Appunti di Matematica Applicata T-A

Corso tenuto dalla Professoressa Francesca Brini

J.Nanni

A.A. 2010/2011

Indice

Capitolo 1

Introduzione alla Probabilità e alla Statistica

1.1 Storia della Probabilità

La statistica nasce nel Rinascimento come "descrittiva". Essa serviva principalmente per il conteggio delle nascite, dei morti di uno **stato** (deriva appunto dalla parola "stato").
La probabilità nasce invece per ragioni pratiche nei giochi d'azzardo. Nel '600 abbiamo tre grandi studiosi: Pascal, Fermat e Huygens. Verso la fine del '600 abbiamo Bernoulli e Laplace, verso l'800 Gauss e Poisson.
Tra fine '800 e inizio '900 abbiamo Chebychev, Markov, Lyapounov, e come ultimo abbiamo Kolmogorov, colui che renderà la probabilità una vera e propria teoria matematica.

1.2 Definizioni base

Definizione 1.2.1 (Popolazione)

Insieme di individui e oggetti studiati rispetto ad una determinata caratteristica misurabile.

Definizione 1.2.2 (Campione)

Dalla popolazione si estrae casualmente un campione su cui viene solitamente fatta l'analisi.

Definizione 1.2.3 (Probabilità)

Prende in considerazione la popolazione e propone le aspettative per il campione.

Definizione 1.2.4 (Inferenza statistica)

Effettua il processo contrario della probabilità. Ovvero parte dal campione e in base a quello dice qualcosa sulla popolazione (es. exit poll, ecc...). Non si può fare inferenza statistica se non si conosce la probabilità.

Definizione 1.2.5 (Statistica descrittiva)

Analizza o la popolazione o il campione e sintetizza attraverso numeri o grafici particolari situazioni(Istogrammi, torte, ecc...).

Capitolo 2

Richiami di calcolo combinatorio

Serve principalmente per **contare**. È il primo modo semplice per contare la probabilità.

Teorema 2.0.1 (Principio fondamentale del calcolo combinatorio (o di enumerazione))

*Dati 2 esperimenti per cui il primo ha **m** esiti possibili e il secondo **n** esiti possibili, si ha che le **possibili sequenze ordinate sono***

$$m \cdot n$$

Generalizzando a N esperimenti, con $n_1, n_2, n_3, ..., n_N$ esiti la sequenza può variare in $n_1 \cdot n_2 \cdot n_3 \cdot ... \cdot n_N$ modi possibili.

Vediamo ora una semplice applicazione.

Esempio 2.0.1 (Targhe automobilistiche)

Quante possibili targhe possiamo formare?
Chiamiamo M il numero di targhe possibili, e immaginiamo ogni casella della targa come un esperimento.
In questo modo avremo che nella prima casella ci sono 26 *esiti possibili (ovvero le lettere dell'alfabeto), nella seconda ancora* 26 *esiti, nella terza* 10 *esiti (le cifre da 0 a 9), e così via.*
Applicando il Principio di enumerazione otteniamo:

$$26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 26 \cdot 26 = 456.976.000$$

Definizione 2.0.6 (Disposizioni semplici)

Dati n oggetti distinti si definiscono disposizioni di n elementi di classe k (con k≤ n) gli allineamenti che si possono formare prendendo k elementi distinti tra gli n oggetti dati.
Per determinare il numero $D_{n,k}$ delle disposizioni semplici si usa la formula seguente.

$$D_{n,k} = \frac{n!}{(n-k)!} = n(n-1)(n-2)\cdots(n-k+1)$$

Osservazione: gli elementi formati differiscono tra loro per qualche elemento o per l'ordine secondo cui gli oggetti sono stati disposti.

Esempio 2.0.2

Quanti allineamenti si possono formare prendendo 3 vocali diverse per volta?
Poichè le vocali sono 5, gli allineamenti richiesti corrispondono al numero delle disposizioni di 5 elementi di classe 3.

$$D_{5,3} = 5 \cdot 4 \cdot 3 = 60$$

Definizione 2.0.7 (Disposizioni con ripetizione)

Dati n oggetti distinti si definiscono disposizioni con ripetizione di n elementi di classe k, gli allineamenti che si possono formare prendendo k elementi non necessariamente distinti tra gli n oggetti dati.
Per determinare il numero $D^r_{n,k}$ delle disposizioni con ripetizione si usa la formula seguente.

$$D^r_{n,k} = n^k$$

Osservazione: gli allineamenti formati differiscono tra loro per qualche elemento, per il numero di volte in cui uno o più oggetti compaiono e per l'ordine secondo cui gli oggetti sono disposti.

Esempio 2.0.3

Determinare quanti numeri telefonici con 8 cifre si possono formare con i numeri 1,2,3,4,5,6,7,8,9.
Gli allineamenti richiesti possono, in questo caso, contenere più volte la stessa cifra, quindi il loro numero è dato da

$$D^r_{9,8} = 9^8 = 43.046.421$$

Definizione 2.0.8 (Permutazioni semplici)

Dati n oggetti distinti si definiscono permutazioni semplici di n elementi gli allineamenti che si possono formare prendendo tutti gli n elementi dati. Per determinare il numero P_n delle permutazioni semplici si usa la seguente formula.

$$P_n = n!$$

Osservazione: gli allineamenti formati differiscono tra loro per l'ordine.

Esempio 2.0.4

Quanti allineamenti si possono fare prendendo tutte e 5 le vocali?

$$P_5 = 5! = 120$$

Definizione 2.0.9 (Permutazioni con ripetizione)

Dati n oggetti, non necessariamente distinti, si definiscono permutazioni con ripetizione di n elementi che si possono formare prendendo tutti gli n elementi dati. Il numero delle permutazioni con ripetizione è dato dalla seguente formula.

$$P^r_{n=k_1+k_2+...+k_p} = \frac{n!}{k_1!k_2!...k_p!}$$

Dove $k_1, k_2, ..., k_p$ rappresentano le molteplicità dei p oggetti distinti.

Osservazione: gli allineamenti formati differiscono tra loro per l'ordine.

Esempio 2.0.5

Quanti numeri si possono ottenere allineando le cifre del numero 22466?
Notiamo che i numeri 2 e 6 si ripetono due volte, e il 4 una volta sola. Quindi abbiamo il seguente risultato.

$$P^r_{5=2+1+2} = \frac{5!}{2!1!2!} = 30$$

Definizione 2.0.10 (Combinazioni semplici)

Dati n oggetti distinti, si definiscono combinazioni semplici di n elementi di classe k, gli insiemi che si possono formare prendendo k elementi distinti fra gli n oggetti. Il numero di combinazioni semplici è dato dalla formula seguente.

$$C_{n,k} = \frac{n!}{(n-k)!k!} = \binom{n}{k} \quad \left(= \frac{D_{n,k}}{k!}\right)$$

Questo risultato viene chiamato anche ***coefficiente binomiale***

Osservazione: le combinazioni semplici differiscono l'una dall'altra per almeno un elemento contenuto.

Proprietà del coefficiente binomiale

- $$\binom{n}{0} = \binom{n}{n} = 1$$
- $$\binom{n}{1} = \binom{n}{n-1} = n$$
- $$\binom{n}{k} = \binom{n}{n-k} \quad \text{proprietà delle classi complementari}$$
- $$\binom{n}{k} = \binom{n}{k-1} \quad \text{proprietà di ricorrenza}$$

Esempio 2.0.6

Determinare il numero di ambi che si possono formare con i 90 numeri del lotto.
Dato che non importa l'ordine in cui si succedono gli ambi, il numero degli ambi sarà dato da

$$C_{90,2} = \frac{90 \cdot 89}{2!} = 4005$$

Definizione 2.0.11 (Combinazioni con ripetizione)

Dati n oggetti distinti, si definiscono combinazioni con ripetizione di n elementi di classe k, gli insiemi che si possono formare prendendo k elementi, non necessariamente distinti, tra gli n elementi dati.
Il numero di combinazioni con ripetizione è dato dalla seguente formula.

$$C^r_{n,k} = \frac{(n+k-1)!}{(n-1)!k!} = \binom{n+k-1}{k}$$

Esempio 2.0.7

Un chimico deve preparare composti utilizzando 3 dosi ricavate da 6 ingredienti diversi. Quanti composti, anche puri (cioè costituiti da 3 dosi dello stesso ingrediente) può preparare?
Il numero complessivo dei composti si ricava applicando la seguente formula.

$$C^r_{6,3} = \frac{(6+3-1)!}{5!3!} = 56$$

Capitolo 3

Fondamenti del calcolo della probabilità

Definizione 3.0.12 (Definizione classica di probabilità)

Dato un esperimento ed un certo evento A legato ad esso diciamo che

$$P(A) = \frac{n^\circ \text{ casi favorevoli}}{n^\circ \text{ casi totali}}$$

Questa definizione in certi casi risulta molto efficace, ma presenta alcuni problemi in altri casi.

- Se il numero di casi totali è infinito?
- Non sempre è possibile contare i casi;
- La definizione implica che gli esiti siano tutti equiprobabili, e questo non accade sempre.

Definizione 3.0.13 (Definizione frequentista di probabilità)

$$P(A) = \frac{n^\circ \text{ successi}}{n^\circ \text{ esperimenti effettuati}}$$

con n° successi=A verificato.

Questa definizione supera il problema 3 della definizione classica, in compenso ne genera altri.

- Il numero di esperimenti fatti deve essere abbastanza elevato;
- Se il numero di esiti è infinito non funziona la definizione;
- L'esperimento deve essere ripetibile e lo devo poter fare.

Queste definizioni presentano come si è visto molti problemi, la definizione di Kolmogorov attraverso i 3 assiomi ovvierà a tutti questi. Prima di introdurre gli assiomi diamo alcune definizioni di carattere insiemistico.

Definizione 3.0.14 (Spazio campione)

Chiamiamo Ω spazio campione, ovvero l'insieme di tutti gli esiti di un esperimento.

Definizione 3.0.15 (Evento dello spazio campione)

$A \subset \Omega$ si dice evento dello spazio campione.

Definizione 3.0.16 (Unione di eventi)

Siano $A, B \subset \Omega$, allora chiamiamo unione di eventi $A \cup B$, ovvero tutti gli esiti dell'esperimento che stanno in A o in B.

Definizione 3.0.17 (Intersezione di eventi)

Siano $A, B \subset \Omega$, allora chiamiamo intersezione di eventi $A \cap B$, ovvero gli esiti dell'esperimento che stanno in A e in B.

Definizione 3.0.18 (Eventi disgiunti)

Siano $A, B \subset \Omega$, allora A e B si dicono disgiunti se la loro intersezione ~~non~~ è vuota, ovvero

$$A \cap B = \emptyset$$

Definizione 3.0.19 (Evento complementare)

Sia $A \subset \Omega$, chiamiamo A^c complementare di A, per cui

$$A^c = \Omega \setminus A$$

Introduciamo ora gli assiomi di Kolmogorov.

Definizione 3.0.20 (Assiomi di Kolmogorov)

Dato uno spazio campione Ω e $A \subset \Omega$ diciamo che

1. $P(A) \in [0, 1]$;
2. $P(\Omega) = 1$;
3. *Dati $A_1, A_2, ..., A_n \subset \Omega$ eventi disgiunti si ha che*

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$$

Vediamo le proprietà di questi assiomi.

Teorema 3.0.2

Dati Ω e $A \subset \Omega$ allora

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

Dimostrazione:
Per la definizione di complementare sappiamo che $\Omega = A \cup A^c$, e che $A \cap A^c = \emptyset$, quindi applicando il secondo e il terzo assioma abbiamo che

$$P(\Omega) = P(A \cup A^c) = 1 \implies P(A) + P(A^c) = 1 \implies P(A^c) = 1 - P(A)$$

Teorema 3.0.3

Sia Ω uno spazio campione, allora

$$P(\emptyset) = 0$$

Dimostrazione:
Sappiamo che $\Omega^c = \emptyset$, allora per il secondo assioma abbiamo che

$$P(\Omega \cup \Omega^c) = 1 \implies P(\emptyset) = 0$$

Teorema 3.0.4

Dato Ω e $A, b \subset \Omega$, se $A \subset B$ allora

$$P(A) \leq P(B)$$

Dimostrazione:

Sappiamo che $B = A \cup (A^c \cap B)$ e che $(A^c \cap B) \cap A = \emptyset$, allora applicando il terzo assioma otteniamo che

$$P(B) = P(A \cup (A^c \cap B)) = P(A) + P(A^c \cap B) \geq P(A)$$

Teorema 3.0.5

Dato Ω e $A, B \subset \Omega$, e sia $A \cap B \neq \emptyset$ allora

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

Dimostrazione:

Sappiamo che $A \cup B = I \cup II \cup III$ eventi disgiunti a 2 a 2, per cui $P(A \cup B) = P(I) + P(II) + P(III)$

$$P(A) = P(I) + P(II) \quad P(B) = P(II) + P(III)$$

allora

$$P(A) + P(B) = P(I) + 2P(II) + P(III) = P(I) + P(II) + P(III) \Longrightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

3.1 Caso degli esiti equiprobabili

Dato uno spazio campione Ω, contenente un numero finito di esiti equiprobabili, la probabilità di un singolo esito è

$$P(e_k) = \frac{1}{N} \quad k = 1, ..., N$$

Se N è il numero di esiti di Ω, inoltre se $A \subset \Omega$

$$P(A) = \frac{\sharp A}{N}$$

Dove $\sharp A$ è la **cardinalità** di A, ovvero il numero di esiti che stanno in A.
Dimostrazione:
Sappiamo che $\Omega = \{e_1, e_2, ..., e_N\}$ e che $P(e_1) = P(e_2) = ... = P(e_N) = P$.
Applicando il secondo e il terzo principio possiamo dire che:

$$1 = P(\Omega) = P(e_1 \cup e_2 \cup ... \cup e_N) = P(e_1) + P(e_2) + ...P(e_N) = N \cdot P$$

Questo perchè gli eventi sono disgiunti tra loro.

$$1 = N \cdot P \Longrightarrow P = \frac{1}{N} = P(e_k), \quad k = 1, ..., N$$

Dimostriamo ora la seconda parte.

$$\emptyset \neq A \subset \Omega, \quad A = \{e_1, e_2, e_3, ..., e_m\}, \quad m \leq n$$

Gli eventi che compongono A sono disgiunti, quindi:

$$P(A) = P(e_1 \cup e_2 \cup ... \cup e_m) = P(e_1) + P(e_2) + ... + P(e_m) = \frac{m}{N} = \frac{\sharp A}{N}$$

Esempio 3.1.1 (Paradosso dei compleanni)

Date N persone (N≤ 365) nate in anni non bisestili, qual è la probabilità che nessuno compia gli anni lo stesso giorno?
Per risolvere questo esercizio poniamo A='Nessuno ha la stessa data di compleanno', e ci proponiamo di risolverlo in maniera classica, ovvero:

$$P(A) = \frac{n^\circ \, casi\, favorevoli}{n^\circ \, casi\ totali}$$

N° *casi totali* $= 365^N = D^r_{365,N}$
N° *casi favorevoli* $= 365 \cdot 364 \cdot 363 \cdot ... \cdot (365 - N + 1) = D_{365,N}$
Quindi per un N generico la probabilità di A si ottiene in questo modo:

$$P(A) = \frac{365 \cdot 364 \cdot 363 \cdot ... \cdot (365 - N + 1)}{365^N} = \frac{D^r_{365,N}}{D_{365,N}}$$

Qual è la probabilità che su N persone almeno 2 abbiano il compleanno lo stesso giorno?

$$P(A^c) = P(A) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot ... \cdot (365 - N + 1)}{365^N}$$

Anche se può sembrare sorprendente, già con $N = 23$ *la probabilità che almeno 2 persone abbiano il compleanno lo stesso giorno è superiore al 50%. Se* $N = 50$ *la probabilità arriva addirittura a toccare il 97%.*

Esempio 3.1.2 (Chiavi indistinguibili)

Si hanno N chiavi indistinguibili, e solo una apre una porta. Possiamo avere 2 casi, nel primo la chiave sbagliata viene scartata, nel secondo invece si sceglie sempre a caso. ***Qual è la probabilità che occorrano k tentativi per aprire la porta?***

1. ***I caso***
   A_k*='Apertura porta al k-esimo tentativo'.*

   $$P(A_k) = \frac{n^\circ casi\, favorevoli}{n^\circ casi\ totali} = \frac{(N-1)(N-2)(N-3)...(N-k+1)}{N(N-1)(N-2)...(N-k+1)} = \frac{D_{(N-1),k}}{D_N, k} = \frac{1}{N}$$

   Quindi la probabilità non dipende da k!!

2. ***II caso***
   B_k *='La chiave viene ritrovata al k-esimo tentativo con reimmissione'.*

   $$P(B_k) = \frac{n^\circ casi\, favorevoli}{n^\circ casi\ totali} = \frac{(N-1)^{k-1} \cdot 1}{N^k} = \frac{1}{N}\left(\frac{N-1}{N}\right)^{k-1} = \frac{1}{N}\left(1 - \frac{1}{N}\right)^{k-1}$$

Nel primo caso k è finito, nel secondo può essere infinito, ma comunque dipende da N, pertanto se N è molto molto grande i due casi sono equivalenti.

$$\lim_{N\to+\infty} \frac{1}{N}\left(1 - \frac{1}{N}\right)^{k-1} = \frac{1}{N}$$

3.2 Problemi di estrazione

Trattiamo questo tipo di problemi con alcuni esempi.

Esempio 3.2.1 (Senza reimmissione)

In una scatola ci sono 8 palline in totale, 5 sono verdi e 3 sono blu, si estraggono 3 palline senza reimmissione. Qual è la probabilità che le palline estratte siano tutte e 3 verdi?
Abbiamo 2 strategie possibili, in quanto non cambia se estraggo le palline tutte insieme o una alla volta.

1. ***I strategia (tutte insieme)***
$$P('3V') = \frac{C_{5,3}}{C_{8,3}} = \frac{5\cdot4\cdot3}{3!}\cdot\frac{3!}{8\cdot7\cdot6} = \frac{5}{28}$$
2. ***II strategia (una alla volta)***
$$P('3V') = \frac{5}{8}\cdot\frac{4}{7}\cdot\frac{3}{6} = \frac{5}{28}$$

Qual è la probabilità che le palline estratte siano 2 verdi e una blu?

1. ***I strategia (tutte insieme)***
$$P(‘2V+1B') = \frac{C_{5,2}\cdot C_{3,1}}{C_{8,3}} = \frac{15}{28}$$
2. ***II strategia (una alla volta)***
I casi totali sono le $D_{8,3}$, i casi favorevoli in questo caso sono:
$$\textit{Casi favorevoli} = 5\cdot4\cdot3+5\cdot3\cdot4+3\cdot4\cdot5 = 3\cdot(5\cdot4\cdot3) = 180$$
Dove quel 3 rappresenta le permutazioni di 3 elementi con 1 elemento di molteplicità 2. Quindi
$$P(‘2V+1B') = \frac{180}{336} = \frac{15}{28}$$

Esempio 3.2.2 (Con reimmissione)

Teniamo sempre in considerazione l'esempio precedente, con la differenza che ogni volta che una pallina viene estratta poi viene rimessa dentro al contenitore.
$$P(‘3V') = \frac{5^3}{8^3} = \frac{125}{512}$$
$$P(‘2V+1B') = \frac{5^2\cdot3}{8^3}\cdot3 = \frac{75}{512}\cdot3 = \frac{225}{512}$$

3.3 Probabilità condizionata

Definizione 3.3.1 (Probabilità condizionata)

Dati uno spazio campione Ω con $A, B \subset \Omega$, definiamo ***la probabilità di A dato un evento B*** *nel seguente modo.*
$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

Definizione 3.3.2 (Eventi indipendenti)

*Dati uno spazio campione Ω, con $A, B \subset \Omega$, A e B si dicono **indipendenti** se*

$$P(A|B) = P(A) \quad opp. \quad P(A \cap B) = P(A) \cdot P(B) \quad opp. \quad P(B|A) = P(B)$$

Ci chiediamo ora, se $A \cup B = \emptyset$, A e B sono anche indipendenti?
Stando alle definizioni precedenti possiamo dire che

$$0 = P(\emptyset) = P(A \cap B) = P(A) \cdot P(B) \Longrightarrow \text{Almeno uno dei due eventi ha probabilità nulla}$$

Non ha molto senso quindi chiederselo.

Esempio 3.3.1

Dato un mazzo di carte da Poker (52 carte) definiamo 2 eventi.
A='Estrazione di un asso'.
B='Estrazione di una carta di picche'.
Ci chiediamo, A e B sono indipendenti?

$$P(A) = \frac{4}{52}, \quad P(B) = \frac{13}{52}, \quad P(A \cap B) = \frac{1}{52}$$

Applichiamo la definizione di eventi indipendenti.

$$\frac{1}{52} = P(A \cap B) = P(A) \cdot P(B) = \frac{4}{52} \cdot \frac{13}{52} = \frac{1}{52}$$

A e B sono quindi indipendenti, ma esiste l'intersezione.

Esempio 3.3.2

Pierino e Filippo decidono di fare uno scherzo alle persone di un palazzo di 10 appartamenti.
Dei 10 campanelli Pierino ne suona 7 e Filippo ne suona 4.
1-Qual è la probabilità che pierino abbia suonato il campanello del Signor Rossi?
P_r='Pierino suona il campanello del Signor Rossi'.

$$P(P_r) = \frac{7}{10}$$

2-Determinare la probabilità che il campanello sia stato suonato almeno una volta.

$$P(P_r \cup F_r) = P(P_r) + P(F_r) - P(P_r \cap F_r)$$

Notiamo che gli eventi sono indipendenti tra loro, in quanto la scelta dei campanelli di Pierino non influenza in alcun modo la scelta di Filippo, quindi:

$$P(P_r \cup F_r) = P(P_r) + P(F_r) - P(P_r) \cdot P(F_r) = \frac{7}{10} + \frac{4}{10} - \frac{28}{100} = \frac{82}{100}$$

3-Qual è la probabilità che il campanello Rossi non venga suonato?

$$P((P_r \cup F_r)^c) = 1 - \frac{82}{100} = \frac{18}{100}$$

3.4 Formula delle probabilità totali e Teorema di Bayes

Definizione 3.4.1 (Partizione di uno spazio campione)

*Sia Ω uno spazio campione e $H_1, H_2, ..., H_N \subset \Omega$. Se $H_k \cap H_l = \emptyset$ con $k \neq l$ allora $H_1, H_2, ..., H_N$ si dicono **partizione di** Ω, dove si ha che*

$$\bigcup_{k=1}^{N} H_k = \Omega$$

Teorema 3.4.1 (Formula delle probabilità totali)

Dati uno spazio campione Ω, una sua partizione $\{H_1, H_2, ..., H_n\}$ e un evento $A \subset \Omega$ allora:

$$P(A) = \sum_{k=1}^{n} P(A|H_k) \cdot P(H_k)$$

Dimostrazione:
Noi sappiamo che:

$$A = (A \cap H_1) \cup (A \cap H_2) \cup ... \cup (A \cap H_n)$$

$$(A \cap H_k) \cap (A \cap H_l) = \emptyset, \quad con\ k \neq l$$

Per il terzo assioma abbiamo che :

$$P(A) = P(A \cap H_1) + P(A \cap H_2) + ... + P(A \cap H_n) =$$

$$= P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + ... + P(A|H_n) \cdot P(H_n) =$$

$$= \sum_{k=1}^{n} P(A|H_k) \cdot P(H_k)$$

Esempio 3.4.1 (Esami del sangue (parte 1))

Mauro un giorno decide di andare a fare gli esami del sangue per vedere se ha contratto o no un certo tipo di malattia. L'incidenza della malattia è del 0.5%, mentre i falsi positivi e i falsi negativi sono entrambi all' 1%.
Qual è la probabilità che il test risulti positivo?
Possiamo dividere il nostro spazio campione in una partizione di 2 ipotesi, ovvero che un paziente sia malato oppure sia sano, infatti altro non può accadere. H_1='Ipotesi malati'.
H_2='Ipotesi sani'.

$$P(H_1) = 0.995, \quad P(H_2) = 0.005$$

Definiamo T ='Test positivo', per cui:

$$P(T|H_1) = 0.01 = Falsi\ positivi, \quad P(T|H_2) = 0.99$$

Quindi applicando ora la formula delle probabilità totali otteniamo:

$$P(T) = P(T|H_1) \cdot P(H_1) + P(T|H_2) \cdot P(H_2) = 0.01 \cdot 0.995 + 0.99 \cdot 0.005 \simeq 0.015$$

A questo punto sarebbe interessante calcolarsi la probabiltà che dato il risultato positivo del test Mauro sia effettivamente malato. Per rispondere a questa domanda però dobbiamo prima enunciare il Teorema di Bayes.

Teorema 3.4.2 (Teorema di Bayes (o delle probabilità a posteriori))

Dati uno spazio campione Ω, una partizione $\{H_1, H_2, ..., H_N\}$ e un evento $A \subset \Omega$ allora possiamo calcolarci la probabilità di un qualsiasi elemento della partizione dato l'evento A, in questo modo.

$$P(H_k|A) = \frac{P(A|H_k) \cdot P(H_k)}{P(A)}$$

Dove $P(A)$, per la formula delle probabilità totali è esprimibile come:

$$P(A) = \sum_{l=1}^{N} P(A|H_l) \cdot P(H_l)$$

Dimostrazione:

$$P(A \cap H_k) = P(A|H_k) \cdot P(H_k) = P(H_k|A) \cdot P(A)$$

Questo implica che:

$$\frac{P(H_k|A) \cdot P(A)}{P(A)} = \frac{P(A|H_k) \cdot P(H_k)}{P(A)}$$

Quindi possiamo concludere che:

$$P(H_k|A) = \frac{P(A|H_k) \cdot P(H_k)}{P(A)}$$

Esempio 3.4.2 (Esame del sangue (parte 2))

Calcoliamoci a questo punto $P(H_2|T)$ col Teorema di Bayes.

$$P(H_2|T) = \frac{P(T|H_2) \cdot P(H_2)}{P(T)} = \frac{0.99 \cdot 0.005}{0.015} \simeq 33\%$$

3.5 Probabilità ricorsiva

Esempio 3.5.1

Dati N matematici ognuno con un cappello proprio, qual è la probabilità che un matematico, scegliendo a caso un cappello, non prenda il suo?

E_N =‘Su N prof. e cappelli nessuno prende il proprio’.
C_1 =‘Il primo prende il suo cappello’.

$$P(E_N) = P(E_N|C_1) \cdot P(C_1) + P(E_N|C_1^c) \cdot P(C_1^c)$$

Sappiamo che:

$$P(C_1) = \frac{1}{N}, \quad P(C_1^c) = 1 - \frac{1}{N}, \quad P(E_N|C_1) = 0, \quad P(E_N|C_1^c) = P(A \cup B)$$

Dove
A =‘Il secondo matematico non prende il I cappello e gli altri non prendono il loro’.
B =‘Il secondo matematico prende il I cappello e gli altri non prendo il loro’.
Per come sono stati definiti A e B sono sicuramente disgiunti, pertanto:

$$P(A \cup B) = P(A) + P(B), \quad P(A) = P(E_{N-1}), \quad P(B) = \frac{P(E_{N-2})}{N-1}$$

Quindi abbiamo che:

$$P(E_N) = \left(P(E_{N-1}) + \frac{P(E_{N-2})}{N-1}\right) \cdot \left(1 - \frac{1}{N-1}\right)$$

Otteniamo così la formula ricorsiva:

$$P(E_N) - P(E_{N-1}) = \frac{1}{N}(P_{N-2} - P_{N-1})$$

Teorema 3.5.1

Se $E, F \subset \Omega$ *sono indipendenti tra loro allora:*

$$E^c \text{ è indipendente da } F^c$$
$$E \text{ è indipendente da } F^c$$

Dimostrazione:
Sappiamo che possiamo scrivere E in questo modo:

$$E = (E \cap F) \cup (E \cap F^c)$$

e che

$$(E \cap F) \cap (E \cap F^c) = \emptyset$$

$$P(E) = P(E \cap F) + P(E \cap F^c) \Longrightarrow P(E \cap F^c) = P(E) - P(E \cap F)$$

Se E ed F sono indipendenti possiamo dire che:

$$P(E \cap F) = P(E) \cdot P(F) \Longrightarrow P(E) - P(E) \cdot P(F) = P(E)(1 - P(F)) = P(E) \cdot P(F^c)$$

Abbiamo ottenuto quindi che:

$$P(E \cap F^c) = P(E) \cdot P(F^c)$$

Quindi E e F^c *cono indipendenti.*

Definizione 3.5.1 (3 eventi indipendenti)

$A, B, C \subset \Omega$ *sono indipendenti se:*

$$P(A \cap B) = P(A) \cdot P(B), \quad P(B \cap C) = P(B) \cdot P(C)$$

$$P(A \cap C) = P(A) \cdot P(C), \quad P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Esempio 3.5.2 (Problema della rovina del giocatore)

A e B lanciano una moneta. A dà un euro a B se esce Testa, altrimenti se esce croce B dà un euro ad A. Il gioco termina quando uno dei due rimane senza euro. Determinare qual è la probabilità che A vinca se inizialmente possiede k ***euro, mentre B possiede*** $N-k$ ***euro.***
Possiamo cominciare affermando che $0 < k < N$*,ovvero i due casi estremi in cui uno dei due rimane senza soldi. Detto questo definiamo C='Esce croce' e T='Esce testa'.*

$$P(C) = p, \quad P(T) = 1 - p = q$$

A_k *='A vince con k euro iniziali'.*
$P(A_k)$ *=?*

Analizziamo il primo lancio.

$$P(A_k) = P(A_k|T)P(T) + P(A_k|C)P(C)$$

Dove $P(A_k|T) = P(A_{k-1})$*, in quanto A perde un euro se esce testa, e analogamente* $P(A_k|C) = P(A_{k+1})$.
Quindi sostituendo otteniamo

$$P(A_k) = P(A_{k-1})q + P(A_{k+1})p$$

Ora se chiamiamo $P_m = P(A_m)$ *e ricordando che* $p + q = 1$ *possiamo dire che*

$$(p+q)P_k = p(P_{k+1}) + q(P_{k-1}) \Longrightarrow p(P_{k+1} - P_k) = q(P_k - P_{k-1})$$

Dividendo tutto per p otteniamo la seguente formula ricorsiva.

$$P_{k+1} - P_k = \frac{q}{p}(P_k - P_{k-1})$$

Ovviamente da sola questa formula non basta. Infatti ci servono probabilità note da cui poter partire a fare i calcoli. Queste probabilità sono i casi limite, ovvero quando k=0 e k=N.

$$P_0 = 0, \quad P_N = 1$$

Se k=1 abbiamo

$$P_2 - P_1 = \frac{q}{p}P_1$$

Per k=2 sfruttiamo il risultato precedente.

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

Per k=3 procediamo analogamente e otterremo che

$$P_4 - P_3 = \left(\frac{q}{p}\right)^3 P_1$$

E così via finchè non otterremo

$$P_N - P_{N-1} = \left(\frac{q}{p}\right)^{N-1} P_1$$

3.6 Variabili casuali(aleatorie)

Definizione 3.6.1 (Variabile casuale)

Esperimento $\longrightarrow$ *esiti.*
A ciascun esito associo un valore diverso reale, anche a caso. In questa maniera ottengo una variabile X il cui valore ***prima*** *dell'esperimento non è noto. X è chiamata* ***variabile casuale o aleatoria****.*

Esempio 3.6.1

Nel lancio di un dado io posso definire una variabile casuale X in questo modo.

$$X = \{1,2,3,4,5,6\}$$

X in questo caso è più precisamente una variabile casuale discreta.

Esempio 3.6.2

Riprendendo l'esempio 3.1.2 si sarebbe potuto definire una variabile Y in questo modo.

$$Y = \text{'Numero di tentativi effettuati per aprire la porta'}, \quad Y = \{1,2,...,+\infty\}$$

In questo caso Y è una variabile casuale discreta, ma infinita.

Esempio 3.6.3

Prendiamo l'evento Z='Altezza individuo'. Z è una variabile casuale non discreta, in quanto $Z \in [0,3]$ *(metri).*

Definizione 3.6.2 (Variabile casuale discreta)

X è una variabile casuale discreta se assume valori discreti finiti o numerabili.

Definizione 3.6.3

X è una variabile continua se vale la seguente condizione.

$$\exists f : \mathbb{R} \longrightarrow \mathbb{R}^+ : P(X \in B) = \int_B f(s)ds \quad con\ B \subset \mathbb{R}$$

f si dice ***funzione di densità di probabilità***.

Definizione 3.6.4 (Funzione di distribuzione di probabilità)

Sia X una variabile casuale, e sia $a \in \mathbb{R}$.
Si definisce funzione di distribuzione di probabilità la seguente funzione.

$$F : \mathbb{R} \longrightarrow [0,1], \quad F(a) = P(X \leq a)$$

Esempio 3.6.4

Prendiamo in considerazione il solito esempio del lancio del dado.

$$X = \{1,2,3,4,5,6\}, \quad P(X = k) = \frac{1}{6}, \quad k = 1,...,6$$

Calcoliamoci la funzione di distribuzione di probabilità.

$$F(0) = P(X \leq 0) = 0, \quad F(1) = P(X \leq 1) = \frac{1}{6}, \quad \cdots \quad ,F(6) = P(X \leq 6) = 1$$

Grandissima importanza ha anche il grafico di questa funzione, in quanto per le variabili casuali discrete come questa ci permette di localizzare i punti di discontinuità che, come vedremo più avanti, sono sempre presenti per variabili casuali discrete. La funzione è discontinua per $a = 1,2,3,4,5,6$.

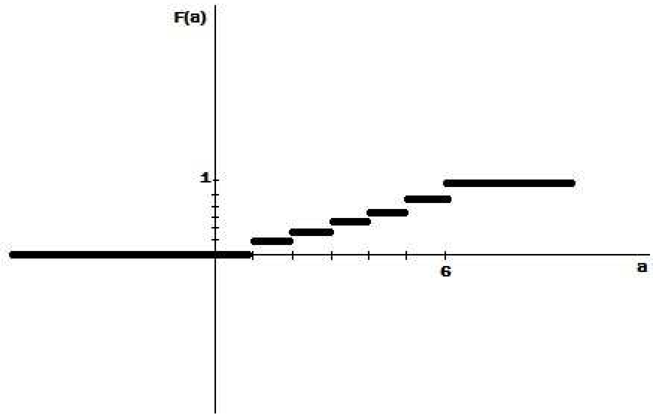


Figura 3.1: Grafico della funzione di distribuzione di probabilità di X nell'esempio 3.6.4

PROPRIETÀ DELLA FUNZIONE DI DISTRIBUZIONE DI PROBABILITÀ

1. $F(-\infty) = P(X \leq -\infty) = 0$;
2. $F(+\infty) = P(X \leq +\infty) = 1$;
3. Se $a, b \in \mathbb{R} : a < b \Longrightarrow F(a) \leq F(b)$, F è quindi una funzione non decrescente;
4. Per X variabile casuale discreta $F(a)$ è discontinua in alcuni punti.
5. Cosa accade per le variabili casuali continue?
   Applichiamo la definizione di variabile casuale continua.
   $$F(a) = P(X \leq a) = \int_{X \leq a} f(s)ds = \int_{-\infty}^{a} f(s)ds \Longrightarrow \text{ Se X è continua anche } f \text{ lo è.}$$
   Deriviamo ora la funzione di distribuzione.
   $$\frac{dF(a)}{da} = \frac{d}{da}\int_{-\infty}^{a} f(s)ds = f(a) \geq 0$$
   NOTA BENE: Non può esistere $f(a)$ per X discreta! Essendo F discontinua in alcuni punti.

Analizziamo ora separatamente i due tipi di variabili casuali.

3.6.1 Variabili casuali discrete

Definizione 3.6.5 (Funzione di massa di probabilità)

Sia X una variabile casuale discreta, si definisce funzione di massa di probabilità di X la seguente funzione.
$$p(a) = P(X = a)$$

Esempio 3.6.5

Prendiamo come esempio il lancio di una moneta.
$$X = \begin{cases} 1 & \text{Testa} \\ 0 & \text{Croce} \end{cases} \qquad P(T) = P(C) = \frac{1}{2}$$
$$p(a) = \begin{cases} \frac{1}{2} & se\ a = 0, 1 \\ 0 & Altrimenti \end{cases}$$

PROPRIETÀ DELLA FUNZIONE DI MASSA DI PROBABILITÀ

1. $0 \leq p(a) \leq 1$;
2. $$\sum_{k=1}^{N} p(x_k) = 1$$
3. $$F(a) = \sum_{x_k \leq a} p(x_k)$$

3.6.2 Variabili casuali continue

Ci chiediamo subito: esiste la funzione di massa di probabilità?

$$p(a) = P(X = a) = \int_{X=a} f(s)ds = \int_{a}^{a} f(s)ds = 0 \quad \forall a \in \mathbb{R}$$

Su infiniti valori la probabilità di un singolo punto è praticamente zero (ma non impossibile).

Esempio 3.6.6

X variabile casuale continua con

$$f(a) = \begin{cases} c(4a-2a^2) & 0 < a < 2 \\ 0 & \textit{altrove} \end{cases}$$

***Quanto vale** c?*
Sappiamo che $f(a) \geq 0$, quindi

$$c(4a-2a^2) \geq 0 \Longrightarrow c \geq 0 \quad \textit{per } 0 < a < 2$$

Infine sappiamo che

$$1 = F(-\infty) = \int_{-\infty}^{+\infty} f(s)ds =$$

$$= \int_{-\infty}^{0} 0ds + \int_{0}^{2} c(4s-2s^2)ds + \int_{2}^{+\infty} 0ds = c\left[2s^2 - \frac{2}{3}s^3\right]_0^2 =$$

$$= c\left(8 - \frac{16}{3}\right) = \frac{8}{3}c \Longrightarrow c = \frac{3}{8} \geq 0$$

***Quanto vale** $F(a)$?*
Se $a \leq 0$:

$$F(a) = \int_{-\infty}^{a} f(s)ds = \int_{-\infty}^{a} 0ds = 0$$

Se $0 < a < 2$:

$$F(a) = \int_{-\infty}^{a} f(s)ds = \int_{-\infty}^{0} 0ds + \int_{0}^{a} \frac{3}{8}(4s-2s^2)ds = \frac{3}{8}\left(2a^2 - \frac{2}{3}a^3\right)$$

Se $a \geq 2$:

$$F(a) = \int_{-\infty}^{a} f(s)ds = \int_{-\infty}^{0} 0ds + \int_{0}^{2} \frac{3}{8}(4s-2s^2)ds + \int_{2}^{a} 0ds = 1 \quad \textit{per la relazione di prima}$$

Riassumendo:

$$F(a) = \begin{cases} 0 & a \leq 0 \\ \frac{3}{8}\left(2a^2 - \frac{2}{3}a^3\right) & 0 < a < 2 \\ 1 & a \geq 2 \end{cases}$$

Verifichiamo la continuità in 0 *e in* 2.

$$F(0) = 0, \quad F(2) = 1 \Longrightarrow F \text{ continua}$$

3.7 Coppia di variabili casuali

Definizione 3.7.1 (Coppia di variabili casuali)

Si definisce coppia di variabili casuali (discrete o continue) due variabili casuali associate allo stesso esperimento.

Analizziamo anche qui separatamente il caso in cui la coppia di variabili sia formata da variabili casuali discrete e il caso in cui sia formata da una coppia di variabili casuali continue.

3.7.1 Coppie di variabili casuali discrete

Prendiamo due variabili casuali discrete (coppie ordinate) $X = \{x_1, x_2, ..., x_n\}$ e $Y = \{y_1, y_2, ..., y_m\}$, con n non necessariamente uguale a m. Per esse possiamo definire la **funzione di massa di probabilità congiunta**, in questo modo:

$$p(a,b) = P(X = a, Y = b)$$

Bisogna notare che la scrittura con la virgola è equivalente all'intersezione, ovvero i due eventi devono accadere tutti e due insieme.

PROPRIETÀ:

$$p : \mathbb{R}^2 \longrightarrow [0,1]$$

Ovviamente come nel caso di una singola variabile la somma di tutte le funzioni di massa deve fare uno.

$$\sum_{k=1}^{n}\sum_{l=1}^{m} p(x_k, y_l) = 1$$

Possiamo definire infine le funzioni marginali di massa di probabilità.

Definizione 3.7.2 (Funzioni marginali di massa di probabilità)

Definiamo funzione marginale di massa di probabilità di X la seguente funzione.

$$p_X(a) = P(X = a) = P(X = a, Y = y_1) + P(X = a, Y = y_2) + ... + P(X = a, Y = y_m) =$$

$$= \sum_{l=1}^{m} p(a, y_l)$$

Analogamente otteniamo anche la funzione marginale della Y.

$$\sum_{k=1}^{n} p(x_k, b)$$

Esempio 3.7.1

Si scelgono a caso 3 batterie tra le 12 presenti in una scatola, che ne contiene 3 nuove, 4 usate e 5 difettose, tutte e 12 indistinguibili.
X='Numero batterie nuove scelte'.
Y='Numero di batterie usate scelte'.
Determinare la funzione congiunta e le funzioni marginali di X e Y.

Cominciamo col calcolare tutte le probabilità in questo modo.

$$P(0,0) = P(X = 0, Y = 0) = \frac{C_{5,3}}{C_{12,3}} = \frac{10}{220}$$

$$P(1,1) = P(X = 1, Y = 1) = \frac{C_{3,1} \cdot C_{4,1} \cdot C_{5,1}}{C_{12,3}} = \frac{60}{220}$$

E così via. Mettiamo tutti i risultati in una tabella come quella in basso. Notiamo come l'ultima colonna e l'ultima riga siano semplicemente le funzioni marginali della X e della Y, mentre la parte interna della tabella rappresenta i valori della funzioni di massa di probabilità congiunta.

X|Y	0	1	2	3	p_X
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
p_Y	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Tabella 3.1: Tabella dell'esempio 3.7.1

Analogamente alla definizione di funzione di massa di probabilità congiunta, possiamo definire la **funzione di distribuzione di probabilità congiunta**.

$$F(a,b) = P(X \leq a, Y \leq b)$$

PROPRIETÀ:

$$F : \mathbb{R}^2 \longrightarrow [0,1]$$

- $$F(-\infty, -\infty) = P(X \leq -\infty, Y \leq -\infty) = 0$$
- $$F(+\infty, +\infty) = P(X \leq +\infty, Y \leq +\infty) = 1$$

Nel caso discreto F ha delle discontinuità.

3.7.2 Coppie di variabili casuali continue

Prendiamo due variabili casuali X e Y. (X,Y) si dicono continue se esiste una funzione di questo tipo.

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}_+ \mid P((x,y) \in B) = \iint_B f(s,t)dsdt$$

Se $B \subset \mathbb{R}^2$.
f si dice **funzione congiunta di densità di probabilità.**
Definiamo ora la **funzione di distribuzione congiunta**.

$$F(a,b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{a}\left(\int_{-\infty}^{b} f(s,t)dt\right)ds$$

Per il teorema fondamentale del calcolo integrale si ha che

$$f(a,b) = \frac{\partial^2}{\partial a \partial b}F(a,b)$$

Definizione 3.7.3 (Funzioni marginali di distribuzione di probabilità)

Definiamo funzione marginale di distribuzione di probabilità di X la seguente funzione.

$$F_X(a) = F(a,+\infty)$$

Analogamente per Y.

$$F_Y(b) = F(+\infty,b)$$

Definizione 3.7.4 (Funzioni marginali di densità di probabilità)

Definiamo funzione marginale di densità di probabilità di X la seguente funzione.

$$f_X(a) = \frac{dF_X(a)}{da} = \frac{dF(a,+\infty)}{da} = \frac{d}{da} = \int_{-\infty}^{a}\left(\int_{-\infty}^{+\infty} f(s,t)dt\right)ds =$$

$$= \int_{-\infty}^{+\infty} f(a,t)dt$$

Lo stesso ragionamento vale per Y.

$$f_Y(b) = \int_{-\infty}^{+\infty} f(s,b)ds$$

3.7.3 Coppia di variabili casuali indipendenti

Prendiamo una coppia (X,Y) di variabili casuali, esse si dicono indipendenti quando

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

con $A, B \subset \mathbb{R}$.
Si dimostra una definizione equivalente, sicuramente più pratica, ovvero: 2 variabili casuali sono indipendenti quando

$$F(a,b) = F_X(a) \cdot F_Y(b) \quad \forall a,b \in \mathbb{R}$$

Ovvero equivalentemente abbiamo che vale per le variabili casuali discrete indipendenti la seguente formula.

$$p(a,b) = p_X(a) \cdot p_Y(b) \quad \forall a,b \in \mathbb{R}$$

Mentre per le variabili casuali continue indipendenti vale questa.

$$f(a,b) = f_X(a) \cdot f_Y(b) \quad \forall a,b \in \mathbb{R}$$

In tutto abbiamo 3 relazioni equivalenti alla prima condizione.

3.8 Valor medio o valore atteso o speranza matematica o media

Data una variabile casuale X il suo valor medio $\mathrm{E}[X]$, se esiste, si definisce per variabili casuali discrete:

$$\mathrm{E}[X] = \sum_{k=1}^{n} x_k p(x_k) \quad \text{con } X = \{x_1, ..., x_n\}$$

Per variabili casuali continue:

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

Ovviamente una condizione necessaria per l'esistenza del valor medio è che la serie (se $N = +\infty$) o l'integrale convergano.

Esempio 3.8.1

Calcoliamo il valor medio per un lancio di un dado non truccato.

$$X = \{1,2,3,4,5,6\}, \quad P(X = k) = \frac{1}{6} \quad \textit{con } k = 1,2,...,6$$

$$E[X] = \sum_{k=1}^{6} x_k p(x_k) = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = 3.5$$

Il valor medio oscilla comunque tra i valori che può assumere X.

Esempio 3.8.2 (Valor medio variabile casuale di Bernoulli)

Sia X una variabile casuale di Bernoulli, definita in questo modo:

$$X = \begin{cases} 1 & \textit{se successo} \\ 0 & \textit{se fallimento} \end{cases} \qquad P(1) = p, \quad P(0) = 1 - p = q$$

Calcoliamo il valor medio.

$$E[X] = 0 \cdot q + 1 \cdot p = p$$

Data una variabile casuale X e definita $Y = g(X)$ il valor medio di Y si definisce, per X discreta in questo modo:

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \sum_{k=1}^{n} g(x_k) p(x_k)$$

Per X continua in quest'altro:

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

Tenendo sempre conto della convergenza o meno della serie o dell'integrale.
PROPRIETÀ:

1. Se $g(X) = aX + b$ con $a, b \in \mathbb{R} \Longrightarrow \mathrm{E}[aX + b] = a\mathrm{E}[X] + b$, quindi il valor medio è lineare.
   Dimostrazione:
   Dimostreremo questa proprietà nel caso continuo, la dimostrazione per il caso discreto è analoga.

$$\mathrm{E}[aX + b] = \int_{-\infty}^{+\infty} (ax + b) f(x) dx =$$

Supponiamo che l'integrale converga (se non convergesse non esisterebbe il valor medio).

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

Come conseguenza si ha che se $a = 0 \Longrightarrow \mathrm{E}[b] = b$.

2. Se $g(X) = X^n$ con $n \geq 0$ allora

$$\mathrm{E}[X^n] = \sum_{k=1}^{n} x_k^n p(x_k) \quad \text{per } X \text{ discreta}$$

$$\mathrm{E}[X^n] = \int_{-\infty}^{+\infty} x^n f(x)dx \quad \text{per } X \text{ continua}$$

$\mathrm{E}[X^n]$ viene chiamato **momento di ordine** n di X.

3. Sia (X,Y) una coppia di variabili casuali e $Z = g(X,Y)$, allora:

$$E[Z] = \mathrm{E}[g(X,Y)] = \sum_{k=1}^{n}\sum_{l=1}^{m} g(x_k,y_l)p(x_k,y_l) \quad \text{per } (X,Y) \text{ discrete}$$

$$E[Z] = \mathrm{E}[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy \quad \text{per } (X,Y) \text{ continue}$$

Poniamo $Z = X+Y$, allora $\mathrm{E}[Z] = \mathrm{E}[X+Y] = \mathrm{E}[X]+\mathrm{E}[Y]$.
Dimostrazione:
Dimostreremo questa proprietà nel caso continuo, la dimostrazione per il caso discreto è analoga.

$$\mathrm{E}[X+Y] = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}(x+y)f(x,y)dx\right)dy =$$

$$= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} xf(x,y)dy\right)dx + \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} yf(x,y)dx\right)dy =$$

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

$$= \int_{-\infty}^{+\infty} xf_X(x)dx + \int_{-\infty}^{+\infty} yf_Y(y)dy = \mathrm{E}[X]+\mathrm{E}[Y]$$

Generalizzando otteniamo la formula seguente.

$$\mathrm{E}\left[\sum_{j=1}^{m} X_j\right] = \sum_{j=1}^{m}\mathrm{E}[X_j]$$

Esempio 3.8.3

Una segretaria possiede N buste e N lettere da riordinare, solo che a causa della sua sbadataggine sia le buste sia le lettere cadono per terra sparse.
Per fare in fretta la segretaria mette a caso le lettere nelle buste, senza controllare che in una busta ci sia già una lettera. Determinare il numero medio di lettere che capitano nella busta corretta.

Y='Numero di lettere che finiscono nella busta giusta'.

$$X_k = \begin{cases} 1 & \text{se la k-esima lettera finisce nella busta giusta} \\ 0 & \text{altrimenti} \end{cases}$$

Calcoliamo le probabilità.

$$P(X_k = 1) = \frac{1}{N}, \quad P(X_k = 0) = \frac{N-1}{N}, \quad Y = \sum_{k=1}^{N} X_k$$

Troviamo il valor medio sfruttando il fatto che X_k è una variabile casuale binomiale.

$$E[X_k] = \frac{1}{N} \Longrightarrow E[Y] = E\left[\sum_{k=1}^{N} X_k\right] = \sum_{k=1}^{N} E[X_k] = \sum_{k=1}^{N} \frac{1}{N} = 1$$

3.9 Varianza e scarto quadratico medio

Data una variabile casuale X si definisce **varianza** di X la seguente espressione.

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2] = \sum_{k=1}^{n} (x_k - \mathrm{E}[X])^2 p(x_k) \quad \text{per } X \text{ discreta}$$

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathrm{E}[X])^2 f(x) dx \quad \text{per } X \text{ continua}$$

PROPRIETÀ:

1. $\mathrm{Var}(X) \geq 0$
2. Dalla varianza si definisce lo **scarto quadratico medio** o **deviazione standard** in questo modo.
$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$
Questo risultato permette di ottenere la misura non al quadrato.
3. $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$
Dimostrazione:
$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2 - 2\mathrm{E}[X]X + \mathrm{E}[X]^2] = \mathrm{E}[X^2] + \mathrm{E}[\mathrm{E}[X]^2] - \mathrm{E}[2\mathrm{E}[X]X] =$$
$$= \mathrm{E}[X^2] + \mathrm{E}[X]^2 - 2\mathrm{E}[X]^2 = \mathrm{E}[X^2] - E[X]^2$$
4. $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$
Dimostrazione:
$$\mathrm{Var}(aX + b) = \mathrm{E}[(aX + b - \mathrm{E}(aX + b))^2] = \mathrm{E}[(aX + b - a\mathrm{E}[X] - b)^2] =$$
$$= \mathrm{E}[(aX - a\mathrm{E}[X]^2)] = \mathrm{E}[a^2(X - \mathrm{E}[X])^2] = a^2\mathrm{E}[(X - \mathrm{E}[X])^2] = a^2\mathrm{Var}(X)$$

Esempio 3.9.1 (Varianza variabile casuale di Bernoulli)

Sappiamo che per X variabile casuale di Bernoulli $E[X] = p$.
Calcoliamoci il momento di ordine 2*, o più in generale il momento di ordine n.*

$$E[X^n] = 0^n \cdot p + 1^n \cdot p = p$$

A questo punto possiamo calcolarci la varianza grazie ala proprietà 3.

$$Var(X) = E[X^2] - E[X]^2 = p - p^2 = p(1-p) = pq$$

Per quale p la varianza è massima? Applichiamo il teorema di Fermat.

$$\frac{d}{dp}(p - p^2) = 1 - 2p = 0 \Longrightarrow p = \frac{1}{2}$$

3.10 Covarianza

Data una coppia di variabili casuali (X,Y) si definisce **covarianza** di (X,Y) la seguente espressione.

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

Più precisamente definendo μ_x e μ_y i valori medi rispettivamente di X e Y possiamo scrivere la seguente espressione.

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_x)(Y-\mu_y)] = \sum_{k=1}^{n}\sum_{l=1}^{n}(x_k-\mu_x)(y_l-\mu_y)p(x_k,y_l) \quad \text{per } X \text{ e } Y \text{ discrete}$$

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_x)(Y-\mu_y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_x)(y-\mu_y)f(x,y)dxdy \quad \text{per } X \text{ e } Y \text{ continue}$$

PROPRIETÀ:

1. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
2. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
3. $\mathrm{Cov}(aX,Y) = \mathrm{Cov}(X,aY) = a\mathrm{Cov}(X,Y)$ con $a \in \mathbb{R}$
4. $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$
   Dimostrazione:

   $$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_x)(Y-\mu_y)] = \mathrm{E}[XY-\mu_xY-\mu_yX+\mu_x\mu_y] = \mathrm{E}[XY]-\mu_X\mathrm{E}[Y]-\mu_y\mathrm{E}[X]+\mu_y\mu_x =$$

   $$= \mathrm{E}[XY] - 2\mu_y\mu_x + \mu_y\mu_y = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

   Esempio 3.10.1 (Covarianza di due variabili casuali di Bernoulli)

   Date due variabili casuali di Bernoulli X e Y calcoliamo la covarianza. Riempiamo la nostra solita tabella come in figura 3.2. Calcoliamoci il valor medio di X e Y.

   $$E[X] = 0 \cdot (p_3 + p_2) + 1 \cdot p_1 = p_1, \quad E[Y] = p_2$$

   Ora calcoliamo $E[XY]$ *in modo da utilizzare la proprietà 4.*

   $$E[XY] = 0 \cdot 0 \cdot p_3 + 1 \cdot 0 \cdot p_1 + 0 \cdot 1 \cdot p_2 + 1 \cdot 1 \cdot 0 = 0$$

   Arriviamo così al risultato finale.

   $$Cov(X,Y) = 0 - p_1p_2 = -p_1p_2$$

Y|X	0	1	
0	p_3	p_1	p_3+p_1
1	p_2	0	p_2
	p_3+p_2	p_1	

Tabella 3.2: Tabella esempio 3.10.1

5. $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)$
   Dimostrazione:

$$\mathrm{Var}(X+Y) = \mathrm{E}\left[(X+Y-\mathrm{E}[X+Y])^2\right] = \mathrm{E}\left[(X+Y-(\mu_x+\mu_y))^2\right] = \mathrm{E}\left[\left((X-\mu_x)+(Y-\mu_y)\right)^2\right] =$$

$$= \mathrm{E}\left[(X-\mu_x)^2+(Y-\mu_y)^2+2(X-\mu_x)(Y-\mu_y)\right] = \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)$$

Definizione 3.10.1 (Variabili casuali scorrelate)

*Data una coppia di variabili casuali (X,Y) esse si dicono **scorrelate** se accade questo:*

$$Cov(X,Y) = 0$$

OSSERVAZIONE:
Se 2 variabili casuali sono indipendenti allora hanno covarianza nulla, ma non sempre il viceversa.
Se 2 variabili casuali sono indipendenti allora sono anche scorrelate, e quindi possiamo dire che per quanto riguarda la varianza si ha questo risultato:

$$\mathrm{Var}(X,Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)$$

Per quanto riguarda il valor medio:

$$\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$$

6.
$$\mathrm{Cov}\left(\sum_{j=1}^{n} X_j, \sum_{i=1}^{m} Y_i\right) = \sum_{j=1}^{n}\sum_{i=1}^{m}\mathrm{Cov}(X_j,Y_i)$$

7.
$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n}\mathrm{Var}(X_k) + \sum_{k=1,j=1,k\neq j}^{n}\mathrm{Cov}(X_k,X_j)$$

Definizione 3.10.2 (Coefficiente di correlazione)

*Data una coppia di variabili (X,Y) si definisce **coefficiente di correlazione** la seguente espressione.*

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Si può dimostrare che

$$Corr(X,Y) \in [-1,1]$$

Esempio 3.10.2

Sia Y una variabile casuale definita in questo modo: $Y = aX + b$ con $a, b \in \mathbb{R}$. Calcolarne la covarianza e il coefficiente di correlazione.

$$Cov(X, aX + b) = Cov(X, aX) + Cov(X, b) = aCov(X, X) + E[(X - \mu_x)(b - E[b])] = aVar(X) + 0$$

Calcoliamo il coefficiente di correlazione.

$$Var(aX + b) = a^2 Var(X) \Longrightarrow Corr(X, Y) = \frac{aVar(X)}{\sqrt{a^2 Var^2(X)}} = \frac{a}{|a|} = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

Se $a > 0$ il coefficiente di correlazione è massimo, e X e Y sono ***direttamente correlate****, se $a < 0$ X e Y sono* ***inversamente correlate****, ovvero al crescere di una l'altra diminuisce.*

Il "corollario" dell'esempio 3.10.2 vale sempre, infatti la covarianza e in particolare il coefficiente di correlazione danno proprio indicazioni su come X e Y sono legate tra di loro.

3.11 Funzione generatrice dei momenti

Data una variabile casuale X viene definita **funzione generatrice dei momenti** la seguente espressione.

$$\Phi(t) = \mathrm{E}[e^{tx}]$$

Più nel dettaglio diciamo che:

$$\Phi(t) = \mathrm{E}[e^{tx}] = \sum_{k=1}^{n} e^{tx_k} p(x_k) \quad \text{per } X \text{ discreta}$$

$$\Phi(t) = \mathrm{E}[e^{tx}] = \int_{-\infty}^{+\infty} f(x) e^{tx} dx \quad \text{per } X \text{ continua}$$

PROPRIETÀ:

1. $\Phi(0) = \mathrm{E}[1] = 1$
2. $$\frac{d}{dt}\Phi(t)\bigg|_{t=0} = \Phi'(0) = \mathrm{E}[xe^{tx}]\bigg|_{t=0} = \mathrm{E}[X]$$
3. $$\frac{d^n}{dt^n}\Phi(t)\bigg|_{t=0} = \Phi^{(n)}(0) = \mathrm{E}[X^n e^{tx}]\bigg|_{t=0} = \mathrm{E}[X^n]$$
4. Se due variabili casuali hanno la stessa funzione generatrice dei momenti esse hanno anche la stessa funzione di distribuzione di probabilità, e ,se discrete la stessa funzione di massa, se continue la stessa funzione di densità.
5. Date due variabili casuali X e Y aventi funzioni generatrici rispettivamente $\Phi_X(t)$ e $\Phi_Y(t)$ allora se X e Y sono indipendenti possiamo affermare che

$$\Phi_{X+Y}(t) = \Phi_X(t) \cdot \Phi_Y(t)$$

Esempio 3.11.1 (Funzione generatrici della variabile di Bernoulli)

Sia X una variabile casuale di Bernoulli, calcoliamone la funzione generatrice dei momenti.

$$\Phi(t) = E[e^{tx}] = e^{t\cdot 0}q + e^{t\cdot 1}p = q + e^t p$$

$$\Phi^{(n)}(0) = \frac{d^n}{dt^n}(q + e^t p)\Big|_{t=0} = p$$

Come verifica della proprietà 1 possiamo far vedere questo risultato:

$$\Phi(0) = q + e^0 p = q + p = 1$$

Capitolo 4

Legge dei grandi numeri, disuguaglianza di Markov e disuguaglianza di Chebychev

4.1 Disuguaglianza di Markov

Data una variabile casuale X tale che $X \geq 0$, allora $\forall a > 0$ si ha il seguente risultato.

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Questo risultato prende il nome di **disuguaglianza di Markov**.
Dimostrazione:
Dimostreremo la disuguaglianza nel caso continuo, ma vale anche per il caso discreto.

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{+\infty} xf(x)dx = \int_{0}^{a} xf(x)dx + \int_{a}^{+\infty} xf(x)dx \geq$$

$$\geq \int_{a}^{+\infty} xf(x)dx$$

Imponiamo la condizione $X \geq a$.

$$\geq \int_{a}^{+\infty} af(x)dx = a\int_{a}^{+\infty} f(x)dx = aP(x \geq a)$$

4.2 Disuguaglianza di Chebychev

Data una variabile casuale X con $\mathrm{E}[X] = \mu$ e $\mathrm{Var}(X) = \sigma^2$ si ha il seguente risultato.

$$\forall \varepsilon > 0 \quad P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Questo risultato prende il nome di **disuguaglianza di Chebychev**
Dimostrazione:

$$P(|X - \mu| > \varepsilon) = P(|X - \mu|^2 \geq \varepsilon^2) \leq$$

Applichiamo Markov.

$$\leq \frac{E\left[(X - \mu)^2\right]}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

4.3 Media campionaria o aritmetica

Date $X_1, X_2, ..., X_n$ variabili casuali identicamente distribuite e indipendenti definiamo la **media campionaria** nel modo seguente.

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Chiamati $\mathrm{E}[X-k] = \mu$ e $\mathrm{Var}(X_k) = \sigma^2$, calcoliamoci valor medio e varianza della media.

$$\mathrm{E}[\overline{X}] = \mathrm{E}\left[\frac{X_1 + X_2 + ... + X_n}{n}\right] = \frac{1}{n}(\mathrm{E}[X_1] + ... + \mathrm{E}[X_n]) = \frac{1}{n} n\mu = \mu$$

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X_1 + X_2 + ... + X_n) =$$

$$= \frac{1}{n^2}(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + ... + \mathrm{Var}(X_n)) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Più è grande n più la varianza è piccola.

4.4 Legge dei grandi numeri

Data una successione di variabili casuali X_k indipendenti e identicamente distribuite, $\forall \varepsilon > 0$ vale che:

$$P\left(\left|\sum_{k=1}^{n} \frac{X_k}{n} - \mu\right| \geq \varepsilon\right) \longrightarrow 0 \quad \text{per } n \to +\infty, \quad \text{se } \mathrm{E}[X_k] = \mu$$

Dimostrazione:
Sappiamo che

$$\mathrm{E}\left[\sum_{k=1}^{n} \frac{X_k}{n}\right] = \mu$$

Per Chebychev diciamo che

$$P\left(\left|\sum_{k=1}^{n} \frac{X_k}{n} - \mu\right| \geq \varepsilon\right) \leq \mathrm{Var}\left(\sum_{k=1}^{n} \frac{X_k}{n}\right)\frac{1}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0 \quad \text{per } n \to +\infty$$

La media quindi converge in probabilità al suo valor medio per valori di n molto grandi.
Questa legge collega la probabilità all'inferenza statistica.

4.4.1 Corollario di Bernoulli

Dati n esperimenti identici e indipendenti, sia n_σ il numero di volte in cui si è verificato l'evento A, la definizione frequentista di probabilità dice che

$$P(A) \approx \frac{n_a}{n} \qquad \Longrightarrow \text{frequenza relativa di A}$$

Associamo ad ogni esperimento una variabile casuale di Bernoulli.

$$X_k = \begin{cases} 1 & \text{se A si verifica nel k-esimo esperimento} \\ 0 & \text{altrimenti} \end{cases}$$

Chiamiamo p la probabilià che si verifichi A, pertanto possiamo concludere che, essendo X_k una variabile casuale di Bernoulli con probabilità p di successo,

$$\mathrm{E}[X_k] = p$$

Ancora sfruttando il fatto che X_k è una Bernoulliana, possiamo dire che $n_a = X_1 + X_2 + ... + X_n$, infatti anche se A non si verifica i valori che si aggiungono alla somma sono degli zeri, che quindi non alterano nulla. detto questo si riconosce subito che

$$\frac{n_a}{n} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Ovvero la media aritmetica. Abbiamo quindi trovato il seguente risultato:

$$\mathrm{E}\left[\frac{n_a}{n}\right] = \mathrm{E}[X_k] = p$$

Ora applichiamo la legge dei grandi numeri.

$$P\left(\left|\frac{n_a}{n} - P(A)\right| \geq \varepsilon\right) \longrightarrow 0 \quad \text{per } n \to +\infty$$

Il che mi garantisce che la definizione frequentista è corretta. Possiamo concludere tutto il discorso enunciando appunto il corollario di Bernoulli:
La frequenza relativa di un evento A in n esperimenti indipendenti e identici converge in probabilità alla P(A).

Capitolo 5

Modelli di variabili casuali discrete

5.1 Variabile casuale di Bernoulli

Sia X una variabile casuale definita in questo modo:

$$X = \begin{cases} 1 & \text{se successo} \\ 0 & \text{altrimenti} \end{cases} \qquad \text{Notazione sintetica: } X \sim Be(p), \quad \text{dove } p = P(X = 1)$$

Abbiamo già visto in precedenza i seguenti risultati che caratterizzano il modello di variabile.

$$\mathrm{E}[X] = p$$

$$\mathrm{Var}(X) = pq$$

$$\Phi(t) = q + e^t p$$

5.2 Variabile casuale binomiale

Si ripete **indipendentemente** e **in maniera identica** un esperimento n volte. Definiamo l'evento X in questo modo:

$$X = \text{'Numero di successi su } n \text{ prove'}, \quad X \in \{0, 1, 2, 3, ..., n\}$$

Notazione sintetica: $X \sim B(n, p)$ dove p è la probabilità di successo di una singola prova

Cerchiamo di calcolarci la funzione di massa di probabilità. Partiamo coi due casi limite, ovvero $p(0)$ e $p(n)$, è abbastanza intuitivo che avranno queste espressioni:

$$p(0) = (1 - p)^n$$

$$p(n) = p^n$$

Cerchiamo ora di arrivare al caso generale, cercando di capire come è fatta $p(1)$.

$$p(1) = p((S_1 \cap F_2 \cap ... \cap F_n) \cup (F_1 \cap S_2 \cap F_3 \cap F_n) \cup (F_1 \cap F_2 \cap F_3 \cap ... \cap S_n)) =$$

$$= p(1 - p)^{n-1} + (1 - p)p(1 - p)^{n-2} + ... + (1 - p)^{n-1} p = np(1 - p)^n$$

Dove S_l e F_l sono rispettivamente i successi o i fallimenti nel l-esimo esperimento. Generalizzando a k otteniamo:

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

Verifichiamo che quest'ultima relazione sia effettivamente un funzione di massa di probabilità.

- $p(k) \geq 0$, sicuramente perchè i 3 fattori sono tutti e tre sempre maggiori o uguali a zero.
- Applicando la formula di Newton dimostriamo che la somma delle funzioni di massa è uguale a 1.

$$\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+q)^n = 1$$

Ora cerchiamo il valor medio.

$$\mathrm{E}[X] = \sum_{k=0}^{n} kp(k) = \sum_{k=1}^{n} k \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

Dimostrare per questa via quanto vale il valor medio risulta essere abbastanza difficile. Procediamo in questo modo.

$$X \sim B(n,p), \quad X = Y_1 + Y_2 + Y_3 + ... + Y_n$$

L'idea è quella di trattare le Y_k come variabili di Bernoulli di parametro p indipendenti tra loro.

$$Y_k \sim Be(p) \Longrightarrow \mathrm{E}[X] = \mathrm{E}\left[\sum_{k=1}^{n} Y_k\right] = \sum_{k=1}^{n} \mathrm{E}[Y_k] = \sum_{k=1}^{n} p = np$$

Calcoliamoci ora la varianza.

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{k=1}^{n} Y_k\right) = \sum_{k=1}^{n} \mathrm{Var}(Y_k) = npq$$

5.3 Variabile casuale geometrica

Si ripete un esperimento **in maniera identica** e **indipendentemente finchè si ottiene un successo.** Definiamo l'evento X in questo modo:

$$X = \text{‘Numero tentativi effettuati per avere un successo’}, \quad X \in \{1, ..., +\infty\}$$

Notazione sintetica: $X \sim G(p)$, dove p è la probabilità di successo di un esperimento

La funzione di massa di probabilità è la seguente.

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

In quanto per i primi $(k-1)$ tentativi si fallisce e al k-esimo tentativo si ha il successo. La funzione di massa è tale, in quanto è maggiore di zero poichè i fattori sono sempre positivi e poichè la somma di tutte le $p(k)$ è 1. Infatti

$$\sum_{k=1}^{+\infty} p(k) = \sum_{k=1}^{+\infty} (1-p)^{k-1} p =$$

Ponendo $l = k-1$

$$= p \sum_{l=0}^{+\infty} (1-p)^l =$$

Ci siamo ricondotti alla serie geometrica di ragione $(1-p)$, quindi

$$= \frac{p}{1-(1-p)} = \frac{p}{p} = 1$$

Cerchiamo ora il valor medio.

$$\mathrm{E}[X] = \sum_{k=1}^{+\infty} k(1-p)^{k-1} p = p \sum_{k=0}^{+\infty} k(1-p)^{k-1} =$$

Posso aggiungere $k = 0$ perchè tanto non altera la somma.

$$= p \sum_{k=0}^{+\infty} \frac{dq^k}{dq} = p \frac{d}{dq} \sum_{k=0}^{+\infty} dq^k = p \frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{p}{(1-q^2)} = \frac{p}{p^2} = \frac{1}{p}$$

Calcoliamoci il momento di ordine 2 per poi calcolarci la varianza.

$$\mathrm{E}[X^2] = \sum_{k=1}^{+\infty} k^2(1-p)^{k-1} p = p \sum_{k=0}^{+\infty} k^2(1-p)^{k-1} = p \sum_{k=0}^{+\infty} (k^2 + k - k) q^{k-1} =$$

$$= p\left[\sum_{k=0}^{+\infty} (k^2 - k) q^{k-1} + \sum_{k=0}^{+\infty} k q^{k-1}\right] = p\left[q \sum_{k=0}^{+\infty} k(k-1) q^{k-2} + \frac{d}{dq}\left(\frac{1}{1-q}\right)\right] =$$

$$= p\left[q \frac{d^2}{dq}\left(\frac{1}{1-q}\right) + \frac{1}{(1-q)^2}\right] = p\left[\frac{2q}{(1-q)^3} + \frac{1}{(1-q)^2}\right] = p\left(\frac{2q+1-q}{(1-q)^3}\right) =$$

$$= p\left(\frac{q+1}{(1-q)^3}\right) = \frac{2p-p}{p^3} = \frac{2-p}{p^2}$$

Calcoliamoci la varianza.

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

5.4 Variabile casuale di Poisson

Definiamo una variabile casuale X in questo modo.

$$X = \{0, 1, 2, ..., +\infty\} \qquad \text{Notazione sintetica: } X \sim Po(\lambda)$$

$$P(X = k) = p(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{per } \lambda > 0$$

Verifichiamo che sia effettivamente una funzione di massa di probabilità, sfruttando il fatto che la serie

$$\sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} = e^{\alpha}$$

Quindi dimostriamo che la somma delle funzioni di massa è uguale a 1.

$$\sum_{k=0}^{+\infty} p(k) = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

Cerchiamo il valor medio.

$$\mathrm{E}[X] = \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} =$$

Poniamo $m = k - 1$.

$$= e^{-\lambda} \sum_{m=0}^{+\infty} \frac{\lambda^{m+1}}{m!} = \lambda e^{-\lambda} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Calcoliamoci il momento di ordine 2 per poi calcolarci la varianza.

$$\mathrm{E}[X^2] = \sum_{k=0}^{+\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \left[\sum_{k=1}^{+\infty} \frac{k-1}{(k-1)!} \lambda^k + \sum_{k=1}^{+\infty} \frac{1}{(k-1)!} \lambda^k \right]$$

$$= e^{-\lambda} \left[\sum_{k=2}^{+\infty} \frac{1}{(k-2)!} \lambda^k + \lambda e^{\lambda} \right]$$

Poniamo $j = k - 2$.

$$= e^{-\lambda} \left[\sum_{j=0}^{+\infty} \frac{\lambda^{j+2}}{j!} + \lambda e^{\lambda} \right] = e^{-\lambda} \left[\lambda^2 \sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} + \lambda e^{\lambda} \right] = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda$$

A questo punto calcoliamoci la varianza.

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Troviamo così un importante risultato che caratterizza le variabili casuali di Poisson.

$$\mathrm{Var}(X) = \mathrm{E}[X] = \lambda$$

Trovate valor medio e varianza vediamo come si comporta la funzione generatrice dei momenti.

$$\Phi(t) = \mathrm{E}[e^{tX}] = \sum_{k=0}^{+\infty} e^{tx} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Come verifica possiamo ricavarci il momento di ordine uno tramite questa funzione.

$$\mathrm{E}[X] = \frac{d}{dt} \Phi(t) \bigg|_{t=0} = e^{\lambda(e^t-1)\lambda e^t} \bigg|_{t=0} = \lambda$$

Dopo aver verificato che valor medio e varianza sono uguali andiamo ad analizzare un'altra proprietà fondamentale della variabile casuale di Poisson, la **riproducibilità**.

PROPRIETÀ DI RIPRODUCIBILITÀ
Prese $X_1 \sim Po(\lambda_1)$ e $X_2 \sim Po(\lambda_2)$ indipendenti, allora:

$$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$$

Dimostrazione:
Date X_1 e X_2 sappiamo che le rispettive funzioni generatrici sono della forma

$$\Phi_{X_1}(t) = e^{\lambda_1(e^t-1)} \qquad \Phi_{X_2} = e^{\lambda_2(e^t-1)}$$

Essendo due variabili indipendenti possiamo applicare la proprietà della funzione generatrice dei momenti, per cui avremo

$$\Phi_{X_1+X_2}(t) = \Phi_{X_1}\Phi_{X_2}(t) = e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-2)} = e^{(\lambda_1+\lambda_2)(e^t-1)} \Longrightarrow X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$$

5.4.1 Distribuzione di Poisson (o Legge degli eventi rari)

Sia Y una variabile casuale definita in questo modo.

$$Y \sim B(n, p)$$

Imponiamo però le seguenti condizioni:

- $n >> 1$
- $np = \lambda$
  Il che implica $p = \frac{\lambda}{n}$, che stando alla prima condizione implica $p \to 0$, da qui il nome "Legge degli eventi rari".

Date queste condizioni, vediamone gli effetti su Y.

$$p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Facendo tendere n a $+\infty$ abbiamo che

$$\lim_{n\to+\infty} p_Y(k) = \lim_{n\to+\infty} \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Il primo fattore tende a 1 in quanto il grado del numeratore è uguale a quello del denominatore, il secondo fattore è una costante, mentre per il terzo fattore possiamo ragionare così

$$\lim_{n\to+\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \lim_{n\to+\infty} \left(1 - \frac{\lambda}{n}\right)^{n} = \lim_{n\to+\infty} e^{\ln\left(1-\frac{\lambda}{n}\right)^n} = \lim_{n\to+\infty} e^{n\ln\left(1-\frac{\lambda}{n}\right)} =$$

$$= \lim_{n\to+\infty} e^{n\left(-\frac{\lambda}{n}\right)} = e^{-\lambda}$$

Abbiamo quindi ottenuto il seguente risultato.

$$\lim_{n\to+\infty} p_Y(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Ovvero la funzione di massa di probabilità di Poisson. Infatti se ci calcoliamo valor medio e varianza di una binomiale otteniamo, per n che tende all'infinito:

$$\mathrm{E}[Y] = \lim_{n\to+\infty} np = \lim_{n\to+\infty} \lambda = \lambda$$

$$\mathrm{Var}(Y) = \lim_{n\to+\infty} npq = \lim_{n\to+\infty} \lambda\left(1 - \frac{\lambda}{n}\right) = \lambda$$

5.4.2 Processo stocastico di Poisson

Definizione 5.4.1 (Processo stocastico)

Famiglia di variabili casuali parametrizzate, ovvero dipendenti da un certo parametro.

$$X(\eta)$$

Un esempio di processo stocastico sono il numero di telefonate in anno (X), dove η potrebbe essere il numero di telefonate in un giorno.
Il processo stocastico di Poisson si propone di contare il numero di eventi a partire da un tempo $t = 0$. Essa è rappresentata dalla funzione $N(t)$.
In questo processo vi sono 5 proprietà richieste:

1. $N(0) = 0$
2. **INDIPENDENZA DEGLI INCREMENTI**
   Il numero di eventi in intervalli disgiunti è indipendente.

3. **STAZIONARIETÀ DEGLI INCREMENTI**
   Il numero di eventi in un dato intervallo dipende dalla lunghezza dell'intervallo, ma non dalla sua posizione.

4. $$\lim_{h\to 0} \frac{P(N(h)=1)}{h} = \lambda \Longrightarrow P(N(h)=1) \sim \lambda \cdot h$$

5. $$\lim_{h\to 0} \frac{P(N(h)=2)}{h} = 0 \Longrightarrow P(N(h)\geq 2) \sim 0$$

Cosa possiamo dire di $N(t)$?
Prendiamo un intervallo di tempo qualsiasi $[0,t]$ e dividiamolo in n sottointervalli, ognuno di lunghezza $\frac{t}{n}$. $N(t)$ è una variabile casuale discreta tale per cui $N(t) \in \{0, ..., +\infty\}$.
La probabilità che ci siano k eventi in [0,t] corrisponde a

$$P(N(t)=k) = P(A_k \cup B_k)$$

dove A_k = 'Ci sono sottointervalli con 2 o più eventi, ci sono sottointervalli con un evento e sottointervalli senza eventi, per un totale di k eventi.' e B_k ='Ci sono k sottointervalli con un evento e i restanti n/k senza eventi.'

$$A_k \cap B_k = \emptyset \Longrightarrow P(A_k \cup B_k) = P(A_k) + P(B_k) \simeq P(B_k)$$

Quest'ultima relazione vale se $n >> 1$, infatti se vale ciò per la proprietà 5 vale che $P(A_k) \simeq 0$.
B_k si comporta come una binomiale, quindi

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

Per la proprietà 4 si ha che $p = \lambda \cdot \frac{t}{n}$, quindi

$$= \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \longrightarrow \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad \text{per } n \to +\infty$$

Quindi $N(t) \sim Po(\lambda t)$.
Il processo stocastico di Poisson è utile anche per studiare gli intertempi, ovvero i tempi che trascorrono tra un evento e l'altro.
Definiamo X_k ='Intertempo trascorso tra l'evento (k-1)-esimo e l'evento k-esimo'. Esse sono variabili casuali continue in quanto si sta parlando di tempi.

$$P(X_1 \leq s) = 1 - P(X_1 > s) = 1 - P(N(s) = 0) = 1 - \frac{(\lambda s)^0}{0!} e^{-\lambda s} = 1 - e^{-\lambda s} = F_{X_1}(s)$$

Passiamo ora a X_2

$$P(X_2 > s | X_1 = t) = P(N(s) = 0) = e^{-\lambda s}$$

Il secondo passaggio deriva dalla terza proprietà del processo stocastico di Poisson.

5.5 Variabili casuali binomiali negative

Si ripetono esperimenti identici in maniera indipendente. Definita con p la probabilità di successo in un singolo esperimento, definiamo X in questo modo.

$$X = \text{'Numero prove effettuate per ottenere } n \text{ successi'} \qquad X \in \{n, ..., +\infty\}$$

Analizziamo la funzione di massa.

$P(X = k) = p(k) = P\Big($‘In $(k-1)$ prove si ottengono $(n-1)$ successi’ $\cap$ ‘Nell’ultima prova c’è un successo’$\Big)$

Essendo le prove indipendenti avremo che $p(k)$ corrisponde al prodotto delle probabilità dei due eventi. Il primo evento è schematizzabile come una variabile casuale binomiale di parametri $(k-1)$ e p, il secondo avrà probabilità p di successo.

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

Per trovare il valor medio usiamo la stessa tecnica usata per le variabili binomiali. Infatti X si può vedere come somma di tanti (n) esperimenti indipendenti.

$$X = Y_1 + Y_2 + ...Y_n$$

Dove le Y_l non sono altro che variabili casuali geometriche. Quindi possiamo calcolarci il valor medio nel modo seguente.

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{l=1}^{n} Y_l\right] = \sum_{l=1}^{n} \mathrm{E}[Y_l] = \sum_{l=1}^{n} \frac{1}{p} = \frac{n}{p}$$

Esempio 5.5.1 (Problema dei fiammiferi di Banach)

Il celebre matematico Stefan Banach soleva acquistare due scatole di cerini per volta. Ciascuna scatola nuova conteneva n fiammiferi. Se le metteva in tasca e prendeva i cerini scegliendo a caso una delle due scatole e rimettendola in tasca subito dopo. Quando cercando un cerino trovava che la scatola scelta era vuota, gettava via le due scatole e ne comprava altre due.
Qual è la probabilità che, quando il matematico estrae per la prima volta una scatola vuota, ci siano esattamente k fiammiferi nell’altra scatola?

Chiamiamo per semplicità le due tasche una A e l’altra B, e diciamo che $P(A) = P(B) = 0.5$ Possiamo dire che $k = \{0, ..., n\}$, in quanto un pacchetto al massimo contiene n sigarette e al minimo zero, inoltre il numero di volte in cui il matematico ha messo la mano in tasca sarà $n+1+(n-k) = 2n+1-k$, dove $n+1$ indica le volte che ha messo la mano nella tasca in cui poi il pacchetto finisce, quindi il numero dei fiammiferi più la volta in cui scopre che il pacchetto è vuoto, e dove $n-k$ indica le volte in cui ha preso il fiammifero dall’altra tasca in cui ne sono rimaste k.
Definiamo l’evento T_a =‘Il matematico scopre che la tasca A è vuota e in B ci sono k fiammiferi’.

$$P(T_a) = P(\text{‘ Su } (2n-k+1) \text{ volte ha scelto A, } n+1 \text{ volte ’})$$

La quale probabilità non è altro che la probabilità di una binomiale negativa.

$$P(T_a) = \binom{2n-k}{n} p^{n+1}(1-p)^{(2n-k+1)-(n+1)} = \binom{2n-k}{n}\left(\frac{1}{2}\right)^{n+1}\left(\frac{1}{2}\right)^{n-k} =$$

$$= \binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k+1}$$

Se definiamo un evento T_b in maniera analoga a T_a, e che avrà la stessa probabilità per ovvi motivi, possiamo risolvere il problema calcolandoci la probabilità dell’unione, e dato che i due eventi sono disgiunti possiamo dire che

$$P(T_a \cup T_b) = P(T_a) + P(T_b) = 2p(T_a) = \binom{2n-k}{n}\left(\frac{1}{2}\right)^{n-k}$$

5.6 Variabile casuale ipergeometrica

Dato un gruppo di N oggetti di cui, m oggetti sono del tipo A e $(N-m)$ del tipo B, si estraggono **senza reimmissione** n oggetti.

$$X = \text{'Numero oggetti di tipo } A \text{ estratti'} \qquad \max(0, n-(N-m)) \leq X \leq \min(n, m)$$

$$P(X=k) = p(k) = \frac{\binom{m}{n}\binom{N-m}{n-k}}{\binom{N}{n}}$$

Capitolo 6

Modelli di variabili casuali continue

6.1 Variabile casuale uniforme

Prendiamo una variabile casuale X definita nell'intervallo $[\alpha,\beta]$ e in cui tutti i valori dell'intervallo sono equiprobabili. Quindi la sua funzione di densità sarà la seguente.

$$f(x) = \begin{cases} k & x \in [\alpha,\beta] \\ 0 & \text{altrove} \end{cases} \qquad \text{Notazione sintetica: } X \sim U(\alpha,\beta)$$

Proviamo a dire qualcosa di più su k.
Sappiamo che è maggiore di zero, poichè la funzione f deve esserlo per definizione. In più possiamo calcolarcelo in questo modo.

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{\alpha}^{\beta} kdx = k(\beta-\alpha) = 1 \Longrightarrow k = \frac{1}{\beta-\alpha}$$

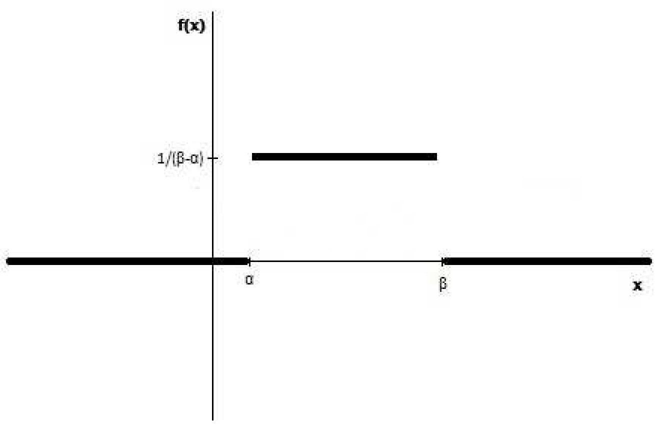


Figura 6.1: Grafico di $f(x)$

Trovato k calcoliamoci la funzione di distribuzione di X.

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds = \begin{cases} 0 & x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \alpha \leq x \leq \beta \\ 1 & x \geq \beta \end{cases}$$

Ora ci chiediamo, se prendiamo un ulteriore intervallo definito in $[\alpha,\beta]$, qual è la probabilità che X stia in quell' intervallo?

Per rispondere a questa domanda prendiamo un'intervallo $[\gamma,\delta]$ tale che $\alpha < \gamma < \delta < \beta$.

$$P(x \in [\gamma,\delta]) = P(\gamma \le x \le \delta) = P(x \le \delta) - P(x \le \gamma) = F(\delta) - F(\gamma) =$$
$$= \frac{\delta-\alpha}{\beta-\alpha} - \frac{\gamma-\alpha}{\beta-\alpha} = \frac{\delta-\gamma}{\beta-\alpha}$$

Quindi semplicemente il rapporto tra l'intervallo desiderato e l'intervallo in cui è definita la variabile.
Calcoliamo il valor medio.

$$\mathrm{E}[X] = \int_{-\infty}^{+\infty} x f(x)dx = \int_{\alpha}^{\beta} x \frac{1}{\beta-\alpha}dx = \frac{1}{\beta-\alpha}\left[\frac{\beta^2-\alpha^2}{2}\right] = \frac{\beta-\alpha}{2}$$

Come al solito per calcolare la varianza calcoliamoci prima il momento di ordine due di X.

$$\mathrm{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{\alpha}^{\beta} x^2 \frac{1}{\beta-\alpha}dx = \frac{1}{\beta-\alpha}\left[\frac{\beta^3-\alpha^3}{3}\right] = \frac{\beta^2+\alpha\beta+\alpha^2}{3}$$

La varianza risulterà quindi

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\beta-\alpha)^2}{4} = \frac{\beta-\alpha}{12}$$

6.2 Variabile casuale esponenziale

Sia X una variabile casuale avente come funzione di densità la seguente.

$$f(x) = \begin{cases} 0 & x \le 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases} \qquad \text{con } \lambda > 0$$

Indichiamo la variabile casuale esponenziale in questo modo: $X \sim E(\lambda)$.
Essa descrive tempi di durata ad esempio, oppure degli intertempi. Calcoliamoci la funzione di distribuzione di probabilità .

$$F(x) = \begin{cases} 0 & x \le 0 \\ \int_0^{+\infty} \lambda e^{-\lambda s} ds = \left[-e^{\lambda s}\right]_0^x = 1 - e^{-\lambda x} & x > 0 \end{cases}$$

Calcoliamoci la funzione generatrice dei momenti.

$$\Phi(t) = \mathrm{E}[e^{tx}] = \int_{-\infty}^{+\infty} f(x)e^{tx}dx = \int_0^{+\infty} \lambda e^{-\lambda x}e^{tx} = \lambda \int_0^{+\infty} e^{-(\lambda-t)x}dx =$$

Ponendo $\lambda - t > 0$

$$= \lambda\left[-\frac{1}{\lambda-t}e^{-(\lambda-t)x}\right]_0^{+\infty} = \frac{\lambda}{\lambda-t}$$

Calcoliamo il valor medio.

$$\mathrm{E}[X] = \left.\frac{d\Phi(t)}{dt}\right|_{t=0} = \left.\frac{\lambda}{(\lambda-t)^2}\right|_{t=0} = \frac{1}{\lambda}$$

Al solito prima di calcolarci la varianza calcoliamoci il momento di ordine 2.

$$\mathrm{E}[X^2] = \left.\frac{d^2\Phi(t)}{dt}\right|_{t=0} = \left.\frac{2\lambda}{(\lambda-t)^3}\right|_{t=0} = \frac{2}{\lambda^2}$$

A questo punto calcoliamoci la varianza.

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Andiamo ad analizzare più nel profondo questa distribuzione andando a vedere quali sono le sue proprietà.

PROPRIETÀ DELLA DISTRIBUZIONE ESPONENZIALE:

1. **Dispositivi in serie**
   Sia dato un dispositivo formato da N elementi in serie. Siano $X_1, X_2, X_3, ..., X_N =$'Durata di funzionamento di ogni elemento'; e siano esse indipendenti tra loro. Definiamo $Y =$'Durata funzionamento del dispositivo'.
   Per le proprietà di un dispositivo in serie $Y = \min(X_1, X_2, ..., X_N)$.
   Per come sono state definite $X_1, X_2, ...X_N$ sono variabili casuali esponenziali di parametro rispettivamente $\lambda_1, ..., \lambda_N$. Vogliamo dimostrare come anche Y sia una variabile casuale esponenziale di parametro $(\lambda_1 + \lambda_2 + ...\lambda_N)$.

$$F_Y(y) = P(\min(X_1, ...X_N) \leq y) = 1 - P(\min(X_1, ..., X_N) > y) =$$

$$1 - P(X_1 > y \cap X_2 > y \cap ... \cap X_N > y) = 1 - P(X_1 > y)P(X_2 > y) \cdots P(X_N > y) =$$
$$= 1 - e^{-\lambda_1 y} e^{-\lambda_2 y} \cdots e^{-\lambda_N y} = 1 - e^{-(\lambda_1 + \lambda_2 + ... + \lambda_N)y}$$
$$\Downarrow$$
$$Y \sim E(\lambda_1 + \lambda_2 + ... + \lambda_N)$$

2. **Dispositivi in parallelo**
   Sia dato un dispositivo formato da N elementi in parallelo. Siano $X_1, X_2, X_3, ..., X_N =$'Durata di funzionamento di ogni elemento'; e siano esse indipendenti tra loro. Definiamo $Y =$'Durata funzionamento del dispositivo'.
   Per le proprietà di un dispositivo in parallelo $Y = \max(X_1, X_2, ..., X_N)$.
   Per come sono state definite $X_1, X_2, ...X_N$ sono variabili casuali esponenziali di parametro rispettivamente $\lambda_1, ..., \lambda_N$

$$F_Y(y) = P(Y \leq y) = P(\max(X_1, X_2, ..., X_N) \leq y) = P(X_1 \leq y) \cap P(X_2 \leq y) \cap ... \cap P(X_N \leq y) =$$

$$= P(X_1 \leq y)P(X_2 \leq y) \cdots P(X_N \leq y) = F_{X_1}(y)F_{X_2}(y) \cdots F_{X_N}(y)$$

   Quindi possiamo dire che

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ \left(1 - e^{-\lambda_1 y}\right)\left(1 - e^{-\lambda_2 y}\right) \cdots \left(1 - e^{-\lambda_N y}\right) & y > 0 \end{cases}$$

3. **Assenza di memoria**
   Sia X una variabile casuale esponenziale di parametro λ e che rappresenta il tempo di durata.
   Siano $s, t > 0$, andiamo a calcolarci la seguente probabilità.

$$P(X > s + t | X > t) = \frac{P(X > s + t \cap X > t)}{P(X > t)} = \frac{P(X > s + t)}{P(X > t)}$$

   Analizziamo prima il numeratore.

$$P(X > s + t) = 1 - P(X \leq s + t) = 1 - F_x(s + t) = 1 - \left(1 - e^{-\lambda(s+t)}\right) = e^{-\lambda(s+t)}$$

   Ora il denominatore.

$$P(X > t) = 1 - P(X \leq t) = 1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t}$$

Quindi possiamo dire che

$$\frac{P(X > s+t)}{P(X > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$$

Quindi la probabilità non dipende da t!!! Vi è appunto assenza di memoria.

4. Sia $Y = cX$, dove $X \sim E(\lambda)$ e $c \in \mathbb{R}_+^* \Longrightarrow Y \sim E\left(\frac{\lambda}{c}\right)$
   Dimostrazione:

$$\Phi_X(t) = \mathrm{E}[e^{tx}] = \frac{\lambda}{\lambda - t} \qquad \text{se } t < \lambda$$

$$\Phi_Y(t) = \mathrm{E}[e^{ty}] = \mathrm{E}[e^{tcx}] = \Phi_X(ct) = \frac{\lambda}{\lambda - tc} = \frac{\frac{\lambda}{c}}{\frac{\lambda}{c} - t} \qquad \text{se } ct < \lambda$$

   L'ultima relazione implica proprio che $Y \sim E\left(\frac{\lambda}{c}\right)$.

Ci chiediamo ora se valga la proprietà di riproducibilità per $X \sim \mathrm{E}(\lambda)$.
Prendiamo $X_1 \sim E(\lambda_1)$ e $X_2 \sim E(\lambda_2)$ variabili casuali indipendenti.

$$\Phi_{X_1+X_2}(t) = \Phi_{X_1}(t)\Phi_{X_2}(t) = \frac{\lambda_1}{(\lambda_1 - t)}\frac{\lambda_2}{(\lambda_2 - t)} = \frac{\lambda_1\lambda_2}{(\lambda_1 - t)(\lambda_2 - t)} =$$

$$\frac{\lambda_1 - \lambda_2}{\lambda_1\lambda_2 - \lambda_1 t - \lambda_2 t + t^2} \qquad \text{se } t < \min(\lambda_1, \lambda_2)$$

Il quale risultato ci dice che la somma di due esponenziali **non** è ancora una variabile esponenziale.

6.3 Variabile casuale gaussiana o normale

Sia X una variabile casuale continua avente la seguente funzione di densità di probabilità.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{Notazione sintetica: } N \sim (\mu, \sigma)$$

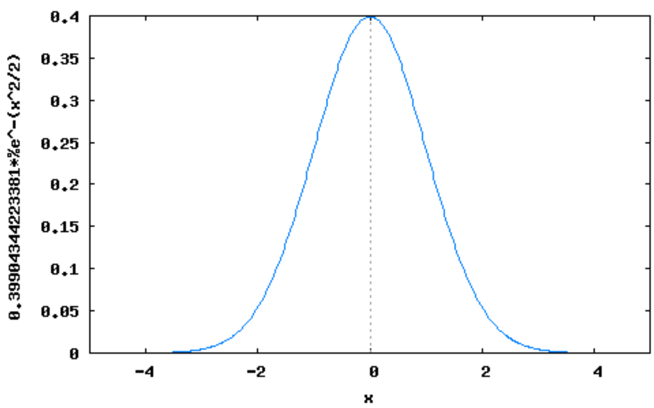


Figura 6.2: Grafico di una variabile $X \sim N(0,1)$

Questo grafico è noto come "Campana di Gauss". Per le proprietà della funzione di densità l'area del grafico sappiamo essere uguale a 1.
Proviamo ora a calcolarci la funzione di distribuzione della gaussiana.

$$F(a) = \int_{-\infty}^{a} f(x)dx = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Siamo arrivati a dover integrare una funzione di cui non conosciamo la primitiva. Quindi possiamo dire ben poco sulla funzione di distribuzione, se non che

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = 1$$

e che, per simmetria

$$F(\mu) = \int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{2}$$

Calcoliamoci la funzione generatrice dei momenti.

$$\Phi(t) = \mathrm{E}[e^{tx}] = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx =$$

Ponendo $y = \frac{x-\mu}{\sigma}$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t(\sigma y+\mu)} e^{-\frac{y^2}{2}} \sigma dy = \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{2t\sigma y - y^2}{2}} dy = \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(y-t\sigma)^2 + t^2\sigma^2}{2}} dy =$$

$$= \frac{e^{t\mu + \frac{t^2\sigma^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(y-t\sigma)^2}{2}} dy = \frac{e^{t\mu + \frac{t^2\sigma^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-s^2}{2}} ds$$

L'ultimo integrale sappiamo essere uguale a $\sqrt{2\pi}$, quindi la nostra espressione diventa

$$\Phi(t) = e^{t\mu + \frac{t^2\sigma^2}{2}}$$

Trovata la funzione generatrice dei momenti calcoliamoci valor medio e varianza.

$$\mathrm{E}[X] = \left.\frac{d\Phi(t)}{dt}\right|_{t=0} = (\mu + t\sigma^2) e^{t\mu + \frac{t^2\sigma^2}{2}}\Big|_{t=0} = \mu$$

$$\mathrm{E}[X^2] = \left.\frac{d^2\Phi(t)}{dt^2}\right|_{t=0} = \sigma^2 e^{t\mu + \frac{t^2\sigma^2}{2}} + (\mu + t\sigma^2)^2 e^{t\mu + \frac{t^2\sigma^2}{2}}\Big|_{t=0} = \sigma^2 + \mu^2$$

Possiamo calcolarci quindi la varianza.

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

PROPRIETÀ DELLA GAUSSIANA:

1. Sia $X \sim N(\mu, \sigma)$, definiamo $Y = \alpha X + \beta$, con $\alpha, \beta \in \mathbb{R}$. Calcoliamo valor medio, varianza e funzione generatrice di questa trasformazione lineare di X.

   $$\mathrm{E}[Y] = \mathrm{E}[\alpha X + \beta] = \alpha\mu + \beta$$

   $$\mathrm{Var}(Y) = \mathrm{Var}(\alpha X + \beta) = \alpha^2 \mathrm{Var}(X) = \alpha^2\sigma^2$$

   $$\Phi_Y(t) = \mathrm{E}[e^{ty}] = \mathrm{E}[e^{t(\alpha x + \beta)}] = \mathrm{E}[e^{t\alpha x} e^{t\beta}] = e^{t\beta}\mathrm{E}[e^{t\alpha X}] = e^{t\beta}\Phi_X(t\alpha)$$

   Andando a sostituire l'espressione calcolata prima otteniamo

   $$= e^{t\beta} e^{t\alpha\mu + \frac{t^2\alpha^2\sigma^2}{2}} = e^{t(\alpha\mu+\beta) + \frac{t^2}{2}(\alpha^2\sigma^2)}$$

   Notiamo che Y ha le caratteristiche di una gaussiana, in cui $\mathrm{E}[Y] = (\alpha\mu + \beta)$ e $\mathrm{Var}(X) = (\alpha^2\sigma^2)$, quindi $Y \sim N(\alpha\mu + \beta\, , \, \alpha\sigma)$

2. Sia $X \sim N(\mu, \sigma)$, dalla proprietà 1 definiamo la variabile detta **normale standard** in questo modo.
$$\frac{X-\mu}{\sigma} = Z \sim N(0,1)$$
Essa non è altro che una trasformazione lineare di X, in cui $\alpha = \frac{1}{\sigma}$ e $\beta = -\frac{\mu}{\sigma}$. Applicando la prima proprietà ricaviamo anche i valori del valor medio e della varianza.
$$\alpha\mu + \beta = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0, \qquad \alpha^2\sigma^2 = \frac{1}{\sigma^2}\sigma^2 = 1$$
La normale standard ha la proprietà che da essa ci si può ricondurre ad un qualsiasi calcolo delle probabilità di variabili gaussiane. Infatti
$$F_X(a) = P(X \le a) = P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = F_Z\left(\frac{a-\mu}{\sigma}\right)$$
Inoltre
$$P(a < X < b) = P(X \le b) - P(X \le a) = P\left(\frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)$$

3. **Proprietà di riproducibilità**
Prese due variabili casuali $X_1 \sim N(\mu_1, \sigma_1)$ e $X_2 \sim N(\mu_2, \sigma_2)$ indipendenti, allora anche
$$X_1 + X_2 \sim N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$
Dimostrazione: Sappiamo che
$$\Phi_{X_1}(t) = e^{t\mu_1 + \frac{t^2}{2}\sigma_1^2} \qquad \Phi_{X_2}(t) = e^{t\mu_2 + \frac{t^2}{2}\sigma_2^2}$$
$\Phi_{X_1+X_2}(t) = \Phi_{X_1}(t)\Phi_{X_2}(t)$ per indipendenza. Quindi
$$\Phi_{X_1+X_2}(t) = e^{t(\mu_1+\mu_2) + \frac{t^2}{2}(\sigma_1^2+\sigma_2^2)}$$
Quest'ultimo risultato dimostra la proprietà.

6.4 Variabile casuale χ^2 a n gradi di libertà

Sia C una variabile casuale tale per cui
$$C = \sum_{k=1}^{n} Z_k^2 \qquad \text{Notazione sintetica: } C \sim \chi_n^2$$
Dove $Z_k \sim N(0,1)$ sono indipendenti tra loro.
Vengono definiti per questo tipo di variabili i valori critici, ovvero quelli per cui si ha che, se α è uno di questi valori,
$$P(C > \chi^2_{n,\alpha}) = \alpha$$
Calcoliamo il valor medio.
$$\mathrm{E}[C] = \sum_{k=1}^{n} \mathrm{E}[Z_k^2] = \sum_{k=1}^{n} \left(\mathrm{Var}(X) + \mathrm{E}[Z_k]^2\right) = n$$

6.5 Variabile casuale di Student a n gradi di libertà

Sia T una variabile casuale tale per cui

$$T = \frac{Z}{\sqrt{\frac{C}{n}}} \qquad \text{Notazione sintetica: } T \sim t_n$$

Dove $C \sim \chi^2_n$ e $Z \sim N(0,1)$.

$$\mathrm{Var}(T) = \frac{n}{n-2}, \qquad \mathrm{E}[T] = 0$$

Anche per la variabile di studente viene definito il valore critico.

$$P(T \geq t_{n,\alpha}) = \alpha$$

Essa è una funzione simmetrica, per cui

$$P(T \leq -t_{n,\alpha}) = \alpha$$

6.6 Funzioni di una variabile casuale continua

Sia X una variabile casuale nota. Ci proponiamo di analizzare la seguente situazione.

$$Y = \varphi(X)$$

Per calcolare il valor medio e il momento di ordine generico rimangono valide le solite formule. Per calcolare la funzione di densità di probabilità invece dobbiamo distinguere 3 casi.

1. φ **monotona crescente**

$$F_Y(y) = P(Y \leq y) = P(\varphi(X) \leq y)$$

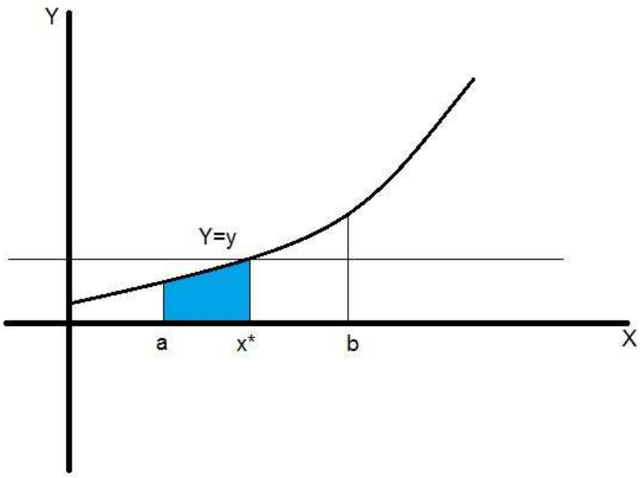


Figura 6.3: Grafico funzione monotona crescente

Se $X \in [a,b]$ scegliamo un x^* tale che $F_Y(y) = P(a \leq X \leq x^*)$. Sapendo che $\varphi(x^*) = y$ e che quindi $x^* = \varphi^{-1}(y)$, diciamo che

$$P(Y \leq y) = \int_a^{x^*} f_X(s)ds = \int_a^{\varphi^{-1}(y)} f_X(s)ds$$

Trovata la funzione di ripartizione ci calcoliamo la funzione di densità facendone la derivata.

$$f(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}\int_a^{\varphi^{-1}(y)} f_X(s)ds = f_X\left(\varphi^{-1}(y)\right)\frac{d}{dy}\varphi^{-1}(y)$$

2. φ **monotona decrescente**

$$F_Y(y) = P(Y \leq y) = P(\varphi(X) \leq y)$$

Il caso è analogo al precedente, con l'unica differenza che la X dovrà essere maggiore di x^*.

$$P(Y \leq y) = P(X \geq x^*) = \int_{x^*}^{+\infty} f_X(s)ds = \int_{\varphi^{-1}(y)}^{+\infty} f_X(s)ds$$

Quindi

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}\int_{\varphi^{-1}(y)}^{+\infty} f_X(s)ds = -f_X\left(\varphi^{-1}(y)\right)\frac{d}{dy}\varphi^{-1}(y) = f_X\left(\varphi^{-1}(y)\right)\left(-\frac{d}{dy}\varphi^{-1}(y)\right)$$

Si noti che la funzione di densità mantiene la sua proprietà di positività.
In generale per una funzione monotona possiamo dire che

$$f_Y(y) = f_X\left(\varphi^{-1}(y)\right)\left|\frac{d}{dy}\varphi^{-1}(y)\right|$$

3. φ **non monotona**

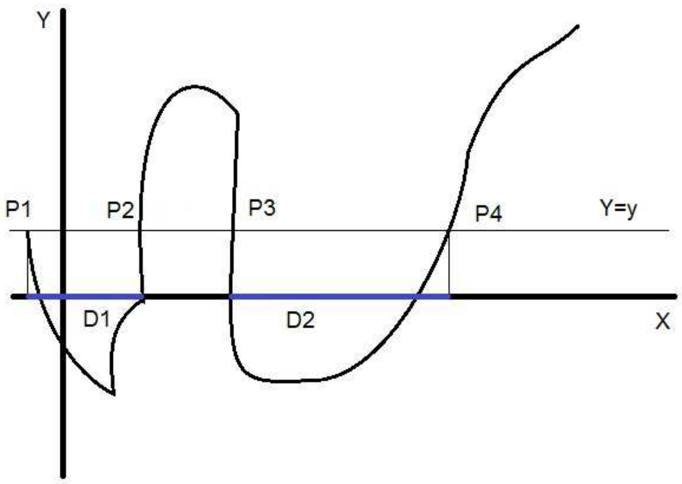


Figura 6.4: Grafico funzione non monotona

$$F_Y(y) = P(\varphi(x) \leq y) = P(X \in D_1(y) \cup X \in D_2(y) \cup ...) = P(X \in D_1(y)) + P(X \in D_2(y)) + ... =$$

$$= \int_{D_1(y)} f_X(x)dx + \int_{D_2(y)} f_X(x)dx + ...$$

Ovviamente varrà sempre la relazione

$$f_Y(y) = \frac{dF_Y}{dy}$$

6.7 Funzione di più variabili casuali

Siano X e Y due variabili casuali note. Ci proponiamo di analizzare la seguente situazione.

$$Z = \psi(X, Y)$$

Per semplicità analizzeremo solo il caso in cui $Z = X + Y$.
Abbiamo già visto come calcolare valor medio e varianza di una somma. Proviamo a dire qualcosa di più sulla funzione di densità e sulla funzione di distribuzione.

$$F_Z(a) = P(Z \leq a) = P(X + Y \leq a) = P(X \leq a - Y) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{a-Y} f(x, y) dx \right) dy$$

Oppure possiamo anche dire che

$$F_Z(a) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{a-X} f(x, y) dy \right) dx$$

Essendo $P(X \leq a - Y) = P(Y \leq a - X)$.
La funzione di densità la calcoliamo come al solito come la derivata della funzione di distribuzione.

$$f_Z(a) = \frac{dF_Z(a)}{da} = \frac{d}{da} \left(\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{a-Y} f(x, y) dx \right) dy \right) = \int_{-\infty}^{+\infty} f(a - y, y) dy$$

Oppure per il ragionamento precedente

$$f_Z(a) = \int_{-\infty}^{+\infty} f(x, a - x) dx$$

Se X eY sono indipendenti allora

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

Quest'ultimi integrali sono detti **integrali di convoluzione**.

Capitolo 7

Metodo Montecarlo

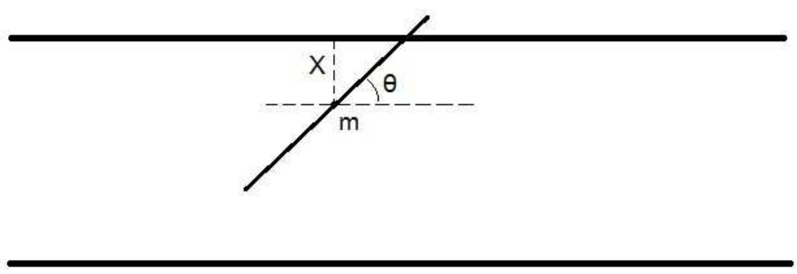
Il metodo Montecarlo consiste nello stimare dei valori attraverso la definizione frequentista di probabilità utilizzando punti o oggetti generati casualmente.
Illustriamo il metodo attraverso due famosi esempi: il problema dell'ago di Buffon e il problema della stima dell'area di una regione di spazio.

7.1 Problema degli spilli di Buffon

L'obiettivo di Buffon era quello di calcolare una stima del valore di π facendo degli esperimenti. Egli suppose di avere un pavimento in legno formato da strisce di legno parallele, ognuna a distanza $2b$ dall'altra. In seguito egli immaginò di lanciare un numero n di spilli lunghi $2a < 2b$ sul pavimento.
Chiamato m il punto medio dello spillo, definisco due variabili casuali X e θ.

$$X = \text{'Distanza da } m \text{ dalla retta più vicina'} \qquad X \sim U(0,b)$$

$$\theta = \text{'Angolo con cui lo spillo interseca la retta'} \qquad \theta \sim U(0,\pi)$$



$$P(\text{'Spillo lanciato a caso intersechi una retta'}) = P(a\sin\theta \geq X) = P(X \leq a\sin\theta) =$$

$$= \iint_{X \leq a\sin\theta} f(x,\theta)dxd\theta = \int_0^{\pi}\left(\int_0^{a\sin\theta}\frac{1}{\pi b}dx\right)d\theta = \int_0^{\pi}\frac{1}{\pi b}[a\sin\theta]d\theta = \frac{2a}{\pi b}$$

Ora se chiamo n_i ='Numero di spilli lanciati casualmente che intersecano una retta ' e n_{tot} ='Numero spilli lanciati ' ottengo

$$\text{Frequenza relativa di intersezioni} = \frac{n_i}{n_{tot}} \approx \frac{2a}{\pi b} \quad \text{con } n_{tot} >> 1$$

Questo risultato discende dal corollario di Bernoulli. Buffon trovò quindi che

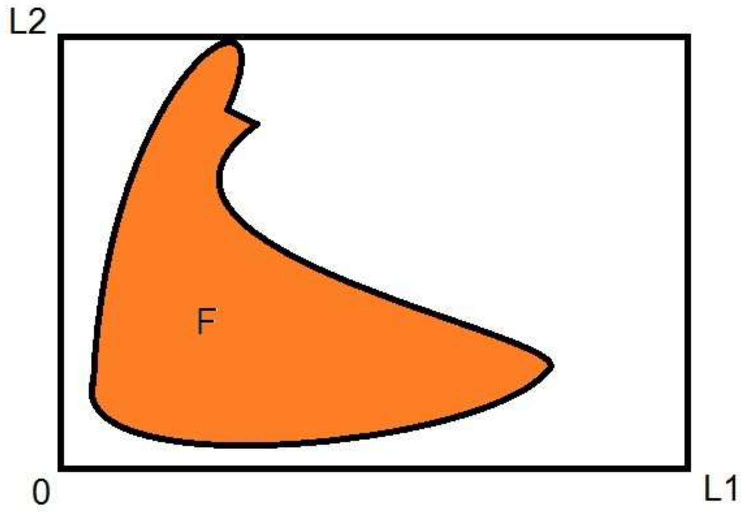
$$\pi \approx \frac{2an_{tot}}{bn_i}$$

Il matematico italiano Mario Lazzarini realizzò l'esperimento di Buffon nel 1901. Lanciando un ago 3408 volte, ottenne la nota stima 355/113 per π, che è un valore molto accurato, differendo da π per non più di $3 \cdot 10^{-7}$.

7.2 Problema della stima dell'area di una regione

Ci poniamo il problema di calcolare l'area della regione F. Definiamo due variabili casuali X e Y indipendenti, corrispondenti alla lunghezza del rettangolo.

$$X \sim U(0, L_1) \quad Y \sim U(0, L_2)$$



Sappiamo che

$$P((X,Y) \in F) = \iint_F f(x,y)dxdy$$

Inoltre

$$f(x,y) = f_X(x)f_Y(y) = \begin{cases} \frac{1}{L_1 L_2} & 0 \leq x \leq L_1, \quad 0 \leq y \leq L_2 \\ 0 & \text{altrimenti} \end{cases}$$

Quindi

$$P((X,Y) \in F) = \iint_F \frac{1}{A_r} dxdy = \frac{1}{A_r} \iint_F dxdy = \frac{A_F}{Ar}$$

Supponiamo di generare casualmente punti sulla regione rettangolare. Chiamiamo N_{tot} ='Numero totale di punti generati ' e n_f ='Numero di punti che finiscono in F '.

$$\frac{n_f}{n_{tot}} \approx \frac{A_F}{A_r} \Longrightarrow A_F \approx \frac{n_F}{n_{tot}} A_r \quad \text{con } n >> 1$$

Approssimazione ottenuta grazie al Corollario di Bernoulli.

Capitolo 8

Inferenza statistica o statistica referenziale

8.1 Introduzione e Teorema del limite centrale

Fondamentali per fare inferenza statistica sono i concetti di *popolazione* e *campione*, già visti nel primo capitolo. Per semplificarci la vita introduciamo un'ipotesi fondamentale:

> *X, che rappresenta la caratteristica che andiamo a studiare, è distribuita in* ***maniera uguale*** *per ogni individuo, inoltre variabili casuali per individui diversi sono indipendenti tra loro.*

Esistono 2 tipi di inferenza statistica:

- **Non parametrica**, se F è incognita, ovvero se non so quale distribuzione caratterizza X.
- **Parametrica**, se la forma di F è nota a meno di qualche parametro.

Il campione non è altro che un vettore di variabili casuali indipendenti e identicamente distribuite.

$$\text{Campione} = (X_1, X_2, ..., X_n)$$

La statistica è una funzione delle variabili casuali che compongono il campione.

$$Y = \varphi(X_1, X_2, ..., X_n)$$

Prendiamo in esame ora la media campionaria.

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

Per quanto riguarda valor medio e varianza abbiamo già visto che

$$\mathrm{E}[X_i] = \mu, \quad \mathrm{Var}(X_i) = \sigma^2$$

$$\mathrm{E}[\overline{X}] = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

Questo ovviamente se le X_i sono indipendenti.
Ora ci chiediamo, come è distribuita la media campionaria?

- **Caso 1**
  Se X_i hanno distribuzioni riproducibili (Poisson o Gauss) allora anche $\overline{X}$ avrà la stessa forma di distribuzione.
- **Caso 2**
  Se X_i non hanno distribuzioni riproducibili è possibile applicare il teorema del limite centrale (se $n >> 1$).

Teorema 8.1.1 (del limite centrale)

Data una successione di variabili casuali $X_1, X_2, ..., X_n$ identicamente distribuite e indipendenti, dove $E[X_k] = \mu,\ Var(X_k) = \sigma^2$

$$\Downarrow$$

$$Y = \sum_{k=1}^{n} X_k \quad \textit{converge in ditribuzione ad una gaussiana per } n \to +\infty$$

Oppure

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

Oppure

$$P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}} \le a\right) \longrightarrow F_z(a) \quad \textit{se } n \to +\infty$$

OSSERVAZIONI:

$$Y = \sum_{k=1}^{n} X_k$$

$$\mathrm{E}[Y] = n\mu$$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = n\sigma^2$$

Questo vuol dire che Y è già una gaussiana.
Conseguenze:

1. $$\overline{X} = \frac{Y}{n} = \frac{\sum_{k=1}^{n} X_k}{n}$$

   $$\text{se } Y \sim N\left(n\mu, \sigma\sqrt{n}\right) \Longrightarrow \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \qquad \text{in accordo con prima!}$$

2. $$S \sim B(n,p) \quad \text{con } n >> 1\ (n > 30)$$

   $$S = \sum_{k=1}^{n} X_k \quad \text{indipendenti e identicamente distribuite} \quad \text{dove } X_k \sim Be(p)$$

   Applicando il Teorema del limite centrale

   $$S \sim N(np, \sqrt{npq})$$

   Quindi possiamo vedere S sia come variabile casuale continua che discreta.
   Sorge un problema. Infatti se $S \sim B(n,p)$

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

Se $S \sim N\left(np, \sqrt{npq}\right)$
$$P(S = k) = 0$$
Essendo continua!! Per ovviare a questo contrasto si ricorre ad una approssimazione alla continuità, ovvero se $S \sim B(n, p)$
$$P(S = k) = P(k - 0,5 < S < k + 0,5)$$
Quindi per S continua
$$P(k - 0,5 < S < k + 0,5) \neq 0$$
È un modo per stimare $P(S = k)$.

Dimostrazione del teorema del limite centrale:
Per facilitare la dimostrazione introduciamo 2 ipotesi aggiuntive, ovvero che $\sigma^2 = 1$ e che $\mu = 0$.
$$\text{Tesi:} \quad Y \sim N\left(0, \sqrt{n}\right) \Leftrightarrow \frac{Y}{\sqrt{n}} \sim N(0, 1)$$
Dimostriamo che $\frac{Y}{\sqrt{n}}$ si comporta come una gaussiana.
$$\frac{Y}{\sqrt{n}} = \sum_{k=1}^{n} \frac{X_k}{\sqrt{n}} \Longrightarrow \Phi_{\frac{Y}{\sqrt{n}}}(t) = \mathrm{E}\left[e^{\frac{tY}{\sqrt{n}}}\right] = \Phi_Y\left(\frac{t}{\sqrt{n}}\right)$$
Y è la somma di n variabili indipendenti tra loro e identicamente distribuite, quindi
$$\Phi_Y\left(\frac{t}{\sqrt{n}}\right) = \Phi_{X_1+X_2+\ldots+X_n}\left(\frac{t}{\sqrt{n}}\right) = \Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right) \cdot \Phi_{X_2}\left(\frac{t}{\sqrt{n}}\right) \cdots \Phi_{X_n}\left(\frac{t}{\sqrt{n}}\right) =$$
$$= \left[\Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$
Voglio dimostrare che
$$\lim_{n\to+\infty} \Phi_{\frac{Y}{\sqrt{n}}}(t) = \Phi_Z(t)$$
Io so già che
$$\Phi_Z(t) = e^{\frac{t^2}{2}}$$
Quindi se è vero questo deve essere vero anche questo
$$\lim_{n\to+\infty} \ln \Phi_{\frac{Y}{\sqrt{n}}}(t) = \frac{t^2}{2}$$
Sappiamo che
$$\Phi_{\frac{Y}{\sqrt{n}}}(t) = \left[\Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n \Longrightarrow \ln \Phi_{\frac{Y}{\sqrt{n}}}(t) = n \ln \Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)$$
Quindi ora il nostro limite diventa
$$\lim_{n\to+\infty} n \ln \Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2}$$
Per verificare questo limite ci è comodo definire $L(s) = \ln \Phi_{X_1}(s)$ Verifichiamo che
$$L(0) = 0$$
$$L'(0) = \left.\frac{\Phi'_{X_1}(s)}{\Phi_{X_1}(s)}\right|_{s=0} = 1 \cdot \mathrm{E}[X_1] = 0 \quad \text{avendo supposto } \mu = 0$$

$$L''(0) = -\frac{\Phi'_{X_1}(s)}{(\Phi_{X_1}(s))^2}\cdot\Phi'_{X_1}(s)+\frac{\Phi''_{X_1}(s)}{\Phi_{X_1}(s)}\bigg|_{s=0} = \mathrm{E}[X_1^2]-\mathrm{E}[X_1]^2 = \mathrm{Var}(X_1) = 1$$

Ora procediamo con la risoluzione del limite.

$$\lim_{n\to+\infty} n\ln\Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \lim_{n\to+\infty} nL\left(\frac{t}{\sqrt{n}}\right) = \lim_{n\to+\infty}\frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} =$$

Ci troviamo davanti ad una forma indeterminata del tipo $\frac{0}{0}$, applichiamo il Teorema di De L'Hopital.

$$= \lim_{n\to+\infty}\frac{L'\left(\frac{t}{\sqrt{n}}\right)\left(-\frac{t}{2}n^{-\frac{3}{2}}\right)}{-n^{-2}} = \lim_{n\to+\infty}\frac{L'\left(\frac{t}{\sqrt{n}}\right)\left(-\frac{t}{2}\right)}{-n^{-\frac{1}{2}}} =$$

Troviamo ancora una forma indeterminata, allora riapplichiamo il Teorema.

$$= \lim_{n\to+\infty}\frac{\left(-\frac{t}{2}\right)^2 L''\left(\frac{t}{\sqrt{n}}\right)n^{-\frac{3}{2}}}{\frac{1}{2}n^{-\frac{3}{2}}} = \lim_{n\to+\infty}\frac{\frac{t^2}{4}}{\frac{1}{2}} = \frac{t^2}{2}$$

Abbiamo dimostrato così il teorema.

8.2 Inferenza statistica negli esperimenti

L'inferenza statistica pensa a come fare un esperimento, non lo fa e basta, non scende nel dettaglio.

8.2.1 Varianza campionaria

La varianza campionaria indica quanto i dati si discostano dalla media campionaria.
È definita nel modo seguente.

$$S^2 = \sum_{k=1}^{n}\frac{\left(X_k-\overline{X}\right)^2}{n-1}$$

Se $n = 1$ e $\overline{X} = X_1$

$$S^2 = \frac{0}{0}$$

Ovvero una forma indeterminata!!! Questo ragionevolmente indica che per avere un valore della varianza campionaria servono almeno due dati. Calcoliamo il valor medio della varianza campionaria.

$$\mathrm{E}\left[S^2\right] = \mathrm{E}\left[\sum_{k=1}^{n}\frac{\left(X_k-\overline{X}\right)^2}{n-1}\right] = \frac{1}{n-1}\mathrm{E}\left[\sum_{k=1}^{n}\left(X_k-\overline{X}\right)^2\right] = \frac{1}{n-1}\mathrm{E}\left[\sum_{k=1}^{n}X_k^2+\sum_{k=1}^{n}\overline{X}^2-2\overline{X}\sum_{k=1}^{n}X_k\right] =$$

$$= \frac{1}{n-1}\mathrm{E}\left[\sum_{k=1}^{n}X_k^2+n\overline{X}^2-2n\overline{X}^2\right] = \frac{1}{n-1}\left(\sum_{k=1}^{n}\mathrm{E}\left[X_k^2\right]-n\mathrm{E}\left[\overline{X^2}\right]\right) =$$

Supponiamo ora che $\mathrm{E}[X_k] = \mu$ e $\mathrm{Var}(X_k) = \sigma^2$

$$= \frac{1}{n-1}\left(\sum_{k=1}^{n}\left(\sigma^2+\mu^2\right)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) = \frac{1}{n-1}\left(n\left(\sigma^2+\mu^2\right)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) =$$

$$= \frac{1}{n-1}\left(n\sigma^2 - \sigma^2\right) = \frac{n-1}{n-1}\sigma^2 = \sigma^2$$

Questo è il secondo motivo per cui nella definizione di varianza campionaria a denominatore si ha $n-1$ e non n.
Per quanto riguarda la varianza campionaria si può dimostrare questo importante risultato che ci sarà utile più avanti.
Se abbiamo una popolazione gaussiana allora

$$\frac{S^2}{\sigma^2}(n-1) \sim \chi^2_{n-1}$$

I gradi di libertà sono $(n-1)$ perchè in realtà $\overline{X}$ dipende da X_k. Ora, come possiamo sfruttare la varianza campionaria?
Noi conosciamo già questo risultato

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

però non sempre conosciamo σ, mentre conosciamo sempre S, quindi ci è utile capire la distribuzione della variabile seguente.

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

Presa $Z \sim N(0,1)$ e $C \sim \chi^2_{n-1}$, noi sappiamo che

$$\frac{Z}{\sqrt{\frac{c}{n-1}}} \sim t_{n-1}$$

Andando a sostituire

$$\frac{\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\frac{\sigma}{S} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

8.2.2 Stima parametrica

Parto da una popolazione con funzione di distribuzione F_θ conosciuta a meno del parametro θ. La nostra incognita è μ, oppure sia μ che σ^2, in una gaussiana. Si cerca di stimare θ attraverso una particolare statistica chiamata **stimatore**$\left(\hat{\theta}\right)$.
Lo stimatore per essere accettabile deve rispettare 3 proprietà:

1. **Correttezza**

   $$\mathrm{E}\left[\hat{\theta}\right] = \theta$$

   Altro motivo per cui nella definizione di varianza campionaria a denominatore vi è $n-1$; altrimenti non potrei mai utilizzarla.

2. **Efficienza**

   $$\mathrm{Var}\left(\hat{\theta}\right) \quad \text{più piccola possibile}$$

3. **Consistenza**

   $$\lim_{n\to+\infty} \mathrm{E}\left[\left(\hat{\theta}-\theta\right)^2\right] = 0$$

   Nel caso dello stimatore $\overline{X}$ infatti

   $$\lim_{n\to+\infty} \mathrm{E}\left[\left(\overline{X}-\mu\right)^2\right] = \lim_{n\to+\infty} \mathrm{Var}\left(\overline{X}\right) = \lim_{n\to+\infty} \frac{\sigma^2}{n} = 0$$

8.2.3 Intervallo di confidenza

In un esperimento trovare un valore preciso di μ o di σ^2 è impossibile, pertanto cerchiamolo in un **intervallo**, detto **di confidenza**, centrato nel valore dello stimatore, detto **stima**. Per trovare un intervallo di μ partiamo dal fatto che sicuramente possiamo dire che

$$\mu \in \left[\overline{X}-\beta,\ \overline{X}+\beta\right]$$

Come scelgo β?
Ipotizziamo che la popolazione sia gaussiana e che σ^2 sia nota.
Per $Z \sim N(0,1)$ definisco Z_γ come valore critico, quindi tale che

$$P(Z > Z_\gamma) = \gamma$$

Poniamo ora $\gamma = \frac{\alpha}{2}$

$$P\left(-Z_{\frac{\alpha}{2}} \le Z \le Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Se ora prendiamo una Z di questo tipo

$$Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

allora possiamo dire che

$$P\left(-Z_{\frac{\alpha}{2}} \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le Z_{\frac{\alpha}{2}}\right) = 1-\alpha \Longrightarrow P\left(-\overline{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le -\mu \le -\overline{X}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\Longrightarrow P\left(\overline{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Abbiamo quindi dimostrato che

$$\mu \in \left[\overline{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \overline{X}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

Questo intervallo non dipende da $\overline{X}$ ma da σ.
Su σ però non possiamo operare, essendo data. Abbiamo 2 possibilità:

1. Possiamo operare su n, più n è grande più l'intervallo di confidenza è piccolo.
2. Possiamo operare anche su α. Se α cresce $Z_{\frac{\alpha}{2}}$ decresce, così diventa però più piccola la probabilità che μ stia nell'intervallo, ovvero $(1-\alpha)$, **la confidenza** appunto, quindi non può essere grandissimo.

In generale si parla di confidenza e non di probabilità perchè una volta fatto l'esperimento non si hanno più variabili casuali, quindi nemmeno probabilità. La probabilità si può avere solo prima dell'esperimento.
Consideriamo ora il caso in cui sia μ che σ^2 sono incognite.
Il procedimento di prima non può essere utilizzato in quanto non conosciamo σ. Allora stimiamolo col suo stimatore.

$$Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

che non è più una gaussiana ma una variabile di Student.
Ragioniamo su questa distribuzione. Se $T \sim t_{n-1}$ allora

$$P\left(-t_{n-1,\frac{\alpha}{2}} \le T \le t_{n-1,\frac{\alpha}{2}}\right) = 1-\alpha$$

Sostituendo otteniamo

$$P\left(-t_{n-1,\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \leq t_{n-1,\frac{\alpha}{2}}\right) = 1-\alpha \Longrightarrow P\left(\overline{X} - t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

Quindi

$$\mu \in \left[\overline{X} - t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}},\ \overline{X} + t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

Osservazione:
Se in un esperimento con σ conosciuta ottengo un certo valore $\overline{X}_1$ e se nel corrispondente esperimento con σ incognito riottengo $\overline{X}_1$ e $S = \sigma$, come sono gli intervalli di confidenza nei due casi?
Si dimostra graficamente che

$$t_{n-1,\frac{\alpha}{2}} > Z_{\frac{\alpha}{2}}$$

Quindi l'intervallo di confidenza in cui è presente la variabile di Student è più grande di quello in cui compare la normale standard. Tuttavia se $n >> 1$ la normale standard converge alla variabile di Student.
Cerchiamo ora un'intervallo di confidenza anche per σ^2, sempre nell'ipotesi che μ sia incognita.
Sappiamo che

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Ragioniamo quindi su questa distribuzione.
Se $C \sim \chi^2_{n-1}$ allora

$$P\left(C > \chi^2_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

E, poichè non è simmetrico il grafico della χ^2

$$P(C \leq ?) = \frac{\alpha}{2} \Longrightarrow P(C > ?) = 1 - \frac{\alpha}{2}$$

Quindi

$$P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} \leq C \leq \chi^2_{n-1,\frac{\alpha}{2}}\right) = 1-\alpha$$

Sostituendo otteniamo

$$P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,\frac{\alpha}{2}}\right) = 1-\alpha \Longrightarrow P\left(\frac{\chi^2_{n-1,1-\frac{\alpha}{2}}}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{n-1,\frac{\alpha}{2}}}{(n-1)S^2}\right) = 1-\alpha$$

$$\Longrightarrow P\left(\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

Abbiamo quindi trovato un intervallo per σ^2.

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}},\ \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right]$$

Per ridurre l'intervallo conviene aumentare n.

8.2.4 Metodo dei minimi quadrati lineare semplice

Il metodo dei minimi quadrati o *metodo della regressione* è una tecnica di ottimizzazione che permette di trovare una funzione che si avvicini il più possibile ad un'interpolazione di un insieme di dati.
Supponiamo di avere un sistema ad un ingresso X e un'uscita Y, legate dalla seguente relazione

$$Y = \beta X + \alpha + e$$

Questo è il tipico caso *lineare semplice*, in quanto la relazione è lineare e come ingresso abbiamo una sola variabile.
Introduciamo degli stimatori per β e α, rispettivamente B e A. Ci proponiamo di trovare la retta che meglio approssima i risultati di Y. Dobbiamo trovare $Y = BX + A$.
Ipotesi:

1. Le X_k non sono variabili casuali, ma valori fissati.
2. Le Y_k sono variabili casuali indipendenti tra loro.
3. Le Y_k sono variabili casuali gaussiane.

$$Y_k \sim N\left(\beta X_k + \alpha, \sigma^2\right)$$

Osservazione:

$$\overline{X} = \sum_{k=1}^{n} \frac{X_k}{n}, \qquad \overline{Y} = \sum_{k=1}^{n} \frac{Y_k}{n}$$

Queste due NON sono medie campionarie!!! Infatti le X_k non sono variabili casuali e le Y_k appartengono a popolazioni diverse.
Il nostro obiettivo è cercare i valori di B e A che minimizzano la somma di quadrati (abbreviata con **SS**: **S**quare **S**um).

$$SS = \sum_{k=1}^{n} (Y_k - (BX_k + A))^2$$

Si tratta dunque di risolvere questo sistema.

$$\begin{cases} \frac{\partial SS}{\partial A} = 0 \\ \\ \frac{\partial SS}{\partial B} = 0 \end{cases} \Longrightarrow \begin{cases} -2\sum_{k=1}^{n} (Y_k - BX_k - A) = 0 \\ \\ -2\sum_{k=1}^{n} X_k (Y_k - BX_k - A) = 0 \end{cases} \Longrightarrow \begin{cases} n\overline{Y} - Bn\overline{X} - nA = 0 \\ \\ \sum_{k=1}^{n} X_k Y_K - B\sum_{k=1}^{n} X_k^2 - An\overline{X} = 0 \end{cases}$$

$$\Longrightarrow \begin{cases} A = \overline{Y} - B\overline{X} \\ \\ \sum_{k=1}^{n} X_k Y_k - B\sum_{k=1}^{n} X_k^2 - n\overline{XY} + nB\overline{X}^2 \end{cases}$$

Usciamo dal sistema per ricavarci B.

$$B\left(\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\right) = \sum_{k=1}^{n} X_k Y_k - n\overline{XY} \Longrightarrow B = \frac{\sum_{k=1}^{n} Y_k \left(X_k - \overline{X}\right)}{X_k^2 - n\overline{X}^2}$$

B è una gaussiana essendo una trasformazione lineare di gaussiane. Stessa cosa per A.
Verifichiamo la correttezza di A e B.

$$\mathrm{E}[B] = \mathrm{E}\left[\frac{\sum_{k=1}^{n} Y_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2}\right] = \frac{\sum_{k=1}^{n} \left(X_k - \overline{X}\right)}{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2}\mathrm{E}\left[Y_k\right] = \frac{\sum_{k=1}^{n} Y_k \left(X_k - \overline{X}\right)}{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2}(\beta X_k + \alpha) =$$

$$= \beta \frac{\sum_{k=1}^{n} \left(X_k^2 - X_k \overline{X}\right)}{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2} + \alpha \frac{\sum_{k=1}^{n} \left(X_k - \overline{X}\right)}{\sum_{X_k - n\overline{X}^2}}$$

Calcoliamo a parte le sommatorie.

$$\sum_{k=1}^{n} \left(X_k - \overline{X}\right) = n\overline{X} - n\overline{X} = 0$$

$$\sum_{k=1}^{n} \left(X_k^2 - X_k \overline{X}\right) \sum_{k=1}^{n} X_k^2 - \overline{X} \sum_{k=1}^{n} X_k = \sum_{k=1}^{n} X_k^2 - n\overline{X}^2$$

Quindi

$$\mathrm{E}[B] = \beta \frac{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2}{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2} = \beta$$

Quindi B è corretto.
Verifichiamo la correttezza di A.

$$\mathrm{E}[A] = \mathrm{E}\left[\overline{Y} - B\overline{X}\right] = \mathrm{E}\left[\overline{Y}\right] - \overline{X}\mathrm{E}[B] = \mathrm{E}\left[\sum_{k=1}^{n} \frac{Y_k}{n}\right] - \beta\overline{X} =$$

$$= \frac{1}{n} \sum_{k=1}^{n} \mathrm{E}[Y_k] - \beta\overline{X} = \frac{1}{n} \sum_{k=1}^{n} (\beta X_k + \alpha) - \beta\overline{X} = \beta\overline{X} - \beta\overline{X} + \frac{\alpha n}{n} = \alpha$$

Quindi A è corretto.
Consideriamo ora le varianze dei due stimatori. Per semplicità definiamo $\gamma_k = \frac{\left(X_k - \overline{X}\right)}{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2}$.

$$\mathrm{Var}(B) = \mathrm{Var}\left(\sum_{k=1}^{n} \gamma_k Y_k\right) = \sum_{k=1}^{n} \mathrm{Var}(\gamma_k Y_k) = \sum_{k=1}^{n} \gamma^2 \mathrm{Var}(Y_k) = \sigma^2 \sum_{k=1}^{n} \gamma_k^2 =$$

Sostituendo otteniamo

$$= \sigma^2 \frac{\sum_{k=1}^{n} \left(X_k^2 + \overline{X}^2 - 2X_k\overline{X}\right)}{\left(\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\right)^2} = \frac{\sigma^2 \left(\sum_{k=1}^{n} X_k^2 + n\overline{X}^2 - 2\overline{X} \sum_{k=1}^{n} X_k\right)}{\left(\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\right)^2} =$$

$$= \frac{\sigma^2 \left(\sum_{k=1}^{n} X_k 2 - n\overline{X}^2\right)}{\left(\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\right)^2} = \frac{\sigma^2}{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2} = \frac{\sigma^2}{\sum_{k=1}^{n} \left(X_k - \overline{X}\right)^2}$$

Questo risultato indica che se i dati sono molto sparpagliati è più facile determinare B, poichè $\mathrm{Var}(B)$ è molto piccola. Diamo ora il valore della varianza di A.

$$\mathrm{Var}(A) = \frac{\sigma^2 \sum_{k=1}^{n} X_k^2}{n \sum_{k=1}^{n} \left(X_k - \overline{X}\right)^2}$$