Elementi di Matematica e Statistica

Le Variabili Casuali

Docente: Riccardo Ievoli
riccardo.ievoli@unife.it

Corso di Laurea in Biotecnologie

12/11/2025

Indice

Introduzione alle Variabili Casuali

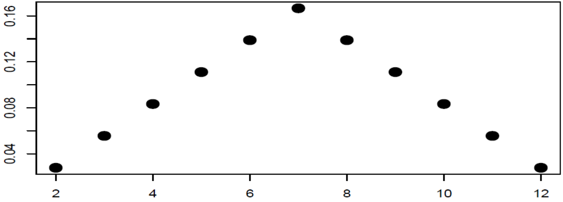
- Si considerano tutti i possibili eventi di una prova
- Si trasformano gli eventi in valori numerici ($E_1 = 1$, $E_2 = 2$, $\ldots$)
- Si associa una probabilità a ciascun evento $\Rightarrow$ variabile casuale "vc", oppure rv (*random variable*).

Due tipologie di variabili casuali

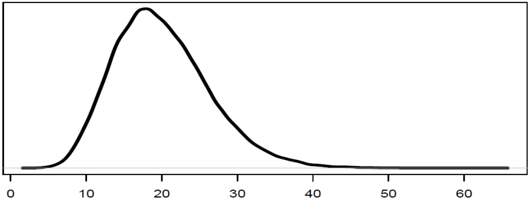
1. **Discrete**: considerano un numero finito di eventi
   - Presentano una distribuzione di probabilità
   - $f(x)$ è un vettore di probabilità
2. **Continue**: considerano un numero infinito di eventi
   - Presentano una funzione di densità
   - $f(x)$ è una funzione continua

In ogni caso, la somma delle probabilità associate a tutti i possibili eventi **DEVE** essere pari a 1. Nel caso continuo si tratta di un **integrale**

- Variabile casuale discreta (somma di due dadi)



- Variabile casuale continua (Chi quadrato)



- Un esempio: variabile casuale "lancio di un dado"
- Sei possibili eventi (Esce il numero 1,...,Esce il numero 6)
- Ecco la distribuzione di probabilità:

Eventi	Probabilità
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$

Nelle variabili casuali sono importanti alcuni valori "medi" detti momenti.
Questi valori caratterizzano le variabili aleatorie

Valore Atteso

- Detto anche media o speranza.
- Si indica con E(.) (*expectation*)
- Rappresenta la media **ponderata** di tutti i possibili risultati

Varianza

- Si può indicare con Var(.) oppure con σ^2
- Si dice anche momento rispetto alla media
- Rappresenta un parametro di variabilità (*scale parameter*)

- **Valore Atteso**

$$E(X) = \mu_X = \sum xf(x)$$

$$E(X) = \mu_X = \int xf(x)dx$$

- **Varianza**

$$\begin{aligned} Var(X) &= \sigma_X^2 = E[(X - E(X))^2] \\ &= E(X^2) - E(X)^2 \\ &= E[(X - \mu_X)^2] \end{aligned}$$

Valore Atteso

- Ricordiamo che la media può essere espressa come $\sum_i x_i f_i$
- Il valore atteso di una costante è la costante: $E(c) = c$
- Linearità: $E(a + bX) = a + bE(X)$
- Se X e Y sono indipendenti: $E(XY) = E(X) \cdot E(Y)$

Varianza

- La varianza di una costante è zero: $Var(c) = 0$
- Se $Y = a + bX$ allora $Var(Y) = Var(a + bX) = b^2 Var(X)$

- Variabile casuale “lancio di un dado”
- Ci sono sei eventi con probabilità pari a $1/6$
- Ciò vuol dire che $f_i = p_i = 1/6$ per $i = 1, 2, \ldots, 6$
- Calcolare il suo valore atteso:

$$E(X) = \sum_{i=1}^{6} x_i p_i$$

- Calcolare il valore atteso dei quadrati $E(X^2) = \sum_{i=1}^{6} x_i^2 p_i$
- Calcolare la varianza come:

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

Esercizio 1

- X è la variabile aleatoria "somma di due dadi"
- Scrivere la distribuzione di probabilità
- Calcolare E(X) e Var(X)

Esercizio 2

- Y è la variabile aleatoria $Y = 20 - 7 \cdot X$, dove X è la v.c. dell'esercizio precedente
- Calcolare E(X) e Var(X)

Variabili Casuali Discrete

Introduzione

La caratterizzazione di una variabile casuale (o aleatoria) avviene attraverso la funzione di probabilità (o di massa di probabilità):

$$f(x) = P(X = x)$$

Tale funzione associa, ad ogni **modalità** della variabile X, la **probabilità** con cui tale evento si verifica nell'esperimento aleatorio.

Può essere vista come il corrispettivo teorico della **frequenza relativa**.

Le variabili casuali possono essere sintetizzate attraverso i **momenti**:

- **Valore atteso**: corrispettivo della media aritmetica.
- **Varianza**: corrispettivo della varianza descrittiva

Variabili Casuali Discrete

Modelli

- I valori assunti da X possono essere inclusi in un sottoinsieme dell'insieme dei numeri interi chiamato **dominio**
- È possibile esprimere la funzione di probabilità mediante un'espressione matematica che dipende sia da X che da uno o più **parametri**
- I parametri hanno un ruolo cruciale: diversi valori possono portare a distribuzioni teoriche aventi caratteristiche completamente diverse
- Valore atteso e varianza rappresentano alcuni tra i parametri più noti nella teoria delle variabili aleatorie.

Ogni modello di probabilità deve sempre e comunque rispettare la seguente equivalenza:

$$\sum_{x} P(X = x) = 1.$$

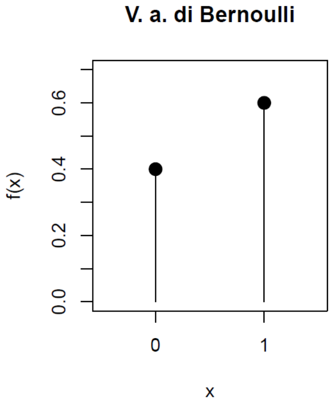
Variabili Casuali Discrete

Distribuzione di Bernoulli

- Modello più semplice: il suo **dominio** è $x \in 0, 1$
- Rispettivamente *insuccesso* (0) e *successo* (1)
- Il **parametro** è la probabilità di successo p (in figura $p = 0,6$)

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

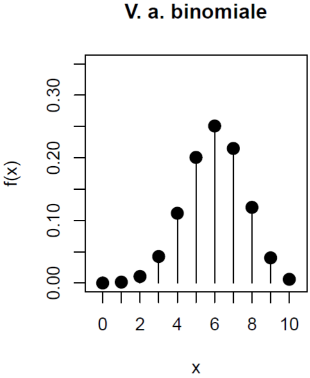
con $x \in \{0, 1\}$



È definita come la sequenza di n prove di Bernoulli **indipendenti** aventi probabilità di successo p (in figura $p = 0,6$ e $n = 10$). Il valore di X è il numero di prove in cui si è registrato il successo, quindi $x \in \{0, 1, ..., n\}$.

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

con $x = 1, 2, \ldots, n$



Variabili Casuali Discrete

Distribuzione Binomiale: Focus Fattoriali

Funzione di probabilità della Distribuzione Binomiale:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 1, 2, \ldots, n$$

- $\binom{n}{x}$ è il coefficiente binomiale:

$$\frac{n!}{x!(n-x)!}$$

- $n!$ è detto Fattoriale di n ed è pari a:

$$\prod_{i=1}^{n} i = 1 \cdot 2 \cdots (n-1) \cdot n$$

Si consideri che $0! = 1$

- I momenti della Distribuzione Binomiale si ottengono a partire da quelli della Bernoulli, ossia:

$$E(X) = p; \quad Var(X) = p(1-p)$$

- Dimostrazione: basta partire dalla definizione di **valore atteso** e di **varianza**

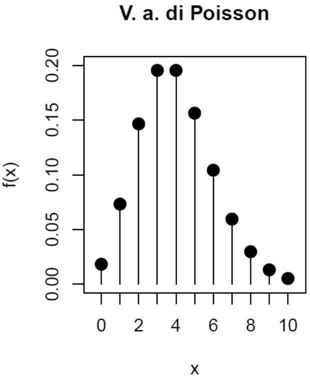
- Pertanto, essendo la Binomiale una somma di Bernoulli, i momenti sono i seguenti:

$$E(X) = n \cdot p; \quad Var(X) = np(1-p)$$

Detta anche **legge degli eventi rari**. Spesso è utilizzata per modellare conteggi $x \in \{0, 1, \ldots\}$. La sua espressione è governata da un parametro (λ) che coincide con **valore atteso** e **varianza** (nel grafico $\lambda = 4$).

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

con $x = 1, 2, \ldots$



La probabilità di successo di un'operazione chirurgica (senza complicanze post-operatorie) è pari a $p = 0,80$. Considerando un campione di $n = 5$ operazioni effettuate (ossia 5 pazienti operati), si risponda alle seguenti domande

- Qual è la probabilità che non abbia successo nessuna operazione?
- Qual è la probabilità che abbia successo esattamente una operazione?
- Qual è la probabilità che abbiano successo esattamente due operazioni?
- Qual è la probabilità che abbiano successo esattamente tre operazioni?

Suggerimento: bisogna partire dalla funzione di probabilità della Binomiale con $n = 5$ e $p = 0,8$.

In uno studio di un certo organismo acquatico, si ha che il numero medio di organismi per ogni campione d'acqua è pari a 2. Assumendo che tale conteggio segua una distribuzione di Poisson, calcolare:

- La probabilità che in un campione si trovino esattamente 0 organismi
- La probabilità che in un campione si trovi un solo organismo
- La probabilità che in un campione si trovino esattamente due organismi

Suggerimento: considerando $\lambda = 2$, nel primo esercizio bisogna partire da:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{2^0 e^{-2}}{0!}$$

Variabili Casuali Discrete

Funzioni in Excel

Come calcolare la probabilità di un evento (o di una unione eventi) tramite Excel.

- `DISTRIB.POISSON(x;MEDIO;CUMULATIVO)`
- `DISTRIB.BINOM(x;PROVE;Probabilità_s; CUMULATIVO)`

`MEDIO`: λ; `PROVE`: n; Probabilità_s: p

Valori per `CUMULATIVO`:

- Se `FALSO` $\Rightarrow P(X = x)$
- Se `VERO` $\Rightarrow P(X \leq x)$

- Se $X \sim Poi(2)$, calcolare P(X=x) con x=2:
  `DISTRIB.POISSON(2;2;FALSO)`

- Se $X \sim Bin(5; 0, 7)$, calcolare P(X=x) con x=3:
  `DISTRIB.BINOM(3;5;0.7;FALSO)`

- Se $X \sim Poi(2)$, calcolare P($X \leq x$) con x=2:
  `DISTRIB.POISSON(2;2;VERO)`

- Se $X \sim Bin(5, 0, 7)$, calcolare P($X \leq x$) con x=2:
  `DISTRIB.BINOM(2;2;VERO)`

Variabili Casuali Continue

Introduzione

In questo caso X può assumere qualunque valore reale nell'intervallo in cui è definita.

La caratterizzazione avviene attraverso la funzione di densità di probabilità, $f(x)$, che può essere vista come il corrispettivo teorico dell'istogramma. Si noti che:

- Non ha più senso valutare la probabilità che un unico valore x sia assunto: ognuno ha massa di probabilità nulla!
- Ha senso ragionare in termini di probabilità in un certo intervallo di valori: sarà definito in termini di area sotto una curva.
- L'area al di sotto della curva definita da $f(x)$ è pari ad 1.
- Anche in questo caso si possono definire **valore atteso** e **varianza**.

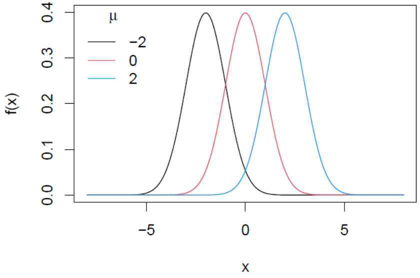
Variabili Casuali Continue

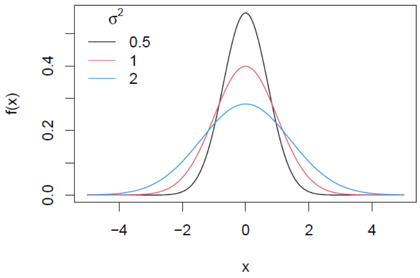
La Distribuzione Gaussiana

Detta anche distribuzione normale, la sua forma distributiva è un po' complessa, ma è specificata da due parametri:

- μ che definisce il **valore atteso**
- σ^2 che definisce la **varianza**.

Si tratta di una distribuzione *unimodale* e *simmetrica* e solitamente si indica con $X \sim N(\mu, \sigma^2)$.





X si distribuisce come una Normale, si scrive: $X \sim N(\mu; \sigma^2)$

Una v.c. normale è specificata attraverso i due parametri:

1. La media μ con $\mu \in (-\infty, +\infty)$
2. La varianza σ^2 con $\sigma^2 \geq 0$

In una v.c. normale $\mu =$ Mediana $=$ Moda

La varianza è non negativa e finita $\sigma^2 < +\infty$

Molti fenomeni discreti possono essere approssimati ad una normale per n "grande"

Variabili Casuali Continue

La Distribuzione Gaussiana: altre caratteristiche

- La distribuzione è simmetrica, cioè ha un coefficiente di **asimmetria** pari a 0
- Questo significa che $\mu = Med$
- Ha una forma campanulare $\rightarrow$ Kurtosi $= 3$
- Le code decadono esponenzialmente
- Il 68% degli eventi ricade nell'intervallo $\mu \pm \sigma$
- Il 95% degli eventi ricade nell'intervallo $\mu \pm 2 \cdot \sigma$
- $\mu \pm \sigma$ sono anche punti di esso (cambia la concavità)
- La sua distribuzione di probabilità è pari a:

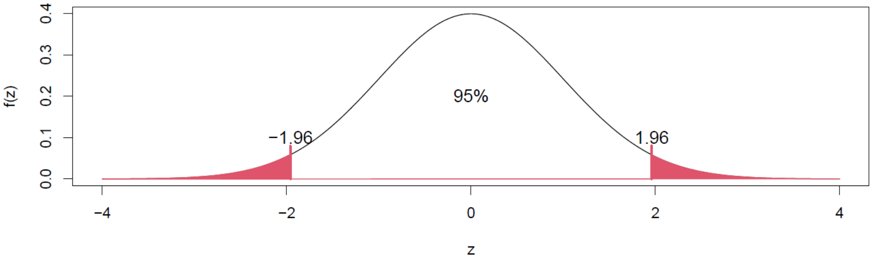
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Se $X \sim N(\mu, \sigma^2)$, allora:

$$Z = \frac{X - \mu}{\sigma}$$

è una normale standardizzata, avente quindi valore atteso nullo ($\mu = 0$) e varianza unitaria $\sigma^2 = 1$.

Utilità pratica: invece di calcolare ogni volta valori specifici, si riconduce tutto alla variabile standardizzata che assume sempre gli stessi valori.



Variabili Casuali Continue

La Distribuzione Gaussiana (2)

Bisogna utilizzare le tavole della gaussiana per risolvere gli esercizi dopo aver standardizzato. Infatti i valori della $\mathrm{N}(0,1)$ sono tabulati. Possibili casi:

- $P(X < x) = P(Z < z) = \phi(z)$
- $P(X > x) = P(Z > z) = 1 - P(Z < z) = 1 - \phi(z)$
- $P(x_1 < X < x_2) = P(x_1 < X < x_2) = P(z_1 < Z < z_2) = \phi(z_2) - \phi(z_1)$

Variabili Casuali Continue

La Distribuzione Gaussiana (3)

Come calcolare $P(X < x)$ per la Gaussiana in Excel?
Media e <u>deviazione standard</u> note:

```
DISTRIB.NORM.N (x; MEDIA; DEVIAZIONE STANDARD; VERO)
```

Normale Standard N$(0, 1)$:

```
DISTRIB.NORM.N.ST (x; VERO)
```

Esempi

- $X \sim N(10, 4)$, voglio calcolare $P(X < 12)$
  `DISTRIB.NORM.N (12; 10; 2; VERO)`
- $X \sim N(0, 1)$, voglio calcolare $P(X < -1)$
  `DISTRIB.NORM.N.ST (-2; VERO)`

Avendo dei dati, è sempre possibile ricondurli ad una distribuzione che presenti un valore medio (stimato o descrittivo) $\hat{\mu} = 0$ e una deviazione standard (e varianza) pari a $\hat{\sigma}^2 = 1$.

Questo procedimento permette anche di **confrontare** due o più distribuzioni con diverse unità di misura.

Data una variabile X con valori osservati $x_1, \ldots, x_n$, e avente media e varianza (poniamo per semplicità calcolate sui dati) pari a $\hat{\mu}$ e $\hat{\sigma}^2$, il procedimento è il seguente $\forall i$:

$$z_i = \frac{x_i - \hat{\mu}}{\hat{\sigma}}$$

Variabili Casuali Continue

Approfondimento: La standardizzazione in statistica (2)

La costruzione di **indici compositi**: il processo

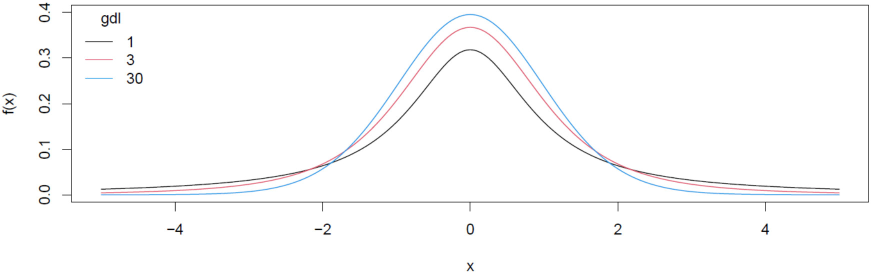
1. Scelgo alcune variabili ritenute rilevanti a misurare un fenomeno (Es. "Benessere Economico" o "Ambientale") su alcune unità statistiche (es. le Regioni Italiane, le Nazioni Europee)
2. Applico a tutte le variabili la **standardizzazione** (o "normalizzazione") $z_i = (x_i - \hat{\mu})\hat{\sigma}^{-1}$
3. Per ogni unità statistica applico una funzione di **aggregazione**, come ad esempio una **media aritmetica** semplice o ponderata

Esempio: il BES di ISTAT `https://www.istat.it/statistiche-per-temi/focus/benessere-e-sostenibilita/la-misurazione-del-benessere-bes/gli-indicatori-del-bes/`

Variabili Casuali Continue

La Distribuzione t-Student

- È simile alla normale: unimodale e simmetrica.
- È **standardizzata**: ha valore atteso nullo e varianza unitaria.
- Possiede un solo parametro: i **gradi di libertà** (gdl) legati alla numerosità del campione. Essi gestiscono l'andamento delle code: più il valore è basso più le code decadono lentamente



La sua funzione di densità è (si provi a disegnare il grafico):

$$f(x) = \frac{1}{b-a} \quad \forall a \leq x \leq b$$

Momenti dell'Uniforme

- **Valore Atteso**: $E(X) = \dfrac{a+b}{2}$
- **Varianza**: $Var(X) = \dfrac{b-a}{12}$

Esercizi

L'età misurata in anni dei dipendenti della Pubblica Amministrazione di distribuisce come una gaussiana con valore atteso pari a 48 e varianza pari a 36. Estraendo a caso 100 dipendenti pubblici, qual è il valore atteso del numero di dipendenti con età inferiore a 40 anni (approssimando all'intero più vicino)?

L'età misurata in anni dei dipendenti della Pubblica Amministrazione di distribuisce come una gaussiana con valore atteso pari a 48 e varianza pari a 36. Estraendo a caso un dipendente pubblico, qual è la probabilità che abbia un'età superiore ai 45 anni?

L'età misurata in anni dei dipendenti della Pubblica Amministrazione di distribuisce come una gaussiana con valore atteso pari a 48 e varianza pari a 36. Estraendo a caso un dipendente pubblico, qual è la probabilità che abbia un'età compresa tra i 45 e i 55 anni?

Cosa abbiamo imparato oggi?

1. Concetti fondamentali legati alle variabili casuali
2. Momenti di una variabile casuale: valore atteso e varianza
3. Variabili casuali discrete: Bernoulli, Binomiale, Poisson
4. Variabili casuali continue: La gaussiana (e l'uniforme continua)

Esercizi per casa

Variabili Casuali Continue

Si supponga di avere un campione di cellule provenienti da un esperimento. La concentrazione di una particolare proteina all'interno di queste cellule segue una distribuzione gaussiana con una media di 200 microgrammi per millilitro (µg/ml) e una deviazione standard di 25 µg/ml.

- Qual è la probabilità che una cellula scelta a caso abbia una concentrazione (di proteina) superiore a 230 µg/ml?
- Qual è la probabilità che una cellula scelta a caso abbia una concentrazione superiore a 250 µg/ml?
- Qual è la probabilità che una cellula scelta a caso abbia una concentrazione compresa tra 210 µg/ml e 225 µg/ml?

Esercizi per casa

Variabili Casuali discrete

Il numero di accessi ad un sito web in un minuto segue una distribuzione di Poisson con parametro (valore atteso) $\lambda = 5$ (cinque accessi in "media" per minuto). Calcolare:

- La probabilità che in un minuto ci siano esattamente 6 accessi
- La probabilità che in un minuto ci siano esattamente 3 accessi
- Il numero di accessi atteso in un'ora
- La deviazione standard del numero di accessi
- La probabilità che in un minuto ci siano meno di 4 accessi
- La probabilità che in un minuto ci siano almeno 3 accessi.
- La probabilità che in un minuto ci sia un numero di accessi compreso tra 3 e 6 (entrambi inclusi nell'intervallo).
- La probabilità che in un minuto ci sia un numero di accessi pari a 4 o 5