Probabilità di Dirac

Per un certo $\omega_0 \in \Omega$, la funzione $P_0 : \mathcal{E} \to \mathbb{R}_+$ data da:

$$P_0(E) := \begin{cases} 1 & se\ \omega_0 \in E \\ 0 & se\ \omega_0 \notin E \end{cases}$$

è una probabilità finitamente additiva su Ω tale che $P_0(\omega_0) = 1$.

Verifichiamo che si tratti effettivamente di una probabilità, e consideriamo due eventi $E, F \in \mathcal{E}$ tali che $E \cap F = \emptyset$. Sono possibili solo due casi:

- $\omega_0 \in E \cup F \implies P_0(E \cup F) = 1$
  D'altra parte, poiché E, F sono incompatibili, ω_0 deve appartenere o a E o a F. Ciò implica $P_0(E) + P_0(F) = 1$.
- $\omega_0 \notin E \cup F \implies P_0(E \cup F) = 0$
  D'altra parte, ω_0 non può appartenere né a E né a F. Ciò implica $P_0(E) + P_0(F) = 0$.

Di conseguenza, in entrambi i casi abbiamo:

$$P_0(E \cup F) = P_0(E) + P_0(F)$$

Probabilità di Bernoulli

Assumiamo che il sample space Ω contenga due sample point (ω_0, ω_1). Siano $p \in (0, 1)$ e $q \equiv 1 - p$. Allora la funzione $P : \mathcal{P}(\Omega) \to \mathbb{R}_+$ data da:

$$P(E) := \begin{cases} 1 & se\ \omega_0 \in E\ \wedge\ \omega_1 \in E \\ p & se\ \omega_0 \notin E\ \wedge\ \omega_1 \in E \\ q & se\ \omega_0 \in E\ \wedge\ \omega_1 \notin E \\ 0 & se\ \omega_0 \notin E\ \wedge\ \omega_1 \notin E \end{cases}$$

è una probabilità finitamente additiva su Ω tale che $P(\omega_1) = p$ e $P(\omega_0) = q$.

Verifichiamo che si tratti effettivamente di una probabilità, e consideriamo due eventi $E, F \in \mathcal{E}$ tali che $E \cap F = \emptyset$. Se poniamo attenzione sul solo evento E, sono possibili 4 casi:

- $\omega_0 \in E\ \wedge\ \omega_1 \in E \implies \omega_0, \omega_1 \in E \cup F\ \wedge\ \omega_0, \omega_1 \notin F \implies$

  $$P(E) = 1, \quad P(F) = 0, \quad P(E \cup F) = 1$$

- $\omega_0 \in E\ \wedge\ \omega_1 \notin E \implies$ abbiamo 2 sottocasi:
  - $\omega_1 \in F \implies \omega_0, \omega_1 \in E \cup F \implies$

    $$P(E) = q, \quad P(F) = p, \quad P(E \cup F) = 1$$

- $\omega_1 \notin F \implies \omega_0 \in E \cup F \ \wedge \ \omega_1 \notin E \cup F \implies$
$$P(E) = q, \quad P(F) = 0, \quad P(E \cup F) = q$$

- $\omega_0 \notin E \ \wedge \ \omega_1 \in E$; si tratta di un caso del tutto analogo al precedente.
- $\omega_0 \notin E \ \wedge \ \omega_1 \notin E \implies$ abbiamo 4 sottocasi:
  - $\omega_0 \in F \ \wedge \ \omega_1 \in F \implies \omega_0, \omega_1 \in E \cup F \implies$
$$P(E) = 0, \quad P(F) = 1, \quad P(E \cup F) = 1$$
  - $\omega_0 \in F \ \wedge \ \omega_1 \notin F \implies \omega_0 \in E \cup F \ \wedge \ \omega_1 \notin E \cup F \implies$
$$P(E) = 0, \quad P(F) = q, \quad P(E \cup F) = q$$
  - $\omega_0 \notin F \ \wedge \ \omega_1 \in F \implies \omega_0 \notin E \cup F \ \wedge \ \omega_1 \in E \cup F \implies$
$$P(E) = 0, \quad P(F) = p, \quad P(E \cup F) = p$$
  - $\omega_0 \notin F \ \wedge \ \omega_1 \notin F \implies \omega_0, \omega_1 \notin E \cup F \implies$
$$P(E) = 0, \quad P(F) = 0, \quad P(E \cup F) = 0$$

In definitiva, in tutti i casi abbiamo:
$$P(E \cup F) = P(E) + P(F)$$

Probabilità discreta uniforme

Assumiamo che il sample space Ω contenga un numero finito di sample point ($\Omega \equiv \{\omega_1, .., \omega_n\}$ per qualche $n \in \mathbb{N}$). Sia $P_k : \mathcal{P}(\Omega) \to \mathbb{R}_+$ la probabilità di Dirac concentrata in ω_k, per $k = 1, ..., n$. Allora la funzione $P : \mathcal{P}(\Omega) \to \mathbb{R}_+$ data da:
$$P(E) := \frac{1}{n} \sum_{k=1}^{n} P_k(E)$$
è una probabilità finitamente additiva su Ω tale che $P(\omega_k) = \frac{1}{n} \ \ \forall k = 1, ..., n$.

Verifichiamo che si tratti effettivamente di una probabilità. Poiché P_k è una probabilità, per ogni $k = 1, ..., n$ abbiamo:
$$P(\Omega) = \frac{1}{n} \sum_{k=1}^{n} P_k(\Omega) = \frac{1}{n} \sum_{k=1}^{n} 1 = \frac{1}{n} \cdot n = 1$$
Inoltre, grazie alle proprietà delle somme finite:
$$P(E \cup F) = \frac{1}{n} \sum_{k=1}^{n} P_k(E \cup F) = \frac{1}{n} \sum_{k=1}^{n} (P_k(E) + P_k(F)) = \frac{1}{n} \sum_{k=1}^{n} P_k(E) + \frac{1}{n} \sum_{k=1}^{n} P_k(F) = P(E) + P(F)$$
per ogni coppia di eventi $E, F \in \mathcal{P}(\Omega)$ tali che $E \cap F = \emptyset$.

Probabilità binomiale

Assumiamo che il sample space Ω contenga un numero finito di sample point ($\Omega \equiv \{\omega_0, \omega_1, .., \omega_n\}$ per qualche $n \in \mathbb{N}$). Siano $p \in (0,1)$ e $q \equiv 1-p$. Allora la funzione $P : \mathcal{P}(\Omega) \to \mathbb{R}_+$ data da:

$$P(E) := \sum_{\{k \in \{0,1,...,n\} : \omega_k \in E\}} \binom{n}{k} p^k q^{n-k}$$

è una probabilità finitamente additiva su Ω tale che $P(\omega_k) = \binom{n}{k} p^k q^{n-k}$ $\forall k = 0, 1, ..., n$.

Verifichiamo che si tratti effettivamente di una probabilità. Considerando la formula del binomio di Newton, otteniamo:

$$P(\Omega) = \sum_{\{k \in \{0,1,...,n\} : \omega_k \in \Omega\}} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1$$

Ora, dati $E, F \in \mathcal{P}(\Omega)$ tali che $E \cap F = \emptyset$, abbiamo:

$$\{k \in \{0,1,...,n\} : \omega_k \in E\} \cap \{k \in \{0,1,...,n\} : \omega_k \in F\} = \emptyset$$

$$\{k \in \{0,1,...,n\} : \omega_k \in E \cup F\} = \{k \in \{0,1,...,n\} : \omega_k \in E\} \cup \{k \in \{0,1,...,n\} : \omega_k \in F\}$$

Di conseguenza otteniamo:

$$P(E \cup F) = \sum_{\{k \in \{0,1,...,n\} : \omega_k \in E \cup F\}} \binom{n}{k} p^k q^{n-k} = \sum_{\{k \in \{0,1,...,n\} : \omega_k \in E\} \cup \{k \in \{0,1,...,n\} : \omega_k \in F\}} \binom{n}{k} p^k q^{n-k} =$$

$$\sum_{\{k \in \{0,1,...,n\} : \omega_k \in E\}} \binom{n}{k} p^k q^{n-k} + \sum_{\{k \in \{0,1,...,n\} : \omega_k \in F\}} \binom{n}{k} p^k q^{n-k} = P(E) + P(F)$$

Proposizione

$$\int_\Omega ||X||_2 \, dP \; < +\infty \iff \int_\Omega |X_n| \, dP \; < +\infty \quad \forall n = 1, ..., N$$

Dimostrazione

$\Rightarrow$) $|X_n| = (X_n^2)^{\frac{1}{2}} \leq \left(\sum_{n=1}^N X_n^2 \right)^{\frac{1}{2}} = ||X||_2 \quad \forall n = 1, ..., N$

$\Leftarrow$) $||X||_2 = \left(\sum_{n=1}^N X_n^2 \right)^{\frac{1}{2}} \leq \sum_{n=1}^N |X_n|$

(Questo perché se elevo $\left(\sum_{n=1}^N X_n^2 \right)^{\frac{1}{2}}$ al quadrato ottengo la somma dei quadrati, mentre se elevo $\sum_{n=1}^N |X_n|$ al quadrato ottengo la somma dei quadrati e i doppi prodotti).

Proposizione

$$\int_\Omega ||X||_2^2 \, dP \ < +\infty \iff \int_\Omega |X_m X_n| \, dP \ < +\infty \quad \forall m, n = 1, ..., N$$

Dimostrazione

$\Rightarrow$) $(X_n^2) \leq \sum_{n=1}^{N} X_n^2 = ||X||_2^2 \quad \forall n = 1, ..., N \implies$
Assumendo che $||X||_2^2$ sia integrabile su Ω, tutte le componenti $X_1, ..., X_N$ di X hanno momento di ordine 2 finito $\implies$
Per la disuguaglianza di Cauchy-Schwarz ($\mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{\frac{1}{2}}\mathbb{E}[Y^2]^{\frac{1}{2}}$), tutti i prodotti $X_m X_n$ hanno momento di ordine 1 finito $\forall m, n = 1, ..., N$.
$\Leftarrow$) Se tutti i prodotti $X_m X_n$, al variare di $m, n = 1, ..., N$, hanno momenti di ordine 1 finito $\implies$
Tutti i quadrati $X_1^2, ..., X_N^2$ hanno momento di ordine 1 finito $\implies$
$||X||_2^2$ ha momento di ordine 1 finito.

Proposizione

$$\mathbf{MSE}(\hat{\theta}_n) = \mathbb{D}^2[\hat{\theta}_n] + \mathbf{Bias}^2(\hat{\theta}_n)$$

Dimostrazione

$\mathbf{MSE}(\hat{\theta}_n) = \mathbb{E}[(\hat{\theta}_n - \theta)^2] = \mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n] + \mathbb{E}[\hat{\theta}_n] - \theta)^2] =$
$= \mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])^2 + 2(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])(\mathbb{E}[\hat{\theta}_n] - \theta) + (\mathbb{E}[\hat{\theta}_n] - \theta)^2] =$
$= \mathbb{E}[([\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])^2] + 2\mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])(\mathbb{E}[\hat{\theta}_n] - \theta)] + \mathbb{E}[(\mathbb{E}[\hat{\theta}_n] - \theta)^2]$

Ora, poiché θ e $\mathbb{E}[\hat{\theta}_n]$ sono numeri reali, abbiamo che:

$$\mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])(\mathbb{E}[\hat{\theta}_n] - \theta)] = (\mathbb{E}[\hat{\theta}_n] - \theta)\mathbb{E}[\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]] = (\mathbb{E}[\hat{\theta}_n] - \theta)(\mathbb{E}[\hat{\theta}_n] - \mathbb{E}[\hat{\theta}_n]) = 0$$

$$\mathbb{E}[(\mathbb{E}[\hat{\theta}_n] - \theta)^2] = (\mathbb{E}[\hat{\theta}_n] - \theta)^2$$

In definitiva:
$\mathbf{MSE}(\hat{\theta}_n) = \mathbb{E}[([\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])^2] + (\mathbb{E}[\hat{\theta}_n] - \theta)^2 = \mathbb{D}^2[\hat{\theta}_n] + \mathbf{Bias}^2(\hat{\theta}_n)$

Proposizione

Se uno stimatore $\hat{\theta}_n$ corretto ($\equiv$ non distorto) di θ è consistente in media quadratica, allora è consistente anche in probabilità.

Dimostrazione

Poiché lo stimatore è corretto, applicando la disuguaglianza di Tchebyshev, possiamo scrivere:

$$P(|\hat{\theta}_n - \theta| \geq \epsilon) = P(|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])^2] = \frac{1}{\epsilon^2}\mathbb{E}[(\hat{\theta}_n - \theta)^2]$$

Di conseguenza, dalla convergenza di $\hat{\theta}_n$ a θ in media quadratica segue la convergenza di $\hat{\theta}_n$ a θ in probabilità.

Proposizione

Sia X una variabile aleatoria gaussiana con varianza σ^2 nota. Allora, fissato $\alpha \in (0,1)$, un intervallo di confidenza al livello di confidenza $1-\alpha$ per il valore vero del parametro μ è dato dalle seguenti statistiche:

$$\overline{X}_n - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad ; \quad \overline{X}_n + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

dove $z_{\frac{\alpha}{2}} \equiv z^+_{\frac{\alpha}{2}}$ è il valore critico superiore di livello $\frac{\alpha}{2}$ della variabile aleatoria gaussiana.
Le realizzazioni di tale intervallo di confidenza sono della forma

$$\left(\overline{x}_n - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad , \quad \overline{x}_n + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

dove $\overline{x}_n$ è il valore preso dallo stimatore $\overline{X}_n$ su qualunque delle realizzazioni possibili $x_1, ..., x_n$ del campione $X_1, ..., X_n$.

Dimostrazione

Poiché X è una variabile aleatoria gaussiana con media μ e varianza σ^2, la statistica $\overline{X}_n$ è a sua volta una variabile aleatoria gaussiana con media μ e varianza $\frac{\sigma^2}{n}$. Di conseguenza, $\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ è una variabile aleatoria gaussiana standard.
Considerando i valori critici superiore e inferiore di livello $\frac{\alpha}{2}$ della distribuzione gaussiana standard, abbiamo:

$$P\left(z^-_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z^+_{\frac{\alpha}{2}}\right) \geq 1 - \alpha$$

D'altra parte:

$$z^-_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z^+_{\frac{\alpha}{2}} \iff z^-_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \overline{X}_n - \mu \leq z^+_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \iff$$

$$-\overline{X}_n + z^-_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\overline{X}_n + z^+_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \iff \overline{X}_n - z^+_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n - z^-_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Di conseguenza, abbiamo che:

$$P\Big(\overline{X}_n - z^+_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n - z^-_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\Big) \geq 1-\alpha$$

Perciò, le statistiche $\overline{X}_n - z^+_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ e $\overline{X}_n - z^-_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ costituiscono un intervallo di confidenza per il parametro μ al livello di confidenza $1-\alpha$.
Se ricordiamo che $z^-_{\frac{\alpha}{2}} = -z^+_{\frac{\alpha}{2}}$ e poniamo $z_{\frac{\alpha}{2}} := z^+_{\frac{\alpha}{2}}$, la dimostrazione si conclude.

Proposizione

Sia X una variabile aleatoria gaussiana con varianza σ^2 ignota. Allora, fissato $\alpha \in (0,1)$, un intervallo di confidenza al livello di confidenza $1-\alpha$ per il valore vero di σ^2 è dato dalle seguenti statistiche:

$$\frac{(n-1)S^2_{X,n}}{\chi^2_{n-1,\frac{\alpha}{2},+}} \quad ; \quad \frac{(n-1)S^2_{X,n}}{\chi^2_{n-1,\frac{\alpha}{2},-}}$$

dove $\chi_{n-1,\frac{\alpha}{2},-}$ è il valore critico inferiore di livello $\frac{\alpha}{2}$ e $\chi_{n-1,\frac{\alpha}{2},+}$ è il valore critico superiore di livello $\frac{\alpha}{2}$ della variabile aleatoria gaussiana.
Le realizzazioni di tale intervallo di confidenza sono della forma

$$\left(\frac{(n-1)s^2_n(X)}{\chi^2_{n-1,\frac{\alpha}{2},+}} \quad , \quad \frac{(n-1)s^2_n(X)}{\chi^2_{n-1,\frac{\alpha}{2},-}}\right)$$

dove $s^2_n(X)$ è il valore preso dallo stimatore $S^2_{X,n}$ su qualunque delle realizzazioni possibili $x_1,...,x_n$ del campione $X_1,...,X_n$.

Dimostrazione

Poiché X è una variabile aleatoria gaussiana con varianza σ^2, la statistica $\frac{(n-1)S^2_{X,n}}{\sigma^2}$ ha una distribuzione χ^2_{n-1}.
Abbiamo inoltre che il valore critico inferiore $\chi^2_{n-1,\alpha,-}$ di livello α della variabile aleatoria χ^2_{n-1} è l'α-quantile di χ^2_{n-1}, mentre il valore critico superiore $\chi^2_{n-1,\alpha,+}$ di livello α è l'$1-\alpha$-quantile di χ^2_{n-1}.
Perciò, considerando i valori critici superiore e inferiore di livello $\frac{\alpha}{2}$ della distribuzione χ^2_{n-1}, abbiamo:

$$P\left(\chi^2_{n-1,\frac{\alpha}{2},-} \leq \frac{(n-1)S^2_{X,n}}{\sigma^2} \leq \chi^2_{n-1,\frac{\alpha}{2},+}\right) \geq 1-\alpha$$

Da qui otteniamo facilmente che:

$$P\left(\frac{(n-1)S^2_{X,n}}{\chi^2_{n-1,\frac{\alpha}{2},+}} \leq \sigma^2 \leq \frac{(n-1)S^2_{X,n}}{\chi^2_{n-1,\frac{\alpha}{2},-}}\right) \geq 1-\alpha$$

Perciò, le statistiche $\frac{(n-1)S^2_{X,n}}{\chi^2_{n-1,\frac{\alpha}{2},+}}$ e $\frac{(n-1)S^2_{X,n}}{\chi^2_{n-1,\frac{\alpha}{2},-}}$ costituiscono un intervallo di confidenza per il parametro σ^2 al livello di confidenza $1-\alpha$.