



# Dispense di Inferenza statistica e teoria dell'informazione

Emanuele Izzo, 0307385

Corso di laurea magistrale Informatica  
Università Tor Vergata, Facoltà di Scienze MM.FF.NN.  
30/03/2022

Documento realizzato in L<sup>A</sup>T<sub>E</sub>X



# Indice

0.1	Le basi della probabilità . . . . .	2
<b>1</b>	<b>Introduzione</b>	<b>3</b>
1.1	Il problema del compleanno . . . . .	3
1.2	L'inclusione-esclusione . . . . .	4
1.3	Il problema degli ombrelli . . . . .	5
<b>2</b>	<b>Elementi di probabilità</b>	<b>6</b>
2.1	La teoria degli insiemi . . . . .	6
2.2	Le basi della teoria della probabilità . . . . .	7
2.3	La probabilità condizionata e l'indipendenza . . . . .	10
2.3.1	Regola di Bayes . . . . .	11
2.4	Le variabili aleatorie e le funzioni di distribuzione . . . . .	12
2.4.1	Variabili aleatorie continue e discrete . . . . .	12
2.4.2	Variabili aleatorie identicamente distribuite . . . . .	12
2.5	Le funzioni di densità e di massa . . . . .	13
2.5.1	Funzione di massa . . . . .	13
2.5.2	Funzione di densità . . . . .	13
<b>3</b>	<b>Trasformazioni e valori attesi</b>	<b>14</b>
3.1	Distribuzioni di una funzione di una variabile aleatoria . . . . .	14
3.2	I valori attesi . . . . .	16
3.3	I momenti e la funzione generatrice dei momenti . . . . .	17
<b>4</b>	<b>Famiglie comuni di distribuzioni</b>	<b>19</b>
4.1	Le distribuzioni discrete . . . . .	19
4.1.1	Distribuzione uniforme . . . . .	19
4.1.2	Distribuzione ipergeometrica . . . . .	20
4.1.3	Distribuzione binomiale . . . . .	21
4.1.4	Distribuzione di Poisson . . . . .	23
4.1.5	Distribuzione binomiale negativa . . . . .	25
4.1.6	Distribuzione geoemtrica . . . . .	26
4.2	Le distribuzioni continue . . . . .	29
4.2.1	Distribuzione uniforme . . . . .	29
4.2.2	Distribuzione Gamma . . . . .	31
4.2.3	Distribuzione normale . . . . .	32
<b>5</b>	<b>Varaibili aleatorie multiple</b>	<b>34</b>
5.1	Distribuzione multinomiale . . . . .	35
5.2	Variabili mutualmente indipendenti . . . . .	36

<b>6</b>	<b>Proprietà di un campionamento casuale</b>	<b>38</b>
6.1	Concetti base del campionamento aleatorio . . . . .	38
6.2	Somma di variabili aleatorie da un campionamento aleatorio . . . . .	38
6.3	Campionamento dalla distribuzione normale . . . . .	39
6.4	Concetti di convergenza . . . . .	40
6.4.1	Convergenza di probabilità . . . . .	40
6.4.2	Convergenza di distribuzione . . . . .	41
6.4.3	Metodo Delta . . . . .	43

## 0.1 Le basi della probabilità

# Capitolo 1

## Introduzione

L'**inferenza statistica** (o **statistica inferenziale**) è il procedimento per cui si inducono le caratteristiche di una *popolazione* dall'osservazione di una parte di essa (detta *campione*), selezionata solitamente mediante un esperimento casuale (aleatorio). Si considereranno principalmente campioni casuali semplici di dimensione  $n > 1$ , che possono venire interpretati come  $n$  realizzazioni indipendenti di un esperimento di base, nelle medesime condizioni. Dal momento che si considera un esperimento casuale, si coinvolge il calcolo delle probabilità. Nell'inferenza statistica c'è, in un certo senso, un rovesciamento di punto di vista rispetto al calcolo delle probabilità. Nell'ambito di quest'ultimo, noto il processo di generazione dei dati sperimentali (*modello probabilistico*) siamo in grado di valutare la probabilità dei diversi possibili risultati di un esperimento. Nella statistica il processo di generazione dei dati sperimentali non è noto in modo completo (il processo in questione è, in definitiva, l'oggetto di indagine) e le tecniche statistiche si prefiggono di indurre le caratteristiche di tale processo sulla base dell'osservazione dei dati sperimentali da esso generati.

### 1.1 Il problema del compleanno

Il **problema del compleanno** è formalizzato come segue: sono presenti  $k$  persone all'interno di una stanza, quale è la probabilità che almeno due persone compiano gli anni nello stesso giorno? Per rispondere a questa domanda risulta più semplice calcolare la probabilità che nessuna delle  $k$  persone compia gli anni nello stesso giorno.

$E := "$  $i$  persone compiono gli anni lo stesso giorno"

$$P(E = 0) = \frac{365}{365} \cdot \frac{364}{365} \cdot \dots \cdot \frac{365 - (k - 1)}{365} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{365}\right)$$

$$P(E \geq 1) = 1 - P(E = 0) = 1 - \prod_{i=0}^{k-1} \left(1 - \frac{i}{365}\right)$$

Questo prodotto è incredibilmente difficile da calcolare. È possibile però approssimarlo: usando l'approssimazione  $e^{-x} \cong 1 - x$ , possiamo riscrivere il prodotto sopra in un'altra

forma.

$$\begin{aligned}
 P(E = 0) &= \prod_{i=0}^{k-1} \left(1 - \frac{i}{365}\right) \cong \prod_{i=0}^{k-1} e^{-\frac{i}{365}} = \\
 &= e^{-\sum_{i=0}^{k-1} \frac{i}{365}} = \\
 &= e^{-\frac{1}{365} \sum_{i=0}^{k-1} i} = \\
 &= e^{-\frac{1}{365} \cdot \frac{k(k-1)}{2}} = \\
 &= e^{-\frac{k(k-1)}{730}}
 \end{aligned}$$

Da cui abbiamo che  $P(E \geq 1) = 1 - e^{-\frac{k(k-1)}{730}}$ . Possiamo fare alcune osservazioni dal valore appena calcolato:

- $P(E \geq 1) = \frac{1}{2} \Rightarrow k \cong 23$ , ossia se sono presenti circa 23 persone, la probabilità che almeno due persone compino gli anni lo stesso giorno è del 50%.
- $P(E \geq 1) = \frac{19}{20} \Rightarrow k \cong 48$ , ossia se sono presenti circa 48 persone, la probabilità che almeno due persone compino gli anni lo stesso giorno è del 95%.

## 1.2 L'inclusione-esclusione

L'**inclusione-esclusione** è la generalizzazione della probabilità dell'unione di due eventi; la formula generale è definibile come segue: la somma delle probabilità dei singoli eventi meno la probabilità di tutte le possibili intersezioni a due degli insiemi più la probabilità di tutte le possibili intersezioni a tre degli insiemi, e così via, fino a sommare o sottrarre la probabilità dell'intersezione tra tutti gli eventi. Esiste una dimostrazione interessante per la formula di inclusione-esclusione: partiamo dal definire una variabile aleatoria *indicatrice* o *unitaria* per un certo evento  $A$ .

$$1_A = \begin{cases} 1 & \text{se } A \text{ avviene} \\ 0 & \text{altrimenti} \end{cases}$$

Un'osservazione importante è che la media di una variabile aleatoria unitaria di un evento è uguale alla probabilità dell'evento, ossia  $\mathbb{E}[1_A] = P(A)$ . Possiamo inoltre notare alcune cose riguardo queste variabili aleatorie unitarie:

$$\begin{aligned}
 1_{A^C} &= 1 - 1_A \\
 1_{A \cap B} &= 1_A 1_B \\
 1_{A \cup B} &= 1_A + 1_B
 \end{aligned}$$

Possiamo sfruttare queste variabili aleatorie unitarie in combinazione con le leggi di De Morgan:

$$\begin{aligned}
 P((A_1 \cup A_2 \cup \dots \cup A_n)^C) &= \\
 &= 1 - P(A_1 \cup A_2 \cup \dots \cup A_n) = \\
 &= P(A_1^C \cap A_2^C \cap \dots \cap A_n^C) = \\
 &= \mathbb{E}((1 - 1_{A_1})(1 - 1_{A_2}) \dots (1 - 1_{A_n}))
 \end{aligned}$$

Essendo quest'ultimo però un prodotto tra quantità numeriche può essere espresso come:

$$\mathbb{E}[(1-1_{A_1})(1-1_{A_2})\dots(1-1_{A_n})] = \mathbb{E}[1 - (1_{A_1} + 1_{A_2} + \dots + 1_{A_n}) + (1_{A_1}1_{A_2} + 1_{A_1}1_{A_3} + \dots + 1_{A_1}1_{A_n} + 1_{A_2}1_{A_3} + \dots + 1_{A_{n-1}}1_{A_n}) - \dots \pm 1_{A_1}1_{A_2}\dots 1_{A_n}]$$

Operando poi sulla media si ottiene la formula descritta prima.

### 1.3 Il problema degli ombrelli

Il **problema degli ombrelli** è formalizzato come segue: sono presenti  $n$  persone all'interno di una stanza, ognuno con il proprio ombrello. Tutte le persone posano il proprio ombrello all'interno di un portaombrelli; una volta posati tutti gli ombrelli nel portaombrelli, ogni persona, una alla volta, prende uno degli ombrelli al suo interno casualmente. Quale è la probabilità che nessuno abbia preso il proprio ombrello? Questo problema corrisponde, partendo da una sequenza iniziale di  $n$  elementi, alla probabilità di scegliere una delle  $n!$  permutazioni possibili di essi e avere che un elemento abbia stessa posizione nella permutazione e nella sequenza iniziale. Definiamo le variabili aleatorie  $A_k := \text{"l'elemento } k\text{-esimo è } k\text{"}$  e  $E := \text{"}n \text{ elementi sono nella posizione corretta"}$ , la probabilità a cui siamo interessati è  $P(E = 0) = 1 - P(A_1 \cup A_2 \cup \dots \cup A_n)$ . Poichè gli eventi non sono indipendenti, è necessario utilizzare la formula di inclusione-esclusione; una cosa interessante che si può appuntare è che le intersezioni sono interpretabili come "1 o più elementi sono al punto giusto" (ossia  $E \geq 1$ ) e le relative probabilità sono facili da calcolare data la struttura delle permutazioni: la probabilità che  $k$  sugli  $n$  elementi siano nella posizione giusta corrisponde a fissare una delle possibili combinazioni di  $k$  elementi e permutare i restanti  $n - k$ , e la probabilità di ciò è pari a :

$$P(E = k) = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{n!}{k!(n-k)!} \frac{(n-k)!}{n!} = \frac{1}{k!}$$

Pertanto la probabilità che a noi interessa, considerando i segni delle varie intersezioni, sarà uguale a:

$$P(E = 0) = 1 - \sum_{i=1}^n (-1)^{i-1} P(E = i) = 1 - \left( \sum_{i=1}^n (-1)^{i-1} \frac{1}{i!} \right) = \sum_{i=0}^n (-1)^i \frac{1}{i!} \cong e^{-1}$$

Possiamo notare pertanto che la probabilità che ci interessa si avvicina, con  $n$  crescente, molto rapidamente a  $e^{-1} \cong \frac{1}{3}$

# Capitolo 2

## Elementi di probabilità

### 2.1 La teoria degli insiemi

**Definizione:** L'insieme  $S$  di tutti i possibili risultati di un particolare esperimento è chiamato lo **spazio campionario** dell'esperimento. Un **evento** è una qualsiasi collezione di possibili risultati di un esperimento, ossia un qualsiasi sottoinsieme di  $S$  (incluso  $S$  stesso).

Sia  $A$  un evento, ossia un sottoinsieme di  $S$ . Diciamo che l'evento  $A$  avviene se il risultato dell'esperimento è contenuto nell'insieme  $A$ . Definito un secondo evento  $B$ , possiamo definire le seguenti relazioni e operazioni:

- *Contenimento:*  $A \subset B \Leftrightarrow [x \in A \Rightarrow x \in B]$ .
- *Uguaglianza:*  $A = B \Leftrightarrow A \subset B \text{ e } B \subset A$ .
- *Unione:*  $A \cup B := \{x | x \in A \text{ o } x \in B\}$ .
- *Intersetzione:*  $A \cap B := \{x | x \in A \text{ e } x \in B\}$ .
- *Complementare:*  $A^C := \{x | x \notin A\}$ .



**Teorema (2.1):** Per qualsiasi tre eventi  $A, B, C \in S$ :

- *Commutatività:*

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

- *Associatività:*

$$A \cup (B \cup C) = (A \cup B) \cup C$$

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

- *Leggi distributive:*

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

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

- *Leggi di De Morgan:*

$$(A \cup B)^C = A^C \cap B^C$$

$$(A \cap B)^C = A^C \cup B^C$$

**Definizione:** Due eventi  $A, B \in S$  sono **disgiunti** (o **mutualmente esclusivi**) se  $A \cap B = \emptyset$ . Gli eventi  $A_1, A_2, \dots, A_n \in S$  sono **disgiunti a due a due** se per ogni  $i \neq j$  si ha che  $A_i \cap A_j = \emptyset$ .

**Definizione:** Se  $A_1, A_2, \dots, A_n \in S$  sono disgiunti a due a due e  $\cup_{i=1}^n A_i = S$ , allora la collezione  $A_1, A_2, \dots, A_n$  forma una partizione di  $S$ .

## 2.2 Le basi della teoria della probabilità

**Definizione:** Una collezione di sottoinsieme  $S$  è detta una  **$\sigma$ -algebra**, indicata con  $\mathcal{A}$ , se soddisfa le seguenti tre proprietà:

- $\emptyset \in \mathcal{A}$ .
- Se  $A \in \mathcal{A}$ , allora  $A^C \in \mathcal{A}$ .
- Se  $A_1, A_2, \dots, A_n \in \mathcal{A}$ , allora  $\cup_{i=1}^n A_i \in \mathcal{A}$ .

**Definizione:** Dato un spazio di probabilità  $S$  e una  $\sigma$ -algebra associata  $\mathcal{A}$ , una **funzione di probabilità**  $P$  è una funzione nei reali con dominio  $\mathcal{A}$  che soddisfa:

- $P(A) \geq 0$  per tutti gli  $A \in \mathcal{A}$ .
- $P(S) = 1$ .
- Se  $A_1, A_2, \dots, A_n \in \mathcal{A}$  sono disgiunti a coppie, allora  $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ .

Le tre proprietà date nella definizione sopra sono di solito chiamate *assiomi di probabilità*. Le definizioni assiomatiche non fanno alcun tentativo nel dire quale particolare funzione  $P$  bisogna scegliere.

**Teorema (2.2):** Sia  $S = \{s_1, s_2, \dots, s_n\}$  un insieme finito. Sia  $\mathcal{A}$  una  $\sigma$ -algebra dei sottoinsiemi di  $S$ . Sia  $p_1, p_2, \dots, p_n$  numeri non negativi che sommano a 1. Per ogni  $A \in \mathcal{A}$ ,  $P(A)$  è definita come:

$$P(A) = \sum_{i: s_i \in A} p_i$$

**Teorema (2.2):** Poniamo che  $S$  sia finito. Per ogni  $A \in \mathcal{A}$ ,  $P(A) = \sum_{i: s_i \in A} p_i \geq 0$ , poiché ogni  $p_i \geq 0$ . Pertanto l'assioma 1 è vero. Inoltre:

$$P(S) = \sum_{i: s_i \in S} p_i = \sum_{i=1}^n p_i = 1$$

Pertanto l'assioma 2 è vero. Siano  $A_1, A_2, \dots, A_k$  eventi disgiunti a coppie, allora:

$$P(\cup_{i=1}^k A_i) = \sum_{j: s_j \in \cup_{i=1}^k A_i} p_j = \sum_{i=1}^k \sum_{j: s_j \in A_i} p_j = \sum_{i=1}^k P(A_i)$$

E pertanto anche il terzo assioma è vero.

Il terzo assioma della definizione della funzione di probabilità, detto *assioma di additività numerabile*, è sostituibile con un assioma equivalente, ossia *assioma dell'additività finita*: se  $A \in \mathcal{A}$  e  $B \in \mathcal{A}$  sono disgiunti, allora:

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

**Teorema (2.3):** Se  $P$  è una funzione di probabilità e  $A$  è un qualsiasi insieme in  $\mathcal{A}$ , allora:

- $P(\emptyset) = 0$ .
- $P(A) \leq 1$ .
- $P(A^C) = 1 - P(A)$ .

**Teorema (2.3):** Gli insiemi  $A$  e  $A^C$  formano una partizione dello spazio di campionamento, ossia  $S = A \cup A^C$ . Pertanto:

$$P(A \cup A^C) = P(S) = 1$$

Inoltre,  $A$  e  $A^C$  sono disgiunti, e pertanto:

$$P(A \cup A^C) = P(A) + P(A^C)$$

Da cui abbiamo che:

$$1 = P(A) + P(A^C) \rightarrow P(A^C) = 1 - P(A)$$

Dimostrando il terzo assioma. Siccome  $P(A^C) \geq 0$ , il secondo assioma è verificato tramite il terzo. Per dimostrare il primo assioma, basta porre  $S = S \cup \emptyset$ , e siccome  $S$  e  $\emptyset$  sono disgiunti, abbiamo che:

$$1 = P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$$

E pertanto  $P(\emptyset) = 0$ .

**Teorema (2.4):** Se  $P$  è una funzione di probabilità e  $A$  e  $B$  sono dei qualsiasi insiemi in  $\mathcal{A}$ , allora:

- $P(B \cap A^C) = P(B) - P(A \cap B)$ .
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .
- Se  $A \subset B$ , allora  $P(A) \leq P(B)$ .

**Teorema (2.4):** Per stabilire il primo assioma dobbiamo notare che:

$$B = \{B \cap A\} \cup \{B \cap A^C\}$$

E pertanto:

$$P(B) = P(\{B \cap A\} \cup \{B \cap A^C\}) = P(B \cap A) + P(B \cap A^C)$$

Dimostrando quindi il primo assioma. Per il secondo assioma dobbiamo notare che:

$$A \cup B = A \cup \{B \cap A^C\}$$

Usando il fatto che  $A$  e  $B \cap A^C$  sono disgiunti (dato che  $A$  e  $A^C$  sono disgiunti), abbiamo che:

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

Dimostrando quindi il secondo assioma. Se  $A \subset B$ , allora  $A \cap B = A$ , allora utilizzando il primo assioma abbiamo che:

$$0 \leq P(B \cap A^C) = P(B) - P(A)$$

Dimostrando il terzo assioma.

**Teorema (2.5):** Se  $P$  è una funzione di probabilità, allora:

- $P(A) = \sum_{i=1}^n P(A \cap C_i)$  per una qualsiasi partizione  $C_1, C_2, \dots, C_n$ .
- $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$  per un qualsiasi set  $A_1, A_2, \dots, A_n$  (*disuguaglianza di Boole*).

## 2.3 La probabilità condizionata e l'indipendenza

**Definizione:** Se  $A$  e  $B$  sono due eventi in  $S$ , e  $P(B) > 0$ , allora la **probabilità condizionata** di  $A$  dato  $B$ , scritta  $P(A|B)$ , è:

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

Questa formula ci fornisce due metodi alternativi per calcolare l'intersezione di due eventi, infatti:

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

**Definizione:** Due eventi,  $A$  e  $B$ , sono detti **indipendenti** se:

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

Questa definizione è estendibile anche ad una famiglia di eventi.

**Definizione:** Una collezione di eventi  $A_1, A_2, \dots, A_n$  sono detti **mutualmente indipendenti** se per ogni sottocollezione  $A_1, A_2, \dots, A_k$  abbiamo che:

$$P(\cap_{j=1}^k A_{i_j}) = \prod_{j=1}^k P(A_{i_j})$$

**Teorema (2.6):** Se  $A$  e  $B$  sono due eventi indipendenti, allora le seguenti coppie sono indipendenti:

- $A$  e  $B^C$ .
- $A^C$  e  $B$ .
- $A^C$  e  $B^C$ .

**Teorema (2.4):** È necessario mostrare solo il primo assioma, poiché le dimostrazioni sono simili. Per mostrare il primo assioma vogliamo mostrare che  $P(A \cap B^C) = P(A)P(B^C)$ :

$$\begin{aligned} P(A \cap B^C) &= P(A) - P(A \cap B) = \\ &= P(A) - P(A)P(B) = \\ &= P(A)(1 - P(B)) = \\ &= P(A)P(B^C) \end{aligned}$$

### 2.3.1 Regola di Bayes

**Teorema (2.7):** Siano  $A_1, A_2, \dots, A_n$  una partizione dello spazio di campionamento, e sia  $B$  un qualsiasi insieme. Allora, per ogni  $i \in [1, n]$ , abbiamo che:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

## 2.4 Le variabili aleatorie e le funzioni di distribuzione

**Definizione:** Una **variabile aleatoria** è una funzione dallo spazio  $S$  ai numeri reali. La **funzione di distribuzione cumulativa** (o **CDF**) di una variabile aleatoria  $X$ , chiamata  $F_X(x)$ , è definita come:

$$F_X(x) = P(X \leq x) \quad \text{per tutte le } x$$

**Teorema (2.8):** La funzione  $F_X(x)$  è una CDF se e solo se le seguenti condizioni sono vere:

- $\lim_{x \rightarrow -\infty} F_X(x) = 0$  e  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .
- $F_X(x)$  è un funzione *non decrescente* in  $x$ .
- $F_X(x)$  è *continua a destra*, ossia, per ogni  $x_0$ ,  $\lim_{x \rightarrow x_0^-} F_X(x) = F_X(x_0)$ .

### 2.4.1 Variabili aleatorie continue e discrete

**Definizione:** Una variabile aleatoria  $X$  è **continua** se  $F_X(x)$  è una funzione continua di  $x$ , mentre è detta **discreta** se  $F_X(x)$  è una funzione a salti di  $x$ .

### 2.4.2 Variabili aleatorie identicamente distribuite

**Definizione:** Le variabili aleatorie  $X$  ed  $Y$  sono **identicamente distribuite** se, per ogni set  $A \in \mathcal{B}$ ,  $P(X \in A) = P(Y \in A)$

**Teorema (2.9):** Sia  $X$  ed  $Y$  due variabili aleatorie, le due definizioni sono equivalenti:

- $X$  ed  $Y$  sono identicamente distribuite.
- $F_X(x) = F_Y(x)$  per ogni  $x$ .

## 2.5 Le funzioni di densità e di massa

### 2.5.1 Funzione di massa

**Definizione:** Sia  $X$  una variabile aleatoria discreta, la **funzione di massa di probabilità** (o **PMF**) è definita come:

$$f_X(x) = P(X = x) \quad \text{per ogni } x$$

**Teorema (2.10):** Una funzione  $f_X(x)$  è una PMF di una variabile aleatoria  $X$  se e solo se:

- $f_X(x) > 0$  per tutte le  $x$ .
- $\sum_x f_X(x) = 1$ .

### 2.5.2 Funzione di densità

**Definizione:** Sia  $X$  una variabile aleatoria discreta, la **funzione di densità di probabilità** (o **PDF**) è una funzione  $f_X(x)$  che soddisfa:

$$F_X(x) = \int_{-\infty}^x f_X(x) dx \quad \text{per ogni } x$$

**Teorema (2.11):** Una funzione  $f_X(x)$  è una PDF di una variabile aleatoria  $X$  se e solo se:

- $f_X(x) > 0$  per tutte le  $x$ .
- $\sum_x f_X(x) = 1$ .

# Capitolo 3

## Trasformazioni e valori attesi

### 3.1 Distribuzioni di una funzione di una variabile aleatoria

Se  $X$  è una variabile aleatoria con CDF  $F_X(x)$ , allora una qualsiasi funzione di  $X$  è anch'essa una variabile aleatoria, e viene di solito indicata con  $Y = g(X)$ . Poiché  $Y$  è una funzione di  $X$ , possiamo descrivere il comportamento probabilistico di  $Y$  in termini di  $X$ :

$$P(Y \in A) = P(g(X) \in A)$$

Possiamo anche dire che  $g$  definisce una funzione dallo spazio di campionamento originale di  $X$ ,  $\mathcal{X}$ , ad un nuovo spazio di campionamento  $\mathcal{Y}$ :

$$g(x) : \mathcal{X} \rightarrow \mathcal{Y}$$

È possibile usare tale funzione per effettuare una mappatura inversa come segue:

$$g^{-1}(y) = \{x \in \mathcal{X} : g(x) = y\}$$

Se la variabile aleatoria  $Y$  è ora definita da  $Y = g(X)$ , possiamo allora dire che per ogni set  $A \subset Y$  si ha:

$$\begin{aligned} P(Y \in A) &= P(g(X) \in A) = \\ &= P(\{x \in \mathcal{X} : g(x) \in A\}) = \\ &= P(X \in g^{-1}(A)) \end{aligned}$$

Ciò definisce la distribuzione di probabilità di  $Y$ . Una cosa importante da dire è che se  $X$  è una variabile aleatoria discreta, allora lo è pure  $Y$ ; inoltre, la PMF e la CDF di  $Y$  sono:

$$\begin{aligned} f_Y(y) &= P(Y = y) = \sum_{x \in g^{-1}(y)} P(X = x) = \sum_{x \in g^{-1}(y)} f_X(x) \text{ per ogni } y \in Y \\ F_Y(y) &= P(Y \leq y) = \sum_{Y \leq y} \sum_{x \in g^{-1}(y)} f_X(x) \text{ per ogni } y \in Y \end{aligned}$$



Caso analogo lo si ha se  $X$  è una variabile aleatoria continua, per cui anche  $Y$  lo è; in questo caso la CDF di  $Y$  è:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = \\ &= P(g(X) \leq y) = \\ &= P(\{x \in \mathcal{X} : g(x) \leq y\}) = \\ &= \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x) dx \text{ per ogni } y \in Y \end{aligned}$$

Mentre la PMF è pari alla derivata della CDF. In alcuni casi però non è facile trovare la PMF e la CDF di una variabile aleatoria continua.

**Teorema (3.1):** Sia  $X$  una variabile aleatoria con CDF  $F_X(x)$ , e siano  $\mathcal{X} = \{x : f_X(x) > 0\}$  e  $y = \{y : y = g(x) \text{ per qualche } x \in X\}$ , allora la CDF di  $Y$  è definita come segue:

- Se  $g$  è una funzione crescente su  $\mathcal{X}$ ,  $F_Y(y) = F_X(g^{-1}(y))$  per  $y \in \mathcal{Y}$ .
- Se  $g$  è una funzione decrescente su  $\mathcal{X}$ ,  $F_Y(y) = 1 - F_X(g^{-1}(y))$  per  $y \in \mathcal{Y}$ .

**Teorema (3.2):** Sia  $X$  una variabile aleatoria con PMF  $d_X(x)$ , e sia  $g$  una funzione monotona. Siano  $\mathcal{X} = \{x : f_X(x) > 0\}$  e  $y = \{y : y = g(x) \text{ per qualche } x \in X\}$ . Supponiamo che  $f_X(x)$  sia continua su  $\mathcal{X}$  e  $g^{-1}(y)$  ha una derivata continua in  $\mathcal{Y}$ . Allora la PDF di  $Y$  è data da:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{se } y \in \mathcal{Y} \\ 0 & \text{altrimenti} \end{cases}$$

È importante dire che, nel caso in cui  $g$  non sia né crescente monotona né decrescente monotona, i due teoremi sopra non sono applicabili. È però possibile suddividere la funzione in sottointervalli in cui è monotona, e usarli per trovare un'espressione per  $Y = g(X)$

**Teorema (3.3):** Sia  $X$  una variabile aleatoria con PMF  $d_X(x)$ , e sia  $g$  una funzione monotona. Siano  $\mathcal{X} = \{x : f_X(x) > 0\}$  e  $\mathcal{Y} = \{y : y = g(x) \text{ per qualche } x \in \mathcal{X}\}$ . Supponiamo che esista una partizione  $A_1, A_2, \dots, A_k$  di  $\mathcal{X}$  tale che  $f_X(x)$  è continua su ogni  $A_i$ . Inoltre, supponiamo esistano le funzioni  $g_1(x), g_2(x), \dots, g_l(x)$  definite su  $A_1, A_2, \dots, A_k$  rispettivamente, e che per ogni  $i \in [1, k]$  soddisfano:

- $g(x) = g_i(x)$  per  $x \in A_i$ .
- $g_i(x)$  è monotona su  $A_i$ .
- L'insieme  $\mathcal{Y} = \{y : y = g_i(x) \text{ per qualche } x \in A_i\}$  è lo stesso.
- $g^{-1}(y)$  ha una derivata continua in  $\mathcal{Y}$ .

Allora:

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & \text{se } y \in \mathcal{Y} \\ 0 & \text{altrimenti} \end{cases}$$

**Teorema (3.4):** Sia  $X$  una variabile aleatoria con CDF  $d_X(x)$  continua, e sia definita la variabile aleatoria  $Y = F_X(x)$ . Allora  $Y$  è uniformemente distribuita su  $(0, 1)$ , ossia  $P(Y \leq y) = y$  per  $0 \leq y \leq 1$ .

## 3.2 I valori attesi

**Definizione:** Il **valore atteso** (o **media**) di una variabile aleatoria  $g(X)$ , indicato con  $\mathbb{E}[g(X)]$ , è pari a:

$$\mathbb{E}[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{se } X \text{ è continua} \\ \sum_{x \in X} g(x) f_X(x) = \sum_{x \in X} g(x) P(X = x) & \text{se } X \text{ è discreta} \end{cases}$$

**Teorema (3.5):** Sia  $X$  una variabile aleatoria e siano  $a, b, c \in \mathbb{R}$ . Allora, per ogni  $g_1(x)$  e  $g_2(x)$  per cui esiste la media:

- $\mathbb{E}[ag_1(x) + bg_2(x) + c] = a\mathbb{E}[g_1(x)] + b\mathbb{E}[g_2(x)] + c$ .
- Se  $g_1(x) \geq 0$  per tutte le  $x$ , allora  $\mathbb{E}[g_1(x)] \geq 0$ .
- Se  $g_1(x) \geq g_2(x)$  per tutte le  $x$ , allora  $\mathbb{E}[g_1(x)] \geq \mathbb{E}[g_2(x)]$ .
- Se  $a \leq g_1(x) \leq b$  per tutte le  $x$ , allora  $a \leq \mathbb{E}[g_1(x)] \leq b$ .

### 3.3 I momenti e la funzione generatrice dei momenti

**Definizione:** Per ogni intero  $n$ , l' $n$ -esimo **momento** di  $X$  (o di  $F_X(x)$ ), indicato con  $\mu'_n$ , è pari a:

$$\mu'_n = \mathbb{E}[X^n]$$

Mentre l' $n$ -esimo **momento centrato** di  $X$ , indicato con  $\mu_n$ , è pari a:

$$\mu_n = \mathbb{E}[(X - \mu)^n]$$

Dove  $\mu = \mu'_1 = \mathbb{E}[X]$  indica la media di  $X$ .

Oltre alla media di una variabile aleatoria  $X$ , ossia  $\mu'_1 = \mathbb{E}[X]$ , il momento forse più importante è il secondo momento centrato, conosciuto più comunemente come varianza.

**Definizione:** La **varianza** di una variabile aleatoria  $X$  è il secondo momento centrato:

$$\text{Var}[X] = \mu_2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

La radice quadrato positiva della varianza di  $X$  è detta **deviazione standard** di  $X$  ed è indicata con  $\sigma(X)$ .

Esiste una formula alternativa per calcolare la varianza di una variabile aleatoria  $X$ :

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}^2[X]] = \\ &= \mathbb{E}[X^2] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}^2[X] = \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}^2[X] = \\ &= \mathbb{E}[X^2] - 2\mathbb{E}^2[X] + \mathbb{E}^2[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X]\end{aligned}$$

**Teorema (3.6):** Se  $X$  è una variabile aleatoria con varianza finita, allora per ogni  $a, b \in \mathbb{R}$  si ha:

$$\text{Var}[aX + b] = a^2 \text{Var}[X]$$

**Definizione:** Sia  $X$  una variabile aleatoria con CDF  $F_X(x)$ . la **funzione generatrice dei momenti** (o **MGF**) di  $X$  (o  $F_X(x)$ ), indicata con  $M_X(t)$ , è pari a:

$$M_X(t) = \mathbb{E}[e^{tX}]$$

Più esplicitamente, possiamo scrivere la MGF di  $X$  come:

$$M_X(t) = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx & \text{se } X \text{ è continua} \\ \sum_x e^{tx} P(X = x) & \text{se } X \text{ è discreta} \end{cases}$$

**Teorema (3.7):** Sia  $X$  una variabile aleatoria, e siano  $a, b \in \mathbb{R}$ , allora:

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

**Teorema (3.7):** Se  $X$  ha MGF  $M_X(t)$ , allora:

$$\mathbb{E}[X^n] = M_X^{(n)}(0)$$

Dove definiamo:

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)|_{t=0}$$

Ossia l' $n$ -esimo momento è uguale all' $n$ -esima derivata di  $M_X(t)$  valutata in  $t = 0$ .

**Teorema (3.8):** Siano  $F_X(x)$  e  $F_Y(y)$  due CDF i cui momenti esistono, allora:

- Se  $X$  e  $Y$  hanno un intervallo delimitato in cui non si annullano, allora  $F_X(u) = F_Y(u)$  per ogni  $u$  se e solo se  $\mathbb{E}[X^r] = \mathbb{E}[Y^r]$  per tutti gli interi  $r \geq 0$ .
- Se le MGF esistono e  $M_X(t) = M_Y(t)$  per tutte le  $t$  in qualche vicinato di 0, allora  $F_X(u) = F_Y(u)$  per ogni  $u$ .

**Teorema (3.9):** Siano  $\{X_i\}_{i \geq 1}$  una sequenza di variabili aleatorie, ognuna con MGF  $M_{X_i}(t)$ . Supponiamo inoltre che:

$$\lim_{t \rightarrow \infty} M_{X_i}(t) = M_X(t) \quad \text{per ogni } t \text{ in un vicinato di } 0$$

E  $M_X(t)$  è una MGF. Allora esiste un'unica CDF  $F_X(x)$  i cui momenti sono determinati da  $M_X(t)$  e, per tutte le  $x$  dove  $F_X(x)$  è continua, abbiamo:

$$\lim_{t \rightarrow \infty} F_{X_i}(x) = F_X(x)$$

Ossia la convergenza di tutte le MGF in un'unica implica la convergenza di tutte le CDF in un'unica.

# Capitolo 4

## Famiglie comuni di distribuzioni

### 4.1 Le distribuzioni discrete

#### 4.1.1 Distribuzione uniforme

La **distribuzione uniforme** è una distribuzione di probabilità discreta che descrive la probabilità di estrarre un certo elemento in un insieme di  $n$  elementi.

##### Definizione

Una variabile aleatoria  $X$  ha distribuzione uniforme se, preso  $n \in \mathbb{N}$ , si ha che:

$$P(X = k) = \frac{1}{n} \quad \text{per ogni } k \in [1, n]$$

In questo caso si scriverà che  $X \sim \text{UNIF}(n)$ .

##### Media

La media di una variabile aleatoria  $X$  con distribuzione uniforme è pari a:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=1}^n xP(X = k) = \\ &= \sum_{k=1}^n x \frac{1}{n} = \\ &= \frac{1}{n} \sum_{k=1}^n k = \\ &= \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} \end{aligned}$$

## Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione uniforme è pari a:

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= \sum_{k=1}^n k^2 P(X = k) - \left(\frac{n+1}{2}\right)^2 = \\ &= \sum_{k=1}^n k^2 \frac{1}{n} - \frac{(n+1)^2}{4} = \\ &= \frac{1}{n} \sum_{k=1}^n k^2 - \frac{n^2 + 2n + 1}{4} = \\ &= \frac{1}{n} \frac{n(n+1)(2n+1)}{6} - \frac{n^2 + 2n + 1}{4} = \\ &= \frac{(n+1)(2n+1)}{6} - \frac{n^2 + 2n + 1}{4} = \\ &= \frac{4n^2 + 6n + 2 - 3n^2 - 6n - 3}{12} = \\ &= \frac{n^2 - 1}{12} = \frac{(n+1)(n-1)}{12} \end{aligned}$$

### 4.1.2 Distribuzione ipergeometrica

La **distribuzione ipergeometrica** è una distribuzione di probabilità discrete che descrive la probabilità di  $k$  successi in  $n$  tentativi, senza rimpiazzo, da una popolazione finita  $N$  che contiene esattamente  $K$  oggetti con quel tratto.

#### Definizione

Una variabile aleatoria  $X$  ha distribuzione ipergeometrica se, presi  $p \in [0, 1]$  e  $0 \leq n \leq N$ , si ha che:

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

In questo caso si scriverà che  $X \sim \text{IPER}(N, K, n)$

## Media

La media di una variabile aleatoria  $X$  con distribuzione ipergeometrica è pari a:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^K k P(X = k) = \\ &= \sum_{k=0}^K k \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = \\ &= \sum_{k=1}^K k \frac{K \binom{K-1}{k-1} \binom{N-K}{n-k}}{\frac{N}{n} \binom{N-1}{n-1}} = \\ &= \frac{Kn}{N} \sum_{k=1}^K k \frac{\binom{K-1}{k-1} \binom{N-K}{n-k}}{\binom{N-1}{n-1}} =\end{aligned}$$

Effettuando il cambio di variabile  $h = k - 1$  e definendo la nuova variabile aleatoria  $Y \sim \text{IPER}(N - 1, K - 1, n - 1)$ , abbiamo:

$$\begin{aligned}\mathbb{E}[X] &= \frac{Kn}{N} \sum_{k=1}^K k \frac{\binom{K-1}{k-1} \binom{N-K}{n-k}}{\binom{N-1}{n-1}} = \\ &= \frac{Kn}{N} \sum_{h=0}^{K-1} k \frac{\binom{K-1}{h} \binom{(N-1)-(K-1)}{n-1-h}}{\binom{N-1}{n-1}} = \\ &= \frac{Kn}{N} \sum_{h=0}^{K-1} k P(Y = y) = \frac{Kn}{N}\end{aligned}$$

## Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione ipergeometrica è pari a:

$$\text{Var}[X] = \frac{Kn}{N} \left( \frac{(N-n)(N-K)}{N(N-1)} \right)$$

### 4.1.3 Distribuzione binomiale

La **distribuzione binomiale** è una distribuzione di probabilità discreta che descrive il numero di successi in una sequenza di  $n$  esperimenti indipendenti, ognuno che prevede una domanda sì-no ed avente la propria probabilità di successo  $p$

#### Definizione

Una variabile aleatoria  $X$  ha distribuzione binomiale se, presi  $n \in \mathbb{N}$  e  $p \in [0, 1]$ , si ha che:

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

In questo caso si scriverà che  $X \sim \text{BIN}(n, p)$ .

## Media

La media di una variabile aleatoria  $X$  con distribuzione binomiale è pari a:

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \\
 &= \sum_{k=0}^n k \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k} = \\
 &= \sum_{k=1}^n k \frac{n \cdot (n-1)!}{k \cdot (k-1)! (n-1-(k-1))!} p \cdot p^{k-1} (1-p)^{n-1-(k-1)} = \\
 &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)! (n-1-(k-1))!} p^{k-1} (1-p)^{n-1-(k-1)} = \\
 &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)}
 \end{aligned}$$

Effettuando il cambio di indice  $j = k - 1$  si ha:

$$\begin{aligned}
 \mathbb{E}[X] &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} = \\
 &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = \\
 &= np(p + 1 - p)^{n-1} = np
 \end{aligned}$$

## Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione binomiale è pari a:

$$\begin{aligned}
 Var[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\
 &= \sum_{k=0}^n k^2 P(X = k) - (np)^2 = \\
 &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} - (np)^2 = \\
 &= \sum_{k=0}^n (k^2 - k + k) \binom{n}{k} p^k (1-p)^{n-k} - (np)^2 = \\
 &= \sum_{k=0}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + \underbrace{\sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}}_{= \mathbb{E}[X] = np} - (np)^2 = \\
 &= \sum_{k=0}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + np - (np)^2
 \end{aligned}$$



Supponendo che  $n \geq 2$  (poichè per  $n = 1$  si avrebbe una Bernoulliana):

$$\begin{aligned}
Var[X] &= \sum_{k=0}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^n (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^n k(k-1) \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^n n(n-1) \frac{(n-2)!}{(k-2)! (n-k)!} p^{k-2+2} (1-p)^{n-k} + np - (np)^2 = \\
&= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)! (n-2-(k-2))!} p^{k-2} (1-p)^{n-2-(k-2)} + np - (np)^2
\end{aligned}$$

Effettuando il cambio di indice  $j = k - 2$  si ha:

$$\begin{aligned}
Var[X] &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)! (n-2-(k-2))!} p^{k-2} (1-p)^{n-2-(k-2)} + np - (np)^2 = \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \frac{(n-2)!}{(j)! (n-2-j)!} p^j (1-p)^{n-2-j} + np - (np)^2 = \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} + np - (np)^2 = \\
&= n(n-1)p^2 (p + 1-p)^{n-2} + np - (np)^2 = \\
&= n(n-1)p^2 + np - (np)^2 = \\
&= n^2 p^2 - np^2 + np - n^2 p^2 = \\
&= -np^2 + np = np(1-p)
\end{aligned}$$

### Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria  $X$  con distribuzione binomiale è pari a:

$$\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] = \\
&= \sum_{k=0}^n e^{kt} \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (1-p + pe^t)^n
\end{aligned}$$

### 4.1.4 Distribuzione di Poisson

La **distribuzione di Poisson** è una distribuzione di probabilità discreta che esprime le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che mediamente se ne verifica un numero  $\lambda$ .

### Definizione

Una variabile aleatoria  $X$  ha distribuzione di Poisson se, preso  $\lambda > 0$ , si ha che:

$$P(X = k) = \frac{\lambda^n}{n!} e^{-\lambda} \quad \text{per ogni } n \in \mathbb{N}$$

In questo caso si scriverà che  $X \sim \text{POISSON}(\lambda)$ .

### Media

La media di una variabile aleatoria  $X$  con distribuzione di Poisson è pari a:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \sum_{n=1}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} n \frac{\lambda^{n-1}}{(n-1)!} = \\ &= \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda \end{aligned}$$

### Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione di Poisson è pari a:

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} - \lambda^2 = \\ &= \sum_{n=0}^{\infty} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} + \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} - \lambda^2 = \\ &= \lambda^2 e^{-\lambda} \sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} + \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} - \lambda^2 = \\ &= \lambda^2 e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} + \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} - \lambda^2 = \\ &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} - \lambda^2 = \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

## Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria  $X$  con distribuzione di Poisson è pari a:

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] = \\ &= \sum_{n=0}^{\infty} e^{tn} \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

### 4.1.5 Distribuzione binomiale negativa

La **distribuzione binomiale negativa** è una distribuzione di probabilità discreta che descrive il numero di esperimenti indipendenti e identicamente distribuiti, ognuno che prevede una domanda sì-no e avente la propria probabilità di successo  $p$ , necessari per ottenere l' $r$ -esimo successo. In alcuni casi viene anche definita come una distribuzione di probabilità discreta che descrive il numero di esperimenti indipendenti e identicamente distribuiti, ognuno che prevede una domanda sì-no e avente la propria probabilità di successo, che hanno restituito esito negativo prima di ottenere l' $r$ -esimo successo.

#### Definizione

Una variabile aleatoria  $X$  ha distribuzione binomiale negativa se, presi  $p \in [0, 1]$  e  $r \geq 1$ , si ha che:

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad \text{per ogni } k \geq r$$

In questo caso si scriverà che  $X \sim \text{NBIN}(r, p)$ . Se si usa la definizione alternativa di distribuzione binomiale negativa, ponendo  $Y = X - r$ , si ha che:

$$P(Y = k) = \binom{r+k-1}{k} p^r (1-p)^k$$

Anche in questo caso si scriverà che  $Y \sim \text{NBIN}(r, p)$ .

## Media

La media di una variabile aleatoria  $Y$  con distribuzione binomiale negativa è pari a:

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{k=0}^{\infty} k P(X = k) = \\ &= \sum_{k=0}^{\infty} k \binom{r+k-1}{k} p^r (1-p)^k = \\ &= \sum_{k=0}^{\infty} k \frac{(r+k-1)!}{k! (r-1)!} p^r (1-p)^k = \\ &= \sum_{k=0}^{\infty} \frac{(r+k-1)!}{(k-1)! (r-1)!} p^r (1-p)^k = \\ &= \sum_{k=0}^{\infty} r \frac{(r+k-1)!}{(k-1)! r!} p^r (1-p)^k = \\ &= \sum_{k=0}^{\infty} r \binom{r+k-1}{k-1} p^r (1-p)^k\end{aligned}$$

Effettuando il cambio di variabile  $h = k - 1$  otteniamo:

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{k=0}^{\infty} r \binom{r+k-1}{k-1} p^r (1-p)^k = \\ &= \sum_{h=0}^{\infty} r \binom{r+h}{h} p^r (1-p)^{h+1} = \\ &= r \sum_{h=0}^{\infty} \binom{r+h}{h} p^r (1-p)^{h+1} = \\ &= r \sum_{h=0}^{\infty} \frac{1-p}{p} \binom{r+h}{h} p^{r+1} (1-p)^h = \\ &= r \frac{1-p}{p} \sum_{h=0}^{\infty} \binom{r+h}{h} p^{r+1} (1-p)^h = \\ &= r \frac{1-p}{p} \sum_{h=0}^{\infty} \binom{r+1+h-1}{h} p^{r+1} (1-p)^h = r \frac{1-p}{p}\end{aligned}$$

## Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione ipergeometrica è pari a:

$$Var[Y] = \frac{r(1-p)}{p^2}$$

### 4.1.6 Distribuzione geoemtrica

La **distribuzione geometrica** è una distribuzione di probabilità discreta che descrive il numero di fallimenti che precedono il primo successo in una serie di prove indipendenti, le cui probabilità di successo sono pari a  $p$ .

## Definizione

Una variabile aleatoria  $X$  ha distribuzione geometrica se, preso  $p \in [0, 1]$ , si ha che:

$$P(X = k) = (1 - p)^{k-1}p \quad \text{con } k \in [1, \infty)$$

In questo caso si scriverà che  $X \sim \text{GEOM}(p)$ . Notare che la seguente distribuzione calcola la probabilità di effettuare  $k - 1$  fallimenti su  $k$  prove, le quali precedono il primo successo al  $k$ -esimo tentativo. È possibile modificarla in modo da calcolare la probabilità di avere  $k$  fallimenti prima di ottenere il primo successo:

$$P(X = k) = (1 - p)^k p \quad \text{con } k \in [0, \infty)$$

In questo caso si scriverà che  $X \sim \text{GEOM}^+(p)$ .

## Media

La media di una variabile aleatoria  $X$  con distribuzione geometrica è pari a:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \\ &= p \sum_{k=1}^{\infty} k(1-p)^{k-1} = \\ &= p \sum_{k=1}^{\infty} ((1-p)^k (-1))' = \\ &= -p \sum_{k=1}^{\infty} ((1-p)^k)' = \\ &= -p \left( \sum_{k=1}^{\infty} (1-p)^k \right)' = \\ &= -p \left( \frac{(1-p)^1}{1 - (1-p)} \right)' = \\ &= -p \left( \frac{1-p}{p} \right)' = -p \left( \frac{1}{p} - 1 \right)' = \\ &= -p(p^{-1} - 1)' = \\ &= -p \left( -\frac{1}{p^2} \right) = \frac{1}{p} \end{aligned}$$

## Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione geometrica è pari a:

$$\begin{aligned}
 Var[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\
 &= \sum_{k \geq 1} k^2 P(X = k) - \left(\frac{1}{p}\right)^2 = \\
 &= \sum_{k \geq 1} k^2 (1-p)^{k-1} p - \frac{1}{p^2} = \\
 &= \sum_{k \geq 1} (k^2 - k + k) (1-p)^{k-1} p - \frac{1}{p^2} = \\
 &= \sum_{k \geq 1} (k^2 - k) (1-p)^{k-1} p + \underbrace{\sum_{k \geq 1} k (1-p)^{k-1} p}_{\mathbb{E}[X] = \frac{1}{p}} - \frac{1}{p^2} = \\
 &= \sum_{k \geq 1} (k^2 - k) (1-p)^{k-1} p + \frac{1}{p} - \frac{1}{p^2} = \\
 &= \sum_{k \geq 2} (k^2 - k) (1-p)^{k-1} p + \frac{1}{p} - \frac{1}{p^2} = \\
 &= p(1-p) \sum_{k \geq 2} (k^2 - k) (1-p)^{k-2} + \frac{1}{p} - \frac{1}{p^2}
 \end{aligned}$$

Possiamo utilizzare una serie notevole e le sue derivate per semplificare i calcoli. In particolare:

$$\begin{aligned}
 \sum_{k \geq 0} x^k &= \frac{1}{1-x} && \text{per ogni } |x| < 1 \\
 \sum_{k \geq 0} k x^{k-1} &= \sum_{k \geq 1} k x^{k-1} = \frac{1}{(1-x)^2} && \text{derivata prima} \\
 \sum_{k \geq 1} k(k-1) x^{k-2} &= \sum_{k \geq 2} k(k-1) x^{k-2} = \frac{2}{(1-x)^3} && \text{derivata seconda}
 \end{aligned}$$

Possiamo utilizzare l'ultima serie notevole trovata ponendo  $x = 1 - p$ :

$$\begin{aligned}
 Var[X] &= p(1-p) \sum_{k \geq 2} (k^2 - k) (1-p)^{k-2} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= p(1-p) \frac{2}{(1-(1-p))^3} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= p(1-p) \frac{2}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \\
 &= \frac{2 - 2p + p - 1}{p^2} = \frac{1-p}{p^2}
 \end{aligned}$$

## Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria  $X$  con distribuzione geometrica è pari a:

$$\begin{aligned} M_X(t) &= \sum_{n \geq 1} e^{tn} P(X = n) = \\ &= \sum_{n \geq 1} e^{tn} (1-p)^{n-1} p = \\ &= \frac{p}{1-p} \sum_{n \geq 1} (e^t(1-p))^n = \\ &= \begin{cases} \frac{p}{1-p} \frac{e^t(1-p)}{1 - e^t(1-p)} & \text{se } e^t(1-p) < 1 \\ \infty & \text{se } e^t(1-p) \geq 1 \end{cases} = \\ &= \begin{cases} \frac{pe^t}{1 - e^t(1-p)} & \text{se } e^t(1-p) < 1 \\ \infty & \text{se } e^t(1-p) \geq 1 \end{cases} \end{aligned}$$

## Mancanza di memoria della distribuzione geometrica

**Teorema (4.1):** Sia  $X$  una variabile aleatoria con distribuzione geometrica, allora si ha:

$$P(X = n + k | X > k) = P(X = n) \quad \text{per ogni } n, k > 1 \text{ interi}$$

**Teorema (4.1):** Calcoliamo la probabilità condizionata:

$$P(X = n + k | X > k) = \frac{P(\{X = n + k\} \cap \{X > k\})}{P(X > k)} = \frac{P(X = n + k)}{P(X > k)}$$

Osserviamo che:

$$P(X > k) = \sum_{i \geq k+1} (1-p)^{i-1} p = p \sum_{i \geq k+1} (1-p)^{i-1} = p \left( \frac{(1-p)^k}{1 - (1-p)} \right) = (1-p)^k$$

Da qui abbiamo che:

$$P(X = n + k | X > k) = \frac{P(X = n + k)}{P(X > k)} = \frac{(1-p)^{n+k-1} p}{(1-p)^k} = (1-p)^{n-1} p = P(X = n)$$

## 4.2 Le distribuzioni continue

### 4.2.1 Distribuzione uniforme

La **distribuzione uniforme** è una distribuzione di probabilità continua che è uniforme su un insieme, ovvero che attribuisce la stessa probabilità a tutti i punti appartenenti

ad un dato intervallo  $[a, b]$  contenuto nell'insieme.

### Definizione

Una variabile aleatoria  $X$  ha distribuzione uniforme se, presi  $a, b \in \mathbb{R}$  tali che  $a < b$ , si ha che:

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{b-a} & \text{per ogni } x \in [a, b] \\ 0 & \text{altrimenti} \end{cases}$$

In questo caso si scriverà che  $X \sim \text{UNIF}(a, b)$ .

### Media

La media di una variabile aleatoria  $X$  con distribuzione uniforme è pari a:

$$\begin{aligned} \mathbb{E}[X] &= \int_a^b \frac{x}{b-a} dx = \\ &= \left[ \frac{x^2}{2(b-a)} \right]_a^b = \\ &= \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} = \\ &= \frac{b^2 - a^2}{2(b-a)} = \\ &= \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2} \end{aligned}$$

### Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione uniforme è pari a:

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= \int_a^b \frac{x^2}{b-a} dx - \left( \frac{b+a}{2} \right)^2 = \\ &= \left[ \frac{x^3}{3(b-a)} \right]_a^b - \frac{b^2 + 2ab + a^2}{4} = \\ &= \frac{b^3}{3(b-a)} - \frac{a^3}{3(b-a)} - \frac{b^2 + 2ab + a^2}{4} = \\ &= \frac{b^3 - a^3}{3(b-a)} - \frac{b^2 + 2ab + a^2}{4} = \\ &= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} - \frac{b^2 + 2ab + a^2}{4} = \\ &= \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} = \\ &= \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2}{12} = \\ &= \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12} \end{aligned}$$



### 4.2.2 Distribuzione Gamma

La **distribuzione Gamma** è una famiglia distribuzioni di probabilità continue; di questa famiglia fanno parte la distribuzione esponenziale, la distribuzione di Erlang e la distribuzione chi quadrato.

#### Definizione

Una variabile aleatoria  $X$  ha distribuzione Gamma se, presi  $\alpha, \beta \in \mathbb{R}$ , dove  $\alpha$  è detto il *parametro di forma* mentre  $\beta$  è detto il *parametro di scala*, si ha che:

$$f_X(x) = P(X = x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha) \beta^\alpha} \quad \text{per } x > 0$$

In questo caso si scriverà che  $X \sim \text{GAMMA}(\alpha, \beta)$ . La funzione  $\Gamma(\alpha)$  è detta *funzione Gamma* di Eulero ed è definita come:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

#### Media

La media di una variabile aleatoria  $X$  con distribuzione Gamma è pari a:

$$\begin{aligned} \mathbb{E}[X] &= \int_a^b x \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha) \beta^\alpha} dx = \\ &= \frac{1}{\Gamma(\alpha) \beta^\alpha} \int_a^b x x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \\ &= \frac{1}{\Gamma(\alpha) \beta^\alpha} \int_a^b x^\alpha e^{-\frac{x}{\beta}} dx = \\ &= \frac{1}{\Gamma(\alpha) \beta^\alpha} \Gamma(\alpha + 1) \beta^{\alpha+1} = \\ &= \frac{\alpha \Gamma(\alpha) \beta^{\alpha+1}}{\Gamma(\alpha) \beta^\alpha} = \alpha \beta \end{aligned}$$

## Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione Gamma è pari a:

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= \int_a^b x^2 \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} dx - (\alpha\beta)^2 = \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_a^b x^2 x^{\alpha-1} e^{-\frac{x}{\beta}} dx - \alpha^2 \beta^2 = \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_a^b x^{\alpha+1} e^{-\frac{x}{\beta}} dx - \alpha^2 \beta^2 = \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha+2)\beta^{\alpha+2} - \alpha^2 \beta^2 = \\ &= \frac{\alpha(\alpha+1)\Gamma(\alpha)\beta^{\alpha+2}}{\Gamma(\alpha)\beta^\alpha} - \alpha^2 \beta^2 = \\ &= \alpha(\alpha+1)\beta^2 - \alpha^2 \beta^2 = \\ &= \alpha^2 \beta^2 + \alpha\beta^2 - \alpha^2 \beta^2 = \alpha\beta^2 \end{aligned}$$

## Distribuzioni collegate

Questi sono alcuni esempi di distribuzioni appartenenti alla famiglia della distribuzione Gamma:

- La *distribuzione di Poisson*, infatti, sia  $X \sim \text{GAMMA}(\alpha, \beta)$  con  $\alpha \in \mathbb{N}$ , abbiamo che:

$$f_X(x) = P(X \leq x) = P(Y \geq \alpha) \quad \text{per } x > 0$$

Dove  $Y \sim \text{POISSON}(\frac{x}{\beta})$ .

- La *distribuzione chi quadrato con  $p$  gradi di libertà*, indicata con  $\mathcal{X}_p^2$ , infatti, sia  $X \sim \text{GAMMA}(\alpha = \frac{p}{2}, \beta = 2)$  con  $p \in \mathbb{N}$ , abbiamo che:

$$f_X(x) = P(X = x) = \frac{x^{\frac{p}{2}} e^{-\frac{x}{2}}}{\Gamma(\frac{p}{2}) 2^{\frac{p}{2}}} \quad \text{per } x > 0$$

- La *distribuzione esponenziale*, infatti, sia  $X \sim \text{GAMMA}(\alpha = 1, \beta)$ , abbiamo che:

$$f_X(x) = P(X = x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \quad \text{per } x > 0$$

### 4.2.3 Distribuzione normale

La **distribuzione normale** è una distribuzione di probabilità continua che permette la rappresentazione di variabili aleatorie di valore reale le cui distribuzioni sono sconosciute, ma di cui si conosce la media  $\mu$  e la varianza  $\sigma^2$ .

## Definizione

Una variabile aleatoria  $X$  ha distribuzione uniforme se, presi  $\mu, \sigma \in \mathbb{R}$  tali che  $a < b$ , si ha che:

$$f_X(x) = P(X = x) = \frac{e^{\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$$

In questo caso si scriverà che  $X \sim N(\mu, \sigma^2)$ . Nel caso in cui  $\mu = 0$  e  $\sigma = 1$ , si ha che:

$$f_X(x) = P(X = x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

In questo caso si dice che  $X$  ha **distribuzione normale standard**.

## Media

La media di una variabile aleatoria  $X$  con distribuzione normale è pari a:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{e^{\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dx = [\dots] = \mu$$

Notare che nel caso in cui  $X$  ha distribuzione normale standard la media è pari a

$$\mathbb{E}[X] = \mu = 0$$

## Varianza

La varianza di una variabile aleatoria  $X$  con distribuzione normale è pari a:

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \\ &= \int_{-\infty}^{\infty} x^2 \frac{e^{\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dx - \mu^2 = [\dots] = \sigma^2 \end{aligned}$$

Notare che nel caso in cui  $X$  ha distribuzione normale standard la media è pari a

$$\text{Var}[X] = \sigma^2 = 1$$

# Capitolo 5

## Varaibili aleatorie multiple

**Definizione:** Una **distribuzione multivariabile** è definita come un vettore  $(X_1, X_2, \dots, X_n)$ , dove  $\{X_i\}_{i \in [1, n]}$  sono variabili aleatorie, con spazio di campionamento che è un sottoinsieme di  $\mathbb{R}^n$ .

Se  $X_1, X_2, \dots, X_n$  è composto da variabili aleatorie discrete, allora la **PMF congiunta** per ogni  $(x_1, x_2, \dots, x_n) \in \mathbb{R}$  è la funzione definita da:

$$f(x) = f(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Allora possiamo dire che per ogni  $A \in \mathbb{R}$  si ha che:

$$P(X \in A) = \sum_{x \in A} f(x)$$

Se  $X_1, X_2, \dots, X_n$  è composto da variabili aleatorie continua, allora la **PDF congiunta** per ogni  $(x_1, x_2, \dots, x_n) \in \mathbb{R}$  è la funzione che soddisfa:

$$P(X \in A) = \int_A f(x) dx = \int \int \dots \int_A f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Questi integrali sono *integrali ad n pieghe* con limiti di integrazione in modo che l'integrazione sia su tutti i punti  $x \in A$ . Sia  $g(x) = g(x_1, x_2, \dots, x_n)$  sia una funzione a valori reali definita sullo spazio di campionamento  $X$ . Allora  $g(X)$  è una variabile aleatoria e il valore atteso di  $g(X)$  è:

$$\begin{aligned} \mathbb{E}[g(X)] &= \sum_{x \in \mathbb{R}^n} g(x) f(x) && \text{se } X \text{ è discreta} \\ \mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x) f(x) dx && \text{se } X \text{ è continua} \end{aligned}$$

La *PMF marginale* di un qualsiasi sottoinsieme delle coordinate di  $X_1, X_2, \dots, X_n$  può essere calcolata sommando le PMF congiunte su tutti i possibili valori delle altre coordinate; ossia, preso il sottoinsieme  $(X_1, X_2, \dots, X_k)$ , la sua PMF marginale per ogni  $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$  è pari a:

$$f(x_1, x_2, \dots, x_k) = \sum_{(x_{k+1}, x_{k+2}, \dots, x_n) \in \mathbb{R}^{n-k}} f(x_1, x_2, \dots, x_n)$$

La *PDF marginale* di un qualsiasi sottoinsieme delle coordinate di  $X_1, X_2, \dots, X_n$  può essere calcolata integrando le PMF congiunte su tutti i possibili valori delle altre coordinate; ossia, preso il sottoinsieme  $(X_1, X_2, \dots, X_k)$ , la sua PMF marginale per ogni  $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$  è pari a:

$$f(x_1, x_2, \dots, x_k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_{k+1} dx_{k+2} \dots dx_n$$

La *PMF/PDF condizionale* di un sottoinsieme delle coordinate di  $(X_1, X_2, \dots, X_n)$  dati i valori delle rimanenti coordinate è ottenuta dividendo la PMF/PDF congiunta per la PMF/PDF marginale delle rimanenti coordinate; ossia, sia  $X_{k+1}, X_{k+2}, \dots, X_n$  con  $X_{k+1} = x_{k+1}, X_{k+2} = x_{k+2}, \dots, X_n = x_n$ , e sia  $f(x_1, x_2, \dots, x_k) > 0$ , si ha che:

$$f(x_{k+1}, x_{k+2}, \dots, x_n | x_1, x_2, \dots, x_k) = \frac{f(x_1, x_2, \dots, x_n)}{f(x_1, x_2, \dots, x_k)}$$

## 5.1 Distribuzione multinomiale

**Definizione:** Siano  $n, m \in \mathbb{N}$  e siano  $p_1, p_2, \dots, p_n$  tali che per ogni  $i \in [1, n], p_i \in [0, 1]$  e  $\sum_{i=1}^n p_i = 1$ . Allora il vettore aleatorio  $(X_1, X_2, \dots, X_n)$  ha **distribuzione multinomiale** con  $m$  prove e probabilità  $p_1, p_2, \dots, p_n$  se la PMF congiunta di  $(X_1, X_2, \dots, X_n)$  sull'insieme  $(x_1, x_2, \dots, x_n)$ , con  $x_1, x_2, \dots, x_n$  non negativi tali che  $\sum_{i=1}^n x_i = m$ , è pari a:

$$f(x_1, x_2, \dots, x_n) = \frac{m!}{x_1! x_2! \dots x_n!} p_1^{x_1} p_2^{x_2} \dots p_n^{x_n} m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$$

**Teorema (5.1):** Siano  $n, m \in \mathbb{N}$ . Sia  $A$  l'insieme di vettori  $x = (x_1, x_2, \dots, x_n)$  tali che ogni  $x_i$  è non negativi e  $\sum_{i=1}^n x_i = m$ . Allora, per ogni  $p_1, p_2, \dots, p_n$  tali che per ogni  $i \in [1, n], p_i \in [0, 1]$ , è pari a:

$$(p_1 + p_2 + \dots + p_n)^m = \sum_{x \in A} \frac{m!}{x_1! x_2! \dots x_n!} p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}$$

## 5.2 Variabili mutualmente indipendenti

**Definizione:** Siano  $X_1, X_2, \dots, X_n$  variabili aleatorie con PMF/PDF congiunta  $f(x_1, x_2, \dots, x_n)$ . Sia  $f_{X_i}(x_i)$  la PMF/PDF marginale di  $X_i$ . Allora  $(X_1, X_2, \dots, X_n)$  è un **vettore aleatorio mutualmente indipendente** se, per ogni  $(x_1, x_2, \dots, x_n)$ , si ha che:

$$f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\dots f_{X_n}(x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

**Teorema (5.2):** Sia  $(X_1, X_2, \dots, X_n)$  un vettore aleatorio mutualmente indipendente, siano  $g_1, g_2, \dots, g_n$  funzioni a valori reali tali che, per ogni  $i \in [1, n]$ ,  $g_i(x_i)$  è una funzione solo per  $x_i$ , allora:

$$\mathbb{E}[g(X_1)g(X_2)\dots g(X_n)] = \mathbb{E}[g(X_1)]\mathbb{E}[g(X_2)]\dots \mathbb{E}[g(X_n)]$$

**Teorema (5.3):** Sia  $(X_1, X_2, \dots, X_n)$  un vettore aleatorio mutualmente indipendente, e siano  $M_{X_1}(t), M_{X_2}(t), \dots, M_{X_n}(t)$  le MGF di  $(X_1, X_2, \dots, X_n)$ . Sia  $Z$  la variabile aleatoria definita come  $Z = X_1 + X_2 + \dots + X_n$ , allora la MGF di  $Z$  è pari a

$$M_Z(t) = M_{X_1}(t)M_{X_2}(t)\dots M_{X_n}(t)$$

In particolare, se  $X_1, X_2, \dots, X_n$  hanno tutte la stessa distribuzione con MGF  $M_X(t)$ , allora:

$$M_Z(t) = (M_X(t))^n$$

**Teorema (5.4):** Sia  $(X_1, X_2, \dots, X_n)$  un vettore aleatorio mutualmente indipendente, e siano  $M_{X_1}(t), M_{X_2}(t), \dots, M_{X_n}(t)$  le MGF di  $(X_1, X_2, \dots, X_n)$ . Siano  $a_1, a_2, \dots, a_n \in \mathbb{R}$  e  $b_1, b_2, \dots, b_n \in \mathbb{R}$  delle costanti fisse, e sia  $Z$  la variabile aleatoria definita come  $Z = (a_1X_1 + b_1) + (a_2X_2 + b_2) + \dots + (a_nX_n + b_n)$ , allora la MGF di  $Z$  è pari a:

$$M_Z(t) = (e^{t \sum b_i})M_{X_1}(a_1t)M_{X_2}(a_2t)\dots M_{X_n}(a_nt)$$

**Teorema (5.5):** Sia  $(X_1, X_2, \dots, X_n)$  un vettore aleatorio mutualmente indipendente con  $x_i \sim N(\mu, \sigma^2)$  per ogni  $i \in [1, n]$ . Siano  $a_1, a_2, \dots, a_n \in \mathbb{R}$  e  $b_1, b_2, \dots, b_n \in \mathbb{R}$  delle costanti fisse, allora:

$$Z = \sum_{i=1}^n (a_iX_i + b_i) \sim N\left(\sum_{i=1}^n a_i\mu_i + b_i, \sum_{i=1}^n a_i^2\sigma_i^2\right)$$

**Teorema (5.6):** Sia  $(X_1, X_2, \dots, X_n)$  un vettore aleatorio. Allora  $(X_1, X_2, \dots, X_n)$  è un vettore aleatorio mutualmente indipendente se e solo se esistono per ogni  $i \in [1, n]$  una funzione  $g(x_i)$  tale che la PMF/PDF congiunta può essere scritta come:

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

**Teorema (5.7):** Sia  $(X_1, X_2, \dots, X_n)$  un vettore aleatorio mutualmente indipendente. Sia  $g_i(x_i)$  una funzione solo di  $x_i$  per ogni  $i \in [1, n]$ . Allora le variabili aleatorie  $Y_i = g(X_i)$ , con  $i \in [1, n]$ , sono mutualmente indipendenti.

# Capitolo 6

## Proprietà di un campionamento casuale

### 6.1 Concetti base del campionamento aleatorio

**Definizione:** Le variabili aleatorie  $X_1, X_2, \dots, X_n$  sono dette un **campionamento aleatorio** di grandezza  $n$  dalla popolazione  $f(x)$  se  $X_1, X_2, \dots, X_n$  sono variabili aleatorie mutualmente indipendenti e la PMF/PDF marginale di ciascuna  $X_i$  è la stessa funzione  $f(x)$ . Alternativamente,  $X_1, X_2, \dots, X_n$  sono dette **indipendenti e identicamente distribuite** con PMF/PDF  $f(x)$ .

### 6.2 Somma di variabili aleatorie da un campionamento aleatorio

**Definizione:** Sia  $X_1, X_2, \dots, X_n$  un campionamento aleatorio di grandezza  $n$  da una popolazione e sia  $T(x_1, x_2, \dots, x_n)$  sia una funzione a valori reali o a valori vettoriali il cui dominio include lo spazio di campionamento  $(X_1, X_2, \dots, X_n)$ . Allora la variabile aleatoria o il vettore aleatorio  $Y = T(X_1, X_2, \dots, X_n)$  è chiamata **statistica**. La distribuzione di probabilità di una statistica  $Y$  è chiamata la **funzione di campionamento** di  $Y$ .

**Definizione:** Il **valore atteso di campionamento** è la media aritmetica dei valori in un campionamento aleatorio. È di solito definita come:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$



**Definizione:** La **varianza di campionamento** è la statistica definita da:

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

La **deviazione standard del campionamento** è la statistica definita da  $S = \sqrt{S^2}$ .

### 6.3 Campionamento dalla distribuzione normale

**Teorema (6.1):** Sia  $X_1, X_2, \dots, X_n$  un campionamento aleatorio da una distribuzione normale  $N(\mu, \sigma^2)$ , e sia  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  e  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ .

Allora:

- $\bar{X}$  e  $S^2$  sono variabili aleatorie indipendenti.
- $\bar{X}$  ha una distribuzione  $N(\mu, \frac{\sigma^2}{n})$ .
- $\frac{(n-1)}{\sigma^2} S^2$  ha una distribuzione chi quadrato con  $n-1$  gradi di libertà.

**Teorema (6.1a):**

- Se  $Z \sim N(0, 1)$  è una variabile aleatoria, allora  $Z^2 \sim \mathcal{X}_1^2$ , ossia il quadrato di una variabile aleatoria con distribuzione normale standard è una variabile aleatoria con distribuzione chi quadrato.
- Se  $X_1, X_2, \dots, X_n$  sono variabili aleatorie indipendenti e, per ogni  $i \in [1, n]$ ,  $X_i \sim \mathcal{X}_{p_i}^2$ , allora  $X_1 + X_2 + \dots + X_n \sim \mathcal{X}_{p_1+p_2+\dots+p_n}^2$ , ossia delle variabili aleatorie indipendenti con distribuzione chi quadrato sommate danno una variabile aleatoria con distribuzione di chi quadrato, il cui grado di libertà è la somma dei gradi di libertà delle singole variabili aleatorie.

**Teorema (6.1b):** Siano  $\{X_i \sim N(\mu_i, \sigma_i^2)\}_{i \in [1, n]}$  variabili aleatorie indipendenti. Definiamo per le costanti  $a_{hi}$  e  $b_{ki}$ , con  $i \in [1, n]$ ,  $h \in [1, m_1]$ ,  $k \in [1, m_2]$  e  $m_1 + m_2 \leq n$ , le seguenti variabili aleatorie:

$$U_h = \sum_{i=1}^n a_{hi} X_i \quad h \in [1, m_1]$$

$$V_k = \sum_{i=1}^n b_{ki} X_i \quad k \in [1, m_2]$$

Allora:

- Le variabili aleatorie  $U_h$  e  $V_k$  sono indipendenti se e solo se  $Cov(U_h, V_k) = 0$ .  
Inoltre,  $Cov(U_h, V_k) = \sum_{i=1}^n a_{hi} b_{ki} \sigma_i^2$ .
- I vettori aleatori  $U_1, U_2, \dots, U_{m_1}$  e  $V_1, V_2, \dots, V_{m_2}$  sono indipendenti se e solo se  $U_h$  è indipendente da  $V_k$  per tutte le coppie  $h, k$  con  $h \in [1, m_1]$  e  $k \in [1, m_2]$ .

## 6.4 Concetti di convergenza

### 6.4.1 Convergenza di probabilità

**Definizione:** Una sequenza di variabili aleatorie  $X_1, X_2, \dots, X_n$  **converge in probabilità** ad una variabile aleatoria  $X$  se, per ogni  $\epsilon > 0$ , si ha che:

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0 \quad \text{o equivalentemente} \quad \lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1$$

Notare come nella definizione  $X_1, X_2, \dots, X_n$  non sono indipendenti e identicamente distribuite, come in un campionamento aleatorio; questo perché tipicamente non succede quando si parla di convergenza.

**Definizione:** Una sequenza di variabili aleatorie  $X_1, X_2, \dots, X_n$  **converge quasi sicuramente** ad una variabile aleatoria  $X$  se, per ogni  $\epsilon > 0$ , si ha che:

$$P\left(\lim_{n \rightarrow \infty} |X_n - X| \geq \epsilon\right) = 0 \quad \text{o equivalentemente} \quad \lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1$$

Seppur le due definizioni risultano molto simili, la seconda presenta un'implicazione più forte. Definiamo con  $X_n(s)$  e  $X(s)$  le funzioni definite sullo spazio di campionamento  $S$ ; la definizione sopra dice che  $X_n$  converge ad  $X$  se e solo se  $X_n(s)$  converge ad  $X(s)$  per tutti gli  $s \in S$ , escluso forse per  $s \in N$ , dove  $N \subset S$  e  $P(N) = 0$ .

**Teorema (6.2):** Siano  $X_1, X_2, \dots, X_n$  delle variabili aleatorie che convergono in probabilità ad una variabile aleatoria  $X$ , e sia  $h$  una funzione continua. Allora  $h(X_1), h(X_2), \dots, h(X_n)$  convergono in probabilità ad  $h(X)$ .

### Legge debole dei grandi numeri

**Teorema (6.3):** Siano  $X_1, X_2, \dots, X_n$  delle variabili aleatorie indipendenti e identicamente distribuite con  $\mathbb{E}[X_i] = \mu$  e  $Var(X_i) = \sigma^2 < \infty$ . Definito  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , si ha per ogni  $\epsilon > 0$  che:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

Ossia  $\bar{X}_n$  converge in probabilità a  $\mu$ .

### Legge forte dei grandi numeri

**Teorema (6.4):** Siano  $X_1, X_2, \dots, X_n$  delle variabili aleatorie indipendenti e identicamente distribuite con  $\mathbb{E}[X_i] = \mu$  e  $Var(X_i) = \sigma^2 < \infty$ . Definito  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , si ha per ogni  $\epsilon > 0$  che:

$$P\left(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| < \epsilon\right) = 1$$

Ossia  $\bar{X}_n$  converge quasi sicuramente a  $\mu$ .

## 6.4.2 Convergenza di distribuzione

**Definizione:** Una sequenza di variabili aleatorie  $X_1, X_2, \dots, X_n$  converge in distribuzione ad una variabile aleatoria  $X$  se:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

Su tutti i punti  $x$  dove  $F_X(x)$  è continua.

**Teorema (6.5):** Se la sequenza di variabili aleatorie  $X_1, X_2, \dots, X_n$  converge in probabilità ad una variabile aleatoria  $X$ , allora converge anche in distribuzione ad  $X$ .

**Teorema (6.6):** Se la sequenza di variabili aleatorie  $X_1, X_2, \dots, X_n$  converge in probabilità ad una costante  $\mu$  e solo se la sequenza converge anche in distribuzione a  $\mu$ , allora l'affermazione:

$$P(|X_n - \mu| > \epsilon) \rightarrow 0 \quad \text{per ogni } \epsilon > 0$$

È equivalente a:

$$P(X_n \leq x) = \begin{cases} 0 & \text{se } x < \mu \\ 1 & \text{se } x > \mu \end{cases}$$

### Teorema del limite centrale

**Teorema (6.7):** Siano  $X_1, X_2, \dots, X_n$  una sequenza di variabili aleatorie indipendenti e identicamente distribuite le cui MGF esistono in un vicinato di 0. Sia  $\mathbb{E}[X] = \mu$  e  $Var[X] = \sigma^2 > 0$ . Sia  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , e sia  $G_n(x)$  la CDF di  $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ . Allora, per ogni  $x \in (-\infty, \infty)$ , si ha che:

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy$$

Ossia  $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$  è una distribuzione normale standard limitante.

Una cosa interessante che possiamo notare da questo teorema è che la normalità deriva dalla somma di "piccole" perturbazioni indipendenti. L'assunzione di varianza finita è essenzialmente necessaria per la convergenza alla normalità. Va detto che il teorema del limite centrale risulta valido anche se vengono applicate delle generalizzazioni: in particolare possiamo sostituire le MGF con le funzioni caratteristiche.

**Teorema (6.8):** Siano  $X_1, X_2, \dots, X_n$  una sequenza di variabili aleatorie indipendenti e identicamente distribuite con  $\mathbb{E}[X] = \mu$  e  $Var[X] = \sigma^2 \in (0, \infty)$ . Sia  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , e sia  $G_n(x)$  la CDF di  $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ . Allora, per ogni  $x \in (-\infty, \infty)$ , si ha che:

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy$$

Ossia  $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$  è una distribuzione normale standard limitante.

## Teorema del limite centrale

**Teorema (6.9):** Se  $X_n$  converge in distribuzione ad  $X$  e  $Y_n$  converge in probabilità ad  $a$  costante, allora:

- $Y_n X_n$  converge in distribuzione ad  $aX$ .
- $X_n + Y_n$  converge in distribuzione ad  $X + a$ .

### 6.4.3 Metodo Delta

Ci sono casi in cui non siamo specificatamente interessati alla distribuzione di una variabile aleatoria, ma bensì a qualche funzione della variabile aleatoria. Un metodo per procedere è basato sull'usare una serie approssimante di Taylor, che permette di approssimare la media e la varianza di una funzione di una variabile aleatoria.

**Definizione:** Se una funzione  $g(x)$  ha derivata di ordine  $n$ , ossia  $g^{(n)}(x) = \frac{d^n}{dx^n} g(x)$  esiste, allora, per qualsiasi costante  $a$ , il **polinomio di Taylor** di ordine  $r$  in  $a$  è pari a:

$$T_r(x) = \sum_{i=1}^r \frac{g^{(i)}(a)}{i!} (x - a)^i$$

**Teorema (6.10):** Se  $g^{(r)}(a) = \frac{d^r}{dx^r} g(x)|_{x=a}$  esiste, allora:

$$\lim_{x \rightarrow a} \frac{g(x) - T_r(x)}{(x - a)^r} = 0$$

Possiamo utilizzare la definizione sopra per calcolare una stima della media e della varianza di una serie di variabili aleatorie, utilizzando una serie di Taylor del *primo ordine*. Siano  $T = (T_1, T_2, \dots, T_n)$  variabili aleatorie con media  $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ . Supponiamo esista una funzione differenziabile  $g(T)$  per la quale vogliamo una stima della varianza. Definiamo:

$$g'_i(\theta) = \frac{d}{dt_i} g(t)|_{t_1=\theta_1, t_2=\theta_2, \dots, t_n=\theta_n}$$

L'approssimazione tramite l'espansione della serie di Taylor del primo ordine di  $g$  su  $\theta$  è:

$$g(t) \cong g(\theta) + \sum_{i=1}^n g'_i(\theta)(t_i - \theta_i)$$

Da cui abbiamo che:

$$g(t) - g(\theta) \cong \sum_{i=1}^n g'_i(\theta)(t_i - \theta_i)$$

Calcolando la media da entrambi i lato otteniamo:

$$\mathbb{E}[g(T)] \cong \mathbb{E}[g(\theta) + \sum_{i=1}^n g'_i(\theta)(t_i - \theta_i)] = g(\theta)$$

Possiamo ora approssimare la varianza di  $g(T)$  come:

$$\begin{aligned} \text{Var}[g(T)] &\cong \mathbb{E}[(g(T) - g(\theta))^2] \cong \\ &\cong \mathbb{E}[(\sum_{i=1}^n g'_i(\theta)(t_i - \theta_i))^2] = \\ &= \sum_{i=1}^n (g'_i(\theta))^2 \text{Var}[T_i] + 2 \sum_{j < i} g'_i(\theta) g'_j(\theta) \text{Cov}(T_i, T_j) \end{aligned}$$

Nel caso in cui le variabili siano indipendenti, l'approssimazione della varianza diventa pari a:

$$\text{Var}[g(T)] \cong \sum_{i=1}^n (g'_i(\theta))^2 \text{Var}[T_i]$$

### Metodo Delta del primo ordine

**Teorema (6.11):** Sia  $X_n$  una sequenza di variabili aleatorie le quali soddisfano che  $\sqrt{n}(X_n - \theta)$  converge in distribuzione ad una distribuzione normale  $N(0, \sigma^2)$ . Per una data funzione  $g$  ed uno specifico valore *theta*, suppondo che  $g'(\theta)$  esista e non sia 0, allora  $\sqrt{n}(g(X_n) - g(\theta))$  converge in distribuzione ad una distribuzione normale  $N(0, \sigma^2(g'(\theta))^2)$ .

### Metodo Delta del secondo ordine

**Teorema (6.12):** Sia  $X_n$  una sequenza di variabili aleatorie le quali soddisfano che  $\sqrt{n}(X_n - \theta)$  converge in distribuzione ad una distribuzione normale  $N(0, \sigma^2)$ . Per una data funzione  $g$  ed uno specifico valore *theta*, suppondo che  $g'(\theta) = 0$  e che  $g''(\theta)$  esista e non sia 0, allora  $\sqrt{n}(g(X_n) - g(\theta))$  converge in distribuzione a  $\sigma^2 \frac{g''(\theta)}{2} \chi_1^2$ .

### Metodo Delta multivariato

**Teorema (6.13):** Sia  $X_1, X_2, \dots, X_n$  un campionamento aleatorio con  $\mathbb{E}[X_{ij}] = \mu_i$  e  $\text{Cov}(X_{ik}, X_{jk}) = \sigma_{ij}$ . Per una data funzione  $g$  con derivate parziali del primo ordine continue ed uno specifico valore di  $\mu = (\mu_1, \mu_2, \dots, \mu_p)$  per cui  $\tau^2 = \sum_i \sum_j \sigma_{ij} \frac{d}{d\mu_i} g(\mu) \frac{d}{d\mu_j} g(\mu) > 0$ , allora  $\sqrt{n}(g(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_s) - g(\mu_1, \mu_2, \dots, \mu_p))$  converge in distribuzione ad una distribuzione normale  $N(0, \tau^2)$ .