cheetSheet

Probabilità e (in)dipendenza

Probabilità condizionata e formula di Bayes

Probabilità condizionata

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \ \cap B) = P(A|B) \ \cdot \ P(B) = P(B|A) \ \cdot \ P(A)$$

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \ \cdot \ P(A_3|A_1 \cap A_2) \ \cdots P(A_n|A_1 \cap \dots A_{n-1})$$

Formula della probabilità totale o formula della fattorizzazione

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \ \cdot \ P(B_i)$$

dove $B_1, \dots, B_n$ è un sistema di alternative e A un evento

Formula di Bayes

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A|B_j) \cdot P(B_j)}$$

Dove A e B sono due eventi non trascurabili e $B_1, \dots, B_n$ è un sistema di alternative

Eventi indipendenti

Gli eventi A e B sono indipendenti se vale

$$P(A \ \cap \ B) = P(A) \ \cdot \ P(B)$$

Gli eventi $A_1, \dots, A_n$ sono indipendenti se $\forall \ 2 \leq k \leq n$ e $\forall \ 1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$ vale

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

Variabili Aleatorie

b-quantile

Un intero r t.c. $P(X \leq r) \geq \beta$ e $P(X \geq r) \leq 1 - \beta$

Valore Atteso

Caso Discreto	Caso con Densità
$E[X] = \sum_i x_i \cdot (x_i)$	$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx$

Proprietà del valore atteso

- $E[g(X)] = \sum_i g(x_i) \cdot p(x_i)$ oppure $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$
- $E[aX + b] = aE[X] + b$, in particolare $E[b] = b$
- $|E[X]| \leq E[|X|]$
- se $P(X \geq 0) = 1 \implies E[X] \geq 0$
- $E[X + Y] = E[X] + E[Y]$
- se $X \geq Y \implies E[X] \geq E[Y]$
- $E[XY] = E[X] \cdot E[Y]$ (se sono indipendenti)
- Disuguaglianza di Schwartz: $E[|XY|] \leq \sqrt{E[X^2]} \cdot \sqrt{E[Y^2]}$

Momento

Si dice momento di ordine $n = 1, 2, \ldots$ se $E[|X|^n] < \infty$

Disuguaglianza di Markov

$$a \cdot P(X \geq a) \leq E[X] \quad \forall\, a > 0$$

Varianza

$$Var(X) = E[X^2] - E[X]^2$$

Scarto quadratico medio o Deviazione standard

$$\sigma(X) = \sqrt{Var(X)}$$

Disuguaglianza di Chebyshev

$$P(|X - E[X]| > d) \leq \frac{Var(X)}{d^2} \qquad \forall\, d > 0$$

Variabili Aleatorie notevoli

Variabili Binomiali

Probabilità di ottenere h successi in un esperimento ripetuto n volte con probabilità di successo p. Detta variabile $B(n, p)$

$$P(X = h) = \binom{n}{h} \cdot p^h \cdot (1 - p)^{n-h}$$

VALORE ATTESO

$E[X] = np$

VARIANZA

$Var(X) = np(1 - p)$

Variabili di Bernulli

Se una variabile binomiale ha $n = 1$, per semplicità la chiamiamo di Bernulli e scriviamo $B(p)$

VALORE ATTESO

$E[X] = p$

VARIANZA

$Var(X) = p(1 - p)$

Variabili Geometriche

Conta il numero di volte in cui dobbiamo ripetere un esperimento con probabilità di successo p, per ottenere il primo successo

$$P(X = h) = (1 - p)^{h-1} \cdot p$$

Variabile di Poisson

É una buona approssimazione della distribuzione binomiale quando n è grande, p è piccolo e $n \cdot p$ vale circa λ

$$P(X = h) = e^{-\lambda} \cdot \frac{\lambda^h}{h!}$$

VALORE ATTESO

$E[X] = \lambda$

VARIANZA

$Var(X) = \lambda$

Variabili uniformi su intervalli

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & altrimenti \end{cases}$$

VALORE ATTESO

$E[X] = \frac{a+b}{2}$, $E[X^2] = \frac{a^2+ab+b^2}{3}$

VARIANZA

$Var(X) = \frac{(b-a)^2}{12}$

Variabili esponenziali

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

VALORE ATTESO

$E[X] = \frac{1}{\lambda}$

VARIANZA

$Var(X) = \frac{1}{\lambda^2}$

Variabili Gaussiane

GAUSSIANE STANDARD

Indicate come $N(0, 1)$, hanno densità

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

e funzione di ripartizione

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

Valore atteso

Momenti di ordine pari	Momenti di ordine dispari
$E[X^n] = 0$	$E[X^{2h+2}] = (2h+1) \cdot E[X^{2h}]$

in particolare $E[X^2] = 1$

Varianza

$Var(X) = 1$

GAUSSIANE NON STANDARD

Indicate come $N(m, \sigma^2)$, dove m e σ sono i suoi parametri, ha densità

$$f(t) = \frac{1}{\sigma} \cdot \varphi(\frac{t-m}{\sigma})$$

e funzione di ripartizione

$$F(t) = \Phi(\frac{t-m}{\sigma})$$

VALORE ATTESO

$E[X] = m$

VARIANZA

$Var(X) = \sigma^2$

Cambio di Variabile

Data X con densita f_X, una funzione $h : A \to B$ biunivoca e $Y = h \circ X$,

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \cdot \left| \frac{dh^{-1}(y)}{dy} \right| & y \in B \\ 0 & y \notin B \end{cases}$$

Variabili Aleatorie Multivariate

Indipendenza di v.a.

Dati gli eventi $A_1, \ldots, A_n$ e le v.a. $X_1, \ldots, X_n$, queste sono indipendenti se vale

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$$

Caso Discreto	Caso con Densità
$p(x_i, y_j) = p_X(x_i) \cdot P_Y(y_j)$	$f(x, y) = f_X(x) \cdot f_Y(y)$

Combinazione di v.a. discrete

Dati X, Y discrete, a valori naturali e indipendenti, sia $Z = X + Y$

$$p_Z(n) = \sum_{h=0}^{n} p_X(h) \cdot p_Y(n-h)$$

COMBINAZIONE DI V.A. BINOMIALI

Dati $X : B(n, p)$ e $Y : B(m, p)$, la v.a. $Z = X + Y$ è binomiale $B(n + m, p)$

Formula della convoluzione

Dati X, Y indipendenti e con densità, la v.a. $Z = X + Y$ ha densità

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z - y) dy$$

Covarianza e Correlazione

Covarianza

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

Coefficiente di correlazione

(misura della dipendenza lineare tra X e Y)

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

PROPRIETÀ

- $|\rho(X,Y)| \leq 1$
- $min_{a,b \in R^2} E[(Y-a-bX)^2] = Var(Y) \cdot (1-\rho(X,Y)^2)$

Indipendenti e equidistribuite (i.i.d.)

le v.a. $X_1, \ldots, X_2$ sono i.i.d. se hanno tutte la stessa c.d.f.

$$P_{X_n}(t) = P(X_n \leq t) = F(t) \quad \text{per } n = 1, 2, \ldots$$

e se vale

$$P(X_{k_1} \leq t_1, \ldots, X_{k_n} \leq t_n) = F_{X_1}(t1) \cdots F_{X_n}(t_n)$$

Legge dei grandi numeri

Data una successione di variabili i.i.d. $X_1, X_2, \ldots$ e il loro valore atteso $\mu = E[X_i]$, allora X_n converge in distribuzione a μ, si ha quindi

$$\lim_{n \to \infty} P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mu\Big| > \varepsilon\Big) = 0$$

Teorema centrale del limite

Data una successione di variabili i.i.d. $X_1, X_2, \ldots$, il loro valore atteso $\mu = E[X_i]$ e la loro varianza $\sigma^2(X_i) = \sigma^2 > 0$, presi $-\infty \leq a \leq b \leq +\infty$, si ha

$$\lim_{n \to \infty} P(a \leq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq b) = \Phi(b) - \Phi(a)$$

Diciamo quindi che $\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ converge in distribuzione a una gaussiana standard.

PROPOSIZIONE

$$\lim_{n \to \infty} P(a \leq \sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \leq b) = \Phi(b) - \Phi(a)$$

Densità Gamma

Indicata con $\Gamma(r, \lambda)$ dove r e λ sono i parametri

$$f(x) = \begin{cases} \frac{1}{\Gamma(r)}\lambda^r x^{r-1} e^{-\lambda x} & x > 0 \\ 0 & x \geq 0 \end{cases}$$

dove $\int_0^{+\infty}(\lambda x)^{r-1}e^{-\lambda x}\lambda dx = \int_0^{+infty}t^{r-1}e^{-t}dt = \Gamma(r)$

MOMENTI

Una variabile X con densità $\Gamma(r,\lambda)$ ha tutti i momenti, e $\forall\ \beta > 0$ si ha

$$E[X^\beta] = \frac{\Gamma(r+\beta)}{\Gamma(r)\lambda^\beta}$$

COMBINAZIONE

Siano X con densità $\Gamma(r,\lambda)$ e Y con densità $\Gamma(s,\lambda)$, la variabile $X+Y$ avrà densità $\Gamma(r+s,\lambda)$

Densità chiquadro

Date $X_1,\dots,X_n$ v.a. gaussiane standard indipendenti, indichiamo la variabile $(X_1^2+\dots+X_n^2)$ come

$$\chi^2(n) \text{ ed avrà densità } \Gamma(\frac{n}{2},\frac{1}{2})$$

e la chiamiamo *Densità chi-quadro a n gradi di libertà*

Densità di student

Date X, C_n due variabili indipendenti con densità $N(0,1)$ e $\chi^2(n)$ $(n \geq 1)$, La variabile $T_n = \frac{X}{\sqrt{\frac{C_n}{n}}} = \sqrt{n}\frac{X}{\sqrt{C_n}}$ ha densità

$$f_{T_n}(t) = \frac{\Gamma(n/2+1/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+\frac{t^2}{n})^{-n/2-1/2}$$

e la chiamiamo *densità di Student a n gradi di libertà*

Campioni di Variabili Aleatorie

Funzione di Verosimiglianza

la funzione $L(\theta;\dots)$ definita come

Caso Discreto	Caso con densità
$L(\theta;x_1,\dots,x_n) = \prod_{i=1}^n p_\theta(x_i)$	$L(\theta;x_1,\dots,x_n) = \prod_{i=1}^n f_\theta(x_i)$

Stima di massima verosimiglianza

Se esiste, è una statistica campionaria, indicata con $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$, t.c. valga l'eguaglianza

$$L(\hat{\theta}; x_1, \ldots, x_n) = \max_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n) \quad \forall\, (x_1, \ldots, x_n)$$

Stima col metodo dei momenti

Se esiste, è una statistica campionaria $\tilde{\theta} = \tilde{\theta}(x_1, \ldots, x_n)$ che permetta di eguagliare alcuni momenti teorici con i corrispondenti moomenti empirici, cioè di scrivere, per uno o più interi positivi k

$$E_{\tilde{\theta}}[X^k] = \frac{1}{n} \sum_{i_1}^{n} x_i^k, \quad \forall\, (x_1, \ldots, x_n)$$

Intervalli di fiducia

Intervalli di fiducia per la media di un campione Bernulli

$$\left[\bar{X}_n \pm \sqrt{\tfrac{\bar{X}_n(1-\bar{X}_n)}{n}} q_{1-\frac{\alpha}{2}}\right]$$

è un intervallo di fiducia per la media p del campione $X_1, \ldots, X_n$ con livello di fiducia approssimativamente $1-\alpha$

Intervalli di fiducia per la media di un campione di taglia grande

INTERVALLO DI FIDUCIA PER LA MEDIA, CAMPIONI GRANDI, VARIANZA NOTA

$$\left[\bar{X}_n \pm \tfrac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$$

è un intervallo di fiducia per la media m del campione con livello di fiducia approssimativamente $1-\alpha$

INTERVALLO DI FIDUCIA PER LA MEDIA, CAMPIONI GRANDI, VARIANZA NON NOTA

$$\left[\bar{X}_n \pm \tfrac{S_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$$

è un intervallo di fiducia per la media m del campione con livello di fiducia apprssimativamente $1-\alpha$

Intervalli di fiducia per la varianza di un campione gaussiano

$$\left(0, \frac{\sum_{i=1}^n (X_i - \bar{X}_i)^2}{\chi^2_{\alpha,n-1}}\right] = \left(0, \frac{(n-1)S_n^2}{\chi^2_{\alpha,n-1}}\right], \quad \left[\frac{\sum_{i=1}^n (X_i - \bar{X}_i)^2}{\chi^2_{1-\alpha,n-1}}, +\infty\right) = \left[\frac{(n-1)S_n^2}{\chi^2_{1-\alpha,n-1}}, +\infty\right)$$

sono intervalli di fiducia unilateri per σ^2 con livello di fiducia $1-\alpha$

Test Statistici

Z-test

Test sulla media di un campione gaussiane con varianza nota

REGIONE CRITICA

$$\frac{d\sqrt{n}}{\sigma} = q_{1-\frac{\alpha}{2}}, \quad C = \left\{\sqrt{n}\frac{|\bar{X}-m_0|}{\sigma} > q_{1-\frac{\alpha}{2}}\right\} = \left\{|\bar{X} - m_0| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right\}$$

In pratica di fronte a dati concreti $x_1, \dots, x_n$ si calcola la media empirica $\bar{x}_n$ e si rifiuta l'ipotesi $\mathcal{H}_0$ se $|\bar{x} - m_0| > \frac{\sigma \cdot q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$, si accetta $\mathcal{H}_0$ in caso contrario

CALCOLO DEL P-VALUE

$$\begin{aligned} \bar{\alpha} = \bar{\alpha}(x_1, \dots, x_n) = P_{m_0}\left\{\sqrt{n}\frac{|\bar{X}-m_0|}{\sigma} > \frac{\sqrt{n}}{\sigma}|\bar{x} - m_0|\right\} = \\ = 2\left[1 - \Phi\left(\frac{\sqrt{n}}{\sigma}|\bar{x} - m_0|\right)\right] \end{aligned}$$

CALCOLO DELLA CURVA OPERATIVA

(probabilità di accettare quando m è il valore del parametro) al livello α

$$\begin{aligned} \beta(m) = P_m\left\{-q_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}}{\sigma}(\bar{X} - m_0) \leq_{1-\frac{\alpha}{2}}\right\} = \\ = \Phi\left(\sqrt{n}\frac{(m_0-m)}{\sigma} + q_{1-\frac{\alpha}{2}}\right) - \Phi\left(\sqrt{n}\frac{(m_0-m)}{\sigma} - q_{1-\frac{\alpha}{2}}\right) =: \\ =: h\left(\sqrt{n}\frac{m-m_0}{\sigma}\right) \end{aligned}$$

Test sulla media di un campione gaussiane con varianza sconosciuta o T-test

REGIONE CRITICA

$$C = \left\{ \sqrt{n} \frac{|\bar{X} - m_0|}{S} > \tau_{(1-\frac{\alpha}{2}, n-1)} \right\}$$

P-VALUE

$$\bar{\alpha} = P\left\{ \sqrt{n} \frac{|\bar{X} - m_0|}{S} > \frac{\sqrt{n}}{s} |\bar{x} - m_0| \right\} = 2 \left[1 - F_{n-1} \left(\frac{\sqrt{n}}{s} |\bar{x} - m_0| \right) \right]$$

dove $\bar{x}$ e $\bar{s}$ sono la media e la varianza campionaria di $(x_1, \ldots, x_n)$ e F_k è la c.d.f. di $T(k)$

Test approssimato sulla media di un campione di taglia grande

Per il teorema centrale del limite si può approssimare a un campione gaussiano

REGIONE CRITICA

$$C = \left\{ \sqrt{n} \frac{|\bar{X} - m_0|}{\sigma} > q_{1-\frac{\alpha}{2}} \right\}$$

P-VALUE

$$\bar{\alpha} = P\left\{ |Z| > \frac{\sqrt{n}}{\sigma} |\bar{x} - m_0| \right\} \approx 2 \left[1 - \Phi \left(\frac{\sqrt{n}}{\sigma} |\bar{x} - m_0| \right) \right]$$

Test sulla varianza di un campione gaussiano

REGIONE CRITICA

$$C = \left\{ \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2} > \chi^2_{1-\alpha, n-1} \right\}$$

P-VALUE

$$\bar{\alpha} = P_{\sigma_0} \left\{ \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2} > \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} \right\} = 1 - G_{n-1} \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} \right)$$

dove G_n è la c.d.f. della variabile $\chi^2(n)$

CURVA OPERATIVA

$$\beta(\sigma) = P_\sigma \left\{ \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma_0^2} \leq \chi^2_{1-\alpha,n-1} \right\} = G_{n-1}\left(\frac{\sigma_0^2}{\sigma^2} \chi^2_{1-\alpha,n-1} \right)$$