Probabilità e Statistica

Isadora Antoniano-Villalobos
isadora.antoniano@unive.it

Laurea in Informatica
(Data science/ Tecnologie e scienze dell'informazione)
Università Ca' Foscari di Venezia

Anno academico 2024/2025

Variabili Casuali Discrete
Testo Capitolo 3

Introduzione

Abbiamo già visto vari esempi di **fenomeni aleatori** e **spazi campionari** ad essi collegati.

- Uno spazio campionario relativo ad un esperimento o ad un fenomeno casuale può essere di varia natura.

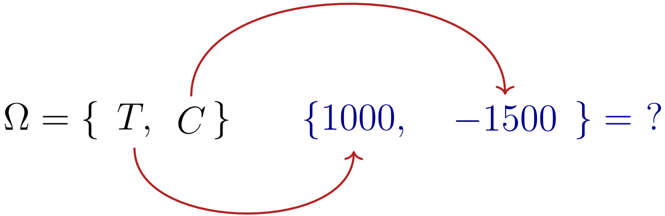
  ☞ Non è detto che sia un insieme numerico!

- In molte situazioni, anziché essere interessati allo specifico risultato di un esperimento, siamo interessati ad una sua **funzione numerica**.

Introduzione

- **Esempio:** si fa una scommessa in cui si vincono 1000 euro se lanciando una moneta si ottiene testa e se ne perdono 1500 se il risultato è croce.

☞ Chiaramente ciò che ci preme di più è l'importo della vincita (negativo se si tratta della perdita) e non il risultato esatto del lancio della moneta.



$$\Omega = \{\ T,\ \ C\,\} \qquad \{1000, \quad -1500\ \} = ?$$

Variabili aleatorie

Variabile aleatoria

Una **variabile aleatoria** o **casuale** X è una funzione che assume valori numerici determinati dall'esito di un certo fenomeno aleatorio.

Formalmente, se Ω è lo spazio campionario relativo al fenomeno di interesse, X è una particolare funzione

$$X : \Omega \to \mathbb{R} \ .$$

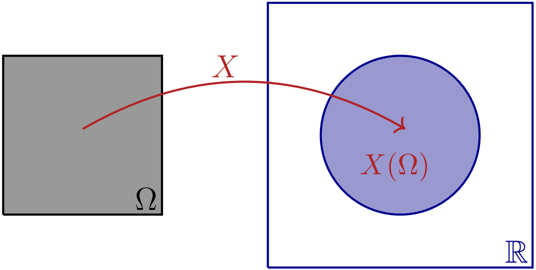
☞ Le variabili aleatorie si indicano di solito con una lettera maiuscola.

Esempi:

1. $S_4 =$ numero di teste in 4 lanci consecutivi di una moneta; i possibili valori di S_4 sono: 0, 1, 2, 3 e 4.
2. $X =$ vincita nella scommessa descritta precedentemente. I possibili valori di X sono 1000 e -1500.
3. $T =$ tempo di vita di un componente elettronico prodotto da una ditta. I possibili valori di T sono tutti i numeri reali maggiori di 0.

Spazio campionario indotto

Una variabile aleatoria associata ad un esperimento definisce dunque un nuovo spazio campionario numerico, costituito da tutti i possibili valori assunti dalla variabile stessa.



In questo modo si passa da un generico spazio campionario ad un sottoinsieme di $\mathbb{R}$.

Spazio campionario indotto

Bisogna ora assegnare le probabilità agli eventi del nuovo spazio campionario.

☞ Vedremo due modi differenti per fare quest'assegnazione, a seconda che lo spazio campionario indotto sia:

- un sottoinsieme numerabile di $\mathbb{R}$ (esempi 1 e 2) ➔ variabile **discreta**
- un sottoinsieme non numerabile di $\mathbb{R}$ (esempio 3) ➔ variabile **continua**

Variabili aleatorie discrete

Variabile aleatoria discreta

Una **variabile aleatoria discreta** X assume valori in un insieme numerabile (o finito) di punti, $\{x_1, x_2, \dots, x_i, \dots\}$.

Un modello probabilistico per X è un'assegnazione di probabilità ad ogni suo possibile valore:

$$\mathbb{P}[X = x_i] = p_i, \quad \forall i = 1, 2, \dots \ .$$

Le probabilità p_i sono tali che:

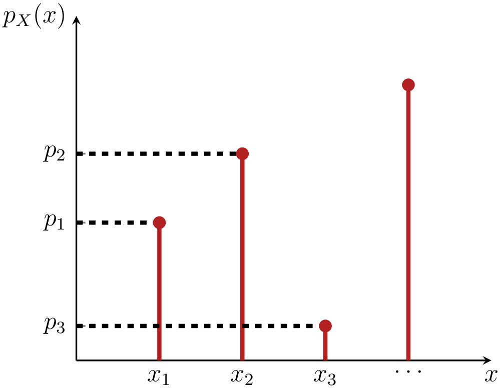
1. $0 \leq p_i \leq 1, \quad \forall i = 1, 2, \dots$
2. $\sum_i p_i = 1$.

Per calcolare $\mathbb{P}[X \in A]$, si sommano le probabilità dei singoli valori che appartengono ad A:

$$\mathbb{P}[X \in A] = \sum_{i: x_i \in A} p_i \ .$$

Funzioni di probabilità

Un'assegnazione di probabilità per X, $p_X(x) = \mathbb{P}[X = x]$ viene chiamata **funzione di probabilità** e può essere rappresentata graficamente tramite un diagramma a bastoncini:



Funzioni di probabilità

➔ **Esempio:** Si ricordi l'esempio dell'urna composta da quattro palle bianche numerate da 1 a 4 e tre palle nere numerate da 1 a 3, dalla quale si campiona casualmente una palla

- Sia $X =$ numero estratto dall'urna.

$$X \in \{1, 2, 3, 4\} = X(\Omega) \subset \mathbb{R}$$

$$\mathbb{P}[X = 1] = 2/7, \quad \mathbb{P}[X = 2] = 2/7,$$

$$\mathbb{P}[X = 3] = 2/7, \quad \mathbb{P}[X = 4] = 1/7$$

Si scrive anche

$$p_X(x) = \begin{cases} 2/7 & \text{se } x = 1, 2, 3; \\ 1/7 & \text{se } x = 4; \\ 0 & \text{altrimenti} \end{cases}$$

Funzioni di probabilità

👉 Se $A = \{1, 2\}$, allora

$$\mathbb{P}\left[X \in A\right] = \mathbb{P}\left[X \leq 2\right] = p_X(1) + p_X(2) == \frac{2}{7} + \frac{2}{7} = \frac{4}{7}$$

- Supponiamo di vincere 30 euro se la pallina estratta è bianca e di perderne 20 se è nera. La variabile casuale che rappresenta questo scenario è: $Y =$ importo vinto nel gioco.

$$p_Y(y) = \begin{cases} 4/7 & \text{se } y = 30, \\ 3/7 & \text{se } y = -20, \\ 0 & \text{altrimenti} \end{cases}$$

Funzioni di probabilità

- Consideriamo ora un'estrazione di due palline senza reinserimento. Sia $S =$ somma dei due numeri estratti.

$$p_S(s) = \begin{cases} 2/42 & \text{se } x = 2, \\ 8/42 & \text{se } x = 3, \\ 10/42 & \text{se } x = 4, \\ 12/42 & \text{se } x = 5, \\ 6/42 & \text{se } x = 6, \\ 4/42 & \text{se } x = 7, \\ 0 & \text{altrimenti} \end{cases}$$

Infatti, ad esempio,

$$\begin{aligned} \mathbb{P}[S = 2] &= \mathbb{P}[X_1 = 1 \cap X_2 = 1] \\ &= \mathbb{P}[X_1 = 1]\,\mathbb{P}[X_2 = 1 | X_1 = 1] \\ &= \frac{2}{7} \times \frac{1}{6} \end{aligned}$$

➔ **Esercizio:** Cosa cambia se l'estrazione avviene con reinserimento?

Funzioni di ripartizione e di sopravvivenza

Funzioni di ripartizione e di Sopravvivenza

Si dice **funzione di ripartizione** (o di distribuzione cumulativa) di una variabile aleatoria X la funzione $F_X : \mathbb{R} \to [0,1]$ così definita:

$$F_X(x) = \mathbb{P}\left[X \leq x\right], \quad \forall x \in \mathbb{R}.$$

La **coda** o **funzione di sopravvivenza** di X è la funzione $\bar{F}_X : \mathbb{R} \to [0,1]$ data da:

$$\bar{F}_X(x) = \mathbb{P}\left[X > x\right] = 1 - F_X(x), \quad \forall x \in \mathbb{R}.$$

- La funzione di ripartizione in un dato punto x è una probabilità, per questo il suo codominio è $[0,1]$

Funzioni di ripartizione e di sopravvivenza

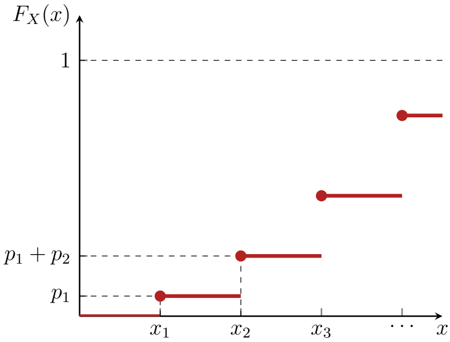
- La funzione di ripartizione F_X ha le seguenti proprietà:
  1. F_X è non decrescente,
  2. F_X è continua a destra,
  3. $\lim_{x\to-\infty} F_X(x) = 0$ e $\lim_{x\to+\infty} F_X(x) = 1$.

Se X è una v.a. discreta con valori $\{x_1, x_2, \ldots\}$ e funzione di probabilità p_X, allora

$$F_X(x) = \sum_{i:x_i\leq x} \mathbb{P}\left[X = x_i\right] = \sum_{i:x_i\leq x} p_X(x_i).$$

☞ La funzione di ripartizione di una v.a. discreta è una funzione costante a tratti, con salti in corrispondenza dei punti massa $x_1, x_2, \ldots$.

Funzioni di ripartizione e di sopravvivenza



Dalla funzione di ripartizione si può risalire alla funzione di probabilità della variabile così:

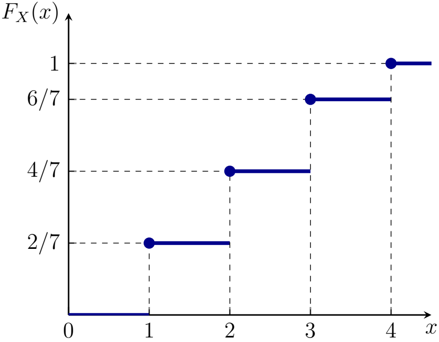
$$p_X(x) = \mathbb{P}[X = x] = F_X(x) - F_X(x^-), \quad \forall x \in \mathbb{R},$$

cioè la probabilità che X assuma il valore x è uguale al salto della funzione di ripartizione nel punto x.

Funzioni di ripartizione e di sopravvivenza

- **Esempio:** Per la variabile $X =$ numero estratto dalla solita urna, si ha:

$$F_X(x) = \begin{cases} 0 & x < 1 \\ 2/7 & 1 \leq x < 2 \\ 4/7 & 2 \leq x < 3 \\ 6/7 & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$



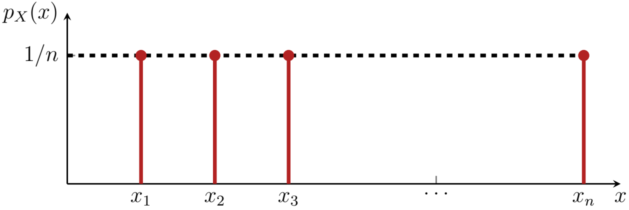
- **Esercizio:** trovare la funzione di probabilità di X a partire dalla funzione di ripartizione.

Distribuzione Uniforme Discreta

Distribuzione Uniforme (Discreta)

Consideriamo una variabile aleatoria X che assume un numero finito di valori, $\{x_1, \ldots, x_n\}$, tutti con la stessa probabilità $p_i = 1/n$, $i = 1, \ldots, n$. Si dice allora che X ha una **distribuzione uniforme** e si scrive

$$X \sim \mathsf{U}\{x_1, \ldots, x_n\}$$



Distribuzione Uniforme Discreta

Esempi:

1. X descrive il risultato del lancio di un dado (non truccato). I valori possibili per X sono gli interi fra 1 e 6, ciascuno con probabilità $1/6$ di verificarsi.

2. Si fa una scommessa in cui si guadagnano 1000 euro se lanciando una moneta equilibrata si ottiene testa e se ne perdono 1500 se il risultato è croce. La variabile X che indica il guadagno ottenuto, ha distribuzione uniforme sull'insieme $\{x_1 = 1000, x_2 = -1500\}$.

Distribuzione ipergeometrica

Distribuzione ipergeometrica

Sia X la variabile aleatoria che conta il numero di successi su k estrazioni senza reinserimento da una popolazione con $m+n$ elementi dei quali m sono considerati successi. Si dice allora che X ha una **distribuzione ipergeometrica** di parametri m, n e k e si scrive

$$X \sim \mathsf{Ig}\,(m, n, k)$$

Si ha che

$$p_X(x) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}} \ ,$$

per $\max\{0, k-m\} \leq x \leq \min\{k, m\}$, e $k \leq m+n$.

Distribuzione ipergeometrica

- **Esempio:** Un *software* consiste di 12 programmi, 5 dei quali necessitano di *upgrade*. Se vengono scelti a caso 4 programmi per un test qual è la probabilità che almeno 2 di essi siano da aggiornare? Sia X la variabile che conta il numero di programmi da aggiornare fra i 4 scelti. Allora $X \sim \mathsf{Ig}\left(m=5, n=7, k=4\right)$ e la probabilità richiesta è

$$\begin{aligned}\bar{F}_X(2) &= \mathbb{P}\left[X \geq 2\right] = 1 - \mathbb{P}\left[X \leq 1\right] \\ &= 1 - \mathbb{P}\left[X = 0\right] - \mathbb{P}\left[X = 1\right] \\ &= 1 - \frac{\binom{5}{0}\binom{7}{4-0}}{\binom{5+7}{4}} - \frac{\binom{5}{1}\binom{7}{4-1}}{\binom{5+7}{4}}.\end{aligned}$$

Usando R:

- $\mathbb{P}\left[X \geq 1\right]$

```
> phyper(q=2, m=5, n=7, k=4, lower.tail = FALSE)
[1] 0.1515152
```

Distribuzione di Bernoulli

Supponiamo di eseguire un esperimento i cui possibili risultati possono essere classificati come *successo*, con probabilità p o *insuccesso*, con probabilità $1 - p$. Questo tipo di esperimento viene comunemente chiamato **prova bernoulliana**.

☞ **Jakob Bernoulli (1654–1705):** Noto anche come Jacques, James o Giacomo Bernoulli, è stato un matematico e scienziato svizzero. La sua opera principale, *Ars Conjectandi*, pubblicata postuma nel 1713, è un lavoro fondamentale per la teoria delle probabilità. I concetti *campionamento bernoulliano*, *teorema di Bernoulli*, *variabile casuale bernoulliana* e *numeri di Bernoulli* sono legati ai suoi lavori e nominati in suo onore. Inoltre il primo teorema centrale del limite, ovvero la legge dei grandi numeri, venne formulata da Jakob. (Wikipedia)

Jakob Bernoulli

Distribuzione di Bernoulli

Distribuzione di Bernoulli

Sia X la variabile casuale che prende il valore $X = 1$ quando l'esito di una prova bernoulliana è un successo e $X = 0$ quando questo è un insuccesso. Si dice allora che X ha una **distribuzione di Bernoulli** di parametro $p \in (0, 1)$ e si scrive

$$X \sim \text{Ber}\,(p)$$

Si ha che,

$$p_X(x) = \mathbb{P}\left[X = x\right] = p^x (1-p)^{1-x} \mathbf{1}_{\{0,1\}}(x) = \begin{cases} 1-p & \text{se } x = 0; \\ p & \text{se } x = 1; \\ 0 & \text{altrimenti} \end{cases}$$

Distribuzione di Bernoulli

Esempio: L'estrazione di un singolo individuo da una popolazione con N elementi dei quali K sono considerati successi è una prova bernoulliana. La variabile casuale $X =$ "Numero di successi" ha una distribuzione di Bernoulli con parametro $p = K/N$.

☞ Anche la variabile casuale $Y =$ "Numero di insuccessi" ha una distribuzione di Bernoulli, ma con parametro $q = 1 - p = (N - K)/N$.

Distribuzione binomiale

Distribuzione binomiale

Supponiamo di eseguire n prove bernoulliane indipendenti, ognuna con probabilità di successo p. Sia X la variabile aleatoria che conta il numero totale di successi ottenuti nelle n prove. Si dice allora che X ha una **distribuzione binomiale** di parametri n e $p \in (0, 1)$ e si scrive

$$X \sim \text{Bin}\,(n, p)$$

Si ha che

$$p_X(x) = \mathbb{P}\,[X = x] = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots, n\ .$$

Distribuzione binomiale

Si osservi che si tratta effettivamente di una distribuzione di probabilità:

$$\begin{aligned}\sum_{x=0}^{n} \mathbb{P}\left[X = x\right] &= \sum_{x=0}^{n} \binom{n}{x} p^{x} (1-p)^{n-x} \\ &= (p + 1 - p)^{n} = 1,\end{aligned}$$

per la formula del binomio di Newton.

- Sia X la variabile aleatoria che conta il numero di successi su n estrazioni con reinserimento da una popolazione con N elementi dei quali K sono considerati successo. La probabilità di successo rimane invariata, uguale a $p = K/N$, ad ogni estrazione successiva. Quindi, X ha una distribuzione binomiale di parametri n e $p = K/N$.
- La distribuzione di Bernoulli è un caso particolare della binomiale con $n = 1$.

Distribuzione binomiale

Esempio: Si consideri la solita urna con 4 palline bianche e 3 nere. Sia S_3 la variabile che conta il numero di palline bianche ottenute in tre estrazioni con reinserimento. Qual è la $\mathbb{P}\left[S_3 = 2\right]$?

➔ Considerando l'estrazione di pallina bianca come successo e visto che $\mathbb{P}\left[\text{bianca}\right] = 4/7$, si può affermare che $S_3 \sim \text{Bin}\left(3, 4/7\right)$.
Allora,

$$\mathbb{P}\left[S_3 = 2\right] = \binom{3}{2} \left(\frac{4}{7}\right)^2 \left(\frac{3}{7}\right)^{(3-2)} = 0.4198 \ .$$

La distribuzione binomiale si utilizza anche nel caso di **Campionamento da popolazioni infinite** $(N \to \infty)$, quando si conosce ad ogni estrazione la probabilità di successo p.

☞ Se la popolazione viene considerata infinita, non si distingue nemmeno fra estrazioni con o senza reinserimento perché si assume che in ogni caso la probabilità di successo in estrazioni successive non cambia, e si usa sempre la binomiale.

Distribuzione binomiale

Esempio: Un motore di ricerca su internet cerca una parola in una sequenza (finita, ma così numerosa da considerarsi infinita) di siti web indipendenti. Si pensa che il 20% di questi siti possa contenere la parola cercata. Qual è la probabilità che 5 dei primi 10 siti visitati contengano la parola cercata? E la probabilità che al meno 5 dei primi 10 siti la contengano?

➔ Sia Y la variabile che conta il numero di siti, fra i primi 10 visitati, che contengono la parola cercata. Allora $Y \sim \text{Bin}\,(n = 10, p = 0.2)$ e

$$\mathbb{P}\left[Y = 5\right] = \binom{10}{5} 0.2^5 (1 - 0.2)^{10-5} = 0.0264$$

$$\begin{aligned}\mathbb{P}\left[Y \leq 5\right] &= \mathbb{P}\left[Y = 0\right] + \mathbb{P}\left[Y = 1\right] + \ldots + \mathbb{P}\left[Y = 5\right] \\ &= 0.1074 + 0.2684 + \ldots + 0.0264 = 0.9936\end{aligned}$$

Distribuzione binomiale

Usando R:

➔ $\mathbb{P}\left[Y = 0\right]$

```
> dbinom(x=0,size=10,prob=0.2)
[1] 0.1073742
```

➔ $\mathbb{P}\left[Y = 1\right]$

```
> dbinom(x=1,size=10,prob=0.2)
[1] 0.2684355
```

$\vdots$

➔ $\mathbb{P}\left[Y = 5\right]$

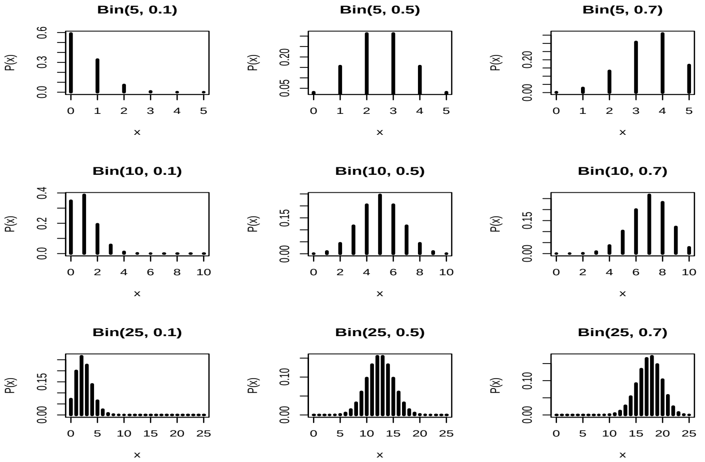
```
> dbinom(x=5,size=10,prob=0.2)
[1] 0.02642412
```

➔ O direttamente, $\mathbb{P}\left[Y \leq 5\right]$

```
> pbinom(q=5,size=10,prob=0.2)
[1] 0.9936306
```

Distribuzione binomiale

I parametri della binomiale determinano la forma della distribuzione:



Distribuzione geometrica

Distribuzione geometrica

Sia X una variabile aleatoria che conta il numero di ripetizioni indipendenti necessarie per osservare il primo successo in un esperimento binario che ha probabilità di successo p. Si dice allora che X ha una **distribuzione geometrica** di parametro $p \in (0,1)$ e si scrive

$$X \sim \text{Geo}\,(p)$$

Si ha che

$$p_X(x) = \mathbb{P}\left[X = x\right] = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots \ .$$

Si tratta effettivamente di una distribuzione di probabilità:

$$\sum_{x=1}^{\infty} \mathbb{P}\left[X = x\right] = p\sum_{x=1}^{\infty}(1-p)^{x-1} = p\,\frac{1}{1-(1-p)} = 1.$$

Distribuzione geometrica

Mancanza di memoria

Se $X \sim \text{Geo}\left(()\, p\right)$, allora

$$\mathbb{P}\left[X > m+n | X > m\right] = \frac{\mathbb{P}\left[X > m+n\right]}{\mathbb{P}\left[X > m\right]}$$
$$= \mathbb{P}\left[X > n\right] \ .$$

Questa proprietà si dimostra tenendo conto che

$$\bar{F}_X(x) = \mathbb{P}\left[X > k\right] = (1-p)^k \ , \quad k = 0, 1, 2, \ldots \ .$$

☞ La distribuzione geometrica è l'unica distribuzione discreta con la proprietà di *mancanza di memoria*.

Distribuzione geometrica

Esempio Si consideri ancora il motore di ricerca su internet. Se il 20% dei siti visitati contengono la parola cercata,

1. qual è la probabilità di dover visitare 15 siti per trovare la parola?
2. Dato che i primi 4 siti visitati non contenevano la parola cercata, qual è la probabilità di doverne visitare più di 10 in tutto per trovare la parola cercata?

➔ Supponiamo che X conti il numero di pagine da visitare per trovare per la prima volta la parola cercata.
Allora $X \sim \text{Geo}\,(0.2)$ e

$$\mathbb{P}\left[X = 15\right] = 0.2 \times 0.8^{14} = 0.0088,$$

$$\mathbb{P}\left[X > 10 | X > 4\right] = \mathbb{P}\left[X > 10 - 4 = 6\right] = 0.8^{6} = 0.2621,$$

Distribuzione geometrica

Attenzione! In alcuni contesti, si definisce la variabile aleatoria geometrica come il numero di ripetizioni indipendenti necessarie **prima di** osservare il primo successo. In questo caso, i possibili valori della variabile sono $\{0, 1, 2, \ldots\}$. Questa è la definizione usata da **R**. Infatti:

➔ Ricordiamo che $\mathbb{P}[X = 15] = 0.008796093$,

```
> dgeom(x=15, prob=0.2)
[1] 0.007036874
> dgeom(x=15-1, prob=0.2)
[1] 0.008796093
```

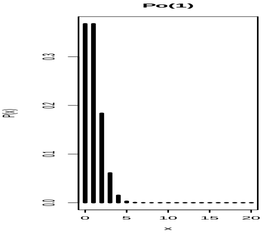
Distribuzione di Poisson

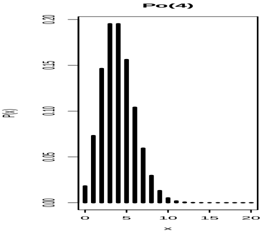
Distribuzione di Poisson

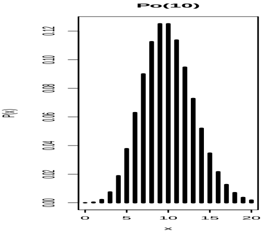
Una variabile X che assume valori nell'insieme dei numeri naturali $\mathbb{N}$, ha **distribuzione di Poisson** di parametro $\lambda > 0$ se

$$p_X(x) = \mathbb{P}\left[X = k\right] = \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, 2, \ldots \; .$$

Scriveremo allora $X \sim \mathrm{Po}\left(\lambda\right)$







Distribuzione di Poisson

☞ E' possibile vedere che:

$$\sum_{i=0}^{+\infty} p_X(i) = \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1 \ .$$

La variabile di Poisson viene utilizzata come modello per il conteggio di manifestazioni di un certo fenomeno di interesse. Ad esempio:

1. chiamate in arrivo ad un centralino in un certo intervallo di tempo;
2. macchine transitanti ad un casello autostradale in un certo periodo del giorno;
3. difetti rilevati in un pezzo di filo d'acciaio prodotto da una ditta;
4. terremoti manifestatisi in una data area nell'arco degli ultimi 10 anni.

Distribuzione di Poisson

Esempio: Al mio account di posta elettronica arrivano messaggi con una media di 10 ogni mezz'ora. Si può inoltre supporre che il numero di messaggi in arrivo segua una distribuzione di Poisson. Qual è la probabilità che nella prossima mezz'ora mi arrivino non più di 3 messaggi? E almeno 12?

➔ $X =$"n. messaggi ogni mezz'ora"$\sim$ Po$\,(\lambda = 10)$.

$$\begin{aligned}\mathbb{P}[X \leq 3] =& \mathbb{P}[X = 0] + \mathbb{P}[X = 1] + \mathbb{P}[X = 2] + \mathbb{P}[X = 3] \\ =& \frac{10^0}{0!}\,e^{-10} + \frac{10^1}{1!}e^{-10} + \frac{10^2}{2!}\,e^{-10} + \frac{10^3}{3!}\,e^{-10} \\ =& 0.0103\end{aligned}$$

➔ **Esercizio:** $\mathbb{P}[X \geq 12] = 1 - \mathbb{P}[X \leq 11] = \ldots \quad .$

Distribuzione di Poisson

Calcolo delle probabilità:

- Utilizzando **R** (o un qualsiasi pacchetto statistico) si può oggi lavorare con le principali funzioni di probabilità, di densità e di ripartizione.
- Una volta (e ancora adesso a volte...) si usavano invece delle tabelle preconfezionate, le **tavole della distribuzione**. Nelle tavole sono registrati i valori di alcune funzioni di ripartizione relativi a diversi valori dei parametri caratterizzanti la distribuzione.

Esempio:

$$X \sim \mathsf{Po}\,(10)$$

➔ $\mathbb{P}\,[X = 12] = 0.0948$

```
> dpois(x=12, lambda = 10)
[1] 0.09478033
```

Distribuzione di Poisson

- $\mathbb{P}\left[X \leq 12\right] = 0.7916$

```
> ppois(q=12, lambda = 10)
[1] 0.7915565
```

- $\mathbb{P}\left[X > 12\right] = 1 - 0.7916 = 0.2084$

```
> 1-ppois(q=12, lambda = 10)
[1] 0.2084435
```

- $\mathbb{P}\left[X \geq 12\right] = \mathbb{P}\left[X > 11\right] = 1 - 0.6968 = 0.3032$

```
> 1-ppois(q=11, lambda = 10)
[1] 0.3032239
```

Approssimazione Poisson per la binomiale

Quando $n \to \infty$ e $p \to 0$ ma in modo tale che il prodotto $np \to \lambda$ rimane costante, allora la funzione di probabilità di una v.a. binomiale di parametri n e p si può approssimare con la funzione di probabilità di una Poisson di parametro λ:

Approssimazione Poisson per la binomiale

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}$$

- L'approssimazione viene utilizzata nella pratica quando $n \geq 100$ e $p \leq 0.05$. Si sceglie allora $\lambda = np$.

Approssimazione Poisson per la binomiale

- Questa è nota anche come la **legge degli eventi rari**, perchè si può interpretare come la distribuzione del numero di eventi di tipo "successo" quando gli esperimenti bernoulliani sono tanti $(n \to \infty)$ ma la probabilità di successo è molto piccola $(p \to 0)$, ovvero, i successi sono rari.

Esempio: Una fabbrica di componenti elettronici fornisce il 3% dei chip acquistati da un produttore di telefoni cellulari. Qual è la probabilità che su 100 chip acquistati ve ne siano al massimo 3 provenienti da quella fabbrica?
$X =$ Numero di chip provenienti dalla fabbrica su 100. Then
$X \sim$ Bin $(100, 0.03)$ e:

$$\mathbb{P}\left[X \leq 3\right] = \sum_{k=0}^{3} \binom{100}{k} 0.03^k 0.97^{100-k} = 0.64724921$$

☞ È difficile da calcolare senza computer!

Approssimazione Poisson per la binomiale

Dato che $n = 100$ è grande e $p = 0.03$ è piccolo, si può utilizzare l'approssimazione con la Poisson di parametro $\lambda = np = 3$,

$$\sum_{k=0}^{3} \frac{3^k}{k!} e^{-3} = 0.64723189,$$

che è più facile da calcolare e molto precisa.

Distribuzioni congiunte

Introduzione

Fino ad ora ci siamo occupati delle distribuzioni di probabilità di variabili aleatorie prese singolarmente. Invece nella pratica si è spesso interessati a formulare delle affermazioni di tipo probabilistico riguardanti **due o più variabili aleatorie contemporaneamente**.

- Di seguito introduciamo i concetti fondamentali sulle distribuzioni congiunte, considerando il semplice caso di **coppie** di variabili aleatorie (X, Y) ➔ **distribuzioni bivariate**.
- Quello che diremo si può facilmente generalizzare a **vettori** di variabili aleatorie $(X_1, \ldots, X_n)$ di dimensione $n > 1$ ➔ **distribuzioni multivariate**.

Distribuzioni congiunte

funzione di ripartizione congiunta

La **funzione di ripartizione congiunta** della coppia di variabili aleatorie (X, Y) è una funzione $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ definita così:

$$F_{X,Y}(x, y) = \mathbb{P}\left[X \leq x, Y \leq y\right], \quad \forall (x, y) \in \mathbb{R}^2.$$

Distribuzioni congiunte

Dalla f.r. congiunta si possono facilmente ottenere le f.r. delle singole v.a. X e Y:

$$\begin{aligned} F_X(x) &= \mathbb{P}[X \leq x] \\ &= \mathbb{P}[X \leq x, Y < \infty] \\ &= \mathbb{P}\left[\lim_{y\to\infty}\{X \leq x, Y \leq y\}\right] \\ &= \lim_{y\to\infty}\mathbb{P}[X \leq x, Y \leq y] \\ &= \lim_{y\to\infty} F_{X,Y}(x,y), \end{aligned}$$

In modo simile,

$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y).$$

Distribuzioni congiunte

- Si noti quindi che la f.r. congiunta di (X, Y) contiene anche tutta l'informazione riguardante le distribuzioni di probabilità delle singole v.a. X e Y, ovvero le **distribuzioni marginali**.

- Invece il viceversa non è vero: non è possibile ricavare la f.r. congiunta a partire dalle marginali, a meno che non si abbia ulteriore informazione sulla relazione fra le due variabili.

Calcolo di probabilità congiunte

A tutte le domande riguardanti probabilità congiunte di (X, Y) si può rispondere usando la f.r. congiunta $F_{X,Y}(x, y)$.

Ad esempio,

$$\begin{aligned}&\mathbb{P}[x_1 < X \leq x_2,\, y_1 < Y \leq y_2]\\ &\qquad = F_{X,Y}(x_2, y_2) + F_{X,Y}(x_1, y_1) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1).\end{aligned}$$

Funzione di probabilità congiunta

Funzione di probabilità congiunta

Se le v.a. X e Y sono entrambe discrete, si definisce la **funzione di probabilità congiunta**

$$p_{X,Y}(x,y) = \mathbb{P}\left[X = x, Y = y\right], \quad \forall (x,y) \in \mathbb{R}^2.$$

Le **funzioni di probabilità marginali** si ottengono sommando rispetto a tutti i valori possibili dell'altra variabile:

$$p_X(x) = \sum_{y : p(x,y) > 0} p_{X,Y}(x,y)$$

e

$$p_Y(y) = \sum_{x : p(x,y) > 0} p_{X,Y}(x,y).$$

Funzione di probabilità congiunta

La funzione di probabilità congiunta e le marginali si possono rappresentare nella stessa **tabella a due entrate**:

		Y			
		y_1	y_2	$\dots$	$p_X(x)$
X	x_1	$p_{X,Y}(x_1, y_1)$	$p_{X,Y}(x_1, y_2)$	$\dots$	$p_X(x_1)$
	x_2	$p_{X,Y}(x_2, y_1)$	$p_{X,Y}(x_2, y_2)$	$\dots$	$p_X(x_2)$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	$p_Y(y)$	$p_Y(y_1)$	$p_Y(y_2)$	$\dots$	1

Funzione di probabilità congiunta

Esempio: Un programma consiste di due moduli. Il numero di errori X nel primo modulo e il numero di errori Y nel secondo modulo hanno distribuzione congiunta: $p_{X,Y}(0,0) = p_{X,Y}(0,1) = p_{X,Y}(1,0) = 0.2$, $p_{X,Y}(1,1) = p_{X,Y}(1,2) = p_{X,Y}(1,3) = 0.1$, $p_{X,Y}(0,2) = p_{X,Y}(0,3) = 0.05$.

☞ Per prima cosa rappresentiamo la funzione di probabilità congiunta e le marginali in una tabella:

		Y				
		0	1	2	3	$p_X(x)$
X	0	0.20	0.20	0.05	0.05	0.5
	1	0.20	0.10	0.10	0.10	0.5
	$p_Y(y)$	0.40	0.30	0.15	0.15	1

Funzione di probabilità congiunta

- La probabilità che non ci siano errori nel primo modulo, si calcola sommando tutte le probabilità congiunte della riga corrispondente a $X = 0$:

$$\mathbb{P}[X = 0] = \sum_{y=0}^{3} p_{X,Y}(0, y) = 0.20 + 0.20 + 0.05 + 0.05 = 0.5.$$

- La probabilità che nel primo modulo ci siano meno errori che nel secondo modulo è

$$\begin{aligned}\mathbb{P}[X < Y] &= 1 - \mathbb{P}[X \geq Y] \\ &= 1 - \mathbb{P}[X = 0, Y = 0] - \mathbb{P}[X = 1, Y = 0] - \mathbb{P}[X = 1, Y = 1] \\ &= 1 - (0.2 + 0.2 + 0.1) = 0.5.\end{aligned}$$

Variabili aleatorie indipendenti

Il concetto di indipendenza di eventi, già visto in probabilità elementare, si estende anche alle v.a.

Variabili aleatorie indipendenti

Si dice che due v.a. sono **indipendenti** se

$$\mathbb{P}\left[X \in A, Y \in B\right] = \mathbb{P}\left[X \in A\right]\mathbb{P}\left[Y \in B\right], \quad \forall A, B \subset \mathbb{R}.$$

Si può dimostrare che questo è equivalente a dire che:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \quad \forall (x,y) \in \mathbb{R}^2,$$

ovvero

$$p_{X,Y}(x,y) = p_X(x)p_Y(y), \quad \forall (x,y) \in \mathbb{R}^2$$

Esempio: Torniamo all'esempio del numero di errori nei due moduli di un programma:

Variabili aleatorie indipendenti

		Y				
		0	1	2	3	$p_X(x)$
X	0	0.20	0.20	0.05	0.05	0.5
	1	0.20	0.10	0.10	0.10	0.5
	$p_Y(y)$	0.40	0.30	0.15	0.15	1

Sono indipendenti le due v.a. X e Y?

☞ No, perchè, ad esempio,

$$p_{X,Y}(1,1) = 0.10 \neq p_X(1)p_Y(1) = 0.50 \cdot 0.30.$$

Variabili aleatorie indipendenti

Esempio: Consideriamo il caso in cui X e Y sono v.a. indipendenti con funzioni di probabilità marginali:

x	0	1	2	3
$p_X(x)$	0.5	0.3	0.1	0.1
$p_Y(x)$	0.7	0.2	0.1	0

Se sono note le distribuzioni marginali di X e Y e inoltre si sa che le due variabili sono indipendenti, allora è possibile ricostruire la distribuzione congiunta di (X, Y).

Variabili aleatorie indipendenti

- La funzione di probabilità congiunta si può calcolare sfruttando l'indipendenza, in questo modo:

$$p_{X,Y}(0,0) = \mathbb{P}[X = 0, Y = 0] = p_X(0) \cdot p_Y(0) = 0.5 \cdot 0.7 = 0.35,$$

$$p_{X,Y}(0,1) = p_X(0) \cdot p_Y(1) = 0.5 \cdot 0.2 = 0.1,$$

e così via.
La tabella delle probabilità congiunte è dunque:

		Y				
		0	1	2	3	$p_X(x)$
	0	0.35	0.10	0.05	0	0.5
X	1	0.21	0.06	0.03	0	0.3
	2	0.07	0.02	0.01	0	0.1
	3	0.07	0.02	0.01	0	0.1
	$p_Y(y)$	0.7	0.2	0.1	0	1

Variabili aleatorie indipendenti

- Con la funzione di probabilità congiunta possiamo calcolare tutte le probabilità riguardanti X e Y congiuntamente.
  Possiamo anche calcolare, ad esempio, la distribuzione della v.a. $X + Y$. Infatti:

$$\mathbb{P}\left[X + Y = 0\right] = \mathbb{P}\left[X = 0, Y = 0\right] = 0.35,$$

$$\begin{aligned} \mathbb{P}\left[X + Y = 1\right] &= \mathbb{P}\left[X = 0, Y = 1\right] + \mathbb{P}\left[X = 1, Y = 0\right] \\ &= 0.21 + 0.10 = 0.31 \end{aligned}$$

  e così via.

Variabili aleatorie indipendenti

☞ Completate i calcoli e controllate che

$$p_{X+Y}(w) = \begin{cases} 0.35 & \text{if } w = 0 \\ 0.31 & \text{if } w = 1 \\ 0.18 & \text{if } w = 2 \\ 0.12 & \text{if } w = 3 \\ 0.03 & \text{if } w = 4 \\ 0.01 & \text{if } w = 5. \end{cases}$$

Distribuzioni condizionate

Funzioni di probabilità condizionata

Siano X e Y due v.a. discrete con funzione di probabilità congiunta $p_{X,Y}(x,y)$.
La **funzione di probabilità condizionata** di X dato $Y = y$ è

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \quad \forall y \text{ t.c. } p_Y(y) > 0.$$

Allo stesso modo, la **funzione di probabilità condizionata** di Y dato $X = x$ è

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}, \quad \forall x \text{ t.c. } p_X(x) > 0.$$

Distribuzioni condizionate

☞ In effetti, ad esempio, se $p_Y(y) > 0$,

$$\begin{aligned} p_{X|Y}(x|y) &= \mathbb{P}\left[X = x | Y = y\right] \\ &= \frac{\mathbb{P}\left[X = x, Y = y\right]}{\mathbb{P}\left[Y = y\right]} = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \end{aligned}$$

➔ Se X e Y sono indipendenti, allora

$$p_{X|Y}(x|y) = p_X(x) \text{ e } p_{Y|X}(y|x) = p_Y(y).$$

Esempio: Siano X e Y due v.a. discrete tali che: $p_{X,Y}(0,0) = 0.4$, $p_{X,Y}(0,1) = 0.2$, $p_{X,Y}(1,0) = 0.1$, $p_{X,Y}(1,1) = 0.3$. Allora,

		Y		
		0	1	$p_X(x)$
X	0	0.4	0.2	0.6
	1	0.1	0.3	0.4
	$p_Y(y)$	0.5	0.5	1

Distribuzioni condizionate

- La funzione di probabilità di X dato $Y = 1$ si calcola utilizzando le informazioni contenute nella colonna corrispondente a $Y = 1$:

$$p_{X|Y}(0|1) = \frac{p_{X,Y}(0,1)}{p_Y(1)} = \frac{0.2}{0.5} = 0.4$$

e

$$p_{X|Y}(1|1) = \frac{p_{X,Y}(1,1)}{p_Y(1)} = \frac{0.3}{0.5} = 0.6,$$

in modo che

$$X|Y = 1 \sim p_{X|Y}(x|1) = \begin{cases} 0.4 & \text{if } x = 0 \\ 0.6 & \text{if } x = 1 \end{cases}$$

Essendo questa distribuzione diversa dalla marginale di X, concludiamo che le due variabili non sono indipendenti.

Distribuzioni condizionate

☞ **Nota:** La **legge delle probabilità totali** applicata alle variabili casuali ci permete di trovare la distribuzione marginale di una variabile, diciamo X a partire dalla marginale dell'altra, diciamo Y e la corrispondente distribuzione condizionata:

$$\begin{aligned}\mathbb{P}\left[X=x\right] &= \sum_y \mathbb{P}\left[X=x, Y=y\right] \\ &= \sum_y \mathbb{P}\left[X=x|Y=y\right]\mathbb{P}\left[Y=y\right].\end{aligned}$$