

# **Elementi di Probabilità e Statistica:**

*Giulio Del Corso*

**Indice:**

- 3 Spazi di Misura
- 4 Introduzione e spazi discreti
- 6 Introduzione alle variabili aleatorie
- 7 Legge di Probabilità
- 8 Probabilità condizionale e indipendenza
- 10 Leggi discrete (Binomiale, Bernoulli, Geometrica, Poisson)
- 13 Ripasso serie numeriche
- 14 Probabilità su di uno spazio numerabile
- 15 Speranza
- 16 Varianza
- 17 Strumenti teorici utili
- 18 Teoremi limite per variabili di Bernoulli
- 19 Funzione di ripartizione della variabile  $N(0,1)$ :
- 20 Variabili aleatorie a più dimensioni
- 22 Formula della Convoluzione discreta
- 23 Funzione generatrice della Probabilità
  
- 24 Inferenza statistica su spazi di probabilità numerabili**
- 27 Stima
- 28 Stima corretta, consistente e di massima verosimiglianza. Modelli esponenziali
- 30 Rischio e Riassunto esaustivo
- 31 Regione di fiducia
- 32 Test statistici
- 33 Test ad ipotesi semplice
- 34 Modello a rapporto di verosimiglianza crescente
  
- 36 Probabilità generale**
- 38 Applicazioni misurabili e funzioni semplici
- 40 Densità di probabilità
- 42 Variabili aleatorie reali su spazi generali
- 44 Variabili aleatorie con densità
- 46 Esempi di densità (Uniforme, gamma, gaussiana, chi-quadro, Student, Fisher)
- 49 Convergenza in probabilità
  
- 50 Inferenza statistica su di uno spazio di probabilità generale**
- 52 Inferenza statistica su modelli gaussiani
- 56 Test sui campioni gaussiani
- 60 Confronti fra campioni gaussiani indipendenti
- 62 Modelli statistici lineari

### Spazi di misura:

Gli spazi di probabilità sono una particolare famiglia di spazi di misura definiti come una terna  $(E, \varepsilon, \mu)$  con  $E$  insieme non vuoto,  $\varepsilon$  tribù (i cui elementi sono detti **insiemi misurabili**) e  $\mu: \varepsilon \rightarrow [0, +\infty]$   $\sigma$ -additiva una misura.

#### **Notazione:**

$\mu(E) < \infty$  Misura finita

$\mu(E) = 1$  Misura normalizzata (fra cui la Misura di Probabilità)

$\mu(x) = 0, x \in \varepsilon$  si dice trascurabile

La misura si dice concentrata su  $H$  se  $H^c$  è trascurabile

#### **Proprietà di una misura:**

$\mu(B) = \mu(A) + \mu(B \setminus A)$  se  $A \subseteq B$

$\mu(A) \leq \mu(B)$  se  $A \subseteq B$  (**Isotonia**)

$\mu(A \cap B) + \mu(A \cup B) = \mu(A) + \mu(B)$  (**Modularità**)

Sia  $(A_n)_{n \geq 1}$  successione con  $A_n \in \varepsilon \mid A_n \subseteq A_{n+1} \forall n$ ; se  $A = \bigcup A_n$  allora:

$\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$  (**Continuità**)

#### **Proprietà delle Tribù:**

Intersezione di tribù è una tribù.

La **tribù generata da  $H$**  è l'intersezione di tutte le tribù su  $E$  contenenti  $H$ .

Dato  $(E, \varepsilon)$  se  $H \subseteq \varepsilon$  e la tribù generata da  $H$  coincide con  $\varepsilon$  si dice che  $H$  è un sistema di generatori.

Un sistema di generatori si dice **base** di  $\varepsilon$  se  $\begin{cases} \forall A, B \in H \quad A \cap B \in H \\ E = \bigcup A_i \mid A_i \in H \end{cases}$

#### **Esempio:**

$(E, P(E))$  allora  $H = \{x\} \cup \emptyset$  è una base di  $P(E)$

#### **Criterio di coincidenza di due misure:**

$u, v$  misure su  $(E, \varepsilon)$ ;  $H$  base di  $\varepsilon$ , se  $u(A) = v(A) \forall A \in H \rightarrow$  Le due misure coincidono.

#### **Esempio:**

Questo è il motivo per il quale se assegno la probabilità dell'uscita dei singoletti al gioco del lotto ho una sola probabilità che la rispetti.

### Applicazioni misurabili:

Dati  $(E_1, \varepsilon_1); (E_2, \varepsilon_2)$   $f: E_1 \rightarrow E_2$  si dice misurabile se  $\forall A \in \varepsilon_2 \quad f^{-1}(A) \in \varepsilon_1$  Parallelo a v.a.

#### **Proprietà:**

Composizione di applicazioni misurabili è misurabile.

Data una misura  $\lambda$  su  $\varepsilon_1$  e una  $f$  misurabile allora  $\mu(A) = \lambda(f^{-1}(A))$  è una misura su  $\varepsilon_2$  detta misura immagine e indicata con  $f_\lambda$  (Parallelo a quello di Legge di probabilità).

### **Introduzione:**

Quando ci apprestiamo a studiare un esperimento aleatorio dobbiamo contestualizzarlo precisando gli oggetti con cui stiamo lavorando.

Per prima cosa costruiamo un insieme ambiente  $\Omega$  detto delle eventualità o degli esiti.

Detto questo individuiamo una famiglia  $F$  di sottoinsiemi di  $\Omega$  (Detti **eventi**) che rappresentano i casi che ci interessa studiare. Non sempre prendiamo  $F = P(\Omega)$  in quanto non sempre siamo interessati ad ogni evento.

Vogliamo che questo insieme sia stabile, deve quindi rispettare ( **$\sigma$ -algebra o tribù**):

1.  $\Omega \in F$
2. Se  $A \in F \rightarrow A^c \in F$
3. Se  $\forall n, A_n \in F \rightarrow \bigcup_{n=1}^{\infty} A_n \in F$

### **Notazione:**

$(\Omega, F)$  è detto **Spazio Probabilizzante**

#### **Esempio:**

*Dato un lancio di un dado a 6 facce il mio insieme ambiente sarà  $\Omega = \{1,2,3,4,5,6\}$  e una scelta di una tribù potrebbe essere  $F = P(\Omega)$ .*

Su di uno spazio probabilizzante vogliamo misurare il grado di fiducia di un dato evento, una **Misura di Probabilità** (o misura normalizzata) è:

$$P: F \rightarrow [0,1] \mid P(\Omega) = 1$$

Se gli  $A_n \in F$  sono due a due disgiunti allora:  $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$  ( **$\sigma$ -additività**)

#### **Proprietà:**

$$P(\emptyset) = 0$$

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

$$\text{Se } B \subseteq A \rightarrow P(A \setminus B) = P(A) - P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \text{ Generale con principio di Inclusione-Esclusione}$$

$$B \setminus A = B \cap A^c$$

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

**Notazione:**  $(\Omega, F, P)$  è detto **Spazio di Probabilità**

#### **Esempio:**

*Nonostante  $P$  possa essere scelta in maniera arbitraria spesso se definisco la  $P$  su di una sottofamiglia il resto diventa vincolato. Ad esempio se assegno ad ogni numero del lotto probabilità  $\frac{1}{90}$  allora la probabilità di ogni singolo evento segue in maniera automatica.*

### **Notazione:**

$w$  realizza l'evento  $B$  significa che  $w \in \Omega \wedge w \in B$  con  $B \in F$  su  $(\Omega, F)$

Evento  $C$  o  $B$  significa l'evento  $C \cup B \in F$

Evento  $C$  e  $B$  significa l'evento  $C \cap B \in F$

Quasi impossibile  $\sim$  Trascurabile e Quasi certo  $\sim P(A) = 1$

### **Spazio discreto:**

$(\Omega, F)$  spazio probabilizzante si dice discreto se  $\begin{cases} \Omega \text{ numerabile} \\ F = P(\Omega) \end{cases}$

### **Osservazione:**

Nel caso di spazi discreti possiamo definire una  $p: \Omega \rightarrow [0,1] \mid \sum_{w \in \Omega} p(w) = 1$  detta **densità discreta** (di probabilità) strettamente correlata alla Misura di probabilità definita su  $F$ .

Data la densità discreta definiamo  $P \mid P(A) = \sum_{w \in A} p(w) \quad \forall A \in F$

Data  $P$  su  $F$  la densità discreta associata sarà  $p: \Omega \rightarrow [0,1] \mid p(w) = P(\{w\})$

### **Esempio:**

*Avendo un dado truccato sia  $\Omega = \{1,2,3,4,5,6\}$ , definiamo una densità discreta su  $\Omega$  data da  $p(i) = 0 \quad \forall i \neq 6$  e  $p(6) = 1$ , questa induce naturalmente una  $P$  su  $P(\Omega) \mid$  ogni evento contenente il 6 è quasi certo mentre ogni evento non contenente il 6 è quasi impossibile.*

### **Osservazione:**

Se due densità discrete di probabilità sono uguali a meno di una costante allora coincidono.

### **Ripartizione uniforme:**

La ripartizione uniforme su  $\Omega \neq \emptyset$  e finito è l'unica misura di probabilità su  $P(\Omega) \mid$  tutti i singoletti hanno lo stesso valore. In questo caso  $P(A) = \frac{|A|}{|\Omega|} \quad \forall A \in P(\Omega)$

### **Formule fondamentali di probabilità:**

Il numero di applicazioni da un insieme di cardinalità  $k$  in uno di cardinalità  $n$  è  $n^k$

Il numero di modi in cui si possono ordinare gli elementi di un insieme di cardinalità  $n$  è  $n!$

Il numero di sottoinsiemi di  $k$  elementi di un insieme di  $n$  elementi è  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

### **Introduzione alle variabili aleatorie:**

Quando lavoriamo con le probabilità spesso ci poniamo delle domande relative in maniera non diretta all'insieme sui cui è definita la probabilità.

Vogliamo quindi spostare il nostro studio su uno spazio di misura (Ad esempio  $\mathbb{R}$ ).

#### ***Esempio:***

*Lancio di un dado con vari guadagni a seconda del numero, probabilità che io vinca tot. euro?*

### **Definizione generale (Variabile aleatoria):**

Una variabile aleatoria è un'applicazione misurabile  $X: (\Omega, F) \rightarrow (E, \varepsilon)$

In pratica ad ogni eventualità  $w \in \Omega$  associo  $X(w) \in E$  sotto la condizione che la tribù  $F$  contenga le immagini inverse di  $\varepsilon$ .

#### **Notazione:**

Dato  $B \in \varepsilon$  allora  $X^{-1}(B) = \{w \in \Omega \mid X(w) \in B\} := \{X \in B\}$  che si legge "X cade in B"

L'insieme degli  $X^{-1}(B)$  forma una tribù indotta da  $X$  su  $\Omega$ , denotata con  $T(X) = T(X, \varepsilon)$ , detta degli elementi esprimibili mediante  $X$ .

#### **Osservazione (Variabile aleatoria discreta reale):**

Una variabile aleatoria reale discreta è una  $X: \Omega \rightarrow \mathbb{R}$  v.a. |  $\Omega$  numerabile.

#### **Osservazione:**

D'ora in avanti lavoreremo quasi esclusivamente con variabili aleatorie reali (Nella prima parte discrete).

#### **Notazione pratica:**

Date due variabili aleatorie  $X, Y$  indichiamo con:

$$(aX + b)(x) = aX(x) + b$$

$$(X + Y)(x) = X(x) + Y(x)$$

$$(XY)(x) = X(x)Y(x)$$

**Definizione (Legge di probabilità):**

Si chiama legge o distribuzione di probabilità della variabile aleatoria reale  $X$  la probabilità definita sui sottoinsiemi di  $\mathbb{R}$  mediante la formula:  $P_X(A) = P(X^{-1}(A))$

$P_X$  è anche detta **probabilità immagine** e indicata con  $X(P)$

Lavorando con variabili aleatorie discrete  $X$  possiamo definire una **densità discreta** nell'insieme immagine come  $p(r) = P\{X = r\}$

**Notazione:**

$$\{X \in A\} = \{w_i \mid X(w_i) \in A\} = X^{-1}(A)$$

**Esempio:**

Assegnando ai lanci di un dado il valore in euro del numero uscito la probabilità di vincere 5 euro diventa  $P_X(5) = P\{X = 5\} = P(X^{-1}(5)) = \frac{1}{6}$

**Osservazione (Esistenza di una variabile aleatoria associata ad una probabilità):**

Assegnata una probabilità discreta  $Q$  su  $\mathbb{R}$  ossia fissati  $(x_1, x_2, \dots)$  e  $(p(x_1), p(x_2), \dots) \mid \sum_i p(x_i) = 1$  esiste sempre una variabile aleatoria  $X$  la cui legge di probabilità sia  $Q$ .

Scegliamo  $\Omega = \{x_1, x_2, \dots\}$ ;  $P \mid P(\{x_i\}) = p(x_i)$ ;  $X: \Omega \rightarrow \mathbb{R}$  l'applicazione identica.

### Probabilità condizionale:

Dato uno spazio di probabilità  $(\Omega, F, P)$  supponiamo realizzato un evento  $H$  di probabilità non trascurabile.

Vogliamo ridefinire lo spazio di probabilità alla luce di questa nuova informazione sostituendolo con  $(\Omega, F, P_H) \mid P_H(B) := P(B|H) = P(H)^{-1}P(B \cap H) \quad \forall B \in F$

#### **Condizionamento ripetuto:**

$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$  con  $A_1 \cap \dots \cap A_{n-1}$  non trascurabile.

### Definizione (Sistema di alternative):

È un insieme  $B_1, \dots, B_n$  di eventi non trascurabili che forma una partizione (Finita/numerabile) di  $\Omega$ .

#### **Proprietà (D1):**

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

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)} \quad \text{Formula di Bayes}$$

### Idea Indipendenza:

Ogni variabile aleatoria fornisce un'informazione su di un evento, possiamo quindi definire diverse variabili aleatorie sullo stesso spazio di probabilità.

Queste variabili possono a loro volta essere totalmente indipendenti l'una dall'altra oppure correlate.

#### **Esempio indipendenti:**

*Gioco a dadi fra due amici,  $X_1$  che associa alla partita 1  $\{1,2\}$  a seconda che abbia vinto il primo o il secondo e  $X_2$  che restituisce invece la vincita della 2° partita.*

#### **Esempio dipendenti:**

*$X_3$  che nello stesso gioco di prima associa  $\{0,1,2\}$  a seconda che abbia vinto più partite (Su due lanci) nessuno, il primo o il secondo. Questa è ovviamente dipendente da  $X_1$  e da  $X_2$ .*

### Indipendenza fra variabili aleatorie (Definizione formale):

Dato uno spazio di probabilità  $(\Omega, F, P)$  e due variabili aleatorie  $X, Y$  a valori in due arbitrari spazi di misura si dice che sono indipendenti se  $\forall H$  della tribù indotta da  $X$  vale  $Y(P) = Y(P_H)$ .

#### **Caso particolare (Indipendenza fra eventi di una tribù):**

Dato uno spazio di probabilità  $(\Omega, F, P)$  allora  $H, K \in F$  si dicono indipendenti se:

$$P(H \cap K) = P(H)P(K)$$

#### **Proprietà:**

1.  $A, B$  indipendenti  $\rightarrow A^C, B$  indipendenti,  $A, B^C$  indipendenti,  $A^C, B^C$  indipendenti.
2. Se  $P(A) = \begin{cases} 0 \\ 1 \end{cases} \rightarrow A$  è indipendente da ogni altro  $B$ .
3. Due eventi incompatibili ( $A \cap B = \emptyset$ ) sono dipendenti a meno che  $P(A) = 0$  o  $P(B) = 0$

**Definizione (Stocasticamente indipendenti):**

$A_1, \dots, A_n$  sono detti stocasticamente indipendenti se  $\forall k, 2 \leq k \leq n$  e  
 $\forall$  scelta  $1 \leq i_1 < i_2 < \dots < i_k \leq n$  si ha:  $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$

**Osservazione:**

Degli eventi possono essere a due a due indipendenti ma non "globalmente" indipendenti.

**Esempio:**

$\Omega = \{1,2,3,4\}$  con probabilità uniforme gli eventi  $A = \{1,2\}$ ;  $B = \{1,3\}$ ;  $C = \{2,3\}$  sono a due a due indipendenti ma non globalmente indipendenti.

**Osservazione Importante:**

Dalla definizione di indipendenza fra eventi si può ricavare una formulazione equivalente di indipendenza fra variabili aleatorie (reali):

$X_1, \dots, X_n$  v.a. reali su  $(\Omega, F, P)$  sono indipendenti  $\leftrightarrow \forall A_1, \dots, A_n \in \mathbb{R}$  si ha  
 $P\{X_1 \in A_1, \dots, X_n \in A_n\} = P\{X_1 \in A_1\} \dots P\{X_n \in A_n\}$

**Caso particolare (Indipendenza fra tribù):**

Due tribù si dicono indipendenti se ogni elemento dell'una è indipendente da ogni elemento dell'altra.

**Definizione (Indipendenti variabili aleatorie reali):**

Due variabili aleatorie  $X, Y$  si dicono indipendenti se scelti due sottoinsiemi  $A, B$  di  $\mathbb{R}$  gli eventi  $X^{-1}(A)$ ;  $Y^{-1}(B)$  sono indipendenti.

Vale la formula:

$$P\{X \in A, Y \in B\} = P\{X \in A\} P\{Y \in B\}$$

**Notazione:**

$$P\{X \in A, Y \in B\} := P\{X^{-1}(A)\} \cap P\{Y^{-1}(B)\}$$

$$P\{X \in A\} := P\{X^{-1}(A)\}$$

**Osservazione (D2):**

Due variabili aleatorie discrete  $X, Y$  sono indipendenti  $\leftrightarrow$  vale la formula

$$p(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j)$$

**Definizione generalizzata:**

Data famiglia di variabili aleatorie queste si dicono indipendenti se ogni sottofamiglia finita è formata da variabili indipendenti.

### **Leggi discrete:**

Come oggetti stiamo lavorando con uno spazio di probabilità, ed uno di misura dotato come insieme dei suoi misurabili quelle delle sue parti, tipicamente  $\mathbb{R}$ .

Definiremo le principali variabili aleatorie studiandone le probabilità così da avere degli esempi a cui rifarci nel corso degli studi.

#### **Osservazione (legge fondamentale):**

Data  $X$  v.a. discreta su uno spazio di misura  $E$  se  $\mu$  è la sua legge allora si dice legge fondamentale  $\mu$  ristretta alle parti di  $E$  di misura non nulla. Tipicamente per legge si intende legge fondamentale.

#### **Osservazione (Importante/Formale):**

Ai fini di comodità di lettura ho inserito un dominio ad esempio in ogni definizione (Nella binomiale  $\{0,1\}^n$ , etc.) dal punto di vista formale però queste variabili aleatorie è importante ricordarsi che vanno da uno spazio di probabilità discreto qualsiasi nell'immagine.

### **Variabile binomiale:**

Una variabile binomiale di parametri  $n \in \mathbb{N}$  e  $p \in ]0,1[$  è una funzione  $X: \{0,1\}^n \rightarrow \{0,1, \dots, n\}$  che considera  $n$  ripetizioni di un esperimento aleatorio con probabilità  $p$  di successo e conta il numero di successi ottenuti.

#### **In pratica:**

Definito come  $\Omega$  l'insieme delle stringhe ordinate di 0 e 1 di lunghezza  $n$  e come probabilità

$$P(\text{stringa}) = p^{\text{numero di } 1}(1-p)^{\text{numero di } 0}$$

$X(\text{stringa}) = \#1$  nella stringa e  $P_X(k) = P\{X = k\}$  è la probabilità che fra tutte le stringhe possibili si ottenga proprio una di quelle con  $k$  volte il numero 1.

La **legge binomiale** è la legge indotta da una variabile binomiale, viene indicata spesso con  $B(n, p)$

Per queste leggi vale:  $P_X(k) = P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$

Per le leggi binomiali  $B(1, p)$  si usa il nome di **Legge di Bernoulli** di parametro  $p$ .

#### **Osservazione carina:**

Una variabile binomiale di parametro  $n$  può essere vista come somma di  $n$  variabili di Bernoulli.

Speranza di una variabile di Bernoulli  $E[X] = p$ ;  $Var(X) = p(1-p)$

Speranza di una variabile binomiale  $E[X] = np$ ;  $Var(X) = np(1-p)$

#### **Esempio:**

*Tipico esempio di uso di una legge binomiale è l'esercizio "Lancio di un dado  $n$  volte, determinare la probabilità che escano esattamente due 6".*

*Ci limitiamo a considerare il caso 6 come "giusto" e il caso "non 6" come sbagliato.*

*Diventa quindi una legge binomiale  $B\left(n, \frac{1}{6}\right)$  e dunque  $B\left(n, \frac{1}{6}\right)(2) = \binom{n}{2} p^2 (1-p)^{n-2}$*

### **Variabile geometrica:**

Una variabile geometrica di parametro  $p \in ]0,1[$  è una funzione  $X$  che considera il numero di volte che è dovuto essere reiterato l'esperimento con probabilità di riuscita  $p$  affinché si ottenga un esito positivo.  $X$  (stringa di 0 terminanti con un 1) = Lunghezza della stringa

La **Legge geometrica** associata ad una variabile geometrica è la probabilità che io ottenga proprio una stringa di lunghezza  $k$  formata da 0 con in fondo un 1.

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

$$\text{Ovviamente per } n, h \in \mathbb{N} \text{ vale } P\{X = n + h \mid X > n\} = P\{X = h\}$$

$$\text{Speranza di una variabile geometrica } E[X] = \frac{1}{p}; E[X^2] = \frac{1+(1-p)}{p^2}; \text{Var}(X) = \frac{1-p}{p^2}$$

### **Osservazione:**

$$P\{X > k\} = (1 - p)^k; P\{X \leq k\} = 1 - (1 - p)^k$$

### **Osservazione (Caratterizzazione alternativa):**

Una variabile geometrica è caratterizzabile come una variabile aleatoria a valori positivi

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

### **Osservazione generale:**

Quando ho assegnato il dominio a queste variabili aleatorie l'ho fatto per fornire un esempio comprensibile. In realtà queste sono definite dalla tribù degli eventi in sottoinsiemi di  $\mathbb{R}$  e quali siano questi eventi è spesso secondario conoscendone la legge di probabilità su  $\mathbb{R}$ .

### **Densità o legge di Poisson:**

Definire una variabile di Poisson non è una cosa facile quindi ci limitiamo a descrivere la probabilità che essa induce su  $\mathbb{R}$  (Ossia la sua legge indotta che, essendo la probabilità concentrata su  $\mathbb{N}$ , sarà discreta).

Si indica con  $P(\lambda)$  la distribuzione di Poisson, ossia la misura di probabilità indotta su  $\mathbb{R}$  (concentrata su di un insieme numerabile) che associa ad ogni  $n$ :

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

#### **Esempio 1:**

*Dato un centralino telefonico con un numero di guasti che si distribuiscono secondo la legge di Poisson con  $\lambda = 3$  quale è la probabilità che in un anno si verifichino esattamente tre guasti?*

$$P\{X = 3\} = P_3(3) = e^{-3} \frac{3^3}{3!} = 0,224$$

### **Costruzione di una variabile di Poisson:**

Consideriamo una legge binomiale di parametro  $\frac{\lambda}{n}$  |  $0 < \lambda < n$  e  $n$  il numero di prove.

La probabilità di un dato evento sarà  $P\{X = x\} = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \rightarrow_{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^x}{x!}$

#### **Quindi:**

Se stiamo lavorando con numeri molto alti di prove di probabilità molto bassa allora possiamo approssimare la distribuzione Binomiale con quella di Poisson.

Approssimazione discreta:	$n \geq 20$	$\frac{\lambda}{n} \leq 0.05$
Approssimazione ottima:	$n \geq 100$	$\frac{\lambda}{n} \leq \frac{10}{n}$

#### **Esempio 2:**

*Un telefono può squillare in ogni istante ma sappiamo che in media squilla 5 volte in un ora. Se vogliamo lavorare sul discreto possiamo suddividere l'ora in 60 parti e considerare che la probabilità che squilli proprio in uno di quegli intervalli è di  $\frac{5}{60}$ .*

*La probabilità che squilli sarà data da legge binomiale  $B\left(60, \frac{5}{60}\right)$*

*Se invece avessi considerato la suddivisione in secondi avremmo ottenuto la legge  $B\left(3600, \frac{5}{3600}\right)$*

*Portando al limite la suddivisione (O accettando un'approssimazione minima) otteniamo la legge di Poisson di parametro  $\lambda$ .*

### **Proprietà:**

Date due variabili aleatorie indipendenti  $X, Y$  di leggi  $P(\lambda_1)$ ;  $P(\lambda_2)$  allora  $X + Y$  avrà legge  $P(\lambda_1 + \lambda_2)$

Data una variabile aleatoria  $X$  con distribuzione di probabilità coincidente con una distribuzione di probabilità di Poisson di parametro  $\lambda$  (O legge di Poisson).

Speranza:  $E[X] = \lambda$

Varianza:  $\text{Var}(X) = \lambda$

### Ripasso serie numeriche:

Data una successione di numeri reali  $(a_i)_{i \in \mathbb{N}}$  si chiama **somma della serie** il limite (Se esiste) delle somme parziali.

La serie converge se questo limite esiste.  $\sum_{k=1}^{\infty} a_k = \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k = \lim_{n \rightarrow \infty} S_n$   
Se la serie converge la successione  $(a_n)$  è infinitesima.

### Serie a termini positivi.

Essendo monotona crescente il limite esiste  $\sum_{k=1}^{\infty} a_k \in [0, +\infty]$

**Assoluta convergenza:** Una serie si dice assolutamente convergente se  $\sum_{n=1}^{\infty} |a_n| < +\infty$

Se una serie è assolutamente convergente valgono i seguenti risultati.

#### **Cambiare l'ordine dei termini:**

Data  $v: \mathbb{N} \rightarrow \mathbb{N}$  biunivoca allora:  $\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} v(a_k)$

#### **Sommare a pacchetti:**

Sia  $A_1, A_2, \dots$  una partizione di  $\mathbb{N}$  allora vale:  $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \sum_{k \in A_n} a_k$

### Criteri di convergenza per serie a termini positivi:

#### **Criterio della radice (Cauchy):**

Se  $a_n \geq 0$  e se  $\exists \lim_{n \rightarrow \infty} \sqrt[n]{a_n} = L < 1$  allora  $\sum a_n$  converge.

#### **Criterio del rapporto (D'Alembert):**

Se  $a_n \geq 0$  e se  $\exists \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = L < 1$  allora  $\sum a_n$  converge.

#### **Criterio del confronto Asintotico:**

$a_n, b_n > 0$  Se  $\exists \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = L$  allora:

- Se  $0 < L < \infty$  le serie  $\sum a_n$  e  $\sum b_n$  hanno lo stesso comportamento.
- Se  $L = 0$  e  $\sum b_n$  converge (Non)  $\rightarrow \sum a_n$  converge (Non)

Se  $L = \infty$  e  $\sum a_n$  converge  $\rightarrow \sum b_n$  converge

### Esempi di serie importanti:

**Serie telescopica:** Es.  $\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1$  infatti:  $= \sum_{k=1}^n \left( \frac{1}{k} - \frac{1}{k+1} \right) = 1 - \frac{1}{n+1} \rightarrow 1$  per  $n \rightarrow \infty$

**Serie Geometrica:** Sia  $q \in \mathbb{C}$ . Se  $|q| < 1$  allora:  $\sum_{k=0}^{\infty} \frac{1}{1-q}$

**Serie Armonica:**  $E = \left\{ \sum_{i=1}^n \frac{1}{i}; n \in \mathbb{N}^+ \right\}$ ; non è limitata superiormente.

### Sviluppi in serie di Taylor:

$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ ;  $\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$ ;  $\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$

$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}$

### Serie utili:

$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$  per  $x \in ]0,1[$ ;  $\sum_{n=0}^k x^n = \frac{x^{k+1}-1}{x-1}$ ;  $\sum_{n=0}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}$ ;  $\sum_{n=0}^k n^2 = \frac{k(k+1)(2k+1)}{6}$

Scomposizione:  $\sum_{n=12}^{\infty} a_n = \sum_{n=0}^{\infty} a_n - \sum_{n=0}^{11} a_n$ ;  $\frac{1}{(1+x)^{m+1}} = \sum_{n=0}^{\infty} \binom{n+m}{m} x^n, x \in ]-1,1[$

### Probabilità su uno spazio numerabile:

In questa parte svilupperemo la teoria rifacendoci sempre a spazi numerabili (Dunque a probabilità discrete) ma manterremo la notazione con gli integrali per poter dopo estendere queste proprietà a generici spazi di probabilità.

### Integrale rispetto ad una misura discreta:

Dato un insieme numerabile  $E = \{e_1, \dots, e_n\}$  sul quale sia definita una misura  $m$  e supponiamo che

$$\forall i \ m(e_i) := m(\{e_i\}) < +\infty$$

$$\text{Abbiamo che } \forall A \subseteq E \text{ vale } m(A) = \sum_{e_i \in A} m(e_i)$$

### Definizione (Integrale):

Data  $f: E \rightarrow \mathbb{R}$  si dice integrabile se  $\sum_i |f(e_i)|m(e_i) < +\infty$

L'integrale di  $f$  è il numero  $\int(f)dm = \sum_i f(e_i)m(e_i)$

#### **Proprietà:**

$L^1$  è lo spazio delle funzioni integrabili.

Se  $f, g \in L^1 \rightarrow (af + g) \in L^1$  e  $\int(af + g)dm = a \int(f)dm + \int(g)dm$

Se  $0 \leq f \leq g \rightarrow \int(f)dm \leq \int(g)dm$

Se  $0 \leq f \wedge \int(f)dm = 0 \rightarrow f$  vale identicamente 0 eccetto su di un insieme trascurabile.

### **Teorema di Beppo-Levi (Passaggio al limite) (D3):**

$$0 \leq f_n ; f_n \uparrow f \rightarrow \int(f_n)dm \uparrow \int(f)dm$$

### **Teorema di convergenza dominata (D4):**

Sia  $(f_n)_{n \geq 1}$  successione di funzioni convergente puntualmente ad  $f$  e supponiamo che  $\exists g \geq 0$  integrabile  $|f_n| \leq g \ \forall n$ .

$$\text{Vale allora: } \lim_{n \rightarrow \infty} \int(f_n)dm = \int(f)dm$$

### **Disuguaglianza di Schwarz (D5):**

$$|\int(fg)dm| \leq \sqrt{\int(f^2)dm} \sqrt{\int(g^2)dm}$$

Se è un'uguaglianza le funzioni  $f, g$  coincidono a meno di una costante moltiplicativa.

### Osservazione interessante:

Tutte le proprietà fin qui incontrate valgono anche nel caso in cui l'insieme  $E$  sia non numerabile ma la misura sia concentrata su di un insieme numerabile. (Si dice misura **discreta**).

### **Teorema (Integrazione rispetto ad una probabilità immagine) (D6):**

Siano  $X$  una variabile aleatoria discreta (reale),  $P_X$  la sua legge di probabilità e  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ .

Allora  $\varphi$  integrabile rispetto a  $P_X \leftrightarrow \varphi \circ X$  è integrabile rispetto a  $P$

$$\text{In tal caso vale: } \int_{\mathbb{R}}(\varphi(x))dP_X(x) = \int_{\Omega}(\varphi(X(w)))dP(w)$$

### **Definizione (Valore atteso/medio o Speranza):**

Data una variabile aleatoria discreta  $X$  si dice che ha speranza se è integrabile rispetto a  $P$ .

In tal caso la speranza è:  $E[X] = \int_{\Omega} \varphi(X(\omega)) dP(\omega) = \sum_i X(\omega_i) P(\omega_i)$

#### **Esempio:**

Consideriamo il caso di lancio di un dado; nel caso in cui esca un numero pari si vince un numero di euro pari al valore indicato sul dado, nel caso invece sia dispari si perde quel valore.

La variabile aleatoria con cui stiamo lavorando è  $X: \Omega = \{1,2,3,4,5,6\} \rightarrow [-5,6]$

Con  $X(i) = \begin{cases} i & \text{se } i \equiv 0 \pmod{2} \\ -i & \text{se } i \equiv 1 \pmod{2} \end{cases}$  allora la speranza  $E[X] = -\frac{1}{6} + \frac{2}{6} - \frac{3}{6} + \frac{4}{6} - \frac{5}{6} + \frac{6}{6} = \frac{1}{2}$

In casi come questi la speranza viene anche detta guadagno medio e sottolinea il fatto che il gioco non sia equilibrato a favore di chi gioca ( $E$  non del banco).

#### **Osservazione (Lineare):**

La speranza è lineare, ossia:  $E[aX + b] = aE[X] + b$

#### **Osservazione:**

Vale la formula:  $E[X] = \sum_{n \geq 0} P\{X > n\} = \sum_{n \geq 1} P\{X \geq n\}$

#### **Osservazioni pratica:**

Se  $X, Y$  sono due variabili aleatorie integrabili allora:

$$E[X + Y] = E[X] + E[Y]$$

Se  $X, Y$  sono due variabili aleatorie indipendenti e dotate di momento primo allora  $XY$  ammette momento primo e:

$$E[XY] = E[X] \cdot E[Y]$$

### **Definizione (Momento):**

Sia  $1 \leq p < +\infty$  e  $X$  una variabile aleatoria. Si chiama momento assoluto di ordine  $p$  il numero:

$$E[|X|^p] = \sum_i |x_i|^p p(x_i) \in [0, +\infty]$$

Se questo numero è finito si dice che  $X$  ammette momento di ordine  $p$ .

Il momento di ordine  $p$  è il numero  $E[X^p]$

#### **Proposizione (D7):**

Sia  $1 \leq p < q < +\infty$ , se  $X$  ha momento di ordine  $q$ , ammette momento di ordine  $p$ .

**Definizione (Varianza):**

Data  $X$  variabile aleatoria dotata di momento secondo, la varianza di  $X$  è il numero:

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

**Esempio:**

Osserviamo che ogni allontanamento dal valore medio/speranza porta ad un aumento della varianza, considerando il caso precedente

$$\text{Var}(X) = E\left[\left(X - \frac{3}{6}\right)^2\right] = E[X^2] - \left(\frac{3}{6}\right)^2 = \frac{1}{36} + \frac{4}{36} + \frac{9}{36} + \frac{16}{36} + \frac{25}{36} + \frac{36}{36} - \frac{9}{36} = \frac{41}{18}$$

Una variabile  $Y: \Omega = \{1,2,3,4,5,6\} \rightarrow [-5,6] \mid Y(i) = \frac{3}{6}$  apparentemente simile in quanto ha lo stesso guadagno medio presenta invece una varianza nulla.

**Osservazione (Relazione utile):**

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

**Osservazione:**

La formula utilizzata per calcolare la varianza deriva dalla linearità della speranza:

$$E[aX + b] = \sum_{i=1}^{\infty} (ax_i + b)P\{X = x_i\} = a \sum_{i=1}^{\infty} x_i P\{X = x_i\} + \sum_{i=1}^{\infty} bP\{X = x_i\} = aE[X] + b$$

Dunque:

$$E[(X - E[X])^2] = E[X^2 - 2E[X]X + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

### **Strumenti utili di teoria:**

#### **Disuguaglianza di Markov (D8):**

Sia  $X$  una variabile aleatoria a valori positivi e  $t$  una costante positiva, vale allora la disuguaglianza:

$$t \cdot P\{X \geq t\} \leq E[X]$$

#### **Disuguaglianza di Chebishev (D9):**

Sia  $X$  una variabile aleatoria dotata di momento seconda, vale la disuguaglianza:

$$t^2 \cdot P\{|X - E[X]| \geq t\} \leq \text{Var}(X)$$

#### **Corollario (D10):**

$\text{Var}(X) = 0 \leftrightarrow X$  è costante q.o.

### Teoremi limite per variabili di Bernoulli (Moivre-Laplace):

Date  $X_1, \dots, X_n$  variabili di Bernoulli indipendenti di parametro  $p$  e  $S_n = X_1 + \dots + X_n$

#### **Legge dei grandi numeri per variabili Binomiali (D11):**

$$\forall \varepsilon > 0 \text{ vale } \lim_{n \rightarrow \infty} P \left\{ \left| \frac{S_n}{n} - p \right| > \varepsilon \right\} = 0$$

#### **Osservazione:**

Le ipotesi possono essere indebolite  $X_1, \dots, X_n$  possono essere semplicemente indipendenti, equidistribuite, dotate di momento secondo e varianza positiva. A questo punto basta sostituire  $p$  con  $E[X_i]$ .

### Teorema (Limite centrale per Variabili Binomiali) (D12):

Presi due numeri  $-\infty \leq a < b \leq +\infty$  si ha:

$$\lim_{n \rightarrow \infty} P \left\{ a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

#### **Osservazione:**

Il vero uso di questo teorema è per approssimare conti complessi su variabili binomiali grandi. Il calcolo esplicito dell'integrale non è richiesto in quanto esistono delle tavole per un rapido calcolo approssimato.

#### **Esempio:**

Sia  $X \sim B(400, 0.05)$  vogliamo calcolare  $P\{X > 30\}$  sommiamo a sinistra e a destra i valori necessari a portarci nelle ipotesi del teorema:

$$P\{X > 30\} = P \left\{ \frac{X-20}{\sqrt{400 \cdot 0.05 \cdot 0.95}} > \frac{30-20}{\sqrt{400 \cdot 0.05 \cdot 0.95}} \right\} = 1 - P \left\{ \frac{X-20}{\sqrt{400 \cdot 0.05 \cdot 0.95}} \leq 2.29 \right\} \text{ che si può approssimare con } 1 - \Phi(2.29)$$

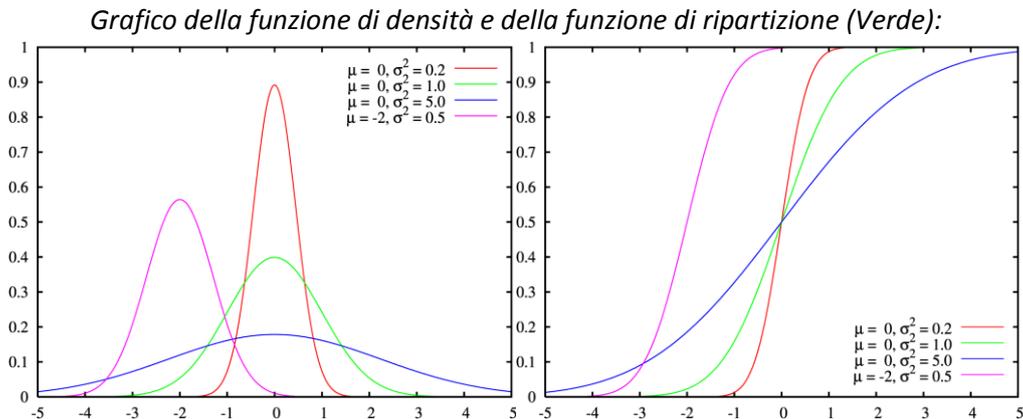
### Legge dei grandi numeri di Bernoulli:

Sia  $X_1, X_2, \dots$  successione di variabili indipendenti (Bernoulliane di parametro  $p$ ),

$$S_n = X_1 + \dots + X_n \text{ allora } \forall \varepsilon > 0 \text{ vale } \lim_{n \rightarrow +\infty} P \left\{ \left| \frac{S_n}{n} - p \right| < \varepsilon \right\} = 0$$

## Funzione di ripartizione della variabile N(0,1):

La funzione di ripartizione è ottenuta come l'integrale di una distribuzione normale o gaussiana.



La funzione di ripartizione è: 
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

### **Idea:**

Questa funzione ci serve in due casi, il primo è per il Limite centrale di variabili binomiali grazie al quale calcoliamo in maniera approssimata la probabilità di eventi complessi ("Tirando 400 dadi la probabilità che ottenga almeno 30 volte il dado 6") la seconda è a determinare la dimensione dell'intervallo in un test statistico (Usandole al contrario).

### Utilizzo della tavola statistica N(0, 1):

1. La tavola restituisce valori  $\Phi(x)$  per  $x \in [0, 4.49]$  per valori superiori si approssima  $\Phi(x) = 1$
2. Per calcolare  $\Phi(x)$  si scorre l'asse delle y fino a trovare l'unità e i decimi corrispondenti, una volta individuata la riga si scorrono le colonne fino al centesimo più vicino al valore che vogliamo calcolare.
3. Sfruttiamo la tavola e il Limite centrale per variabili binomiali per calcolare probabilità del tipo  $P\left\{\frac{X-k}{h} \leq s\right\} \cong \Phi(s)$  se ci troviamo a calcolare un oggetto del tipo opposto basta ricordarsi che è il complementare  $P\left\{\frac{X-k}{h} > s\right\} = 1 - P\left\{\frac{X-k}{h} \leq s\right\} \cong 1 - \Phi(s)$
4.  $\Phi(-x) = 1 - \Phi(x)$
5. Se dobbiamo usarle al contrario, ossia trovare  $q_a$  (Quantile) |  $P\{X \leq q_a\} = a$  scorriamo lungo le tavole (Il senso è da sinistra a destra e, una volta completata la riga, scendere a quella sottostante) fino a trovare il valore che meglio approssimi  $a$ .

### **Attenzione:**

La tavola mostra come risultati solamente gli  $a \in [\frac{1}{2}, 1]$  quindi se cerchiamo gli  $a \in [0, \frac{1}{2}]$  dobbiamo sfruttare la disuguaglianza:  $q_a = -q_{1-a}$

### **Variabili aleatorie a più dimensioni:**

I valori di una variabile aleatoria possono essere presi in ogni  $\mathbb{R}^n$ , senza perdere di generalità lavoreremo con il caso  $\mathbb{R}^2$ .

La variabile aleatoria sarà dunque un'applicazione  $(X, Y): \Omega \rightarrow \mathbb{R}^2$  e la sua legge di probabilità  $P_{X,Y} = (X, Y)_\#$  sarà definita sui sottoinsiemi di  $\mathbb{R}^2$ .

La densità discreta sull'insieme immagine (funzione di probabilità) viene definita come:

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\} = P\{(X, Y)^{-1}(x_i, y_j)\} = P\{X = x_i\} \cap P\{Y = y_j\}$$

Da cui per un generico  $B \subseteq \mathbb{R}^2$  abbiamo:

$$P_{X,Y}(B) = P\{(X, Y) \in B\} = \sum_{(x_i, y_j) \in B} p(x_i, y_j)$$

### **Definizione (Covarianza):**

Supponiamo  $X, Y$  ammettano momento secondo:

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

#### **Osservazione:**

La covarianza è bilineare:  $\text{Cov}(aX + bY, Z) = a \cdot \text{Cov}(X, Z) + b \cdot \text{Cov}(Y, Z)$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$$

Se  $\text{Cov}(X, Y) = 0$  le variabili si dicono **incorrelate**, in tal caso:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

#### **Proposizione (D13):**

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}$$

### **Definizione (Scarto quadratico medio):**

$$q(X) = \sqrt{\text{Var}(X)}$$

### **Definizione (Coefficiente di correlazione):**

Se  $X, Y$  sono due variabili aleatorie non costanti che ammettano momento secondo, il loro coefficiente di correlazione è:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{q(X)q(Y)}$$

**Proposizione (D14):**

Nel caso discreto valgono le formule:

$$p_X(x_i) = \sum_{y_j} p(x_i, y_j); p_Y(y_j) = \sum_{x_i} p(x_i, y_j)$$

**Idea:**

Avendo la relazione globale che lega le probabilità marginali sono capace di ricavarle, non vale però il contrario a meno che le due variabili siano indipendenti (Vedi pagina 8)

**Definizione (Probabilità prodotto):**

Prese  $P_1; P_2$  due probabilità discrete definite su  $E_1, E_2 \exists!$  Probabilità prodotto

$$P := P_1 \otimes P_2 \mid P(A \times B) = P_1(A) \cdot P_2(B) \text{ con } A \subseteq E_1; B \subseteq E_2$$

**Osservazione (Seconda caratterizzazione indipendenza fra v.a.) (D):**

$X, Y$  variabili aleatorie sono indipendenti  $\leftrightarrow P_{X,Y} := P((X,Y)^{-1}(A \times B)) = P_X \otimes P_Y$

Vale per un qualsiasi numero finito di variabili.

**Proposizione (D15):**

Siano  $X, Y$  due v.a. indipendenti e  $f, g$  due funzioni reali. Allora le variabili  $f \circ X; g \circ Y$  sono indipendenti.

**Generalizzazione:**

Funzioni di variabili indipendenti che non coinvolgono la stessa variabile sono indipendenti.

Quindi se  $(X, Y, Z)$  sono indipendenti allora  $f(X, Y); g(Z)$  sono indipendenti mentre  $f(X, Y); g(Y, Z)$  non lo sono.

**Osservazione:**

Nel caso di probabilità discrete, concentrate nei punti  $x_1, x_2, \dots$  per  $P_1$  e  $y_1, y_2, \dots$  per  $P_2$ ,

$$\text{vale: } p(x_i, y_j) = P_1 \otimes P_2(\{x_i, y_j\}) = p_1(x_i) \cdot p_2(y_j)$$

**Teorema (D16):**

$X, Y$  variabili aleatorie indipendenti dotate di momento primo allora anche  $XY$  ammette momento primo secondo la relazione:

$$E[XY] = E[X]E[Y]$$

**Corollario:**

Due variabili indipendenti dotate di momento secondo sono incorrelate.

**Formula della convoluzione discreta (D17):**

Siano  $X, Y$  variabili aleatorie indipendenti a valori interi e sia  $Z = X + Y$  vale la formula:

$$p_Z(n) = P\{Z = n\} = \sum_{h=-\infty}^{+\infty} p_X(h)p_Y(n-h)$$

**Notazione:**

$$\{X + Y = n\} = \bigcup_{h=-\infty}^{+\infty} \{X = h, Y = n - h\}$$

***Esempio applicazione:***

*Se  $X \sim B(n, p); Y \sim B(m, p)$  sono indipendenti, allora  $(X + Y) \sim B(n + m, p)$  da cui si può dedurre la speranza e la varianza di una variabile binomiale.*

### Riepilogo proprietà delle serie di potenze:

Data una successione di numeri  $(a_n)_{n \geq 0}$  si chiama serie di potenze ad essa associata  $g(t) = \sum_{n=0}^{+\infty} a_n t^n$

Il raggio di convergenza  $R := \frac{1}{\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}}$

La serie di potenze converge per  $|t| < R$  e non converge per  $|t| > R$

Supponendo  $R > 0$  vale la relazione  $a_n = \frac{1}{n!} \frac{d^n(g(t))}{dt^n}$

### Definizione (Funzione generatrice della probabilità):

Data una variabile aleatoria  $X$  a valori interi positivi  $G_X(t) = \sum_{n=0}^{+\infty} t^n p(n) = E[t^X]$

#### Osservazione pratica:

Si calcola spesso direttamente utilizzandole note proprietà delle serie, sia ad esempio  $X$  v.a.

$$| P(X = k) = (1 - \theta)^2 \theta^k (k + 1) \rightarrow G_X(t) = \sum_{n=0}^{+\infty} t^n (1 - \theta)^2 \theta^n (n + 1) = \frac{(1 - \theta)^2}{(1 - t\theta)^2}$$

#### Proprietà:

$G_X(t) = G_Y(t) \leftrightarrow X$  e  $Y$  sono equidistribuite

$X$  e  $Y$  sono indipendenti  $\rightarrow G_{X+Y}(t) = G_X(t)G_Y(t)$

#### Proposizione (D18):

Sia  $X$  una variabile aleatoria reale, allora vale:

$$E[X] = \lim_{t \rightarrow 1^-} G'_X(t)$$

$$E[X(X - 1)] = \lim_{t \rightarrow 1^-} G''_X(t)$$

#### Osservazione:

Questa proposizione può essere utilizzata sfruttando la seguente catena di passi:

Ho due v.a. che so essere indipendenti, conosco la loro funzione generatrice, allora la generatrice della somma è data dal prodotto delle generatrici da cui possiamo con questa proposizione ricavare la speranza.

### Funzioni generatrici comuni:

$$X \sim B(n, p) \rightarrow G_X(t) = [1 + p(t - 1)]^n$$

$$X \text{ geometrica di parametro } p \rightarrow G_X(t) = \frac{tp}{1 - t(1-p)}$$

$$X \text{ di Poisson di parametro } \lambda \rightarrow G_X(t) = e^{\lambda(t-1)}$$

## Inferenza statistica su di uno spazio di Probabilità numerabile:

### **Statistica descrittiva:**

Utilizzo dei dati di un'indagine statistica senza costruire formalmente un modello probabilistico che li interpreti.

Un'indagine statistica può essere vista come una funzione  $X: \{1, 2, \dots, n\} \rightarrow C$

Se  $C$  è piccolo si parla di indagine **qualitativa**.

Se  $C = \mathbb{R}$  o  $C = \mathbb{R}^k$  si parla di indagine **quantitativa**.

#### **Esempio:**

$\{1, 2, \dots, n\}$  corrisponde al numero del questionario o, di conseguenza, alla persona che lo ha compilato,  $C$  potrebbe ad esempio essere una  $k$ -upla ordinata di elementi dove ogni elemento è una risposta ad un questionario.

Su di un'indagine quantitativa possiamo sfruttare i concetti paralleli a quelli già introdotti in precedenza:

Data  $\{x_1, \dots, x_n\}$  immagine di  $X$  in  $\mathbb{R}$  ( $\{X(i) \mid i \in \{1, 2, \dots, n\}\}$ )

**Media empirica:**  $\bar{x} = \frac{x_1 + \dots + x_n}{n}$

**Varianza empirica:**  $\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$

### **Inferenza statistica:**

L'idea è partire dall'esperienza (studio di un campione) per individuare la migliore legge di probabilità su quel modello (Migliore nel senso che meglio lo rappresenta).

#### ***Esempio Controllo di qualità:***

*Dato un insieme molto numeroso di computer (Popolazione) vogliamo stimare la percentuale di pezzi guasti.*

*Per mancanza di tempo/motivi di costo è possibile testare solamente un sottoinsieme della popolazione (Le informazioni da esse ricavate mediante variabili aleatorie sono dette Campione) di dimensione  $n$ .*

#### **Osservazione:**

La percentuale deve essere sconosciuta perché altrimenti siamo in grado di assegnare direttamente una densità di probabilità (Bernoulliana) su questo insieme di parametro

$$p = \frac{\#computer\ guasti}{\#totale\ computer}$$

*Vorremo individuare la probabilità su di esso (significherebbe infatti conoscere la percentuale dei pezzi guasti) ma l'unica cosa che possiamo dire è che appartiene alla famiglia delle probabilità di Bernoulli.*

*Formalmente  $\Omega = \{0,1\}^n$  dove per convenzione 1 significa guasto, gli eventi sono le parti (Probabilità di pescare una determinata famiglia di pezzi) e la famiglia di probabilità è  $(P^\theta, \theta \in ]0,1[) \mid P^\theta(k_1, \dots, k_n) = \theta^{k_1+\dots+k_n} (1-\theta)^{n-(k_1+\dots+k_n)}$*

*Al solito gli eventi con cui stiamo lavorando sono stringhe ordinate di 0 e di 1 mentre a noi interessa studiare il "numero di pezzi guasti", definiamo quindi  $n$  v.a. Bernoulliane  $X_i(k_1, \dots, k_n) = k_i$  proiezioni canoniche indipendenti fra loro.*

#### **Osservazione:**

Le proiezioni così definite sono bernoulliane e indipendenti per le proprietà derivate da  $P^\theta$ , se cambiassi lo spazio le loro proprietà cambierebbero.

### **Definizione (Modello statistico):**

Un modello statistico è una terna  $(\Omega, F, (P^\theta, \theta \in \Theta))$  con  $\Omega$  insieme,  $F$  tribù su  $\Omega$  e  $(P^\theta, \theta \in \Theta)$  una famiglia di probabilità su  $(\Omega, F)$

Per comodità se  $\theta_1 \neq \theta_2 \rightarrow P^{\theta_1} \neq P^{\theta_2}$  (**Modello identificabile**)

Un evento **trascurabile** in un modello deve essere trascurabile per ogni scelta della probabilità nella famiglia.

#### **Osservazione.**

Per il resto del capitolo lavoreremo con  $\Omega$  numerabile in modo tale da poter definire la probabilità sui singoletti (Lavorando con una densità discreta di probabilità).

**Definizione (Verosimiglianza):**

Assegnato un modello statistico  $(\Omega, F, (P^\theta, \theta \in \Theta))$  con  $\Omega$  numerabile la verosimiglianza è la funzione:  $L: \Theta \times \Omega \rightarrow \mathbb{R}^+ \mid L(\theta, w) = P^\theta(\{w\})$

**Osservazione:**

Descritta la verosimiglianza possiamo ricostruire ogni probabilità della famiglia.

Infatti  $P^\theta(A) = \sum_{w_i \in A} L(\theta, w_i) \quad \forall A \in F, \forall \theta \in \Theta$

Nel caso discreto la verosimiglianza è a valori in  $[0,1]$ .

**Definizione (Campione):**

Data  $(m^\theta, \theta \in \Theta)$  una famiglia di leggi di probabilità discrete concentrate su un sottoinsieme numerabile  $C$  di  $\mathbb{R}$ .

Si chiama **campione** di taglia  $n$  e legge  $m^\theta$  una famiglia di variabili aleatorie indipendenti  $(X_1, \dots, X_n)$  tutte di legge  $m^\theta$ .

**Idea:**

Sebbene intuitivamente un campione sia un sottoinsieme di  $\Omega$  in realtà lo definiamo come un insieme di variabili aleatorie in quanto a noi interessa l'informazione associata ad un sottoinsieme di  $\Omega$ .

**Esempio:**

*Nel nostro esempio del controllo di qualità il campione sarà di taglia  $n$ , con  $n$  il numero di pezzi che testiamo, e di legge Bernoulli (Funzionante/difettoso) di parametro  $\theta$  da stimare.*

### **Definizione (Stima):**

Dato un modello statistico  $(\Omega, F, (P^\theta, \theta \in \Theta))$  la **stima** è una variabile aleatoria reale  $U: \Omega \rightarrow \mathbb{R}$

Alla stima è (Quasi sempre) accoppiata una funzione  $g: \theta \rightarrow \mathbb{R}$

Scopo della stima è valutare  $g(\theta)$  (Non  $\theta$  in quanto non sappiamo nemmeno se sia un numero), assegnare  $g(\theta)$  significa lavorare con parametri reali.

Una data probabilità della famiglia è caratterizzata da un  $\theta \in \Theta$  che è associato ad una  $g(\theta) \in \mathbb{R}$ .

### **Idea stima:**

La stima assegna ad ogni esito un valore (Tipicamente giusto/sbagliato nel controllo di qualità),  $g(\theta)$  d'altro canto è l'informazione completa della probabilità vera, stimare significa che ho informazione su di un tot di esiti (Il campione) e da esso generalizzo a tutto. Ovviamente generalizzare significa ipotizzare che la probabilità indotta su  $\mathbb{R}$  abbia una data distribuzione e, di conseguenza, significa stimare una probabilità della famiglia  $P^\theta$ .

### **Ricapitolando:**

$P^\theta$  è la vera probabilità, caratterizzata da un parametro  $\theta$  che associo mediante  $g$  ad un  $g(\theta) \in \mathbb{R}$ , la stima è una v.a. reale  $U$  il cui grafico si può interpretare nel seguente modo: mediante la regola da noi definita la curva descrive come scegliere  $\theta$ , se sto lavorando con delle informazioni parziali prendo  $\theta$  come la media integrale dei valori assunti dal grafico in quei punti.

Di conseguenza se il grafico è abbastanza stabile attorno al valore  $g(\theta)$  che sto cercando distimare significa che con un campione piccolo scelto a caso ho comunque buone informazioni.

Con alte oscillazioni significa che è facile che con un campione ridotto si ottenga un valore abbastanza distante dall'originale.

### **Esempio:**

*Se nel controllo di Qualità prendo come campione tutto l'insieme ho una  $n$ -upla di variabili aleatorie la cui somma fratto  $n$  mi descrive completamente il parametro (In pratica diventa uno spazio di probabilità). Se invece il campione avesse taglia  $k$  allora la somma delle  $k$  variabili aleatorie che mi descrivono l'informazione su di esso fratto  $k$  è una stima, è ovvio che a seconda di "dove" prendo il campione la stima può assumere valori diversi ed è altrettanto ovvio che potrei definire la stima in maniera diversa (Se ad esempio sapessi che la macchina per testare i pezzi diventa progressivamente meno efficace allora non farei la semplice media definita prima ma potrei fare valere di più le informazioni ottenute dai primi test).*

### **Esempio di Stima:**

Vogliamo stimare quale sia  $\theta$  la cui probabilità descriva al meglio la mia popolazione, preso il campione  $w$  di taglia  $n$  e le variabili  $X_i$  prima definite (Le proiezioni sull' $i$ -esimo termine del campione) una stima di  $\theta$  potrebbe essere  $\theta = \bar{X}(w) = \frac{X_1(w) + \dots + X_n(w)}{n}$

### **Esempio applicazione:**

*Se ho un campione di taglia 100 sulla popolazione 1000 dei computer e di questi 100 8 sono guasti allora un'approssimazione potrebbe essere la variabile binomiale di parametri  $B\left(1000, \frac{8}{100}\right)$ .*

*Nel caso specifico  $\theta \in [0,1]$ , se così non fosse sfrutterei  $g$  per collegare  $\theta$  a  $[0,1]$*

### **Definizione (Stima corretta):**

Assegnata un  $g: \theta \rightarrow \mathbb{R}$  la stima  $U$  di  $g(\theta)$  (Del parametro quindi della probabilità) è detta corretta se  $\forall \theta, U$  è  $P^\theta$  integrabile e  $E^\theta[U] = g(\theta)$

#### **Esempio:**

In un campione di taglia  $n$  e legge geometrica di parametro  $\theta$  la stima  $U = \frac{X_1 + \dots + X_n}{n}$  è corretta.

#### **Idea:**

Assegnata una stima della probabilità  $P^\theta$  per essere corretta deve prima di tutto essere integrabile.

Quindi  $\sum U(w)P^\theta(w) = \sum rP_U(r) < +\infty$  ossia non voglio che la variabile aleatoria mi dia valori assurdi in  $\mathbb{R}$ .

Dopodiché deve valere  $E^\theta[U] = g(\theta)$  ossia il valor medio (speranza) della mia variabile aleatoria deve essere proprio il parametro che stiamo cercando.

Stiamo semplicemente dicendo che nel caso in cui io abbia tutti i dati la stima che devo ottenere è proprio  $g(\theta)$

### **Definizione (Stima consistente):**

Consideriamo  $\forall n$  un campione  $X_1, \dots, X_n$  di taglia  $n$  e una stima  $U_n = h_n(X_1, \dots, X_n)$  di  $g(\theta)$   
Si dice che questa successione di stime è consistente se scelti comunque  $\theta \in \Theta; \varepsilon > 0$  si ha:  
 $\lim_{n \rightarrow \infty} P^\theta\{|U_n - g(\theta)| > \varepsilon\} = 0$

#### **Esempio:**

In un campione infinito di leggi di Poisson di parametro  $\theta$  la successione delle medie empiriche  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$  è una stima consistente di  $\theta$ .

#### **Idea:**

Una successione di stime è consistente se progressivamente tendono a diventare più precise. Inoltre (Pratica) se  $U_n \rightarrow g(\theta); \text{Var}^\theta(U_n) \rightarrow 0$  la stima è consistente

### **Definizione (Stima di massima verosimiglianza):**

Assegnato un modello statistico  $(\Omega, F, (P^\theta, \theta \in \Theta)) \mid \theta \subseteq \mathbb{R}$  si dice che  $U$  è una stima di massima verosimiglianza se  $\forall w \in \Omega$  si ha  $L(U(w), w) = \sup_{\theta \in \Theta} L(\theta, w)$ . Se esiste viene indicata con  $\hat{\theta}(w)$   
(Va bene anche se l'uguaglianza è a meno di un insieme trascurabile)

#### **Osservazione pratica:**

Quello che si fa per calcolare la stima di massima verosimiglianza è fissare l'elemento  $w$  (Di cui spesso utilizziamo solo una parte dell'informazione come ad esempio la somma delle componenti), calcolare la stima  $L(\theta, w)$  e cercare l' $\theta$  che massimizzi la funzione, se appartiene all'intervallo quella è una stima di massima verosimiglianza.

Ci riduciamo in pratica a studiare una funzione di variabile  $\theta$  e parametro  $w$ .

**Esempio stima di massima verosimiglianza:**

Considerato un campione  $(X_1, \dots, X_n)$  di legge Geometrica di parametro  $\theta$  e taglia  $n$  e lo spazio  $(\mathbb{N}^*)^n$

La verosimiglianza è:  $L(\theta; k_1, \dots, k_n) = (1 - \theta)^{k_1 + \dots + k_n - n} \cdot \theta^n$

Cerchiamo il massimo di questa funzione (Teniamo  $n$  e  $k_1, \dots, k_n$  fissati facendo variare solamente  $\theta$ )

Il massimo lo otteniamo nel punto  $\frac{n}{k_1 + \dots + k_n}$ .

Quindi la stima di massima verosimiglianza sarà data da  $\hat{\theta}_n(k_1, \dots, k_n) = \frac{n}{k_1 + \dots + k_n}$

Equivalente:

$$\hat{\theta}_n = \frac{n}{X_1 + \dots + X_n}$$

Se consideriamo un campione infinito (Vedi sotto) allora la successione delle stime  $(\hat{\theta}_n)_{n \geq 1}$  è consistente.

**Modelli esponenziali:**

Un modello nel quale la funzione di probabilità è del tipo  $p(\theta, k) = c(\theta) \cdot \exp(\theta, T(k)) \cdot g(k)$

con  $\theta \in \Theta$  intervallo su  $\mathbb{R}$ .

**Esempio:**

Un modello nel quale  $p(\theta, w) = (1 - \theta)^2 \cdot \theta^{T(w)} \cdot \prod (k_i + 1)$  con  $T(w) = \sum k_i$  e  $w = \{k_1, \dots, k_n\}$  è esponenziale.

**Teorema (Stima consistente):**

Dato un modello esponenziale nel quale per ogni  $n$  (Taglia del campione) esiste la stima di massima verosimiglianza  $\hat{\theta}_n$  allora la successione di stime  $(\hat{\theta}_n)_{n \geq 1}$  è consistente.

**Attenzione:**

Molti dei modelli prima descritti possono essere ricondotti ad un modello esponenziale.

Infatti nel caso di leggi geometriche:  $p(\theta, k) = \theta \cdot \exp((k - 1), \log(1 - \theta))$

Ed invece per una legge di Poisson:  $p(\theta, k) = e^{-\theta} \cdot \theta^k \cdot (k!)^{-1} = e^{-\theta} \cdot \exp(k, \log(\theta)) \cdot (k!)^{-1}$

### **Definizione (Rischio):**

Data  $U$  stima del parametro  $g(\theta)$  si chiama **rischio** (quadratico) il numero:

$$R(\theta, U) = E^\theta \left( (U - g(\theta))^2 \right)$$

#### **Idea:**

Ricordiamoci che  $U$  stima (In funzione degli esiti che prendo come campione) il parametro  $\theta$ , di conseguenza il rischio è un numero che rappresenta quanto la funzione si discosti dal parametro  $g(\theta)$ . Detto in altra maniera se ipotizziamo che la probabilità giusta sia quella di parametro  $\theta$  e data una stima  $U$  stiamo cercando quanto ci si può discostare da quel parametro se io prendo un campione.

#### **Osservazione:**

Se  $U$  è corretta  $R(\theta, U) = \text{Var}^\theta(U)$

### **Possibilità di ordinare le stime in funzione del rischio:**

$U$  è **preferibile** a  $V$  se  $\forall \theta R(\theta, U) \leq R(\theta, V)$

$U$  è **strettamente preferibile** a  $V$  se è preferibile ed  $\exists \bar{\theta} \mid R(\bar{\theta}, U) < R(\bar{\theta}, V)$

$U$  è **ammissibile** se non esistono stime strettamente preferibili.

$U$  è **ottimale** se preferibile ad ogni altra stima.

#### **Idea:**

Significa che non ho idea di quale sia la probabilità ma se per ogni elemento della famiglia la stima 1 è più vicina ad essa della stima 2 allora comunque sceglierò la stima 1.

Può darsi che due stime abbiano rischi diversi a seconda del parametro.

Ad esempio la stima 1 potrebbe essere molto migliore nel caso in cui il parametro sia minore di un tot mentre peggiore della stima 2 nel caso il parametro sia maggiore. In questi casi se siamo in grado di determinare a quale delle due famiglie appartenga la probabilità vera possiamo stabilire quale sia la stima migliore, altrimenti rimangono non confrontabili.

### **Definizione (Riassunto esaustivo):**

Sia  $T: \Omega \rightarrow E$  una variabile aleatoria, si dice che  $T$  è un riassunto esaustivo se la verosimiglianza si può scrivere nella forma:  $L(\theta, w) = h(\theta, T(w)) \cdot k(w)$

#### **Teorema (D19):**

Sia  $T$  un riassunto esaustivo e  $U$  una stima di  $g(\theta)$  allora  $\exists$  una stima  $V$  della forma  $V(w) = f(T(w))$  preferibile ad  $U$  e inoltre se  $U$  non è già della forma  $f \circ T$  è strettamente preferibile. Se  $U$  è corretta allora  $V$  è corretta.

#### **Idea:**

Il riassunto esaustivo contiene tutte le informazioni rilevanti per ottenere una buona stima (Come il numero di pezzi difettosi nel controllo di qualità).

**Definizione (Regione di fiducia):**

Sia assegnato  $\forall w \in \Omega$  un sottoinsieme di parametri  $C(w) \subseteq \Theta$ , si dice che  $C(w)$  è una regione di fiducia per il parametro  $\theta$  al livello  $(1 - \alpha)$  se qualunque sia  $\theta$  vale  $P^\theta\{w \mid \theta \in C(w)\} \geq 1 - \alpha$

**Osservazione:**

Si usa la notazione **intervallo di fiducia** se  $\Theta \subseteq \mathbb{R}$  e  $C(w)$  è un intervallo.

**Obiettivo:**

Definire un insieme dei parametri (Ad esempio centrato attorno alla stima) per il quale la probabilità che il parametro “vero” appartenga all’intervallo sia maggiore di  $1 - \alpha$ .

**Esempio Controllo di qualità:**

Consideriamo un campione  $X_1, \dots, X_n$  di variabili di Bernoulli di parametro  $\theta$  su tutta la popolazione, una stima (corretta ma funzione di un campione di taglia  $n$ )  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  e la varianza  $Var^\theta(\bar{X}) = \frac{\theta(1-\theta)}{n}$

Vogliamo individuare un intervallo di fiducia  $I = [\bar{X} - d, \bar{X} + d]$  per il quale la probabilità che  $\theta \in I$  sia  $> 1 - \alpha$

Questo è equivalente a calcolare  $P^\theta\{|\bar{X} - \theta| > d\} \leq \frac{\theta(1-\theta)}{nd^2}$  per la disuguaglianza di Chebishev.

Siccome non sappiamo il vero valore di  $\theta$  deve valere sempre, siccome il massimo di  $\theta(1 - \theta) = \frac{1}{4}$

Basta porre  $d = \frac{1}{\sqrt{4n\alpha}}$  e si ottiene l’intervallo di fiducia  $[\bar{X} - \frac{1}{\sqrt{4n\alpha}}, \bar{X} + \frac{1}{\sqrt{4n\alpha}}]$

## Test statistici:

### **Idea:**

In pratica quello che stiamo facendo è formulare un'ipotesi (che porti ad una riduzione della famiglia di parametri su cui stiamo lavorando) e verificare, mediante un test, se possiamo fidarci di essa.

Caso reale: un'azienda rivenditrice di computer ne acquista una partita da un fornitore che garantisce una percentuale di pezzi difettosi al di sotto dell'1%.

Per decidere se possiamo fidarci della garanzia del fornitore dobbiamo eseguire un test su un numero ridotto di pezzi e decidere da esso se accettare per vera l'affermazione di prima.

### **Formalmente:**

Dividiamo l'insieme dei parametri  $\Theta$  in due famiglie  $(\Theta_0, \Theta_1)$ , quelle che rispettano l'ipotesi  $H_0$ ) e quelle che non la rispettano  $H_1$ ).

Definire un test significa scegliere un evento  $D \in \mathcal{F}$  (Detto **regione critica**) i cui elementi sono i risultati che ci spingono a rifiutare l'ipotesi.

### **Esempio semplice:**

*Se testo più dell'1% dei pezzi e risultano tutti difettosi allora l'ipotesi è falsa.*

*Un altro modo per definirla potrebbe essere  $D = \left\{ w \in \Omega \mid \bar{X}(w) > \frac{1}{100} \right\}$  con  $\bar{X}$  la stima.*

## Definizione (Livello):

La **taglia** di un test di regione critica  $D$  è il  $\sup_{\theta \in \Theta_0} P^\theta(D)$

Il test è di **livello**  $\alpha$  se la sua taglia è  $\leq \alpha$

### **Idea:**

La taglia mi indica la peggior situazione possibile, ossia la massima probabilità che il test restituisca falso un risultato vero (**Errore di prima specie**)

## Definizione (Potenza):

La **potenza** di un test di regione critica  $D$  è la funzione  $\pi_D: \Theta_1 \rightarrow [0,1]$  definita da  $\theta \rightarrow P^\theta(D)$

La regione critica  $D$  è più potente della regione critica  $D^*$  se  $\forall \theta \in \Theta_1$  vale  $P^\theta(D) \geq P^\theta(D^*)$

### **Idea:**

Se è più potente significa che è più probabile che, nel caso in cui l'ipotesi sia falsa, ossia  $\theta_{vero} \in \Theta_1$ , venga fuori uno degli eventi (quelli appartenenti alla regione critica) che mi permetta di considerare falsa l'ipotesi. (Evitare un **errore di seconda specie**)

## **Osservazione pratica:**

Nella pratica fissiamo un livello che ci garantisca di non eliminare troppo spesso dei casi "buoni" da cui ricaviamo una forma e un limite per la regione critica. Dopodiché cerchiamo di incrementare la potenza aumentandone il più possibile la dimensione.

### **Test ad ipotesi semplice:**

Un test si dice ad ipotesi (Alternativa) semplice se l'ipotesi (Alternativa) è della forma  $\{\theta_0\}$ .

### **Regione critica e regione di fiducia nei test ad ipotesi semplice:**

Dato  $\forall w \in \Omega$  una regione di fiducia  $C(w)$  al livello  $(1 - \alpha)$  e un test di ipotesi  $H_0$ ) semplice (Quindi l'unico caso che mi interessa e se il parametro è uguale a  $\theta_0$ ) di alternativa  $\theta \neq \theta_0$

La condizione di rifiuto dell'ipotesi è  $\theta_0 \notin C(w)$  quindi la regione critica è della forma:

$$D = \{w \in \Omega \mid \theta_0 \notin C(w)\} \text{ e vale } P^{\theta_0}(D) \leq \alpha \text{ ossia il test è di livello } \alpha.$$

Riassumendo: regione di fiducia al livello  $(1 - \alpha) \rightarrow$  test di livello  $\alpha$ .

### ***Esempio interessante:***

*Dato un campione  $X_1, \dots, X_n$  con legge di Bernoulli pianifichiamo il test con ipotesi semplice  $H_0) \{\theta_0\}$  Siccome un intervallo di fiducia si può scrivere nella forma  $C(w) = \{\theta \in \Omega \mid -d \leq \bar{X}(w) \leq d\}$  da cui otteniamo la regione critica  $D = \{w \mid |\bar{X}(w) - \theta_0| > d\}$  con  $d$  da calcolare*

*(Attenzione: questa costruzione non mi serve a calcolare la grandezza dell'intervallo ma solo a determinarne la forma).*

*Siccome per massimizzare la potenza la regione critica deve essere più grande possibile cerchiamo il minimo  $d$  per cui valga  $P^{\theta_0}\{|\bar{X} - \theta_0| > d\} \leq \alpha$  con  $\alpha$  il livello richiesto dal test.*

*Per ricavare  $d$  possiamo sfruttare la disuguaglianza di Chebishev oppure usare il teorema di De Moivre-Laplace.*

### **Lemma di Neyman-Pearson (D20):**

Dato un modello statistico con l'insieme dei parametri ridotto a due punti ( $\Theta = \{\theta_0, \theta_1\}$ ) e un test di ipotesi semplice  $H_0) \theta = \theta_0$ .

Consideriamo l'insieme  $D = \{w \in \Omega \mid L(\theta_0, w) \leq c \cdot L(\theta_1, w)\}$  con  $c$  costante positiva.

Allora:

1.  $D$  è la regione critica di un test più potente di ogni altro test di livello  $P^{\theta_0}(D)$
2. Vale la disuguaglianza  $P^{\theta_1}(D) \geq P^{\theta_0}(D)$

### **Osservazione:**

La generalizzazione di questo risultato permette di studiare con facilità i test detti unilateri.

### **Test unilatero:**

Un test si dice unilatero se l'ipotesi è della forma  $H_0) \theta \leq \theta_0$  oppure  $H_0) \theta \geq \theta_0$ .

**Attenzione:** l'insieme dei parametri deve essere un intervallo di  $\mathbb{R}$

### **Rapporto di verosimiglianza crescente:**

Dato un modello statistico con  $\Theta$  intervallo di  $\mathbb{R}$  e  $T$  una variabile aleatoria reale.

Si dice che il **modello è a rapporto di verosimiglianza crescente rispetto a  $T$**  se scelti  $\theta_1 < \theta_2 \exists$  una funzione reale strettamente crescente a valori positivi per cui valga l'uguaglianza:

$$\frac{L(\theta_2, w)}{L(\theta_1, w)} = f_{\theta_1, \theta_2}(T(w))$$

#### **Osservazione 1:**

La funzione è definita rispetto alla stima e non rispetto al campione perché non ha senso dire se un campione è maggiore o minore di un altro mentre la stima valutata nel campione (Essendo un intervallo di  $\mathbb{R}$ ) è ordinata.

#### **Osservazione 2:**

Avere un modello con stima di verosimiglianza crescente significa che man mano che all'aumentare della stima è maggiore (Nel caso del controllo di qualità che aumenti il numero di pezzi difettosi nel campione)

Il rapporto fra verosimiglianze (Che altro non sono che la probabilità rispetto ai due parametri distinti  $\theta_1, \theta_2$  che esca proprio il campione  $w$  grazie al quale sto costruendo una stima) può aumentare se il numeratore cresce di più del denominatore.

Attenzione: va bene anche decresce di meno.

Quindi all'aumentare della stima il fatto che il rapporto aumenti vuol dire che in proporzione diventa sempre più alta la verosimiglianza all'aumentare del parametro, più alta verosimiglianza vuol dire che è sempre più probabile che il parametro sia più alto. Dunque ha senso pensare che se stiamo facendo un test la regione critica sarà del tipo  $\geq d$ .

#### **Osservazione 3:**

Questa notazione è alquanto pesante tuttavia nella pratica ci limitiamo a studiare il rapporto fra verosimiglianze da cui, ricordandoci come abbiamo definito la stima, cerchiamo di capire se aumentandola aumenta anche il valore.

### ***Esempio Controllo di qualità:***

Dato il solito spazio  $\Omega = \{0,1\}^n$  con campione di Bernoulli il rapporto fra verosimiglianze è

$$\frac{L(\theta_2, k_1, \dots, k_n)}{L(\theta_1, k_1, \dots, k_n)} = \left(\frac{\theta_2}{\theta_1}\right)^{k_1 + \dots + k_n} \left(\frac{1-\theta_2}{1-\theta_1}\right)^{n - (k_1 + \dots + k_n)}$$

Il rapporto, a prescindere dalla scrittura esatta della funzione, all'aumentare della stima (E di conseguenza della somma  $k_1 + \dots + k_n$ ) aumenta.

Questo significa che il modello così costruito è a verosimiglianza crescente.

### **Teorema (D21):**

Dato un modello a verosimiglianza crescente rispetto ad una stima  $T$  e il test unilatero di ipotesi  $H_0) \theta \leq \theta_0$ .

Consideriamo l'insieme  $D = \{w \in \Omega \mid T(w) \geq d\}$ .

Il test di regione critica  $D$  è tale che:

1. Vale l'uguaglianza  $\sup_{\theta \leq \theta_0} P^\theta(D) = P^{\theta_0}(D)$

#### **Osservazione:**

Sta dicendo che la massima probabilità che io prenda per falso un risultato vero è proprio  $P^{\theta_0}(D)$

2.  $D$  è più potente di qualsiasi altro test con lo stesso livello ( $P^{\theta_0}(D)$ ).

#### **Osservazione:**

Significa che una volta che ho stabilito che il modello è a verosimiglianza crescente rispetto alla stima scelta ho in automatico la forma della regione critica.

Inoltre, scelto il livello del test (Dal quale ricaviamo l'intervallo effettivo usando ad esempio il teorema di De Moivre-Laplace), non dobbiamo preoccuparci di ottimizzare la potenza perché è già la massima possibile.

### **Attenzione (Invertire):**

Nel caso in cui il modello abbia verosimiglianza decrescente oppure se l'ipotesi è della forma  $H_0) \theta \geq \theta_0$  la regione critica è della forma:  $D = \{w \in \Omega \mid T(w) \leq d\}$

#### **Esempio Controllo di qualità:**

*Avendo stabilito che il modello è a verosimiglianza crescente per la stima data sappiamo che la regione critica della forma  $D = \{w \in \Omega \mid \bar{X}(w) \geq d\}$*

#### **Esempio super pratico:**

*Questo è il caso di un test su pezzi difettosi, se trovo nel campione un numero di pezzi difettosi superiore ad un tot. questo mi spinge a rifiutare l'ipotesi "i pezzi difettosi sono pochi".*

*Fissato il livello  $\alpha$  e usando il teorema del limite centrale per variabili binomiali possiamo ricavare:*

$$\sup_{\theta \leq \theta_0} P^\theta(D) = P^{\theta_0}(D) = P^{\theta_0} \left\{ \sqrt{n} \frac{\bar{X} - \theta_0}{\sqrt{\theta_0(1-\theta_0)}} \geq \sqrt{n} \frac{d - \theta_0}{\sqrt{\theta_0(1-\theta_0)}} \right\} \approx 1 - \Phi \left( \sqrt{n} \frac{d - \theta_0}{\sqrt{\theta_0(1-\theta_0)}} \right) = \alpha$$

*Utilizzando le tavole della distribuzione normale  $N(0,1)$  otteniamo:  $d = \theta_0 + \frac{q_{1-\alpha} \sqrt{\theta_0(1-\theta_0)}}{\sqrt{n}}$*

## **Probabilità generale:**

### **Introduzione:**

Vogliamo definire in maniera più generale i concetti introdotti nei capitoli precedenti. Per fare questo necessitiamo di uno spazio canonico con cui lavorare ( $\mathbb{R}$ ) che doteremo di una  $\sigma$ -algebra di parti detta dei boreliani.

### **Definizione (Borelliani o $\sigma$ -algebra di Borel):**

I boreliani (Sulla retta  $\mathbb{R}$ ) sono la  $\sigma$ -algebra generata dagli aperti o dai chiusi di  $\mathbb{R}$ .

#### **Osservazione:**

Una  $\sigma$ -algebra generata da un insieme è la piccola fra tutte quelle che lo contengono.

**Equivalente:** è l'intersezione di tutte quelle che lo contengono.

#### **Osservazione:**

$B(\mathbb{R}^n)$  è definita come la  $\sigma$ -algebra su  $\mathbb{R}^n$  generata dagli aperti di  $\mathbb{R}^n$  o, equivalentemente dai prodotti cartesiani dei Boreliani di  $\mathbb{R}$

Osservazione (Lebesgue misurabili):

### **Teorema di unicità di Probabilità:**

Siano  $P, Q$  due probabilità definite su una  $\sigma$ -algebra di parti  $F$  di un insieme  $E$  e supponiamo che coincidano su di un insieme  $I$  di parti  $| I$  genera  $F$  e  $I$  è stabile per intersezione finita.

Allora  $P$  e  $Q$  coincidono su tutto  $F$ .

### **Teorema di esistenza di Probabilità:**

Sia  $A$  un'algebra di parti di un insieme  $E$ . Sia  $P: A \rightarrow [0,1]$   $\sigma$ -additiva (Con  $P(E) = 1$ ).

Allora  $P$  si prolunga (In un solo modo) alla  $\sigma$ -algebra  $F$  generata da  $A$ .

## Costruzione delle probabilità su $\mathbb{R}$ :

### Definizione (Funzione di ripartizione):

Data  $P$  probabilità definita su  $(\mathbb{R}, B(\mathbb{R}))$  si chiama funzione di ripartizione la funzione:

$$F: \mathbb{R} \rightarrow [0,1] \mid F(x) = P(]-\infty, x])$$

### **Proprietà (D):**

La funzione di ripartizione è:

Crescente

Continua a destra

$$F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1 ; F(-\infty) = 0$$

### **Teorema di esistenza di una probabilità su $B(\mathbb{R})$ (D22):**

Assegnata  $F: \mathbb{R} \rightarrow [0,1]$  che rispetti le proprietà elencate.

$$\text{Allora } \exists! P \text{ probabilità su } (\mathbb{R}, B(\mathbb{R})) \mid \forall x \in \mathbb{R} F(x) = P(]-\infty, x])$$

### **Osservazione: probabilità discrete**

Sono probabilità concentrate su di una successione di punti  $(x_1, x_2, \dots)$ .

Su queste probabilità vale l'uguaglianza  $P(A) = \sum_{x_i \in A} p(x_i) \forall A \in B(\mathbb{R})$  con  $p(x_i) = P(\{x_i\})$

### Applicazioni misurabili:

Dati  $(E_1, \varepsilon_1); (E_2, \varepsilon_2)$   $f: E_1 \rightarrow E_2$  si dice misurabile se  $\forall A \in \varepsilon_2 \ f^{-1}(A) \in \varepsilon_1$

### Definizione (Funzione semplice):

Dato  $(E, \varepsilon)$  spazio misurabile si chiama **semplice** una funzione misurabile  $\varphi: E \rightarrow \mathbb{R}$  che assume un numero finito di valori (Imm( $\varphi$ ) ha dimensione finita).

#### **Osservazione:**

Sia  $\text{Imm}(\varphi) = \{a_1, \dots, a_n\}$  e poniamo per definizione  $A_i = \{\varphi = a_i\}$ , allora la funzione può essere scritta come combinazione lineare di indicatori di insiemi misurabili.

$$\varphi = \sum_{i=1}^n a_i \cdot I_{A_i}$$

#### **Viceversa:**

Ogni combinazione lineare di funzioni indicatori su insiemi misurabili è semplice.

Osservazione aggiuntiva:

Fissata una coppia di funzioni allora esistono  $A_1, \dots, A_n$  disgiunti per cui si possa scrivere:

$$f_1 = \sum_{i=1}^n a_i \cdot I_{A_i}; \quad f_2 = \sum_{i=1}^n b_i \cdot I_{A_i}$$

Quindi l'insieme delle funzioni semplici è un **spazio vettoriale** ed un reticolo (Max e Min sono semplici).

### Definizione (Integrale delle funzioni semplici):

Dato uno spazio di misura (normalizzato)  $(E, \varepsilon, m)$  e una funzione semplice della forma

$\varphi = \sum_{i=1}^n a_i \cdot I_{A_i}$  allora definiamo l'integrale di  $\varphi$  come:

$$\int_E \varphi(x) d\mathbf{m} = \sum_{i=1}^n a_i \cdot m(A_i)$$

#### **Proprietà:**

$$\int_E (a\varphi + \Psi) d\mathbf{m} = a \cdot \int_E \varphi d\mathbf{m} + \int_E \Psi d\mathbf{m}$$

$$\text{Se } \varphi \leq \Psi \text{ allora } \int_E \varphi d\mathbf{m} \leq \int_E \Psi d\mathbf{m}$$

### **Proprietà di Beppo-Levi per funzioni semplici:**

Sia  $(\varphi_n)_{n \geq 1}$  una successione di funzioni semplici e supponiamo che  $\varphi_n \uparrow \varphi$  e che  $\varphi$  sia semplice.

$$\text{Allora: } \int \varphi_n d\mathbf{m} \uparrow \int \varphi d\mathbf{m}$$

### **Approssimazione con funzioni semplici (D23):**

Sia  $f$  una funzione misurabile a valori positivi.

Allora esiste una successione di funzioni semplici  $(\varphi_n)_{n \geq 1} \mid \varphi_n \uparrow f$

**Definizione (Integrale delle funzioni a valori positivi):**

Sia  $f$  una funzione misurabile a valori positivi e consideriamo una successione di funzioni semplici  $(\varphi_n)_{n \geq 1} \mid \varphi_n \uparrow f$

Allora si definisce:

$$\int f d\mathbf{m} := \lim_{n \geq 1} \int \varphi_n d\mathbf{m}$$

**Proprietà di Beppo-Levi (D24):**

Se  $(\varphi_n)_{n \geq 1}$  e  $(\psi_n)_{n \geq 1}$  sono due successioni di funzioni semplici convergenti alla funzione  $f$  si ha:

$$\lim_{n \rightarrow \infty} \int \varphi_n d\mathbf{m} = \lim_{n \rightarrow \infty} \int \psi_n d\mathbf{m}$$

**Osservazione:**

Questa proprietà mi garantisce che l'integrale prima definito non dipenda dalla scelta della successioni delle funzioni semplici.

**Definizione (Funzione integrabile ed integrale):**

Si dice che la funzione misurabile  $f$  è integrabile se  $\int |f| d\mathbf{m} < +\infty$  e in tal caso si chiama integrale di  $f$  il numero:

$$\int f d\mathbf{m} = \int f^+ d\mathbf{m} - \int f^- d\mathbf{m}$$

**Teorema di convergenza dominata:**

Sia  $(f_n)_{n \geq 1}$  una successione di funzioni misurabili convergente puntualmente ad  $f$  e supponiamo che esista  $g$  integrabile a valori positivi | si abbia  $\forall x \in E \mid f_n(x) \leq g(x)$ , allora si ha:

$$\lim_{n \rightarrow \infty} \int f_n d\mathbf{m} = \int f d\mathbf{m}$$

**Osservazione:**

Vale ugualmente la disuguaglianza di Schwarz.

**Osservazione (Misure discrete):**

Quando l'insieme  $E$  è numerabile (o a misura concentrata su di un insieme numerabile) allora questa definizione di integrale coincide con quella di pagina 13.

**Definizione (Densità di probabilità):**

Si chiama densità di probabilità su  $\mathbb{R}$  una funzione  $f: \mathbb{R} \rightarrow \mathbb{R}$  misurabile, a valori positivi, integrabile secondo Lebesgue e tale che  $\int_{-\infty}^{+\infty} f(x)dx = 1$

**Probabilità associata su  $B(\mathbb{R})$ :**

$P(A) = \int_A f(x)dx$  è una probabilità sui boreliani indotta da questa funzione.

**Osservazione:**

Questa probabilità è  $\sigma$ -additiva.

$$P(\mathbb{R}) = 1$$

**Idea:**

È la generalizzazione sul continuo del distribuire ad ogni punto di un insieme finito (o al più numerabile) un valore associato.

**Esempio:**

*Se voglio studiare la probabilità che una persona scelta a caso dalla folla abbia altezza compresa fra 1.83 e 1.88 non posso non considerare che ogni valore di  $\mathbb{R}$  compreso fra i due estremi possa essere assunto. Con conoscenze successive potremmo affermare che la funzione sarà Gaussiana attorno all'altezza media del paese da cui sto selezionando una persona.*

**Teorema (D25):**

Detta  $P$  la probabilità indotta da una densità di probabilità  $f$ .

Una funzione misurabile  $g$  definita su  $\mathbb{R}$  è integrabile rispetto a  $P$  se e solo se il prodotto  $gf$  è integrabile rispetto alla misura di Lebesgue.

$$\text{In tal caso: } \int g(x)dP(x) = \int g(x)f(x)dx$$

**Osservazione:**

Queste definizioni di probabilità e di integrazione possono essere generalizzate al caso  $(\mathbb{R}^n, B(\mathbb{R}^n))$

**Osservazione (Funzione di ripartizione):**

La funzione di ripartizione di una probabilità definita da una densità  $f$  è:

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

**Proposizione (Funzioni assolutamente continue):**

La probabilità associata ad una funzione di ripartizione  $F$  è definita da una densità  $\leftrightarrow F$  è assolutamente continua.

$$\text{Cioè } \forall \varepsilon > 0 \exists \delta \mid \sum_{i \leq n} |x_i - y_i| < \delta \rightarrow \sum_{i \leq n} |F(x_i) - F(y_i)| < \varepsilon$$

***Criterio pratico:***

Se  $F$  è una funzione di ripartizione continua e  $C^1$  a tratti (Ossia derivabile con derivata continua eccetto in un insieme finito di punti  $a_1, \dots, a_n$ ) allora la probabilità associata ad  $F$  è definita da una densità e una versione della densità  $f$  è data da (Eccetto in  $\{a_1, \dots, a_n\}$ ):

$$f(x) = \frac{dF(x)}{dx}$$

**Definizione (Variabile aleatoria reale):**

Fissato  $(\Omega, F, P)$  spazio di probabilità, una variabile aleatoria reale è un'applicazione misurabile:

$$X: (\Omega, F) \rightarrow (\mathbb{R}, B(\mathbb{R}))$$

**Attenzione:**

Questa volta siccome lo spazio di probabilità è generale la variabile aleatoria deve rispettare la seguente proprietà (Sappiamo che i semi intervalli generano i boreliani):

$$\{X \leq x\} = X^{-1}([-\infty, x]) \in F$$

Quindi non è più vero che assegnata una generica funzione  $f: \mathbb{R} \rightarrow \mathbb{R}$  allora  $f \circ X$  è ancora una v.a., per fare sì che rimanga una variabile aleatoria abbiamo bisogno che  $f$  sia boreliana.

**Definizione (Legge di probabilità):**

Si chiama legge (o distribuzione) di probabilità di una variabile aleatoria reale  $X$  la probabilità  $P$  da essa indotta.

Alla stessa maniera la funzione di ripartizione di  $X$  è la funzione di ripartizione della sua legge di probabilità.

**Osservazione (Costruzione variabile aleatoria associata ad una probabilità):**

Fissata una probabilità  $Q$  su  $(\mathbb{R}, B(\mathbb{R}))$ , individuuiamo una  $X$  v.a. reale la cui legge di probabilità coincida con  $Q$ .

Sia  $\Omega = \mathbb{R}$ ;  $F = B(\mathbb{R})$  e  $P = Q$  e consideriamo  $X: \mathbb{R} \rightarrow \mathbb{R}$  l'identità.

$$\text{Allora } P_X = X(P) = Q$$

Questo significa che non dobbiamo studiare ogni volta la variabile aleatoria associata in quanto sappiamo che ne esiste sicuramente almeno una che induca quella legge di probabilità.

**Teorema (Integrazione rispetto ad una probabilità immagine) (D26):**

Sia  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  boreliana.

$\varphi$  è integrabile rispetto a  $P_X \leftrightarrow \varphi \circ X$  è integrabile rispetto a  $P$ .

$$\text{In tal caso vale: } \int_{\mathbb{R}} \varphi(x) dP_X(x) = \int_{\Omega} \varphi(X(\omega)) dP(\omega)$$

**Osservazione:**

Le definizioni di speranza, varianza, etc. coincidono con quelle date nella prima parte.

**Osservazione:**

Consideriamo il caso di variabili doppie  $(X, Y)$  che si può generalizzare al caso di variabili  $n$ -esime.

**Definizione (Variabile doppia):**

Prendiamo un'applicazione  $(X, Y): (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$  misurabile |  $X, Y$  siano due funzioni definite su  $\Omega$  a valori in  $\mathbb{R}$

**Proposizione (D27):**

La coppia  $(X, Y)$  è una v.a. (misurabile come applicazione a valori in  $\mathbb{R}^2$ )  $\leftrightarrow$  entrambe le componenti  $X, Y$  sono variabili aleatorie reali.

**Osservazione (Indipendenza):**

Rimane invariata dal caso discreto la definizione di indipendenza.

**Proposizione:**

Se  $X, Y$  sono indipendenti e  $f, g$  sono due funzioni boreliane allora anche  $f \circ X, g \circ Y$  sono indipendenti.

**Definizione (Probabilità prodotto):**

Siano  $P, Q$  due probabilità su  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , si chiama probabilità prodotto ( $P \otimes Q$ ) la probabilità su  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ , tale che presi due sottoinsiemi boreliani  $A, B$  di  $\mathbb{R}$  allora:

$$P \otimes Q(A \times B) = P(A) \cdot Q(B)$$

**Osservazione:**

Questa probabilità è unica.

**Teorema di Fubini-Tonelli:**

$$\int \int_{\mathbb{R}^2} \varphi(x, y) dP \otimes Q(x, y) = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} \varphi(x, y) dQ(y) \right] dP(x)$$

**Teorema (D28):**

Supponiamo che  $X, Y$  siano indipendenti e dotate di momento primo, anche  $XY$  ha valore atteso e vale la formula:

$$E[XY] = E[X] \cdot E[Y]$$

**Definizione (Variabile aleatorie con densità):**

$X$  v.a. reale ha densità  $f$  se la sua legge di probabilità  $P_X$  ha densità  $f$ .

**Equivalente:**

$\forall A$  boreliano vale la formula:  $P\{X \in A\} = P_X(A) = \int_A f(x)dx$

**Osservazione:**

Una variabile aleatoria con densità ha come funzione di ripartizione:  $F(x) = \int_{-\infty}^x f(x)dx$

**Osservazione:**

La definizione di variabile aleatoria con densità può essere generalizzata al caso  $n$ -dimensionale.

**Attenzione:**

Modificare la densità su di un insieme trascurabile (Secondo la misura di Lebesgue) non modifica l'integrale. Sarebbe più corretto dunque parlare di famiglia di densità  $f$  associate.

**Proposizione (Pratica) (D29):**

Sia  $X$  una variabile aleatoria reale. Sono equivalenti:

1.  $X$  ha densità  $f$
2.  $\forall \varphi$  reale, boreliana, limitata (Per garantire la finitezza) vale la formula:  $E[\varphi(X)] = \int_{\mathbb{R}} \varphi(x)f(x)dx$

**Proposizione (Densità marginali da densità congiunta) (D30):**

Sia  $(X, Y)$  una variabile doppia con densità  $f(x, y)$ , allora le componenti  $X, Y$  ammettono densità  $f_1, f_2$  che soddisfano:

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y)dy ; f_2(y) = \int_{-\infty}^{+\infty} f(x, y)dx$$

**Attenzione:**

Come prima il viceversa è falso, non è possibile conoscendo le densità marginali delle componenti ricavare la densità congiunta di cui non possiamo garantire nemmeno l'esistenza.

**Formula della convoluzione (D31):**

Siano  $X, Y$  due variabili aleatorie indipendenti con densità rispettivamente  $f_1, f_2$ .

Allora la somma  $(X + Y)$  ha densità  $g(x) = \int_{-\infty}^{+\infty} f_1(x - y) \cdot f_2(y)dy$

**Idea:**

Ci stiamo interrogando su cosa succede alle variabili aleatorie se modifichiamo in maniera stabile (Mediante diffeomorfismi) la loro densità.

**Definizione (Diffeomorfismo):**

Un diffeomorfismo è un'applicazione bigettiva fra due aperti  $A$  e  $B$  di  $\mathbb{R}^k$  che sia differenziabile con inversa differenziabile.

**Proposizione (D32):**

Sia  $X$  una v.a. con densità  $f$  diversa da 0 su un aperto  $A \subseteq \mathbb{R}$  e sia  $h: A \rightarrow B$  un diffeomorfismo.

Consideriamo la variabile  $Y = h(X)$ , essa ha densità  $g \mid g(y) = \begin{cases} 0 & \text{se } y \notin B \\ f(h^{-1}(y)) \cdot \left| \frac{d h^{-1}(y)}{dy} \right| & \text{se } y \in B \end{cases}$

**Generalizzazione:**

Nel caso n-dimensionale sostituiamo  $\left| \frac{d h^{-1}(y)}{dy} \right|$  con il valore assoluto del determinante della matrice Jacobiana delle funzione  $h^{-1}$

**Notazione (Matrice Jacobiana):**

La matrice Jacobiana di una funzione  $f$  è la matrice delle derivate parziali:

$$\begin{pmatrix} \frac{df_1}{dx_1} & \dots & \frac{df_n}{dx_1} \\ \vdots & & \vdots \\ \frac{df_1}{dx_n} & \dots & \frac{df_n}{dx_n} \end{pmatrix}$$

**Esempio carino:**

Un esempio di densità da poter assegnare è la **densità uniforme** sull'intervallo  $[0,1]$  definita come

$$f(x) = \begin{cases} 1 & \text{per } 0 < x < 1 \\ 0 & \text{altrimenti} \end{cases}$$

Scegliamo  $X$  v.a. con densità  $f$ .

Possiamo allora definire la v.a.  $Y = \log(X)$  con densità  $g$  definita grazie alla proposizione precedente come:

$$g(y) = \begin{cases} e^y & \text{se } y < 0 \\ 0 & \text{se } y \geq 0 \end{cases}$$

**Esempio applicazione:**

Sia  $(X, Y)$  una variabile doppia con densità  $f$  diversa da 0 sull'aperto  $A$  di  $\mathbb{R}^2$ .

Consideriamo  $h: A \rightarrow B$  diffeomorfismo, definiamo  $(U, V) = h(X, Y)$ .

La coppia  $(U, V)$  per la proposizione precedente ha una densità  $g$  che si annulla al di fuori di  $B$  mentre su  $B$  soddisfa:

$$g(u, v) = f(x(u, v), y(u, v)) \cdot \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix}$$

## Esempi di densità:

### Densità uniforme:

La densità uniforme sull'intervallo  $]a, b[$  è una funzione che è costante su quell'intervallo e nulla fuori.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{per } a < x < b \\ 0 & \text{altrimenti} \end{cases}$$

### **Proprietà:**

Data  $X$  v.a. con densità uniforme:

$$E[X] = \frac{a+b}{2}$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

### Densità gamma:

La densità gamma di parametri  $r, \lambda > 0$  indicata con  $\Gamma(r, \lambda)$  è definita come:

$$f(x) = \begin{cases} \frac{1}{\Gamma(r)} \cdot \lambda^r \cdot x^{r-1} \cdot e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

### **Osservazione:**

Quando  $r = 1$  allora  $\Gamma(1, \lambda)$  è detta densità esponenziale di parametro  $\lambda$ .

Attenzione:

$$\Gamma(n) = (n - 1)!$$

### **Proprietà:**

Data  $X$  v.a. con densità gamma e  $\beta > 0$ :

$$E[X^\beta] = \frac{\Gamma(r+\beta)}{\Gamma(r) \cdot \lambda^\beta}$$

$$\text{In particolare: } E[X] = \frac{r}{\lambda}$$

### **Proposizione (D33):**

Se  $X \sim \Gamma(r_1, \lambda)$  e  $Y \sim \Gamma(r_2, \lambda)$  sono indipendenti, allora  $(X + Y) \sim \Gamma(r_1 + r_2, \lambda)$

### Legge di Fisher:

Siano  $C_n \sim \chi^2(n)$ ;  $C_m \sim \chi^2(m)$  v.a. indipendenti, si chiama legge di Fisher  $F_{n,m}$  la legge di:

$$\frac{C_n/n}{C_m/m}$$

**Densità:**

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ c(n, m) \cdot \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}} & \end{cases}$$

**Densità gaussiana (o normale  $N(0, 1)$ ):**

La densità gaussiana su  $\mathbb{R}$  è la funzione:

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

**Osservazione:**

Segue dal fatto che  $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$

**Proprietà:**

Questa densità ha funzione di ripartizione:  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$

Data  $X$  v.a. con densità gaussiana:

$$E[X] = 0$$

$$\text{Var}(X) = E[X^2] = 1$$

**Definizione (Variabile Gaussiana):**

$X$  v.a. ha legge gaussiana  $N(m, \sigma^2)$  se  $\frac{X-m}{\sigma}$  ha legge standardizzata  $N(0,1)$

**Proprietà:**

$X$  si può scrivere nella forma  $X = \sigma \cdot Y + m$

Dove  $Y$  ha distribuzione  $N(0,1)$  e  $X$  ha densità  $g(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}$

$$E[X] = m$$

$$\text{Var}(X) = \sigma^2$$

**Proposizione (D34):**

Se  $X \sim N(m_1, \sigma_1^2)$ ;  $Y \sim N(m_2, \sigma_2^2)$  sono indipendenti, allora  $(X + Y) \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$

**Grafico:**

Il grafico della densità associata alla variabile aleatoria  $X = \sigma \cdot Y + m$  è simmetrico (A campana) con punto di massimo in  $m$  e due flessi nei punti  $m \pm \sigma$

### Legge chi-quadro:

Si dice legge chi-quadro a  $n$  gradi di libertà ( $\chi^2(n)$ ) la legge  $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

#### **Idea:**

Si studia questa legge in quanto è quella della variabile  $X_1^2 + \dots + X_n^2$  con  $(X_1, \dots, X_n)$  indipendenti gaussiane  $N(0,1)$

#### **Attenzione (Tavole):**

Esistono tavole per il calcolo approssimato della legge.

La notazione utilizzata dalle tavole è la seguente:

$$\chi^2_{(\alpha,n)} := \alpha\text{-quantile della legge } \chi^2(n), \text{ ossia il valore per il quale (Assegnata } X \sim \chi^2(n) \text{ v.a.)}$$
$$P\{X \leq \chi^2_{(\alpha,n)}\} = \alpha$$

#### **Definizione (Quantile):**

Data una funzione di ripartizione  $F$  ed un numero  $\alpha \in ]0,1[$  si chiama  $\alpha$ -quantile di  $F$  il numero definito come:  $r_\alpha = \inf\{x \in \mathbb{R} \mid F(x) > \alpha\}$

#### **Idea:**

In pratica è il minimo valore per il quale la probabilità che accada un evento fino a quel valore sia maggiore dell'indice fissato  $\alpha$  (è l'integrale fino a quel punto)

### Legge di Student:

Siano  $X \sim N(0,1)$ ;  $Y \sim \chi^2(n)$  indipendenti, la legge di Student a  $n$  gradi di libertà ( $T(n)$ ) è la legge di:

$$\frac{\sqrt{n} \cdot X}{\sqrt{Y}}$$

#### **Idea:**

Studiamo questa legge in quanto la incontreremo frequentemente studiando i modelli gaussiani.

#### **Osservazione:**

Se  $T$  è una legge di Student, allora è simmetrica (Ossia  $T$  e  $-T$  sono equidistribuite)

#### **Densità:**

La densità di questa legge è la funzione  $g(x) = c_n \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$  con  $c_n$  costante opportuna.

#### **Attenzione (Tavole):**

Anche per la legge di Student esistono le tavole in funzione di  $n$  (Gradi di libertà) e di  $\alpha$  (Quantile).

Dunque  $t_{(\alpha,n)}$  è il valore dell' $\alpha$ -quantile della legge  $T(n)$

#### **Osservazione ( $T$ simmetrica):**

$$t_{(\alpha,n)} = -t_{(1-\alpha,n)}$$

Quindi se cerchiamo  $t$  numero tale che  $P\{|T| > t\} = \alpha$ , allora  $t = t_{(1-\frac{\alpha}{2},n)}$

**Definizione (Convergenza in probabilità):**

Si dice che la successione di variabili aleatorie  $(X_n)_{n \geq 1}$  converge in probabilità alla v.a.  $X$  se,  $\forall \varepsilon > 0$  si ha:

$$\lim_{n \rightarrow +\infty} P\{|X_n - X| > \varepsilon\} = 0$$

**Teorema (Legge dei grandi numeri) (D35):**

Sia  $X_1, X_2, \dots$  una successione di variabili aleatorie dotate di momento secondo, incorrelate; supponiamo che  $E[X_i] = m \forall i$  (Tutte hanno la stessa speranza) e che esista una costante  $K \mid \text{Var}(X_i) \leq K \forall i$ .

Allora posto  $S_n = X_1 + \dots + X_n$ , la successione  $\left(\frac{S_n}{n}\right)_{n \geq 1}$  converge in probabilità ad  $m$ .

**Criterio di convergenza 1:**

Sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie dotate di momento secondo e supponiamo che:

$$\lim_{n \rightarrow +\infty} E[X_n] = c ; \lim_{n \rightarrow +\infty} \text{Var}(X_n) = 0$$

Allora la successione converge in probabilità a  $c$

**Criterio di convergenza 2:**

Sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie e siano  $F_n(\cdot)$  le relative funzioni di ripartizione.

Sono equivalenti le seguenti affermazioni:

1.  $(X_n)_{n \geq 1}$  converge in probabilità a  $c$
2. Per  $x < c, \lim_{n \rightarrow +\infty} F_n(x) = 0$  e per  $x > c, \lim_{n \rightarrow +\infty} F_n(x) = 1$

**Criterio di convergenza 3 (D36):**

Sia  $(X_n)_{n \geq 1}$  una successione di variabili aleatorie convergente in probabilità a  $c$  e sia  $g$  una funzione boreliana continua nel punto  $c$ .

Allora  $Y_n = g(X_n)$  converge in probabilità a  $g(c)$

**Definizione (Convergenza in legge):**

Si dice che la successione di v.a.  $(X_n)_{n \geq 1}$  converge in legge (o in distribuzione) alla v.a.  $X$  se  $\forall f: \mathbb{R} \rightarrow \mathbb{R}$  continua e limitata si ha:

$$\lim_{n \rightarrow +\infty} E[f(X_n)] = E[f(X)]$$

**Proposizione (D37):**

Siano  $X_n$  e  $X$  v.a.,  $F_n$  e  $F$  le relative funzioni di ripartizione con  $F$  continua (La legge di  $X$  si dice **diffusa**). Allora sono equivalenti le seguenti affermazioni:

1.  $(X_n)_{n \geq 1}$  converge in legge a  $X$
2.  $\forall x \in \mathbb{R}$  si ha,  $\lim_{n \rightarrow +\infty} F_n(x) = F(x)$

**Teorema (Limite centrale di Paul Lévy):**

Sia  $X_1, X_2, \dots$  una successione di variabili indipendenti equidistribuite dotate di momento primo  $\mu$  e di varianza  $\sigma^2$ , posto  $S_n = X_1 + \dots + X_n$ , la successione:

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} = \sqrt{n} \cdot \left(\frac{\bar{X}_n - \mu}{\sigma}\right)$$

Converge in legge alla variabile gaussiana  $N(0,1)$

## Inferenza statistica su uno spazio di Probabilità generale:

### **Idea:**

Riprendere le nozioni già introdotte nel caso in cui si stia lavorando con spazi di probabilità numerabili e generalizzarle ad uno spazio di probabilità generale.

In particolare studieremo i modelli con densità, quelli caratterizzabili mediante una densità di cui conosciamo già numerose proprietà.

### Definizione (Modello statistico):

Un modello statistico è una terna  $(\Omega, F, (P^\theta, \theta \in \Theta))$  con  $\Omega$  insieme,  $F$  tribù su  $\Omega$  e  $(P^\theta, \theta \in \Theta)$  una famiglia di probabilità su  $(\Omega, F)$

### Definizione (Modello con densità):

Un modello statistico che soddisfi le seguenti condizioni:

1.  $\Omega$  è uno spazio euclideo  $\mathbb{R}^n$  (o un sottoinsieme numerabile di uno spazio euclideo).
2.  $F$  è la  $\sigma$ -algebra di Borel su  $\Omega$
3. Le probabilità  $P^\theta$  ammettono densità (Rispetto alla misura di Lebesgue  $n$ -dimensionale)

### Definizione (Verosimiglianza):

Assegnato un modello statistico  $(\Omega, F, (P^\theta, \theta \in \Theta))$  la verosimiglianza è una funzione:

$L: \Theta \times \Omega \rightarrow \mathbb{R}^+$  | fissato  $\theta$  allora  $L(\theta, \cdot)$  sia una versione della densità di  $P^\theta$  (Ossia che appartenga alla famiglia delle funzioni che a meno di insiemi trascurabili secondo misura di Lebesgue siano la densità della probabilità associata al parametro).

### **Osservazione:**

Non c'è incongruenza con la definizione data nel caso numerabile in quanto la funzione che associa  $\omega \rightarrow P^\theta(\{\theta\})$  è la densità di  $P^\theta$  rispetto ad  $m$  (Con  $m$  la misura che conta i punti di un sottoinsieme)

### Definizione (Campione):

Data  $(f(\theta, \cdot), \theta \in \Theta)$  una famiglia parametrizzata di densità di probabilità su  $\mathbb{R}$ , si chiama **campione** di taglia  $n$  e densità  $f(\theta, \cdot)$  una famiglia di variabili aleatorie indipendenti, equidistribuite, aventi tutte densità  $f(\theta, \cdot)$  (Sotto  $P^\theta$ ).

### **Costruzione canonica di un campione:**

Si prende  $\Omega = \mathbb{R}^n$  e si considera come verosimiglianza la funzione:  $L(\theta, x_1, \dots, x_n) = \prod_{i=1}^n f(\theta, x_i)$

Le variabili  $X_i$  sono definite come le proiezioni di indice  $i$ .

Ponendo su  $\Omega$  la probabilità  $P^\theta$  definita dalla densità  $L(\theta, \cdot)$  queste sono indipendenti e di densità  $f(\theta, \cdot)$

**Osservazione:**

Stima, stima corretta, rischio, regione di fiducia, test etc. vengono estese banalmente al caso di uno spazio di probabilità generale.

Valgono gli stessi risultati, in particolare quelli relativi ai modelli esponenziali, alla stima preferibile se in funzione del riassunto esaustivo e il lemma di Neyman-Pearson.

## Inferenza statistica sui modelli gaussiani:

### Idea

Studiamo in particolare i modelli gaussiani perché sono i più comuni in ambito biologico. Inoltre considerando variabili aleatorie date da combinazioni di variabili di densità gaussiano riusciamo a coprire un numero impressionante di casi.

Osserviamo inoltre che sebbene possa sembrare in alcuni casi assurdo l'utilizzo di una densità gaussiana dal punto di vista numerico otteniamo valori sensati.

In pratica la gaussiana centrata a 1.80 e di varianza 100 rappresentante le altezze medie delle persone in Italia dal punto di vista formale può assumere anche valori negativi (Altezza -3 cm) ma possiamo sfruttare il seguente risultato per affermare che a meno di eventi abbastanza trascurabili gli eventi sono compresi fra 1.45 e 2.15

### Osservazione:

I valori di una v.a. con densità gaussiana  $N(0,1)$  sono compresi fra  $-3.5$  e  $3.5$  con probabilità  $> 0.999$

In generale data una variabile aleatoria con densità  $N(m, \sigma^2)$  i suoi valori sono compresi (A meno di eventi di probabilità minore di  $10^{-3}$ ) fra  $m - 3.5 \cdot \sigma$  e  $m + 3.5 \cdot \sigma$

### Lemma (D38):

Sia  $X = (X_1, \dots, X_n)$  un vettore aleatorio formato da  $n$  v.a. indipendenti con densità  $N(0,1)$ .

Sia  $A$  una matrice ortogonale di dimensione  $n$  (Quindi una matrice di cambio di base ortogonale) e sia  $Y = AX$

Allora le componenti  $(Y_1, \dots, Y_n)$  sono indipendenti con densità  $N(0,1)$

### Proposizione (Media) (D39):

Siano  $(X_1, \dots, X_n)$  v.a. indipendenti con densità  $N(0,1)$ , definiamo  $\bar{X} := \frac{X_1 + \dots + X_n}{n}$

Allora valgono i seguenti risultati:

1. Le variabili  $\bar{X}$  e  $\sum_{i \leq n} (X_i - \bar{X})^2$  sono indipendenti.
2.  $\bar{X}$  ha densità  $N\left(0, \frac{1}{n}\right)$  e  $\sum_{i \leq n} (X_i - \bar{X})^2$  ha densità  $\chi^2(n-1)$
3. La variabile  $\sqrt{n} \cdot \sqrt{n-1} \cdot \frac{\bar{X}}{\sqrt{\sum_{i \leq n} (X_i - \bar{X})^2}}$  ha densità di Student  $T(n-1)$

### Definizione:

Sia  $(X_1, \dots, X_n)$  un campione di  $n$  v.a., definiamo:  $S^2 = \frac{\sum_{i \leq n} (X_i - \bar{X})^2}{n-1}$

### Osservazione:

$$S = \sqrt{\frac{\sum_{i \leq n} (X_i - \bar{X})^2}{n-1}}$$

**Teorema (D40):**

Siano  $(X_1, \dots, X_n)$  v.a. indipendenti con densità  $N(m, \sigma^2)$ . Allora valgono i seguenti risultati:

1. Le variabili  $\bar{X}$  e  $S^2$  sono indipendenti.
2.  $\bar{X}$  ha densità  $N\left(m, \frac{\sigma^2}{n}\right)$  e  $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$  ha densità  $\chi^2(n - 1)$
3. La variabile  $\frac{\sqrt{n}(\bar{X} - m)}{S}$  ha densità di Student  $T(n - 1)$

## Studio di un modello statistico di taglia $n$ e densità $N(m, \sigma^2)$ :

### **Idea:**

Per descrivere questa densità necessitiamo di entrambi i parametri, di conseguenza se uno di questi non è noto a priori dobbiamo avere due stime.

Consideriamo come modello statistico un campione di taglia  $n$  e densità  $N(m, \sigma^2)$  sullo spazio  $\Omega = \mathbb{R}^n$   
Su questo modello l'insieme dei parametri è  $\Theta = \mathbb{R} \times ]0, +\infty[$

### **Definizione (Media/Varianza nota):**

Si dice che la media è nota se il parametro  $m$  è fisso, in questo caso  $\Theta = ]0, +\infty[$

Si dice che la varianza è nota se il parametro  $\sigma^2$  è fisso, in questo caso  $\Theta = \mathbb{R}$

### ***Esempio:***

*La media è nota se stiamo prendendo un campione ristretto da una popolazione di cui conosciamo già le proprietà, ad esempio se prendiamo i pazienti di Pisa affetti da una malattia già studiata sulla popolazione italiana.*

### **Verosimiglianza:**

$$L(m, \sigma^2; x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot \sigma^n} \cdot \exp\left(-\frac{\sum_i (x_i - m)^2}{2 \cdot \sigma^2}\right) = (2\pi)^{-\frac{n}{2}} \cdot \exp\left(-\frac{\sum_i x_i^2}{2 \cdot \sigma^2} + \frac{m}{\sigma^2} (\sum_i x_i) - \frac{n \cdot m^2}{2 \cdot \sigma^2} - n \cdot \log \sigma\right)$$

### **Riassunto esaustivo:**

Si ottiene un riassunto esaustivo con la v.a. doppia  $(\sum_i X_i, \sum_i X_i^2)$

### **Osservazione:**

Se la media è nota un riassunto esaustivo è:  $\sum_i (X_i - m)^2$

Se la varianza è nota un riassunto esaustivo è:  $\sum_i X_i$

**Proprietà pratica (Stima di massima verosimiglianza):**

Valgono le seguenti stime di massima verosimiglianza:

$$\hat{m} = \bar{X}$$

$$\hat{\sigma}^2 = \begin{cases} \frac{\sum_i (X_i - m)^2}{n} & \text{se } m \text{ è nota} \\ \frac{\sum_i (X_i - \bar{X})^2}{n} & \text{se } m \text{ è sconosciuta} \end{cases}$$

**Osservazione:**

Deriva dal fatto che stiamo cercando i punti di massimo (Una volta rispetto ad  $m$  e l'altra rispetto a  $\sigma$ ) dell'espressione:  $\left[-\frac{\sum_i x_i^2}{2 \cdot \sigma^2} + \frac{m}{\sigma^2} (\sum_i x_i) - \frac{n \cdot m^2}{2 \cdot \sigma^2} - n \cdot \log \sigma\right]$ . Dopo aver studiato le espressioni al limite si studiano le derivate parziali e si uguagliano a 0.

$$\begin{cases} 0 = \frac{d[\dots]}{dm} = \frac{\sum_i x_i}{\sigma^2} - \frac{nm}{\sigma^2} \\ 0 = \frac{d[\dots]}{d\sigma} = \frac{\sum_i (x_i - m)^2}{\sigma^3} - \frac{n}{\sigma} \end{cases}$$

Da questo ricaviamo i valori sopra descritti.

**Osservazione (Stime consistenti):**

Queste sono tutte stime corrette in quanto la densità gaussiana  $N(m, \sigma^2)$  si può scrivere nella forma:

$c(m, \sigma^2) \cdot \exp\left(-\frac{x^2}{2 \cdot \sigma^2} + \frac{m}{\sigma^2} \cdot x\right) e^{-\frac{x^2}{2 \cdot \sigma^2} + \frac{m}{\sigma^2} \cdot x}$  è in corrispondenza biunivoca con il parametro naturale  $(m, \sigma^2)$ . Quindi è un modello esponenziale.

**Osservazione (Stime corrette):**

$\bar{X}$  è una stima corretta della speranza.

$\frac{\sum_i (X_i - \bar{X})^2}{n}$  non è una stima corretta della varianza (Ha legge  $\chi^2(n-1)$  e dunque ha speranza  $(n-1)$ )

Una stima corretta della varianza sarà dunque:  $S^2 = \frac{\sum_{i \leq n} (X_i - \bar{X})^2}{n-1}$

## Test sui campioni gaussiani:

### **Idea:**

Vogliamo studiare come effettuare un test sulla media nel caso di varianza nota/sconosciuta e come effettuarne uno sulla varianza.

Supponiamo dunque assegnato un campione  $(X_1, \dots, X_n)$  di taglia  $n$  e densità gaussiana.

### **Caso 1 (Test per la media con varianza nota):**

In questi casi si sfrutta il fatto che sotto  $P^m$   $\bar{X}$  ha densità  $N\left(m, \frac{\sigma^2}{n}\right)$  o equivalentemente  $\frac{\sqrt{n} \cdot (\bar{X} - m)}{\sigma}$  ha densità  $N(0,1)$

### **Esempio:**

Trovare un intervallo di fiducia al livello 0.95 per la media di un campione gaussiano con varianza nota.

*Osserviamo che abbiamo indicato una funzione del parametro e della variabile  $\bar{X}$  la cui legge non dipende dal parametro  $m$  (condizioni per il metodo della quantità di pivot).*

*Cerchiamo un intervallo di fiducia della forma  $[\bar{X}(w) - d, \bar{X}(w) + d]$  con  $d$  |*

$$P^m\{|\bar{X} - m| > d\} = P^m\left\{\frac{\sqrt{n}}{\sigma} \cdot |\bar{X} - m| > \frac{d \cdot \sqrt{n}}{\sigma}\right\} \leq 0.05$$

*Perché sia il più piccolo possibile imponiamo l'uguaglianza (Ricordando che  $\frac{\sqrt{n}}{\sigma} \cdot |\bar{X} - m|$  ha densità  $N(0,1)$ , quindi  $\frac{d \cdot \sqrt{n}}{\sigma} = q_{0.975} = 1.96$ .*

*L'intervallo è dunque  $\bar{X}(w) \pm \frac{1.96 \cdot \sigma}{\sqrt{n}}$*

### **Osservazione (Metodo della quantità di pivot):**

Si parla di metodo della quantità di pivot quando si individua una funzione di una v.a.  $X$  e del parametro  $\theta$  che sia invertibili rispetto al parametro  $\theta$  e la cui legge di probabilità non dipenda da  $\theta$

### **Esempio test unilatero:**

Individuare la regione critica di un test al livello 0.02 della forma  $H_0) m \leq m_0$  contro  $H_1) m > m_0$ , con varianza nota.

*Sia  $m_1 < m_2$  e scriviamo il rapporto delle verosimiglianze:*

$$\frac{L(m_2; x_1, \dots, x_n)}{L(m_1; x_1, \dots, x_n)} = \exp\left(\frac{m_2 - m_1}{\sigma^2} \cdot (\sum_i x_i) - \frac{n(m_2^2 - m_1^2)}{2 \cdot \sigma^2}\right) \text{ che è crescente rispetto alla v.a. } \bar{X},$$

*dunque la regione critica sarà della forma:  $D = \{\bar{X} \geq c\}$  con  $c$  |  $P^{m_0}\{\bar{X} \geq c\} = 0.02$*

*Se non conosciamo il quantile di questa variabile aleatoria possiamo ricondurci alla forma normalizzata:*

$$0.02 = P^{m_0}\{\bar{X} - m_0 \geq d\} = P^{m_0}\left\{\frac{\sqrt{n}}{\sigma} \cdot (\bar{X} - m_0) \geq \frac{\sqrt{n}}{\sigma} \cdot d\right\}$$

*Quindi  $\frac{\sqrt{n}}{\sigma} \cdot d = q_{0.98} = 2.055$*

*Si rifiuta l'ipotesi se  $\bar{X}(w)$  (La media aritmetica dei valori osservati) supera  $\left(m_0 + \frac{2.055 \cdot \sigma}{\sqrt{n}}\right)$*

## Caso 2 (Test per la media con varianza sconosciuta/Test di Student):

**Idea:**

Conoscendo la varianza potevamo studiare la variabile  $\frac{\sqrt{n}\bar{X}}{\sigma}$  che ha densità  $N\left(\frac{m\sqrt{n}}{\sigma}, 1\right)$

In questo caso sostituiamo a  $\sigma^2$  la sua stima corretta  $S^2$ .

Studiamo dunque:  $\frac{\sqrt{n}\bar{X}}{S} = \sqrt{n} \cdot \sqrt{n-1} \cdot \frac{\bar{X}}{\sqrt{\sum_{i \leq n} (X_i - \bar{X})^2}}$

### Definizione (Legge di Student decentrata):

Si chiama legge di Student a  $n$  gradi di libertà decentrata di  $\alpha$  la legge di:  $\frac{\sqrt{n}\cdot X}{\sqrt{Y}}$

Con  $X \sim N(\alpha, 1)$ ,  $Y \sim \chi^2(n)$

#### **Proprietà:**

Le densità di Student decentrate di  $\alpha$  sono a rapporto di verosimiglianza crescente rispetto alla v.a. identità ( $T(x) = x$  su  $\mathbb{R}$ )

#### **Osservazione:**

La v.a.  $\frac{\sqrt{n}\bar{X}}{S}$  (Sotto  $P^{m,\sigma^2}$ ) ha legge di Student  $T(n-1)$  decentrata di  $\frac{m\sqrt{n}}{\sigma}$

**Viene da:**

$$\frac{\sqrt{n}\bar{X}}{S} = \sqrt{n-1} \cdot \frac{\frac{\sqrt{n}\bar{X}}{\sigma}}{\sqrt{\frac{\sum_{i \leq n} (X_i - \bar{X})^2}{\sigma^2}}}$$

### **Test di Student unilatero**

Consideriamo al livello  $\alpha$  la regione critica di un test dell'ipotesi  $H_0) m \leq 0, \sigma$  qualsiasi contro  $H_1) m > 0 \sigma$  qualsiasi.

*Il test può essere scritto come:*

$$H_0) \frac{m}{\sigma} \leq 0 \text{ contro } H_1) \frac{m}{\sigma} > 0$$

*Avendo individuato una v.a.  $\left(\frac{\sqrt{n}\bar{X}}{S}\right)$  il cui valore dipende esclusivamente da  $\frac{m}{\sigma}$  la cui distribuzione di probabilità ha legge di Student  $T(n-1)$  decentrata di  $\frac{m\sqrt{n}}{\sigma}$  ed è a rapporto di verosimiglianza crescente rispetto a  $\frac{m}{\sigma}$ . Dunque la regione critica è della forma:*

$$D = \left\{ \frac{\sqrt{n}\bar{X}}{S} \geq d \right\} = \left\{ w \in \Omega \mid \frac{\sqrt{n}\bar{X}(w)}{S(w)} \geq d \right\} \text{ con } d \mid P^{m,\sigma^2} \left\{ \frac{\sqrt{n}\bar{X}}{S} \geq d \right\} = \alpha$$

*Scegliamo dunque  $d = t_{(1-\alpha, n-1)}$*

*In base ai corollari del Lemma di Neyman-Pearson questo test è ottimale fra quelli basati*

*sull'osservazione della variabile  $\frac{\sqrt{n}\bar{X}}{S}$*

**Osservazione pratica:**

Se il test è della forma  $H_0) m \leq m_0, \sigma$  qualsiasi contro  $H_1) m > m_0, \sigma$  qualsiasi non si può utilizzare il trucchetto di prima su  $\frac{m}{\sigma}$ .

Dunque ci riconduciamo al caso precedente considerando le variabili  $(X_i - m_0)$  che hanno legge  $N(m - m_0, \sigma^2)$ .

Sviluppando i calcoli (A questo punto come prima) otteniamo una regione critica della forma:

$$D = \left\{ \frac{\sqrt{n} \cdot (\bar{X} - m_0)}{s} \geq t_{(1-a, n-1)} \right\}$$

**Esempio 2:**

Consideriamo il test al livello  $\alpha$   $H_0) m = 0, \sigma$  qualsiasi contro  $H_1) m \neq 0, \sigma$  qualsiasi

Come prima si arriva ad una regione critica della forma  $D = \left\{ \frac{\sqrt{n} \cdot |\bar{X}|}{s} \geq d \right\}$  con  $d |$

$$P^{0, \sigma^2} \left\{ \frac{\sqrt{n} \cdot |\bar{X}|}{s} \geq d \right\} = \alpha$$

Quindi per ottimizzare il risultato scegliamo  $d = t_{(1-\frac{\alpha}{2}, n-1)}$

**Osservazione:**

Nel caso simile nel quale l'ipotesi è  $H_0) m = m_0, \sigma$  qualsiasi arriviamo ad una regione

$$\text{critica della forma: } D = \left\{ \frac{\sqrt{n} \cdot |\bar{X} - m_0|}{s} \geq t_{(1-\frac{\alpha}{2}, n-1)} \right\}$$

**Esempio 3 (Pratico):**

Sapendo che il tempo medio di guarigione da una polmonite con i farmaci usuali è di 14 giorni sperimentiamo un nuovo farmaco su 17 pazienti e ne rileviamo i tempi di guarigione  $x_1, \dots, x_{17}$  ottenendo i seguenti risultati:

$$\sum_{i=1}^{17} x_i = 197; \sum_{i=1}^{17} x_i^2 = 2596$$

Possiamo affermare che il nuovo farmaco non è più efficace?

**Soluzione:**

Consideriamo  $x_1, \dots, x_{17}$  come valori osservati da un campione  $X_1, \dots, X_{17}$  con legge gaussiana  $N(m, \sigma^2)$  sul quale effettuiamo il test di ipotesi:

$H_0) m \geq 14, \sigma$  qualsiasi contro  $H_1) m < 14, \sigma$  qualsiasi

$$\text{Otteniamo la regione critica } D = \left\{ \frac{\sqrt{17} \cdot (\bar{X} - 14)}{s} \geq t_{(a, 16)} \right\}$$

**Osservazione:**  $t_{(0.05, 16)} = -1.746$ ;  $t_{(0.01, 16)} = -2.58$

Calcoliamo  $\bar{x} = \frac{\sum_i x_i}{17} = 11.58$ ;  $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{2} = 19.56$  e dunque  $\frac{\sqrt{17} \cdot (\bar{x} - 14)}{s} = -2.25$

Perciò (**Attenzione**) l'ipotesi viene rifiutata al livello 0.05 ma accettata al livello 0.01

### **Caso 3 (Test sulla varianza):**

#### **Osservazione 1:**

Per studiare questa categoria di test sfruttiamo la seguente proprietà:

Se  $m$  è noto:  $\frac{\sum_i (X_i - m)^2}{\sigma^2}$  ha densità  $\chi^2(n)$

Se  $m$  è sconosciuto:  $\frac{\sum_i (X_i - \bar{X})^2}{\sigma^2}$  ha densità  $\chi^2(n - 1)$

#### **Osservazione 2:**

$\sum_i (X_i - \bar{X})^2$  ha densità (Sotto  $P^{m_0, \sigma^2}$ )  $f(x) = c(n) \cdot \sigma^{-(n+1)} \cdot x^{\frac{n-3}{2}} \cdot e^{-\frac{x}{2 \cdot \sigma^2}}$

Inoltre queste densità sono a verosimiglianza crescente.

#### **Esempio:**

Consideriamo il test  $H_0) \sigma^2 \leq \sigma_0^2$ ,  $m$  qualsiasi, contro  $H_1) \sigma^2 > \sigma_0^2$ ,  $m$  qualsiasi al livello  $\alpha$ .

Si tratta di un test unilatero che ha una regione critica del tipo  $D = \{\sum_i (X_i - \bar{X})^2 \geq c\}$

Con  $c$  scelto in modo tale che:  $P^{m, \sigma_0^2} \left\{ \frac{\sum_i (X_i - \bar{X})^2}{\sigma_0^2} \geq \frac{c}{\sigma_0^2} \right\} = \alpha$

Siccome  $\frac{\sum_i (X_i - \bar{X})^2}{\sigma_0^2}$  non dipende da  $m$  e, per  $\sigma = \sigma_0$ , è  $\chi^2(n - 1)$  si considera  $\frac{c}{\sigma_0^2} = \chi_{(1-\alpha, n-1)}^2$

Quindi raccolti i dati  $x_1, \dots, x_n$  si rifiuta l'ipotesi se  $\sum_i (x_i - \bar{x})^2 \geq \chi_{(1-\alpha, n-1)}^2 \cdot \sigma_0^2$

## Confronti tra campioni gaussiani indipendenti (Analisi della varianza):

### Idea:

Studiare due campioni indipendenti che non posso accomunare per motivi reali (Ad esempio considero delle persone di due fasce di età distinte rispetto ad un problema funzione dell'età o recupero dei manufatti da due siti archeologici distinti).

Formalmente consideriamo come primo campione  $X_1, \dots, X_n$  di legge  $N(m_1, \sigma_1^2)$  e come secondo  $Y_1, \dots, Y_k$  di legge  $N(m_2, \sigma_2^2)$

Lo spazio lo scegliamo in maniera naturale come  $\Omega = \mathbb{R}^{n+k}$  e l'insieme dei parametri è:

$\Theta = (\mathbb{R}^2 \times ]0, +\infty[^2)$  (Così definiti i parametri sono della forma  $(m_1, m_2, \sigma_1^2, \sigma_2^2)$ )

Verosimiglianza:

$$L(m_1, m_2, \sigma_1^2, \sigma_2^2; x_1, \dots, x_n, y_1, \dots, y_k) = \prod_{i=1}^n f_{m_1, \sigma_1^2}(x_i) \cdot \prod_{j=1}^k f_{m_2, \sigma_2^2}(y_j)$$

Con  $f_{m_1, \sigma_1^2}(x_i)$  densità di  $N(m_1, \sigma_1^2)$ ;  $X_i$  proiezioni della  $i$ -esima coordinata e  $Y_j$  proiezione della  $(n + j)$ -esima coordinata.

### **Esempio 1 (Confronto tra due varianze):**

Studiamo il test  $H_0) \sigma_1^2 \leq \sigma_2^2$  contro  $H_1) \sigma_1^2 > \sigma_2^2$  ad un livello  $\alpha$  prescelto.

**Osservazione:**

La stima corretta di  $\sigma_1^2$  è  $S^2(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ , idem per  $\sigma_2^2$

Dunque  $\frac{S^2(X)}{S^2(Y)} = \frac{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}{\frac{\sum_{j=1}^k (Y_j - \bar{Y})^2}{k-1}}$  che ha legge di Fisher  $F_{n-1, k-1}$

Andando ad intuito rifiutiamo l'ipotesi se il rapporto è troppo grande.

Dunque la regione critica del test è data da:

$$D = \left\{ \frac{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}{\frac{\sum_{j=1}^k (Y_j - \bar{Y})^2}{k-1}} \geq q \right\} \text{ con } q \text{ } (1 - \alpha)\text{-quantile della legge } F_{n-1, k-1}$$

**Definizione (Problema di Behrens-Fisher):**

Si chiama problema di Behrens-Fisher l'individuazione della regione critica del test di ipotesi:

$$H_0) m_1 = m_2 \text{ contro } H_1) m_1 \neq m_2$$

**Lemma (D41):**

Se  $m_1 = m_2$  e  $\sigma_1^2 = \sigma_2^2$  la variabile  $Z_{n,k} = \frac{\bar{X} - \bar{Y}}{\sqrt{\sum_{i \leq n} (X_i - \bar{X})^2 + \sum_{j \leq k} (Y_j - \bar{Y})^2}} \cdot \frac{\sqrt{n+k-2}}{\sqrt{\frac{1}{n} + \frac{1}{k}}}$  ha densità di Student  $T(n + k - 2)$

**Conseguenza:**

Risolvere il problema di Behrens-Fisher sotto l'ulteriore ipotesi che  $\sigma_1^2 = \sigma_2^2$  diventa un'estensione del test di Student.

Dunque la regione critica al livello  $\alpha$  sarò  $D = \left\{ |Z_{n,k}| \geq t_{\left(1-\frac{\alpha}{2}, n+k-2\right)} \right\}$  per l'ipotesi  $H_0) m_1 = m_2$  oppure sarò  $D = \left\{ Z_{n,k} \geq t_{(1-\alpha, n+k-2)} \right\}$  per l'ipotesi  $H_0) m_1 \leq m_2$

**Esempio 2 pratico:**

Le misurazioni delle tibie da scheletri dalle tombe etrusche di Cerveteri danno i seguenti risultati:

$$13 \text{ misurazioni, } \bar{x} = 47.2 ; \frac{\sum(x_i - \bar{x})^2}{12} = 7.92$$

mentre analoghe misurazioni dalle tombe di Ladispoli portano a:

$$8 \text{ misurazioni, } \bar{y} = 44.9 ; \frac{\sum(y_j - \bar{y})^2}{7} = 9.27$$

Si può affermare (Al livello 0.05) che gli abitanti di Cerveteri erano effettivamente più alti?

**Soluzione:**

I due campioni possono essere considerati Gaussiani e indipendenti, vogliamo capire se possiamo portarci nelle ipotesi del lemma, eseguiamo dunque un test sulla varianza.

$$H_0) \sigma_1^2 = \sigma_2^2 \text{ contro } H_1) \sigma_1^2 < \sigma_2^2$$

(Il caso minore non lo consideriamo perché vorrebbe dire che il risultato è casuale)

Ricaviamo dalle tavole il quantile  $F_{(0.95, 7, 12)} = 2.91$  e siccome  $\frac{9.27}{7.92} = 1.17 < 2.91$  dunque accettiamo al livello 0.05 l'uguaglianza fra varianze.

Effettuiamo ora il test di ipotesi:

$$H_0) m_1 = m_2 \text{ contro } H_1) m_1 > m_1$$

Osserviamo che i valori per la variabile  $Z_{13,8}$  portano a  $1.761 > 1.729 = t_{(0.95, 19)}$  rifiutiamo l'ipotesi di uguaglianza e concludiamo al livello 0.05 che gli abitanti di Cerveteri erano effettivamente più alti.

**Osservazione:**

In pratica per risolvere questo caso effettua un test del livello richiesto sulla probabilità di trovarci nelle ipotesi semplificate del lemma. Una volta verificato la situazione diviene standard.

## Modelli statistici lineari:

### Definizione (Modello lineare):

Si chiama modello statistico lineare un modello nel quale l'osservazione è data da  $n$  variabili aleatorie  $X_1, \dots, X_n$  che si possono scrivere nella forma:

$$X_i = \sum_{j=1}^k a_{ij} \cdot \theta_j + \sigma \cdot W_i$$

Con le seguenti proprietà:

1.  $k < n, (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$  e  $\sigma > 0$
2. La matrice  $n \times k$ ,  $A = [a_{ij}]$  è di rango massimo, dunque  $A: \mathbb{R}^k \rightarrow \mathbb{R}^n$  è iniettiva.
3. Le variabili  $W_1, \dots, W_n$  sono gaussiane e indipendenti.

### **Notazione:**

Per i modelli lineari utilizziamo anche la notazione:  $\mathbf{X} = A\boldsymbol{\theta} + \sigma\mathbf{W}$

### **Proprietà:**

Le variabili aleatorie che costituiscono l'osservazione non formano un campione in quanto non sono equidistribuite ma sono indipendenti e  $X_i \sim N(\sum_j a_{ij} \cdot \theta_j, \sigma^2)$

### **Insieme dei parametri:**

$$\Theta = \mathbb{R}^k \times ]0, +\infty[$$

### **Verosimiglianza:**

$$L(\boldsymbol{\theta}; \sigma^2; x_1, \dots, x_n) = (2 \cdot \pi)^{-\frac{n}{2}} \cdot \exp\left(-\frac{\|\mathbf{x} - A\boldsymbol{\theta}\|^2}{2 \cdot \sigma^2} - n \cdot \log \sigma\right)$$

### **Lemma (D42):**

Sia  $A: \mathbb{R}^k \rightarrow \mathbb{R}^n$  un'applicazione lineare iniettiva. Dato  $\mathbf{x} \in \mathbb{R}^n$ , il punto  $\mathbf{y} \in \mathbb{R}^k$  che minimizza  $\|\mathbf{x} - A\mathbf{y}\|^2$  è dato da  $\mathbf{y} = U\mathbf{x}$  essendo  $U = (A^t A)^{-1} A^t$

### **Osservazione:**

$AU = P$  proiezione ortogonale da  $\mathbb{R}^n$  sul sottospazio  $A(\mathbb{R}^k)$

### **Stime di massima verosimiglianza:**

La stima di  $\boldsymbol{\theta}$  è  $\hat{\boldsymbol{\theta}}(\mathbf{x}) = U\mathbf{x}$

La stima di  $\sigma^2$  è  $\hat{\sigma}^2 = \frac{\|\mathbf{x} - A\hat{\boldsymbol{\theta}}\|^2}{n} = \frac{\|\mathbf{x} - AU\mathbf{x}\|^2}{n}$

### **Teorema di Gauss-Markov (D43):**

$U\mathbf{x}$  è una stima corretta di  $\boldsymbol{\theta}$  di rischio minimo tra tutte le stime lineari corrette.

Inoltre  $\frac{\|\mathbf{x} - AU\mathbf{x}\|^2}{n-k}$  è una stima corretta di  $\sigma^2$ .