

Probabilità e Statistica per l'Informatica

github.com/martinopiaggi/polimi-notes

2021-2022

Contents

1	Probabilità	4
1.1	Evento	4
1.1.1	Fattoriale e permutazioni	4
1.1.2	Coefficiente binomiale	5
1.1.3	Indipendenza Stocastica	5
1.1.4	Probabilità condizionata	5
1.2	Teorema di Bayes	5
1.2.1	Formula di fattorizzazione	6
1.2.2	Indipendenza condizionale	6
2	Variabili aleatorie discrete	6
2.1	Media, Valore Atteso o Speranza Matematica	6
2.2	Legge di Bernoulli	6
2.3	Legge di Poisson	7
2.4	Distribuzione binomiale	7
2.4.1	Approssimazione binomiale con Normale	7
2.5	Indicatrice evento	7
2.6	Distribuzione Geometrica	7
2.7	Distribuzione Ipergeometrica	8
3	Variabili Aleatorie Assolutamente Continue	8
3.1	Legge uniforme del continuo	8
3.2	Legge di Weibull	9
3.3	Legge Gamma	9
3.4	Legge Gaussiana	9
3.5	Funzione di sopravvivenza	9
3.6	Modello scala posizione	10
3.7	Introduzione trasformazioni variabili aleatorie	10
3.8	Quantile	11
4	Medie e varianze notevoli	12
4.1	Variabile Aleatoria Uniforme su α e β	12
4.2	Variabile Aleatoria Gaussiana	12
4.3	Variabile Aleatoria Esponenziale	12
4.4	Variabile Aleatoria Binomiale	12
4.5	Variabile Aleatoria Poissoniana	12
4.6	Variabile Aleatoria Geometrica	12
4.6.1	Disuguaglianza di Markov	12
4.6.2	Disuguaglianza di Chebychev	12
5	Vettori aleatori	13
5.1	Funzione densità congiunta	13
5.2	Densità marginale	13
5.3	Funzione di ripartizione congiunta	13
5.4	Momenti	13
5.4.1	Generatrice dei momenti	14
5.5	Vettori di v.a. indipendenti	14
5.6	Trasformazioni vettori aleatori	14
5.6.1	Max E Min di v.a. indipendenti	15
5.6.2	Valore atteso di funzione di vettori aleatori	15
5.6.3	Media campionaria/empirica	16

5.6.4	Varianza campionaria	16
5.6.5	Deviazione standard campionaria	16
5.6.6	Legge debole dei grandi numeri	16
5.6.7	Legge forte dei grandi numeri	16
5.7	Vettori e matrici	16
5.8	Matrice di varianza-covarianza	16
5.9	Gaussiana multivariata	16
5.10	Teorema Centrale del Limite	17
5.10.1	Correzione del continuo	17
6	Distribuzioni in Statistica	17
6.1	Gamma	17
6.2	Chi quadrato	17
6.3	T di student	17
6.4	F di fisher	18
6.5	Somma di variabili Gamma indipendenti	18
6.6	Somma di Chi Quadrato	18
6.7	Quantili della F di Fisher	18
7	Inferenza parametrica	18
7.1	Stima puntuale	18
7.2	Distorsione di uno stimatore	18
7.2.1	Errore quadratico medio MSE	19
7.3	Consistenza	19
7.3.1	Consistenza debole	19
7.3.2	Consistenza in media 2	19
7.3.3	Consistenza forte	19
8	Metodi per la stima puntuale	19
8.1	Principio di sostituzione	19
8.2	Principio dei momenti	20
8.3	MLE, Maximum likelihood estimation	20
9	Stima intervallare IC	20
9.1	Quantità pivotale	20
10	Test delle ipotesi	21
10.1	Singola popolazione	21
10.1.1	Ipotesi statistica	21
10.1.2	Ipotesi nulla H_0	21
10.1.3	Ipotesi semplice e composta	21
10.1.4	Errore di prima specie e di seconda specie	21
10.1.5	P-value	22
10.1.6	Curva OC	22
10.1.7	Potenza del test	22
10.1.8	Test Z	22
10.1.9	Test T	22
10.2	Test su due popolazioni	22
10.2.1	Test T	22
10.2.2	Test F	23
10.2.3	Test χ^2 di buon adattamento	23

1 Probabilità

- 01. Variabili Aleatorie Discrete
- 02. Variabili Aleatorie Assolutamente Continue
- 03. Medie e varianze notevoli
- 03. Vettori aleatori

1.1 Evento

Il mattone elementare della logica probabilistica: **evento**. Formalmente, un evento è un insieme di esiti possibili, ovvero un sottoinsieme dello spazio campionario. Lo spazio campionario è l'insieme di tutti i possibili risultati di un esperimento.

$E \in \Omega$ dove Ω è l'insieme universale dei casi elementari.

$A \cup B$ indica che si verifica o A o B

$A \cap B$ indica che si verifica sia A che B

$A \setminus B$ indica che si verifica A ma non B

$A \triangle B$ indica che $(A \setminus B) \cup (B \setminus A)$

Legge di De Morgan interpretata con apice 'c' .. cioè la negazione è il complementare dell'evento.

La misura di Probabilità è una funzione che associa un evento a un numero P .

La funzione probabilità ha le seguenti proprietà:

- P è compreso tra 0 e 1
- $P(\Omega)$ è 1
- Gli eventi sono a 2 a 2 incompatibili, cioè disgiunti. $\sum \cup E_n = \sum P(E_n)$

Def: probabilità uniforme = ogni evento ha la stessa probabilità di accadere

cioè $P(E) = \frac{|E|}{|\Omega|}$

in gergo: casi favorevoli / casi possibili

$$P(A) + P(A^c) = 1$$

Importante la formula inversa!!

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

1.1.1 Fattoriale e permutazioni

$n!$ cioè tutte le possibili permutazioni, cioè disposizioni (senza ovviamente la ripetizione).

Ese: in quanti modi posso disporre 4 lettere? $4 * 3 * 2 * 1 = 24$

La prima lettera la posso scegliere tra n possibilità, la seconda tra $n - 1 \dots$

1.1.2 Coefficiente binomiale

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

Ci permette di contare tutti i possibili sottoinsiemi di cardinalità k da un insieme di cardinalità n . Ovviamente a livello insiemistico $\{a, b\} = \{b, a\}$ di conseguenza è ovvio anche capire che il coefficiente binomiale non considera l'ordine e di conseguenza vede elementi come ab e ba come lo stesso.

1.1.3 Indipendenza Stocastica

$$P(A \cap B) = P(A)P(B)$$

Due elementi stocasticamente indipendenti hanno questa relazione.

La cui generalizzazione risulta intuitivamente:

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

ma attenzione! La relazione sopra vale solo se valgono le condizioni a 2 a 2:

$$P(A \cap C) = P(A)P(C)$$

$$P(A \cap B) = P(A)P(B)$$

$$P(B \cap C) = P(B)P(C)$$

1.1.4 Probabilità condizionata

$$P(A \cap B) = P(A|B)P(B)$$

La probabilità condizionata di A rispetto a B è la probabilità che si verifichi A sapendo che si è verificato B .

Da questa relazione abbiamo un importante risultato, il teorema di Bayes:

1.2 Teorema di Bayes

Il teorema di Bayes, anche noto come regola di Bayes, è una formula matematica utilizzata per determinare la probabilità condizionata degli eventi. In sostanza, il teorema di Bayes descrive la probabilità di un evento basata sulla conoscenza pregressa delle condizioni che possono essere rilevanti per l'evento in questione.

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

e sappiamo anche che

$$P(B|A) = 1 - P(B^c, A)$$

Questa relazione vale **SOLO** per l'elemento condizionato, non il condizionante!!!

Il teorema di Bayes ha poi un'ulteriore generalizzazione:

$$P(A) = \sum P(A|B_i)P(B_i)$$

1.2.1 Formula di fattorizzazione

$$P(A \cap B) = P(A|B)P(B)$$

A quanto pare la formula della probabilità condizionata di un evento rispetto ad un altro evento è una ‘formula di fattorizzazione’, buona a sapersi.

In effetti risulta effettivamente buona da sapere in piccolo esercizietti.

1.2.2 Indipendenza condizionale

$$P(A \cup B|E) = P(A|E)P(B|E)$$

Come suggerisce il nome questa formula sta ad indicare la indipendenza condizionale. Cioè A e B sono sempre **dipendenti** eccetto per la condizione E. In tal caso allora sono indipendenti .. appunto una indipendenza condizionata da E.

2 Variabili aleatorie discrete

Da adesso in poi non parliamo più di eventi ma di variabili aleatorie, cioè passiamo dalla descrizione di eventi (ad esempio, il lancio di un dado e l’evento “ottenere un 4”) alla descrizione di quantità numeriche (ad esempio, la variabile aleatoria associata al lancio del dado e il valore numerico 4 che essa può assumere).

Questo passaggio è utile in quanto permette di formalizzare il concetto di probabilità: concetto semplice ma di difficile e sottile comprensione, fondamentale per il resto del corso.

Una variabile aleatoria è una funzione da Ω a \mathbb{R} , dove Ω è lo spazio degli eventi.

Con l’introduzione di variabili aleatorie possiamo quindi parlare di Funzioni:

- funzione di ripartizione discreta $F_x(X) = P(X < x)$
 - monotona
 - $\lim_{x \rightarrow -\infty} F(X) = 0$
 - $\lim_{x \rightarrow +\infty} F(X) = 1$
 - continuità da destra $\lim_{y \rightarrow x+} F(Y) = F(X)$
- funzione di densità discreta $f(X) = P(X = x)$
 - detta anche di massa, talvolta indicata con $p(x)$
 - $p(x) = F(X) - F(X^-)$

2.1 Media, Valore Atteso o Speranza Matematica

$$E(X) = \sum_{i=1} x_i f(x)$$

Si commenta da solo, profanamente la si può vedere come ‘media pesata delle probabilità’ e corrisponde ad appunto il valore che ci si aspetta. Con un po’ di fantasia può essere visto come il ‘baricentro’ di un.. stronzate varie.. La cosa più importante che ti devi portare a casa del valore atteso è la sua **linearità**, è la svolta.

2.2 Legge di Bernoulli

X ha legge bernoulliana di parametro p, se è una var. aleatoria con densità discreta definita da:

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

2.3 Legge di Poisson

La legge di Poisson è molto importante in statistica e in particolare in analisi dei dati quando ci si occupa di eventi rari, cioè eventi che hanno una bassa probabilità di accadere es: le telefonate in arrivo ad un call center, gli errori in un processo di produzione, gli incidenti stradali in una determinata zona, etc.

X var. aleatoria si dice che rispetta la legge di Poisson se ha distribuzione che rispetta la relazione:

$$P(X) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Sta roba risulta rilevante poichè essa può approssimare asintoticamente la distribuzione binomiale. Prende infatti il nome di ‘distribuzione degli eventi rari’ poichè tale approssimazione è giustificata nei casi in cui i parametri n e p siano rispettivamente molto grande e molto piccolo.

2.4 Distribuzione binomiale

Dato un processo di Bernoulli, esso può essere definito binomiale. La sua distribuzione viene definita da:

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Da interpretare come la probabilità di k successi con n tentativi.

NB: Processo di Bernoulli ci dice proprio che noi stiamo valutando una variabile che è o “positiva” o “negativa” o “1” o “0” ... “successo” “fallimento”. Quindi il successo ha una certa probabilità p , il fallimento $1-p$.

2.4.1 Approssimazione binomiale con Normale

Per $np \geq 5$ e $np(1-p) \geq 5$ la binomiale si può approssimare a una normale.

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

Approssimazione binomiale con Poisson

Se n è grande e p è piccolo, la distribuzione binomiale può essere invece approssimata con quella di Poisson con parametro $\lambda = np$.

2.5 Indicatrice evento

L’indicatrice è quindi una variabile aleatoria discreta che assume valori 0 o 1 a seconda che l’evento si verifichi o meno.

$$I(A) = \begin{cases} 0 & \text{quando } \omega \notin A \\ 1 & \text{quando } \omega \in A \end{cases}$$

2.6 Distribuzione Geometrica

domanda: “qual è la probabilità di avere il **PRIMO** successo in una serie di n tentativi al k -esimo tentativo?”

risposta:

$$P(X = k) = p(1-p)^{k-1}$$

2.7 Distribuzione Ipergeometrica

domanda: “qual è la probabilità di trovare k volte con n tentativi (senza reinserimento) solo r elementi in un insieme fatto di $r+b$ elementi?”

risposta:

$$P(X = k) = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{r+b}{n}}$$

Ese: Io voglio estrarre $k=3$ palline di tipo r =rosse. ci sono 7 di tipo b =bianco e ho un totale di n tentivi.. $r+b$ sono le palline iniziali ma ricorda che a ogni tentativo non reinserisco la pallina presa.

3 Variabili Aleatorie Assolutamente Continue

Level Up del caso discreto per tutti i problemi nel quale abbiamo a che fare con var. aleatorie di infiniti valori continue e non discreti (ese: temperatura, bho)

Abbiamo tutte le funzioni precedentemente riportate nel caso discreto:

- Funzione di ripartizione

$$F(X < x) = \int_{-\infty}^x f(t) dt$$

- Funzione di densità

$$F(X)' = f(x)$$

Ora funzioni di densità di probabilità continue notevoli:

3.1 Legge uniforme del continuo

La più semplice: **uniforme** su un insieme, ovvero che attribuisce la stessa probabilità a tutti i punti appartenenti ad un dato intervallo $[a,b]$ contenuto nell'insieme. In generale per $a > b$ si ha che

$$f_x(x) = \frac{1}{a-b}$$

Mentre la sua funzione di ripartizione $F(x)$, cioè $P(X \leq x)$ è

$$F(x) = x$$

Distribuzione esponenziale

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

errore comune ... non dimenticarti della funzione nel caso di $x < 0$, soprattutto quando fai l'integrale per trovare la funzione di ripartizione.

3.2 Legge di Weibull

$$f_x(x) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha}$$

Come la distribuzione esponenziale descrive la “durata di vita” di un fenomeno privo di memoria, così la distribuzione di Weibull può descrivere la durata di vita per un fenomeno la cui “probabilità di morire” può variare nel tempo, in funzione di α . La distribuzione di Weibull con parametro $\alpha = 1$ è una distribuzione esponenziale, la quale infatti prevede tassi di guasto costanti nel tempo.

$\alpha < 1$ il tasso di guasto diminuisce nel tempo (alta ‘mortalità infantile’) $\alpha = 1$ tasso di guasto è invariante nel tempo $\alpha > 1$ tasso di guasto aumenta con il tempo

3.3 Legge Gamma

La distribuzione Gamma è la distribuzione di probabilità della variabile aleatoria definita come la somma di variabili aleatorie indipendenti e con distribuzione esponenziale.

$$f_x(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

con

$$\Gamma = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

3.4 Legge Gaussiana

Legge gaussiana $N(\mu, \sigma)$:

$$f(x) = \frac{e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}}{\sqrt{2\pi\sigma^2}}$$

con μ valore atteso, mentre σ^2 la varianza (di conseguenza σ è la deviazione standard). La legge Gaussiana di notevole importanza infatti come funzione di densità di probabilità prende il nome di distribuzione normale. Semplice modello per fenomeni complessi. nb: - simmetrica - $\int_{-\infty}^{+\infty} N(\dots) = 1$

La distribuzione normale gaussiana è simmetrica rispetto al valore atteso μ

3.5 Funzione di sopravvivenza

Utilizzata soprattutto per studiare la vita l'usura o il rodaggio d componenti, ma anche la mortalità di organismi.

$$S(x) = 1 - F(X) = P(X > x)$$

Possiamo quindi descrivere comportamenti come l' **assenza di memoria** , cioè che la probabilità di sopravvivere fino a un tempo $t+s$, sapendo che tale oggetto è sopravvissuto fino al tempo t , è uguale alla probabilità di un oggetto nuovo.

assenza di memoria $S(t+s)|S(t) = S(s)$ **usura** $S(t+s)|S(t) < S(s)$ **rodaggio** $S(t+s)|S(t) > S(s)$

La funzione di sopravvivenza ‘next level’ è l’Hazard Rate o Intensità di rischio o Intensità di Guasto, indicata come $h(t)$. Introduciamo le seguenti relazioni:

$$e^{-H(t)} = S(t)$$

dove

$$H(t) = \int_0^t h(t) dt$$

Di conseguenza le relazioni riportate qualche riga fa le possiamo riscrivere osservando la nostra funzione $h(s)$, cioè l'*intensità di guasto*. **assenza di memoria** $h(s)$ costante **rodaggio** $h(s)$ decrescente **usura** $h(s)$ crescente modellizzare una misurazione di una variabile aleatoria?

3.6 Modello scala posizione

$$X = \mu + \sigma\epsilon$$

Dove μ è la misura d'interesse, σ è la precisione di misura e ϵ è la variabile aleatoria che indica l'errore di misura. La funzione di ripartizione risulta essere:

$$F_{\mu,\sigma} = F_{\epsilon}\left(\frac{x - \mu}{\sigma}\right)$$

Valore atteso di una variabile ass. continua

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

in analogia con (ricordiamo) il caso discreto:

$$E[x] = \sum_x x_i p_i$$

Importante proprietà del valore atteso è la linearità.

$$E[aX + b] = aE[X] + b$$

NB: Il valore atteso di una costante, è una costante. quindi, per il caso del modello scala posizione:

$$E[x] = \mu + \sigma E[\epsilon]$$

altra proprietà degna di nota del valore atteso:

$$se P(a \leq x \leq b) allora a \leq E(x) \leq b$$

3.7 Introduzione trasformazioni variabili aleatorie

Le trasformazioni tra variabili aleatorie sono v.a. definite $y = g(x)$, cioè funzioni di v.a. In esercizi di questo tipo scappano sempre errorini di distrazione e di ragionamento riguardo il dominio di definizione delle funzioni.

Le linee da seguire sono:

- 1) ragionare preventivamente sul dominio delle funzioni
- 2) spesso è utile cercare la $F_y(Y)$ e poi da lì tramite semplici uguaglianze ricondursi a $F_x(g(y))$
- 3) da $F_y(Y)$ con una bella derivata trovi $f_y(Y)$ se necessario

Finchè la trasformazione è invertibile e cioè $g(x)$ è monotona, non c'è nessun problema, infatti le relazioni sono come te lo aspetti. Quando invece g non è invertibile bisogna fare qualche magheggio in più, ad esempio suddividi in intervalli la funzione. Ad ogni modo il ragionamento da applicare è lo stesso della composizione di funzioni. Come scritto precedentemente bisogna ragionare su domini e sulle relazioni tra le variabili aleatorie. Per quanto riguarda le trasformazioni nel **caso discreto**: si ha densità:

$$f_y(y_j^*) = \sum_{i=g(x_i)=y_j^*} f_x(x_i)$$

e ripartizione come:

$$F(Y) = F(g(x)) = P(g(x) < y) = P(x < g^{-1}(y)) = F_x(g^{-1}(y))$$

Per quanto riguarda le trasformazioni nel **caso continuo**: solita def:

$$F(Y) = F(g(x)) = P(g(x) < y) = P(x < g^{-1}(y)) = F_x(g^{-1}(y))$$

e se voglio calcolarmi densità:

$$f_y(y) = \frac{dF_Y(y)}{dy} = \frac{dg^{-1}(y)}{dy} f_y(g^{-1}(y))$$

3.7.0.0.1 Valore atteso per le trasformazioni caso discreto:

$$\mathbb{E}[y] = \mathbb{E}[g(x)] = \sum g(x) f_x(x)$$

caso continuo:

$$\mathbb{E}[y] = \mathbb{E}[g(x)] = \int_{\mathbb{R}} g(x) f_x(x) dx$$

NB: nessuna trasformazione di funzione .. la formula è ‘come te la aspetteresti’

3.7.0.0.2 Varianza La varianza è definita come

$$Var(x) = \sigma^2 = E((x - m)^2)$$

dove $m = E(X)$. Si tratta quindi del valore atteso della ‘distanza quadratica’ tra la v.a. x e la sua media/valore atteso. **caso discreto**:

$$Var(x) = \sum_x (x - m)^2 f(x)$$

caso continuo:

$$Var(x) = \int_{-\infty}^{+\infty} (x - m)^2 f(x) dx$$

Deviazione standard: $\sigma = \sqrt{Var(x)}$ Importante proprietà dalla varianza: $var(Ax + b) = a^2 var(x)$ *NB: per dimostrare questa roba serve ricordarsi che il $\mathbb{E}[a] = a$ per ogni a costante.*

Può capitare di avere una certa varianza/media di una certa variabile aleatoria di un certo esperimento. Si ha la necessità di normalizzare la variabili aleatoria per evitarne variazioni causate ad esempio dal cambio di unità di misura. La varianza la si normalizza con il valore atteso.

$$\frac{\sqrt{Var(x)}}{\mathbb{E}[x]}$$

A proposito di normalizzare, standardizzazione/normalizzazione di v.a:

$$T := \frac{x - \mu}{\sigma}$$

T è appunto la var.a. standardizzata, con μ e σ rispettivamente la media e la deviazione standard. Il cuore fondamentale della standardizzazione è che se T è una var. a. normalizzata allora avrà **sempre** $\mu = 0$ e $\sigma = 1$

3.8 Quantile

z_α si dice che è il quantile di α se

$$P(X > z_\alpha) = \alpha$$

Il quantile di 0.5 prende il nome di *mediana*. I quantili di ordine 1/4 e 2/4 e 3/4 prendono il nome di *quartili*.

4 Medie e varianze notevoli

4.1 Variabile Aleatoria Uniforme su α e β

$$\mathbb{E}[x] = \frac{\alpha + \beta}{2}$$
$$Var[x] = \frac{(\beta - \alpha)^2}{12}$$

4.2 Variabile Aleatoria Gaussiana

$$\mathbb{E}[x] = \mu$$
$$Var[x] = \sigma^2$$

4.3 Variabile Aleatoria Esponenziale

$$\mathbb{E}[X] = \frac{1}{\lambda}$$
$$var[x] = \frac{1}{\lambda^2}$$

4.4 Variabile Aleatoria Binomiale

$$\mathbb{E}[X] = np$$
$$var[x] = np(1 - p)$$

4.5 Variabile Aleatoria Poissoniana

$$\mathbb{E}[X] = \lambda$$
$$var[x] = \lambda$$

4.6 Variabile Aleatoria Geometrica

$$\frac{1}{p}$$
$$\frac{(1 - p)}{p^2}$$

In alcune situazioni può succedere di avere a disposizione il valore atteso e/o la varianza di una var. ale. , ma di non averne la funzione di probabilità. Le disuguaglianze di Markov e Chebychev fanno comodo in queste situazioni, perchè ci permettono di ‘stimare’ / ‘maggiorare’ $p(x)$.

4.6.1 Disuguaglianza di Markov

$$P(X > c) \leq \frac{\mathbb{E}[x]}{c}$$

4.6.2 Disuguaglianza di Chebychev

$$P(|X - \mathbb{E}[X]| > r) \leq \frac{var[x]}{r^2}$$

5 Vettori aleatori

Introduzione vettori di più variabili aleatorie. Nel caso di $n=2$ è comodo utilizzare tabelle per rappresentare i valori di probabilità.

$$(X_1, X_2, X_3, \dots) : \Omega \rightarrow \mathbb{R} \times \mathbb{R} \times \mathbb{R} \dots$$

5.1 Funzione densità congiunta

caso discreto:

$$f(x, y) = \sum_i \sum_j P(x = X_i, y = Y_j)$$

caso continuo:

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P(x, y) dx dy$$

5.2 Densità marginale

Fun Fact: si chiama così perchè (in situazioni con due vettori di dimensione $n = 2$) è la probabilità che si trova ‘ai margini’ della tabella tipicamente utilizzata per rappresentare le probabilità. Molto semplicemente è la probabilità di una delle variabili fissate le altre.

Nel caso $n=2$: - caso discreto:

$$f_{x_i}(x) = \sum_y f(x_i, y)$$

- caso continuo:

$$f_{x_i}(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

5.3 Funzione di ripartizione congiunta

$$F(x, y) = P(X \leq x, Y \leq y)$$

caso discreto:

$$F(x, y) = \sum_{x < X} \sum_{y < Y} f(x, y)$$

caso continuo:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy$$

Posso ricavare la funzione di ripartizione di x dalla funzione di ripartizione congiunta:

$$F_x(x) = \lim_{y \rightarrow \infty} F_{x,y}(x, y)$$

5.4 Momenti

Sia X una v.a. ass. cont. o discreta tale che $|X^k|$ ammetta valore atteso. Il numero $\mathbb{E}[X^k]$ è detto momento k -esimo.

5.4.1 Generatrice dei momenti

Funzione generatrice degli eventi, dove il momento n-esimo è la derivata ennesima della funzione ϕ . caso discreto:

$$\phi(X) = \sum_x e^{tx} p(x)$$

caso continuo:

$$\phi(X) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx$$

5.5 Vettori di v.a. indipendenti

Cosa si intende per indipendenza?

$$f(x, y) = f_x(x) f_y(y)$$

L'indipendenza di vettori aleatori risulta comodo in diversi conti. Come si dimostra/controlla se c'è la dipendenza (nel pratico)? Controlli che il prodotto delle densità marginali (ai margini della tabella) sia uguale al rispettivo elemento dentro la tabella. *lo devi fare per ognuno degli elementi, DEVE VALERE PER TUTTI* NB: (NB forse scontato ma occhio) negli esercizi risulta particolarmente furbo controllare le caselle dove c'è lo zero ... così che si 'sgama' subito se le variabili non sono indipendenti.

5.6 Trasformazioni vettori aleatori

ysy:

$$g(x_1, x_2, \dots, x_n) = (g(x_1), g(x_2), \dots, g(x_n))$$

Esempio classicone: v.a. S definita come $S = x + y$ caso discreto:

$$f(S) = \sum_x f_{x,y}(x, S-x)$$

caso discreto con v.a. indipendenti:

$$f(S) = \sum_x f_x(x) f_y(S-x)$$

caso continuo:

$$f(S) = \int_{-\infty}^{+\infty} f_{x,y}(x, S-x) dx$$

caso continuo con v.a. indipendenti:

$$f(S) = \int_{-\infty}^{+\infty} f_x(x) f_y(S-x) dx$$

NB: 'Questo procedimento' è valido per le somme di variabili aleatorie, non confonderti con ad esempio calcolare l'integrale doppio per trovare la funzione di partizione in una 'zona' di piano.

5.6.1 Max E Min di v.a. indipendenti

Sia $\max(x) = \max(X_1, \dots, X_n)$ e $\min(x) = \min(X_1, \dots, X_n)$ allora:

$$F_{\max}(x) = \prod_{i=1}^n F_i(x)$$

$$F_{\min}(x) = 1 - \prod_{i=1}^n S_i(x)$$

NB: se i.i.d. (indipendenti identicamente distribuite, cioè stessa legge e stessi parametri) allora: $F_{\max}(x) = F_1(x)^n$ e $F_{\min}(x) = S_1(x)^n$

5.6.2 Valore atteso di funzione di vettori aleatori

Molto molto gradito il fatto che non è necessario alcun tipo di ragionamento su funzioni o trasformazioni. caso discreto:

$$\mathbb{E}[g(x, y)] = \sum_{(x, y)} g(x, y) f(x, y)$$

caso continuo:

$$\mathbb{E}[g(x, y)] = \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy$$

NB: questi esempi sono fatti con due $n = 2$.. ma ovviamente i vettori possono essere fatti di n variabili aleatorie.

5.6.2.1 Somma di varie funzioni notevoli Somma di gaussiane fa gaussiane. Somma di binomiali con p comuni sono ancora binomiali. Somma di v.a. con distribuzione di Poisson (con parametri uguali) sono Poisson con parametri somma di parametri.

5.6.2.2 Covarianza Definizione generale:

$$\text{Cov}(X, Y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)]$$

Definizione che ci piace:

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

Da notare come il caso particolare di $X=Y$ corrisponde alla definizione di varianza.

$$\text{Cov}(X, X) = \text{Var}(X) = \mathbb{E}[(x - \mu_x)^2]$$

La covarianza è un 'indice' che indica di quando le variabili si influenzano a vicenda, di quanto variano insieme e quindi della loro reciproca dipendenza.

Se due variabili sono **indipendenti** allora varrà sicuramente che la $\text{Cov}[X, Y] = 0$ MA NON VALE VICEVERSA! Se la covarianza è uguale a zero non si può dire nulla! Sulla dipendenza delle variabili.

5.6.2.2.1 Indice di correlazione di Pearson

$$\rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sigma_x \sigma_y}$$

L'indice di correlazione di Pearson (anche detto coefficiente di correlazione lineare) tra due variabili aleatorie è un indice che esprime un'eventuale relazione di linearità tra esse. Ha sempre un valore compreso tra +1 e -1, dove +1 e -1 corrispondono alla perfetta correlazione lineare (positiva/negativa) e 0 corrisponde a un'assenza di correlazione lineare.

5.6.3 Media campionaria/empirica

Media empirica delle variabili aleatorie X_1, X_2, \dots è la variabile aleatoria:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

Per linearità del **valore atteso**, se tutte le v.a. hanno **stessa legge** allora:

$$\mathbb{E}[\bar{X}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X_i]$$

5.6.4 Varianza campionaria

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

5.6.5 Deviazione standard campionaria

$$S_n = \sqrt{S_n^2}$$

5.6.6 Legge debole dei grandi numeri

$$\lim_{n \rightarrow \infty} P\left(\left| \frac{\sum X_i}{n} - \mu \right| \leq \epsilon\right) = 1$$

$\forall \epsilon > 0$

5.6.7 Legge forte dei grandi numeri

$$P\left(\lim_{n \rightarrow \infty} \frac{\sum X_i}{n} = \mu\right) = 1$$

5.7 Vettori e matrici

Sia $Y = [Y_1, Y_2, Y_3, Y_4, \dots]^T$ vettore aleatorio, allora:

$$\mathbb{E}[Y] = \mu = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \dots \\ \dots \end{bmatrix}$$

5.8 Matrice di varianza-covarianza

La matrice Σ rappresenta la variazione di ogni variabile rispetto alle altre (inclusa se stessa). È simmetrica.

5.9 Gaussiana multivariata

Un vettore aleatorio $X = [X_1, X_2, \dots]$ si dice gaussiano multivariato $N(\mu, \Sigma)$ se esiste una matrice A tale che $AA^T = \Sigma$ e un vettore $Z = (Z_1, \dots, Z_m)^T$ iid $N(0, 1)$ tale che

$$X = AZ + \mu$$

È una generalizzazione della distribuzione normale a dimensioni più elevate. La sua importanza deriva principalmente dal fatto che è spesso utilizzata per descrivere, almeno approssimativamente, un qualunque insieme di v.a a valori reali (possibilmente) correlate, ognuna delle quali è clusterizzata attorno ad un valore medio.

5.10 Teorema Centrale del Limite

$$\lim_{n \rightarrow \infty} \frac{x - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx N(0, 1)$$

quindi

$$P\left(\lim_{n \rightarrow \infty} \frac{x - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq \epsilon\right) \approx \phi(\epsilon)$$

5.10.1 Correzione del continuo

La correzione di continuità è una modifica dell'intervallo di integrazione che si applica quando si approssima una distribuzione discreta con una distribuzione continua (nel nostro caso la normale). La correzione di continuità consiste nell' ampliare / allargare di 0.5 (convenzionalmente) gli estremi dell'intervallo sul quale si integra la densità di probabilità continua usata per approssimare una distribuzione discreta. Con tale correzione, la precisione dell'approssimazione è maggiore.

05. Distribuzioni in Statistica

06. Inferenza parametrica

07. Test delle ipotesi

6 Distribuzioni in Statistica

Principali distribuzioni utilizzate in statistica:

6.1 Gamma

$$\Gamma(\alpha, \lambda) := \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{I}_{(0, +\infty)}(x)$$

dove Γ è la **Gamma di Eulero**. La distribuzione Gamma è utilizzata per modellare il tempo di attesa di un certo numero di eventi rari.

6.2 Chi quadrato

$$X^2(n) := Z_1^2 + \dots + Z_n^2$$

con $Z_i \sim N(0, 1)$ Inoltre $X^2(n)$ può essere scritto come $\Gamma(n/2, 1/2)$. Chi quadrato viene utilizzata per il test di ipotesi sulle varianze.

6.3 T di student

La distribuzione T di Student viene utilizzata per il test di ipotesi sulla media di una popolazione quando la deviazione standard della popolazione non è nota.

Con $Z \sim N(0, 1)$ e $X_n^2 = X^2(n)$ è:

$$T_n := \frac{Z}{\sqrt{\frac{X_n^2}{n}}}$$
$$\frac{\bar{X} - u}{\frac{S_n}{\sqrt{n}}} \sim t(n-1)$$

Belle storielle fantasy del perchè si chiama ‘..di student’

6.4 F di fisher

La distribuzione F di Fisher viene utilizzata per il test di ipotesi sulla differenza tra le varianze di due popolazioni.

$$F_{n,m} := \frac{X_n^2/n}{X_m^2/m}$$

6.5 Somma di variabili Gamma indipendenti

Se $X_1 \sim \Gamma[\alpha_1, \lambda]$ e $X_2 \sim \Gamma[\alpha_2, \lambda]$ v.a. indipendenti. Allora

$$X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$$

6.6 Somma di Chi Quadrato

Se $X_1 = X^2(n)$ e $X_2 = X^2(m)$ e sono indipendenti allora

$$X_1 + X_2 = X^2(n + m)$$

6.7 Quantili della F di Fisher

$$f_{1-\alpha, m, n} = \frac{1}{f_{\alpha, n, m}}$$

Questa formula è importante, ad esempio, nella costruzione di intervalli di confidenza per la varianza di due popolazioni normali.

7 Inferenza parametrica

Dire qualcosa (FARE INFERENZA) sui parametri incogniti, usando i dati (OSSERVAZIONI CAMPIONARIE).

7.1 Stima puntuale

Gli stimatori puntuali forniscono, per ogni realizzazione campionaria, un solo valore come “stima” del parametro incognito θ (o di una sua funzione $k(\theta)$). Lo stimatore è quindi una v.a. (o vettore aleatorio) e la stima è un numero (o un vettore di numeri).

7.2 Distorsione di uno stimatore

La distorsione di uno stimatore è una misura della bontà di uno stimatore puntuale, definita come la differenza tra il valore atteso dello stimatore e il parametro che si sta cercando di stimare. Uno stimatore si dice non distorto se la sua distorsione è nulla.

$$bias = \mathbb{E}(\hat{\theta}) - \theta$$

Quando uno stimatore **non** è distorto, allora si dice corretto.

$$\mathbb{E}(\hat{\theta}) - \theta = 0$$

Uno stimatore può essere anche **asintoticamente non distorto**. Cioè è distorto magari, ma il limite del valore atteso dello stimatore per n-osservazioni che tendono ad infinito è uguale al parametro.

7.2.1 Errore quadratico medio MSE

L'errore quadratico medio (MSE) è un'altra misura della bontà di uno stimatore, definita come la somma della varianza dello stimatore e della sua distorsione al quadrato.

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \tau(\theta))^2]$$

Ma anche (decomposizione):

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (bias(\hat{\theta}, \theta))^2$$

Un buon stimatore dovrebbe avere un MSE basso.

7.3 Consistenza

7.3.1 Consistenza debole

La consistenza debole è una proprietà desiderabile di uno stimatore $\hat{\theta}$, che indica che lo stimatore converge al parametro che si sta cercando di stimare al crescere del numero di dati campionari.

$$\lim_{n \rightarrow \infty} P_{\theta}(|\hat{\theta} - \theta| > \epsilon) \rightarrow 0$$

In poche parole uno stimatore è consistente se tende a fornire stime sempre più precise del parametro al crescere del campione.

7.3.2 Consistenza in media 2

La consistenza in media 2 ci dice che lo stimatore converge in media quadratica al valore del parametro da stimare.

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\theta}[(\hat{\theta} - \theta)^2] \rightarrow 0$$

Lo stimatore consistente in media 2 è preferibile a uno stimatore solo debolmente consistente perché fornisce una stima più precisa del parametro. Se la successione di stimatori è consistente in media quadratica allora è anche debolmente consistente.

Tuttavia, la consistenza in media 2 non è sempre garantita, anche se lo stimatore è consistente in modo debole.

7.3.3 Consistenza forte

La consistenza forte implica che lo stimatore converge al parametro quasi certamente, cioè con probabilità pari a 1. Questa è una proprietà ancora più forte rispetto alla consistenza debole.

$$\mathbb{P}(\lim_{n \rightarrow \infty} \hat{\theta} = \theta) = 1$$

8 Metodi per la stima puntuale

8.1 Principio di sostituzione

Stimo il mio parametro con la 'sua versione empirica'. mi baso sulla legge forte dei grandi numeri, e pongo quindi la mia media empirica uguale al parametro da stimare. Funzione cumulata empirica, e quindi anche di sopravvivenza. Metodo che funziona nel mondo della fantasia.

8.2 Principio dei momenti

Determinare il parametro θ univocamente dai momenti. Cioè cerchiamo il parametro che eguaglia i momenti ‘teorici’ con i momenti empirici. Ci troviamo quindi un sistema.

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

L’idea è simile al metodo di sostituzione.

8.3 MLE, Maximum likelihood estimation

Criterio della massima verosimiglianza. Si basa su determinare il θ che con maggior probabilità ha generato i dati osservati. La funzione di verosimiglianza:

$$L(\theta, x) = \prod_i^n f(x_i)$$

Ma usiamo quasi sempre la log-verosimiglianza. $l(\theta, x) = \ln(L(\theta, x))$ La log-verosimiglianza ha gli stessi punti di massimo di L ma ha alcuni vantaggi: i prodotti diventano somme e gli esponenti diventano moltiplicazioni.

$$l(\theta, x) = \sum_i^n \ln(f(x_i))$$

- asintoticamente non distorta
- consistenza in media quadratica
- asintoticamente gaussiana

8.3.0.1 Principio d’invarianza funzionale per gli MLE Se $\hat{\Theta}$ è lo stimatore di θ allora lo stimatore di massima verosimiglianza per $\alpha = g(\theta)$ è $\hat{\alpha} = g(\hat{\theta})$

9 Stima intervallare IC

Per quantificare il grado di precisione associato alla stima puntuale di un parametro θ si ricorre alla cosiddetta stima intervallare, cioè si costruisce un intervallo di valori in cui si ha una elevata fiducia che cada il valore incognito del parametro. Gli estremi dell’intervallo sono calcolati per mezzo del campione dei dati osservati.

9.1 Quantità pivotale

Sia Q la quantità pivotale per la costruzione di intervalli di confidenza. Allora:

$$P_\theta(q_1 < Q < q_2)$$

dove q_1 e q_2 dipendono dal livello di confidenza α ma non dal parametro θ . Ci sono intervalli bilateri e unilateri. IC più stretto implica più precisione.

**Intervalli di confidenza per la media μ , applicato nel caso di una popolazione normale:

- varianza σ nota (Z):

$$P(\mu \in (\bar{X}_n \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}})) = 1 - \alpha$$

- varianza σ incognita (T):

$$P(\mu \in (\bar{X}_n \pm t_{1-\alpha/2} \frac{S_n}{\sqrt{n}})) = 1 - \alpha$$

NB: in caso di campione numeroso, il caso della T di Student è approssimabile da una normale

****Intervalli di confidenza per la varianza σ , applicato nel caso di una popolazione normale:**

- media μ nota:

$$P\left(\frac{S^2(n)}{X_{1-\alpha/2,n}^2} < \sigma^2 < \frac{S^2(n)}{X_{\alpha/2,n}^2}\right) = 1 - \alpha/2$$

- media μ incognita :

$$P\left(\frac{S^2(n-1)}{X_{1-\alpha/2,n-1}^2} < \sigma^2 < \frac{S^2(n-1)}{X_{\alpha/2,n-1}^2}\right) = 1 - \alpha/2$$

Proporzione p di una popolazione bernoulliana

$$P\left(\hat{p} - \frac{\sqrt{\hat{p}(1-\hat{p})}}{n} z_{1-\alpha/2} < p < \hat{p} + \frac{\sqrt{\hat{p}(1-\hat{p})}}{n} z_{1-\alpha/2}\right) = 1 - \alpha$$

Parametro λ di una popolazione poissoniana

$$P\left(\bar{X} - \frac{\sqrt{\bar{X}}}{n} z_{1-\alpha/2} < p < \bar{X} + \frac{\sqrt{\bar{X}}}{n} z_{1-\alpha/2}\right) = 1 - \alpha$$

Parametro λ di una popolazione esponenziale

$$P\left(\frac{1}{2n\bar{X}} X_{\alpha/2,2n}^2 < \lambda < \frac{1}{2n\bar{X}} X_{\alpha/2,2n}^2\right)$$

10 Test delle ipotesi

10.1 Singola popolazione

10.1.1 Ipotesi statistica

Un'ipotesi statistica è un'affermazione riguardo la media o la varianza di una determinata distribuzione.

10.1.2 Ipotesi nulla H_0

Un'ipotesi nulla/alternativa è una coppia di ipotesi nel quale la veridicità di una ipotesi sui campioni esclude l'altra.

10.1.3 Ipotesi semplice e composta

Un'ipotesi semplice è un'ipotesi la quale, se vera con i nostri campioni, ci permette di descrivere completamente la nostra distribuzione. Composta viceversa.

10.1.4 Errore di prima specie e di seconda specie

Un **errore di prima specie** è quando consideriamo H_0 non compatibile con i nostri dati ma in realtà è vera. Un **errore di seconda specie** è quando accettiamo l'ipotesi H_0 compatibile con la popolazione ma in realtà è falsa. Consideriamo gli errori di prima specie più gravi degli errori di seconda.

10.1.5 P-value

P-value è il valore più basso di α per cui **non accettiamo la nostra ipotesi** H_1 , quindi per α maggiori del p-value continueremo a non accettare l'ipotesi.

Un P-value quasi uguale a zero significa che siamo praticamente certi di non sbagliare rifiutando l'ipotesi H_0 . Un P-value molto maggiore indica invece che a qualsiasi livello ragionevole di significatività, non rifiutiamo l'ipotesi nulla; in questo caso si può anche dire che il test ci porta ad accettare l'ipotesi. Il P-value può essere difficile da calcolare con precisione usando le tavole, quindi negli esercizi ci limiteremo a fare stime intervallari di quest'ultime (in un software ovviamente possiamo facilmente trovare una stima puntuale).

10.1.6 Curva OC

β è definita come la probabilità di accettare l'ipotesi **nulla** in funzione del parametro che stiamo testando. Di conseguenza se il parametro incognito soddisfa l'ipotesi alternativa, la curva OC rappresenta una probabilità d'errore di II specie.

10.1.7 Potenza del test

$\pi(\theta) = 1 - \beta$ è la funzione in funzione del parametro incognito θ che è definita come la probabilità di rifiutare l'ipotesi nulla e quindi accettare l'ipotesi alternativa. *La potenza del test è la possibilità di rigettare l'ipotesi nulla quando è giusto farlo (cioè quando H_0 è falsa).* **Esempio:** Nel caso in l'ipotesi alternativa viene soddisfatta per $T_0 > t$ $\pi(\theta) = P_\theta(T_0 > t_\theta)$ con $\theta = \frac{\Theta}{\Theta_0}$

La potenza del test aumenta quanto maggiore è il numero dei soggetti del campione.

La potenza del test aumenta passando da un α di 0.05 a uno di 0.1 ma così facendo aumenta anche il rischio di rigettare una ipotesi nulla vera.

10.1.8 Test Z

Test su media di una popolazione gaussiana con varianza nota.

10.1.9 Test T

Test su media una popolazione gaussiana con varianza incognita.

10.2 Test su due popolazioni

10.2.1 Test T

Test su medie con varianze incognite ma uguali (f di Fisher).

$$T_0 = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{S_p^2(\frac{1}{n_x} + \frac{1}{n_y})}}$$

dove S_p^2 è la stima combinata della deviazione standard generale:

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

10.2.2 Test F

Il test F viene utilizzato per confrontare la varianza di due popolazioni normali indipendenti, assumendo che le popolazioni siano normalmente distribuite e abbiano la stessa media. Il test F é basato sulla distribuzione F di Fisher.

Caso unilaterale:

$$P\left(\frac{\frac{S_x^2}{\sigma_x^2}}{\frac{S_y^2}{\sigma_y^2}} > f_{0.01,n,m}\right) = P\left(\frac{\sigma_y^2}{\sigma_x^2} > f_{0.01,n,m} \frac{S_y^2}{S_x^2}\right) = P\left(\frac{\sigma_x^2}{\sigma_y^2} \leq f_{0.99,m,n} \frac{S_x^2}{S_y^2}\right)$$

10.2.3 Test χ^2 di buon adattamento

Il test chi-quadrato consiste nel verificare se le frequenze osservate corrispondono alle frequenze attese. Nel caso di una sola variabile di misura, si utilizza il test della bontà di adattamento del chi-quadrato. L'idea alla base del test è di confrontare i valori osservati nei dati e quelli attesi qualora l'ipotesi nulla fosse vera.

Dato f_i le frequenze osservate di ciascuna modalità, p_i le probabilità teoriche di ciascuna modalità e n la numerosità del campione e k il numero delle modalità/classi del campione.

$$X^2 = \sum_{i=1}^k \frac{(f_i - np_i)^2}{np_i}$$

Rifiuto H_0 se $X^2 \geq X_{v,\alpha}^2$ (dove α è la significatività e v sono $k - 1$).