Appunti del corso di Statistica Matematica

Gadotti Andrea, Nardin Michele, Peruzzetto Marco

cba

Quest’opera `e distribuita con Licenza Creative Commons Attribuzione - Con- dividi allo stesso modo 3.0 Italia. Per leggere una copia della licenza visita il sito web http://creativecommons.org/ licenses/by-sa/3.0/it/ o spedisci una lettera a Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Indice

1 Prima parte del corso 2 1.1 Introduzione. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2 1.1.1 Funzione generatrice dei momenti . . . . . . . . . . . . . . . . . . . . 2 1.1.2 Famiglia Esponenziale a k parametri . . . . . . . . . . . . . . . . . . 4 1.1.3 Trasformazioni di variabili casuali . . . . . . . . . . . . . . . . . . . 5 1.1.4 Convergenze . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6 1.1.5 Teoria asintotica . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7 1.2 Approccio applicativo alla Statistica Matematica . . . . . . . . . . . . . . . 10 1.2.1 Statistiche d’ordine . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10 1.2.2 Intervalli di confidenza . . . . . . . . . . . . . . . . . . . . . . . . . . 17 1.2.3 Test per la verifica di ipotesi . . . . . . . . . . . . . . . . . . . . . . 24 1.2.4 Esempi di statistiche test (generali e particolari) . . . . . . . . . . . 28

2 Seconda parte del corso 34 2.1 Efficienza . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38 2.1.1 Teorema di Rao-Cramér . . . . . . . . . . . . . . . . . . . . . . . . . 39 2.1.2 Estensione a un vettore di parametri: . . . . . . . . . . . . . . . . . 43 2.2 Sufficienza . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48 2.2.1 Statistiche sufficienti minimali . . . . . . . . . . . . . . . . . . . . . 54 2.2.2 Principio di verosimiglianza . . . . . . . . . . . . . . . . . . . . . . . 55 2.2.3 Famiglie esponenziali e sufficienza . . . . . . . . . . . . . . . . . . . 57 2.2.4 Teorema di Rao-Blackwell . . . . . . . . . . . . . . . . . . . . . . . . 57 2.3 Completezza . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60 2.3.1 Teorema di Lehmann-Scheffé . . . . . . . . . . . . . . . . . . . . . . 60 2.4 Propriet`a degli stimatori di massima verosimiglianza . . . . . . . . . . . . . 62 2.5 Metodi Numerici per la Stima di Massima Verosimiglianza . . . . . . . . . . 65 2.6 Teoria dei Test pi`u Potenti ed Uniformemente pi`u Potenti . . . . . . . . . . 66 2.6.1 Rapporto di massime verosimiglianze . . . . . . . . . . . . . . . . . . 71

1

Capitolo 1

Prima parte del corso

1.1 Introduzione

In questa prima sezione vengono presentati i richiami di teoria della probabilit`a, affrontati nelle primissime lezioni del corso.

1.1.1 Funzione generatrice dei momenti

Lezione del 18/02, ultima modifica 09/04, Andrea Gadotti

Definizione 1. Sia X una variabile casuale (discreta o assolutamente continua). Se esiste t0 > 0 tale per cui E(etX) < +∞ ∀t ∈ (−t0,t0), chiameremo la funzione

MX := E(etX) funzione generatrice dei Momenti di X.

Esempi

1. X ∼ b(1,p) con p ∈ (0,1). Si ha:

MX(t) = E(etX) =

∑1x=0etxP(X = x)

=

∑1x=0etxpx(1 − p)1−x = pet + (1 − p)

2. X ∼ P(λ) con λ > 0. Si ha:

MX(t) = E(etX) =

∑+∞x=0etxe−λx! λx

= eλ(et−1)

= eλ(et−1)

3. X ∼ G(α, β), ovvero

fX(x;α, β) := Γ(α)β1

αxα−1e− β 1x

2

con α > 0, β > 0, x > 0 e

Γ(α) :=

∫ +∞

0 xα−1e−xdx

(nota: α ∈ N =⇒ Γ(α)=(α − 1)!)

Abbiamo che:

MX(t) = E(etX) =

∫ +∞

0 etx Γ(α)β1

αxα−1e− β 1xdx

= ...[sostituzione σ := x( β 1− t)]...

= (1 − 1

βt)α

con t < 1β

Momenti di una variabile casuale

Definizione 2. Se una variabile casuale ammette FGM derivabile infinite volte in un intorno di t = 0 e se tutti i suoi momenti sono finiti, allora definiamo il momento di ordine s non centrato:

μ s := E(Xs) = dtdssMX(t)|t=0 Il momento di ordine s centrato in a ∈ R `e:

μs(a) := E((X − a)s)

Ovvero μ s centrato (senza = μs(0). specificare E’ chiaro altro, che intenderemo μ 1 = E(X). centrato Chiameremo in μ 1):

infine momento di ordine s

μs := E((X − μ 1)s)

Teorema 1. Dall’ultima definizione si vede facilmente (sviluppando l’elevamento a poten- za) che vale la seguente relazione tra momenti centrati e non:

μs = E((X − μ 1)s) =

∑s(−1)mm=0( ms)μ s−m(μ 1)m

Osserviamo che μ2 = E((X − μ 1)2) = V ar(X) = E(X2) − (E(X))2 = μ 2 − (μ 1)2

Teorema 2. Date due (o pi`u) v.c. X e Y indipendenti aventi f densit`a / f massa fX e fY e fgm MX(t) e MY (t) rispettivamente, allora si ha

MX+Y (t) = MX(t)MY (t)

3

Teorema 3. Siano X e Y v.c. con funzioni di ripartizione FX(x) e FY (y) rispettivamente. Siano MX(t) e MY (t) le fgm di X e Y . Se MX(t) = MY (t) per ogni t in un intorno dell’origine, allora

X = dY

Osservazione 1. Il teorema appena visto ci dice sostanzialmente che, se esiste, la fgm caratterizza la distribuzione della corrispondente v.c.

Esempio Siano (X1, ..., Xn) mico (Xi ∼ b(1,p)). Vogliamo risultati della replicazione di un esperimento trovare la distribuzione di Sn :=

∑ncasuale dicoto-

i=1

Xi. Calcoliamo quindi

la sua fgm:

MSn(t) = E(etSn) = E(et∑ni=1 Xi)

TEO2 =

∏ni=1E(etXi) =

∏ni=1MXi(t) =

∏n(pet + (1 − p)) = (pet + (1 − p))n i=1ovvero Sn `e distribuita come b(n, p) per il Teorema 2. Esercizio Ripetere il calcolo precedente supponendo Xi ∼ P(λ),∀i.

1.1.2 Famiglia Esponenziale a k parametri

Una famiglia di f densit`a / f massa `e detta essere una Famiglia Esponenziale a k parametri θ1, ..., θk se la corrispondente f densit`a / f massa (che `e indicizzata da θ1, ..., θk) pu`o essere scritta come

fX(x;θ) = C∗(x)D∗(θ) exp{

∑km=1Am(θ)Bm(x)} dove C∗(x) `e una funzione della sola x, D∗(θ) `e una funzione del solo θ, Am(θ) `e una funzione del solo θ e Bm(x) `e una funzione della sola x. Esempi

1. X ∼ G(α, β) =⇒ fX(x;α, β) = 1 Γ(α)βαxα−1e− β 1x1R+(x), α > 0, β > 0 1R+ `e detto

supporto della distribuzione. Quindi possiamo riscrivere fX(x;α, β) come

fX(x;α, β) = Γ(α)β1

α1R+(x)exp((α − 1)ln(x) − β1x)

e quindi ponendo D∗(α, β) := Γ(α)β1 α, C∗(x) := 1R+(x), A1(α, β) := (α−1), B1(x) := ln(x), A2(α, con k = 2.

β) := − β 1e B2(x) := x, otteniamo G(α, β) come famiglia esponenziale

2. X ∼ b(n, p) =⇒ fX(x;n, p) = (nx)px(1 − p)n−x1{0,1,...,n}(x) con n ∈ N noto. Quindi

possiamo riscrivere fX(x;n, p) come

fX(x;n, p) =

(nx)1{0,1,...,n}(x)(1 − p)nexp(xln( 1 − p

p))

4

con p1−p Quindi detto odd ratio ponendo D∗(p) o parametro naturale := (1 − p)n, C∗(x) := della (nxfamiglia esponenziale. )1{0,1,...,n}(x), A1(p) := B1(x) := x, otteniamo b(n, p) come famiglia esponenziale con k = 1.

ln( 1−pp),

3. X vc con fX(x, θ) = esponenziale. Il fatto e1−x/θ

che θ 1(θ,∞)(x): la distribuzione di X non appartiene a famiglia il supporto di fX dipenda dal parametro θ NON permette a fX di appartenere ud una famiglia esponenziale!

Osservazione 2. Le famiglie di esponenziali hanno interessanti propriet`a matematiche (propriet`a di regolarit`a).

Dal punto di vista statistico, ci`o si traduce in un’interessante conseguenza: tutta l’in- formazione contenuta nei dati a disposizione (X1, ..., Xn) relativa alla funzione fX(x;θ) pu`o essere sintetizzata attraverso k quantit`a (funzioni di (X1, ..., Xn)) che potranno essere impiegate per costruire procedure inferenziali (stima, test per la verifica di ipotesi) riguar- danti il parametro θ. Ovvero, l’appartenenza ad una famiglia esponenziale permette una riduzione dei dati (X1, ... ,Xn) via Bm.

1.1.3 Trasformazioni di variabili casuali

Lezione del 01/03, ultima modifica 09/04, Michele Nardin

Discrete

Teorema 4. Sia X una vc con funzione di massa fX(x) = P(X = x), e sia AX il suo supporto. Sia W=h(X) una nuova vc. Allora

P(W = w) = ∑

{x∈AX:h(x)=w}P(X = x)

Esempi

1. Sia X ∼ b(n, p) con relativa funzione di massa fX(x, p) = (nx)px(1−p)n−x10,1,...,n(x),

n noto e p ∈ (0,1).

Considero quindi W = n − X. Come si distribuisce W?

P(W = w) = P(X = n − w) =

( n

n − w)pn−w(1 − p)w10,1,...,n(w)

2. Sia X una vc tale che fX(x) = P(X = x) = (12)x1N(x), W = X3.

P(W = w) = P(X3 = w) = P(X = √3w) =

(12) √3w

11,8,27,64,...(w)

5

Assolutamente continue

Teorema 5. Sia X una variabile casuale (ass continua) con funzione di densit`a fX(x) e sia W = h(X), ove h `e una funzione monotona. Supponiamo inoltre che fX(x) sia continua sul supporto di X e che h−1(w) abbia derivata continua sul supporto di W. Allora fW(w) = fX(h−1(w))∣∣∣∣ ddwh−1(w)∣∣∣∣1AW(w)

Esempio (Standardizzazione di una vc normale) Sia X ∼ N(m, s2). Considero W = h(X) = X−m s . Allora, dato che h−1(w) = sw + m, che ha derivata continua su tutto R,

fW(w) = fX(sw + m)|s| = e

√−w2π 22

= fN(0,1)

Teorema 6. Se W = h(X) ove h `e monotona a tratti (un numero di tratti finito k) e valgono le condizioni del teorema precedente (su ogni tratto), allora

fW(w) =

∑kn=1fX(h−1 n (w))∣∣∣∣ dwdh−1 n (w)∣∣∣∣1AW(w)

Esempio (Chi-quadro) Sia X ∼ N(0,1) e W = h(X) = X2. h `e monotona sui tratti A0 = 0, A1 = (−∞,0), A2 = (0,+∞). Considero Trovo Adwd2∀w h−1 1 che ≥ 0).

hh−1 1 1(x) (w) = = x−2 √per w x < 0 mentre (NB:h−1 1 (w) ∈ hA2(x) 1∀w = ≥ x2 0), per mentre x > 0. h−1 2 (w) = √w (NB:h−1 2 (w) = − 2√1

w, dwdh−1 2 (w) = 2√1

w sono entrambe continue su R+.

(w) ∈

fW(w) = √12πe

∣∣∣∣ 2√1

w∣∣∣∣ + √12πe

∣∣∣∣ 2√1

w∣∣∣∣

= √2πw1 e

−(−√w)2 2

−(√w)2 2

−w2 1R+(w) = 21/2Γ(1/2)1

w

12−1e

−w2

Si riconosce che W ∼ G(α = 1/2,β = 2) e si chiama Chi quadrato con ν = 1 gradi di libert`a. In generale, una vc Chi Quadro con ν X1,X2, ..., Xn sono vc iid N(0,1). Per il Teorema = n gradi di libert`a `e W = 2 sulla FGM di una somma ∑ni=1 di Xvc i 2, iid ove si trova immediatamente che W ∼ G(α = n · 1/2,β = 2).

1.1.4 Convergenze

Convergenza in probabilit`a

Definizione 3. Sia {Xn}n∈N una successione di variabili casuali e sia X un’altra varia- bile casuale, tutte definite probabilit`a a X (scriviamo sullo Xn → pstesso X) se spazio ∀ε > 0

campionario. Diciamo che Xn converge in

n→∞lim P(|Xn − X| ≥ ε)=0

6

Osservazione 3. Se Xn → pX diciamo che la massa della differenza |Xn − X| converge a 0. Inoltre, quando scriviamo Xn → pX, stiamo sottintendendo tutta la parte iniziale della definizione precendete, cio`e il sia {Xn}n∈N una successione di variabili casuali....

Teorema 7. Alcuni risultati utili:

1. Supponiamo che Xn → pX e Yn → pY . Allora Xn + Yn → pX + Y 2. Supponiamo che Xn → pX e sia a una costante. Allora aXn → paX

3. Supponiamo che Xn g(Xn) → pg(a)

→ pa costante, e sia g una funzione reale continua in a. Allora

4. (Corollario di 3.) Se Xn → pa, allora Xn 2(a ≥ 0).

→ p1a (se a = 0), √Xn → p√a

5. Xn → pX e Yn → pY allora XnYn → pXY

Convergenza in distribuzione

→ pa2, 1Xn

→ pa2, 1Xn

Definizione 4. Sia {Xn}n∈N una successione di variabili casuali e sia X un’altra variabile casuale, tutte definite sullo stesso spazio campionario.

Siano FXn e FX le relative funzioni di ripartizione (dette anche ”di distribuzione”). Sia C(FX) l’insieme dei punti ove FX `e continua. Diciamo che Xn converge in distribuzione (o in legge) a X (scriviamo Xn → dX) se

n→∞lim FXn(x) = FX(x) ∀x ∈ C(FX)

Teorema 8. Se Xn → pX allora Xn → dX.

Osservazione 4. Il contrario in generale non vale, tranne nel caso in cui X `e una vc degenere (cio`e costante).

Teorema 9. Supponiamo che Xn → dX e sia g una funzione continua sul supporto di X. Allora g(Xn) → dg(X)

Teorema 10 (Slutsky). Supponiamo che Xn → dX, An → pa costante e Bn → pb costante. Allora An + BnXn → da + bX

1.1.5 Teoria asintotica

Lezione del 04/03, ultima modifica 09/04, Michele Nardin

Teorema √n(Xn − θ) 11. → d(∆-method) N(0,σ2). Sia {Xn}n∈ N una successione di vc tale che Supponiamo che una funzione g(X) sia derivabile in θ e che g (θ) = 0. Allora √n(g(Xn) − g(θ)) → dN(0,σ2(g (θ))2)

7

Esempio Considero

Yn = χ√2n − 2n n = √n( √χ2n 2n− √12) ove χ2n `e la chiquadro con n gradi di libert`a. Ricordiamo (discende dal fatto che χ2n ∼ G(α = n/2,β = 2). Affermiamo che E(χche 2n) Y= n n → de N(0,1). che V ar(χInfatti:

2n)=2n

Yn = χ√2n 2n − n =

∑ni=1 √Xni 2√− 2 n · 1 dove Xi ∼ N(0,1), Quindi per il Teorema Scrivendo ora Yn e nella quindi centrale forma XYi 2n del ∼ = χ√Limite 21, nquindi ( √χ2n 2n(vedi − le √X1sotto) 2i 2hanno media μ = 1 e varianza σ2 = 2.

si ha quanto voluto. ) ipotesi θ = 1/√del 2, g ∆-method (t) = sono soddisfatte. 2√1t|θ=1/√2 = 2−3/4. riconosciamo Considero quindi g(t) = che √t, la che prima `e derivabile parte delle in Allora

√n(g

( √χ2n2n)

− g(θ)) = √n(√

√χ2n 2n −

√

√12)

→ dN(0,12 · 2−3/2)

Teorema 12. (Teorema centrale del limite) Siano X1, ...Xn vc iid dotate di media μ e varianza finita σ2. Allora∑ni=1 √n X· i σ − nμ =

√n(Xσ

n − μ)

→ dN(0,1)

con Xn media aritmetica delle Xi.

Esempi/Applicazioni

1. XQuando n ∼ b(n, scriviamo p), Xn ∼ a∼ aN(np, stiamo np(1−p)) considerando (ricordiamo un che Xn ∼ ∑ni=1 bi, ove bi ∼ b(1,p)). andamento asintotico, ossia sottinten- diamo un’approssimazione (via via migliore con l’aumentare di n) giustificata dal TLC (il senso `e che per n ’grandi’ la distribuzione ’funziona circa cos`ı’).

2. XY ̄n 1, := ..., YnXn

n ∼ avc N(1,1/n) P(λ = 1). Considero Yn = ∑Xi. Dato che Yn ∼ aN(nλ, nλ) e λ = 1, Considero quindi Wn = √n(Yn n− 1) = 1/Ynn √−1

n = √Y ̄n−E(V ar(YYn) ̄ ̄n) , trovo che Wn ∼ aN(0,1)

Teorema 13. Sia {Xn} una succ di vc iid, ognuna con con FGM MXn(t) definita e < ∞ per t ∈ (−h, h), e sia X un’altra vc con FGM MX(t) definita e < ∞ per t ∈ (−h1,h1), h1 ≤ h. Se

n→+∞lim MXn(t) = MX(t) ∀|t| ≤ h1

allora Xn → dX.

Applicazione

8

1. Sia Xn ∼ b(n, p). Ricordiamo che Xn = ∑Xi ove Xi ∼ b(1,p), ed inoltre μ =

E(X) = np. Siccome MXn(t) = E(etXn) = [(1 − p) + pet]n = [1 + μn(et − 1)]n,

MXn(t) n→∞ −→ eμ(et−1)

che `e la FGM di una Poisson di parametro μ, ovvero Xn d→ P(n, p).

9

1.2 Approccio applicativo alla Statistica Matematica

Questa sezione corrisponde alla parte di corso svolta dalla seconda settimana di marzo fino a met`a aprile, che riguarda gli aspetti pratici della statistica: verranno introdotte le statistiche d’ordine, gli intervalli di confidenza e i test per verifiche d’ipotesi.

Definizione 5. (Campione Casuale) Il vettore casuale (X1, ..., Xn) si dice Campione Casuale relativamente ad una vc X ∼ FX(x, θ) se i suoi elementi sono vc i.i.d.

Osservazione Il fatto che le vc siano i.i.d. implica che

FX1,...,Xn(X1, ..., Xn) =

∏ni=1FXi(Xi) e

fX1,...,Xn(X1, ..., Xn) =

∏ni=1fXi(Xi)

Definizione 6. (Statistica) Sia (X1, ..., Xn) un campione casuale da una distribuzione associata alla vc X, e sia Ω lo spazio campionario di (X1, ..., Xn) Ogni funzione

T(X1, ..., Xn):Ω −→ Rk

che NON dipende da parametri incogniti `e detta Statistica.

1.2.1 Statistiche d’ordine

Lezioni 08 e 11 Marzo, ultima modifica 21/03, Scritte da: Marco Peruzzetto

Definizione 7. Sia (X1,...,Xn) un campione casuale con distribuzione FX (x, θ), densit`a fX (x, θ) e supporto supp{X} := (a, b) ⊂ R ove X ∈ {X1,...,Xn} e −∞ ≤ a<b ≤ +∞. Definiamo ricorsivamente le seguenti variabili casuali:

• X(1) := min ({X1,...,Xn});

• X(i) := min({X1,...,Xn}\{X(1),...,X(i−1)}) ∀1 < i ≤ n.

Chiameremo allora X(i) la i-esima Statistica d’Ordine del campione.

Osservazione: La statistica d’ordine consiste semplicemente nel vettore per il quale le variabili casuali vengono appunto ordinate, in base al valore che assumono in un de- terminato punto del loro dominio comune, in ordine crescente. In particolare X(i) sar`a l’ i-esima max ({Xvariabile 1,...,Xn}). pi`u piccola. Osserviamo Naturalmente, che la funzione se il (Xcampione 1,...,Xha n) ↦− lunghezza → (X(1),...,Xn, allora (n)) X`e (n) essa =

stessa una Statistica.

Teorema 14. Sia (X1,...,Xn) un campione casuale come sopra. Allora si ottiene ∀ 1 ≤ m ≤ n che la densit`a dell’ m-esima statistica d’ordine `e data da:

fX(m) (x, θ) = (m − 1)!(n n!

− m)!fX (x, θ) · FX (x, θ)m−1 · (1 − FX (x, θ))n−m

10

Daremo due dimostrazioni, la seconda pi`u bella della prima.

Dimostrazione. (a cura di Marco Perruzzetto) Innanzi tutto si ha che il supporto (a, b) pu`o essere partizionato in n parti, per cui evidentemente si ha:

f(X(1),...,X(n))(x(1),...,x(n),θ) =

{ n! 0 ∏ni=1 fX(x(i),θ) se a<x(1) < x(2) <...<x(n) < b altrimenti.

ove la produttoria `e giustificata dal fatto che le variabili sono tutte indipendenti e che de- vono essere ciascuna minore dell’altra per l’ordinamento assegnato; il coefficiente fattoriale `e presente poiché le n parti dell’intervallo (a, b) possono essere assegnate alle n variabili in tale numero di modi, dato che ciascuna Adesso per trovare la distribuzione di fcon (X(1)i ,...,X= (n)m. ) necessariamente bligatoriamente scrivere la distribuzione funzione di distribuzione, Xciascuna i nei domini possibili di tutte le ∀1 ≤ i ≤ n ha la stessa distribuzione.

altre funzioni X(m) sar`a di distribuzione dunque sufficiente di ciascuna integrare X(i) In particolare, inferiori a superiori a come otterremo fquelli Xciascuna (m)∫ a x, fviceversa di X(θ, quindi fXft)dt (m) X(i) in che = ogni per ogni F∀a<xX(θ, fi Xpunto. < (i) x) (m) per m essendo dovr`a < i>m Ricordando b la assumere a piacere valori dovr`a assumere valori ob- allora che possiamo la densit`a la derivata della distribuzione sar`a data da:

fX(m)(x(m),θ) =

=

∫ x(2)

a dx(1) ···∫ x(m)

a dx(m−1)

∫ bx(m)

dx(m+1) ···∫ bx(n−1)

dx(n)f(X(1),...,X(n))(x(1),...,x(n),θ)

=

∫ x(2)

a dx(1) ···∫ x(m)

a dx(m−1)

∫ bx(m) dx(m+1) ···∫ bx(n−1) dx(n)n!

∏ni=1fX(x(i),θ) = fX(x(m))∫ x(2)

a dx(1) ···∫ x(m)

a dx(m−1)

∫ bx(m)

dx(m+1) ···∫ bx(n−1)

∏ ni=1,i =mfX(x(i),θ) = n!

(m − 1)!(n − m)!fX (x, θ) · FX (x, θ)m−1 · (1 − FX (x, θ))n−m,

dove `e stato usato il fatto che ∫ a bF Xα(θ, t)fX(θ, t)dt = F α+1 Xα+1

, ∀α = −1.

Dimostrazione. Sia Ω il dominio comune del campione casuale. Definiamo per x ∈ R la nuova variabile casuale Yx come:

Ω −→ {0,...,n}

Yx(ω) :=

dx(n)n!

∑ni=1

1{Xi(ω)≤x}(ω)=#{i ∈ {1,...,n} : Xi ≤ x},

funzione che, per cos`ı dire, “conta” il numero di variabili casuali Xi che non superano x. Si vede immediatamente che ∀1 ≤ m ≤ n, si ha la distribuzione

FX(m)(θ, x) = P[X(m) ≤ x] = P[almeno X(1), ..., X(m) stanno sotto x] =

= P[Yx ≥ m] =

∑nk=mP[Yx = k] =

=

∑nk=m(nk)F Xk(θ, x)(1 − FX(θ, x))n−k.

11

Come nella prima dimostrazione usiamo il fatto che la densit`a si pu`o vedere come derivata della funzione di ripartizione. Ne segue che per calcolare la densit`a sar`a sufficiente calco- lare la derivata in ciascun punto x della distribuzione appena trovata. In particolare si potr`a vedere che coesisteranno il termine che vogliamo ottenere con altre due sommatorie, che tuttavia si elidono l’una con l’altra lasciando quindi la relazione espressa dal teorema. Si ha infatti che:

f(m)(θ, x) = ∂x∂FX(m)(θ, x) =

=

(nk)

· fX(θ, x){kF X k−1 (θ, x)(1 − FX(θ, x))n−k − (n − k)F Xk(θ, x)(1 − FX(θ, x))n−k−1} = m(mn)

· fX(θ, x)F X m−1 (θ, x)(1 − FX(θ, x))n−m+

∑ nk=m+1k(nk∑nk=m

)fX(θ, x)F X k−1 (θ, x)(1 − FX(θ, x))n−k−

k=mn−1∑(n − k)(nk)fX(θ, x)F Xk(θ, x)(1 − FX(θ, x))n−k−1 = (m − 1)!(n n!

− k)! · fX(θ, x)F X m−1 (θ, x)(1 − FX(θ, x))n−m+ j=mn−1∑(j + 1)( j + n

1)fX(θ, x)F Xj(θ, x)(1 − FX(θ, x))n−j−1− k=mn−1∑(n − k)(nk)fX(θ, x)F Xk(θ, x)(1 − FX(θ, x))n−k−1 = (m − 1)!(n n!

− k)! · fX(θ, x)F X m−1 (θ, x)(1 − FX(θ, x))n−m+ n−1∑j=m

j!(n − n!

j − 1)!fX(θ, x)F Xj(θ, x)(1 − FX(θ, x))n−j−1−

n−1∑k=m

k!(n − n!

k − 1)!fX(θ, x)F Xk(θ, x)(1 − FX(θ, x))n−k−1

= (m − 1)!(n n!

− k)! · fX(θ, x)F X m−1 (θ, x)(1 − FX(θ, x))n−m.

Definizione. Sia (X(1),...,X(n)) una statistica d’ordine di un campione casuale. Allora possiamo definire le nuove seguenti variabili:

• X(n) − X(1), detta Range oppure Misura di Dispersione;

• •

X(1)+X2 (n)

detta Mid Range oppure Misura di Centralit`a;

∀n pari ∀n dispari XX( ( n2 n+1

)2 +X2 )

( n2 +1)

}dette ciascuna Mediana campionaria;

• Sia 2(n+1) 1 <p< 1− 2(n+1)1 , che possiamo in ogni caso pensare molto grande. A questo punto possiamo definire l’intero kp := ⌊come p(n+ 0 1)<p< ⌋+⌊21 (p(n+ per n

12

1) − 2⌊p(n + 1)⌋)⌋, che risulta essere cos`ı ben definito in quanto compreso tra 1 e n e restituisce l’approssimazione all’intero pi`u vicino al variare di p del reale p(n + 1).

• A questo punto, se scegliamo ξp ∈ F −1X (p), chiameremo ξp Quantile di popolazione di ordine p. In seguito troveremo utile stimare tale valore. Perci`o introduciamo la variabile casuale ad esso collegata X(kp), detta Quantile campionario di ordine p. Se p = im, allora X(kp) `e detta anche i-esimo m-ile campionario. In particolare con Q1 e Q3 si indicano rispettivamente il primo e il terzo quartile. Intuitivamente, X(kp) mi d`a la v.c. che sta al kp-esimo posto, ovvero al p(n+1)-esimo posto (se p(n + 1) `e intero). Ad esempio, se p = 13, X(k 13 ) `e la v.c. che nel vettore ordinato sta alla posizione n+1 3 .

• Le variabili LF := Q1 − h e UF := h + Q3, ove h := 32(Q3 − Q1) sono dette rispettivamente Lower e Upper Fence.

Osservazione: Osserviamo che pi`u la misura di centralit`a si discosta dalla mediana, pi`u vi `e asimmetria nella funzione di densit`a f (i.e.: una funzione di distribuzione `e simmetrica :⇐⇒ ∃x0 ∈ R : f(x0 + x) = f(x0 − x),∀x ∈ Dom(f).) Inoltre, ponendo che la funzione di ripartizione sia iniettiva e la funzione di densit`a sia simmetrica, si vede immediatamente che la media di popolazione, ovvero il quantile di popolazione di ordine p = 12 coincide con il valore di aspettazione della variabile casuale, il quale a sua volta deve coincidere con x0.

Dato un campione casuale di parametro θ ∈ R fissato, sappiamo che una qualsiasi funzione di statistiche su tali variabili `e, proprio per definizione, uno stimatore del parametro θ. L’esistenza di un’infinit`a non numerabile di stimatori `e sicuramente un problema da ov- viare in merito alla scelta tra essi di uno stimatore che effettivamente permetta di stimare il pi`u correttamente possibile il parametro θ. Cercheremo dunque di individuare alcune propriet`a che possano effettivamente giustificare la scelta di un determinato stimatore, affinché esso risulti il pi`u possibile affidabile.

Definizione. Sia (X1 ...,Xn) un campione di parametro θ e Tn(X1,...,Xn) uno sti- matore. La funzione Bθ[Tn(X1,...,Xn)] := Eθ[Tn(X1,...,Xn)] − θ si dice distorsione di Tn (nota: con Eθ formalmente intendiamo semplicemente E). In particolare Tn si dir`a non distorto se e solo se la sua distorsione `e nulla ∀θ ∈ R (nel senso che uno stimatore -e.g. la media campionaria- non pu`o stimare bene solo alcune medie, ma qualsiasi media reale, ad esempio la media di qualsiasi normale centrata in qualsiasi punto). Altrimenti si dice distorto. Se infine si ottiene che limn→+∞ Bθ[Tn(X1,...,Xn)] = 0, Tn si dice asintotica- mente non distorto.

Esempio Sia (X(1),...,X(n)) un campione casuale con distribuzione simmetrica (sen- za perdita di generalit`a, la assumiamo simmetrica rispetto all’origine) e scegliamo come stimatore Tn proprio la mediana campionaria. `E chiaro innanzi tutto che essa in generale gode delle seguenti due propriet`a:

• ∀b ∈ R, Tn(X1 + b,...,Xn + b) = Tn(X1,...,Xn) + b;

• Tn(−X1,...,−Xn) = −Tn(X1,...,Xn).13

Abbiamo inoltre che la distribuzione di (X1,...,Xn) e del vettore (−X1,...,−Xn) coin- cidono (ricordando che l’origine `e il centro di simmetria). Si avr`a dunque:

E[Tn] = E[Tn(X1,...,Xn)] = E[Tn(−X1 ...,−Xn)] = E[−Tn(X1,...,Xn)] = −E[Tn(X1 ...,Xn)] = −E[Tn]

perci`o, in definitiva, 2E[Tn] = 0 ovvero E[Tn] = 0. Quindi, nel caso di una distribuzione simmetrica, la media campionaria `e uno stimatore non distorto del valore di aspettazione, del punto di simmetria e della media della popolazione (dato che tutti loro nel nostro caso coincidono).

Esempio Sia (X1,...,Xn) un campione casuale di parametro θ ∈ R fissato e con μ := E[X], σ2 := Var[X]. Vogliamo provare a calcolare la distorsione di due stimatori “classici”:

1. Scegliamo Eθ[Xn]−θ ∀n ∈ N. In come = particolare n

1∑stimatore ni=1 Eθ[X`e uno i]−θ la media stimatore = n1·nEcampionaria θ[X] non Xn := n

1∑= μ−θ. Dunque ni=1 la distorsione Xi. Allora `e Bcostante θ[Xn] =

distorto per il valore di aspettazione μ. Possiamo calcolare facilmente anche la varianza di Xn media campionaria si rivela essere quindi un buon stimatore. che (nota: risulta nel essere calcolo σn 2. della La

varianza stiamo trattando lo stimatore come una variabile casuale essa stessa, quindi pi`u la varianza `e piccola megliore `e lo stimatore).

2. Prendiamo 1ora come stimatore la varianza campionaria, data dalla variabile Sn 2:=

n−1 E[S= ∑n2] 1n−1

= ni=1(∑n−1

(X1ni=1 i ∑E[(X− ni=1 Xn)E[(Xi − 2. μ)Allora: i 2− ] − X∑n)uno stimatore non distorto di σ2ni=1 2] . = n−1

1∑ni=1 E[((Xi − μ) − (Xn − μ))2 ] =

`e avrebbe asintoticamente distorsione non −σn distorto.

2, e E[(Xn − μ)2]) = 1n−1 Notiamo che lo stimatore ·(n−1)σSn ∗:= 2 n

1= ∑σni=12. Perci`o (Xi−XSn)n 2`e 2

dunque `e peggiore della varianza campionaria, anche se

Calcoleremo adesso la varianza della varianza campionaria. Assumiamo per il momento che n−1 σ2 il campione provenga da una distribuzione normale N(μ, σ2). Sn 2∼ χ2n−1 e dunque si avr`a subito Var[Sn2] = n−12σ4

. Infatti, n−1 σ2 In tal Sn 2= caso ∑ni=1

mostriamo (Xi−Xσ

n )2 che =

∑ni=1

(Xn−μ √σn

)2

⇒ ∑ni=1

(Xi−μ

σ − Xn−μ

σ

)2 = ∑ni=1

(Xi−μ

σ

)2 − n(Xn−μ

σ

)2 = ∑ni=1

(Xi−μ

σ

)2 − (∼ χ21) − (∼ χ21) ⇒ n−1 σ2 Sn 2∼ χ2n−1, ove abbiamo usato il seguente teorema:

Teorema 15. Sia (X1,...,Xn) un scuna Xi, 1 ≤ i ≤ n `e MX(t). Allora campione casuale ove la funzione generatrice di cia-

MXn(t)=(MX( nt))n. per mostrare che Xn ∼ N(μ, σn 2). Infatti:

MXn(t)=(MX( nt))n =

(eμ nt+ σ2n2t2 2

)n

= eμt+ σn 2· 2, tda cui la tesi.

14

Teorema 16. Sia (X1,...,Xn) un campione casuale da una popolazione con distribuzione discreta C(x)D(θ) definite o assolutamente exp{∀1 ≤ ∑km m=1 ≤ (T1,...,Tk) sar`a ancora continua dove la densit`a associata sia della forma f(x, θ) =

Am(θ)Bm(x)} con k naturale k da Tm(X1,...,Xn) := della forma esponenziale:

∑ni=1 positivo. Bm(Xi). Siano T1,...,Tk statistiche Allora la distribuzione di

f(T1,...,Tk)(θ, t1,...,tk) = C(t1,...,tk)D(θ)nexp{ ∑ki=1

Am(θ)tm}

.

Esempio. Sia (X1,...,Xn) un campione casuale. ra la densit`a sar`a discreta, ossia sar`a f(p, x) = P[X = Supponiamo x] = 1{0,1}(x)(1−p) che X ∼ expBin(1,p). {xlog( 1−pAllo- pApplicando il teorema otteniamo duce 1{0,...,n}subito (t1)(che tn1T1 ∼ Bin(n, p). T1(X1,...,Xn) = In particolare possiamo ∑ni=1 scriverne B1(Xi) = la ∑densit`a: ni=1 Xi )}. da cui si de-

Esempio. Sia )(X(1 1,...,X− p)n exp{tn) un 1 logcampione ( 1−pp)}.

fT1(p, t1) =

casuale e supponiamo che il nostro campione ca- suale abbia distribuzione uniforme Unif([0,θ]). Vogliamo stimare θ. Supponiamo che, es- sendo θ il massimo valore che ciascuna variabile pu`o assumere, un plausibile buon stimatore possa essere proprio il massimo della statistica ordinata, ovvero Tn(X1,...,Xn) := X(n). Per il teorema di pagina 8, ne conosciamo gi`a la distribuzione:

fTn(θ, x) = (n − 1)!(n n!

− n)!

(xθ)n−1 (1 − xθ)n−n = θnnxn−1 Allora E[Tn] = asintoticamente E[T n2] − E[Tn]2 = ∫ non 0 θ∫ θ0 nθnxdistorto ndx = nn+1per θ = θ θ. Possiamo =⇒ Bθ[Tn] anche = calcolarne n+1−θ . Ne segue che `e distorto, ma la varianza: Var[Tn] = θnnxn+1dx − nn+1 = nθ2 (n+1)2(n+2) mane in ogni caso uno stimatore affidabile. Osserviamo −−−→ n→∞ che 0. Perci`o il massimo X(n) ri- possiamo tuttavia introdurre un nuovo stimatore propriet`a cercate.

che ci assicura la non distorsione, ovvero Tn ∗:= n+1n T, che possiede le

Definizione. Sia (X1,...,Xn) un campione casuale con distribuzione F(θ, x), ove θ ∈ Θ ⊂ R. Sia poi Tn una statistica ∀n ∈ N. Diremo Tn essere uno stimatore consistente di θ :⇐⇒ Tn(X1,...,Xn) −−−→ n→∞ P θ.

Esempio. Sia (X1,...,Xn) un campione casuale, ove X ∈ L2(R). Indichiamo come al solito media e varianza rispettivamente con μ e σ2. Allora abbiamo:

1. La media campionaria Xn := n

11θ ∑ni=1 Xi −−−→ n→∞ P μ, grazie alla legge debole dei grandi numeri poiché limn→+∞ P[(Xn − μ) > ε] = 0, ∀ε > 0.

2. Consideriamo adesso la varianza campionaria

Sn 2:= n − 1

1

∑n(i=1(Xi − Xn)2 = n

n − 1

1n

) Xi 2− X2n. Abbiamo ora i seguenti tre termini:

15

∑ni=1

• limn→+∞ n−1 n= 1, un semplice limite;

• 1n

∑ni=1 Xi 2fatto che X−−−→ 2 n→∞ rimane P E[X2], ancora grazie alla legge debole dei grandi numeri e al

ancora sommabile;

• X2n

−−−→ n→∞ P μ2 = E[X]2 grazie al Teorema 4 sulla convergenza.

Ne prodotto segue e quindi prodotto che Sper n

2−−−→ n→∞ costanti P σ2, sempre per i teoremi sulla convergenza di somma,

di variabili casuali.

3. Consideriamo ancora il campione casuale distribuito uniformemente Unif([0,θ]) con stimatore Tn(X1 ...,Xn) := X(n). Troviamo che anch’esso `e consistente per la stima del FX(n)massimo. (θ − ε) = (Infatti, 1 − θεP[|Tn − θ| > ε] = P[θ − Tn > ε] = P[X(n) ≤ θ − ε] = consistente per θ.

)n −−−→ n→∞ 0. Allo stesso modo si pu`o verificare che anche T n ∗`e

Definizione. Sia (X1,...,Xn) un campione casuale e Tn : X −→ YTn una statistica (stimatore). Vi sia inoltre una funzione di parametri a : Θ −→ YΘ. Allora la funzione non negativa Loss : (YTn ∪ YΘ) × YΘ −→ R≥0 viene detta Funzione di Perdita se soddisfa alle seguenti 1. Losscondizioni:

(a(θ),a(θ)) = 0, ∀θ ∈ Θ; 2. Per ogni Tn Rischio, tale ∈ T , che Riskesiste (Tn,a(θ)una funzione ) = Eθ[LossRisk (Tn,a(θ): YTn × )], Y∀θ Θ −→ R, detta Funzione di

∈ Θ.

Osservazione. La funzione di perdita pu`o essere pensata come una misura della discre- panza tra l’azione Tn e lo stato della natura a(θ).

Definizione. Possiamo gi`a definire due tipologie di funzioni di perdita che spesso vengono utilizzate in statistica:

1. Loss1 (Tn,a(θ)) := |Tn − a(θ)|, chiamata Errore assoluto; 2. Loss2 (Tn,a(θ)) := (Tn − a(θ))2. Essa ammette anche come possibile funzione di rischio Risk2 (Tn,a(θ)) := Eθ [(Tn − a(θ))2]; se tuttavia a = idΘ, allora la funzione MSEθ(Tn) := Risk2 (Tn,θ) prende il nome di Mean Square Error (oppure Errore Quadratico Medio).

Osservazione. Semplicemente aggiungendo e sottraendo il valore E[Tn]2 si ottiene subito la seguente uguaglianza: MSEθ(Tn) = Varθ[Tn] + Bθ[Tn]2. Teorema 17. Sia Tn uno stimatore di θ (non necessariamente non distorto). Allora si ha che limn→+∞ MSEθ(Tn) = 0 `e condizione sufficiente (ma non necessaria) per la consistenza di Tn. Dimostrazione. Si ha infatti la seguente semplice catena di diseguaglianze:

P[|Tn − θ| > ε] =

∫|Tn−θ|>ε fTn(θ, tn)dtn

<

∫|Tn−θ|>ε

(tn ε− 2 θ)2

fTn(θ, tn)dtn < ε12 MSEθ(Tn).

16

1.2.2 Intervalli di confidenza

Lezione del 15/03, ultima modifica 26/03, Michele Nardin

Sia (X1, ..., Xn) un campione casuale definito da una variabile casuale avente funzione di ripartizione FX(x, θ). Vogliamo stimare l’incognita θ, e per farlo ci serviamo di uno stimatore Tn. Una volta estratto il campione casuale, e quindi in possesso di una n-upla di valori reali (x1, ..., xn) che ne rappresenta una determinazione, possiamo effettivamente calcolare valore della nostra stima: `e impensabile per`o che la stima coincida esattamente con il valore incognito (se X ha distribuzione continua P(Tn = θ) = 0!). Dobbiamo quindi associare a Tn un margine di errore.

Introduciamo innanzitutto il concetto di Statistica Pivot:

Definizione 8. Sia (X1, ..., Xn) un campione casuale da una distribuzione con funzione di ripartizione FX(x, θ), θ ∈ Θ. Definiamo Statistica Pivot una funzione Q((X1, ..., Xn),θ) tale che

1. Q `e funzione del campione casuale e del parametro θ (parametro su cui si vuol fare

inferenza)

2. Q non contiene parametri incogniti oltre a θ

3. la distribuzione di Q, FQ, `e completamente nota (ossia non dipende da θ)

4. Q `e invertibile rispetto a θ

Esempi: Campione casuale da N(μ, σ2):

1. Supponiamo di conoscere la varianza: allora un esempio di statistica pivot `e

Zn =  ̄Xn − μ σ/√n

la quale, grazie all’ipotesi di campionamento da vc normale, ha distribuzione N(0,1) (che non dipende da μ).

2. Supponiamo di non conoscere la varianza: in tal caso, al posto della varianza usiamo lo stimatore varianza campionaria S2n, il quale `e non distorto (gi`a dimostrato) e consistente (infatti MSEσ2(S2n) = V ar(S2n)+B2(S2n) = 2σ4

n−1 → 0) e quindi la statistica pivot in questione sar`a

Q =  ̄Xn − μ §n/√n la quale (dimostreremo che) ha distribuzione t-student con n-1 gradi di libert`a.

Esempio introduttivo (exit poll)

Vogliamo stimare la proporzione pi dei voti ricevuti dall’iesimo partito sul totale. Il no- stro problema sar`a quello di trovare un intervallo centrato nella stima ˆpi , ed un margine d’errore, ME, tale per cui, ad una fissata soglia di probabilit`a α si abbia

P[pi ∈ (ˆpi − ME, ˆpi + ME)] = 1 − α

17

Costruzione generale

In generale, sia θ0 il valore vero del parametro θ che vogliamo stimare, e per semplicit`a assumiamo che Tn sia un suo stimatore tale che

√n(Tn − θ0) → dN(0,σT2n)

Per il momento assumiamo di conoscere σTn2, sicché

Zn =

√n(Tn − θ0) σTn

∼ aN(0,1)

Bisogna notare che Zn `e una statistica pivot. Fissato α ∈ (0,1), consideriamo i quantili della X distribuzione N(0,1), ±zα/2 (ossia quei valori tali per cui, se X ≤ zα/2)=1− α). Possiamo affermare che, per n sufficientemente ∼ N(0,1), P(−zα/2 grande, (il simbolo ≤ = .indica un’uguaglianza approssimata)

P(−zα/2 ≤ Zn ≤ zα/2) = .1 − α

da cui

P(−zα/2 ≤

√n(Tn − θ0)

σTn ≤ zα/2) = .1 − α e ancora

P(Tn − zα/2σ√Tnn ≤ θ0 ≤ Tn + zα/2σ√Tnn) = .1 − α Possiamo quindi definire un intervallo casuale,

IC =

[Tn − zα/2σ√Tnn,Tn + zα/2σ√Tnn]

(`e casuale P(θ ∈ IC) perch`e = .per 1 − α.

Tn `e una vc). IC `e uno Stimatore Intervallare. Si pu`o affermare che

Nomenclatura : zα/2 si dice Fattore di Affidabilit`a, σ√Tnn si dice Standard Error dello sti- matore Tn.

Sia ora (x1, ..., xn) una determinazione campionaria (ossia i dati effettivamente osservati da un campione casuale) (cio`e una n-upla) e sia Tn(x1, ..., xn) = tn l’effettivo valore assunto dallo stimatore.

Definiamo di seguito l’intervallo di confidenza con probabilit`a di copertura 1 − α

ICθ(1 − α) :=

[tn − zα/2σ√Tnn,tn + zα/2σ√Tnn]

La probabilit`a di copertura viene anche detta livello di confidenza.

18

Nella campionaria il teorema pratica, 10 di (di Tn, Slutsky), σST2n Tn2non `e , il quale troviamo noto a priori. Possiamo per`o usare sappiamo che converge in probabilit`a che

lo stimatore a σTn2varianza . Allora, per

Zn =

√n(Tn − θ0)

STn =

√n STn Tn −

√n STn θ0 → dN(0,1) Possiamo quindi ripetere il ragionamento fatto poco sopra usando la varianza campionaria al pari posto a 1 − di α Scome

Tn2, e quindi costruire l’intervallo di confidenza con probabilit`a di copertura

ICθ(1 − α) :=

[tn − zα/2S√Tnn,tn + zα/2S√Tnn

]

Intervallo di confidenza per la media μ

Sia (X1, ..., Xn) un campione casuale, media e varianza incognite. Siano X ̄n e Sn 2gli stimatori di media e varianza della popolazione. Allora per il TLC e per il teorema di Slutsky si ha che √n( X ̄n − μ)

Sn

→ dN(0,1)

che `e una statistica pivot. Quindi l’intervallo di confidenza con probabilit`a di copertura 1 − α (sempre approssimato) sar`a

ICμ(1 − α) =

[

X ̄n − zα/2 √Snn, X ̄n + zα/2 √Snn]

Intervallo di confidenza per una proporzione p

Sia (X1, ..., Xn) un e consistente) di p. campione casuale da Troviamo che per il b(1,p) TLC e e per sia la ˆpn WLLN = n

1∑ni=1 (legge Xi lo stimatore dei grandi (corretto numeri) √√ˆpn(ˆpn(1 n − − ˆpp) n)

→ dN(0,1)

e quindi l’intervallo di confidenza con probabilit`a di copertura 1 − α approssimato sar`a

ICp(1 − α) =

[ˆpn − zα/2√ ˆpn(1 n − ˆpn)

, ˆpn + zα/2√ ˆpn(1 − ˆpn)

n

]

Distribuzione esatta della statistica pivot: distribuzione t di Student

Lezione del 18/03, ultima modifica 26/03, Michele Nardin

La distribuzione t di Student con ν gradi di libert`a `e definita come T = √SZ2/ν ove Z ∼ N(0,1) mentre S2 ∼ χ2ν (chiquadro con ν gradi di libert`a). La funzione di densit`a `e

ftν(t, ν) = Γ((ν Γ(ν/2)

+ 1)/2)

√1πν

1 [1 + t2/ν]v+1

2 1R(t)

19

tale funzione `e simmetrica, ha la classica forma a campana come la normale, ma a differenza di quest’ultima ha le code pi`u pesanti. Risulta che la statistica pivot per la media in campioni poco numerosi 1 (in caso di campionamento da normale) ha distribuzione esatta t di Student. Infatti

Q = XSn/n − √n μ

=

X√σ/n−μ √n Sσn22

troviamo al numeratore Xn−μ

σ/√n ∼ N(0,1), (grazie al fatto che le Xi sono equi distribuite normalmente) mentre al denominatore abbiamo che

√Sn2σ2 =

√(n − 1)Sn

2(n − 1)σ2 =

√

H (n − 1)

Abbiamo gi`a dimostrato che H = abbiamo la radice di una chiquadro (n−1)Sdiviso σ2 n 2i ∼ suoi χ2n−1gradi , quindi in definitiva al denominatore di libert`a, ovvero siamo proprio in presenza di una distribuzione t di Student.

Osservazione importante: Quindi, quando il campione casuale `e poco numeroso, `e conveniente usare i quantili della distribuzione t di student per costruire gli intervalli di confidenza. Per numerosit`a campionarie n > 30, approssimare la distribuzione t di stu- dent con la distribuzione normale offre risultati soddisfacenti. Ricordiamo che per il tlc Q → N(0,1))

Intervallo di confidenza esatto

Fissato un livello di confidenza 1 − α, consideriamo i quantili della distribuzione t di student (con n-1 gradi di libert`a, ove n `e la dimensione campionaria) ±t(α/2;n−1), troviamo

P

(−t(α/2;n−1) ≤ XSn/n − √n μ

≤ t(α/2;n−1))

= 1 − α

Notiamo che questa volta vale l’uguaglianza ’vera’, poiché non stiamo considerando ap- prossimazioni asintotiche. In presenza del campione effettivamente estratto, (x1, ..., xn), scriviamo xn e s2n i valori assunti da media e varianza campionaria, l’intervallo di confidenza `e

ICμ(1 − α) =

[xn − t(α/2;n−1)√sn 2n,xn + t(α/2;n−1)√s2nn

]

Osservazione 5. Alcune osservazioni che, pur sembrando banali, `e bene tenere a mente:

1. Al anche crescere l’ampiezza del livello di IC

di confidenza (1−α) e/o della varianza campionaria Sn 2cresce

2. Al crescere dell’ampiezza campionaria n, (fermo restando il livello di confidenza)

l’ampiezza di IC diminuisce

1In realt`a vale per tutti i campioni, `e solo che da un certo punto in poi la differenza con la normale `e davvero trascurabile! Sulle tavole si riporta solo per ν < 120

20

Intervalli di confidenza per la varianza

Sia (X1, ..., Xn) un campione casuale da N(μ, σ2). Consideriamo la statistica pivot

W = n σ− 2 1

Sn 2Abbiamo gi`a mostrato che W ∼ χ2n−1. Ma allora, dato che noi cerchiamo q1,q2 t.c.

P

(q1 ≤ n σ− 2 1

Sn 2≤ q2)

= 1 − α

troviamo che essi sono i quantili di ordine α/2e1 − α/2 della chiquadro con n-1 gradi di libert`a, che indicheremo q1 otteniamo:

= χ2(n−1,α/2) e q2 = χ2(n−1,1−α/2). Con qualche passaggio

P

( 1q2 ≤ (n − σ2

1)Sn 2≤ q11)

= 1 − α

P

((n − 1)Sn

2q2 ≤ σ2 ≤ (n − q1

1)Sn

2)

= 1 − α

Troviamo cos`ı l’intervallo casuale (e di conseguenza il relativo intervallo di confidenza, una volta estratto il campione e trovato un valore a Sn2)

IC =

[(n − 1)Sn

2q2 ; (n − q1

1)Sn

2]

Intervalli di confidenza per la differenza di medie

Vogliamo confrontare due distribuzioni: sintetizziamo la differenza tra due popolazioni tramite la differenza delle loro media.

Supponiamo inizialmente di avere due campioni casuali tra loro indipendenti: (X(YNB: 1, 1, ..., non ..., YXnnecessariamente 2n) 1) da da una una distribuzione distribuzione n1 D1, con media D2, con media dev’essere uguale a μμ2 n2

1 (ignota) (ignota) e e varianza varianza σσ2 21 2(nota)

(nota)

Consideriamo gli stimatori media campionaria per le due medie, che indicheremo con X e Y . La statistica pivot che ci interessa per ∆ = μ1 − μ2 sar`a

Z = (X − [ Y σn121 ) + − σn(μ222] 1 1− μ2) 2

Notiamo Z ∼ ache var(X −Y ) = σn121 + nσ222 N(0,1), quindi possiamo trovare dato che cov(X,Y ) = 0 per l’indipendenza. un intervallo di confidenza 2

Ma allora

IC∆(1 − α) = [(X − Y ) − ME;(X − Y ) + ME]

2Approssimato, visto che conosciamo solo l’andamento asintotico di Z! D1 e D2 non `e detto che siano mormali!

21

[ove ME = zα/2

σn1 12+ σn222] 12. Al posto delle varianze possiamo usare anche gli stimatori corretti e consistenti varianza campionaria, e giungere allo stesso risultato per il teorema di Slutsky.

In generale non conosciamo la varianza delle distribuzioni: in base al problema che dob- biamo affrontare, pu`o essere plausibile supporre di conoscere la distribuzione delle due popolazioni a meno di uno o pi`u parametri.

Location Model: Supponiamo di avere μ1 e varianza con media μ2 σe 1 2varianza (ignote), campioni siano tra loro i loro stimatori indipendenti σ2 2(ignote) e ed (X1, ..., Xi loro X e stimatori S1 2inoltre che n1) da distribuzione e (Yσ1 1, = ..., Y σYe 2 nS2= ) 22. da σ. normale con media distribuzione Supponiamo Possiamo normale che i due ’fondere’ le informazioni contenute in S1 2e S22:

(P ooledV ariance) Sp 2:= (n1 − 1)Sn1 + 1 2+ n2 (n− 2 2

− 1)S2 2che risulta essere uno stimatore corretto e consistente di σ2 (esercizio). La statistica Pivot che prendiamo in considerazione sar`a

T = (X − Y ) − (μ1 − μ2)

Sp

( 1n1 + n12) 12

la quale risulta essere distribuita come tn1+n2−2. Ricalcando i passaggi delle applicazioni precedenti, fissato α troviamo l’intervallo casuale per ∆

IC =

[(X − Y ) − t(n1+n2−2;α/2)Sp

( 1n1 + n12) 1] 2Intervalli di confidenza per la differenza di proporzioni Supponiamo di avere da distribuzione b(1,p2), ( 1n1 (Xcon 1, ..., stimatore Xn1) da ˆpdistribuzione 2. b(1,p1), Supponiamo che i due con stimatore campioni siano ˆp1 e (Ytra 1, ..., loro Yn2)

indipendenti. Allora∆= ˆp1 − ˆp2 ∼ aN

+ n12) 12 ;(X − Y ) + t(n1+n2−2;α/2)Sp

(p1 − p2, p1(1 n− 1 ) p1)

quindi usando la statistica Pivot

Z = √(p1 ( ˆp− 1 p− 2, ˆpp2) 1(1−pn− 1 (p1 − p2) 1)

+ p2(1−pn2

2)

+ p2(1 n− 2

p2)

∼ aN(0,1)

trovo l’intervallo di confidenza

IC∆(1 − α) =

)

[( ˆp1 − ˆp2) − zα/2√A(p1,p2); ( ˆp1 − ˆp2) + zα/2√A(p1,p2)]

22

( p1(1−pn1 1)

+ p2(1−pn2

2)

). Ovviamente al posto di p1 e p2 uso gli stimatori corretti e consistenti ˆp1 e ˆp2.

Lezione del 20 marzo, ultima modifica 26 marzo, Michele Nardin

Intervalli di confidenza per rapporti di varianze

Introduciamo per prima cosa la Distribuzione F di Snedecor-Fisher.

Siano rapporto W1 tra ∼ due χ2ν1 chiquadrato e W2 ∼ χ2ν2 divise indipendenti. per La distribuzione Fν1,ν2 `e definita come il

i rispettivi gradi di libert`a, in formule

W = WW1/ν2/ν1 2

ed ha funzione di densit`a

fW(w;ν1,ν2) = Γ(Γ(ν2 1) ν+ 1+ν2 Γ(2 )

ν2 2)(ν1/ν2)ν1/2wν1/2−1 [1 − νν12w] ν1+ν2 2 1R+(w) Vale la seguente propriet`a, utile per calcolare i quantili non tabulati:

Se W ∼ Fν1,ν2 ⇒ 1W ∼ Fν2,ν1

Le tavole (comunemente) forniscono i valori dei quantili per (1−α) ∈ {0.80,0.90,0.95,0.975,0.99,0.999}. Quindi possiamo sfruttare la propriet`a sopra scritta per trovare che

wα;ν1,ν2 = 1

w1−α;ν2,ν1

Intervallo di confidenza per rapporti di varianze Supponiamo (ignote), varianza di avere (Xi loro stimatori σ2 2(ignote) e i 1, ..., X e XSn11 2) da distribuzione e (Y1, ..., Yn2) normale con media loro stimatori Y e S22. ove A(p1,p2) =

μ1 da distribuzione normale con e varianza media μ2 σ1 2e

per Fissato S22).

Ricordiamo che (n1−1)Sσ1 21 2∼ χ2n1−1 (idem

1− α consideriamo una statistica pivot e w1, w2 tc P(w1 ≤ W ≤ w2)=1 − α. La statistica pivot in questione sar`a

W =

(n1−1)Sσ1 21 2/n1 − 1 (n2−1)Sσ2 22 2/n2 − 1 ∼ F(n1−1),(n2−1) poich`e rapporto di due chiquadro. Risulta inoltre, semplificando:

W = Sσ11

22Sσ22 22S12S2 223 = σσ21 22

w1 e w2 saranno esplicitamente troviamo che

i w1 = quantili di ordine α/2 e 1 − α/2 w(n1−1,n2−1;α/2) = 1 w(n2−1,n1−1;1−α/2) della distribuzione F(n1−1),(n2−1), e w2 = w(n1−1,n2−1;1−α/2) e quindi

1 − α = P(w1 ≤ σσ21

22(S12/S2

2w2 ≤ σσ12 22≤ S12w/S1

2 2)

da cui troviamo anche il relativo intervallo casuale. Quando invece abbiamo un campione effettivamente di confidenza sar`a

estratto, posti s21 e s22 i valori assunti dalla varianza campionaria, l’intervallo

IC σσ12

22S12S2 2≤ w2) = P

[s21/s22

 w(n1−1,ns212−1;1−α/2)/s22

; s21/s22 1 w(n2−1,n1−1;1−α/2) w2 ; s21w/s1 (1 − α) =

22

]

=



1.2.3 Test per la verifica di ipotesi

Lezione del 25/03, ultima modifica 20/05, Andrea Gadotti

La procedura di test per la verifica di ipotesi che descriveremo a breve cerca di fornire una soluzione ai seguenti problemi:

1. Determinare quanto un’ipotesi `e realistica, verosimile, compatibile con l’informazione

empirica a disposizione.

2. Trovare un ragionamento oggettivo (matematico) per inferire dall’informazione di- sponibile (ovvero il contenuto di un campione) circa la veridicit`a dell’ipotesi formu- lata.

3. Misurare in qualche modo questa ”vicinanza” tra ipotesi e realt`a.

Useremo statistiche pivot in ambito parametrico: la distribuzione da cui proviene il cam- pione casuale (X1, ..., Xn) `e nota a meno di uno o pi`u parametri. Di seguito si pu`o vedere una descrizione generale della situazione che si riproporr`a per tutta questa sezione:

X ∼ FX(x, θ), θ ∈ Θ, (X1, ..., Xn) iid. Le nostre due ipotesi saranno della forma:

{ H0 : θ ∈ Θ0 H1 : θ ∈ Θ1

con Θ0 ∪ Θ1 = Θ.

Chiameremo H0 ipotesi nulla e H1 ipotesi alternativa. Di solito H0 rappresenta la cono- scenza pregressa, la supposizione vera fino a prova contraria. Al contrario, H1 `e l’ipotesi di lavoro, quella su cui ripieghiamo nel momento in cui il nostro test risulta in contraddizione con H0. Il test si riduce a una regola di decisione in merito a H0 e H1 sulla base del campione

24

casuale (X1, ..., Xn) da X ∼ FX(x;θ). Dividiamo lo spazio dei campioni in due regioni disgiunte: C (regione critica del test) e Cc. La decisione pu`o chiaramente essere corretta, ma anche errata, poiché il campione costituisce un’informazione non completa. Risulta quindi necessario formulare delle conclusioni in probabilit`a, ovvero associare alla nostra conclusione la probabilit`a che questa sia corretta, cercando ovviamente di massimizzarla. Possiamo riassumere le varie possibilit`a nella tabella e nel disegno sottostanti:

H0 `e vera H0 `e falsa Rifiuto H0 errore di I specie nessun errore Non rifiuto H0 nessun errore errore di II specie

Esempio Lancio di una moneta onesta. Consideriamo il campione casuale (X1, ..., Xn) e il numero di teste Sn = ∑ni=1 Xi. Vorremmo stimare la probabilit`a che esca testa con la media campionaria Xn. In questo caso potremmo avere:

{ H0 : p = 1/2 H1 : p = 1/2

La regola di decisione consiste quindi nel rifiutare H0 se (X1, ..., Xn) ∈ C e invece rifiutare H0 se (X1, ..., Xn) ∈ Cc. Ci piacerebbe trovare una regola di decisione che permetta di minimizzare la probabilit`a di commettere errori di I o II tipo. Purtroppo questo non `e possibile, per la natura stessa della relazione che corre tra gli errori di I e II tipo. Di seguito un esempio che ci d`a un’idea del perché:

Esempio Consideriamo un campione casuale (X1, ..., Xn) da N(μ, σ2) con σ2 noto. Sup- poniamo che le nostre due ipotesi siano:

{ H0 : μ = μ0 ovvero N(μ = μ0,σ2) H1 : μ = μ1 ovvero N(μ = μ1,σ2) con μ1 > μ0.

25

Consideriamo α := P(rifiutare H0 | H0 vera) = P(il nostro campione appartiene a C | H0 vera) = P(il nostro campione `e ≥ c | la distribuzione corretta `e quella di sinistra) = P(commettere un errore di I tipo) e β := P(non rifiutare H0 | H0 falsa) = P(il nostro campione appartiene a Cc | H0 falsa) = P(il nostro campione `e ≤ c | la distribuzione corretta `e quella di destra) = P(commettere un errore di II tipo). (Nota: α `e detto livello di significativit`a del test) E `evidente che non `e possibile annullare contemporaneamente sia α che β. La procedura si divide quindi in due passi: il primo consiste nel fissare α, il secondo nell’individuare la regola di decisione che minimizza β, in modo da trovare un test ottimo.

Lezione del 05/04, ultima modifica 20/05, Andrea Gadotti In generale una statistica test si pu`o descrivere come di seguito:

{ H0 : θ ∈ Θ0 H1 : θ ∈ Θ1

dove Θ `e lo spazio dei possibili parametri della distribuzione e Θ = Θ0 ∪ Θ1. Quello che vogliamo trovare `e la regola di partizionamento che divida lo spazio dei cam- pioni C in C0 e C1 in funzione di α (deciso da noi). Per farlo imponiamo la condizione α = P(x ∈ C | θ ∈ Θ0). Vediamo ora un esempio con il lancio di una moneta:

Esempio Consideriamo il campione casuale (X1, ..., Xn) dove Xi = 0 con probabilit`a p e Xi = 1 con probabilit`a 1 − p. Facciamo le nostre ipotesi:

{ H0 : p ≤ 1/2 H1 : p > 1/2

Prediamo ora come regola di decisione imponiamo l’equazione in k:

Sn =

∑ n Xi . Deciso un α a nostra discrezione,

∑ n Xi . Deciso un α a nostra discrezione,

α = P(S>k | p ≤ 1/2) (= P(S>k | H0 vera))

26

A questo punto risolvendo l’equazione troveremo il k per il quale rifiuteremo H0 se S>k.

Esempio Sia (X1, ..., Xn) un campione casuale con Xi ∼ b(1,p) e sia

{ H0 : p = p0 H1 : p<p0

Sulla base dell’informazione circa p contenuta in (X1, ..., Xn) vogliamo sottoporre a verifica il sistema di ipotesi in questione. Procediamo in questo ordine:

(a) Prendiamo S := ∑Xi ∼ b(n, p), che `e di fatto il numero di successi. Sotto H0

abbiamo che S ∼ b(n, p0).

(b) Scegliamo una regola di decisione (usando anche la distribuzione -nota- di S sotto H0). Ovvero, individuiamo la regione di rifiuto del test. A questo punto vorremo rifiutare H0 a favore di H1 quando S ≤ k, dove k `e l’incognita che troveremo nel punto (c).

(c) Scelto il nostro α, si ha che il valore di k deve essere tale per cui

α = P(S ≤ k | p = p0) =

(ns)ps0(1 − p0)n−s

A questo punto, essendo α fissato, abbiamo un’equazione in k che risolta ci restituisce il suo valore.

Esempio particolare Nella situazione generale sopra descritta prendiamo un caso par- ticolare ∑kcon n = 20 e p0 = 0,7. Decidiamo α = 0.15. L’equazione diventa: 0,15 = s=0

∑ks=0

(20s

)0,7s0,320−s. Osserviamo che il valore di P(S ≤ k | p = 0,7) per k = 11 risulta 0,1133, mentre per k = 12 `e 0,2277. Quindi il nostro k `e compreso tra 11 e 12. In conclu- sione, se il nostro test dovesse presentare 12 (o pi`u) successi, allora non rifiuteremmo H0. In caso contrario scarteremmo H0 a favore di H1. Definizione 9. Sia β := P(x ∈ C0 | θ ∈ Θ1), ovvero β `e la probabilit`a (fissato α) di commettere un errore di II specie. Chiamiamo potenza del test il valore γ := 1 − β. Un test risulta ottimale quando la sua potenza `e massima. Notiamo che possiamo definire una funzione di potenza γ(t) := 1 − β(t)

Osservazione 6. Prendendo di nuovo in considerazione l’esempio precedente sul campione casuale normale, `e chiaro che una volta fissato α, ovvero c, minore `e μ1, pi`u piccola `e l’area sottesa dalla coda della relativa normale, ovvero β.

Esempio Prendendo di nuovo in considerazione l’esempio precedente sulla bernoulliana, abbiamo che

γ(p)=1 − β(p)=1 − P(S>k | p<p0) = P(S ≤ k | p<p0) =

(ns)ps(1 − p)n−s

Di seguito possiamo osservare il grafico della funzione:

27

∑ks=0

Osservazione Il test in merito al precedente sistema di ipotesi relative a p `e un test esat- to, perché poggia sulla distribuzione esatta di S (S ∼ b(n, p)). Questo per`o non accade sempre, e quindi talvolta `e necessario ricorrere alla teoria asintotica e dei test approssimati.

1.2.4 Esempi di statistiche test (generali e particolari)

Test per la media di una popolazione qualsiasi Supponiamo di avere un campione casuale (X1, ..., Xn) proveniente da una distribuzione non nota di media μ e varianza σ2 (finita) non note. Le nostre ipotesi sono { H0 : μ = μ0

H1 : μ = μ1 con μ1 > μ0

Decidiamo di ”condensare” l’informazione presente nel campione circa μ e σ2 tramite Xn e Sn 2(che ricordiamo essere stimatori non distorti e consistenti), sapendo che Xn → Pμ e Sn

2→ Pσ2. A questo punto, la nostra regola di decisione consister`a nel rifiutare H0 a favore di H1 se Xn `e molto pi`u grande di μ0. Noi sappiamo che Xn ∼ aN

(μ, Sn n2), ovvero Usando questo risultato possiamo individuare SXn/n−μ la √n

regione → DN(0,1) =: Z

critica del test di livello α fissato. Imponiamo la seguente uguaglianza:

α = P(x ∈ C | μ = μ0) = P(Xn ≥ k | μ = μ0)

= P

(XSn/n − √n μ

≥ Sk n/− √μ

n | μ = μ0)

= P

(XSn n/− √μn 0

≥ Sk n/− √μn0

)

28

Quindi

1−α = 1−P

(XSn n/− √μn 0

≥ Sk n/− √μn0

)

= P

(XSn n/− √μn 0

< Sk n/− √μn0

)

= P

(N(0,1) < Sk n/− √μn0

)

ovvero Sk n/− √μn 0

= z1−α = −zα

Dove z1−α `e il valore da cercare sulle tavole relative alla distribuzione normale in funzio- ne dell’α scelto. Nota: l’ultima uguaglianza di probabilit`a `e in realt`a un’approssimazione che `e tanto pi`u corretta quanto pi`u grande `e n. In definitiva, abbiamo che C = {x ∈ κ : Xn−μSn/√n 0 ≥ −zα} = {x ∈ κ : Xn ≥ μ0 − zα √Sn}

Possiamo anche considerare la funzione di potenza approssimata:

γ(μ1)=1 − β(μ1) = P(Xn ≥ μ0 − zασ/√n | μ = μ1)

= P

(Xσ/n − √n μ1

≥ μ0 − zασ/σ/√√n

n − μ) 1

)

= 1 − P = Φ(Z ≥ −zα + (−zα +

√n(μ0 − μ1)

)

σ

dove Φ `e la funzione di ripartizione di N(0,1). Notiamo che il valore di γ(μ1) tende a 1 per n → ∞. Intuitivamente, questo `e esattamente ci`o che ci aspettiamo, in quanto pi`u `e grande μ1, pi`u esso `e distante dal nostro μ0, e di conseguenza `e lecito aspettarsi che la probabilit`a che un campione abbia media vicina a μE `0 chiaro quando quindi invece che μ = una μ1 funzione sar`a bassa, di potenza ovvero la `e tanto potenza migliore del test quanto `e elevata. pi`u il suo grafico sta vicino alla retta y = 1.

Esempio In riferimento al caso generale appena trattato, supponiamo di avere μ0 = 12, Xn = 14,3, Sn 2= 22,5, n = 50. normale N(0,1) troviamo zα = Se 1,645. fissiamo Ne α segue = 0,05, che usando k = 12+1,645le tavole per √la 22,5/50, distribuzione che `e minore di 14,3. Concludiamo quindi rifiutando H0.

Esempio di test esatto con t di Student Abbiamo (X1, ..., Xn) campione casuale da N(μ, σ2) con μ e σ2 non noti. Le nostre ipotesi sono:

{ H0 : μ = μ0 H1 : μ>μ0

Sappiamo che Xn H∼ 0N(μ0,σ2/n) e quindi:

T := XSn n/− √μn 0

= Xσ/n − √n

μ0

√n(μ0 − μ1) σ

√S1 n2/σ2 ∼ Sn/Z

√n ∼ tn−1

29

(vedi pag. 17 e pag. 12) A questo punto, possiamo trovare il nostro valore critico k usando le tavole della distribu- zione tn−1. Notiamo che quello appena mostrato `e un test esatto, in quanto non si basa su un’appros- simazione dello stimatore per valori elevati di n (usando ad esempio il TLC), bens`ı usa la sua distribuzione reale (in questo caso la distribuzione tn−1).

Lezione 08/04, ultima modifica 20/05, Michele Nardin

Esempio di test bilaterale Sia (X1, ..., Xn) un campione casuale da una distribuzione avente media μ e varianza σ2 finite. Considerato il fatto che non abbiamo informazio- ni sulla distribuzione delle variabili casuali, ci appoggeremo su di un test approssimato. Supponiamo di voler verificare la seguente ipotesi relativa alla media della distribuzione: { H0 : μ = μ0 H1 : μ = μ0

Seguendo la falsariga di quanto visto sopra, procediamo come segue:

(a) Prendiamo X =

la distribuzione sotto H0 abbiamo esatta ∑che n Xi

, X delle che ∼ a`e stimatore della media del variabili, consideriamo la N(μ0,σ2/n) (questo, come campione. Non conoscendo distribuzione approssimata: gi`a detto varie volte, per il TLC).

(b) Scegliamo una regola di decisione: usando anche la distribuzione asintotica di X sotto H0, individuiamo la regione di rifiuto del test. Dobbiamo trovare quindi due valori h, k ∈ R tali per cui rifiuto H0 se Xn ≤ h o Xn ≥ k.

(c) Scelto il livello di confidenza α, si ha che h, k vanno scelti in modo tale per cui

α = P(X ≤ h ∨ X ≥ k | μ = μ0)

Grazie alla normalit`a della distribuzione asintotica, possiamo supporre che la distri- buzione di Xn sia simmetrica, per lo meno da un certo n in poi (ricordiamo che, per campioni con numerosit`a superiori a 30, l’approssimazione `e gi`a buona). Questo ci permette di scrivere

α2 = P(X ≤ h | μ = μ0) = P(X ≥ k | μ = μ0)

Dato che X Sn/− √μn 0

→ DN(0,1)

denotato con zα/2 il quantile di ordine α/2 della normale standard, la regione critica (di rifiuto) sar`a

C =

{x = (x1, ..., xn) ∈ X :

∣∣∣∣X Sn/− √μn0

∣∣∣∣ ≥ zα/2}

30

da cui

k = √Snnzα/2 + μ0(= −h)

Consideriamo inoltre la funzione di potenza associata, sempre sfruttando l’approssi- mazione normale ed il fatto che, per n grande, Sn ≈ σ:

γ(μ) = P(X ≤ μ0 − √Snnzα/2 | μ = μ0) + P(X ≥ μ0 + √Snnzα/2 | μ = μ0)

≈ Φ(√n(μσ 0 − μ)

− zα/2)

+

[1 − Φ(√n(μσ 0 − μ)

− zα/2)]

Di seguito possiamo osservare il grafico della funzione:

Osservazione: i test costruiti usando t di Student sono pi`u conservativi rispetto a quelli costruiti con approssimazione Normale, nel senso che `e pi`u facile che un test venga accettato usando la prima rispetto alla seconda. Questo `e dato dal fatto che la distribu- zione t di student ha le code pi`u pesanti rispetto alla normale! (in particolare fissato un α ∈ (0,1), i quantili della normale standard zα/2 sono pi`u piccoli dei quantili della t di Student con n − 1 gradi di libert`a tα/2;n−1, cio`e |zα/2| < |tα/2;n−1|)

Lezione del 12/04, ultima modifica 20/05, Andrea Gadotti

Esempio di test t-Student per due campioni Campione (X1, ..., Xn1) da N(μ1,σ2)e(Y1, ..., Yn2) da N(μ2,σ2) indipendenti. La varianza σ `e la stessa ma non `e nota. Supponiamo di avere elementi per pensare che:

{ H0 : μ1 = μ2 H1 : μ1 > μ2

Abbiamo che X1 − X2 ∼ N(μ1 − μ2,σ1/n1 + σ2/n2). Prendiamo ora

S2p := (n1 − 1)S21 + (n2 − 1)S22

n − 2

31

dove n = n1 + n2. Abbiamo che:

T := (X1 Sp− √X1/n2) − 1 + (μ1/n1 − 2

μ2)

H∼ 0tn−2

dove abbiamo usato anche il fatto che Sp 2`e una chi-quadro con n−2 gradi di libert`a (vedi pag. 19). In conclusione, avremo che la nostra regione critica sar`a C = {(x, y) | T ≥ tn−2;α}.

Esempio con bernoulliana Abbiamo (X1, ..., Xn) campione casuale da b(1,p) (ricor- diamo: media p, varianza p(1 − p)). Le nostre ipotesi sono:

{ H0 : p = p0 H1 : p = p1

con p1 < p0. Consideriamo

ˆpn :=

∑Xi n

)

Abbiamo quindi:

Z := √ ˆpn − p0 ˆpn(1−ˆpn)

n A questo punto imponiamo: α = P(Z ≤ zα | H0).

Esempio di test sulla varianza Campione casuale (X1, ..., Xn) da N(μ, σ2), μ e σ2 non noti. Le nostre ipotesi sono:

{ H0 : H1 : σσ2 2 = = σσ0 1 22con σ1 2> σ02. Notiamo che n−1

σ0 2Sn

2∼ aN

(p, p(1 n

− p)

H∼ 0χ2n−1 =: W. (Nota: come in molti esempi precedenti, il fatto che sia intelligente tirare fuori queste osservazioni che portano all’analisi di distri- buzioni conosciute `e calato dall’alto, almeno per ora). Imponiamo:

α = P(Sn 2≥ k | H0) = P

(W ≥ n σ− 0 21

k)

Quindi n−1 σ0 2In conclusione, k = wα;n−1. 3

rifiuto H0 se W ≥ wα;n−1, ovvero se Sn 2≥ k = σ0 2wn−1 α,n−1

.

Esempio In riferimento al caso generale appena trattato, supponiamo di avere n = 25, σ0 2= 15, σ1 2= 20, s2n = 17,4 e α = 0,05. Allora:

k = w0,05;(25−1) = w0,05;24 = 36,415 > 27,84 = 25 15 − 1

17,4 = n σ− 0 21

s2n = w

3Ricordiamo che per la distribuzione chi-quadro, con wα;n−1 intendiamo il valore tale che P(χ2n−1 ≥ wα;n−1)

32 (n − (n − σ2 σ0 21

Sn 2≥ n σ− 2 1

Sn 2≥ n σ− 0 21

k | σ2 = σ021

k)

) = P

= P

In conclusione, non rifiutiamo H0.

Esempio con due campioni normali Campione da N(μ2,σ22) indipendenti. μi e σi 2non noti. Le nostre (X1, ..., ipotesi Xn1) sono:

da N(μ1,σ12)e(Y1, ..., Yn2)

{ H0 : H1 : σσ1 1 22= = σσ2 2 22Nota: l’ipotesi dell’uguaglianza delle due varianze prende il nome di omoschedasticit`a. Sappiamo che Si

2→ Pσi 2. Inoltre

W := SS2122/σ2

2/σ1 2=

(n2−1)Sσ2 22 21 (n1−1)Sσ1

2n2−1

1 2∼ F(n2−1),(n1−1)

(vedi pag. 20) Sotto H0 consister`a abbiamo chiaramente nel rifiutare H0 che W := a favore di H1 se SS2122SS/σ/σ21 22W >k2, dove k1 e k2 dipendono dalla distribuzione dividendo equamente in due parti la probabilit`a 2 1 22`e = SS21 22. ”lontano” di W di errore, La nostra regola di decisione da 1, ovvero se W <k1 o = le due SS21 22e dal valore di α. Perci`o,

equazioni risultano:

α/2 = P(W >k2 | σ1 2= σ22) e 1 − α/2 = P(W <k1 | σ1 2= σ22)

In conclusione,

C = {(x1,x2) : S22S1 2< w(n2−1),(n1−1);α/2 o S22S1 2> w(n2−1),(n1−1);1−α/2

In riferimento al caso generale appena trattato, supponiamo di avere n1 = 14, n2 = 10 sCome 21 = 17,4, nel caso σ1 2= generale, 20, s22 = diciamo 37,9 e Abbiamo che w(10−1),(14−1);0,025 α = 0,05.

= W 3,31 := Se 22/Sw(10−1),(14−1);0,975 1 2∼ F(n2−1),(n1−1)= .

1/w13,19;0,025 = 1/3,76 = 0,26. Poiché s22/s21 = 37,9/17,4=2,178, decidiamo di non rifiutare H0.

33

1 n1−1

Capitolo 2

Seconda parte del corso

Lezioni dal 15/04 al 06/05 comprese. Autore: Marco Peruzzetto. Questa parte comprende ed amplia le cose viste a lezione, estendendo alcune dimostrazioni e osservazioni. Quanto non fatto in classe verr`a denotato da un (\*).

Definizione: Sia X := (X1,...Xn) un vettore casuale da distribuzione F( x, θ) per θ ∈ Θ e con sia x fX= i(x(xi,θ) 1,...,xla corrispondente n) funzione densit`a di ciascuna Xi, ∀1 ≤ i ≤ n. Indicheremo una qualsiasi possibile determinazione del vettore X. Essa conterr`a tutta l’informazione in merito a θ. Possiamo allora definire la Funzione di Verosimiglianza come la funzione:L(θ, x) := f X( x, θ) = f(X1,...,Xn)(xi,...,xn,θ) , θ ∈ Θ, che rappresenta quindi la funzione di densit`a dell’intero vettore in dipendenza del para- metro θ. Nel caso in cui il vettore casuale sia un campione casuale, allora tutte le varia- bili casuali di cui esso `e composto saranno i.i.d., ragion per cui la funzione di massima verosimiglianza assumer`a la seguente tipica forma:

L(θ, x) =

∏ni=1fXi (xi,θ) , θ ∈ Θ. Esiste anche la Funzione di Log-Verosimiglianza definita come l(θ, x) := log(L(θ, x)).

Esempio: Sia (X1,...,Xn) ∼ P oisson(θ). Allora

L(θ, x) =

∏nei=1

−θxi! θxi

1N(xi) = e−nθ∏ni=1 θ∑ni=1 xi!

x1 ∏ni=11N(xi) da cui

l(θ, x) = log(θ)

∑ni=1

xi − nθ −

∑nlog(xi!). i=1

Osservazioni:

• La funzione di verosimiglianza d`a un valore alla probabilit`a che x provenga da F X( x, θ) per tutti i differenti valori di θ ∈ Θ.

34

• Nella funzione di verosimiglianza `e stato volontariamente invertito il parametro θ con il parametro x rispetto, ad esempio, alla funzione densit`a. La ragione si basa sulla diversa interpretazione della stessa: a tutti gli effetti la funzione di verosimiglianza non `e altro che la funzione densit`a del vettore casuale X. Quindi essa pu`o essere vista in due modi diversi: il primo interpreta la funzione L come una funzione di x, e quindi del risultato, una volta fissato il valore del parametro (perci`o L esattamente la densit`a), mentre il secondo la interpreta come una funzione del parametro θ, per un fissato valore del risultato x. Proprio in quest’ultimo caso ha senso parlare di verosimiglianza: il valore assunto da L indica quanto verosimilmente il valore di un parametro (θ) sia corretto rispetto al risultato che si possiede ( x).

• (\*) Data una determinazione x di X, la funzione L(·, x), essendo la densit`a del vet- tore casuale, esprime la probabilit`a che X assuma proprio il valore x. Ci`o avviene in modo diretto se le variabili componenti il vettore sono discrete e tramite integrazione se continue. Ha senso allora chiedersi, data una determinazione x0 di X, quale sia (se esiste) un possibile valore θ0 ∈ Θ capace di massimizzare il valore di L(θ0, x0). Massimizzare tale valore significa infatti per quanto detto, andare a massimizzare la probabilit`a che X assuma il valore x0. Ci`o avverr`a direttamente se il vettore casuale `e discreto, ma anche se `e continuo, e ci`o banalmente grazie alla monotonia dell’in- tegrale, in quanto, se riusciamo a massimizzare la funzione con θ anche l’integrale (ovviamente integrando in d x) sar`a massimo (rivedere).

• L’importanza di cercare il valore del parametro che massimizzi L fissata la determina- zione risiede nel fatto che spesso in statistica si ha a che fare con poche determinazioni e si parte dunque dall’evidente presupposto che se il campionamento effettuato ci ha fornito quelle specifiche determinazioni, esse debbano essere mediamente le pi`u pro- babili. Tale presupposto viene in effetti denominato Principio di “Rational Belief”. La probabilit`a che dato quel campione casuale si ottengano quelle determinazioni la immagineremo quindi come la massima possibile. Cercheremo dunque un θ ∈ Θ che soddisfi a ci`o. `E inevitabile che attraverso la verosimiglianza si possano ottenere degli stimatori del parametro.

• La funzione di log-verosimiglianza `e stata introdotta pressoché per il semplice motivo di semplificare i calcoli quando si cerca di andare a massimizzare la funzione di verosimiglianza. Essa risulta dunque essere comoda, in quanto, essendo il logaritmo una funzione strettamente crescente, il massimizzante di l(·,·) coincider`a con quello di L(·,·).

Esempio (Problema dei Pesci): Dato un lago, lo scopo `e cercare di stimare la grandezza N della popolazione dei pesci che vi vivono. Un modo pu`o essere il seguente: si pescano esattamente N1 pesci, i quali vengono in qualche modo marcati. In seguito, dopo aver permesso un mescolamento, si esegue un’ulteriore pesca, di n pesci. Si nota che fra questi ve ne sono n1 marcati. Vogliamo capire quale sia il valore di N pi`u plausibile. Nel nostro caso avremo un vettore casuale composto da una sola variabile, ovvero X = (X), la quale ha valori in N (ed `e quindi discreta) e restituisce i possibili valori di n1. La sua densit`a sar`a allora fornita in modo diretto e coincide con la funzione di verosimiglianza in quan- to vi `e una singola variabile casuale nel vettore. L’insieme dei parametri sar`a anch’esso N. Chiaramente vogliamo stimare il pi`u plausibile valore di θ = N. Avremo dunque:

35

L(N) := L(N,n1) = P[X = n1] = N1 = 300 e n = 80, se la nostra determinazione che massimizza la probabilit`a sarebbe ( nN1)((NN−Nnn−n) 1 1 )

. Per ottenuta N ∼ 1200. E `effettuare un esempio concreto: con fosse n1 = 30, allora il parametro quindi plausibile che nel lago viva una quantit`a di pesci che si aggira effettivamente intorno ai 1200 esemplari.

Definizione: Assumiamo che la funzione di verosimiglianza sia dervabile per il para- metro θ. Allora la funzione S(θ) := S(θ) = 0 `e chiamata Equazione di Stima.

∂θ∂l(θ, x) viene detta Score Function. L’equazione

Osservazione: Osserviamo che poiché la funzione densit`a di una qualsiasi variabile ca- suale `e sempre positiva o nulla, in quanto prodotto, lo dovr`a essere anche la parte di L(·,·) che non dipende da θ. Ne segue che, se vogliamo massimizzare la funzione di verosimi- glianza, possiamo direttamente limitarci a considerare solo i valori di θ ∈ Θ che rendano L(·,·) strettamente positiva per ciascuna determinazione x fissata o scelta. Dunque si pu`o restringere senza perdere generalit`a l’insieme Θ in modo da avere valori che non permet- tono a L(·,·) di annullarsi. S(θ) risulta quindi avere una buona definizione, in quanto non `e necessario effettuare ulteriori ipotesi su l(·,·), essendo:

S(θ) = ∂θ∂l(θ, x) = ∂θ ∂log(L(θ, x)) = L(θ, 1

x)

∂θ∂L(θ, x).

Definizione: La funzione di verosimiglianza induce uno stimatore del parametro θ. Esso sar`a arg{chiamato maxθ∈Θ{L(θ, Stimatore X)}} = di argMassima {maxθ∈Θ{l(θ, Verosimiglianza X)}}.

ed `e cos`ı definito: θˆn = θˆn( X) :=

Osservazioni:

• Da ora in poi l’argomento delle funzioni L(·,·) e l(·,·) verr`a interpretato a seconda della convenienza e del senso sia come (θ, x), ovvero come determinazione, oppure come (θ, X), ovvero vettore casuale. Osserviamo che in quest’ultimo caso, le funzioni L e l diventano esse stesse automaticamente variabili casuali o stimatori, che dir si voglia. Ci`o si ripercuote inevitabilmente sulle funzioni ad esse collegate, ad esempio su S(θ).

• Se L(·,·) o l(·,·) sono derivabili rispetto a θ, la funzione indipendente da θ che risolve ∀ x l’equazione di stima S(θ) = 0 fornisce effettivamente lo stimatore di mas- sima massima verosimiglianza. verosimiglianza Oppure, θˆn( X) equivalentemente, `e quello che soddisfa potremmo l’equazione dire che S(θˆn( lo X)stimatore ) = 0.

di

• In generale non vi `e garanzia che lo stimatore di massima verosimiglianza esista, op- pure, se esiste, che esso sia unico. Tuttavia nel caso di famiglie di densit`a che rispet- tino certe ipotesi di regolarit`a (per esempio le famiglie esponenziali) tale problema non si pone.

• Anche assumendo che tale stimatore esista e sia unico, non `e detto che sia sempre ottenibile analiticamente. Talvolta sar`a necessario ricorrere a metodi numerici per la risoluzione dell’equazione di stima.

Esempi:

36

1. Riprendiamo l’esempio precedente ove avevamo il campione casuale con variabili distribuite come poissoniane di parametro θ. La funzione di log-verosimiglianza era data da

l(θ, x) = log(θ)

∑ni=1

xi − nθ −

∑nlog(xi!) i=1 sicché otteniamo subito che

S(θ) = 1θ

∑ni=1

xi − n

Si deduce allora immediatamente l’equazione di stima 1θ ∑1ni=1 xi − n = 0 ⇒ θ = n

∑ni=1 xi Ma allora

θˆn( X) = Xn = n

1∑i=1

nXi

che `e la media campionaria.

2. Sia X := (X1,...,Xn) ∼ U[(0,θ)]. Allora

fX(x, θ) := 1θ1[0,θ](x)

Perci`o

L(θ, x) = θ1n

∏ni=11[0,θ](xi) = θ1n1[X(n),+∞](θ) ⇒ θˆn( X) = X(n).

3. Sia X := (X1,...,Xn) ∼ exp(β), β > 0. Allora fX(x, θ) := βe−βx1R+(x). Perci`o

L(θ, x) = (βne−β ∑ni=1 xi)1Rn+( x) ⇒ l(θ, x) = nlog(β) − β

∑ni=1

xi

Otterremo deduce che allora anche l’equazione in questo caso di stima θˆn = X0 n.

= S(β) = nβ − ∑ni=1 xi, da cui subito si

4. (Troncamento) Sia sempre X := (X1,...,Xn) ∼ exp(β). Naturalmente ciascuna variabile casuale ha come codominio i reali non negativi. Possiamo supporre di aver effettuato gli n rilevamenti dal campione casuale e di essere riusciti a individuarne esattamente m puntualmente (che senza perdita di generalit`a immagineremo essere i primi m), mentre dei restanti n −m immaginiamo di aver rilevato solamente che il loro valore supera una certa soglia fissta T > 0. Il campione contiene quindi due tipi di informazione da coniugare nella funzione di massima verosimiglianza, che avr`a stavolta una forma un po’ diversa. La indicheremo con L . Otteniamo:

37

mL (β, x) =

∏fXi(xi,β) ·

∏

nP[Xi > T] i=1i=m+1= ∏mi=1fXi(xi,β) ·

∏ ni=m+1(1 − FXi(T,β))

=

∏mβe−βxi ·

∏ ni=m+1∫ +∞

βe−βxidxi = βi=1me−β ∑mi=1 xi T · e−β(n−m)T

Da cui

l (β, x) := log(L (β, x)) = mlog(β) − β

∑mi=1

xi − β(n − m)T.

Inoltre possiamo definire anche qui una score function nel modo naturale:

S (β) := ∂β∂l (β, x)

da cui, uguagliando a 0 si pu`o ottenere l’equazione di stima

mβ −

∑i=1 mxi − (n − m)T = 0

Si deduce cos`ı lo stimatore di massima verosimiglianza con troncamento a T, dato da

β ˆn( X) =

∑mi=1 Xi + m (n − m)T

.

2.1 Efficienza

Dato uno stimatore Tn di un campione casuale X := (X1,...,Xn) possiamo partire dal concetto di cercare di stimatori errore quadratico che minimizzino medio MSEil pi`u θ(Tpossibile n) = Vartale θ(Tvalore. n)+Bθ2(TIl problema n). Lo scopo sar`a quello presenta alcu- ne difficolt`a: per fare banale Un( X) := θ0. un E `piccolo esempio, sia θ0 ∈ Θ e consideriamo il seguente stimatore ora evidente che se da una altro stimatore pu`o essere uniformemente migliore di parte Un, dall’altra MSEθ0(Uappare n) = 0, sicché nessun chiaro che di un siffatto stimatore non ci si possa attendere molto, e nemmeno fidare, in quanto esso ignora completamente tutta l’informazione contenuta nel vettore casuale. La difficolt`a di trova- re stimatori che abbiano errore quadratico medio minimo `e dunque legata a due aspetti principali: spesso la struttura di MSE `e complicata in quanto contiene aspetti legati al parametro θ; inoltre la classe degli stimatori competitori di θ `e quasi sempre troppo ampia. Cercheremo allora di semplificare il problema restringendo un po’ il campo: considerere- mo solo gli stimatori non distorti, per andare poi a cercare tra questi quelli con varianza minima.

38

Esempio: Sia X := (X1,...,Xn) ∼ P oisson(λ). In tal caso si verifica subito che E[X] = Var[X] = λ. Ne segue che sia lo stimatore media campionaria Xn sia lo sti- matore varianza campionaria Sn 2sono due stimatori non distorti di λ. Si ha tuttavia che Var[Xn] = λn ≤ λn

(1 + n−12nλ )

= Var[Sn2]. Preferiremo dunque la media campio- naria. Ma consideriamo ora il seguente stimatore cos`ı definito, per a ∈ [0,1] fissato, Wn,a( X) := aXn + (1 − a)Sn2. Anch’esso `e non distorto. Sorgono affrontare: ammesso che Xn sia migliore (i.e. con varianza pi`u cos`ı due difficolt`a piccola) di Sn2, da esso `e anche migliore di ogni stimatore Wn,a∀a oppure esso `e il migliore tra tutti gli stimatori non distorti di λ? Esiste un limite inferiore alla varianza? Se infatti esso esistesse, da- rebbe operativit`a alla scelta dello stimatore, in quanto se trovassimo uno stimatore che raggiunge tale limite, sapremo che non `e necessario cercare ulteriormente per migliorare le nostre possibilit`a. Ebbene, tale limite esiste sicuramente, sotto alcune ulteriori ipotesi di regolarit`a da addure alla non distorsione per gli stimatori.

2.1.1 Teorema di Rao-Cramér

Definizione: Una Famiglia Regolare `e una famiglia di densit`a che soddisfa le seguenti condizioni di regolarit`a:

1. Condizione di Indentificabilit`a: i valori delle densit`a sono distinti al variare del

parametro, ovvero θ = θ =⇒ fX(x, θ) = fX(x, θ ). 2. Le funzioni densit`a hanno supporto comune ∀θ ∈ Θ e il loro supporto non dipende

in alcun modo dal parametro θ. 3. Le funzioni sono di classe C2 rispetto alla variabile θ 4. Rispetto a θ, `e lecito lo scambio tra le derivate e l’integrale.

Definizione: Sia X = (X1,...,Xn) un campione casuale. Allora la funzione

I : Θ −→ R I(θ) := Eθ[S(θ)2] = Eθ[( ∂∂θ)2 ] l(θ, X)viene denominata Informazione di Fisher del campione casuale.

Osservazioni:

• Il prossimo teorema ci garantir`a nel caso di famiglie regolari che il limite inferiore della varianza inoltre che di un qualsiasi stimatore pi`u la varianza di uno non distorto stimatore si avvicina di θ a `e tale la quantit`a quantit`a, I(θ)1 pi`u . Notiamo `e signi- ficativa la sintesi dell’informazione circa θ contenuta nel vettore X realizzata dallo stimatore non distorto.

• Spesso si usano anche le seguenti notazioni per l’informazione di Fisher, ovvero I(θ), In(θ), nI1(θ). Infatti dato un vettore casuale qualsiasi X = (X1 ...,Xn), la sua funzione densit`a, ovvero L(·,·) non si spezza necessariamente nel prodotto delle den- sit`a di ciascuna componente Xi. Ci`o avviene invece nel caso in cui tutte le variabili

39

casuali siamo siano indipendenti: di fronte ad un in campione tal caso casuale, si pu`o scrivere allora le variabili I X(θ) casuali = ∑ni=1 sono IXi(θ). addirittura Se poi

i.i.d., e di conseguenza I X(θ) = nIX1(θ), da cui la notazione.

• Si pu`o dimostrare facilmente che, dato un campione casuale X := (X1,...,Xn) con densit`a nella famiglia regolare, vale la seguente uguaglianza:

Eθ[( ∂∂θ)2 ]

l(θ, X)= −Eθ[ ∂θ∂2

2l(θ, X)]

In effetti, come gi`a visto, si ha: ottiene subito che:

∂θ∂L(θ, X) = L(θ, X) ∂θ∂l(θ, X). Perci`o, derivando si

∂θ∂2

2L(θ, X) = L(θ, X) ∂θ∂2

2l(θ, X) + ∂θ∂L(θ, X) ∂θ∂l(θ, X) = L(θ, X) ∂θ∂2

2l(θ, X) + L(θ, X)( ∂∂θ)2

l(θ, X).

Si pu`o quindi ricavare l’asserto baster`a dunque ( ∂θ∂verificare l(θ, X))2 che = il L(θ, valore 1

X)

∂θ∂2

2L(θ, di X) − ∂θ∂2

2l(θ, X). Per provare aspettazione del primo addendo del secondo termine dell’uguaglianza sia nullo. Si ha:

E[ L(θ, 1

X)

∂θ∂2

2L(θ, X)]

∫=

Rn

∂θ∂2

2L(θ, x) · L(θ, x) · d x

=

1

∫Rn

L(θ, x) ∂θ∂2

2L(θ, x) · d x

= ∂2 ∂θ2

∫Rn L(θ, x) · d x = ∂θ∂2

21=0.

Definizione: Sia X := (X1,...,Xn) un campione casuale e Tn = Tn( X), Vn = Vn( X) due stimatori non distorti di θ. Allora:

1

• • • Diremo Efficienza assoluta Diremo Efficienza relativa Diremo che Tn `e Efficiente o di Bahadur di Tn e Vn se eff(Tn) di Tn il valore il valore eff(Tn,V= 1. Nel caso in eff(Tn) :=

Varn) cui := eff(TVarVarθθn) [T[Vn] n]> .

I(θ) θ[Tn]1 lo .

stimatore Tn si dir`a anche Super-Efficiente. In generale, si dir`a che Tn `e pi`u (meno) efficiente di Vn se eff(Tn,Vn) < (>)1.

• Diremo che Tn `e Asintoticamente Efficiente se limn→∞ eff(Tn) = 1.

Teorema 18 (di Rao-Cramér). Sia X := (X1,...,Xn) un campione casuale di densit`a f poi X( x, g θ) appartenente : Θ −→ Θ alla famiglia regolare con θ ∈ Θ ⊂ R un insieme di parametri. Sia una funzione derivabile e assumiamo l’informazione di Fisher I(θ) = 0 ∀θ ∈ Θ. Varθ[Tn] Allora, ≥ (g (θ)per )2 · qualsiasi I(θ)1 .

stimatore Tn = Tn( X) non distorto del parametro g(θ), vale

40

Dimostrazione. g(θ) sotto = il Eparametro θ[Tn] = Poiché ∫Rθ n Te n( grazie Tx)f n `e Xuno ( alle x, θ)d stimatore ipotesi x = ∫di Rn non regolarit`a Tn( x)L(θ, distorto otteniamo:

x)d di x, g(θ), con abbiamo:

θ ∈ Θ. Perci`o, derivando

g (θ) = ∂∂θ

∫Rn Tn( x)( ∂θ∂L(θ, x))d x

=

∫Rn Tn( x)L(θ, x)d x = ∫Rn Tn( x)L(θ, x)( ∂θ∂l(θ, x))d x =

∫Rn Tn( x)( ∂∂θ)l(θ, x)f X( x, θ)d x = Eθ[Tn( X) · ∂∂θ]l(θ, X).

Osserviamo ora per prima cosa che: Eθ[ ∂θ∂l(θ, X)]

=

∫Rn

( ∂θ∂l(θ, x))L(θ, x)d x =

( ∂θ∂L(θ, x))d x = ∂∂θ

∫∫Rn

Rn L(θ, x)d x = ∂θ

∂∫Rn f X( x, θ)d x = ∂θ∂1=0

Ne seguono direttamente le due seguenti relazioni:

• Covθ [Tn( X), ∂θ∂l(θ, X)] = Eθ[Tn( X) · ∂θ∂l(θ, X)] − Eθ[Tn( X)] · Eθ[ ∂θ∂l(θ, X)] = Eθ[Tn( X) · ∂∂θl(θ, X)] − Eθ[Tn( X)] · 0 = Eθ[Tn( X) · ∂θ∂l(θ, X)] = g (θ);

• Varθ [ ∂θ∂l(θ, X)] = Eθ[( ∂θ∂l(θ, X))2 ]

− Eθ[ ∂θ∂l(θ, X)]2 = Eθ[( ∂∂θl(θ, X))2 ], di conseguenza Varθ [ ∂θ∂l(θ, X)] = In(θ).

D’altra parte, dalla disuguaglianza di Cauchy-Schwarz abbiamo:

(g (θ))2 = Covθ

[Tn( X), ∂θ∂l(θ, X)]2

≤ Varθ[Tn( X)] · Varθ

[ ∂∂θ]

l(θ, X)= Varθ[Tn( X)] · In(θ),

grazie alle relazioni appena introdotte. La tesi segue subito, ricordando che sia la varianza che l’informazione di Fisher sono quantit`a positive.

Controesempio: Le ipotesi di regolarit`a del teorema sono necessarie. Consideriamo infatti il cmpione casuale X := (X1,...,Xn) ∼ U([0,θ]). La sua densit`a non appartiene alla famiglia regolare in quanto ha il supporto dipendente dal parametro θ. Uno stimatore non distorto di θ di conseguenza `e abbiamo stimatore gi`a super-efficiente. visto essere Tn( La X) tesi := del n−1n X(n). teorema Tuttavia Var[Tn] < non `e dunque valida 1 I(θ) e in questo caso.

Lemma 1. Sotto le usuali condizioni di regolarit`a, esiste uno stimatore non distorto Tn di θ efficiente, ossia tale che la sua varianza (raggiunge il limite inferiore di Rao-Cramér, se e solo se S(θ) = ∂θ∂l(θ, X) = In(θ)Tn( X) − θ).

41

Dimostrazione. Grazie alla disuguaglianza di Cauchy-Schwarz abbiamo la seguente relazio- ne Covza sse Come 2θ[Tn( X), ∂θ∂vi `e linarit`a gi`a calcolato l(θ, X)] tra i nella due ≤ precedente termini, Varθ[Tn( ovvero X)]·Vardimostrazione, sse θ[ ∂θ∃a, ∂l(θ, b ∈ X)], il R valore tali nella che di quale ∂θ∂aspettazione l(θ, sussiste X) = a l’uguaglian- + del bTprimo n( X).

membro Eθ[bTn( X)] dell’uguaglianza = a + bθ ⇒ a `e = nullo, −bθ. perci`o Quindi 0 = ∂θ∂l(θ, Eθ[ X) ∂θ∂l(θ, = X)] b(Tn( = X) Eθ[a − θ). + bTSe n( moltiplichiamo X)] = Eθ[a] +

tutto per ∂θ∂l(θ, X) abbiamo che

( ∂θ∂l(θ, X))2 = bTn( X) ∂θ∂l(θ, X) − bθ ∂θ∂l(θ, X). Calco- lando infine nuovamente il valore di aspettazione e riprendendo alcuni risultati ottenuti dalla dimostrazione del teorema di Rao-Cramér abbiamo che: I(θ) = Eθ[( ∂θ∂l(θ, X))2 ]

= bEθ[Tn( X) ∂θ∂l(θ, X)] − bθEθ[ ∂θ∂l(θ, X)] = b · 1 − bθ · 0 e di conseguenza si ha b = I(θ), da cui si deduce immediatamente la tesi.

Esempio: Consideriamo ancora X := (X1,...,Xn) ∼ P oisson(λ). Sappiamo che la sua densit`a appartiene alla famiglia regolare. Avevamo introdotto ∀a ∈ [0,1] fissato gli sti- matori non distorti Wn,a( X) := aXn+(1−a)Sn 2e ci eravamo chiesti quale fosse il migliore. Ebbene, come gi`a tra visto, tutti che essi, la score la risposta function `e proprio `e data Wda n,1 S(λ) = X= n, −n+ la media λ

1∑campionaria. Infatti si ha,

ora teniamo calcoliamo che S(λ)=(XIn(λ) = n−λ)· −Eλ[ ∂λ∂nλ 2

2l(λ, = X)] = −Eλ[ dλdS(λ)] = (Xn−λ)·I(λ) e possiamo −Econcludere λ[− ni=1 λ12XXi n] = grazie = n( λ12 λ1·nλ Xil n − 1). Se = nλ, ot- Lemma 1.

Lemma 2. Sotto le usuali ipotesi di regolarit`a, sia che verosimiglianza esista uno stimatore di θ, allora Tn vale non Tdistorto n = θˆn.

di θ efficiente. I(θ) Se θˆ= n `e 0 lo ∀θ stimatore ∈ Θ e supponiamo di massima

Dimostrazione. Il limite inferiore di Rao-Cramér non `e una quantit`a nulla. Inoltre come gi`a osservato e grazie al Lemma 1 si ha:

0 = S(θˆn( X)) = (Tn( X) − θˆn( X))In(θˆn( X)),

da cui Tn( X) − θˆn( X) = 0 e quindi la tesi.

Controesempio: Non sempre lo stimatore di massima verosimiglianza `e anche stimatore efficiente, e dunque, per il Lemma 2, non sempre esiste uno stimatore efficiente. Sia infatti Ora, ∂θ∂X 2

2 log:= ((Xf(x, 1,...,Xθ)) n) ∼ fX(x, θ) := θxθ−1 · 1(0,1)(x) = − θ12 ⇒ I1(θ) = −E[− θ12] = 1θ2 campione casuale, ⇒ In(θ) = θn2. con θ > 0. Per`o S(θ) = ∂θ∂l(θ, X) = ∂∂θ log(∏ni=1 θXi

θ−1 )

= nθ + ∑ni=1 log(Xi). L’equazione di stima S(θ)=0ci fornisce allora θˆn( X) = ∑ni=1 n log(Xi), lo stimatore di massima verosimiglianza. Vogliamo ora trovare la sua distribuzione. Definiamo innanzi tutto il nuovo vettore casuale Y := (Y1,...,Yn) dove ∀i = 1..n si ha Yi := log(Xi). Osserviamo che il logaritmo `e una funzione monotona crescente, e possiamo applicare il teorema 1.1 per ottenere che la densit`a delle nuove variabili `e fY (y,θ) = θ(e−y)θ−1| − e−y| · 1R+(y) = θe−θy1R+(y), e θ > 0. Dunque, Y ∼ G(α = 1,β = 1θ). Poiché X `e un vettore indipendente, segue necessariamente

42

che anche Y lo sia; quindi, grazie alla propriet`a di riproducibilit`a della densit`a Gamma W := ∑ni=1 Yi ∼ G(α = n, β = 1θ). Si pu`o mostrare che:

E[W k] = (n + k − 1)!

θ(n − 1)! .

Ricordando che ˆθn = nW −1 possiamo calcolare subito i valori di aspettazione

• Eθ[ˆθn] = Eθ[nW −1] = nEθ[W −1] = nn−1θ = θ, perci`o `e stimatore distorto, anche se asintoticamente non distorto.

• E[(ˆθn)2] = E[n2W −2] = n2E[W −2] = θ2n2

(n−2)(n−1)

e dunque Var[ˆθn] = E[(ˆθn)2] − E[ˆθn]2 = n2θ2

(n−1)2(n−2)2 > 1 I(θ) = θn. Ne segue che ˆθn non `e stimatore efficiente di θ anche se eff(ˆθn) −−−→ n→∞ 1.

2.1.2 Estensione a un vettore di parametri:

Possiamo, al posto di un singolo parametro, andare a considerare un vettore di parametri θ := (θ1,...,θk) ∈ Θk, Θ ⊂ R che indicizza la distribuzione di una variabile casuale X. Ad esempio la distribuzione Gamma dipende da due parametri solitamente indicati con α e β. In particolare, modellare un fenomeno con un numero di parametri che sia il pi`u piccolo possibile assume un valore importante per quanto riguarda la stabilit`a degli stima- tori. A ci`o `e stato dato il nome piuttosto eloquente di Principio di Parsimonia. Nel caso di un vettore di parametri si potr`a allora estendere il concetto di Informazione di Fisher ottenendo una matrice.

Definizione. Sia X un campione casuale e θ := (θ1,...,θk) un campione di parame- tri. Allora la Matrice di Informazione di Fisher `e la matrice I( θ) ∈ M(k × k,R) il cui i-j-esimo elemento `e dato dal numero

E θ[ ∂∂θil( θ, X) · ∂∂θj l( θ, X)].

Proposizione. (\*) Per ogni vettore casuale X si ha che la matrice di informazione di Fisher `e simmetrica e semidefinita positiva.

Dimostrazione. (\*) Il fatto che sia simmetrica `e pressoché immediato e viene diretta- mente dalla definizione. Per mostrare che `e anche semidefinita positiva, sia w( θ) := ( ∂∂θ1l( θ, X),..., ∂∂θkl( θ, X)). Si vede subito che I( θ) = E θ[w( θ)t·w( θ)]. Sia ora u ∈ Rk\{0}. Dobbiamo mostrare che (u · I( θ) · ut) ≥ 0. Ebbene, sfruttando la linearit`a del valore di aspettazione, otteniamo che:

uI( θ)ut = uE θ[w( θ)t · w( θ)]ut = E θ[u · w( θ)t · w( θ) · ut]

= E θ[u · w( θ)t · ((u · w( θ)t)t] = E θ[ u · w( θ)t 2] ≥ 0.

43

Osservazione: Osserviamo che le ipotesi di regolarit`a definite per il caso unidimen- sionale, possono essere espanse al caso k-dimensionale nel modo pi`u naturale, ovvero supponendo che esse valgano per ciascuno dei parametri θi, ∀1 ≤ i ≤ k. Ebbene, sotto le usuali ipotesi di regolarit`a, si ottiene facilmente con quanto gi`a mostrato che I(θ)ij = −E θdi [ ∂2

∂θi∂θjl( θ, x)]. Osserviamo informazione: essendo le densit`a di che vi `e coerenza con la simmetria della matrice classe C2, vale il teorema di Schwarz sullo scambio delle derivate.

Lemma. (\*) Siano A ∈ M(n×n,R) una matrice simmetrica e definita positiva, e b ∈ Rn. Allora, se definiamo la funzione

f : Rn −→ R

f(x) := x · A · xt − 2b · xt

abbiamo che f ha un unico punto di minimo ˆx := b · A−1.

Dimostrazione. (\*) Scriviamo per semplicit`a x = (x1,...,xn), b = (b1,...,bn) e A = (aij)ij. utilizzeremo il metodo della matrice hessiana. Abbiamo che:

f(x) =

∑ni,j=1xiaijxj − 2

∑nh=1bhxh =

∑nk=1akkx2k + 2

∑ni=2

∑i−1j=1

∑nh=1bhxh

dove l’ultima Definendo ora uguaglianza ∑01 := 0, calcoliamo viene direttamente la r-esima derivata, dal fatto ∀1 che ≤ A r `e ≤ una n:

matrice simmetrica.

∂∂xrf(x) = 2arrxr + 2

xiaijxj − 2

∑r−1j=1

∑ ni=r+1xiair − 2br

= 2arrxr + 2

arjxj + 2

∑r−1j=1

∑ ni=r+1xiari − 2br

= 2arrxr + 2

arjxj + 2

∑ ns=1,s =rasrxs − 2br

= 2

∑ns=1

arsxs − 2br = 2(Ar · xt) − 2br,

ove Ar `e la r-esima riga della matrice A. Ora, per trovare i possibili punti stazionari sar`a necessario eguagliare a 0 tutte le derivate parziali e risolvere il sistema. Non avendo termini quadratici esso sar`a lineare:



... ∂x∂1f(x)=0

∂∂xnf(x)=0



2(A...1 · xt) − 2b1 = 0 2(An · xt) − 2bn = 0  A...1 · xt = b1 ⇔

⇔

⇔ A · xt = bt. An · xt = bn

Allora, poiché la matrice A `e invertibile, otterremo l’unico punto stazionario ˆxt = A−1 ·bt, ed essendo la matrice simmetrica, sar`a ˆx = b · A−1. Per verificare adesso che si tratta di

44

un punto di minimo, andiamo a calcolare la matrice hessiana di tutte le derivate seconde. Palesemente f `e una funzione di classe C∞, ragion per cui deve valere il teorema di Schwarz sullo scambio delle derivate seconde. In generale ∀1 ≤ s, t ≤ n, si avr`a:

∂2 ∂xs∂xtf(x) = ∂∂xs

(2

atixi − 2bt)

= 2ats.

Ne segue quindi che la matrice hessiana di f sar`a ∀x ∈ Rn data da 2A. Essa risulta quindi definita positiva e conferma di conseguenza che ˆx `e un punto di minimo.

Teorema 19 (di Rao-Cramér). Sia X := (X1,...,Xn) un campione casuale di densit`a f poi X( x, g θ) appartenente : Θ −→ R una alla famiglia funzione derivabile Fisher I(θ) sia invertibile ∀θ ∈ Θ. regolare e θ ∈ Θ ⊂ Rk un insieme di parametri. e Definiamo assumiamo ora il vettore che γ( la θ) matrice := ( ∂θ∂1di informazione g( θ),..., ∂θn∂Sia g( θ)di ). Allora, per ogni stimatore T = T( X) non distorto del numero g( θ) si ha che Var[T] ≥ γ( θ) · I−1( θ) · γ( θ)t.

Dimostrazione. (\*) La dimostrazione si articola sfruttando alcuni passaggi gi`a utilizzati nel caso unidimensionale. Innanzi tutto ∀1 ≤ j ≤ k si ha:

∂∂θj g( θ) = ∂∂θj ( ∂∂xtE θ[T] = ∂∂θj

) f(x)= ∂∂xs

∑ni=1

∫Rn T( x)f X( x, θ)d x = ∂∂θj

∫Rn T( x)L( θ, x)d x

=

∫Rn ∫T( x)( ∂Rn [T( x) ∂θ∂j ∂θj )l( θ, x)L( θ, x)d x = E θT( x) ·

L( θ, x)d x = ( ∂∂θj )]l( θ, x).

Osserviamo allora che essendo

E θ[ ∂∂θj ]

l( θ, X)=

∫∂( ∂∂θj = Rn

∂∫∂θj ∂θj

∫)l( θ, x)L( θ, x)d x =

L( θ, x)d x Rn Rn L( θ, x)d x = ∂∂θj

∫Rn f X( x, θ)d x = ∂∂θj 1=0,

Cov ∂θ∂j l( θ

θ, [TX)n( X), ] − E ∂θ∂θj[l( Tn( θ, X)X)] ] · = 0 E θ= [E Tθn( [TX)· n( X) ∂θ∂j· l( ∂θθ, ∂jl( X)θ, ]X)−E ] θ[Tn( X)]·E θ= ∂θ∂j g( θ).

[ ∂θ∂jl( θ, X)] = E θ[Tn( X)·

Fatto ∑ki=1 cci`o, i ∂θ∂il( sia θ, ora x). c E `= chiaro (c1,...,cche k) ∈ Rk e definiamo W( X, θ) sar`a allora ora una la seguente funzione: W( x, θ) := variabile casuale. Vogliamo cal- colare il valore di Var[T( X) − W( X, θ)]. Per prima cosa osserviamo che, sfruttando la

45

linearit`a del valore atteso,

Var[T − W] = E[(T − W)2] − E[(T − W)]2

= E[T 2 − 2TW + W 2] − (E[T] − E[W])2 = E[T 2] − 2E[TW] + E[W 2] − E[T]2 + 2E[T]E[W] − E[W]2 = Var[T] − 2 Cov[T,W] + Var[W]

Adesso osserviamo invece che i conti si possono semplificare in quanto

E[W( X, θ)] = E[ ∑ki=1

ci ∂∂θil( θ, X)]

=

∑ki=1

ciE[ ∂θ∂il( θ, X)]

=

∑ki=1

ci · 0=0.

Perci`o otteniamo subito che:

Cov[T( X),W( X, θ)] = E[T( X)W( X, θ)] − E[T( X)]E[W( X, θ)] = E[T( X)W( X, θ)] = E[ ∑ki=1

ciT( X) ∂θ∂i]

l( θ, X)=

ciE[T( X) ∂θ∂i] l( θ, X)=

∑ki=1 ∑ki=1

ci Cov θ

[Tn( X), ∂θ∂il( θ, X)] =

∑ki=1

ci ∂θ∂ig( θ)

= c · γ( θ)t;

Var[W( X, θ)] = E[(W( X, θ))2] − E[W( X, θ)]2 = E[(W( X, θ))2]

= E[( ∑ki=1

ci ∂∂θil( θ, X))2] = E[ ∑ki=1

(ci ∂∂θil( θ, X))( ∂θ∂j l( θ, X)cj)]

=

∑kj=1 ciE[ ∂θ∂il( θ, X) · ∂∂θj ]l( θ, X)cj

=

∑ki=1

∑kj=1 ∑ki=1

∑kj=1

ciIij( θ)cj = c · I( θ) · ( c)t.

Osservando infine che la varianza di una qualsiasi variabile casuale `e un numero sempre positivo, otteniamo la relazione definitiva:

0 ≤ Var[T( X) − W( X, θ)] = Var[T( X)] + c · I( θ) · ( c)t − 2 c · γ( θ)t.

Poiché ci`o deve valere ∀ c ∈ Rk possiamo allore scrivere:

0 ≤ min c∈Rk{Var[T( X) − W( X, θ)]} = Var[T( X)] + min c∈Rk{ c · I( θ) · ( c)t − 2 c · γ( θ)t}.

46

Ora, la matrice di informazione di Fisher `e per ipotesi invertibile, sicché, grazie alla pre- cedente proposizione, essa `e quindi definita positiva. Possiamo allora applicare il lemma, secondo cui il minimizzatore `e:

ˆc := arg{ min c∈Rk{ c · I( θ) · ( c)t − 2 c · γ( θ)t}} = γ( θ) · I−1( θ).

Sostituendo allora ˆc nella varianza e ricordando che I( θ) `e simmetrica (e quindi anche la sua inversa) si ottiene infine la tesi:

0 ≤ min {Var[T( X) − W( X, θ)]} = Var[T( X)] + ˆc · I( θ) · (ˆc)t − 2ˆc · γ( θ)t c∈Rk= Var[T( X)] − γ( θ) · I−1( θ) · γ( θ)t.

Corollario. Nelle ipotesi del teorema di Rao-Cramér, se ∀1 ≤ i, j ≤ k abbiamo che Tj = Tj( X) `e uno stimatore non distorto del parametro θj, allora Var[Tj( X)] ≥ Ijj −1 ( θ).

Esempio: Sia X := (X1,...,Xn) ∼ N(μ, σ2) un campione casuale. Allora

L(μ, σ2, x) =

∏ni=1

√2πσ1 2e− 2σ12 (x1−μ)2 = (2πσ2)− n2 e

∑ni=1(xi−μ)2

La funzione di log-verosimiglianza sar`a allora

l(μ, σ2, x) = −n2 log(2πσ2) − 1σ2

2σ12

∑ni=1(xi − μ)2.

Facilmente possiamo calcolare la matrice di informazione di Fisher, che risulter`a:

I( θ) =

( nσ0 2 0 n2σ4

( σ2n 0 0 2σ4n

) Osserviamo anche che valori dell’inversa della ), e quindi

Var[Xmatrice n] di = Fisher. σn 2e Var[Sn2] = Si deduce 2σ4

n−1 allora > dal 2σn 4, teorema dove sono stati calcolati i di Rao-Cramér che la media campionaria `e stimatore efficiente per μ, mentre non lo `e la varianza campionaria per σ2, anche se lo `e asintoticamente. Coerentemente, gli stimatori di massima verosimi- glianza si possono ottenere risolvendo il sistema composto dalle rispettive score-functions dei parametri, { ∂μ∂σ∂∂2l(μ, l(μ, σσ22, , X)=0 X)=0

che ci restituisce le due soluzioni ˆμn = Xn e σˆ2n valori non diagonali della matrice di Fisher sono nulli = perché n−1nSn2. Notiamo infine che i due ci troviamo in distribuzione normale, ove la non correlazione implica anche l’indipendenza.

Osservazione: Sia Z ∼ N(0,1) una variabile casuale. Allora vale la seguente relazione:

μ2s := E[Z2s] = (2s)! 2ss!

47

(I( θ))−1 =

Essa risulta comoda per il calcolo dei momenti delle variabili normali, tenendo conto che X := μ + σZ =⇒ X ∼ N(μ, σ2).

Esempio: Sia X := (X1,...,Xn) ∼ fX(x, η) := ηe−η(x−3)1[3,+∞], con η > 0. Voglia- mo individuare calcolare possibilmente il limite inferiore un siffatto di Rao-Cramér stimatore per e, dopo uno stimatore averlo trovato, non distorto calcolare di se g(η) esso := sia η1,

o meno efficiente. In base al teorema di Rao-Cramér si ha che

I(g(η)) = (g (η))2 I(η) 1

= η14

1 I(η).

Per calcolare I(η), troviamo innanzi tutto la funzione di log-verosimiglianza. Si ha per prima cosa che:L(η, x) = ηne−η ∑ni=1(xi−3) ⇒ l(η, x) = nlog(η) − η

∑ni=1(xi − 3) Possiamo calcolare adesso

I(η) = −Eη[ ∂η∂2

2l(η, x)]

= −Eη[− ηn2] = ηn2.

Ne segue subito che il limite inferiore cercato sar`a allora

I(g(η)) = 1η4 · η2n = nη12.

Per trovare un possibile stimatore, ricordiamo la relazione gi`a dimostrata durante la dimostrazione del teorema di Rao-Cramér E[S(η)] = 0. Ne segue che 0 = Eη[ ] Eη[nη − ∑ni=1(Xi − 3)] = nEη[η 1− (Xn − 3)] = n( )

∂η∂l(η, X)= η 1− Eη[Xn − 3]da cui

Tn( X) := (Xn − 3)

`e lo stimatore cercato. Ora, `e semplice calcolare che

Var[Tn] = Var[Xn] = n12

∑ni=1

Var[Xi] = n 1Var[X] = nη12

Ne segue che lo stimatore cercato `e effettivamente anche efficiente.

2.2 Sufficienza

Introduzione alla sufficienza (scritta da Michele Nardin)

Rinfreschiamo il concetto di distribuzione condizionata:

Definizione 10. (Caso discreto) Siano X, Y variabili casuali discrete e supponiamo P(X=x) = 0. La distribuzione condizionata di Y dato X=x `e

PY |X := P(Y = y|X = x) = P(Y P(X = y ∩ = X x)

= x)

48

(Caso continuo) Siano X, Y variabili casuali definite su (Ω,E,P), con funzioni di densit`a fX,fY rispettivamente, e densit`a congiunta fX,Y . Ovviamente `e assurdo richiedere che P(X=x) = 0. Allora si ragiona tramite densit`a condizionate: supponendo che fX(x) = 0

fY |X(y|X = x) := fX,Y (x, y)

fX(x) si dice densit`a condizionata di Y dato X=x

NB: fY |X(y|X = x) = fY (y) per ogni y se e solo se X e Y sono indipendenti.

La sufficienza di una statistica definisce formalmente la capacit`a di tale funzione di rap- presentare in maniera sintetica l’informazione contenuta nel campione, senza perdita di informazione rilevante.

Intuitivamente si pu`o pensare alla propriet`a di conservazione dell’informazione rilevante in questo modo: qualsiasi altra statistica, calcolata a partire dallo stesso campione, non porta pi`u informazioni rispetto a θ di quante ne abbia gi`a portato la statistica sufficiente. Per questo, banalmente, l’intero campione `e sicuramente una statistica sufficiente (molto poco sintetica, ma a volte non c’`e di meglio).

Sia (X1,X2, ..., Xn) un campione casuale da una distribuzione avente funzione di ri- partizione FX(x, θ). Sia inoltre Tn(X1, ..., Xn) uno stimatore di θ avente funzione di ripartizione FTn(t, θ).

Definizione 11. La statistica Tn viene detta sufficiente per θ se la distribuzione di X1, ..., Xn condizionata a Tn = tn non dipende da θ.

Abbiamo quindi formalizzato le richieste fatte sopra: il fatto che la distribuzione con- dizionata (del vettore allo stimatore) non dipenda da θ implica di fatto l’impossibilit`a di perdere informazioni rilevanti su θ stesso, e cio`e, lo stimatore Tn contiene tutte le infor- mazioni necessarie riguardanti θ per fare inferenza sul parametro incognito. In altre parole, le informazioni contenute in Tn riguardo a θ sono le stesse contenute nel- l’intero campione. Vedremo che la sufficienza pu`o essere verificata usando il teorema di fattorizzazione di Neyman, che di fatto fornisce una caratterizzazione ’pi`u comoda’ da verificare (rispetto all’uso brutale della definizione) per assicurarsi che una data statistica abbia la propriet`a desiderata.

Sufficienza (Lezione del 06/05, Scritta da Marco Peruzzetto)

Definizione. Sia X := (X1,...,Xn) un campione casuale avente densit`a fX(x, θ) e sia Tn una statistica. Allora Tn `e detta Statistica Sufficiente per θ se la densita del vettore X condizionata a Tn( X) = tn non dipende da θ.

Esempi:

1. Sia X = (X1,...,Xn) ∼ b(1,p). Allora Tn( X) := ∑ni=1 Xi `e una statistica sufficiente

per p. Infatti, in base alla definizione di statistica sufficiente, si ottiene che

f X|Tn( X)( x, tn) := f( X,Tn( X))( x, tn)

fTn( X)(tn) = ptn(1 − p)n−tn ( ntn)ptn(1 − p)n − tn = 1( ntn)

49

e dunque non dipende dal parametro θ. 2. Sia X := (X1,...,Xn) ∼ G(α = 2,β). Allora Tn( X) := ∑ni=1 Xi `e sufficiente per β. Infatti, si ha innanzi tutto che Tn( X) ∼ G(2n, β) grazie all’indipendenza di X e alla propriet`a di riproducibilit`a di G. Allora

f X|Tn( X)( x, tn) := f( X,Tn( X))( x, tn)

fTn( X)(tn) = f X( x)

fTn( X)(tn)

=

∏ni=1 Γ(2)β1 2x2−1 i e− β 1xi 1 Γ(2n)β2nt2n−1 n e− β 1tn

e si vede quindi che si semplificano tutti i termini in β, per cui `e sufficiente.

3. Sia X := (X1,...,Xn) un campione casuale avente funzione di densit`a fX(x, θ) := e−(x−θ)1(θ,+∞)(x), θ > 0. Allora la statistica d’ordine X(1) `e sufficiente per θ. Infatti:

fX(1)(x, θ) = ne−n(x−θ)

ed otteniamo subito che

f X|X(1)( x, x(1),θ) = ff XX(1)( x, (x(1)θ)

) =

∏ni=1 e−(xi−θ)1(θ,+∞)(xi) ne−n(x(1)−θ)1(θ,+∞)(x(1)) = e− ∑ni=1 xi ne−nx(1)

che non dipende dal parametro θ.

Teorema 20 (di Fattorizzazione (Neyman)). Sia X := (X1,...,Xn) un campione casuale avente densit`a fX(x, θ). Allora Tn = Tn( X) `e statistica sufficiente per θ se e solo se esistono due funzioni non negative g,h, con

(·) h = h( x) e NON dipende da θ

(·) g = g(θ, tn( x)) dipende da θ e da (X1,...,Xn) solo attraverso tn

tali che ∀ x0 ∈ X e ∀θ0 ∈ Θ, la funzione di verosimiglianza possa essere scritta come

L(θ0, x0) = h( x0) · g(θ0,tn( x0))

Dimostrazione. [⇒]. Sia ha:

L(θ, x) = f X|Tn=tn( x, tn,θ) · fTn(tn,θ) = f X|Tn=tn( x, tn) · fTn(tn,θ),

dove la prima uguaglianza viene dalla definizione di densit`a condizionata, mentre la secon- da viene dalla sufficienza di Tn, per cui la densit`a condizionata non dipende dal parametro θ. Possiamo scegliere allora h = [⇐]. Dimostreremo qui il caso in f cui X|Til n=tcampione n e g = fTn.

casuale sia composto da variabili discrete.

50

Sia quindi L(θ, x) = h( x)·g(θ, tn( x)) e definiamo il nuovo insieme Atn := { x ∈ X : Tn( x) = tn}. Allora

fTn(tn) = ∑ x∈Atn

L(θ, x) = ∑ x∈Atn

h( x) · g(θ, tn( x)) = g(θ, tn( x)) · ∑ x∈Atn

h( x),

perci`o, per definizione di densit`a condizionata si ottiene che:

f X|Tn=tn( x, tn) = fL(θ, Tn(tx)

n) = g(θ, h( tn( x) x)) · g(θ, · ∑ x∈Atn tn( x))

h( x) = ∑ x∈Atn h( x) h( x).

Essa non dipende quindi dal parametro θ, sicché `e sufficiente. La dimostrazione per il caso continuo `e analoga.

51

Attenzione: da questo punto in poi, le dispense sono state scritte durante la sessione d’esame. Per questo motivo, gli autori non hanno potuto dedicarvi il tempo e l’attenzione che un lavoro di questo tipo richiede. Crediamo che possano essere comunque pi`u precise e complete degli appunti presi in classe, ma non garantiamo nulla. Appena possibile le dispense verranno ricontrollate e rese definitive.

52

Lezione del 10/05, ultima modifica 10/06, Andrea Gadotti

Il teorema di fattorizzazione costituisce un criterio per l’individuazione, se esiste, di una statistica sufficiente.

Esempio Sia (X1, ..., Xn) da N(μ, σ2).

(a) se μ `e noto, allora cerco una statistica sufficiente per σ2.

Tn(X1, ..., Xn) =

∑ni=1(Xi − μ)2

(b) Se σ2 `e noto, allora cerco una statistica sufficiente per μ.

(c) Se μ e σ2 non sono noti, allora cerco una statistica congiuntamente sufficiente

per μ e σ2. Sn(X1, ..., Xn) =

(∑ni=1 Xi,∑ni=1(Xi − X2n)).

Risulta quindi evidente che il concetto di statistica sufficiente `e problem depen- dent.

Esempio 1. Sia (X1, ..., Xn) da U(θ, θ + 1) con θ ∈ R. Vogliamo trovare una statistica sufficiente per θ. Calcoliamo innanzitutto la funzione di verosimiglianza:

L(θ;x) =

∏ni=1

1[θ;θ+1](xi) (2.1)

1[θ;θ+1](xi) (2.1)

=

∏ni=1

1R(xi)1[θ;θ+1](x(n))1[θ;θ+1](x(1)) (2.2)

1R(xi)1[θ;θ+1](x(n))1[θ;θ+1](x(1)) (2.2)

=

∏ni=1

1R(xi)1[X(n)−1;X(n)](θ) (2.3)

1R(xi)1[X(n)−1;X(n)](θ) (2.3)

Notiamo che l’ultima espressione `e il prodotto di due funzioni dove la prima dipende solo dal campione, mentre la seconda dipende sia dal parametro θ che dal campione. Quindi grazie al teorema di fattorizzazione di Neyman abbiamo che (X(1),X(n)) `e statistica sufficiente per U. Nota: poiché U(θ,θ+1) non appartiene alla famiglia regolare non `e necessariamente vero che dimensione statistica sufficiente = dimensione vettore parametri. Infatti in questo caso abbiamo 2 = 1.

Osservazione 7. Spesso vediamo il campione casuale come diverso da una statistica, che `e funzione del campione. Ma esso porta in s`e tutta l’informazione disponibile relativamente al vettore parametrico. Anche il campione stesso `e una statistica, in particolare `e una statistica sufficiente, l’unico problema `e che non `e per nulla sintetico.

53

Problema: esiste una statistica sufficiente che sia migliore (ovvero pi`u sintetica) delle altre, a parit`a di informazione contenuta circa θ? In effetti una tale statistica esiste e viene detta statistica sufficiente minimale.

2.2.1 Statistiche sufficienti minimali

Esempio introduttivo Pensiamo di replicare n = 4 volte una prova bernoulliana con probabilit`a di successo p. Consideriamo lo spazio campionario κ, composto da 24 = 16 elementi. Vogliamo fare inferenza su p. Definiamo di seguito tre statistiche differenti:

1) Y1 := risultato della prima prova

2) Y2 := numero di successi nelle quattro prove

3) Y3 := (Y1,Y2)

A questo punto notiamo che:

1) Y1 non `e una statistica sufficiente per p (il fatto che ci sia stato un successo o

meno nella prima prova non ci d`a alcuna informazione rispetto a p)

2) Y2 `e statistica sufficiente per p (cfr. Teorema di fattorizzazione) e consiste in

un unico valore, ovvero la somma degli elementi del campione 3) Y3 consiste in due valori, ovvero `e un vettore di dimensione 2. `E statistica sufficiente per p, ma `e eccessivamente raffinata per il problema che vogliamo affrontare, poiché l’apporto di Y1 `e inutile. Infatti Y3 non `e statistica sufficiente minimale per p.

Definizione 12. Una statistica Tn(X1, ..., Xn) `e detta statistica sufficiente minimale per θ se:

1) `e statistica sufficiente per θ

2) assume valori distinti solamente in punti dello spazio campionario κ a cui

corrispondono verosimiglianze non equivalenti, ovvero se ∀x1,x2 ∈ κ vale

Tn(x1) = Tn(x2) ⇐⇒ L(θ, x1) = L(θ, x2) ∀θ ∈ Θ

Teorema 21 (di Lehmann-Scheffé). Sia (X1, ..., Xn) un campione casuale da una distribuzione di densit`a f(x;θ) e sia Sn = Sn(X1, ..., Xn) tale che, prese le deter- minazioni campionarie x e y, il rapporto tra le funzioni di verosimiglianza valutate in x e y, `e funzione che non dipende da θ se e solo se Sn(x) = Sn(y). Allora Sn(X1, ..., Xn) `e statistica sufficiente minimale per θ. Ovvero:

L(θ;x) L(θ;y) = m(x, y) ⇐⇒ Sn(x) = Sn(y)

=⇒ Sn(X1, ..., Xn) `e statistica sufficiente minimale per θ.

54

Esempio 2. Sia (X1, ..., Xn) da X ∼ Γ(α, β),α,β > 0. Vogliamo trovare una statistica congiuntamente sufficiente e minimale per α e β:

1) Cerchiamo una statistica sufficiente:

L(α, β;x) =

((Γ(α)βα)−n (e− β

1∑ni=1 xi)( ∏ni=1

xi))

·

( ∏ni=1

1) R+(xi)`e Osserviamo Quindi statistica per congiuntamente il che teorema abbiamo di scritto fattorizzazione sufficiente L(α, β;x) per abbiamo come α e β.

g (α, che β;S∑n := ni=1 (x∑i,ni=1 ∏ni=1 xi,x∏i)·h(x). ni=1 xi)

2) Verifichiamo Sn(y), ovvero ora (∑che ni=1 xtale i,∏statistica ni=1 xi)=(`e ∑anche ni=1 yminimale. i,∏ni=1 yi). Supponiamo Allora:

che Sn(x) =

L(α, L(α, β;x)

β;y) = e

(∏i

xiyi) β(1∑i xi−∑i yi)

= e0

(∏(1))α

= 1

i

Per il teorema di Lehmann-Scheffé concludiamo che Sn `e statistica sufficiente minimale. Nota: in realt`a noi abbiamo verificato che vale solo una delle due implicazioni richieste dalle ipotesi del teorema (quella da destra a sinistra). Questo `e ci`o che `e stato fatto a lezione, e per il momento scriviamo solo questo.

2.2.2 Principio di verosimiglianza

La verosimiglianza combina due tipi di informazione:

• informazione pre-sperimentale espressa attraverso il modello statistico scelto per descrivere il fenomeno di indagine

Fθ = {f(x;θ) : θ ∈ Θ ⊆ Rk}

• informazione sperimentale contenuta in quella che `e la determinazione campio- naria x = (x1,...,xn)

L(θ;x) =

∏ni=1

fXi(xi;θ)

`e Nota: definita In generale anche nel L(θ;x) caso in = cui fX1,...,Xnle variabili (x1,...,xcasuali n;θ), non densit`a siano congiunta, indipendenti ovvero ed equidistribuite. Quando invece lo sono, possiamo esprimere L(θ;x) nella solita forma, ovvero come produttoria.

Principio {f(x;θ) : di θ ∈ verosimiglianza Θ ⊆ Rk}, Con riferimento due punti x1 ad un dato modello stocastico Fθ = e x2 ∈ κ tali che L(θ;x2) = L(θ;x2) devono condurre alle medesime conclusioni inferenziali circa θ.

55

Teorema 22. L’informazione di Fisher fornita da uno stimatore Tn basato su una statistica sufficiente Wn coincide con l’informazione fornita dall’intero campione

Dimostrazione. Dal teorema di fattorizzazione l’informazione di Fisher sar`a funzione del campione solo attraversoddθ ln [g(Wn);θ] Wn = Tn(x) dal momento che la derivata rispetto a θ di h `e nulla ovvero ddθ ln [h] = 0, dunque

I(θ) = Eθ

[ ddθ ]2 ln(g(Wn);θ)dove il membro di sinistra `e l’informazione associata all’intero campione e il membro di destra `e l’informazione di Fisher restituita dalla statistica sufficiente Wn. Questo implica che I(Wn) `e ugualmente rappresentativo per l’informazione di Fisher di quanto non lo sia l’intero campione.

Esempio 3. Sia (X1,...,Xn) da N(μ, σ2 = 1), vogliamo individuare una statistica sufficiente minimale per μ e la relativa informazione di Fisher.

1) Notiamo che

Wn :=

∑ni=1

Xi ∼ N(nμ, n)

`e statistica sufficiente (si verifica subito usando il teorema di fattorizzazione). Vogliamo mostrare che l’informazione di Fisher che si ottiene usando solo Wn `e la stessa che si trova usando l’intero campione.

2) Calcoliamo I(μ) supponendo il solo valore wn = al punto 1). Iniziamo ∑ni=1 calcolando xi di avere in mano, anziché l’intero campione (x1, ..., xn),

(di cui conosciamo la distribuzione, come visto la funzione di verosimiglianza di wn (che coincide con la sua densit`a, essendo un unico valore e non un vettore):

L(μ;wn) = √2πn1e− 2n1(wn−nμ)2 A questo punto, calcolando la funzione di log-verosimiglianza, con un po’ di conti si trova facilmente che IWn(μ) = n. 3) Calcoliamo I(μ) supponendo di avere in mano l’intero campione. Iniziamo

trovando la funzione di verosimiglianza di (x1, ..., xn):

L(μ;x) =

∏ni=1

( √12πe− 12(xi−μ)2)

Come prima, con un po’ di conti si ottiene Ix(μ) = n, ovvero le due informa- zioni di Fisher coincidono.

Lezione del 13/05, ultima modifica 16/06, Andrea Gadotti

56

2.2.3 Famiglie esponenziali e sufficienza

Notiamo innanzitutto che una qualsiasi funzione di densit`a per una distribuzione appartenente alla famiglia esponenziale pu`o essere scritta come fX(x;θ) = exp{C(x) + D(θ)

∑km=1} Am(θ)Bm(x)Teorema 23. Sia (X1, ..., Xn) un campione casuale nenziale a k parametri. Allora il vettore Tn statistica sufficiente per θ.

:= (appartenente ∑ni=1 B1(xi), ...,alla ∑famiglia ni=1 Bk(xespo- i)) `e

Dimostrazione. Da inserire (comunque `e facile, basta usare il teorema di fattorizza- zione).

Esempio 4. Sia (X1, ...Xn) da N(μ, σ2) con μ e σ2 non noti (dunque X provie- ne che (μ, da σI2n ). una := (Nota: (∑famiglia ni=1 la B1(xseconda esponenziale i),∑ni=1 uguaglianza B2(xa i)) 2 parametri). = si (∑verifica ni=1 xi,∑immediatamente).

Per ni=1 il x2i teorema `e appena visto si ha statistica sufficiente per

Affermiamo (μ, σ2) `e θ ˆ= (senza ( n 1statistica sufficiente ∑ni=1 dimostrazione) In.

Xi, n

1∑ni=1(Xche i − lo Xn)stimatore 2) di massima verosimiglianza per e notiamo che esso `e funzione della

Lezione del 17/05, ultima modifica 21/06, Andrea Gadotti

2.2.4 Teorema di Rao-Blackwell

In questa sezione ci occuperemo del seguente teorema molto importante:

Teorema 24 (di Rao-Blackwell). Sia (X1, ..., Xn) un campione casuale da una di- stribuzione avente funzione di densit`a f(x;θ). Siano inoltre Vn uno stimatore non distorto di θ e Tn una statistica sufficiente per θ. Allora

Vn,Tn = Eθ(Vn|Tn)

ha le seguenti propriet`a:

(1) Eθ(Vn,Tn) = θ,∀θ ∈ Θ, ovvero Vn,Tn `e uno stimatore non distorto di θ

(2) Varθ(Vn,Tn) ≤ Varθ(Vn)

(3) Vn,Tn = φ(Tn), ovvero Vn,Tn `e funzione della statistica sufficiente Tn. Possiamo riassumere l’enunciato dicendo il condizionamento di uno stimatore ri- spetto a una statistica sufficiente preserva la non distorsione e migliora lo stimatore sotto il profilo della varianza.

Dimostrazione.

57

(1) Utilizziamo senza dimostrazione il seguente risultato:

E[E(X|Y )] = E(X)

A questo punto il primo punto `e subito dimostrato. Infatti

Eθ(Vn,Tn) = Eθ[Eθ(Vn|Tn)] = θ

(2) Utilizziamo senza dimostrazione il seguente risultato:

Var(X) = Var[E(X|Y )] + E[Var(X|Y )]

A questo punto anche il secondo punto `e subito dimostrato. Infatti

Var(Vn) = Var[E(Vn|Tn)] + E[Var(Vn|Tn)] = Var(Vn,Tn) + E[Var(Vn|Tn)]

e si conclude osservando che E[Var(Vn|Tn)] ≥ 0 in quanto valore di aspettazione di una quantit`a sicuramente non negativa.

(3)

Vn,Tn = Eθ(Vn|Tn) =

∫D vn·fvn|tn(vn,tn)dvn =

∫Dn vn(x1, ..., xn)·fx|tn(x, tn)dx = φ(Tn)

∫Dn vn(x1, ..., xn)·fx|tn(x, tn)dx = φ(Tn)

dove l’ultima uguaglianza `e giustificata dal fatto da θ per definizione di statistica sufficiente.

che fx|tn(x, tn) non dipende

Esempio 5. Sia (X1, ..., Xn) da b(1,p),p statistica sufficiente minimale per p. ∈ Vogliamo (0,1). Sappiamo che Tn = applicare il teorema di Rao-Blackwell ∑ni=1 Xi `e

con lo stimatore X1. Notiamo che lo stimatore scelto `e estremamente ”grezzo”, dato che il primo risultato di n prove non ci dice in realt`a nulla su p, ma nonostante questo lo stimatore che otterremo grazie al teorema sar`a molto buono.

a) Notiamo innanzitutto che Vn := X1 `e stimatore non distorto per p. Infatti

E(Vn) = E(X1) = p

b) Vogliamo costruire ora Vn,Tn:

58