Introduzione all'inferenza statistica, II lezione

Carla Rampichini

Dipartimento di Statistica "Giuseppe Parenti" - Firenze - Italia

carla@ds.unifi.it - www.ds.unifi.it/rampi/

Dottorato in METODOLOGIA DELLE SCIENZE SOCIALI, settembre 2008 - p. 1/25

$T(\mathbf{X})$ è sufficiente? (2)

TEOREMA Se $p(\mathbf{x} | \theta)$ è la funzione di densità (*pdf*) o di massa di probabilità (*pmf*) congiunta di $\mathbf{X}$ e $q(T(\mathbf{X} | \theta))$ è la *pdf* o la *pmf* di $T(\mathbf{X})$, allora $T(\mathbf{X})$ è una **statistica sufficiente** per θ se, per ogni $\mathbf{x} \in \mathcal{X}$, il rapporto $p(\mathbf{x} | \theta)/q(T(\mathbf{X} | \theta))$ è **costante** al variare di θ .

Alcuni esempi di statistiche sufficienti:

- distribuzione binomiale: n. di *successi* $T(\mathbf{X}) = \sum_{i=1}^n X_i$
- distribuzione normale, σ^2 nota: *media campionaria*
 $T(\mathbf{X}) = (1/n) \sum_{i=1}^n X_i = \bar{x}$
- statistiche d'ordine: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

Dottorato in METODOLOGIA DELLE SCIENZE SOCIALI, settembre 2008 - p. 2/25

Come si trova una statistica sufficiente?

Il teorema precedente non aiuta! Bisognerebbe:

1. **indovinare** quale può essere $T(\mathbf{X})$
2. **verificare** che il rapporto tra le *pdf* o le *pmf* non dipende da θ

Halmos e Savage (1949) hanno proposto il seguente **teorema**, che consente di trovare una statistica sufficiente osservando la *pdf* o la *pmf* del campione.

TEOREMA di FATTORIZZAZIONE Sia $f(\mathbf{x} | \theta)$ la *pdf* o la *pmf* congiunta di un campione $\mathbf{X}$. Una statistica $T(\mathbf{X})$ è una statistica **sufficiente** per θ **se e solo se** esistono le funzioni $g(t | \theta)$ e $h(\mathbf{x})$ tali che, per ogni punto campionario $\mathbf{x} \in \mathcal{X}$ e ogni valore di $\theta \in \Theta$ vale:

$$f(\mathbf{x} | \theta) = g(T(\mathbf{x}) | \theta)h(\mathbf{x})$$

Esempi: distribuzione normale, distribuzione uniforme.

Dottorato in METODOLOGIA DELLE SCIENZE SOCIALI, settembre 2008 - p. 3/25

Teorema di fattorizzazione

Il *Teorema di fattorizzazione* richiede che la fattorizzazione $f(\mathbf{x} | \theta) = g(T(\mathbf{x}) | \theta)h(\mathbf{x})$ valga per ogni $\mathbf{x}$ e ogni θ .

► Se l'insieme di punti campionari $\mathbf{x}$ per i quali $f(\mathbf{x} | \theta)$ è positiva dipende da θ : attenzione che il prodotto tra g e h sia nullo dove è nulla $f(\mathbf{x} | \theta)$!!!

► *Esempio:* distribuzione uniforme discreta

Dottorato in METODOLOGIA DELLE SCIENZE SOCIALI, settembre 2008 - p. 4/25

Statistiche sufficienti multiple

La statistica sufficiente può essere un **vettore** di funzioni:

$$T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}), \dots, T_r(\mathbf{X}))$$

- ▶ Questa situazione si presenta di solito quando anche il parametro è un vettore $\theta = (\theta_1, \theta_2, \dots, \theta_s)$, spesso $r = s$.
- ▶ Ogni parte della $f(x | \theta)$ che contiene un elemento di θ va inclusa in $g(T(\mathbf{x}) | \theta)$.
- ▶ **Esempio:** statistica sufficiente per la distribuzione normale $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X})) = (\bar{X}, S^2)$
- ▶ Ricordare SEMPRE che la definizione di statistica sufficiente dipende dal **MODELLO!!!** Se si calcola $(\bar{X}, S^2)$ e si ignora il resto dei dati $\Rightarrow$ piena fiducia nel modello Gaussiano!

Famiglia esponenziale

Famiglie di distribuzioni per le quali la *pdf* o la *pmf* può assumere la forma:

$$f(x | \theta) = h(x)c(\theta)\exp\left(\sum_{j=1}^n w_j(\theta)t_j(x)\right)$$

dove:

- $h(x) > 0, c(\theta) > 0$
- $t_j(x)$ funzioni a valori reali che **non** dipendono da θ
- $w_j(\theta)$ funzioni a valori reali che **non** dipendono da x

Fanno parte della famiglia esponenziale le seguenti famiglie di distribuzioni:

- *continue:* normale, gamma, beta
- *discrete:* binomiale, poisson, binomiale negativa

Le distribuzioni appartenenti alla famiglia esponenziale hanno **proprietà statistiche** interessanti!

Famiglie di locazione e scala

Consideriamo tre tipi di famiglie di distribuzione continue che godono di proprietà matematiche e statistiche interessanti:

- famiglie di locazione
- famiglie di scala
- famiglie di locazione e scala

▶ Ognuna delle famiglie è costruita specificando un'unica *pdf* $f(x)$, detta **distribuzione standard** della famiglia, dalla quale è possibile generare tutte le altre distribuzioni appartenenti alla famiglia utilizzando una trasformazione opportuna.

TEOREMA Sia $f(x)$ una *pdf* qualunque, e $\mu, \sigma > 0$ due costanti qualunque. Allora la funzione

$$g(x | \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

è una *pdf*.

Famiglie di locazione

Definizione Sia $f(x)$ una *pdf* qualunque. Allora la famiglia di *pdf* $f(x - \mu)$, indicizzata dal parametro $-\infty < \mu < \infty$ è detta **famiglia di locazione** con *pdf* standard $f(x)$. μ è detto **parametro di locazione** della famiglia.

- L'effetto di introdurre il parametro μ è quello di traslare il grafico della distribuzione:
 $P(x_1 < X < x_2 | \mu = 0) = P(x_1 + \mu < X < x_2 + \mu | \mu)$
- le distribuzioni normale, di Cauchy e esponenziale doppia con varianza nota sono famiglie di locazione
- è possibile costruire famiglie di locazione a partire da qualunque *pdf* $f(x)$, introducendo un parametro di locazione
- se X è una v.c. con *pdf* $f(x - \mu)$, possiamo rappresentare X come $X = Z + \mu$, dove Z è una v.c. con *pdf* $f(x)$

Famiglie di scala

Definizione Sia $f(x)$ una pdf qualunque. Allora per ogni $\sigma > 0$, la famiglia di pdf $(1/\sigma)f(x/\sigma)$, indicizzata dal parametro σ è detta **famiglia di scala** con pdf standard $f(x)$. σ è detto **parametro di scala** della famiglia.

- L'effetto di introdurre il parametro σ è quello di *allargare* ($\sigma > 1$) o *restringere* ($\sigma < 1$) il grafico di $f(x)$.
- la distribuzione standard si ottiene ponendo $\sigma = 1$

Famiglie di locazione e scala

Definizione Sia $f(x)$ una pdf qualunque. Allora per ogni $-\infty < \mu < \infty$ e ogni $\sigma > 0$, la famiglia di pdf $(1/\sigma)f((x - \mu)/\sigma)$, indicizzata dai parametri (μ, σ) è detta **famiglia di locazione-scala** con pdf standard $f(x)$. μ è detto **parametro di locazione** e σ è detto **parametro di scala** della famiglia.

- L'effetto di introdurre i parametri σ e μ è quello di traslare e di *allargare* ($\sigma > 1$) o *restringere* ($\sigma < 1$) il grafico di $f(x)$.
- la distribuzione standard si ottiene ponendo $\mu = 0$ e $\sigma = 1$

Teorema Sia $f(\cdot)$ una pdf qualunque. Sia μ un numero reale e $\sigma > 0$. Allora X è una v.c. con pdf $(1/\sigma)f((x - \mu)/\sigma)$ **sse** esiste una v.c. Z con pdf $f(z)$ e $X = Z\sigma + \mu$.

$T(\mathbf{X})$ sufficiente e famiglia esponenziale

TEOREMA Siano $X_1, X_2, \dots, X_n$ osservazioni i.i.d. da una pdf o pmf $f(x | \theta)$ che appartiene alla **famiglia esponenziale**:

$$f(x | \theta) = h(x)c(\theta)\exp(\sum_{j=1}^k w_j(\theta)t_j(x))$$

con $\theta = (\theta_1, \theta_2, \dots, \theta_d)$, $d \leq k$.

Allora la statistica:

$$T(\mathbf{X}) = (\sum_{i=1}^n t_1(x_i), \sum_{i=1}^n t_2(x_i), \dots, \sum_{i=1}^n t_k(x_i))$$

è una **statistica sufficiente** per θ .

Esempi per distribuzione:

1. binomiale: $\theta = p$, $0 < p < 1$
2. normale: $\theta = (\mu, \sigma)$, $-\infty < \mu < \infty$, $0 < \sigma < \infty$

Statistica sufficiente minimale

Per ogni problema di riduzione dei dati possiamo trovare **più di una** statistica sufficiente!

- ▼ Il campione completo $\mathbf{X}$ è **SEMPRE** una statistica sufficiente!
- ▼ Ogni funzione biunivoca r di una statistica sufficiente $T(\mathbf{X})$ è ancora una statistica sufficiente: $T(\mathbf{X})^* = r(T(\mathbf{X}))$.

★ Qual è la statistica sufficiente **migliore**?

↔ Quella che sintetizza maggiormente i dati, conservando tutta l'informazione su θ contenuta nel campione!!

DEFINIZIONE Una statistica sufficiente $T(\mathbf{X})$ è detta **MINIMALE** se, per ogni altra statistica sufficiente $T'(\mathbf{X})$, $T(\mathbf{X})$ è funzione di $T'(\mathbf{X})$.

Statistica sufficiente minimale (2)

In base alla definizione data $T(\mathbf{X})$ è sufficiente minimale se, per ogni altra statistica sufficiente $T'(\mathbf{X})$:

▼ $T'(\mathbf{X}) = T'(\mathbf{Y}) \Rightarrow T(\mathbf{X}) = T(\mathbf{Y})$.

▼ in altri termini, se $\{B_{t'} : t' \in T'\}$ sono gli insiemi della partizione di $\mathcal{X}$ generata da $T'(\mathbf{X})$ e $\{A_t : t \in T\}$ sono gli insiemi della partizione di $\mathcal{X}$ generata da $T(\mathbf{X})$, ogni $B_{t'}$ è un sottoinsieme di qualche A_t .

◆ La definizione di statistica sufficiente minimale non è pratica!!!

Bisogna infatti:

1. indovinare che $T(\mathbf{X})$ è sufficiente minimale
2. verificare la condizione data nella definizione

Statistica sufficiente minimale (3)

Il seguente TEOREMA di Lehmann e Scheffé (1950) fornisce un modo *più pratico* per trovare la **statistica sufficiente minimale**:

TEOREMA Sia $f(\mathbf{X} | \theta)$ la *pmf* o la *pdf* di un campione $\mathbf{X}$. Supponiamo che esista una funzione $T(\mathbf{X})$ tale che per ogni coppia di punti campionari $\mathbf{x}$ e $\mathbf{y}$ il rapporto $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$ sia costante rispetto a θ sse $T(\mathbf{x}) = T(\mathbf{y})$, allora $T(\mathbf{X})$ è una statistica **sufficiente minimale**.

Esempio: distribuzione normale $(\bar{X}, S^2)$.

◆ La statistica sufficiente minimale **non è unica!**

Qualunque funzione biunivoca di una statistica sufficiente minimale è ancora sufficiente minimale!!!

Esempio: distribuzione normale $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$.

Statistica ancillare

Definizione Una statistica $S(\mathbf{X})$ la cui distribuzione non dipende da θ è detta **statistica ancillare**.

► Una statistica ancillare, da sola, non contiene alcuna informazione su θ , può contenere informazione su θ se utilizzata **insieme** ad altre statistiche.

► *Esempi:*

- La statistica $R = X_{(n)} - X_{(1)}$ è una statistica ancillare per una famiglia di locazione con *cdf* $F(x - \theta)$, $-\infty < \theta < \infty$.
- Ogni statistica che dipende dagli $n - 1$ valori campionati $X_1/X_n, \dots, X_{n-1}/X_n$ è una statistica ancillare per una famiglia di scala con *cdf* $F(x/\sigma)$, $\sigma > 0$.

Sufficienza, ancillarità e statistiche complete

Definizione: Sia $f(\mathbf{X} | \theta)$ la *pmf* o la *pdf* di un campione $\mathbf{X}$ e $T(\mathbf{X})$ uno stimatore sufficiente per θ . $T(\mathbf{X})$ è uno stimatore **completo** se, per qualsiasi funzione $q(T(\mathbf{X}))$ tale che $E[q(T(\mathbf{X}))] = 0, \forall \theta$ vale $q(T(\mathbf{X})) = 0$, cioè uno stimatore è completo se l'unica funzione di $T(\mathbf{X})$ il cui valor medio è zero è la funzione identicamente nulla.

- La completezza è una proprietà di una famiglia di distribuzioni
- Per famiglie complete alcune procedure inferenziali sono uniche
- Il legame tra completezza, ancillarità e sufficienza è dato dal Teorema di Basu (1955)

Teorema di Basu

Teorema di Basu: Se $T(\mathbf{X})$ è una statistica sufficiente completa ed è sufficiente minimale per θ , allora $T(\mathbf{X})$ è indipendente da ogni statistica ancillare.

► Il teorema di Basu è utile perchè consente di dedurre l'indipendenza tra due statistiche senza bisogno di ricavarne la distribuzione congiunta

► Vale inoltre la seguente importante proprietà delle statistiche complete:

Teorema: Uno stimatore sufficiente completo è sempre minimale.

Utilizzo del teorema di Basu

- per utilizzare il teorema di Basu è necessario provare la completezza della statistica, il che non è facile
- per fortuna molti dei problemi che si incontrano soddisfano le seguenti

Teorema: Siano $X_1, \dots, X_n$ v.c. da una famiglia esponenziale con pdf o pmf della forma

$$f(x | \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp \left(\sum_{j=1}^k w_j(\boldsymbol{\theta})t_j(x) \right)$$

con $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$,

allora la statistica: $T(\mathbf{X}) = (\sum_{i=1}^n t_1(X_i), \dots, \sum_{i=1}^n t_k(X_i))$

è completa se $\{(w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta})) : \boldsymbol{\theta} \in \Theta\}$ contiene un insieme aperto di $\mathcal{R}^k$.