

Appunti del corso di Affidabilità e Diagnostica dei Sistemi Elettrici

Andrea Cavallini, Gian Carlo Montanari
DIE-Università di Bologna
viale Risorgimento 2, 40136 Bologna
andrea.cavallini@mail.ing.unibo.it
<http://limat.ing.unibo.it>

A.A 1999/2000

Indice

1	Calcolo delle probabilità	6
1.1	Esperimento aleatorio	6
1.2	Eventi e spazi rappresentativi	6
1.3	Algebra degli eventi	7
1.4	Probabilità	9
1.5	Alcune conseguenze degli assiomi (1)-(3)	10
1.5.1	Probabilità di \emptyset	10
1.5.2	Probabilità di $\bar{\mathcal{E}}$	10
1.5.3	Probabilità di $\mathcal{E} - \mathcal{F} = \mathcal{E}\bar{\mathcal{F}}$	10
1.5.4	Probabilità di $\mathcal{E} + \mathcal{F}$	11
1.6	Combinazioni di eventi equiprobabili: il campionamento	13
1.6.1	Campionamento con reintroduzione	17
1.6.2	Campionamento senza reintroduzione	18
1.7	La legge dei grandi numeri	19
2	Indipendenza e dipendenza stocastica	20
2.1	Probabilità condizionata	20
2.2	Indipendenza stocastica	21
2.2.1	Chain rule	22
2.2.2	Esempi	22
2.3	Numero di successi in esperimenti ripetuti	24
2.4	Approssimazioni della distribuzione binomiale	26
2.4.1	Il teorema di deMoivre Laplace	26
2.4.2	Il teorema di Poisson	27
2.5	Probabilità totale	30
3	Variabili aleatorie	32
3.1	Il concetto di variabile aleatoria	32
3.2	Eventi	33
3.3	Funzioni di distribuzione	34
3.4	Densità di probabilità	37
3.4.1	Classificazione delle VA ed eventi delle VA continue	37
3.4.2	Densità di probabilità	39
3.4.3	VA discrete come caso particolare di VA continue	39

3.4.4	VA miste	40
3.5	Percentili	40
3.6	Trasformazioni lineari	41
3.7	Funzioni di uso comune	42
3.7.1	Distribuzione normale (gaussiana)	43
3.7.2	Distribuzione lognormale	48
3.7.3	Distribuzione di Weibull	50
3.7.4	Distribuzione esponenziale	51
3.7.5	Distribuzione chi-quadro	52
3.7.6	Legge di probabilità , di Student	53
3.7.7	La distribuzione F di Snedecor	54
3.8	Distribuzioni condizionate	54
3.9	Appendice 1: L'impulso di Dirac e derivata generalizzata	57
3.9.1	Definizione	57
3.9.2	Derivata di funzioni con discontinuitá	59
3.10	Appendice 2: Tavole della distribuzione normale	60
3.11	Appendice 3: Tavole della distribuzione chi-quadro	63
4	Variabili aleatorie bivariate	65
4.1	Eventi	65
4.2	Distribuzione e densitá di probabilitá	66
4.3	Distribuzioni marginali	66
4.4	Variabili aleatorie congiuntamente normali	69
4.5	Indipendenza stocastica	69
4.6	Alcune funzioni di VA doppie	70
4.6.1	Somma di due variabili aleatorie	70
4.6.2	Differenza di due VA	72
4.6.3	Massimo di due VA	73
4.6.4	Minimo di due VA	74
4.7	Distribuzioni condizionate	75
4.7.1	Variabili aleatorie congiuntamente normali	77
5	Momenti di una variabile aleatoria	79
5.1	Previsione di una variabile aleatoria	79
5.1.1	Previsione di una sequenza di dati	79
5.1.2	Comportamento asintotico: il valore atteso e media	80
5.1.3	La probabilitá come valore atteso	82
5.1.4	Esistenza del valore atteso	82
5.1.5	Linearitá del valore atteso	83
5.1.6	Altre misure di intensitá	83
5.2	Momenti del secondo ordine di VA univariate: varianza	84
5.3	Il lemma di Tchebycheff	86
5.4	Altre misure di dispersione	88
5.5	Momenti di ordine superiore a 2	88

5.6	Momenti del secondo ordine di VA doppie: covarianza	89
5.6.1	Trasformazioni lineari	92
5.7	Il teorema del limite centrale	94
5.8	Valore atteso e varianza condizionati	96
6	Affidabilità	99
6.1	Generalità sul guasto	99
6.2	Sistemi non riparabili	101
6.2.1	Funzioni affidabilistiche empiriche	103
6.2.2	Il tasso di guasto istantaneo	104
6.2.3	Parametri affidabilistici	106
6.3	Tasso di guasto per componenti elettronici	107
6.4	Generalità, concetto di missione	110
6.5	Il diagramma affidabilistico	110
6.6	Strutture semplici	112
6.6.1	Sistemi di tipo serie	112
6.6.2	Sistemi di tipo parallelo (ridondanza)	112
6.6.3	Combinazione di strutture tipo serie e parallelo	113
6.6.4	Influenza del modo di guasto dei dispositivi	113
6.7	Strutture complesse	116
6.7.1	Il metodo della probabilità totale	116
6.7.2	Il metodo dello spazio degli stati	117
7	Disponibilità	119
7.1	Definizioni	119
7.1.1	Analisi con le catene di Markov	120
7.2	Analisi combinatoria	121
7.2.1	Frequenza	121
7.3	Analisi di sistemi serie/parallelo	124
7.3.1	Sistemi con dispositivi a guasti indipendenti	124
7.3.2	Sistemi con dispositivi a guasti dipendenti	127
7.4	Ridondanza	129
7.5	Analisi affidabilistica di un sistema di distribuzione radiale	132
7.5.1	Considerazioni generali	132
7.5.2	Criterio di guasto	134
7.5.3	Sistema radiale semplice (1)	135
7.5.4	Sistema radiale con selezione della linea di alimentazione esterna all'impianto.	138
7.5.5	Sistema radiale con selezione della linea di alimentazione alla sbarra di media tensione dell'utente.	141
7.5.6	Sistema radiale con selezione della linea di alimentazione al primario del trasformatore.	143
7.5.7	Sistema radiale con selezione della linea di alimentazione al secondario del trasformatore.	145

8	Metodi empirici	147
8.1	Stima empirica delle leggi di probabilità	147
8.2	Percentili	152
8.3	Carte probabilistiche	152
8.4	Stima empirica di momenti e percentili	154
8.4.1	Valore atteso	154
8.4.2	Varianza	155
8.4.3	Covarianza e correlazione empiriche	155
8.4.4	Momenti	156
9	Stime puntuali	157
9.1	Introduzione	157
9.2	Proprietà degli stimatori	157
9.3	Il metodo dei momenti	161
9.4	Principio di massima verosimiglianza, ML	163
9.4.1	Proprietà dello stimatore ML	164
9.4.2	Stima ML della probabilità di un evento	165
9.4.3	Stima dei parametri di una distribuzione normale	166
9.4.4	Stima ML del tasso di guasto	166
9.4.5	Stima dei parametri di una distribuzione di Weibull	168
10	Stime per intervalli	170
10.1	Introduzione	170
10.2	Quantità pivotali	170
10.2.1	Il metodo della quantità pivotale	173
10.3	Campionamento da una distribuzione normale	174
10.3.1	Calcolo degli intervalli di confidenza per la media	174
10.3.2	Varianza	176
10.3.3	Rapporto di varianze	176
10.4	Il metodo statistico	177
10.5	Intervallo di confidenza per la probabilità	179
10.5.1	Calcolo mediante l'approssimazione normale	181
10.6	Intervallo di confidenza per λ di una distribuzione esponenziale	182
11	Verifica delle ipotesi	185
11.1	Introduzione	185
11.2	Ipotesi parametriche	185
11.2.1	Esempio di test per la media	188
11.2.2	Ipotesi semplici e composte	189
11.3	Test bidirezionali	192
11.3.1	Intervalli di confidenza	193
11.4	Test unidirezionali	194
11.4.1	Intervalli di confidenza	194
11.5	Test sulla media	195

11.5.1	Test bidirezionale	195
11.5.2	Test unidirezionali	197
11.6	Test sulla varianza per distribuzioni normali	198
11.7	Test sul rapporto delle varianze per distribuzioni normali	198
11.7.1	Test bidirezionali	198
11.7.2	Test unidirezionali	199
11.8	Test su due medie	199
11.8.1	Varianze identiche	199
11.8.2	Varianze diverse	200
11.9	Test bilaterali	201
11.9.1	Test bilaterale sulla probabilità	202
11.9.2	Test bilaterale su <i>MTBF</i>	203
11.9.3	Test sequenziali	205
11.10	Test non parametrici	206
11.10.1	Test di adattamendo del chi quadrato	206

Capitolo 8

Metodi empirici

I metodi empirici consentono una stima agevole ed intuitiva della distribuzione di probabilità (popolazione) di una VA e dei suoi momenti sulla base dei risultati sperimentali (campione). Normalmente, i metodi empirici richiedono un modello probabilistico estremamente ampio. Infatti, il primo metodo presentato, quello delle carte di probabilità, richiede solamente che sia specificata la legge di probabilità dei dati. La stima empirica dei momenti non richiede alcuna ipotesi sul modello probabilistico dei dati.

8.1 Stima empirica delle leggi di probabilità

Sia

$$\mathcal{X} = \{x_k, k = 1 \dots, n\} \quad (8.1)$$

l'insieme dei risultati sperimentali (campione) della variabile aleatoria \mathbf{x} . La stima empirica della distribuzione di probabilità si ottiene ordinando per valori crescenti i dati misurati:

$$x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_n$$

Poiché la migliore stima della probabilità di un evento è la frequenza relativa di esso, la migliore stima della probabilità $\Pr(\mathbf{x} \leq x_i)$ è data dal rapporto i/n :

$$\hat{F}(x_i) = \frac{i}{n} \quad (8.2)$$

Il “cappello” (^) indica che la grandezza calcolata è una stima, ottenuta in base ai dati del campione, della distribuzione di probabilità della popolazione. La procedura seguita è mostrata, schematicamente, in figura 8.1. Tale procedura è, in realtà, comune a tutte le stime di tipo statistico e prevede: (1) la realizzazione di esperimenti necessaria per formare un campione, (2) l'applicazione di una procedura (stimatore) per (3) ottenere la stima del parametro o distribuzione cercata. Si noti che il risultato di questa operazione non è noto a priori e, pertanto, deve essere considerato a tutti gli effetti come la realizzazione di una variabile aleatoria. Quest'ultimo punto, particolarmente importante, sarà approfondito in seguito.

Il grafico dei punti $(x_i, i/n)$ fornisce una rappresentazione visiva dell'andamento della funzione di distribuzione della VA \mathbf{x} . Nella figura 8.2 si mostrano le distribuzioni empirica e teorica di 20 dati estratti da una popolazione normale di media nulla e varianza unitaria.

Un modo alternativo di procedere sia il seguente. Siano $x_{min} = x_1, x_{max} = x_n$ (si ipotizza che i dati siano ordinati per valori crescenti). L'intervallo $[x_{min}, x_{max}]$ può essere suddiviso in C sottointervalli:

$$U_i = \left[x_{C_i} - \frac{h}{2}, x_{C_i} + \frac{h}{2} \right] \quad (8.3)$$

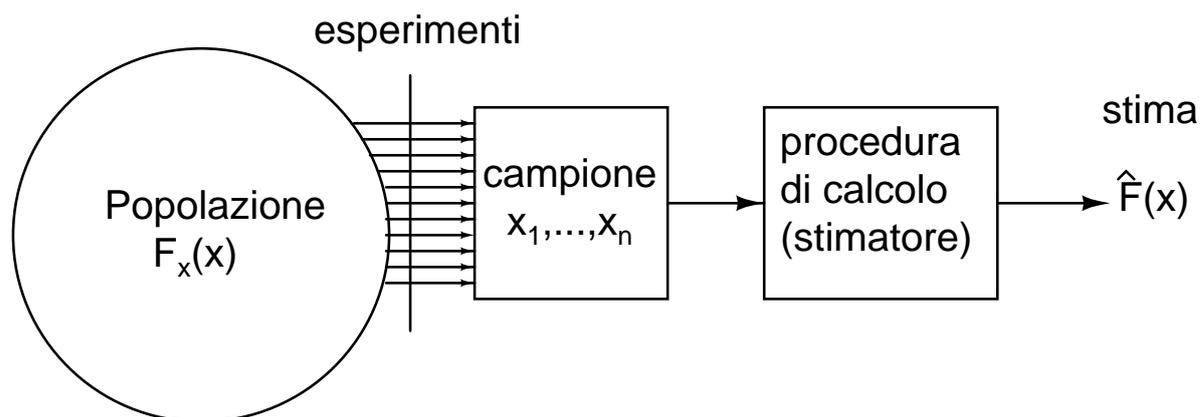


Figura 8.1: Procedura di stima della distribuzione di una variabile aleatoria.

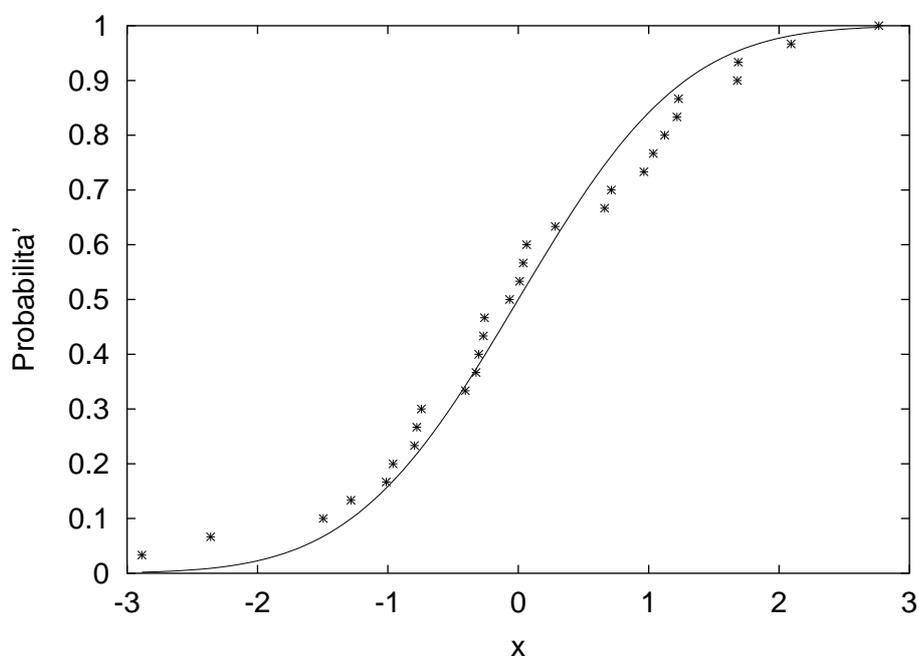


Figura 8.2: Distribuzione empirica (punti) e teorica (linea continua) di un campione di numerosità 20 estratto da una popolazione normale di media nulla e varianza unitaria.

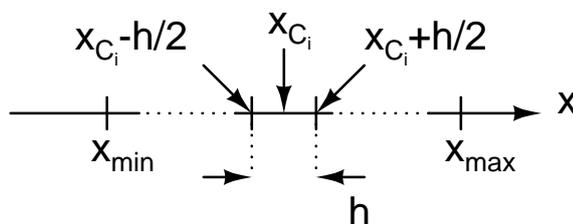


Figura 8.3: Partizione dell'asse reale.

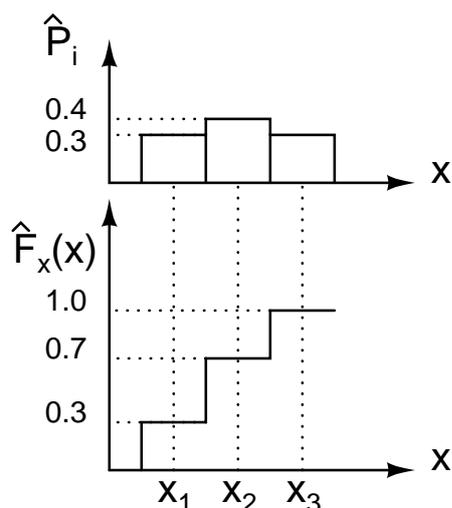


Figura 8.4: Stima della distribuzione di probabilità a partire dai valori calcolati \hat{P}_i .

dove

$$h = \frac{x_{max} - x_{min}}{C} \quad (8.4)$$

e

$$x_{C_i} = x_{min} + \frac{h}{2} \quad (8.5)$$

E' immediato verificare che tali intervalli ricoprono completamente $[x_{min}, x_{max}]$ e non sono sovrapposti, cioè formano una partizione. Conseguentemente, i numeri in $[x_{min}, x_{max}]$ sono suddivisi in classi. Ad esempio, la classe i -esima comprende tutti i numeri appartenenti ad U_i .

Sia ora n_i il numero di volte in cui la variabile aleatoria assume valori in U_i . Poiché U_i definisce un evento della variabile aleatoria x , se $P_i = \Pr(x \in U_i)$, allora, per il teorema di Bernoulli è:

$$\hat{P}_i = \frac{n_i}{n} \quad (8.6)$$

Da \hat{P}_i si può ottenere una stima della distribuzione di probabilità. Infatti, la probabilità di osservare un valore della VA inferiore a $x_{C_i} + h/2$ è la somma della probabilità di osservare la VA in uno qualsiasi (unione, OR logico) degli intervalli di indice inferiore o uguale ad i . Dunque:

$$\hat{F}(x_{C_i} + h/2) = \sum_{j=1}^i \hat{P}_j \quad (8.7)$$

Tale procedimento è mostrato, sinteticamente, in figura 8.4

Inoltre, poiché:

$$\Pr(x \in [a, b]) = F_x(b) - F_x(a)$$

e

$$f_x(x) = \lim_{h \rightarrow 0} \frac{\Pr(x \in [x, x+h])}{h}$$

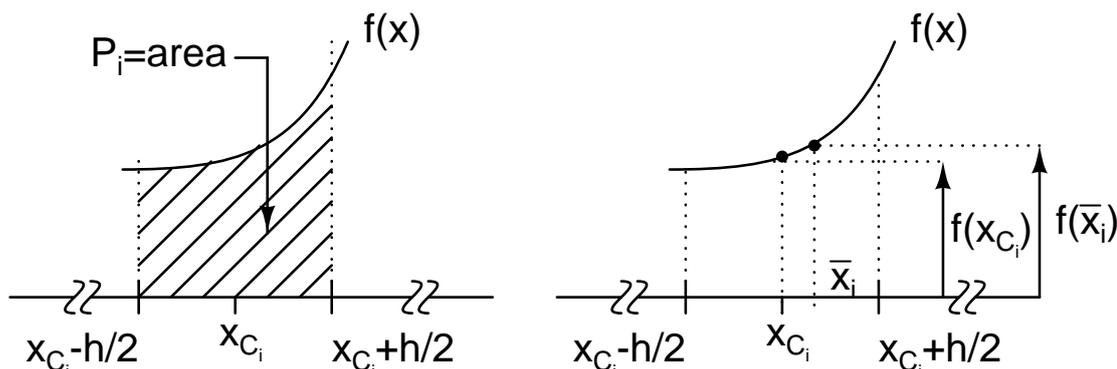


Figura 8.5: Per stimare correttamente f_{C_i} è necessario che l'ampiezza, h , dell'intervallo U_i tenda a zero. In caso contrario, ciò che si stima è in valore medio di f in U_i , che non coincide, in generale, con f_{C_i} .

La stima della densità di probabilità è data da:

$$\hat{f}(x_{C_i}) = \frac{\hat{P}_i}{h} \quad (8.8)$$

Per il teorema di Bernoulli è:

$$P_i = \lim_{n \rightarrow \infty} \hat{P}_i \quad (8.9)$$

pertanto, i metodi empirici descritti forniscono stime della distribuzione di probabilità e della densità di probabilità la cui accuratezza cresce con la dimensione del valore considerato (ciò è vero, in generale, per ogni stima statistica). Per quanto riguarda la densità è, inoltre, richiesto che l'ampiezza dei sottointervalli, h , tenda a zero. In caso contrario, ciò che si stima è in valore medio di f in U_i , che non coincide, in generale, con f_{C_i} , come mostrato, schematicamente, in figura 8.5.

Teoricamente, dunque, il numero di classi deve tendere all'infinito. In generale, il numero di classi può essere legato alla dimensione n del campione attraverso la regola empirica:

$$C = 1 + 3.3 \log n \quad (8.10)$$

Una scelta di C sbagliata può influire significativamente sul risultato della stima. In generale, bassi valori di C non permettono di apprezzare i dettagli delle funzioni. Alti valori di C , al contrario, generano una stima "rumorosa". Per chiarire queste affermazioni si osservino i grafici di figura 8.6. Tali grafici mostrano la stima empirica della densità di probabilità di una popolazione la cui distribuzione è la somma di due normali di varianza unitaria ed aventi media 0 e 5, rispettivamente:

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{\exp(-0.5x^2) + \exp(-0.5(x-5)^2)}{2}$$

Si vede come per $C = 5$ la stima della densità non fornisce i dettagli della distribuzione, per $C = 100$ la stima fornisce dettagli ma l'errore rispetto alla densità vera può essere, localmente, molto forte (rumore di stima). Per $C = 1 + 3.3 \log 200 \approx 18$ la stima è sufficientemente aderente alla densità della popolazione.

Si noti, infine, che la partizione può essere anche realizzata con intervalli di ampiezza disuguale, non uniforme. Ciò è particolarmente utile per distribuzioni aventi "code" fortemente allungate come, ad esempio, la distribuzione esponenziale, lognormale, di Gumbel, ecc.

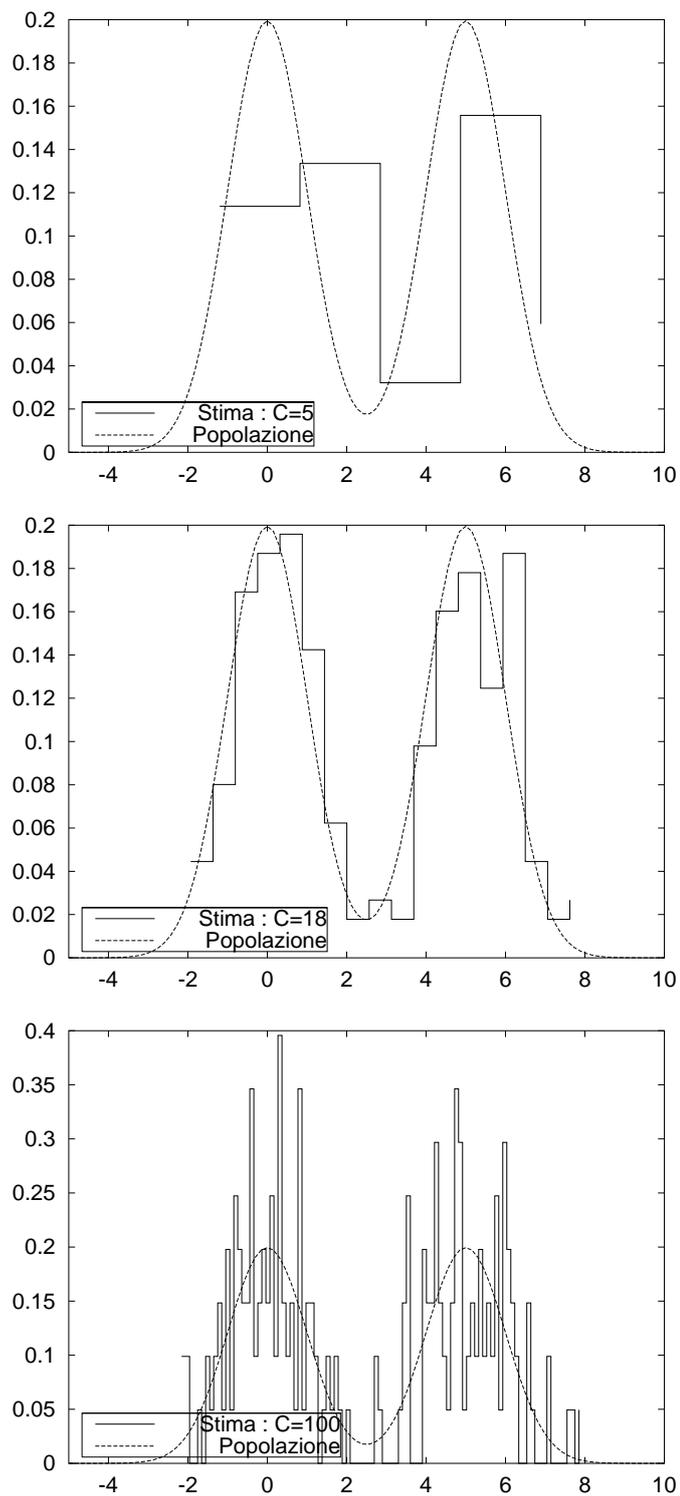


Figura 8.6: Calcolo della distribuzione empirica di una popolazione effettuata a partire da un campione di numerosità $n = 200$.

8.2 Percentili

Il procedimento per la stima empirica dei percentili é analogo a quello fatto per stimare la distribuzione di probabilità. Infatti, il primo passo é quello di ordinare i dati nel campione per ordine crescente. Se $x_1, \dots, x_i, \dots, x_n$ sono i dati ordinati, ad x_i si associa la probabilità i/n , stima della distribuzione in x_i . Il percentile a probabilità P é il numero x_i per cui $i/n \leq P$. Eventualmente, la stima può essere migliorata interpolando la distribuzione di probabilità empirica.

8.3 Carte probabilistiche

Il metodo delle carte probabilistiche consente di verificare rapidamente una ipotesi sulla distribuzione di probabilità della popolazione. Tale ipotesi, detta ipotesi nulla, si indica sinteticamente come:

$$H_0 : F_x(x) \in \mathcal{F}_0 \quad (8.11)$$

cioé si ipotizza che la VA \mathbf{x} segua una distribuzione di probabilità appartenente ad una famiglia parametrica prefissata, \mathcal{F}_0 . Per essere concreti, si può ipotizzare che la popolazione segua una distribuzione normale. Oppure, che la distribuzione segua una legge di Weibull.

Il metodo delle carte probabilistiche si basa sull'esistenza di due opportune funzioni, che indicheremo come Φ e Ψ . La funzione Ψ , se esiste, é tale che i punti $(\Phi(x), \Psi(F(x)))$ si dispongono secondo una retta. Ad esempio, per una distribuzione esponenziale:

$$F(x) = 1 - \exp(-\lambda x) \quad (8.12)$$

la funzione Φ é la funzione identità ($\Phi(x) = x$), la funzione $\Psi(F)$ (si omette la dipendenza da x per semplicitá) é la seguente:

$$\Psi(F) = -\log(1 - F(x)) \quad (8.13)$$

I punti $(x, -\log(1 - F(x)))$ si dispongono infatti secondo rette di pendenza pari a λ .

Generalizzando, si supponga di avere un campione \mathcal{X} consistente di n misure di una VA. Come si applica il metodo delle carte probabilistiche? La procedura é la seguente:

- scegliere una famiglia parametrica \mathcal{F}_0 e, se esistono, le appropriate funzioni Φ e Ψ ;
- ordinare il campione in modo che $x_1 \leq x_2 \dots \leq x_n$;
- stimare $F(x_i)$ mediante, ad esempio, $\hat{F}(x_i) = (i - 0.5)/n$;
- calcolare $\Phi(x)$ e $\Psi(\hat{F})$;
- verificare se le coppie $(\Phi(x), \Psi(\hat{F}))$ si dispongono, approssimativamente, in linea retta. Se ciò é vero, allora si accetta l'ipotesi H_0 ;

L'ultimo punto, cioé la verifica se i punti si dispongono secondo una linea retta o no, può essere fatta, come suggerito sopra, in modo soggettivo. Alternativamente, dopo avere verificato che i punti non siano disposti secondo una legge fortemente non lineare, é possibile calcolare il coefficiente di correlazione campionario \hat{R}_{xy} (il procedimento di stima sará discusso in seguito). Se tale coefficiente é prossimo a 1, allora si accetta l'ipotesi H_0 .

Per determinare $\hat{F}(x_i)$ é possibile utilizzare, alternativamente, le seguenti espressioni:

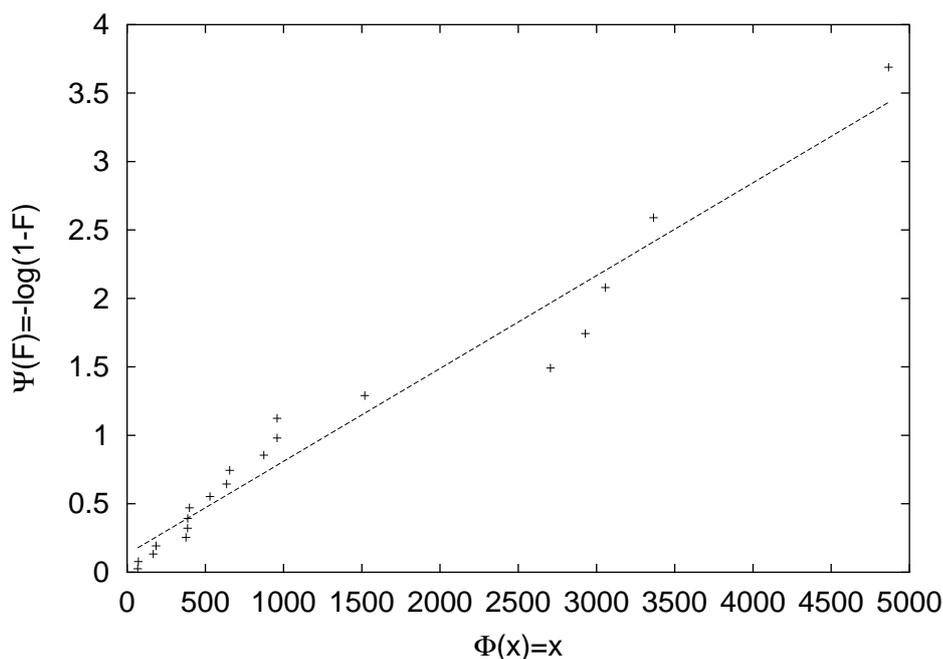


Figura 8.7: Rappresentazione mediante carta probabilistica di dati distribuiti secondo una legge esponenziale di parametro $\lambda = 0.001$.

$$\hat{F}(x_i) = \frac{i - 0.5}{n} \quad (8.14)$$

$$\hat{F}(x_i) = \frac{i}{n + 1} \quad (8.15)$$

$$\hat{F}(x_i) = \frac{i - 0.5}{n + 0.5} \quad (8.16)$$

Tali espressioni sono approssimate e sono state derivate in base all'ipotesi di equiprobabilità degli eventi elementari. Una stima esatta (ma piuttosto complessa) richiede l'utilizzo della funzione Beta.

Un esempio di applicazione della procedura discussa (con $\hat{F}(x_i) = (i - 0.5)/n$) è mostrato in figura 8.7 per dati distribuiti secondo una legge esponenziale di parametro $\lambda = 0.001$.

Alcune coppie di funzioni Φ e Ψ per distribuzioni di uso comune sono riportate in seguito:

$$\text{Esponenziale} \begin{cases} \Phi(x) = x \\ \Psi(F) = -\log(1 - F) \end{cases}$$

$$\text{Weibull} \begin{cases} \Phi(x) = \log(x) \\ \Psi(F) = \log(-\log(1 - F)) \end{cases}$$

$$\text{Normale} \begin{cases} \Phi(x) = x \\ \Psi(F) = \sqrt{2}\text{erf}^{-1}(2F - 1) \end{cases}$$

$$\text{Lognormale} \begin{cases} \Phi(x) = \log(x) \\ \Psi(F) = \sqrt{2}\text{erf}^{-1}(2F - 1) \end{cases}$$

8.4 Stima empirica di momenti e percentili

8.4.1 Valore atteso

Per il momento di ordine 1, cioè il valore atteso, lo stimatore empirico è il valore medio dei dati nel campione. Quindi, a seconda che i dati siano raggruppati o meno in classi, si possono utilizzare le seguenti espressioni:

$$\hat{\eta} = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k \quad (8.17)$$

$$\hat{\eta} = \bar{x} = \sum_{i=1}^C v_i \frac{n_i}{n} \quad (8.18)$$

dove n è la numerosità del campione, C il numero di classi, v_i il valore assunto dalla VA in ciascuna classe ed n_i il numero di osservazioni in ciascuna classe.

La naturale interpretazione dell'esperimento aleatorio multiplo appena compiuto è che sono stati ripetuti n esperimenti indipendenti, ognuno dei quali ha fornito un valore (realizzazione) della VA \mathbf{x} . Al fine di costruire modelli statistici, un esperimento come quello descritto può essere considerato in modo differente. Infatti, è possibile considerare che i dati x_1, \dots, x_n siano ottenuti dall'osservazione di n variabili aleatorie $\mathbf{x}_1, \dots, \mathbf{x}_n$ identicamente distribuite ed indipendenti (i.i.d.). Dunque, qualsiasi stimatore è in realtà una funzione di n variabili aleatorie e la stima è, pertanto, una variabile aleatoria anch'essa. Nel seguito saranno mostrate alcune proprietà degli stimatori. Qui, verifichiamo che il valore atteso della stima del valore atteso coincide con il valore atteso stesso. Infatti:

$$\mathbf{E}[\hat{\eta}] = \mathbf{E}\left[\frac{1}{n} \sum_{k=1}^n \mathbf{x}_k\right] = \frac{1}{n} \sum_{k=1}^n \mathbf{E}[\mathbf{x}_k] = \mathbf{E}[\mathbf{x}] \quad (8.19)$$

dove la penultima eguaglianza è stata ottenuta sfruttando la linearità del valore atteso, l'ultima sfruttando il fatto che la distribuzione (e quindi il valore atteso) di ogni variabile aleatoria \mathbf{x}_k coincide con quella di \mathbf{x} . Uno stimatore il cui valore atteso coincide con il valore del parametro che si intende stimare si dice non polarizzato.

Si ricorda che $\sigma_{\bar{x}}^2 = \sigma_x^2/n$ e che il lemma di Tchebycheff asserisce che, per una VA \mathbf{z} di valore atteso η_z e varianza σ_z^2 è:

$$\Pr(|\mathbf{z} - \eta_z| < \epsilon) \geq 1 - \sigma_z^2/\epsilon^2$$

È possibile quindi affermare che, per il valore medio di una VA \mathbf{x} è:

$$\Pr(|\bar{\mathbf{x}} - \mathbf{E}[\mathbf{x}]| < \epsilon) \geq 1 - \delta \quad (8.20)$$

dove

$$\delta = \frac{\sigma_x^2}{n\epsilon^2} \quad (8.21)$$

Pertanto, lo scostamento fra media campionaria e valore atteso di \mathbf{x} può essere reso arbitrariamente piccolo se è possibile controllare la dimensione del campione.

8.4.2 Varianza

La varianza é definita come:

$$\mathbf{E}[(\mathbf{x} - \mathbf{E}[\mathbf{x}])^2] = \mathbf{E}[\tilde{\mathbf{x}}^2] \quad (8.22)$$

L'ultima espressione, in cui $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{E}[\mathbf{x}]$ suggerisce che, conoscendo il valore atteso della variabile aleatoria, la stima della varianza potrebbe essere ottenuta calcolando $\tilde{x}_k = x_k - \mathbf{E}[\mathbf{x}]$ e quindi utilizzando la stima del valore atteso basata sul valore medio (eq. (8.17)). Tuttavia, salvo casi particolari, la stima della varianza implica anche una stima congiunta del valore medio. E' possibile dimostrare che:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2 \quad (8.23)$$

fornisce una stima priva di errore sistematico della varianza, quindi $\mathbf{E}[\hat{\sigma}^2] = \sigma^2$. La dimostrazione é lasciata al lettore.

Espressioni equivalenti della varianza sono:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n x_k^2 - n\bar{x}^2 \quad (8.24)$$

algebricamente equivalente a (8.23), Se i dati sono raggruppati in C classi di valore v_i e numero di osservazioni n_i si utilizza:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^C v_i^2 n_i - n\bar{x}^2 \quad (8.25)$$

Si noti infine che $\sqrt{\hat{\sigma}^2}$ é una stima dello scarto quadratico medio, quindi:

$$\hat{\sigma}_x = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (8.26)$$

Legge di probabilit  della varianza campionaria

E' possibile dimostrare che se \mathbf{x} segue una distribuzione normale di deviazione standard σ , allora:

$$t = (n-1) \frac{\hat{\sigma}}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \quad (8.27)$$

segue la distribuzione chi-quadrato con $n-1$ gradi di libert .

8.4.3 Covarianza e correlazione empiriche

La covarianza é definita come:

$$C_{xy} = \mathbf{E}[(\mathbf{x} - \mathbf{E}[\mathbf{x}])(\mathbf{y} - \mathbf{E}[\mathbf{y}])] = \mathbf{E}[\mathbf{x}\mathbf{y}] - \mathbf{E}[\mathbf{x}]\mathbf{E}[\mathbf{y}] \quad (8.28)$$

Si osservi che la varianza pu  essere interpretata come la covarianza di una VA con se stessa, dunque, procedendo in modo formalmente analogo a quanto fatto in precedenza, si pu  dimostrare che

$$\hat{C}_{xy} = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}) \quad (8.29)$$

e' una stima non polarizzata della covarianza. La dimostrazione, analoga alla precedente, é lasciata al lettore. La stima della correlazione é ottenibile da quella della covarianza come:

$$\hat{R}_{xy} = \frac{\hat{C}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \quad (8.30)$$

con

$$\hat{\sigma}_x = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (8.31)$$

$$\hat{\sigma}_y = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (y_k - \bar{y})^2} \quad (8.32)$$

8.4.4 Momenti

Il q-esimo momento centrato della variabile aleatoria ($\mathbf{E}[(x - \mathbf{E}[x])^q]$) puó essere stimato utilizzando la seguente espressione:

$$\hat{m}_q = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^q \quad (8.33)$$

Il corrispondente momento non centrato ($\mathbf{E}[x^q]$) é dato da:

$$\hat{k}_q = \frac{1}{n} \sum_{k=1}^n x_k^q \quad (8.34)$$