Cheatsheet di Statistica

Luca Tagliavini

June 6, 2022

NOTA: i seguenti appunti sono **altamente** informali e sono pensati per essere un supporto allo studente come ultimo ripasso prima di una prova. Contengono solo uno scinto molto riassuntivo dei temi centrali al corso.

1 Probabilità condizionata

La probabilità del verificarsi di un evento A sapendo che si è verificando un evento B è data da:

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

2 Regola della catena

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$$

3 Eventi indipendenti

Due eventi A e B si dicono indipendenti se:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

4 Probabilità totali

Sia A un generico evento e $B_1, \dots, B_n$ una partizione di Ω. Allora vale che:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i)$$

5 Formula di Bayes

Siano A, B due eventi tali che $\mathbb{P}(A), \mathbb{P}(B) > 0$. Allora vale:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

6 Distribuzioni semplici

L'insieme delle distribuzioni semplici

$$D_{n,k} = \{(x_1, \dots, x_k) \mid x_i \in E, x_i \neq x_j \forall i, j \text{ t.c. } i \neq j, \text{ con } |E| = n\}$$

ha cardinalità pari a

$$|D_{n,k}| = \frac{n!}{(n-k)!}$$

7 Distribuzioni con ripetizione

L'insieme delle distribuzioni con ripetizione

$$DR_{n,k} = \{(x_1, \dots, x_k) \mid x_i \in E, \text{ con } |E| = n\}$$

ha cardinalità pari a

$$|DR_{n,k}| = n^k$$

8 Combinazioni

L'insieme delle combinazioni semplici

$$C_{n,k} = \{A \subseteq E \mid |A| = k, \text{ con } |E| = n\}$$

ha cardinalità pari a

$$|C_{n,k}| = \binom{n}{k}$$

Le combinazioni sono usate per modellizzare il numero di occorrenze nel caso di estrazioni simultanee.

9 Permutazioni

Il numero di permutazioni di una coppia ordinata E con $|E| = n$ è pari a

$$|P_n| = |D_{n,n}| = n!$$

10 Varianza e Covarianza

La varianza è data da

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

La covarianza è data da

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

La varianza gode della seguente proprietà:

$$Var(aX + b) = a^2 Var(x)$$

11 Probabilità degli intervalli in F_x

$$\begin{aligned}
\mathbb{P}(X = x) &= \mathbb{P}(X \leq x) - \mathbb{P}(X < x) = F_x(x) - F_x(x-) \\
\mathbb{P}(x < X \leq y) &= \mathbb{P}(X \leq y) - \mathbb{P}(X \leq x) = F_x(y) - F_x(x) \\
\mathbb{P}(x \leq X \leq y) &= \mathbb{P}(X \leq y) - \mathbb{P}(X < x) = F_x(y) - F_x(x-) \\
\mathbb{P}(x \leq X < y) &= \mathbb{P}(X < y) - \mathbb{P}(X < x) = F_x(y-) - F_x(x-) \\
\mathbb{P}(x < X < y) &= \mathbb{P}(X < y) - \mathbb{P}(X \leq x) = F_x(y-) - F_x(x)
\end{aligned}$$

12 Variabili aleatorie discrete

Una variabile aleatoria discreta X assume valori appartenenti al suo supporto $S_X = \{x_1, \ldots, x_n\}$. La sua distribuzione prende la forma di

X	x_1	...	x_n
p_X	$p_X(x_1)$	...	$p_X(x_n)$

La funzione di ripartizione F_x è a tratti e della forma

$$F_x(x) = \begin{cases} 0 & \text{se } x < x_1 \\ p_X(x_1) & \text{se } x_1 \leq x < x_2 \\ \vdots & \\ \sum_{i=1}^{n-1} p_X(x_i) & \text{se } x_{n-1} \leq x < x_n \\ \sum_{i=1}^{n} p_X(x_i) = 1 & \text{se } x \geq x_n \end{cases}$$

Il valore atteso è dato da:

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i p_X(x_i)$$

Analogamente, prese due v.a. continue X e Y il valore atteso del loro prodotto è dato da:

$$\mathbb{E}[XY] = \sum_{i=1}^{n} \sum_{i=j}^{m} x_i y_j p_{X,Y}(x_i, y_j)$$

13 Variabili aleatorie continue

Una variabile aleatoria continua X assume valori appartenenti al suo supporto $S_X = [a, b]$. La sua distribuzione prende la forma di una funzione f_X con le seguenti proprietà:

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

La funzione di ripartizione F_x è **continua** e della forma:

$$F_x(x) = \begin{cases} 0 & \text{se } x < x_1 \\ \int_{-\infty}^{x_1} f_X(x)dx & \text{se } x_1 \leq x < x_2 \\ \vdots \\ \int_{-\infty}^{x_{n-1}} f_X(x)dx & \text{se } x_{n-1} \leq x < x_n \\ \int_{-\infty}^{+\infty} f_X(x)dx & \text{se } x \geq x_n \end{cases}$$

Il valore atteso è dato da:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)dx$$

13.1 Standardizzazione

Presa una variabile alteatoria discreta X la sua forma standardizzata Z si ottiene dalla formula

$$Z = \frac{X - \mu}{\sigma}$$

13.2 Teorema centrale del limite

Preso un grande numero di esperimenti aleatori descritte da n variabili aleatorie $X_1, \ldots, X_n$ si può

$$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n}$$

$$\overline{Z_n} = \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Di conseguenza quando viene richiesto di calcolare $\mathbb{P}(X_1 + \ldots + X_n \leq l)$ si può segure la seguente catena di ugualianze:

$$\begin{aligned}\mathbb{P}(X_1 + \ldots + X_n \leq l) &= \mathbb{P}(X_n \leq \frac{l}{n}) \\ &= \mathbb{P}(Z_n \leq \frac{\frac{l}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}) \\ &= \mathbb{P}(Z_n \leq (\frac{l}{n} - \mu) \cdot \frac{\sqrt{n}}{\sigma}) \\ &= \Phi((\frac{l}{n} - \mu) \cdot \frac{\sqrt{n}}{\sigma})\end{aligned}$$

14 Legge di X_n

Presa una catena di markov omogenea e a stati finiti descritta dalla successione $(X_n)_n$, la densità discreta di una qualche X_n vale

$$p_{X_j}(j) = \sum_{i=1}^{n} p_{X_k}(i)\pi_{i,j}^{(n-k)}$$

assumendo X_k come stato iniziale.

15 Distribuzione invariante

Presa una catena di markov descritta dalla matrice Π un vettore generico $\pi = (\pi_1, \ldots, \pi_n)$ si dice *distribuzione invariante* se

$$\pi = \pi\Pi$$