(X))^2] = a^2 V(X)$$

Remark.

$$1. a). - X \text{ discr. pdf } (x_i/p_i)_{i \in I} \quad V(X) = \sum_i x_i^2 p_i - \left(\sum_i x_i p_i\right)^2$$

$$- X \text{ cont. pdf } f: \mathbb{R} \rightarrow \mathbb{R} \quad V(X) = \int_R x^2 f(x) dx - \left(\int_R x f(x) dx\right)^2.$$

2. c) can be generalized: $\forall X_i \forall i=1, m$ indep. $V(\sum_{i=1}^m X_i) = \sum_{i=1}^m V(X_i)$.

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

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

Ex:

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

$$E(X) = p.$$

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

$$E(X^2) = p.$$

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

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

3. (Sem 6). $X \in N(p, \sigma^2)$, $p \in \mathbb{R}$, $\sigma^2 \geq 0$.

$$E(X) = p$$

$$V(X) = \sigma^2 \quad (\text{std. dev. } \sigma = \sqrt{V(X)} = \sigma).$$

Def.

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

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

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

- the absolute moment of order k of X .

$$\gamma_k = E(|X|^k)$$

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

$$\mu_k = E[(X - E(X))^k] \quad \text{centered at } a.$$

Remarks

1. $X \left(\frac{x_i}{p_i} \right)_{i \in I}$ discr. $\gamma_k = \sum_{i \in I} x_i^k p_i = \sum |x_i|^k p_i$.

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

2. X cont. pdf $f: \mathbb{R} \rightarrow \mathbb{R}$. $\gamma_k = \int x^k f(x) dx = \int |x|^k f(x) dx$

$$\mu_k = \int (x - E(X))^k f(x) dx.$$

$$E(X) = \gamma_1.$$

$$V(X) = \gamma_2 = \gamma_2 - \gamma_1^2$$

$$\mu_1 = 0.$$

$$\gamma_3 = \gamma_3 - 3\gamma_1\gamma_2 + 2\gamma_1^3.$$

3^o. If \mathcal{D}_m exists for some $m \in \mathbb{N}$, then \mathcal{D}_k and $\mathcal{V}_k(p_k)$ exist $\forall k = 1, \dots, m$.

3. Covariance, Correlation Coefficient, Quantiles.

Def.

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

- the covariance of X, Y ($\text{cov}(X, Y)$) is the nr: $\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$.

- the corr-coeff. of X, Y is $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}}$.

Remarks

1^o Obviously, $\text{cov}(X, Y) = \text{cov}(Y, X)$.

$$\text{cov}(X, Y) = \rho(Y, X).$$

2^o $\text{cov}(X, Y) > 0$ means X, Y grow at the same time.

3^o - $\text{cov}(X, Y) < 0$ means as one increases, the other decreases.

Theorem 3.

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

b) $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$.

c) If X, Y indep $\Rightarrow \text{cov}(X, Y) = \rho(X, Y) = 0$. (X, Y are uncorrelated).

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

e) $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$.

Remarks

1^o a). X, Y ind $\Rightarrow X, Y$ uncorr.

2^o d). can be generalized.

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

Theorem 4.

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

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

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

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

C7 10:11.14,

Def

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

$$P(X < q_{\alpha X}) \leq \alpha \leq P(X \leq q_{\alpha X}), \text{ or equiv. } F(q_{\alpha X}) \leq \alpha \leq F(q_{\alpha X} + 0).$$

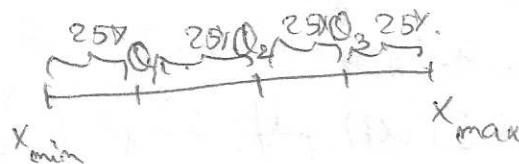
Of these,

the median $m = q_{\frac{1}{2} X}$.

~~Q1~~ the quantiles. $Q_1 = q_{\frac{1}{4} X}$

$$Q_2 = m = q_{\frac{1}{2} X}$$

$$Q_3 = q_{\frac{3}{4} X}$$

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 $F(q_{\alpha X}) = F(q_{\alpha X} + 0)$ so (*) is equiv to $F(q_{\alpha X}) = \alpha$. $q_{\alpha X} = F^{-1}(\alpha)$. So $q_{\alpha X}$ 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, $p, q \geq 1$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$. Then $E(|XY|) \leq (E(|X|^p))^{1/p} \cdot (E(|Y|^q))^{1/q}$

② Schwartz's Ineq.

Let $p=q=2$ in ①.

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

③ Cauchy - Bunakovsky Ineq.

Let $Y = 1$ in ②. $(Y(1))$

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

④ Minkowsky's Ineq.

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

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

5. Markov's Ineq.

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

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

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

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 $e \in S : |X(e)| < a$, 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, $a I_{A(e)} \leq |X(e)|$, i.e., $a I_A \leq |X|$.

Then $E(a I_A) \leq E(|X|)$. p.d.f $a I_A = \begin{pmatrix} 0 & a \\ 1-p(A) & p(A) \end{pmatrix}$.

$E(a I_A) = a p(A)$. 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 $\epsilon > 0$. Then $P(|X - E(X)| \geq \epsilon) \leq \frac{1}{\epsilon^2} V(X)$.

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

Pl.

⑤ for $(X - E(X))^2$ and $a = \epsilon^2 \Rightarrow P(|X - E(X)| \geq \epsilon) = P((X - E(X))^2 \geq \epsilon^2) \leq \frac{1}{\epsilon^2} E((X - E(X))^2) / V(X)$.

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. the probability to X , $X_n \xrightarrow{P} X$, if $\lim_{n \rightarrow \infty} P(|X_n - x| < \epsilon) = 1, \forall \epsilon > 0$.

$$(\quad \geq \epsilon) = 0).$$

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

$$(P(\quad \neq x) = 0)$$

③ $\{X_n\}_n$ conv. in distribution to X , $X_n \xrightarrow{d} X$ ("in repartition", $X_n \rightsquigarrow X$). If

$\lim_{n \rightarrow \infty} F_n(x) = F(x)$, $\forall x$ a point of continuity of F .

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

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

Theorem 1

Let $\{X_n\}_{n=1}^{\infty}$ be a seq. of r.o.'s. and X a r.o. Then the following are equiv:

- $X_n \xrightarrow{a.s.} X$. (almost surely = a.o.)
- $\lim_{n \rightarrow \infty} P(\sup_{k \geq n} |X_k - X| < \varepsilon) = 1$, $\forall \varepsilon > 0$.
 $(P(|X_k - X| > \varepsilon) = 0)$

Theorem 2

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

b) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{a.s.} X$.

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

Want $X_n \xrightarrow{P} X$, i.e. $\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$, $\forall \varepsilon > 0$. (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 \Rightarrow .

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

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

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

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

Theorem 3

Let $\{X_n\}_{n=1}^{\infty}$ be a seq of r.o.'s, let X, Y be r.o.'s, s.t. $X_n \xrightarrow{P} X$,
 $X_n \xrightarrow{P} Y$. Then. $X \stackrel{a.s.}{=} Y$.

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 $\{ |X_n - X| < \frac{\epsilon}{2} \} \cap \{ |X_m - Y| < \frac{\epsilon}{2} \} \subseteq \{ |X - Y| < \epsilon \}$. (30)

Take $(A \subseteq B, \mathcal{U})$.

Remark

T.B holds for $a.s.$, $d.$, $c.$

2. Laws of Large Numbers and Limit Theorems.

We are interested in the stability of the average of "large" nr. of observations.

$$\frac{1}{n} \sum_{k=1}^n |X_k - E(X_k)| \xrightarrow{P} 0$$

$\Rightarrow hX_n \xrightarrow{P} hY_n$ follows the WLLN (weak law of large numbers).

$\xrightarrow{a.s.} hX_n \xrightarrow{a.s.} hY_n$ follows the SLLN (strong L.L.N.)

Let hX_n be a seq. of indep r.v.'s. Notations: $m_k = E(X_k)$ (μ_k),

$$T_k^2 = V(X_k) \quad (\text{T}_k \text{ stand. dev.})$$

$$T_{(m)}^2 = \sum_{k=1}^m T_k^2$$

Define: $X_{(m)} = \frac{1}{T_{(m)}} \sum_{k=1}^m (X_k - m_k)$ (reduced variable, $\frac{X - E(X)}{V(X)}$).

Notice

$$E(X_{(m)}) = \frac{1}{T_{(m)}} \sum_{k=1}^m (E(X_k) - m_k) = 0, \quad V(X_{(m)}) = \frac{1}{T_{(m)}^2} \sum_{k=1}^m V(X_k) = 1.$$

Limit Problem to find X s.t. $X_{(m)} \xrightarrow{d} X$, the answer, a limit theorem. If,

X has a normal distri \Rightarrow central limit theorem.

Theorem (CLT).

Let $\{X_k\}_{k=1}^m$ be a seq of indep and identically distributed, r.v.'s with $E(X_k) = \mu$, $V(X_k) = \sigma^2$. Then, $X_{(m)} \xrightarrow{d} N(0, 1)$

C8 Ch.7. Estimation

7.1 Sample Theory

Notations: Population: characteristic X , $E(X) = \mu$, $V(X) = \sigma^2$

Sample of size n : n , random var's X_1, \dots, X_n indep and having the same pdf as X .

- sample fns $h(X_1, \dots, X_n)$ are random variables.

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

- sample moments: $\bar{J}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$,

- sample central moments: $\bar{\mu}_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$

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

Notations.

	Pop	Sample
Mean	$\mu = E(X)$	\bar{X}
Variance	$\sigma^2 = V(X)$	S^2
Std dev	σ	S
Moment	$\gamma_k = E(X^k)$	$\bar{\gamma}_k$
C. Moment	$\mu_k = E(X - E(X))^k / \bar{\mu}_k$	

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

Pf.

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

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

Prop 2. $E(\bar{\gamma}_k) = \bar{\gamma}_k$ ---

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, independent having the same pdf $f(x; \theta)$.

2 types of est:

- point estimate

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

- interval estimate.

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

$$\hat{\theta}_U(X_1, \dots, X_n)$$

Point Estimators.

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

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

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

Recall:

- \bar{X} is unbiased estimator for μ .

- $\bar{\gamma}_k$ is unbiased estimator for γ_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 Framework.

pop. dist. X whose pdf $f(x; \theta)$ θ - target parameterSample of size m : m sample var's X_1, \dots, X_m indep. and have the same pdf as X , $f(x; \theta)$ Point estimator $\bar{\theta} = \bar{\theta}(X_1, \dots, X_m)$. $\bar{\theta}$ has a mean, $E(\bar{\theta})$ and a variance $T_{\bar{\theta}}^2$ ($T_{\bar{\theta}}$ std. deviation).- $\bar{\theta}$ unbiased if $E(\bar{\theta}) = \theta$ - the standard error $T_{\bar{\theta}}$ to be small, i.e. $T_{\bar{\theta}}$ becomes smaller as m becomes larger.Find an interval estimate: find $\bar{\theta}_L, \bar{\theta}_U$ s.t.

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

- $(\bar{\theta}_L, \bar{\theta}_U)$ - 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,

i.e. $(-\infty, \bar{\theta}_U)$ $(\bar{\theta}_L, \infty)$.

General Framework

We use the pivotal method:We have a pivot, X , (a sample function) satif:

- 1) X depends on X_1, \dots, X_m , other known quantities and θ (the only unknown!)
- 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 ($m > 30$) or comes from an approx normal pop and, T (pop. std. dev) is known then.

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

Let us assume that for a target param. θ and an est $\bar{\theta}$ with mean $E(\bar{\theta}) = \theta$,
and std. error $\sqrt{\text{Var}(\bar{\theta})}$, the $1-\alpha$.

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

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 , $P(X < q_\alpha) = \alpha$).

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

$$\begin{aligned} 1-\alpha &= P(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) = P(z_{\frac{\alpha}{2}} < \frac{\bar{\theta} - \theta}{\sqrt{\text{Var}(\bar{\theta})}} < z_{1-\frac{\alpha}{2}}) = \\ &= P(\sqrt{\text{Var}(\bar{\theta})} \cdot z_{\frac{\alpha}{2}} < \bar{\theta} - \theta < \sqrt{\text{Var}(\bar{\theta})} \cdot z_{1-\frac{\alpha}{2}}) = P(\bar{\theta} - \sqrt{\text{Var}(\bar{\theta})} \cdot z_{1-\frac{\alpha}{2}} < \theta < \bar{\theta} + \sqrt{\text{Var}(\bar{\theta})} \cdot z_{\frac{\alpha}{2}}). \end{aligned}$$

Remark.

$$\left. \begin{array}{l} z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}} \\ \bar{\theta} \in (\bar{\theta} - \sqrt{\text{Var}(\bar{\theta})} \cdot z_{1-\frac{\alpha}{2}}, \bar{\theta} + \sqrt{\text{Var}(\bar{\theta})} \cdot z_{\frac{\alpha}{2}}) \end{array} \right\} \text{So the } 100(1-\alpha)\% \text{ c.i.}$$

$$\Leftrightarrow \bar{\theta} \pm \sqrt{\text{Var}(\bar{\theta})} z_{1-\frac{\alpha}{2}} \text{ only for sym. distr.'s. } N(0,1), T(m-1).$$

B. C.I's for the Parameters of One pop.

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) σ known

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

b) σ unknown

$$\text{pivot } T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \in T(m-1), \text{ i.e. } \mu \in (\bar{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}, \bar{X} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}), \text{ (sym.)}$$

2. C.I for the pop. variance, $\theta = \sigma^2$

If the sample is drawn from the normal pop.

$$\text{pivot } \chi^2 = \frac{(n-1)S^2}{\sigma^2} \in \chi^2(m-1), \text{ i.e. } \sigma^2 \in \left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} \right) \text{ (not sym.)}$$

$$(S \in (\sqrt{S_{1-\frac{\alpha}{2}}}, \sqrt{S_{\frac{\alpha}{2}}}))$$

Ex 1 The shopping times of 64 random selected customers 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 customer.
Sol $\theta = \sigma^2$, $m = 64$, $\bar{X} = 33$, $S^2 = 256$.

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

$$t_1 - \frac{\bar{x}}{2} = t_{0.975} = 1.9983 \quad (\text{time}(0.975, 64)).$$

$$t \frac{\bar{x}}{2} = t_{0.025} = -1.9983.$$

$\mu \in (29.0033, 36.9967)$. So, on average, a customer 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}, \dots, X_{1m_1} from the 1st pop. size m_1 , $\bar{X}_1 = \frac{1}{m_1} \sum_{i=1}^{m_1} X_{1i}$.

$$\Delta_1^2 = \frac{1}{m_1-1} \sum_{i=1}^{m_1} (X_{1i} - \bar{X}_1)^2.$$

same for 2nd pop

In addition: pooled variance (var. centralization).

$$S_p^2 = \frac{\sum_{i=1}^{m_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{m_2} (X_{j\cdot} - \bar{X}_2)^2}{m_1 + m_2 - 2} = \frac{(m_1-1) \Delta_1^2 + (m_2-1) \Delta_2^2}{m_1 + m_2 - 2}$$

□ C.I. for pop. means, $\Theta = \mu_1 - \mu_2$.

- either large samples ($m_1 + m_2 \geq 40$) or normal underlying pop. ($N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$).

a) σ_1, σ_2 known.

$$\text{pivot } Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m_1} + \frac{\sigma_2^2}{m_2}}} \in N(0, 1). \quad \boxed{\text{sym}}$$

b) $\sigma_1 = \sigma_2$ unknown.

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} \in T(m_1 + m_2 - 2). \quad \boxed{\text{sym}}$$

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

$$T^* = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\Delta_1^2}{m_1} + \frac{\Delta_2^2}{m_2}}} \in T(m_1 + m_2 - 2). \quad \boxed{\text{sym}}$$

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

$$F = \frac{\Delta_1^2 / \sigma_1^2}{\Delta_2^2 / \sigma_2^2} \in F(m_1-1, m_2-1) \quad \boxed{\text{not sym}}$$

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. θ .

- a hyp test: the method(s), used to decide if a hyp is true or not.

2. Hypotheses

- 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 param. 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.

		H_0 true	H_1 true
H_0	Type I error α	correct decision.	
H_1	correct decision.	Type II error. β	

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

α - significance level.

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

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.)	GOAL: to decide if - we reject H_0 (in favor of H_1). - do not reject H_0 . Given $\alpha \in (0,1)$ - significance level.
alt H_1	$\theta < \theta_0$ (left-tailed)	
one of	$\theta > \theta_0$ (right-tailed)	
	$\theta \neq \theta_0$ (two-tailed)	

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0).$$

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

- test statistic, TS. (same as the p-value).

- rejection region, RR (critical).

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

Suppose that for the t.par. θ , we have an unbiased point estimator $\bar{\theta}(x_1, \dots, x_n)$.
 $(E(\bar{\theta}) = \theta)$ with standard error $T\bar{\theta}$ s.t. $Z = \frac{\bar{\theta} - \theta}{T\bar{\theta}} \in N(0,1)$.

We choose: $TS = Z$.

- observed value of the TS $TS_0 = TS(\theta = \theta_0)$.

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

For a left-tailed test

$$H_0: \theta = \theta_0.$$

$$H_1: \theta < \theta_0.$$

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).

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

We choose a RR of the form $RR = \{Z < k_{\text{left}}\} = \{ \dots, k_{\text{left}} \}$.

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

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

Right-tailed Test

$$H_0: \theta = \theta_0$$

$$H_1: \theta > \theta_0.$$

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

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

Two-Tailed Test

$$H_0: \theta = \theta_0.$$

$$H_1: \theta \neq \theta_0.$$

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

$$1 - \alpha = P(k_1 < Z < k_2), \quad \text{symm. } Z \frac{\alpha}{2} = -Z_{1-\frac{\alpha}{2}}$$

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

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

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

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

X_1, \dots, X_m sample vari's.

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 ($n \geq 30$), or from a normal pop.

a) σ known, (Z-test)

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

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

$$\text{RR: } \begin{cases} Z_0 < z_{\alpha} & \text{left} \\ Z_0 > z_{1-\alpha} & \text{right} \\ |Z_0| > z_{1-\frac{\alpha}{2}} & \text{two} \end{cases}$$

b) σ unknown, (t-test).

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

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

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

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

If the sample comes from a normal distn.

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

$$T = T_0$$

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

$$T < T_0$$

$$T > T_0$$

$$T \neq T_0$$

$$\text{RR: } \chi^2_0 < \chi^2_{\alpha}$$

$$\chi^2_0 > \chi^2_{1-\alpha}$$

$$\chi^2_0 < \chi^2_{\frac{\alpha}{2}} \text{ or } \chi^2_0 > \chi^2_{1-\frac{\alpha}{2}}$$

$$TS = \chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \in \chi^2(n-1)$$

$$TS_0 = \chi^2_0 = \frac{(n-1)s^2}{\sigma_0^2} \in \text{not sym } \chi^2(n-1)$$

Ex:

- 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 this). 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 contradiction to the VP's suspicion?

$$H_0: \mu = 20$$

$$H_1: \mu < 20.$$

$$\sigma = 4$$

$$n = 36, \bar{x} = 19,$$

$$\alpha = 0.05.$$

$$Z_{0.05} = -1.645 (= \text{norminv}(0.05, 0, 1))$$

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

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

$Z_0 \notin RR \Rightarrow$ 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, do 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 . | $P \leq \alpha \Rightarrow$ reject H_0
 $P > \alpha \Rightarrow$ do not reject.

Remark

1^o 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

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

2^o P can be thought of as the minimum, rejection sign. level.

EX:

② the previous example.

$$Z_0 = -1.5$$

$$\alpha = 0.05.$$

$$P = P(Z < -1.5) (= \text{normcdf}(-1.5, 0, 1)) = 0.0668.$$

$P > \alpha \Rightarrow$ 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. $T(m_1 + m_2 - 2)$ (t-test 2).c) $\sigma_1 \neq \sigma_2$ unknown $T(m)$ (t-test 2 "unequal").

2. $\Theta = \frac{\sigma_1^2}{\sigma_2^2}$ (var-test 2).

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

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

$F(m_1 - 1, m_2 - 1).$

Back to Chapter VIII

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

$$H_0: \theta = \theta_0.$$

$$H_1: \theta < \theta_0 \text{ left.}$$

$\theta > \theta_0$ right.

$\theta \neq \theta_0$ both.

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

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

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

$$H_0: \theta = \theta_0.$$

$$H_1: \theta = \theta_1 \quad \begin{cases} < \theta_0 \\ > \theta_0 \\ \neq \theta_0 \end{cases}$$

$\beta(\theta_1)$.

Ex.

① Recall the Vice President problem,

$$H_0: \mu = 20 (= \mu_0). \quad 5\% \text{ sign. level}$$

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

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

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

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

We found,

$$RR = \left\{ Z_0 < Z_{0.05} \right\} = \left\{ \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -1.645 \right\} \Rightarrow \frac{\bar{X} - 20}{\frac{4}{\sqrt{36}}} < -1.645 \right\}$$

$$= \left\{ \bar{X} < 20 - \frac{4}{3} \cdot 1.645 \right\} = \left\{ \bar{X} < 18.9 \right\}$$

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

Suppose:

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

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

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

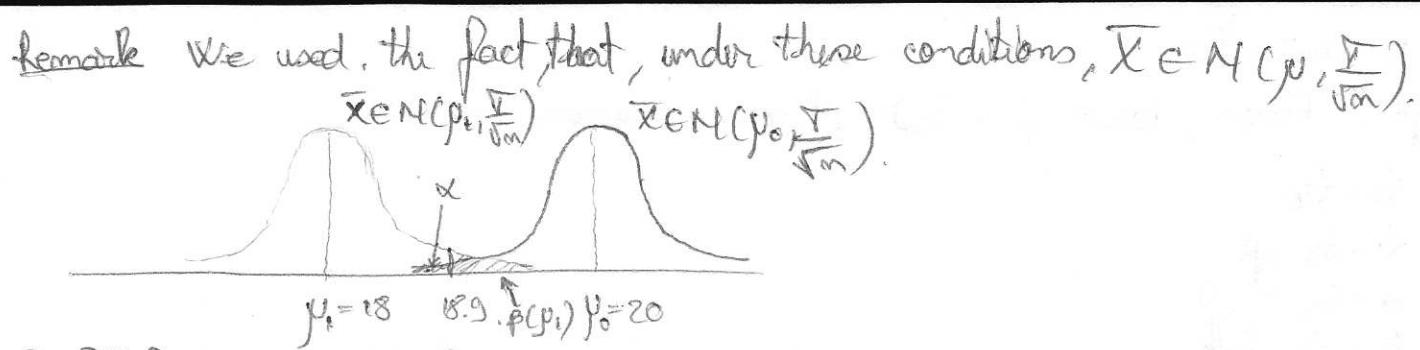
$$= P\left(\frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} > \frac{18.9 - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) \Rightarrow \beta(\mu_1) = P(Z_1 > 1.35 | \mu = \mu_1 = 18) =$$

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

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

$$= 0.0885$$

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



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

$$\Pi(\theta^*) = P(\text{reject } H_0 | \theta = \theta^*) = P(\text{TSE RR} | \theta = \theta^*)$$

$$H_0: \theta = \theta_0$$

$$H_1: \begin{cases} \cdot \\ \cdot \\ \cdot \end{cases}$$

$$\theta = \theta_1$$

$$\Pi(\theta_0) = P(\text{reject } H_0 | H_0) = \alpha$$

For any other value $\theta_1 (\neq \theta_0)$,

$$\Pi(\theta_1) = P(\text{reject } H_0 | \theta = \theta_1) = 1 - P(\text{do not reject } H_0 | H_1) = 1 - \beta(\theta_1)$$

So, making β small \Leftrightarrow making Π large.

$$H_0: \theta = \theta_0$$

$$H_1: \theta = \theta_1$$

We pass $\alpha = \Pi(\theta_0)$ 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 $L(X_1, \dots, X_n; \theta)$ be the likelihood function of the samp. Then for a fixed $\alpha \in (0, 1)$, a most powerful test is the one with RR.

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

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

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

$$H_1: \theta = 30 \quad (= \theta_1) \quad \text{Also find } \beta(\theta_1)$$

(a sample of size $n=1$)

~~that~~ X_1 .

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

In our case $n=1$, X_1 ,

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

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

$$\begin{aligned} \alpha &= P(\text{reject } H_0 | H_0) = P(X_1 \geq K_\alpha | \theta=1) \quad (\text{X}_1 \text{ same pdf with } X_1) \\ &= \int_{K_\alpha}^{\infty} f_{X_1}(x; \theta) dx = \int_{K_\alpha}^{\infty} 1 dx = 1 - K_\alpha. \Rightarrow K_\alpha = 0.95 \end{aligned}$$

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

$$\begin{aligned} \beta(\theta_1) &= P(\text{do not reject } H_0 | H_1) = P(X_1 < 0.95 | \theta=30) = \int_{0.95}^{\infty} f(x; \theta_1) dx = \\ &= \int_0^{0.95} 30 \cdot x^{29} dx = x^{30} \Big|_0^{0.95} = (0.95)^{30} \approx 0.166. \end{aligned}$$

$$\bar{\Pi}(\theta) = 0.834.$$

Remark. In general, $\bar{\Pi}$, 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 .

→ a uniformly most powerful test.

Ex.:

③. Let X_1, \dots, X_n be a r. sample drawn from a $N(\mu, \sigma^2)$ distr. with $\mu \in \mathbb{R}$ unknown. and $\sigma^2 > 0$ known. At the $\alpha \in (0, 1)$ sign. level, find a (unif.) most powerful test for the right-tailed alternative, i.e.

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

$$H_1: \mu > \mu_0$$

$$\text{Sol: } H_0: \mu = \mu_0.$$

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

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

$$\begin{aligned} L(x_1, \dots, x_n; \mu) &= \prod_{i=1}^n f(x_i; \mu) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) = \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \end{aligned}$$

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

Take \ln (increasing function).

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$\frac{1}{2\sigma^2} [\sum x_i^2 - 2\mu_0 \sum x_i + n\mu_0^2 - (\sum \bar{x}_i^2 + 2\mu_1 \sum \bar{x}_i - n\mu_1^2)] \geq \ln k_\alpha.$$

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

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

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

$$K_\alpha.$$

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

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

$$x_1, \dots, x_n \in N(\mu, \sigma^2).$$

$$\bar{x} \in N(\mu, \frac{\sigma^2}{n}). \Rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \in N(0, 1).$$

$$\alpha = P(Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{K_\alpha - \mu}{\sigma/\sqrt{n}} | \mu = \mu_0) = P(Z_0 \geq \frac{K_\alpha - \mu_0}{\sigma/\sqrt{n}} | Z_0 \in N(0, 1))$$

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

\uparrow quantile

$$\Rightarrow \frac{K_\alpha - \mu_0}{\sigma/\sqrt{n}} = z_{1-\alpha}. \quad | K_\alpha = \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} | \text{ indep of } \mu_0,$$

\rightarrow unif most powerful test

Exam. 3. Lists:

- formulas - prob - stat - pdf
- continuous - distr. - pdf
- means - and covariances (Lab 5)

problems, prob and stat

\uparrow
seminar
der mai upgred