

## 10 Il metodo dei momenti

I due estimatori più naturali sono media e varianza empirica; in entrambi i casi possono essere interpretati come lo stimatore che si ottiene sostituendo la distribuzione empirica a quella teorica nel calcolo di parametri che possono esser visti come valori medi di momenti del campione aleatorio. Il metodo dei momenti può essere visto come la generalizzazione di questa idea.

### Spiegazione Semplice

#### Cos'è il "Metodo dei Momenti" (MoM)?

È una "ricetta" (un metodo) per creare degli estimatori per i parametri  $\theta$  di una distribuzione.

L'idea è incredibilmente semplice e intuitiva: **"Facciamo in modo che i 'momenti' del nostro campione siano uguali ai 'momenti' veri della distribuzione."**

- **Momento Teorico (o "della popolazione"):** È una formula che dipende dai parametri veri  $\theta$ . Esempio: Il 1° momento teorico è la media vera,  $\mu_1 = \mathbb{E}[X]$ .
- **Momento Empirico (o "del campione"):** È la stessa cosa, ma calcolata sui nostri dati. Esempio: Il 1° momento empirico è la media campionaria,  $m_1 = \frac{1}{n} \sum X_i$ .

Il metodo consiste nel:

1. Scrivere le equazioni per i primi  $k$  momenti teorici in funzione dei  $k$  parametri  $\theta$ .
2. Calcolare i primi  $k$  momenti empirici dai dati.
3. "Eguagliare" i due:  $\mu_1 = m_1$ ,  $\mu_2 = m_2$ , ecc.
4. Risolvere questo sistema di equazioni per trovare gli estimatori  $\tilde{\theta}$ .

Supponiamo ora di avere un campione aleatorio  $X_1, \dots, X_n$  estratto da una legge di probabilità  $P_\theta$  con  $\theta = (\theta_1, \dots, \theta_k)$ . Supponiamo anche che questa legge ammetta  $k$  momenti finiti, cioè  $\mu_j = E[X_1^j]$  sia ben definito per  $j = 1, \dots, k$ . Chiamiamo  $m_j$  i momenti empirici  $m_j := n^{-1} \sum_{i=1}^n X_i^j$ ; l'idea di fondo è scegliere il valore dei parametri  $\theta_1, \dots, \theta_k$  in modo che valga l'uguaglianza vettoriale:

$$\begin{aligned} m_1 &= \mu_1(\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_k) \\ m_2 &= \mu_2(\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_k) \\ &\dots \\ m_k &= \mu_k(\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_k) \end{aligned}$$

Più precisamente

**Definition 86.** Assumiamo che esista un diffeomorfismo  $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$  tale che  $g(\mu_1, \dots, \mu_k) = (\theta_1, \dots, \theta_k)$ , per ogni valore  $(\theta_1, \dots, \theta_k)$ . Lo stimatore del metodo dei momenti è allora definito da

$$(\tilde{\theta}_1, \dots, \tilde{\theta}_k) := g(m_1, \dots, m_k).$$

### Spiegazione Semplice

#### Chi è la funzione "g"?

La funzione  $g$  è semplicemente la "formula risolutiva" del sistema di equazioni.

- **Il Sistema:**  $m_k = \mu_k(\tilde{\theta})$ . (Passo 3)
- **La Soluzione:**  $\tilde{\theta} = g(m_k)$ . (Passo 4)

$g$  è la funzione (o l'insieme di formule) che ci permette di calcolare gli estimatori  $\tilde{\theta}$  una volta che abbiamo inserito i momenti empirici  $m_k$  calcolati dai dati.

Il termine "diffeomorfismo" è una parola tecnica per dire che la funzione  $g$  è "ben comportata": è invertibile e non ha "salti" o "spigoli", il che ci garantisce che una soluzione esista e sia stabile.

Per determinare le proprietà asintotiche del nostro stimatore dobbiamo rinforzare leggermente le nostre ipotesi.

**Condition 87.** Le variabili aleatorie  $X_1, \dots, X_n$  sono indipendenti ed identicamente distribuite con momenti di ordine  $2k$  finiti.

**Proposition 88.** Sotto la precedente condizione, vale il teorema del limite centrale multivariato

$$\sqrt{n} \begin{pmatrix} m_1 - \mu_1 \\ \vdots \\ m_k - \mu_k \end{pmatrix} \xrightarrow{d} N \left( \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \Omega \right)$$

dove

$$\Omega := \begin{pmatrix} E[(X_1 - \mu_1)^2] & E[(X_1 - \mu_1)(X_1^2 - \mu_2)] & \dots & E[(X_1 - \mu_1)(X_1^k - \mu_k)] \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & E[(X_1^k - \mu_k)^2] \end{pmatrix}$$

### Spiegazione Semplice

#### La "pagella" dei Momenti Empirici

Prima di valutare lo stimatore finale  $\tilde{\theta}$ , valutiamo i "mattoni" con cui è costruito: i momenti empirici  $m_k$ .

Questa proposizione ci dice che i momenti empirici  $m_k$  sono ottimi estimatori dei momenti veri  $\mu_k$ . Come per il Teorema del Limite Centrale (Capitolo 7), il loro errore (la differenza  $m_k - \mu_k$ ), "zoomato" con  $\sqrt{n}$ , non collassa a zero ma si stabilizza in una **Gaussiana multivariata** (una "campana" a più dimensioni).

La matrice  $\Omega$  è semplicemente la matrice di covarianza (la "forma" e la "larghezza") di questa campana multivariata.

**Proposition 89.** Sotto la precedente condizione, lo stimatore del metodo dei momenti converge in probabilità al valore vero dei parametri e vale il teorema del limite centrale:

$$\sqrt{n} \begin{pmatrix} \tilde{\theta}_1 - \theta_1 \\ \vdots \\ \tilde{\theta}_k - \theta_k \end{pmatrix} \xrightarrow{d} N \left( \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, Jg(\mu_1, \dots, \mu_k) \Omega Jg(\mu_1, \dots, \mu_k)^T \right)$$

### Spiegazione Semplice

#### La "pagella" dello Stimatore MoM

Questo è il risultato finale. Ci dice che lo stimatore del Metodo dei Momenti ( $\tilde{\theta}$ ) ha una buona "pagella" (vedi Capitolo 9):

- **È Consistente:**  $\tilde{\theta}$  converge in probabilità a  $\theta$ . (Funziona! Impara dai dati).
- **È Asintoticamente Gaussiano:** Il suo errore "zoomato" ( $\sqrt{n}(\tilde{\theta} - \theta)$ ) segue una curva a campana (Gaussiana). Questo ci permette di calcolare intervalli di confidenza e p-value.

**Come si ottiene questo risultato?** Si applica il **Metodo Delta** (visto nel Capitolo 8) al risultato della Proposizione 88.

- **CLT per i Momenti (Prop 88):**  $\sqrt{n}(m - \mu) \rightarrow N(0, \Omega)$
- **Funzione di Mappa (Def 86):**  $\tilde{\theta} = g(m)$
- **Metodo Delta (Teo 74):**  $\sqrt{n}(g(m) - g(\mu)) \rightarrow N(0, J_g \Omega J_g^T)$

La formula  $Jg\Omega Jg^T$  è la versione multivariata della  $[g'(\mu)]^2\sigma^2$  vista nel Metodo Delta, e ci dà la varianza (la "precisione") del nostro stimatore.

*Dimostrazione.* Delta Method multivariato. □

**Example 90** (Legge Gamma). Supponiamo di avere un campioni di variabili IID con legge Gamma di parametri  $\alpha$  e  $\beta$  cioè con densità

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-\frac{x}{\beta}) I_{[0,\infty)}(x)$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx.$$

Come noto (si veda l'appendice) abbiamo

$$\begin{aligned} E[X_1] &= \alpha\beta, \\ E[X_1^2] &= (\alpha\beta)^2 + \alpha\beta^2 = \alpha\beta^2(\alpha + 1) \\ Var[X_1] &= \alpha\beta^2 \end{aligned}$$

Possiamo pertanto scrivere

$$\begin{aligned} \frac{\mu_2 - \mu_1^2}{\mu_1} &= \frac{(\alpha\beta)^2 + \alpha\beta^2 - (\alpha\beta)^2}{\alpha\beta} = \beta \\ \frac{\mu_1^2}{\mu_2 - \mu_1^2} &= \frac{\alpha^2\beta^2}{\alpha\beta^2} = \alpha. \end{aligned}$$

Lo stimatore del metodo dei momenti è pertanto

$$\begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{pmatrix} = \begin{pmatrix} \frac{m_1^2}{m_2 - m_1^2} \\ \frac{m_2 - m_1^2}{m_1} \end{pmatrix}$$

con matrice Jacobiana

$$\begin{aligned} Jg(\mu_1(\alpha, \beta), \mu_2(\alpha, \beta)) &= \begin{pmatrix} \frac{2\mu_1\mu_2}{(\mu_2 - \mu_1^2)^2} & -\frac{\mu_1^2}{(\mu_2 - \mu_1^2)^2} \\ -\frac{\mu_2}{\mu_1^2} - 1 & \frac{1}{\mu_1} \end{pmatrix} \\ &= \begin{pmatrix} \frac{2\alpha\beta(\alpha^2\beta^2 + \alpha\beta^2)}{(\alpha\beta^2)^2} & -\frac{(\alpha^2\beta^2 + \alpha\beta^2)}{\alpha^2\beta^2} - 1 \\ -\frac{\alpha^2\beta^2}{(\alpha\beta^2)^2} & \frac{1}{\alpha\beta} \end{pmatrix} \\ &= \begin{pmatrix} \frac{2(\alpha+1)}{\beta} & -2 - \frac{1}{\alpha} \\ -\frac{1}{\beta^2} & \frac{1}{\alpha\beta} \end{pmatrix} \end{aligned}$$

(Nota: il testo originale presenta  $\frac{m_1^2}{m_2 - m_1^2}$  e una diversa Jacobiana  $\begin{pmatrix} \dots & -\frac{\mu_2}{\mu_1^2} - 1 \\ \dots & \dots \end{pmatrix}$ . Questa trascrizione segue fedelmente il testo del PDF  $Jg(\mu_1(\alpha, \beta), \mu_2(\alpha, \beta)) = \dots$ )

### Spiegazione Semplice

#### Esempio Pratico: Legge Gamma

Qui applichiamo la "ricetta" del MoM a una distribuzione (la Gamma) che ha 2 parametri,  $\alpha$  e  $\beta$ .

1. **Momenti Teorici (Servono 2 equazioni per 2 parametri):**  $\mu_1 = \mathbb{E}[X] = \alpha\beta$   $\mu_2 = \mathbb{E}[X^2] = \alpha\beta^2(\alpha + 1)$
2. **Momenti Empirici:**  $m_1 = \bar{X}$  (la media campionaria)  $m_2 = \frac{1}{n} \sum X_i^2$  (la media campionaria dei quadrati)
3. **Eguagliare e Risolvere:** Il testo mostra la soluzione:  $\beta = \frac{\mu_2 - \mu_1^2}{\mu_1}$  (che è  $Var[X]/E[X]$ ) e  $\alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$ .
4. **Lo Stimatore Finale (la funzione  $g$ ):** Sostituiamo i momenti teorici ( $\mu_k$ ) con quelli empirici ( $m_k$ ) nelle formule della soluzione. Otteniamo così la "ricetta" per calcolare i parametri  $\tilde{\alpha}$  e  $\tilde{\beta}$  direttamente dai dati.

La "Matrice Jacobiana"  $Jg$  è la derivata (multivariata) di queste formule di soluzione. Serve per applicare la Proposizione 89 e trovare la varianza (la precisione) dei nostri estimatori  $\tilde{\alpha}$  e  $\tilde{\beta}$ .

**Example 91.** Supponiamo che in una certa città avvengano una serie di reati; poiché però non tutti i reati vengono denunciati, la legge del numero dei crimini osservati può essere vista come una binomiale con  $k$  e  $p$  incogniti, cioè il nostro campione aleatorio  $X_1, \dots, X_n$  ha legge

$$\binom{k}{x} p^x (1-p)^{k-x},$$

con  $k, p$  parametri incogniti da stimare. Abbiamo

$$\mu_1 = kp, \mu_2 = k^2 p^2 + kp(1-p),$$

e lo stimatore col metodo dei momenti si ottiene invertendo le relazioni

$$\bar{X}_n = kp$$

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = kp(1-p) + k^2 p^2.$$

Si può ottenere

$$\begin{aligned} \tilde{p}_n &= \frac{\bar{X}_n}{\tilde{k}_n} \\ \tilde{k}_n &= \frac{\bar{X}_n^2}{\bar{X}_n - \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \end{aligned}$$

Il metodo dei momenti fu introdotto da Pearson (1857-1936), matematico ed avvocato, che cambiò il nome da Carl in Karl per la sua ammirazione nei riguardi di Karl Marx. E' da notare che non richiede la conoscenza completa della legge di probabilità delle  $X_i$  a differenza del metodo della massima verosimiglianza che discutiamo nella prossima sezione.

### Spiegazione Semplificata

#### Vantaggi e Svantaggi del Metodo dei Momenti

Questo remark finale è la conclusione perfetta.

- **Vantaggio (Grande):** È facile e "robusto". Come dice il testo, **non richiede la conoscenza completa della legge di probabilità**. Basta conoscere le formule dei primi  $k$  momenti. Spesso è algebricamente semplice.
- **Svantaggio (che si vedrà dopo):** Non è il metodo più *efficiente* (preciso). Il **Metodo della Massima Verosimiglianza** (Maximum Likelihood), che richiede la conoscenza dell'intera funzione di densità  $f(x)$ , è generalmente considerato migliore (ha una varianza più piccola), anche se è computazionalmente più difficile.