

Orali 14 giugno

A Ad ogni orale fa partire un timer (si ferma circa a 15 min).

Mattina

1° orale → 22

- teorema di bayes
 - perché ci serve? cosa significa intuitivamente?
 - scrivi l'enunciato
- cosa vuol dire che uno stimatore è non distorto?
 - qual'è il significato dietro questa formula?
 - fammi un esempio di stimatore non distorto
 - dice che questa domanda è mal posta, perché? (non distorto rispetto a cosa?)
- statistica descrittiva: immagina di avere un campione e di voler andare a valutare la posizione (centralità), quali indicatori (indici) abbiamo visto? (media, mediana, ...)
 - quando ha senso utilizzare la mediana? possiamo utilizzarla sempre?

2° orale → 30

- distribuzione geometrica
- data una variabile aleatoria, come possiamo calcolarne il valore atteso?
- relazione tra X distribuita secondo una geometrica e Y (non ho capito com'è definita Y)
- cosa dice la legge dei grandi numeri?
 - scrivi la formula
 - che cos'è ϵ ?
 - qual'è il significato della formula (forma debole)?
- statistica descrittiva: parlami di concentrazione (indice di Gini,...)
 - qua il prof non fa tante domande perché il ragazzo parla molto

3° orale → 24

- parlami della distribuzione normale
 - la media (valore atteso) e la mediana sono vicine?
 - perché coincidono?
 - disegna il grafico
 - differenza tra standard e non
 - come cambia il grafico con $\sigma = 1$ e $\sigma = 2$? (glielo fa disegnare)
 - accenna all'esame dove veniva richiesto di verificare le proprietà della f_X
 - guardando i grafici, mostra come mediana e media siano uguali
 - e la moda come cambia?
- supponiamo di voler stimare la varianza, che stimatore vorresti usare?
 - e se volessi stimare la deviazione standard?

- posso stimare σ con $\sqrt{\sigma^2}$?
- quando si dice che due eventi sono indipendenti?
 - con che argomento visto in statistica descrittiva possiamo legare il concetto di indipendenza? (covarianza campionaria)
 - com'è definita (formula)?
 - in che senso la covarianza è legata all'indipendenza? (è vicino a 0 per campioni, =0 per variabili aleatorie)

4° orale → 19

- disegna la funzione di ripartizione di una bernoulliana di parametro u
 - calcola $P(X \leq \frac{1}{3})$ (perché si può "calcolare" senza applicare la formula, osservando solo il grafico?)
 - mostrare graficamente μ
 - che informazioni ci dà il valore atteso?
- quali sono gli indici di dispersione che abbiamo visto?
 - cosa vuol dire misurare la dispersione in un campione?
 - perché introduciamo deviazione standard se abbiamo già la varianza che è il suo quadrato?
 - regola empirica → possiamo sempre applicarla?
 - che operazioni faresti su un campione per decidere se puoi usare la regola empirica?
- cosa uso per stimare la varianza di X ?
 - uso la varianza campionario → che cosa descrive X ? (la popolazione)
 - significato di stimatore non distorto (cos'è θ)?
 - perché è bello avere uno stimatore non distorto per θ ?

5° orale → 19

- (riprendiamo domanda di prima) quali sono gli indici di dispersione visti in statistica descrittiva?
 - il ragazzo dice che anche gli indici di gini sono indici di dispersione → il prof dice no → in che senso possono essere legati?
 - come la misuriamo l'eterogeneità?
 - spiegare casi di eterogeneità minima/massima (scrivendo anche formule)
 - problema con l'indice del gini (non normalizzato)? (non ha un limite massimo)
- cosa dicono gli assiomi di kolmogorov?
 - a cosa servono? (per dimostrare i teoremi che abbiamo visto)
- popolazione distribuita in modo esponenziale → come stimeresti λ ?
 - il ragazzo parla di metodo plug-in e metodo di massima verosimiglianza
 - possiamo usare metodo plug-in per stimare λ ?

Pomeriggio

1° orale → 18

- immagina di stimare il parametro λ di una funzione esponenziale, come faresti?
- disegna il grafico della funzione di ripartizione di una bernoulliana di parametro h , mettendo in evidenza quello che secondo te è importante
 - evidenzia il valore atteso della variabile aleatoria X nel grafico disegnato (dice che ci sono due modi)

- considera un dataset con un attributo che sai essere bernoulliana (1 o 0) → è meglio usare media, mediana o moda come indice?
 - ci sono delle situazioni in cui è meglio usare la mediana piuttosto che la media?

2° orale → 18 (risicato)

- (riprende dall'orale prima) → in quali casi è meglio usare la mediana piuttosto che la media?
- statistica inferenziale: che cosa vuol dire che uno stimatore è consistente in media quadratica?
 - che cosa si intende che "uno stimatore stima un parametro in modo accurato"?
 - cosa si intende per accuratezza?
 - che cos'è l'MSE? (lo chiede perché alla domanda prima non sapeva rispondere)
 - che cosa vuol dire che l'MSE tende a 0 quando la dimensione del campione aumenta?

⚠ Non basta la definizione formale, bisogna saper spiegare cosa rappresenta!

- variabile aleatoria binomiale
 - cosa indica il parametro n ?
 - a cosa serve l'ipotesi dell'indipendenza tra i vari esperimenti bernoulliani? (serve nel calcolo della funzione di massa di probabilità)

3° orale → 18

- domanda di prima (x2): in quali casi è meglio usare la mediana piuttosto che la media?
- calcola il valore atteso di una variabile aleatoria binomiale (riprende p_X calcolata dal ragazzo prima)

⚠ Mettere i pedici nelle sommatorie!

- cosa corrisponde ogni addendo nella sommatoria del valore atteso?

$$E(X) = \underbrace{\sum_{i=0}^n x_i \cdot P(X=i)}_{\text{formula generale (indipendente dal modello!)}} \quad \text{questo}$$

- in generale come si calcola il valore atteso di una variabile aleatoria discreta?
- cos'è n ? (n° di osservazioni)
- cosa misura un'osservazione di una binomiale? (n° di successi)
 - quante osservazioni può avere una binomiale? (n+1)
- come si calcola il valore atteso?

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \underbrace{\sum_{i=1}^n \underbrace{E(X_i)}_{X_i \sim B(p)}}_{\text{per linearità del valore atteso}} = \sum_{i=1}^n p = \boxed{np}$$

- consideriamo la distribuzione normale standard
 - disegna il grafico della funzione di densità
 - perché si evidenzia -3 e $+3$ nell'asse x ?

- che significato grafico hanno -1 e $+1$ nel grafico? (deviazione standard)
 - come evidenzieresti la deviazione standard nel grafico?

4° orale → 30

- disegna il grafico della funzione di densità della distribuzione esponenziale
 - specchia il grafico rispetto all'asse y
 - scrivi la formula della funzione di densità ottenuta
 - si può ricavare la funzione di ripartizione? → il ragazzo osserva che la funzione di densità non rispetterebbe la seconda condizione (integrale = 1)
 - si possono fare considerazioni sulla varianza e sul valore atteso?
 - fanno cose sul foglio, non ho capito un cazzo

⚠ Dice che stavano parlando della Distribuzione di Laplace (?)

- che cosa vuol dire che la varianza campionaria è non distorto per la varianza della popolazione?
 - questa "non distorsione" dipende dalla distribuzione? (no, vale per tutte come la media campionaria)

5° orale → 19

- considera un campione preso a coppie → cosa facciamo se vogliamo valutare se c'è una relazione?
 - preso un punto dallo scatterplot, che coordinate ha?
 - un metodo quantitativo? (covarianza campionaria cov , indice di correlazione campionaria r)
 - perchè $cov > 0$ indica una relazione di tipo diretto? (guardare la formula, la maggior parte degli addendi sarebbe positiva)
 - perchè abbiamo introdotto r ?
 - covarianza nelle variabili aleatorie → che informazioni danno?
 - formula e seguente dimostrazione:

$$\begin{aligned}
 \text{Cov}(X, Y) &= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\
 &= E(XY) - \underbrace{\mu_X E(Y)}_{\mu_X \mu_Y} - \underbrace{\mu_Y E(X)}_{\mu_Y \mu_X} + \cancel{\mu_X \mu_Y} \\
 &= E(XY) - E(X)E(Y)
 \end{aligned}$$

- durante la dimostrazione accennano (senza dimostrazione) al fatto che $E(XY) = E(X)E(Y)$ se X e Y sono indipendenti
 - perchè dividiamo per $n - 1$ quando calcoliamo la covarianza campionaria

6° orale → 21

- riprendendo quanto detto prima ($E(XY) = E(X)E(Y)$) → riesci a dimostrare che $E(X + Y) = E(X) + E(Y)$?

$$\begin{aligned}
E(X + Y) &= \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) P(X = x_i, Y = y_j) = \\
&= \sum_{i=1}^n \sum_{j=1}^m x_i \cdot P(X = x_i, Y = y_j) + \sum_{i=1}^n \sum_{j=1}^m y_j \cdot P(X = x_i, Y = y_j) = \\
&= \sum_{i=1}^n x_i \underbrace{\sum_{j=1}^m P(X = x_i, Y = y_j)}_{\text{probabilità marginale di X}} + \sum_{j=1}^m y_j \underbrace{\sum_{i=1}^n P(X = x_i, Y = y_j)}_{\text{probabilità marginale di Y}} = \\
&= \underbrace{\sum_{i=1}^n x_i P(X = x_i)}_{E(X)} + \underbrace{\sum_{j=1}^m y_j P(Y = y_j)}_{E(Y)}
\end{aligned}$$

⚠ Notare che non serve che X e Y siano indipendenti.

- riprende lo scritto → perché hai scelto un diagramma a barre per mostrare il genere e non uno a torta?
 - quando ha senso usare il diagramma a barre e a torta? (dati qualitativi)
 - come abbiamo diviso i dati qualitativi? (nominali e ordinali)
 - quale dei due grafici mostra l'ordine (grafico a barre)
- statistica inferenziale: come stimeresti il parametro p di una binomiale?
 - che cos'è una statistica? (definizione come funzione)

⚠ Una statistica non deve dipendere da nessun parametro ignoto.

7° orale → 25

- cosa misura l'eterogeneità in un campione?
 - che indici abbiamo visto?
 - indici di gini
 - perché questo indice cattura l'eterogeneità?
 - che limiti ha questo indice? (inferiore e superiore) dimostra uno di questi due limiti
 - come è legato questo indici alla dispersione → perché non possiamo usare la varianza? (vale solo per valori numerici)
- proprietà di assenza di memoria
 - per quali distribuzione vale? (geometrica ed esponenziale)
 - scegli una distribuzione per fare la dimostrazione
- metodo della massima verosimiglianza

8° orale → 26

- teorema delle probabilità totali
 - dimostrazione
- il prof scrive sul foglio uno stimatore $T = \frac{1}{n} (\sum_{i=1}^n (X_i - \mu)^2)$ (non sono sicuro sia questo, l'ha detto a voce) chiede se è non distorto per la varianza della popolazione
 - T è uno stimatore?
 - qua fa alcune domande che non mi sono segnato



Come prima: una statistica non deve dipendere da nessun parametro ignoto.