STATISTICA E ANALISI DEI DATI

Capitolo 10 – Stima Puntuale

Dott. Stefano Cirillo
Dott. Luigi Di Biasi

a.a. 2025-2026

Campioni Casuali e Stimatori

- Per ottenere informazioni sui parametri non noti della popolazione:
  - Si può partire da un campione estratto dalla popolazione
  - E si può fare uso dell'inferenza statistica per effettuare delle misure su tale campione
    - il campione deve essere **rappresentativo** della popolazione
- Con «**Osservabile**» indichiamo che:
  - si possono osservare **i valori assunti** dalla variabile aleatoria X (ad esempio, eseguendo un esperimento casuale)
  - e quindi il parametro non noto è presente soltanto nella legge di probabilità (funzione di distribuzione, funzione di probabilità, densità di probabilità)
- Nota: Ovviamente se i parametri sono noti la **legge di probabilità è completamente specificata** e riconducibile ad almeno una di quelle che abbiamo studiato

Stima Statistica

- Uno dei **problemi** centrali dell'inferenza statistica è il seguente:
  - si desidera studiare una popolazione descritta da una variabile aleatoria osservabile X la cui **funzione di distribuzione ha una forma nota ma contiene un parametro $\vartheta \in \theta$** non noto (o più parametri non noti)
- La **stima** è il processo mediante il quale si cerca di determinare il valore di un parametro incognito di una popolazione basandosi sui dati di un campione
- La stima può essere:
  - **Puntuale**: fornisce un unico valore come approssimazione del parametro incognito
    - Esempio: la **media campionaria** come stima della media μ\muμ della popolazione
  - **Intervallare**: fornisce un intervallo di valori entro cui il parametro si trova con un certo livello di confidenza
    - Esempio: **intervallo di confidenza** per la media.

Statistica e Stimatori

- Nei metodi di indagine dell'inferenza statistica si parte dal campione $X_1, X_2, \dots, X_n$
  - si cerca di ottenere informazioni sui parametri non noti facendo uso di alcune variabili aleatorie, che sono funzioni misurabili del campione casuale, dette **statistiche** e **stimatori**
- **Statistica**:
  - Una statistica è una funzione matematica degli elementi di un campione
    - Ad esempio, la media campionaria e la varianza campionaria sono statistiche
  - Le statistiche sono utilizzate per riassumere e descrivere le caratteristiche del campione
- **Stimatore**:
  - Un stimatore è una regola o una formula matematica che determina come calcolare una stima del parametro di interesse basandosi sui dati del campione
- In altri termini è una statistica usata per stimare il valore di uno o più parametri
- Ad esempio, $\bar{X}_n$ (media campionaria) è uno stimatore di una variabile aleatoria normale di valore atteso μ

Statistica e Stimatori

- Nel contesto della stima puntuale, l'obiettivo è ottenere un valore specifico che rappresenti il "miglior indizio" possibile del parametro di popolazione
- Un buon stimatore dovrebbe essere **non distorto** (cioè avere un valore atteso uguale al parametro che si sta cercando di stimare) e avere una bassa variabilità
- **Statistica**:
  - Una statistica $t(X_1, X_2, \dots, X_n)$ è una **funzione misurabile** e osservabile del campione $X_1, X_2, \dots, X_n$
  - I valori da essa assunti **dipendono soltanto dal campione osservato** $x_1, x_2, \dots, x_n$ estratto dalla popolazione
  - I parametri non noti sono presenti soltanto nella funzione di distribuzione della statistica
- **Stimatore** è una statistica tale che:
  - Uno stimatore $\hat{\Theta} = t(X_1, X_2, \dots, X_n)$ è una **funzione misurabile** e osservabile del campione $X_1, X_2, \dots, X_n$ i cui valori possono essere usati per stimare un parametro non noto ϑ della popolazione
  - I valori $\hat{\vartheta}$ assunti da tale stimatore sono detti stime del parametro non noto ϑ
  - Lo stimatore è una **variabile aleatoria**

Campioni casuali e stimatori

- **Definizione**: Sia $X_1, X_2, \dots, X_n$ un campione casuale

Media Campionaria: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

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

Campioni casuali e stimatori

- **Definizione**: Sia $X_1, X_2, \ldots, X_n$ un campione casuale, due **statistiche** sono:

Media Campionaria: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **Varianza Campionaria:** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

- **Proposizione**: Sia $X_1, X_2, \ldots, X_n$ un campione casuale estratto da una popolazione descritta da una variabile aleatoria osservabile X con valore medio $E(X) = \mu$ e varianza $Var(X) = \sigma^2$ si ha che:

$$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

- Dimostrazione: Per la proprietà di linearità del valore medio e l'identica distribuzione delle variabili aleatorie che costituiscono il campione si ha:

$$E(\overline{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

- Inoltre, poiché le variabili aleatorie che costituiscono il campione sono indipendenti ed identicamente distribuite si ha:

$$\mathrm{Var}\left(\overline{X}\right) = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) = \frac{\sigma^2}{n}.$$

Campioni casuali e stimatori

- **Definizione**: Sia $X_1, X_2, \dots, X_n$ un campione casuale, due **statistiche** sono:

Media Campionaria: $\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$ **Varianza Campionaria:** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

- **Proposizione**: Sia $X_1, X_2, \dots, X_n$ un campione casuale estratto da una popolazione descritta da una variabile aleatoria osservabile X con valore medio $E(X) = \mu$ e varianza $Var(X) = \sigma^2$ si ha che:

$$E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

- Dimostrazione: Per la proprietà di linearità del valore medio e l'identica distribuzione delle variabili aleatorie che costituiscono il campione si ha:

$$E(\overline{X}) = E\left[\boxed{\frac{1}{n}\sum_{i=1}^{n} X_i}\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

- Inoltre, poiché le variabili aleatorie che costituiscono il campione sono indipendenti ed identicamente distribuite si ha:

$$\mathrm{Var}\left(\overline{X}\right) = \mathrm{Var}\left[\boxed{\frac{1}{n}\sum_{i=1}^{n} X_i}\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) = \frac{\sigma^2}{n}.$$

Campioni casuali e stimatori

- **Proposizione**: Sia $X_1, X_2, \dots, X_n$ un campione casuale estratto da una popolazione descritta da una variabile aleatoria osservabile X con valore medio $E(X) = \mu$ e varianza $Var(X) = \sigma^2$ si ha che:

$$E(\bar{X}) = \mu \qquad\qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

- La proposizione ci dice inoltre che:
  - al crescere dell'ampiezza del campione la media campionaria fornisce una **stima** sempre più accurata del valore medio della popolazione
  - Dal teorema centrale di convergenza scaturisce che per n sufficientemente grande (ossia per campioni di grande ampiezza) la funzione di distribuzione della media campionaria $\bar{X}$ è approssimativamente una **variabile aleatoria normale** con valore medio μ e varianza $\frac{\sigma^2}{n}$

Stimatore e Stima dei parametri

- Trovare uno stimatore significa costruire una funzione su delle osservazioni campionarie che fornisca una stima del parametro incognito della popolazione
- Come trovare uno stimatore?
  - I principali metodi sono:
    - Il metodo dei momenti
    - Il metodo della massima verosimiglianza

STATISTICA E ANALISI DEI DATI

Metodo dei Momenti

Momenti

- In statistica, un **momento** è una quantità che descrive alcune caratteristiche fondamentali di una distribuzione di probabilità
- **<u>Momenti rispetto all'origine</u>:**
  - Il k-esimo momento di una variabile casuale X rispetto all'origine è: $\mu'_k = \mathrm{E}[X^k]$
    - $\mu'_1 = \mathrm{E}[X]$ →**la media**
    - $\mu'_2 = \mathrm{E}[X^2]$ →legato alla **dispersione**
    - Etc.
- **<u>Momenti centrati (rispetto alla media)</u>:**
  - Il k-esimo momento centrato si definisce come: $\mu_k = \mathrm{E}\left[\left(X - \mathrm{E}[X]\right)^k\right]$
    - $\mu_2 = Var(X)$ →**varianza**
    - μ_3 →**asimmetria**
    - μ_4 →**curtosi**
- I momenti centrati descrivono quindi forma, asimmetria e "coda" della distribuzione

Momenti

- Nel campione, sostituiamo il valore atteso con la media empirica, cioè:
  - Se vogliamo stimare:

$$E[X] = \textit{media della popolazione}$$

  - ma abbiamo solo i dati $X_1, \dots, X_n$,allora prendiamo: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ che è la media empirica.
- Il momento campionario di ordine k è: $M_k' = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
- I momenti campionari sono le quantità che il **metodo dei momenti** utilizza per stimare i parametri di una distribuzione
- **Esempio**: Supponiamo di avere un campione di dati: $X = \{2,3,5,7\}$ con $n = 4$ osservazioni
- **Momento campionario di ordine 1**

$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{2+3+5+7}{4} = \frac{17}{4} = 4.25$$

- **Momento campionario di ordine 2**

$$M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{2^2+3^2+5^2+7^2}{4} = \frac{4+9+25+49}{4} = \frac{87}{4} = 21.75$$

  - Questo momento è spesso usato per ricavare la varianza tramite:

$$Var(X) = M_2' - (M_1')^2$$

Metodo dei Momenti

- Sia $X_1, X_2, \dots, X_n$ un campione casuale estratto da una popolazione con funzione di probabilità (var. discrete) o densità di probabilità (var. continue) $f(X: \vartheta_1, \vartheta_2, \dots, \vartheta_k)$ dove $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ sono i parametri ignoti della popolazione
- I momenti campionari assoluti, cioè i valori '<u>ottimali</u>' (μ'_r) che vogliamo provare a stimare:

$$\mu'_r = E(X^r) = \mu_1^r(\vartheta_1, \vartheta_2, \dots, \vartheta_k)$$

Metodo dei Momenti

- Sia $X_1, X_2, \dots, X_n$ un campione casuale estratto da una popolazione con funzione di probabilità (var. discrete) o densità di probabilità (var. continue) $f(X: \vartheta_1, \vartheta_2, \dots, \vartheta_{\boxed{k}})$ dove $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ sono i parametri ignoti della popolazione

Nota: k e non n perché X potrebbe avere più parametri incogniti

- I momenti campionari assoluti, cioè i valori 'ottimali' (μ_r') che vogliamo provare a stimare:

Momento campionario r-esimo ASSOLUTO ←

$$\boxed{\mu_r' = E(X^r) = \mu_1^r(\vartheta_1, \vartheta_2, \dots, \vartheta_k)}$$

→ Cioè calcoliamo il valore atteso di una distribuzione dove ci sono $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ parametri incogniti.
Il risultato è un valore che contiene i parametri incogniti

Metodo dei Momenti

- Sia $X_1, X_2, \dots, X_n$ un campione casuale estratto da una popolazione con funzione di probabilità (var. discrete) o densità di probabilità (var. continue) $f(x: \vartheta_1, \vartheta_2, \dots, \vartheta_k)$ dove $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ sono i parametri ignoti della popolazione
- I momenti campionari assoluti, cioè i valori ‘ottimali’ (μ'_r) che vogliamo provare a stimare:

Momento r-esimo ASSOLUTO ⟵ $\boxed{\mu'_r = E(X^r) = \mu_1^r(\vartheta_1, \vartheta_2, \dots, \vartheta_k)}$

- L’idea è alla base del metodo dei momenti è di porre il momento campionario assoluto **uguale** al corrispondente momento campionario
- Definiamo quindi il momento campionario come:

Momento campionario r-esimo RELATIVO ⟵ $$M'_r = \frac{1}{n}\sum_{k=1}^{n} X_k^r \qquad (r = 1, \dots, k)$$

  - e poniamo:

$$\mu'_r = M'_r \qquad (r = 1, \dots, k)$$

Metodo dei Momenti

- Sia $X_1, X_2, \ldots, X_n$ un campione casuale estratto da una popolazione con funzione di probabilità (var. discrete) o densità di probabilità (var. continue) $f(x: \vartheta_1, \vartheta_2, \ldots, \vartheta_k)$ dove $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ sono i parametri ignoti della popolazione
- I momenti campionari assoluti, cioè i valori ‘ottimali’ che vogliamo provare a stimare:

$$\mu_r' = E(X^r) = \mu_1^r(\vartheta_1, \vartheta_2, \ldots, \vartheta_k)$$

- L’idea è alla base del metodo dei momenti è di porre il momento campionario assoluto **uguale** al corrispondente momento campionario
- Definiamo quindi:

$$M_r' = \frac{1}{n}\sum_{k=1}^{n} X_k^r \qquad (r = 1, \ldots, k)$$

Sono statistiche perche non contengono i parametri incogniti $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$

- e poniamo:

Contengono i parametri incogniti $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$

$$\boxed{\mu_r'} = \boxed{M_r'} \qquad (r = 1, \ldots, k)$$

Metodo dei Momenti

- Partendo da: $\mu'_r = M'_r$ con $(r = 1, \dots, k)$, definiamo un sistema:

$$\begin{cases} \mu'_1 = M'_1 \\ \dots \\ \mu'_k = M'_k \end{cases}$$

- Se tale sistema è risolvibile in $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ avremo delle funzioni:

$\hat{\vartheta}_1, \hat{\vartheta}_2, \dots, \hat{\vartheta}_k$ funzioni di $X_1, X_2, \dots, X_n$

Metodo dei Momenti

- Partendo da: $\mu'_r = M'_r$ con $(r = 1, \dots, k)$, definiamo un sistema:

$$\boxed{\begin{cases} \mu'_1 = M'_1 \\ \dots \\ \mu'_k = M'_k \end{cases}} \longrightarrow \text{Un sistema di } r \text{ equazioni e } k \text{ incognite}$$

- Se tale sistema è risolvibile in $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ avremo delle funzioni:

Stimatori Cercati dei parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ $\longleftarrow$ $\boxed{\hat{\vartheta}_1, \hat{\vartheta}_2, \dots, \hat{\vartheta}_k}$ funzioni di $X_1, X_2, \dots, X_n$

- Nota: Quello che abbiamo visto è molto generale, ma nella pratica si considerano generalmente al massimo due stimatori $(k = 2)$

Metodo dei Momenti (Esempi)

- **Esempio 1** Caso 1 parametro $(k = 1)$:
  - Consideriamo una variabile aleatoria X
    - μ è il valore che vogliamo stimare, quindi $\mu = \vartheta_1$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$

Metodo dei Momenti (Esempi)

- **Esempio 1** Caso 1 parametro $(k = 1)$:
  - Consideriamo una variabile aleatoria X
    - μ è il valore che vogliamo stimare, quindi $\mu = \vartheta_1$
  - Calcoliamo il momento campionario assoluto:

$$\boxed{\mu'_1 = E(X^1) = \mu} \longrightarrow$$

Momento r-esimo ASSOLUTO

  - Calcoliamo M'_1:

$$M'_1 = \frac{1}{n}\sum_{k=1}^{n} X_k^1 = \bar{X}_n \longrightarrow$$

Momento campionario r-esimo RELATIVO

Metodo dei Momenti (Esempi)

- **Esempio 1** Caso 1 parametro $(k = 1)$:
  - Consideriamo una variabile aleatoria X
    - μ è il valore che vogliamo stimare, quindi $\mu = \vartheta_1$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$

  - Calcoliamo M_1':

$$M_1' = \frac{1}{n}\sum_{k=1}^{n} X_k^1 = \bar{X}_n$$

  - Quindi si ha che (il sistema ci dice che)

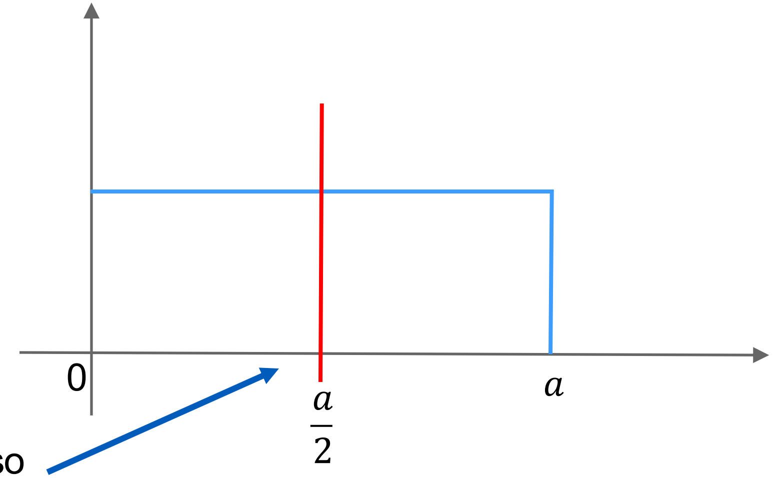
$$\hat{\mu} = M_1' = \bar{X}_n$$

cioè lo stimatore della media μ della popolazione è la media campionaria $\bar{X}_n$

Metodo dei Momenti (Esempi)

- **Esempio 2** Caso 1 parametro $(k = 1)$:
  - Consideriamo una variabile aleatoria uniforme continua nell'intervallo $[0; a]$ con funzione di distribuzione:

$$F_X(x) = \begin{cases} \frac{1}{a} & 0 \leq x \leq a \\ 0 & altrimenti \end{cases}$$



  - Dobbiamo stimare il parametro a, cioe $a = \vartheta_1$
  - Calcoliamo il momento campionario:

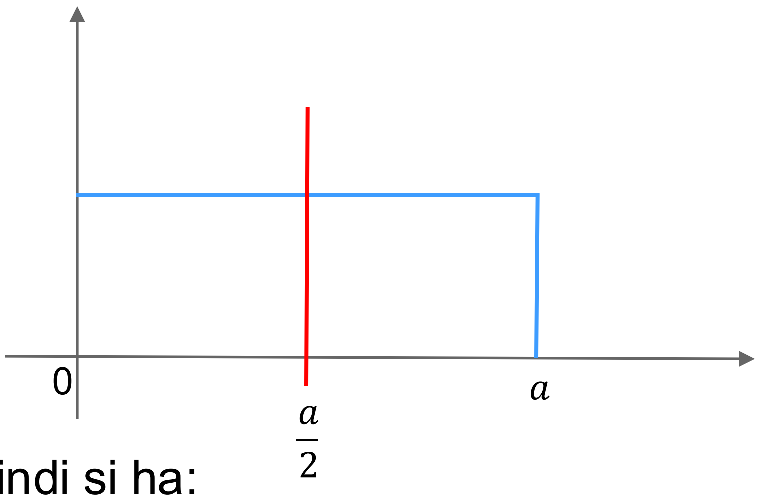
$$\mu_1' = E(X^1) = \frac{a+0}{2} = \boxed{\frac{a}{2}}$$

Valore atteso

Metodo dei Momenti (Esempi)

- **Esempio 2** Caso 1 parametro $(k = 1)$:
  - Consideriamo una variabile aleatoria uniforme continua nell'intervallo $[0; a]$ con funzione di distribuzione:

$$F_X(x) = \begin{cases} \frac{1}{a} & 0 \leq x \leq a \\ 0 & altrimenti \end{cases}$$



  - Calcoliamo il momento campionario:

$$\mu_1' = E(X^1) = \frac{a}{2}$$

  - Il metodo dei momenti ci dice: $\mu_r' = M_r' = \bar{X}_n$ e cioè $\mu_1' = M_1' = \bar{X}_n$, quindi si ha:

$$\mu_1' = \bar{X}_n \qquad \frac{a}{2} = \bar{X}_n \Rightarrow a = 2\bar{X}_n$$

  - Quindi si ha che

$$\hat{a} = \hat{\vartheta}_1 = 2\bar{X}_n$$

cioè il parametro $\hat{\mu}$ deve essere uguale a $\frac{a}{2}$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria **normale** $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria **normale** $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu'_1 = E(X^1) = \mu$$

$$\mu'_2 = E(X^2) = \boxed{\mu^2 + \sigma^2}$$

Nel caso della normale

Poiché $Var(X) = E(X^2) - E(X)^2$
Si ha che:

$$\sigma^2 = Var(X) = E(X^2) - E(X)^2$$

E quindi:

$$\sigma^2 + E(X)^2 = E(X^2)$$

Cioè:

$$\sigma^2 + \mu^2 = E(X^2)$$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria normale $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

  - Quindi viene definito il sistema:

$$\begin{cases} \mu = \bar{X}_n \\ \mu^2 + \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2 \end{cases}$$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria normale $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

  - Quindi viene definito il sistema:

$$\begin{cases} \mu = \bar{X}_n \\ \mu^2 + \sigma^2 = \boxed{\frac{1}{n}\sum_{k=1}^{n} X_k^2} \end{cases}$$

Momento Campionario al quadrato (secondo)

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria normale $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

  - Quindi viene definito il sistema:

$$\begin{cases} \mu = \bar{X}_n \\ \mu^2 + \sigma^2 = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} X_k^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} X_k^2 - \mu^2 \end{cases}$$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria normale $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

  - Quindi viene definito il sistema:

$$\begin{cases} \mu = \bar{X}_n \\ \mu^2 + \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2 - \mu^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2 - \bar{X}_n^{\,2} \end{cases}$$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria normale $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

  - Quindi viene definito il sistema:

$$\begin{cases} \mu = \bar{X}_n \\ \mu^2 + \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2 - \mu^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2 - \bar{X}_n{}^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} (X_k - \bar{X}_n)^2 \end{cases}$$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria normale $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

  - Quindi viene definito il sistema:

Sviluppando il quadrato si ottiene

$$\begin{cases} \mu = \bar{X}_n \\ \mu^2 + \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \mu^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}_n{}^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{X}_n)^2 \end{cases}$$

Metodo dei Momenti (Esempio 2 Parametri)

- **Esempio 3** Caso 2 parametri $(k = 2)$:
  - Consideriamo una variabile aleatoria normale $N(\mu, \sigma^2)$
    - μ e σ sono i valori che vogliamo stimare, quindi $\mu = \vartheta_1$ e $\sigma = \vartheta_2$
  - Calcoliamo il momento campionario assoluto:

$$\mu_1' = E(X^1) = \mu$$
$$\mu_2' = E(X^2) = \mu^2 + \sigma^2$$

  - Quindi viene definito il sistema:

$$\begin{cases} \mu = \bar{X}_n \\ \mu^2 + \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \mu^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}_n{}^2 \end{cases} = \begin{cases} \mu = \bar{X}_n \\ \sigma^2 = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{X}_n)^2 \end{cases}$$

  - Quindi si ha che i due stimatori sono: $\hat{\mu} = \hat{\vartheta}_1 = \bar{X}_n$ e $\hat{\sigma} = \hat{\vartheta}_2 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2}$

Popolazione di Bernoulli

- Vogliamo determinare lo stimatore del parametro p di una popolazione di Bernoulli descritta da una variabile aleatoria $X \sim B(1, p)$ con funzione di probabilità:

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

- Occorre quindi stimare il parametro p considerando che $E(X) = p$:
  - Considerando che $M_r(x_1, x_2, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} X_i^r$
  - Dobbiamo stimare:

$$\hat{p} = \frac{x_1, x_2, \dots, x_n}{n} = \bar{x}$$

- Il metodo dei momenti fornisce quindi come stimatore del parametro p la media campionaria $\hat{X}$
- **Esempio**: Consideriamo un campione di ampiezza 30 contenente i **risultati di lanci indipendenti di una moneta**

```
> campbern<-c(0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0,
+ 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1)
>
> stimap<-mean(campbern)
> stimap
[1] 0.5666667
```

$\boxed{0.5666667} \longrightarrow \hat{p} \approx 0.57$

Popolazione Binomiale

- Vogliamo determinare lo stimatore del parametro p di una popolazione Binomiale descritta da una variabile aleatoria $X \sim B(k,p)$

$$p_X(x) = \binom{k}{x} p^x (1-p)^{\mathrm{k}-x} \quad (x = 0,1,\dots,\mathrm{k})$$

- Occorre stimare il parametro p considerando che $X = Y_1 + Y_2 + \cdots + Y_k$ dove $Y_1, Y_2, \dots, Y_k$ sono di Bernoulli e $E(X) = kp$:
  - Considerando che $M_r(x_1, x_2, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} X_i^r$
  - Dobbiamo stimare:

$$k\hat{p} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{k}\frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\bar{x}}{k}$$

- Il metodo dei momenti fornisce quindi come stimatore del parametro p la media campionaria $\hat{X}$
- **Esempio**: Consideriamo un campione campbinom di ampiezza n = 30 contenente come **risultati il numero di successi ottenuti in k = 10 lanci indipendenti** di una moneta

```
> campbinom<-c(3, 2, 6, 2, 4, 4, 7, 4, 6, 6, 5, 4, 5, 4, 8,
+ 1, 3, 7, 4, 0, 3, 7, 4, 4, 3, 2, 5, 5, 3, 2)
>
> lanci<-10
> stimap<-mean(campbinom)/lanci
> stimap
[1] 0.41
```

$\hat{p} = 0.41$

Popolazione Geometrica

- Vogliamo determinare lo stimatore del parametro p di una popolazione Geometrica descritta da una variabile aleatoria $X \sim B(k, p)$

$$p_Y(y) = p\,(1-p)^y \quad (y = 0,1, \dots)$$

- Occorre quindi stimare il parametro p:
  - Poiché $E(Y) = \frac{(1-p)}{p}$ ponendo $\vartheta = \frac{(1-p)}{p}$
  - Si ha che dobbiamo stimare:

$$\hat{\vartheta} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad ossia \quad \frac{(1-\hat{p})}{\hat{p}} = \bar{x} \quad ossia \quad \frac{1}{1+\bar{x}} = \hat{p}$$

- Il metodo dei momenti fornisce quindi come stimatore del parametro $\hat{\vartheta} = E(Y) = \frac{(1-p)}{p}$ la media campionaria $\hat{X}$
- **Esempio**: Consideriamo un campione campgeom di ampiezza 30 contenente **come risultati il numero di fallimenti prima di ottenere il primo successo in lanci ripetuti di una moneta** si ha

```
> campgeom<-c(7, 2, 4, 0, 1, 1, 0, 2, 8, 2, 1, 0, 8, 1, 0,
+ 10, 9, 1, 5, 8, 0, 4, 3, 7,15, 9, 1, 0, 0, 0)
>
> stimap<-1/(1+mean(campgeom))
> stimap
[1] 0.2158273
```

$\boxed{0.2158273} \rightarrow \hat{p} = 0.21$

Popolazione Geometrica Modificata

- Vogliamo determinare lo stimatore del parametro p di una popolazione Geometrica descritta da una variabile aleatoria $X \sim BN(1, p)$

$$p_X(x) = p\,(1-p)^{x-1} \quad (x = 1,2,\dots)$$

- Occorre quindi stimare il parametro p :
  - Poiché $E(X) = \frac{1}{p}$ ponendo $\vartheta = \frac{1}{p}$
  - Si ha che dobbiamo stimare:

$$\hat{\vartheta} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad ossia \quad \frac{1}{\hat{p}} = \bar{x} \;\; ossia \quad \frac{1}{\bar{x}} = \hat{p}$$

- Il metodo dei momenti fornisce quindi come stimatore del parametro $\vartheta = E(X) = \frac{1}{p}$ la media campionaria $\bar{X}$
- **Esempio**: Consideriamo un campione campgeommod di ampiezza 30 come **risultati i tempi di attesa per ottenere il primo successo in lanci ripetuti di una moneta**

```
> campgeommod<-c(5, 1, 6, 6, 2, 6, 3, 1, 1, 2, 8, 2, 5, 4, 4,
+ 2, 1, 6, 7, 1, 6, 4, 4, 3, 2, 3, 5, 2, 4, 3)
>
> stimap<-1/mean(campgeommod)
> stimap
[1] 0.2752294
```

$\boxed{0.2752294} \longrightarrow \hat{p} = 0.27$

Popolazione Binomiale Negativa

- Vogliamo determinare lo stimatore del parametro p di una popolazione Binomiale Negativa descritta da una variabile aleatoria $Y \sim BN(k,p)$ dove $Y = Y_1 + Y_2 + \cdots + Y_k$ con $Y_1, Y_2, \dots, Y_k$ variabili aleatorie indipendenti e identicamente distribuite in modo geometrico
- Occorre quindi stimare il parametro p:
  - Poiché $E(Y) = \frac{k(1-p)}{p}$ ponendo $\vartheta = \frac{k(1-p)}{p}$
  - Si ha che:

$$\hat{\vartheta} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad ossia \frac{k(1-\hat{p})}{\hat{p}} = \bar{x} \;\; ossia \quad \frac{k}{k+\bar{x}} = \hat{p}$$

- Il metodo dei momenti fornisce quindi come stimatore del parametro $\vartheta = E(Y) = \frac{k(1-p)}{p}$ la media campionaria $\bar{X}$
- **Esempio**: Consideriamo un campione campgeom di ampiezza n = 30 contenente come **risultati il numero di fallimenti prima di ottenere il quinto successo in lanci ripetuti di una moneta** si ha in lanci ripetuti di una moneta si ha

```
> campbinneg<-c(2,3,6,20,9,8,13 ,12 ,10 ,21,  8, 13, 18, 21,  5,
+ 8,  5, 26,  4,  8,  6,  3,  3,  1, 10, 13,  3,  7, 10, 17)
>
> k<-5
> stimap<-k/(k+mean(campbinneg))
> stimap
[1] 0.3386005
```

$\longrightarrow \hat{p} = 0.33$

Popolazione Binomiale Negativa Modificata

- Vogliamo determinare lo stimatore del parametro p di una popolazione Binomiale Negativa Modificata descritta da una variabile aleatoria $Y \sim BN^*(k,p)$ dove $X = Z_1 + Z_2 + \cdots + Z_k$ con $Z_1, Z_2, \dots, Z_k$ variabili aleatorie indipendenti e identicamente distribuite in modo geometrico modificato
- Occorre quindi stimare il parametro p:
  - Poiché $E(Y) = \frac{k}{p}$ ponendo $\vartheta = \frac{k}{p}$
  - Si ha che:

$$\hat{\vartheta} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad ossia \; \frac{k}{\hat{p}} = \bar{x} \; ossia \quad \frac{k}{\bar{x}} = \hat{p}$$

- Il metodo dei momenti fornisce quindi come stimatore del parametro $\vartheta = E(Y) = \frac{k(1-p)}{p}$ la media campionaria $\bar{X}$
- **Esempio**: Consideriamo un campione campgeommod di ampiezza 30 contenente come risultati **i tempi di attesa per ottenere il quinto successo in lanci ripetuti di una moneta**

```
> campbinnegmod<-c(20,22,16,13,12,13,28,23,19, 9,19,15,12,14,16,
+ 15, 9,17, 7,15,17, 8,16,20,13,12,14,16,31,14)
>
> k<-5
> stimap<-k/mean(campbinnegmod)
> stimap
[1] 0.3157895
```

$\longrightarrow \hat{p} = 0.31$

Popolazione Poisson

- Vogliamo determinare lo stimatore del valore medio λ di una popolazione di Poisson descritta da una variabile aleatoria $X \sim P(\lambda)$ con funzione di probabilità:

$$p_X(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x = 0,1, \ldots \quad \lambda > 0$$

- Occorre quindi stimare il parametro λ:
  - Poiché $E(X) = \lambda$ ponendo $\vartheta = \lambda$
  - Si ha che:

$$\hat{\lambda} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \bar{x}$$

- Il metodo dei momenti fornisce quindi come stimatore del parametro $E(X) = \lambda$ la media campionaria $\bar{X}$
- **Esempio**: Consideriamo un campione camppois di ampiezza 30 contenente come risultati il numero di utenti che arrivano ad un centro di calcolo in intervalli di 10 minuti

```
> camppois<-c(2, 2, 1, 2, 2, 1, 3, 3, 4, 1, 3, 6, 2, 2, 2,
+ 1, 2, 4, 2, 2, 6, 0, 1, 8, 2, 3, 4, 1, 1, 2)
>
> stimalambda<-mean(camppois)
> stimalambda
[1] 2.5
```

$\hat{\lambda} = 2.5$

Popolazione Uniforme

- Vogliamo determinare lo stimatore del parametro ϑ di una popolazione Uniforme descritta da una variabile aleatoria $X \sim U(0, \vartheta)$ con funzione densità di probabilità:

$$f(x) = \frac{1}{\vartheta} \qquad 0 < x < \vartheta$$

- Occorre quindi stimare il parametro ϑ considerando che $E(X) = \frac{0+\vartheta}{2} = \frac{\vartheta}{2}$
  - Si ha che: $\frac{\widehat{\vartheta}}{2} = \frac{x_1+x_2+\cdots+x_n}{n} \quad ossia \quad \hat{\vartheta} = 2\bar{x}$
- Il metodo dei momenti fornisce quindi come stimatore del parametro $\vartheta = E(X) = \frac{\vartheta}{2}$ la media campionaria $2\bar{X}$
- **Esempio**: Consideriamo Consideriamo un campione campunif di ampiezza 30 contenente come risultati i tempi, misurati in minuti e supposti uniformi in un intervallo (0, ϑ), necessari per soddisfare le richieste di utenti che arrivano ad un centro di calcolo

```
> campunif<-c(1.556, 1.357, 1.574, 0.133, 1.748, 0.348, 0.566,
+  0.767, 0.374, 1.856, 0.488, 0.327, 0.813, 0.005, 0.191, 1.311,
+  0.345, 0.934, 0.140, 0.796, 0.254, 0.962, 1.318, 1.71, 0.257,
+  0.605, 0.516, 0.083, 0.052, 0.290)
>
> stimatheta<-2.0*mean(campunif)
> stimatheta
[1] 1.445067
```

$\boxed{1.445067} \longrightarrow \hat{\vartheta} = 1.44$

Popolazione Esponenziale

- Si desidera determinare con il metodo dei momenti lo stimatore del valore medio $\frac{1}{\lambda}$ di una popolazione espoenziale descritta da una variabile aleatoria $\mathrm{X} \sim \varepsilon(\lambda)$ con densità di probabilità:

$$f_X(x) = \lambda e^{-\lambda x} \qquad x > 0$$

- Occorre quindi stimare il parametro ϑ considerando che $E(X) = \frac{1}{\lambda}$ e ponendo $\vartheta = \frac{1}{\lambda}$
  - Si ha che: $\frac{\hat{\vartheta}}{2} = \frac{x_1+x_2+\cdots+x_n}{n} \quad ossia \quad \hat{\lambda} = \frac{1}{\bar{x}}$
- Il metodo dei momenti fornisce quindi come stimatore del parametro $\vartheta = \frac{1}{\lambda}$ la media campionaria $\bar{X}$
- **Esempio**: Consideriamo un campione campexp di ampiezza 30 contenente come risultati i tempi di interarrivo (tra arrivi successivi), supposti esponenziali, di utenti che arrivano ad un centro di calcolo

```
> campexp<-c(0.196, 0.409, 0.225, 0.224, 0.248, 0.280, 0.791,
+ 1.165, 0.355, 1.055, 0.393, 0.711, 0.455, 0.066, 0.179,
+ 0.543, 0.067, 0.635, 0.540, 1.454,0.213, 0.532, 0.613, 1.876,
+ 0.047, 2.042, 0.018, 1.105, 0.098, 0.032)
>
> stimatheta<-1.0/mean(campexp)
> stimatheta
[1] 1.810829
```

$\hat{\vartheta} = 1.81$

Popolazione Normale

- **Esempio**: Consideriamo un campione campnorm di ampiezza 30 contenente come risultati le lunghezze in metri, supposte normali, riscontrate nel misurare dei tubi prodotti da un'industria

```
> campnorm<-c(2.86, 3.03, 3.05, 3.32, 3.06, 2.91, 3.11, 3.21,
+ 2.85, 2.86, 2.78, 3.28, 3.39, 3.16, 3.05, 3.01, 3.10,
+ 2.88, 3.25, 2.89, 2.75, 2.99, 3.34, 2.93,3.14, 2.99,
+ 2.97,3.21, 3.27, 2.91)
>
> stimamu<-mean(campnorm)
> stimamu
[1] 3.052461
>
> stimasigma2<-(length(campnorm)-1)*var(campnorm)/length(campnorm)
> stimasigma2
[1] 0.02996195
```

- la stima del parametro μ con il metodo dei momenti è $\hat{\mu} = 3.052461$ e la stima del parametro σ^2 con il metodo dei momenti è $\widehat{\sigma^2} = 0.02996$

STATISTICA E ANALISI DEI DATI

Metodo della massima verosimiglianza

Il metodo della Massima Verosimiglianza

- Il metodo della massima verosimiglianza:
  - consiste nel massimizzare la funzione di verosimiglianza rispetto ai parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_k$
  - E cerca di determinare da quale
    - funzione di probabilità congiunta (nel caso di popolazione <u>discreta</u>)
    - densità di probabilità congiunta (nel caso di popolazione assolutamente <u>continua</u>)
  - è più **verosimile** (è più <u>plausibile</u>) che provenga il campione osservato $(x_1, x_2, \dots, x_n)$
- **Obiettivo**:
  - Determinare i valori $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ che rendono **massima** la funzione di verosimiglianza e che quindi offrano la migliore spiegazione (o **rappresentazione**) del campione osservato $(x_1, x_2, \dots, x_n)$
- I valori $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ che massimizzano la funzione di verosimiglianza sono indicati con $\hat{\vartheta}_1, \hat{\vartheta}_2, \dots, \hat{\vartheta}_k$ essi costituiscono le **stime di massima verosimiglianza** dei parametri non noti $\vartheta_1, \vartheta_2, \dots, \vartheta_k$

Funzione Di Verosimiglianza

- Sia $X_1, X_2, \dots, X_n$ un campione casuale di ampiezza n estratto dalla popolazione
- La funzione di verosimiglianza $L(\vartheta_1, \vartheta_2, \dots, \vartheta_k) = L(\vartheta_1, \vartheta_2, \dots, \vartheta_k; x_1, x_2, \dots, x_n)$ del campione osservato $(x_1, x_2, \dots, x_n)$ è la **funzione di probabilità congiunta** (nel caso di popolazione discreta) oppure la **funzione densità di probabilità congiunta** (nel caso di popolazione continua) del campione casuale $X_1, X_2, \dots, X_n$, ossia

$$L(\vartheta_1, \vartheta_2, \dots, \vartheta_k) = L(\vartheta_1, \vartheta_2, \dots, \vartheta_k; x_1, x_2, \dots, x_n)$$

Funzione Di Verosimiglianza

- Sia $X_1, X_2, \dots, X_n$ un campione casuale di ampiezza n estratto dalla popolazione
- La funzione di verosimiglianza $L(\vartheta_1, \vartheta_2, \dots, \vartheta_k) = L(\vartheta_1, \vartheta_2, \dots, \vartheta_k; x_1, x_2, \dots, x_n)$ del campione osservato $(x_1, x_2, \dots, x_n)$ è la **funzione di probabilità congiunta** (nel caso di popolazione discreta) oppure la **funzione densità di probabilità congiunta** (nel caso di popolazione continua) del campione casuale $X_1, X_2, \dots, X_n$, ossia

$$L(\vartheta_1, \vartheta_2, \dots, \vartheta_k) = L(\vartheta_1, \vartheta_2, \dots, \vartheta_k; x_1, x_2, \dots, x_n)$$

Funzione di verosimiglianza e rappresenta la probabilità di osservare il campione dati i parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_k$

Per l'indipendenza delle variabili si ha che

$$\boxed{L(\vartheta_1, \vartheta_2, \dots, \vartheta_k) = f(x_1; \vartheta_1, \vartheta_2, \dots, \vartheta_k) f(x_2; \vartheta_1, \vartheta_2, \dots, \vartheta_k) \dots f(x_n; \vartheta_1, \vartheta_2, \dots, \vartheta_k) = \prod_{i=1}^{n} f_X(x_i; \vartheta_1, \vartheta_2, \dots, \vartheta_k)}$$

- Dato che L è un prodotto e tutti gli $f(x_i; \vartheta_1, \vartheta_2, \dots, \vartheta_k)$ sono numeri positivi e possiamo fare il $\log$:

$$\log L(\vartheta_1, \vartheta_2, \dots, \vartheta_k) = \sum_{i=1}^{n} \log f(x_i; \vartheta_1, \vartheta_2, \dots, \vartheta_k)$$

- Si vogliono trovare i valori dei parametri $\vartheta_1, \dots, \vartheta_k$che **massimizzano** la log−verosimiglianza:

$$(\hat{\vartheta}_1, \dots, \hat{\vartheta}_k) = \arg \max_{\vartheta_1, \dots, \vartheta_k} \log L(\vartheta_1, \dots, \vartheta_k)$$

Popolazione di Bernoulli

- Vogliamo lo stimatore di massima verosimiglianza del <u>parametro p</u> di una popolazione di Bernoulli descritta da una variabile aleatoria $X \sim B(1, p)$

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

- Si ha che la funzione di verosimiglianza è:

$$L(p) = (p^{x_1}(1-p)^{1-x_1})(p^{x_2}(1-p)^{1-x_2}) \dots (p^{x_n}(1-p)^{1-x_n} = p^{x_1+x_2+\dots+x_n}(1-p)^{n-x_1+x_2+\dots+x_n} = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

Dove le x_i possono assumere il valore 0 o 1

- Si nota che

$$\log L(p) = \log p \sum_{i=1}^{n} x_i + \left[n - \sum_{i=1}^{n} x_i\right] \log(1-p) \qquad (0 < p < 1)$$

Da cui si ottiene:

$$\frac{\boxed{d} \log L(p)}{\boxed{d}p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\left[n - \sum_{i=1}^{n} x_i\right] = \frac{1}{p(1-p)}\sum_{i=1}^{n} x_i - \frac{n}{1-p}$$

Derivata

$$= \frac{n}{p(1-p)}\left[\frac{1}{n}\sum_{i=1}^{n} x_i - p\right].$$

La stima di massima verosimiglianza del parametro p è: $$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$ **<u>La media campionaria</u>**

Popolazione Binomiale

- Vogliamo determinare lo stimatore di massima verosimiglianza del parametro p di una popolazione Binomiale descritta da una variabile aleatoria $X \sim B(k, p)$

$$p_X(x) = \binom{k}{x} p^x (1-p)^{\mathrm{k}-x} \quad (x = 0,1, \ldots, \mathrm{k})$$

- Si ha:

$$L(p) = \binom{k}{x_1} p^{x_1} (1-p)^{k-x_1} \binom{k}{x_2} p^{x_2} (1-p)^{k-x_2} \cdots \binom{k}{x_n} p^{x_n} (1-p)^{k-x_n}$$

$$= \binom{k}{x_1}\binom{k}{x_2} \cdots \binom{k}{x_n} p^{x_1+x_2+\ldots+x_n} (1-p)^{nk-(x_1+x_2+\ldots+x_n)} \qquad (0 < p < 1)$$

- Dove le x_i possono assumere il valore 0 o 1
- Procedendo come per la popolazione di Bernoulli, la stima del parametro kp è:

$$k\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{ossia} \qquad \widehat{p} = \frac{1}{k}\,\frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\overline{x}}{k}.$$

- Quindi, per una popolazione binomiale lo stimatore di massima verosimiglianza e dei momenti del valore medio $E(X) = kp$ è la **media campionaria**

Popolazione Geometrica Modificata

- Vogliamo determinare lo stimatore di massima verosimiglianza del valore medio di una popolazione geometrica modificata descritta da una variabile aleatoria $X \sim BN(1, p)$

$$p_X(x) = p\,(1-p)^{x-1} \quad (x = 1,2,\dots)$$

- Occorre quindi stimare il parametro p, poiché $E(X) = \frac{1}{p}$ ponendo $\vartheta = \frac{1}{p}$ si ha:

$$L(\vartheta) = \left(\frac{1}{\vartheta}\right)^n \left(1 - \frac{1}{\vartheta}\right)^{x_1+x_2+\ldots+x_n-n}$$
$$= \left(\frac{1}{\vartheta}\right)^{x_1+x_2+\ldots+x_n} (\vartheta - 1)^{x_1+x_2+\ldots+x_n-n} \qquad (\vartheta > 1)$$

- Dove le x_i possono sono interi positivi. Da cui: $\log L(\vartheta) = -\log\vartheta \sum_{i=1}^{n} x_i + \log(\vartheta-1)\left(\sum_{i=1}^{n} x_i - n\right) \qquad (\vartheta > 1)$

e quindi:

$$\frac{d\log L(\vartheta)}{d\vartheta} = -\frac{1}{\vartheta}\sum_{i=1}^{n} x_i + \frac{1}{\vartheta-1}\left(\sum_{i=1}^{n} x_i - n\right) = \left(-\frac{1}{\vartheta} + \frac{1}{\vartheta-1}\right)\sum_{i=1}^{n} x_i - \frac{n}{\vartheta-1}$$

$$= \frac{1}{\vartheta\,(\vartheta-1)}\sum_{i=1}^{n} x_i - \frac{n}{\vartheta-1} = \frac{n}{\vartheta\,(\vartheta-1)}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \vartheta\right)$$

Popolazione Geometrica Modificata

- Vogliamo determinare lo stimatore di massima verosimiglianza del valore medio di una popolazione geometrica modificata descritta da una variabile aleatoria $X \sim BN(1,p)$

$$p_X(x) = p\,(1-p)^{x-1} \qquad (x = 1,2,\dots)$$

- Occorre quindi stimare il parametro p, poiché $E(X) = \frac{1}{p}$ ponendo $\vartheta = \frac{1}{p}$ si ha:

$$\begin{aligned} L(\vartheta) &= \left(\frac{1}{\vartheta}\right)^n \left(1-\frac{1}{\vartheta}\right)^{x_1+x_2+\ldots+x_n-n} \\ &= \left(\frac{1}{\vartheta}\right)^{x_1+x_2+\ldots+x_n} (\vartheta-1)^{x_1+x_2+\ldots+x_n-n} \qquad (\vartheta > 1) \end{aligned}$$

- Si ottiene che la stima di massima verosimiglianza del parametro $\vartheta = \frac{1}{p}$ è:

$$\widehat{\vartheta} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{ossia} \qquad \widehat{p} = \frac{1}{\overline{x}}.$$

- Quindi, per una popolazione geometrica modificata lo stimatore di massima verosimiglianza e dei momenti del valore medio $E(X) = \frac{1}{p}$ è la **media campionaria**

Popolazione Poisson

- Vogliamo determinare lo stimatore di massima verosimiglianza del valore medio λ di una popolazione di Poisson descritta da una variabile aleatoria $X \sim P(\lambda)$ con funzione di probabilità:

$$p_X(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad \mathrm{x} = 0,1, \ldots \quad \lambda > 0$$

- Occorre quindi stimare il parametro λ, poiché $E(X) = \lambda$ ponendo $\vartheta = \lambda$, si ha che:

$$L(\lambda) = \frac{\lambda^{x_1 + x_2 + \ldots + x_n}}{x_1!\, x_2! \cdots x_n!} e^{-n\,\lambda} \qquad (\lambda > 0)$$

dove le x_i possono sono interi positivi. Si nota quindi che:

$$\log L(\lambda) = \log \lambda \sum_{i=1}^{n} x_i - n\,\lambda - \log\left[x_1!\, x_2! \cdots x_n!\right]$$

- Da dui segue:

$$\frac{d \log L(\lambda)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = \frac{n}{\lambda} \left(\frac{1}{n} \sum_{i=1}^{n} x_i - \lambda \right)$$

- La stima di massima verosimiglianza del parametro λ è: $\widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$

- Lo stimatore del parametro $E(X) = \lambda$ è la **media campionaria** $\bar{X}$

Popolazione Uniforme

- Vogliamo determinare lo stimatore di massima verosimiglianza del parametro ϑ di una popolazione Uniforme descritta da una variabile aleatoria $\mathrm{X} \sim U(0, \vartheta)$ con funzione densità di probabilità:

$$f(x) = \frac{1}{\vartheta} \qquad 0 < x < \vartheta$$

- Si ha che:

$$L(\vartheta) = \frac{1}{\vartheta^n} \qquad (\vartheta > 0)$$

dove $0 < x_1 < \vartheta, 0 < x_2 < \vartheta, \dots, 0 < x_n < \vartheta$

- Il valore che deve assumere ϑ non può essere inferiore a nessun $x_1, x_2, \dots, x_n$, altrimenti non si rispetta la condizione dell'intervallo
  - Quindi o $\vartheta > \max(x_1, x_2, \dots, x_n)$ la funzione di verosimiglianza è strettamente monotona decrescente e quindi si può assumere la stima di massima verosimiglianza di ϑ:

$$\hat{\vartheta} = \max(x_1, x_2, \dots, x_n)$$

- Lo stimatore di massima verosimiglianza del parametro ϑ è $\hat{\Theta} = \max(X_1, X_2, \dots, X_n)$
  - Dato che $E(X) = \frac{\vartheta}{2}$ lo stimatore di ϑ è diverso da quello ottenuto dal metodo dei momenti: $\hat{\Theta} = 2\bar{X}$

Popolazione Uniforme

- **Esempio**: Consideriamo un campione campunif di ampiezza 30 contenente come risultati i tempi, misurati in ore e supposti uniformi in un intervallo $(0, \vartheta)$ necessari per soddisfare le richieste di utenti che arrivano ad un centro di calcolo

```
> campunif<-c(1.556, 1.357, 1.574, 0.133, 1.748, 0.348, 0.566,
+ 0.767, 0.374, 1.856, 0.488, 0.327, 0.813, 0.005, 0.191, 1.311,
+ 0.345, 0.934, 0.140, 0.796, 0.254, 0.962, 1.318, 1.71, 0.257,
+ 0.605, 0.516, 0.083, 0.052, 0.290)
>
> stimatheta<-max(campunif)
> stimatheta
[1] 1.856
```

$\hat{\vartheta} = 1.856$

Popolazione Esponenziale

- Si desidera determinare con il metodo dei momenti lo stimatore del valore medio $\frac{1}{\lambda}$ di una popolazione espoenziale descritta da una variabile aleatoria $X \sim \varepsilon(\lambda)$ con densità di probabilità:

$$f_X(x) = \lambda e^{-\lambda x} \qquad x > 0$$

- Occorre quindi stimare il parametro ϑ considerando che $E(X) = \frac{1}{\lambda}$ e ponendo $\vartheta = \frac{1}{\lambda}$ si ha che:

$$L(\vartheta) = \left(\frac{1}{\vartheta}\right)^n \exp\left\{-\frac{1}{\vartheta}\sum_{i=1}^{n} x_i\right\} \qquad (\vartheta > 0)$$

Dove le x_i sono positive.

Da cui:

$$\log L(\vartheta) = -n \log \vartheta - \frac{1}{\vartheta}\sum_{i=1}^{n} x_i \longrightarrow \frac{d \log L(\vartheta)}{d\vartheta} = -\frac{n}{\vartheta} + \frac{1}{\vartheta^2}\sum_{i=1}^{n} x_i = \frac{n}{\vartheta^2}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \vartheta\right)$$

- La stima di massima verosimiglianza del parametro $\vartheta = \frac{1}{\lambda}$ è quindi la **media campionaria** $\bar{X}$:

$$\widehat{\vartheta} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Popolazione Normale

- Si è interessati a determinare lo stimatore di massima Verosimiglianza dei parametri μ e σ^2 di una popolazione Normale descritta da una variabile aleatoria $X \sim N(\mu, \sigma)$ di densità di probabilità:

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

- Poiché $E(X) = \mu$ e $E(X^2) = \mu^2 + \sigma^2$ si ha che:

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\,\sigma^2}\right\} \qquad (\mu \in \mathbb{R},\ \sigma > 0)$$

Da cui si ricava

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad (\mu \in \mathbb{R},\ \sigma > 0)$$

e quindi si ha:

$$\boxed{\frac{\partial \log L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{n}{\sigma^2}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right)} \longrightarrow \text{Massimizzazione rispetto a } \mu$$

$$\boxed{\frac{\partial \log L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = -\frac{n}{2\sigma^4}\left(\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right)} \longrightarrow \text{Massimizzazione rispetto a } \sigma$$

Popolazione Normale

- Si è interessati a determinare lo stimatore di massima Verosimiglianza dei parametri μ e σ^2 di una popolazione Normale descritta da una variabile aleatoria $\mathrm{X} \sim \mathrm{N}(\mu, \sigma)$ di densità di probabilità:

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

- Le stime di massima verosimiglianza dei parametri μ e σ^2 sono rispettivamente:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2$$

- Lo stimatore di massima verosimiglianza e dei momenti del valore medio μ è la media campionaria X
- Invece lo stimatore di massima verosimiglianza e dei momenti della varianza $\sigma^2 = \frac{(n-1)S^2}{n}$

Proprietà degli stimatori

X	Metodo dei momenti	Metodo della massima verosimiglianza
Bernoulli $X \sim \mathcal{B}(1,p)$ $E(X)=p$	$\overline{X}$	$\overline{X}$
Binomiale $X \sim \mathcal{B}(k,p)$ $E(X)=k\,p$	$\overline{X}$	$\overline{X}$
Geometrica $X \sim \mathcal{BN}(1,p)$ $E(X)=(1-p)/p$	$\overline{X}$	$\overline{X}$
Geometrica modificata $X \sim \mathcal{BN}^*(1,p)$ $E(X)=1/p$	$\overline{X}$	$\overline{X}$
Binomiale Negativa $X \sim \mathcal{BN}(k,p)$ $E(X)=k(1-p)/p$	$\overline{X}$	$\overline{X}$
Binomiale Negativa modificata $X \sim \mathcal{BN}^*(k,p)$ $E(X)=k/p$	$\overline{X}$	$\overline{X}$
Poisson $X \sim \mathcal{P}(\lambda)$ $E(X)=\lambda$	$\overline{X}$	$\overline{X}$
Uniforme in $(0,\vartheta)$ $X \sim \mathcal{U}(0,\vartheta)$ $E(X)=\vartheta/2$	(1) $\overline{X}$	(2) $\frac{\max(X_1,X_2,\ldots,X_n)}{2}$
Esponenziale $X \sim \mathcal{E}(\lambda)$ $E(X)=1/\lambda$	$\overline{X}$	$\overline{X}$
Normale $X \sim \mathcal{N}(\mu,\sigma)$ $E(X)=\mu$ $\mathrm{Var}(X)=\sigma^2$	(1) $\overline{X}$ (2) $\frac{(n-1)S^2}{n}$	(1) $\overline{X}$ (2) $\frac{(n-1)S^2}{n}$