Probabilità e statistica

Davide Calabrò

January 20, 2020

Probabilità

Funzione probabilità

Definiamo la funzione probabilità come un terna probabilistica (Ω, F, P)

Proprietà

1. $\forall E \in F \to P(E^c) = 1 - P(E)$
2. $\forall E \in F \to P(E) \leq 1$
3. se $E \subseteq F \to P(E) \leq P(F)$ e $P(F\ E) = P(F) - P(E)$ (prop. di monotonia)
4. $\forall E, F \in F \to P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Dimostrazioni

1. $\Omega = E + E^c \to 1 = P(\Omega) = P(E) + P(E^c) \to P(E^c) = 1 - P(E)$
2. $0 \leq P(E) \leq 1 - P(E^c) \leq 1$ quindi le P sono comprese tra 0 e 1
3. Definiamo $F = E \cup F\ E$ (unione disgiunta)
   $P(F) = P(E) + P(F\ E) \to P(F\ E) = P(F) - P(E)$
   Inoltre $P(F\ E) \geq 0$, per $a1 \to P(F) - P(E) \geq 0$, quindi $P(E) \leq P(F)$
4. Definiamo $E \cup F = (E \cap F) \cup [F\ (E \cap F)] \cup [E\ (E \cap F)]$ (forma disgiunta)
   $P(F \cup E) = P(E \cap F) + P(F\ E \cap F) + P(E\ E \cap F) = P(E \cap F) + P(F) - P(E \cap F) + P(E) - P(E \cap F) = P(F) + P(E) - P(E \cap F)$
   Intuitivamente l'idea è che se faccio $P(E)$ e $P(F)$ conto l'intersezione 2 volte.

Definizioni e dimostrazioni

0.1 Legge debole dei grandi numeri

Sia $X_1, X_2, ...$ una successione di variabili aleatoree indipendenti e identicamente distribuite (i.i.d.) con media μ e varianza σ^2 finite. Sia $S_n = X_1 + ... + X_n$ per ogni $n = 1, 2, ...$ Allora $\forall \epsilon > 0$

$$\lim_{n to \infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \tag{1}$$

0.1.1 Media campionaria

Definiamo $\overline{X_n} := \frac{S_n}{n} = \frac{X_1+...+X_n}{n}$. $\overline{X_n}$ è una v.a. detta media campionaria di $X_1 + ... + X_n$. Dunque il risultato precedente si può scrivere come:

$$\lim_{n \to \infty} P(|\overline{X_n} - \mu| > \epsilon) = 0 \tag{2}$$

Dimostrazione Essendo le v.a. i.i.d., per le proprietà della varianza

$$var(\overline{X_n}) = var(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} var(X_i)}{n} = \frac{n \cdot var(X_1)}{n^2} = \frac{\sigma^2}{n} \tag{3}$$

e per le proprietà della media

$$E(\overline{X_n}) = E(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{nE(X_1)}{n} = \mu \tag{4}$$

Inoltre, la disuguaglianza di Chebyshev per una v.a. Y con varianza finita dice che:

$$P(|Y - E(Y)| > \epsilon) \leq \frac{var(Y)}{\epsilon^2} \tag{5}$$

che applicata a $Y = \overline{X_n}$

$$0 \leq P(|\overline{X_n} - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ per } n \to \infty \tag{6}$$

0.2 Legge forte dei grandi numeri

Sia $X_1, X_2, ...$ una successione di v.a. i.i.d. che ammettono media μ. Allora:

$$P(\omega : \lim_{n \to \infty} \overline{X_n}(\omega) = \mu) = 1 \tag{7}$$

dove $\overline{X_n}(\omega) = (X_1(\omega) + ... + X_n(\omega))/n$. La differenza rispetto alla legge debole è che qui P non tende a 1, ma ci converge proprio.

0.3 Teorema centrale del limite (TCL)

Sia $X_1, X_2, \ldots$ una successione di v.a. i.i.d. con media μ e varianza σ^2 finite. Sia, per ogni $n = 1, 2, \ldots$, $S_n = X_1 + \ldots + X_n$, allora $\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{-\frac{u^2}{2}du} \tag{8}$$

1.

$$E(S_n) = E(X_1 + \ldots + X_n) = n\mu \tag{9}$$

$$var(S_n) = var(X_1 + \ldots + X_n) = n\sigma^2 \tag{10}$$

$$\Longrightarrow \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{S_n - E(S_n)}{\sqrt{(}var(S_n))} \text{ standardizzata} \tag{11}$$

La standardizzata di S_n $\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \approx \mathbb{N}(0,1) \implies S_n \approx \mathbb{N}(n\mu, n\sigma^2)$

2.

$$\frac{\frac{S_n - n\mu}{n}}{\frac{\sqrt{n\sigma^2}}{n}} = \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X_n} - E(\overline{X_n})}{\sqrt{var(\overline{X_n})}} \tag{12}$$

Dunque il TCL posso anche scriverlo come

$$\lim_{n\to\infty} P\left(\frac{\overline{X_n} - \mu}{\sigma}\sqrt{n} \leq x\right) = \Phi(x) \tag{13}$$

3. Teorema di Demoivre-Laplace: supponiamo $S_n =$ numero dei successi in n prove di Bernoulli, quindi

$$S_n = X_1 + \ldots + X_n X_i = \begin{cases} 1 \text{ se l'i-esima prova è un successo} \\ 0 \text{ altrimenti} \end{cases} \tag{14}$$

$$S_n \sim Bi(n,p)$$

$$S_n: \text{ somma di n i.i.d. con } X_i \sim Be(p)$$

$$E(X_i) = p(=\mu) var(X_i) = p(i-p)(=\sigma^2)$$

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)} \leq b}\right) = P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) - P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq a\right) \xrightarrow[n\to\infty]{TCL} \Phi(b) - \Phi(a) \tag{15}$$

0.4 Inferenza parametrica - definizioni

0.4.1 Stimatore puntuale

Uno stimatore puntuale di $K(\theta)$ è una statistica $K = d_n(X_1, ..., X_n)$ usata per fare inferenza su $k(\theta)$

Stimatore e stima Se $X_1 = x_1, ..., X_n = x_n$ è un'osservazione campionaria e $K_n = d_n(X_1, ..., X_n)$ è uno STIMATORE di $K(\theta)$ allora $\widehat{K}_n = d_n(X_1, ..., X_n)$ è detta STIMA di $K(\theta)$, che è circa il valore di $\widehat{K}_n$ in corrispondenza delle osservazioni. Dunque lo STIMATORE è una v.a., mentre la STIMA è un numero

0.4.2 Proprietà degli stimatori puntuali

$X_1, ..., X_n$ campione aleatorio da f_θ e sia $K(\theta)$ una caratteristica della popolazione di densità f_θ.

$$D_n = d(X_1, ..., X_2) \text{ stimatore di } K(\theta) \tag{16}$$

Errore quadratico medio Sia $X_1, ..., X_n$ campione aleatorio estratto da f_θ e sia D_n uno stimatore di $K(\theta)$. Si definisce ERRORE QUADRATICO MEDIO di D_n come stimatore di $K(\theta)$ la funzione del parametro θ

$$r(d, K(\theta)) := E_\theta[(D_n - K(\theta))^2] = E_\theta[(d(X_1, ..., X_n) - K(\theta))^2] \tag{17}$$

Mentre nel caso assolutamente continuo abbiamo:

$$r(d, K(\theta)) = \int_{\mathbb{R}^n} (d(x_1, ..., x_n) - K(\theta))^2 f_\theta(x_1)...f_\theta(x_n)dx_1...dx_n \tag{18}$$

Distorsione Sia $X_1, ..., X_n$ campione aleatorio da f_θ e sia D_n uno stimatore di $K(\theta)$. Si definisce DISTORSIONE (o bias) di $D_n = d(X_1, ..., X_n)$ come stimatore di $K(\theta)$ e la funzione di θ

$$b(d, K(\theta)) := E_\theta(D_n) - K(\theta) \tag{19}$$

Lo stimatore D_n si dice non distorto per $K(\theta)$ se

$$b(d, K(\theta)) = 0, \forall\theta \qquad (\Longleftrightarrow E_\theta(D_n) = K(\theta), \forall\theta) \tag{20}$$

Osservazioni

1. Se D_n è non distorto per $K(\theta)$

   $\Rightarrow r(d, K(\theta)) := E_\theta[(D_n - K(\theta))^2] = E_\theta[(D_n - E_\theta(D - n))^2] = var_\theta(D_n)$

2. Decomposizione dell'errore quadratico medio di $D_n = d(X_1, ..., X_n)$ come stimatore di $K(\theta)$

   $$r(d, K(\theta)) = E_\theta[(D_n - K(\theta))^2] \tag{21}$$

   Sapendo che $var(Y) = E(Y^2) - E(Y)^2$, otteniamo:

   $$= var_\theta(D_n - K(\theta)) + [E_\theta(D_n - K(\theta))]^2 = var_\theta(D_n) + [b(d, K(\theta))]^2 \tag{22}$$

Principio di invarianza degli MLE Se $\widehat{\theta}_n$ è MLE di θ e $K(\theta)$ è una caratteristica della popolazione, allora $K(\widehat{\theta}_n)$ è MLE di $K(\theta)$

0.4.3 Proprierà asintotiche degli stimatori

Distorsione Una successione $(D_n = d_n(X_1, ..., X_n))_n$ di stimatori di $K(\theta)$ è detta ASINTOTICAMENTE NON DISTORTA se

$$b(d_n, K(\theta)) = E_\theta(D_n) - K(\theta) \to 0 \text{ per } n \to \infty \quad (\iff E_\theta(D_n) \to K(\theta)) \forall \theta$$

Consistenza debole Una successione $(D_n = d_n(X_1, ..., X_n))_n$ di stimatori di $K(\theta)$ è detta DEBOLMENTE CONSISTENTE se $\forall \epsilon > 0$ vale

$$P_\theta(|D_n - K(\theta)| > \epsilon) \to 0 \text{ per } n \to \infty$$

Consistenza in media quadratica Una successione $(D_n = d_n(X_1, ..., X_n))_n$ di stimatori di $K(\theta)$ è detta CONSISTENTE IN MEDIA QUADRATICA se

$$r(d_n, K(\theta)) := E_\theta[(D_n - K(\theta))^2] \to 0 \text{ per } n \to \infty$$

Osservazioni

- Se $(D_n)_{n \geq 1}$ è CONSISTENTE IN MEDIA QUADRATICA, allora la successione è anche DEBOLMENTE CONSISTENTE
- Data $r(d_n, K(\theta)) = var(D_n) + [b(d_n, K(\theta))]^2$, se $var(D_n) \to 0$ per $n \to \infty$ $\forall \theta$ e $(D_n)_n$ è ASINTOTICAMENTE NON DISTORTA, allora $(D_n)_n$ è CONSISTENTE IN MEDIA QUADRATICA

0.5 Quantità pivotali

Si definisce QUANTITA' PIVOTALE una v.a. $Q = q(X_1, ..., X_n, \theta)$ la cui distribuzione non dipende da θ. Si noti che la quantità pivotale non è una statistica, in quanto dipende da θ

0.6 Verifica di un'ipotesi statistica

Un'IPOTESI STATISTICA è un'affermazione su uno o più parametri delli distribuzione della popolazione.

0.6.1 Regione critica e livello di significatività di un test

Consideriamo una popolazione avente distribuzione F_θ che dipende da un parametro incognito θ e supponiamo di voler verificare una qualunque ipotesi su θ che chiameremo IPOTESI NULLA e che indicheremo con $\mathbb{H}_0$

Def. Un'ipotesi statistica si dice SEMPLICE se caratterizza completamente la distribuzione della popolazione, altrimenti è detta COMPOSTA

Def. Un test per la verifica di un'ipotesi $\mathbb{H}_0$ ha REGIONE CRITICA $C \subseteq \mathbb{R}^n$ se osservando $X_1 = x_1, ..., X_n = x_n$

- ACCETTO $\mathbb{H}_0$ se $(X_1, ..., X_n) \in C$
- RIFIUTO $\mathbb{H}_0$ se $(X_1, ..., X_n) \notin C$

Errori

- Commettiamo un errore di 1^a specie quando i dati ci portano a rifiutare $\mathbb{H}_0$ e $\mathbb{H}_0$ è corretta
- Commettiamo un errore di 2^a specie quando i dati ci portano ad accettare $\mathbb{H}_0$ e $\mathbb{H}_0$ è falsa

Attenzione L'errore di 1^a specie è più grave!

Livello di significatività Si definisce livello di significatività α quel valore tale per cui

$$P(\text{errore di } 1^a \text{ specie}) = P(\text{Rifiutare } \mathbb{H}_0) \leq \alpha$$

p-value Data una famiglia di test, al variare del livello di significatività, si definisce P-VALUE il più piccolo valore della significatività per cui rifuto $\mathbb{H}_0$ con i dati a disposizione. Quindi $\alpha \geq p-value \rightarrow$ rifiuto $\mathbb{H}_0$, accetto altrimenti.

Curva OC Si definisce CURVA OC del test la funzione del parametro incognito μ

$$B(\mu) := P_\mu(\text{accettare } \mathbb{H}_0)$$

quindi se μ soddisfa l'ipotesi alternativa $\mathbb{H}_1$ allora $B(\mu)$

$$B(\mu) = P_\mu \left(-z_{\alpha/2} < \frac{(\overline{X_n} - \mu_0)}{\sigma_0} \sqrt{n} < z_{\alpha/2} \right)$$

Attenzione $\frac{(\overline{X_n}-\mu_0)}{\sigma_0}\sqrt{n}$ non è una gaussiana standard, per la presenza di μ_0 al posto di μ

Potenza Si definisce potenza rispetto al parametro μ

$$\pi(\mu) := 1 - B(\mu)$$