

FormularioSAD

Nicolò Rosati

July 3, 2023

1 Statistica Descrittiva

1.1 Media campionaria

È un **operatore lineare**: se ciascun valore viene traslato e/o scalato per un valore la media campionaria subisce la stessa trasformazione.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

1.2 Scarti

Differenze tra ciascun valore dei dati e la loro media

$$\sum_{i=1}^n (x_i - \bar{x})$$

1.3 Indici di dispersione

1.3.1 Varianza campionaria

Espressa con un'unità di misura al quadrato rispetto all'originale. La varianza campionaria dei dati scalati e/o traslati non è la varianza campionaria scalata e/o traslata dei dati, è bensì elevata al quadrato

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i^2 - n\bar{x}^2)$$

1.3.2 Deviazione standard

$$s = \sqrt{s^2}$$

1.3.3 Covarianza campionaria

Misura la relazione tra due variabili. È > 0 se vi è relazione diretta, < 0 in caso di relazione inversa

$$cov(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

1.3.4 Indice di correlazione lineare

Misura anch'esso la relazione tra due variabili

$$\rho = \frac{cov(x, y)}{s_x s_y}$$

È tale che $-1 \leq \rho \leq 1$. Dove valore positivo indica **relazione diretta** e un valore negativo indica una **relazione inversa**

1.3.5 Coefficiente di variazione

Permette di utilizzare dati molto dispersivi sul box plot, li porta tutti su una certa scala

$$s^* = \frac{s}{|\bar{x}|}$$

1.4 Indici di eterogeneità

È necessario avere un indice che misuri la dispersione della distribuzione delle frequenze, detta eterogeneità. In particolare diremo che una variabile si distribuisce in modo eterogeneo se ogni suo valore si presenta con la stessa frequenza.

1.4.1 Indice di eterogeneità di Gini

$$I = 1 - \sum_{j=1}^s f_j^2$$

È tale che $-1 \leq \rho \leq 1$. In caso di **eterogeneità minima** tutti gli elementi del campione assumono lo stesso valore $\Rightarrow I = 0$. Mentre in caso di **eterogeneità massima** tutte le osservazioni hanno la medesima frequenza $\Rightarrow I = \frac{s-1}{s}$.

La versione normalizzata equivale a

$$I' = \frac{s \cdot I}{s - 1}$$

1.4.2 Entropia

$$H = \sum_{i=1}^s f_i \log \frac{1}{f_i} = - \sum_{i=1}^s f_i \log f_i$$

È tale che $H = 0 \iff f_i = 0 \vee \log_2 \frac{1}{f_i} = 0 \iff f_i = 1 \vee f_i = 0$. Ovvero se e solo se non ho mai osservato quello specifico valore o ho osservato solo quello. La versione normalizzata equivale a:

$$H' = \frac{H}{\log s}$$

2 Calcolo combinatorio

2.1 Permutazioni

Preso un insieme di n oggetti $A = \{a_1, \dots, a_n\}$. Chiamiamo permutazione degli n oggetti una **sequenza ordinata** in cui compaiono tutti e soli gli oggetti di A .

2.1.1 Permutazioni semplici

Se gli n oggetti di A sono tutti distinguibili, allora si parla di permutazione semplice, o soltanto permutazione, degli n oggetti. In particolare sono organizzati in una sequenza in cui **conta l'ordine**. Il numero totale di permutazioni ottenibili è:

$$p_n = \prod_{i=1}^n i = n! \tag{1}$$

2.1.2 Permutazioni di oggetti distinguibili a gruppi (con ripetizione)

Se gli n oggetti di A non sono tutti distinguibili, ma sono distinguibili a gruppi di numerosità n_1, n_2, \dots, n_k , allora una sequenza ordinata di tali oggetti che sia distinguibile dalle altre è detta permutazione di oggetti distinguibili a gruppi.

In particolare, sia $P_{n;n_1, \dots, n_k}$ il numero di configurazioni differenti, si può scrivere

$$n! = P_{n;n_1, \dots, n_k} \cdot n_1! \cdot n_2! \cdot \dots \cdot n_k!$$

e di conseguenza, il numero delle permutazioni di n oggetti distinguibili a gruppi di numerosità n_1, n_2, \dots, n_k è:

$$P_{n;n_1, \dots, n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!} := \binom{n}{n_1!, n_2!, \dots, n_k!}. \tag{2}$$

2.2 Disposizioni

Nelle disposizioni l'ordine conta

2.2.1 Disposizioni senza ripetizione

Si calcola il numero $d_{n,k}$ di possibili disposizioni senza ripetizione di n oggetti distinti su k posti

$$d_{n,k} = \frac{n!}{(n-k)!} \quad (3)$$

2.2.2 Disposizioni con ripetizione

Per calcolare il numero $D_{n,k}$ di possibili disposizioni con ripetizione di n oggetti distinti su k posti:

$$D_{n,k} = \underbrace{n \cdot n \cdot \dots \cdot n}_{k \text{ volte}} = n^k. \quad (4)$$

2.3 Combinazioni

2.3.1 Combinazioni senza ripetizione

$$c_{n,k} = \frac{d_{n,k}}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k} \quad (5)$$

2.3.2 Combinazioni con ripetizione

$$C_{n,k} = \binom{n+k-1}{k} \quad (6)$$

2.4 Algebra degli eventi

Si definisce lo spazio degli esiti come l'insieme di tutti gli esiti possibili Ω . I sottoinsiemi di Ω prendono il nome di eventi.

Si definisce l'algebra degli eventi come $a = \{E_1 \dots E_n\} \forall E \in a \ E \in \Omega$. In particolare:

- $\Omega \in a$
- $\forall E \in a \Rightarrow \bar{E} \in a$
- $\forall E, F \in a \rightarrow E \cup F \in a$

2.5 Assiomi di Kolmogorov

1. $\forall E \in a \ 0 \leq P(E) \leq 1$
2. $P(\Omega) = 1$
3. $\forall E, F \in a$ con $E \cap F = \emptyset \Rightarrow P(E \cup F) = P(E) + P(F)$: ovvero la probabilità dell'unione di eventi disgiunti equivale alla somma delle singole probabilità. Altrimenti: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

2.6 Probabilità condizionata

Si definisce la probabilità condizionata di E dato F come:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \quad (7)$$

assumendo che $P(F)$ non sia l'evento impossibile. Analogamente

$$P(F|E) = \frac{P(E \cap F)P(F)}{P(E)} \quad (8)$$

$$P(E|F) = P(F|E) \frac{P(F)}{P(E)} \quad (9)$$

$$P(E \cap F) = P(E|F)P(F) \quad (10)$$

2.7 Teorema delle probabilità totali

Siano $E, F \in \mathcal{A}$. È possibile esprimere E come: $E = E \cap (F^c \cup F) = (E \cap F^c) \cup (E \cap F) = (E \cap E) \cap (F \cap F^c) = \emptyset$. Essendo quindi E ed F disgiunti è possibile applicare il terzo assioma di Kolmogorov ed ottenere la **formula di Bayes**:

$$\begin{aligned} P(E) &= P(E \cap F^c) + P(E \cap F) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \end{aligned} \quad (11)$$

Il quale può essere esteso a:

$$P(F) = \sum_{i=1}^n P(F|E_i)P(E_i) \quad \text{con } E_i \neq \emptyset \quad (12)$$

2.8 Teorema di Bayes

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} \quad (13)$$

2.9 Eventi indipendenti

Due eventi E ed F sono indipendenti se $P(E|F) = P(E)$, ovvero se la conoscenza che F si è verificato non cambia la probabilità di E . In particolare E è indipendente da F se:

$$P(E \cap F) = P(E)P(F) = P(F \cap E) \quad (14)$$

2.10 Sistemi in serie

Sono sistemi che funzionano solo se tutte le componenti funzionano.

$$\begin{aligned} P(\text{sistema funzioni}) &= P\left(\bigcap_{i=1}^n \{\text{i-esima comp funzioni}\}\right) \\ &= \prod_{i=1}^n P(\text{i-esima comp funzioni}) = \prod_{i=1}^n P_i \end{aligned} \quad (15)$$

2.11 Sistemi in parallelo

Sono sistemi che funzionano fino a che almeno uno degli n componenti distinti funziona. Quindi la probabilità che l'intero sistema funzioni è:

$$\begin{aligned} P(\text{sistema funziona}) &= P(\text{almeno 1 comp funziona}) \\ &= 1 - P(\text{nessuna comp funziona}) \\ &= 1 - P\left(\bigcap_{i=1}^n \text{la comp i-esima non funziona}\right) \\ &= 1 - \prod_{i=1}^n P(\text{la comp non funziona}) \\ &= 1 - \prod_{i=1}^n (1 - P_i) \end{aligned}$$

3 Variabili aleatorie

Si definisce $X : \Omega \rightarrow \mathbb{R}$, ovvero $P(X = \alpha) = P(\{\omega \in \Omega : X(\omega) = \alpha\})$ Quindi: $\forall \alpha \in \mathbb{R} X^{-1}(\alpha) = \{\omega \in \Omega : X(\omega) \leq \alpha\} \in \mathcal{a}$

3.1 Funzione di ripartizione

Esprime la probabilità che la variabile aleatoria X assuma un valore minore o uguale ad x .

$$\begin{aligned} F_X &: \mathbb{R} \rightarrow [0, 1] \\ F_X(x) &:= P(X \leq x) \\ P(a < x \leq b) &= F_X(b) - F_X(a) \end{aligned}$$

3.2 Variabili aleatorie discrete

Assumono una quantità finita o numerabile di valori

3.2.1 Funzione di massa di probabilità

Definita su \mathbb{R} e non sul supporto (dominio) della variabile aleatoria D_X

$$\begin{aligned} p_X &: \mathbb{R} \rightarrow [0, 1] \\ p_X(x) &= P(X = x) \end{aligned}$$

In particolare

1. se $x \notin D_x \Rightarrow p(x) = 0$
2. $\forall x \in \mathbb{R} p_X(x) \geq 0$
3. $\lim_{x \rightarrow \infty} F_X(x) = 1$
4. È continua da destra
5. $\sum_x p_X(x) = 1 = P(\underbrace{\bigcup_x \{X = x\}}_{\Omega})$ (sommatoria di eventi disgiunti)

Avendo la funzione di massa di probabilità è possibile ottenere la funzione di ripartizione:

$$\begin{aligned} p_X &\rightarrow F_x \\ F_X(x) &= P(X \leq x) = P(\bigcup_{a \leq x} \{X = a\}) \\ &= \sum_{a \leq x} P(X = a) = \sum_{a \leq x} p_X(a) \end{aligned}$$

e viceversa:

$$\begin{aligned} F_X &\rightarrow p_X \\ p_X(x) &= P(X = x) \\ &= P(x' < X \leq X) = F_X(x') - F_X(x) \end{aligned}$$

3.2.2 Valore atteso variabile aleatoria discreta

$$E[X] := \sum_i x_i p_X(x) = \sum_i x P(X = x) \quad (16)$$

È un **operatore lineare** difatti: $\forall a, b \in \mathbb{R} Y = aX + b \Rightarrow E[Y] = aE[X] + b$ in particolare se $a = 0 \Rightarrow E(b) = b$. Essendo lineare: $E[aX] = aE[X]$.

Il valore atteso di una funzione di variabile aleatoria equivale a:

$$E[g(x)] = \sum_x g(x_i) P(X = x_i)$$

Il valore atteso della somma di variabili aleatorie equivale a:

$$\sum_x \sum_y (x + y) P_{X,Y}(x, y) = E[X] + E[Y]$$

3.3 Varianza

permette di valutare entro quanto le specificazioni di una variabile aleatoria si discostano rispetto al valore atteso.

Sia X una variabile aleatoria, $\mu = E[X]$ allora la varianza:

$$\text{var}(X) = E[(X - \mu)^2] = E[X^2] - E[X]^2 \quad (17)$$

Per ogni coppia di costanti a, b

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

3.4 Deviazione standard

$$\sigma_X = \sqrt{\text{var}(X)} \quad (18)$$

3.5 Variabili aleatorie multivariate

La specificazione ora è bivariata, ovvero è composta da un vettore di due variabili aleatorie discrete X, Y .

3.5.1 Funzione di ripartizione congiunta

$$F_{X,Y} = P(X \leq x \cap Y \leq y) \quad (19)$$

In particolare:

$$\lim_{y \rightarrow +\infty} F_{X,Y} = \lim_{y \rightarrow +\infty} P(X \leq x \cap \underbrace{Y \leq y}_{\substack{\text{Tende all' \\ evento \\ certo}}}) = P(x \leq X) = F_x(x)$$

3.5.2 Funzione di massa di probabilità congiunta

$$p_{X,Y} = P(X = x \cap Y = y) \quad (20)$$

In particolare

$$\begin{aligned} \sum_x p_{X,Y}(x, y) &= \sum_x P(\{X = x\} \cap \{Y = y\}) \\ &= P(\bigcup_x \{X = x\} \cap \{Y = y\}) = \bigcup_x \{X = x\} \cap \{Y = y\} \\ &= P(Y = y) = p_Y(y) = \{y = Y\} \cap \bigcup_x \{x = X\} = \Omega \end{aligned}$$

3.6 Variabili aleatorie indipendenti

Due variabili aleatorie X, Y sono indipendenti $\Leftrightarrow \forall A, B \subseteq \mathbb{R} X \in A$ e $Y \in B$ sono indipendenti. Ovvero se vale che:

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

$$F_{X,Y} = F_X(x)F_Y(y)$$

Con variabili aleatorie discrete l'indipendenza equivale anche a:

$$p_{X,Y} = p_X(x)p_Y(y)$$

Se sono indipendenti: $E(XY) = E(X)E(Y)$

3.7 Covarianza

Si definisce la covarianza tra due variabili aleatorie X, Y come un valore che fornisce una misura di quanto le due varino assieme, ovvero della loro dipendenza. Equivale a 0 con variabili indipendenti

$$\begin{aligned} \text{Cov}(X, Y) &:= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - E[X]E[Y] \end{aligned} \quad (21)$$

Proprietà

- Simmetria: $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = \text{var}(X)$
- $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$
- $\text{Cov}(X + b, Y) = \text{Cov}(X, Y)$ **non vi è linearità**
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$ generalizzando:
$$\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$$
- $\text{Cov}(2X, Y) = 2\text{Cov}(X, Y)$

3.7.1 Varianza somma di variabili aleatorie

$$\begin{aligned} \text{var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \\ \text{var}(X + Y) &= \text{var}(X) + \text{var}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

3.7.2 Coefficiente di correlazione lineare

Misura la forza della relazione tra due variabili aleatorie X, Y .

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (22)$$

3.8 Disuguaglianza di Markov

$$\forall a > 0 \quad P(X \geq a) \leq \frac{E[X]}{a} \quad (23)$$

Si considera $X \geq 0$. Aumentando a la probabilità diminuisce e di conseguenza il risultato tende ad essere sempre più piccolo. Non sempre il rapporto fornisce informazioni sufficientemente esplicative.

3.9 Disuguaglianza di Chebyshev

Sia X una variabile aleatoria con $E[X] = \mu \in \mathbb{R}$ e $\text{var}(X) = \sigma^2 \in \mathbb{R}$ (quindi esistono e sono finite)

$$P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2} \quad \forall r > 0 \quad (24)$$

Posto $r = k\sigma$

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

Il che indica che la probabilità che una variabile aleatoria differisca dal suo valore atteso per più di k volte non può superare il valore $\frac{1}{k^2}$.

4 Modelli di variabili aleatorie discrete

4.1 Modello di Bernoulli

Una variabile aleatoria viene espressa con il modello di Bernoulli ($X \sim B(p)$) se **può esprimere solo due specificazioni**

$$\begin{cases} X = 0 \text{ fallimento} \\ X = 1 \text{ successo} \end{cases}$$

Quindi

$$\begin{cases} P(X = 0) = 1 - p \\ P(X = 1) = p \end{cases}$$

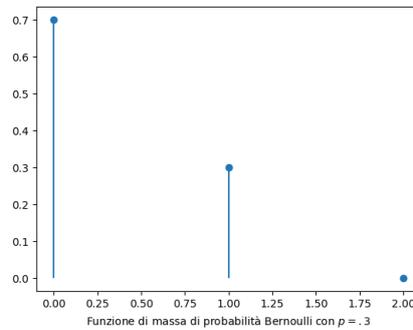
Valore atteso: $E(X) = p$

Varianza: $var(x) = p(1 - p)$

4.1.1 Funzione di massa di probabilità

$$p_X(x) = P(X = x) = p^x(1 - p)^{1-x} I_{\{0,1\}}(x) \quad (25)$$

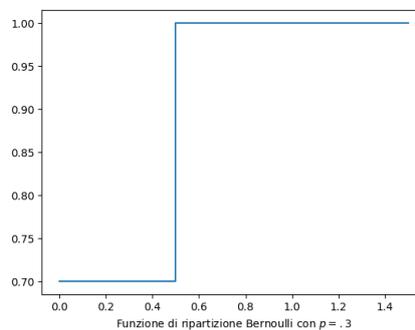
$$= \begin{cases} 1 - p & x = 0 \\ p & x = 1 \\ 0 & \text{altrimenti} \end{cases}$$



4.1.2 Funzione di ripartizione

$$F_X(x) = (1 - p)I_{[0,1)}(x) + I_{[1,+\infty)}(x) \quad (26)$$

$$= \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$



4.2 Modello binomiale

Vengono realizzate n ripetizioni **indipendenti** di un esperimento bernulliano. Si denota con X il numero totale di successi, allora $X \sim B(n, p)$ con $n \in \mathbb{N}^+$ e $p \in [0, 1]$. Di conseguenza $D_X = \{0, 1, \dots, n\}$

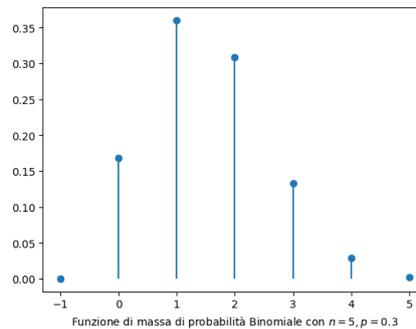
Notare che sommando variabili aleatorie bernulliane si ottengono variabili aleatorie binomiali

Valore atteso: $E[X] = \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i} = np$

Varianza: $var(X) = \sum_{i=1}^n p(1-p) = np(1-p)$

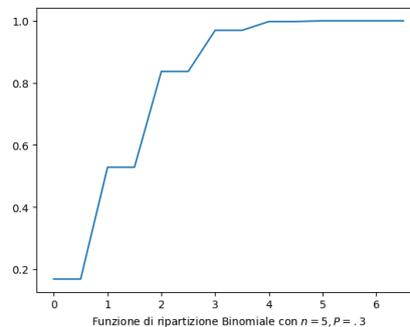
4.2.1 Funzione di massa di probabilità

$$p_X(i) = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} I_{\{1, \dots, n\}}(i) \quad (27)$$



4.2.2 Funzione di ripartizione

$$F_X(x) = P(X \leq x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} I_{[0, n]}(x) + I_{(n, +\infty)}(x) \quad (28)$$



4.3 Modello uniforme discreto

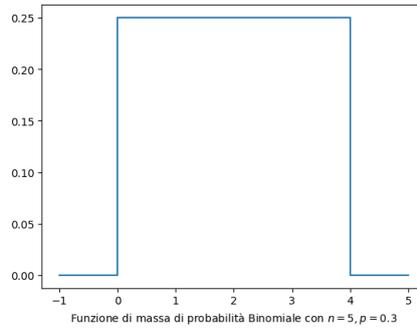
Le variabili aleatorie consistono nell'esito di un esperimento da n possibili esiti **equiprobabili**. Quindi $X \sim U(n)$ con $D_X = \{1, \dots, n\}$

Valore atteso: $E[X] = \frac{n+1}{2}$

Varianza: $var(X) = \frac{n^2-1}{12}$

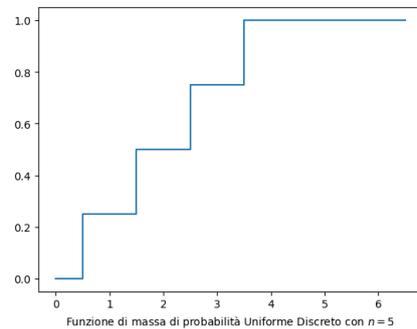
4.3.1 Funzione di massa di probabilità

$$p_X(x) = P(X = x) = \frac{1}{n} I_{\{1, \dots, n\}}(x) \quad (29)$$



4.3.2 Funzione di ripartizione

$$F_X = P(X \leq x) = \sum_{i=0}^{\lfloor x \rfloor} \frac{1}{n} = \frac{\lfloor x \rfloor}{n} \Rightarrow F_X(x) = \frac{\lfloor x \rfloor}{n} I_{[0, n)}(x) + I_{(n, +\infty)}(x) \quad (30)$$



4.4 Modello geometrico

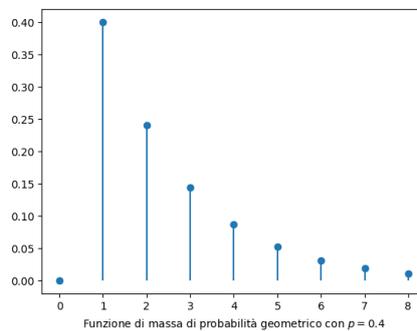
Si basa sull'esecuzione ripetuta di esperimenti bernoulliani di parametro $p \in (0, 1]$ e descrive il **numero di insuccessi necessari affinché si verifichi il primo successo**. Quindi $X \sim G(p)$ con $D_X = \mathbb{N} \cup \{0\}$

Valore atteso: $E[X] = \frac{1-p}{p}$

Varianza: $var(x) = \frac{1-p}{p^2}$

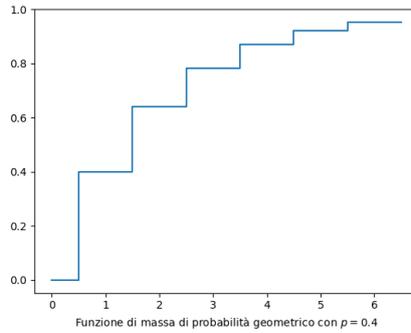
4.4.1 Funzione di massa di probabilità

$$p_X(x) = P(X = x) = (1-p)^x p I_{\mathbb{N} \cup \{0\}}(x) \quad (31)$$



4.4.2 Funzione di ripartizione

$$F_X(n; p) = (1 - (1-p)^{\lfloor x \rfloor + 1}) I_{\mathbb{R}^+}(x) \quad (32)$$



4.4.3 Assenza di memoria

Si nota che: $P(X \geq x) = P(X \geq \lfloor x \rfloor) = P(X > \lfloor x \rfloor - 1) = (1 - p)^{\lfloor x \rfloor}$ e quindi che:

$$\begin{aligned}
 P(X \geq x + y | X \geq x) &= \frac{P(X \geq x + y, X \geq x)}{P(X \geq x)} \\
 &= \frac{P(X \geq x + y)}{P(X \geq x)} \\
 &= \frac{(1 - p)^{\lfloor x \rfloor + \lfloor y \rfloor}}{(1 - p)^{\lfloor x \rfloor}} \\
 &= (1 - p)^{\lfloor y \rfloor} = P(X \geq y)
 \end{aligned}$$

4.5 Modello di Poisson

Adoperato per rappresentare situazioni di conteggio del numero di occorrenze di certi eventi in una unità di tempo o più precisamente numero di “successi” in un certo intervallo.

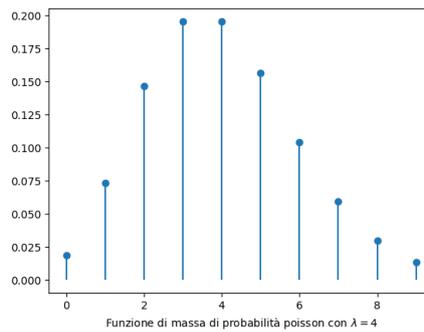
Una variabile descritta secondo tale modello $X \sim P(\lambda)$ con $D_X = \mathbb{N} \cup \{0\}$ e $\lambda > 0$

Valore atteso: $E[X] = \lambda$

Varianza: $var(X) = \lambda$

4.5.1 Funzione di massa di probabilità

$$p_X(i) = P(x = i) = e^{-\lambda} \frac{\lambda^i}{i!} I_{\mathbb{N} \cup \{0\}}(i) \quad (33)$$



È possibile approssimare una variabile aleatoria di Poisson con una binomiale di parametri (n, p) nel caso in cui n cresca e p diminuisca. Infatti:

$$\begin{aligned} P(X = i) &= \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\ &= \frac{n(n-1)\dots(n-i+1)}{i!} \frac{\lambda^i}{n^i} \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n}{n} \frac{n-1}{n} \dots \frac{(n-i+1)}{n} \frac{\lambda^i}{i!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^i} \text{ per } n \rightarrow +\infty \\ &= \frac{\lambda^i}{i!} e^{-\lambda} \end{aligned}$$

Inoltre il modello di Poisson gode della proprietà di **riproducibilità**: $X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2) \Rightarrow X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$

4.5.2 Processo di Poisson

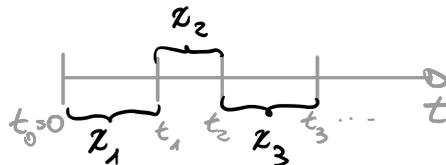
Sia t il tempo con inizio in t_0 . Si definisce $N(t)$ come il numero di eventi che accadono in $(0, t]$ È un processo stocastico che deve soddisfare:

1. $N(0) = 0$
2. Intervalli disgiunti \Rightarrow indipendenza
3. Il conteggio del numero di eventi dipende solo dalla lunghezza dell'intervallo e non dagli estremi.
4. $\lim_{h \rightarrow +\infty} \frac{P(N(h) = 1)}{h} = \lambda = P(N(h) = 1) \approx \lambda h$
5. $\lim_{h \rightarrow +\infty} \frac{P(N(h) \geq 2)}{h} = 0 = P(N(h) \geq 2) \approx 0$ si considera quindi 1 avvenimento o nessun avvenimento.

Si nota che $N(t) \sim P(\lambda t)$ ma, sia k il numero di avvenimenti e n il numero di sottointervalli con $n \geq k$:

$$P(N(t) = k) = \binom{n}{k} \left(\lambda \frac{t}{n}\right)^k \left(1 - \lambda \frac{t}{n}\right)^{n-k}$$

Quindi $N(t) \sim B\left(n, \lambda \frac{t}{n}\right)$ e con $n \rightarrow +\infty$ $N(t) \sim P(\lambda t)$ Se si considerano gli intertempi:



$$\begin{aligned} P(X_1 > t) &= P(N(t) = 0) = e^{-\lambda t} \text{ quindi} \\ F_{X_1} &= P(X_1 \leq t) = 1 - P(X_1 > t) = 1 - e^{-\lambda t} \\ &\Rightarrow X_1 \sim E(\lambda) \end{aligned}$$

Se si considerano gli intertempi:



$$\begin{aligned} P(X_2 > t | X_1 = s) &= P(\text{nessun accadimento tra } s \text{ e } s+t | X_1 = s) \\ &= P(\text{nessun accadimento tra } s \text{ e } s+t) = P(N(t) = 0) = e^{-\lambda t} \\ &\Rightarrow X_2 \sim E(\lambda) \end{aligned}$$

Sono inoltre indipendenti tra di loro in quanto non dipendono da s .

4.6 Modello ipergeometrico

Descrive il problema dell'urna: $X \sim H(n, N, M)$ Con:

- N numero di oggetti funzionanti
- M numero di oggetti difettosi
- n numero di estrazioni (elementi necessari)

X rappresenta il numero di oggetti funzionanti dopo n estrazioni senza reimmissioni da un'urna che contiene N oggetti funzionanti e M oggetti difettosi.

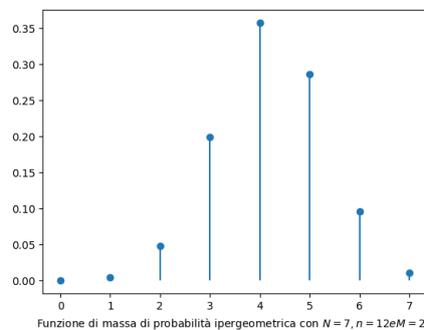
Sia $p = \frac{N}{N+M}$

Valore atteso: $E(X) = np$

Varianza: $var(X) = \underbrace{np(1-p)}_{var \text{ mod binomiale}} \left(1 - \frac{n+1}{N+M-1}\right)$

4.6.1 Funzione di massa di probabilità

$$P(X = i) = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}} I_{\{0, \dots, n\}}(i) \quad (34)$$



Quindi $X_i = \begin{cases} 1 & \text{l'i-esima estrazione ha portato ad un oggetto funzionante} \\ 0 & \text{altrimenti} \end{cases}$

possibile semplificare il modello in un modello binomiale.

Se $N + M$ cresce, è

5 Variabili aleatorie continue

Può assumere una infinità non numerabile di valori. È continua se esiste una funzione non negativa definita su tutto \mathbb{R} avente la proprietà che per ogni insieme B di numeri reali:

$$P(X \in B) = \int_B f_X(x) dx$$

La funzione prende il nome di **funzione di densità di probabilità** e deve soddisfare:

$$1 = P(X \in \mathbb{R}) = \int_{\mathbb{R}} f_X(x) dx$$

in particolare: $P(a \leq X \leq b) = \int_a^b f(x) dx$ se $b = a \Rightarrow P(X = a) = \int_a^a f(x) dx = 0$

5.1 Funzione di ripartizione

$$F(a) := P(X \in (-\infty, a]) = \int_{-\infty}^a f(x) dx$$

Si nota che la funzione di ripartizione è una primitiva della funzione di densità. Quindi:

$$\frac{dF_X(x)}{dx} = f_X(x)$$

Valore atteso: $E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} 1 - F_X(x) dx$

Varianza: $var(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx$

6 Modelli di variabili aleatorie continue

6.1 Modello uniforme continuo

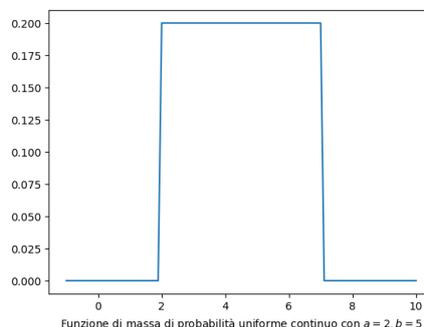
Estende al continuo i concetti per il modello uniforme discreto e viene determinato da un intervallo equivalentemente aperto o chiuso: $X \sim U([a, b])$ con $a < b \Rightarrow D_x = [a, b]$

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

Varianza: $var(x) = \frac{(b-a)^2}{12}$

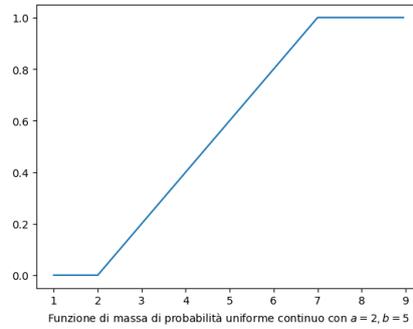
6.1.1 Funzione di densità di probabilità

$$f_X(i) = \frac{1}{b-a} I_{[a,b]}(i) \quad (35)$$



6.1.2 Funzione di ripartizione

$$F_X(x) = P(X \leq x) = \frac{x-a}{b-a} I_{[a,b]}(x) + I_{(b,+\infty)}(x) \quad (36)$$



Le probabilità che $\alpha \leq X \leq \beta$ sono:

$$P(\alpha < X < \beta) = \frac{\beta - \alpha}{b - a}$$

6.2 Modello esponenziale

È un modello usato per misurare il tempo che intercorre tra due eventi. Effettuarne la scalatura, equivale a scalare il suo parametro.

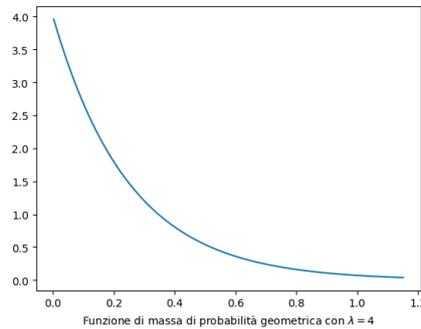
$X \sim E(\lambda)$ con $\lambda > 0$. $D_X = \mathbb{R}^+$

Valore atteso: $E[X] = \frac{1}{\lambda}$

Varianza: $var(X) = \frac{1}{\lambda^2}$

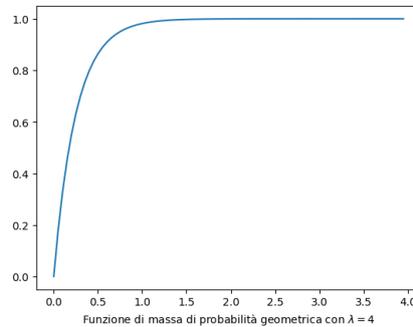
6.2.1 Funzione di densità di probabilità

$$f_X(x) = \lambda e^{-\lambda x} I_{\mathbb{R}^+}(x) \quad (37)$$



6.2.2 Funzione di ripartizione

$$F_X(x) = (1 - e^{-\lambda x}) I_{\mathbb{R}^+}(x) \quad (38)$$



6.2.3 Assenza di memoria

Anche le variabili di modello esponenziale godono della proprietà di assenza di memoria. Ovvero: $P(X > s + t | X > s) = P(X > t)$ dato che:

$$\begin{aligned} P(X > s + t) &= e^{-\lambda(s+t)} \\ &= e^{-\lambda s} e^{-\lambda t} \\ &= P(X > s) P(X > t) \end{aligned}$$

6.3 Modello gaussiano / normale

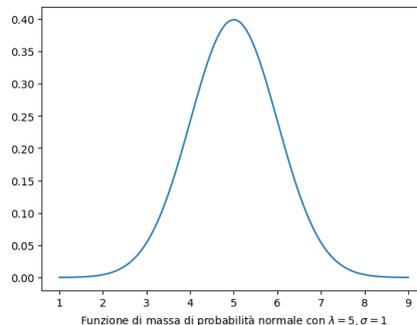
La variabile aleatoria continua è definita da due parametri: $X \sim N(\mu, \sigma)$ con $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$.
 $D_X = \{\mathbb{R}\}$

Valore atteso: $E(X) = \mu$

Varianza: $var(X) = \sigma^2$

6.3.1 Funzione di densità di probabilità

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (39)$$



6.3.2 Funzione di ripartizione

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma}\right)^2\right) dy \quad (40)$$

Risulta non avere una forma analitica. Si calcola una approssimazione del valore. Sia $Y := aX + b$ con $a, b \in \mathbb{R}$, se $X \sim N(\mu, \sigma)$ allora $Y \sim N(a\mu + b, |a|\sigma)$

Gode di riproducibilità: Sia $X_i \sim G(\mu_i, \sigma_i)$ indipendenti, allora $Y := \sum_{i=1}^n X_i \Rightarrow Y \sim N(\sum_i \mu_i, \sqrt{\sum_i \sigma_i^2})$

6.4 Distribuzione normale standard

Partendo da una variabile gaussiana X , si può costruire la variabile Z come standardizzazione di X

$$Z = \frac{x - \mu}{\sigma} = \frac{1}{\sigma}x - \frac{\mu}{\sigma} \sim N(0, 1)$$

La standardizzazione comporta che:

Valore atteso: $E(Z) = 0$

Varianza: $Var(Z) = 1$

6.4.1 Funzione di ripartizione

$$F_X(x) = P(X \leq x) = P\left(\frac{x - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) = F_Z\left(\frac{x - \mu}{\sigma}\right) \quad (41)$$

Si osserva la linearità di una distribuzione normale:

$X \sim N(\mu, \sigma), E[X] = \mu, var(X) = \sigma^2$ $Y = aX + b, a, b \in \mathbb{R} E[Y] = a\mu + b var(X) = a^2\sigma^2$

Nota che $(\Phi(-x) = 1 - \Phi(x))$

$$P(\alpha \leq X \leq \beta) = \Phi\left(\frac{\beta - \mu}{\sigma}\right) - \Phi\left(\frac{\alpha - \mu}{\sigma}\right)$$

7 Teorema di De Moivre-Laplace

Fornisce la possibilità di approssimare una distribuzione binomiale con una gaussiana normale. Siano $X_1, \dots, X_n \sim B(n, p)$ allora:

$$Y := \sum_{i=1}^n X_i \sim B(n, p) \underset{\text{approx}}{\overset{\sim}{\simeq}} N(np, \sqrt{np(1-p)})$$

8 Teorema centrale del limite

Siano X_1, \dots, X_n variabili aleatorie **indipendenti e identicamente distribuite** con $E(X) = \mu$ e varianza $var(X) = \sigma^2$ allora per n grandi (> 30) le variabili sono distribuite in modo approssimativamente normale.

$$\sum_{i=1}^n X_i \overset{\sim}{\simeq} N(n\mu, \sigma\sqrt{n}) \quad (42)$$

9 Statistica inferenziale parametrica puntuale

Popolazione descritta come una variabile aleatoria X . La quale è descritta secondo una certa distribuzione F di cui si conosce tutto a meno dei parametri θ . Quindi $X \sim F(\theta)$.

La stima viene fatta su qualcosa che dipende da θ ovvero $\tau(\theta)$. L'operazione prende il nome di **statistica** \ **stimatore** e non è necessariamente unico. .

Ovvero una funzione $t: D_X^n \rightarrow \mathbb{R}$ quindi $T_n = t(x_1, \dots, x_n) \approx \tau(\theta)$.

Si definisce **campione** un insieme di variabili aleatorie indipendenti ed identicamente distribuite.

9.1 Bias

Si definisce bias \ distorsione di una statistica rispetto ad una quantità ignota:

$$b_{\tau(\theta)}(T) = E(T) - \tau(\theta) \quad (43)$$

9.2 Scarto quadratico medio

Definito come il valore medio dell'errore quadratico quando si usa una certa statistica per stimare un parametro. **Più è vicino a 0 e più è stimato correttamente**

$$MSE_{\tau(\theta)}(T) = E((T - \tau(\theta))^2) = var(T) + b_{\tau(\theta)}(T)^2 \quad (44)$$

Se lo stimatore è non deviato: $MSE_{\tau(\theta)}(T) = var(T)$

9.2.1 Consistenza in media quadratica

Se

$$\forall \theta \lim_{n \rightarrow \infty} MSE_{\tau(\theta)}(T) = 0 \quad (45)$$

9.2.2 Stimatore debolmente consistente

Se

$$\lim_{n \rightarrow +\infty} P(|T_n - \tau(\theta)| < \epsilon) = 1 \quad (46)$$

Fissato ϵ si cerca la probabilità che l'errore si discosti da ϵ . Più P è grande meglio è.

Se è consistente in media quadratica \Rightarrow è debolmente consistente

9.3 Stimatore non deviato

Una statistica è **non deviata** rispetto a $\tau(\theta)$ sse

$$E(t(X_1, \dots, X_n)) = \tau(\theta) \quad (47)$$

Se verificata i valori delle stime tenderanno ad oscillare intorno al valore sconosciuto.

9.3.1 Valore atteso

Fissato $\tau(\theta) = E(X)$. Uno stimatore non deviato è la **media campionaria**

$$t(X_1, \dots, X_n) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (48)$$

Difatti

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n E(X) = E(X)$$

9.3.2 Varianza

Uno stimatore non deviato è:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (49)$$

9.4 Legge dei grandi numeri

9.4.1 Versione forte

$$P\left(\lim_{n \rightarrow +\infty} \bar{X}_n = E(X)\right) = 1 \quad (50)$$

9.4.2 Versione debole

$$\forall \epsilon \lim_{x \rightarrow +\infty} P(|\bar{X}_n - E(X)| > \epsilon) = 0 \quad (51)$$

9.5 Dimensione del campione

Sia X una variabile aleatoria con $E(X) = \mu$ e $var(X) = \sigma^2$, con r fissato come soglia. Si cerca $\tau(\theta) = \mu$

$$P(|\bar{X} - \mu| \leq r) \geq 1 - \delta = 2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 \quad (52)$$

Difatti:

$$\begin{aligned} P(-r \leq \bar{X} - \mu \leq r) &= P\left(-\frac{r}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{r}{\sigma/\sqrt{n}}\right) \\ &\approx P\left(-\frac{r}{\sigma}\sqrt{n} \leq Z \leq \frac{r}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{r}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 + \Phi\left(\frac{r}{\sigma}\sqrt{n}\right) \\ &= 2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 \end{aligned}$$

Quindi

$$\begin{aligned} 2\Phi\left(\frac{r}{\sigma}\sqrt{n}\right) - 1 &\geq 1 - \delta \\ n &\geq \left[\left(\frac{\hat{\sigma}}{r}\Phi^{-1}\left(1 - \frac{\delta}{2}\right)\right)^2\right] \end{aligned} \quad (53)$$

σ va stimata a partire dai dati. Se σ cresce il valore minimo di n aumenta. Se r aumenta saranno sufficienti meno n e se δ è piccolo più si è stringenti quindi Φ^{-1} tende a $+\infty$. Analogamente:

$$\delta \geq 2\left(1 - \Phi\left(\frac{r}{\hat{\sigma}\sqrt{n}}\right)\right)$$

$$r \geq \frac{\hat{\sigma}}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\delta}{2}\right)$$

Si nota che: $TCL \leq TCH$. Difatti

$$\frac{\sigma^2}{r^2} \Phi^{-1} \left(1 - \frac{\delta}{2} \right)^2 \leq \frac{\sigma^2}{\delta r^2}$$

$$\Phi^{-1} \left(1 - \frac{\delta}{2} \right) \leq \sqrt{\frac{1}{\delta}}$$

Che risulta essere sempre vera.

9.6 Stimatori di massima verosimiglianza

Sia X una popolazione $\sim F(\theta)$ con X_1, \dots, X_n i.i.d. Si definisce la verosimiglianza come:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n f_X(x_i) \quad (54)$$

Si cerca quindi un $\hat{\theta}$ tale che:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(x_1, \dots, x_n; \theta) \quad (55)$$

Il valore massimo viene trovato ponendo la derivata uguale a 0. Conviene però passare prima per il $\log \mathcal{L}$ e poi calcolare la derivata di $\log \mathcal{L}$. (Possibile in quando è monotono crescente ed assumono quindi massimo nello stesso valore di θ).

Se $X \sim B(p)$ o distribuita secondo una Poisson:

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; p) &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_i x_i} (1-p)^{\sum_i (1-x_i)} \\ \ln \mathcal{L}(x_1, \dots, x_n; p) &= \sum_i x_i \ln p + \sum_i (1-x_i) \ln(1-p) \\ \frac{d}{dp} \ln \mathcal{L}(x_1, \dots, x_n; p) &= \frac{\sum_i x_i}{p} - \frac{\sum_i (1-x_i)}{1-p} = 0 \\ \hat{p} &= \frac{1}{n} \sum_i x_i \end{aligned}$$

Se $X \sim N(\mu, \sigma)$ entrambi ignoti:

$$\begin{aligned} \mathcal{L}(\mu, \sigma) &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \left(\frac{1}{\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2} \\ \ln \mathcal{L}(\mu, \sigma) &= n \ln \frac{1}{\sqrt{2\pi}} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2 \\ \frac{\partial}{\partial \mu} \ln \mathcal{L}(\mu, \sigma) &= -\frac{1}{\sigma^2} \sum_i (x_i - \mu) \\ \frac{\partial}{\partial \sigma} \ln \mathcal{L}(\mu, \sigma) &= -\frac{n}{\sigma} + \sigma^{-3} \sum_i (x_i - \mu)^2 \\ &\begin{cases} \frac{1}{\sigma^2} \sum_i (x_i - \hat{\mu}) = 0 \\ -\frac{n}{\hat{\sigma}} + \hat{\sigma}^{-3} \sum_i (x_i - \hat{\mu})^2 = 0 \end{cases} \\ &\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \hat{\mu})^2 \end{cases} \end{aligned}$$