VARIABILI CASUALI

Lo spazio campionario può contenere eventi complicati da descrivere e catalogare al fine di costruire su Ω una famiglia di eventi interessante.

In molte applicazioni la complessità dello spazio campionario viene affrontata considerando come EVENTI insiemi di eventi elementari ai quali è associato un valore numerico preciso o compreso in un certo intervallo.

Ciò risponde alla esigenza applicativa che spesso nelle applicazioni ciò che interessa non è l'evento elementare verificatesi ma uno o più valori numerici associati all' evento.

DEFINIZIONE DI VARIABILE CASUALE.

Alle precedenti esigenze fa fronte la seguente definizione composta di **due parti** e dove la seconda è quella fondamentale dal punto di vista del calcolo delle probabilità

Definizione: Dato uno spazio di probabilità Ω , *F*, *P*($\cdot$) si definisce **variabile casuale**:

1 una funzione $X(\cdot)$ con dominio Ω e codominio $\Re$ o un suo sottoinsieme

2 tale che ogni insieme di eventi elementari $\{e : X(e) \leq x\}$

appartiene alla famiglia *F*.

Si osservi che in base alla condizione due

$$\{e : X(e) \le x\}$$

è un evento e quindi esiste la probabilità

$$P\left(\{e : X(e) \le x\}\right).$$

Come conseguenza della seconda parte della definizione e delle proprietà della famiglia di eventi si ha che sono eventi anche gli insiemi:

$\{e : x_1 < X(e) \le x_2\}\,, \{e : x_1 \le X(e) \le x_2\}\,,$

$\{e : x_1 \le X(e) < x_2\}\,, \{e : x_1 < X(e) < x_2\}\,,$

$\{e : X(e) = x_2\}$, i loro complementi e......

FUNZIONE DI RIPARTIZIONE

Data una variabile casuale X *definita su uno spazio di probabilità* Ω , F, $P(\cdot)$ *la seguente funzione*

$$F(x) = P\left(\{e : X(e) \leq x\}\right)$$

viene chiamata funzione di ripartizione della variabile casuale X.

Si noti che una funzione di ripartizione è necessariamente:

- non negativa

- a valori in $[0 \quad 1]$

- monotona non decrescente

- continua o continua da destra

Le prime tre proprietà sono ovvie conseguenze degli assiomi del C.P.. L'ultima proprietà è più delicata e la sua dimo. è omessa.

D'ora in poi ci scorderemo dello spazio di probabilità sottostante alla definizione di variabile casuale $X(\cdot)$ indicheremo una v.c. semplicemente X ed invece di studiare modelli probabilistici attraverso la tripletta Ω , *F*, $P(\cdot)$ *studieremo le coppie X ,F(·).*

Alcune prime regole importanti:

$$P\{e : x_1 < X(e) \leq x_2\} = F(x_2) - F(x_1)$$

$$P\{e : x < X(e)\} = 1 - F(x)$$

In seguito vedremo che per le variabili casuali con funzione di ripartizione continua

$$P\{e : X(e) = x\} = F(x) - F(x) = 0$$

mentre per la variabili casuali la cui funzione di ripartizione in x è solo continua da destra si ha che:

$$P\{e : X(e) = x\} = F(x) - F(x^-) \neq 0$$

V.C.CONTINUE E DISCRETE

Se il codominio di X è finito la variabile casuale è detta discreta semplice

Se il codominio di X è numerabile la variabile casuale è detta discreta

Se il codominio di X ha la potenza del continuo la variabile casuale è detta continua.

VARIABILI CASUALI DISCRETE

possono assumere al più una infinità numerabile di valori

$$x_i, i = 1, 2, 3, ...$$

Funzione di ripartizione (continua da destra monotona non decrescente costante a tratti)

$$F(x) = \sum_{x_i \leq x} p(x_i)$$

Funzione di probabilità:

$$p(x_i) = P(X = x_i)$$

TIPI PRINCIPALI: Binomiale, Ipergeometrica, Binomiale Negativa Poisson

VARIABILI CASUALI CONTINUE

Generalmente assumono valori in tutto $\Re$ o in $\Re^+$.

Noi considereremo quelle con F.d.R. *assolutamente continue* cioè:

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

dove la funzione integranda è detta *funzione di densità di probabilità.*

TIPI PRINCIPALI: Uniforme, Normale, Esponenziale, Gamma, Weibull.

CASI DI INTERESSE STATISTICO: T di Student, F di Fisher, ChiQuadro.

INDICATORI STATISTICI DI POSIZIONE E DISPERSIONE

MEDIA ARITMETICA E VARIANZA

PREMESSA:

media e varianza di vettore di dati sperimentali $\mathbf{x} = [x_1, x_2, ..., x_N]'$:

$$m = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$v^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m)^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2 - m^2$$

media e varianza di una distribuzione di frequenza:

itensità	x_1	x_2	..	x_c
frequenze	f_1	f_2	..	f_c

$$m = \sum_{i=1}^{c} x_i f_i$$

$$v^2 = \sum_{i=1}^{c} (x_i - m)^2 f_i = \sum_{i=1}^{c} x_i^2 f_i - m^2$$

VALORE ATTESO E VARIANZA

di una v.c. DISCRETA SEMPLICE

$$\mu = E(X) = \sum_{i=1}^{c} x_i p(x_i)$$

$$\sigma^2 = Var(X) = \sum_{i=1}^{c} (x_i - \mu)^2 p(x_i) =$$
$$= \sum_{i=1}^{c} x_i^2 p(x_i) - \mu^2$$

se gli x_i son i primi k interi nonegativi spesso conviene usare la

seguente formula facilmente dimostrabile:

$$\sigma^2 = \sum_{i=1}^{c} x_i(x_i - 1)p(x_i) + \mu - \mu^2$$

VALORE ATTESO E VARIANZA

di una v.c. DISCRETA

$$\mu = E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

$$\sigma^2 = Var(X) = \sum_{i=1}^{\infty} (x_i - \mu)^2 p(x_i) =$$
$$= \sum_{i=1}^{\infty} x_i^2 p(x_i) - \mu^2$$

se gli x_i sono gli interi nonegativi spesso conviene usare la

seguente formula facilmente dimostrabile:

$$\sigma^2 = \sum_{i=1}^{\infty} x_i(x_i - 1)p(x_i) + \mu - \mu^2$$

VALORE ATTESO E VARIANZA

di una v.c. continua con F.d.R.
assolutamente continua

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx =$$
$$= \bigvee \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

VARIABILE CASUALE INDICATORE

numero di successi in una prova

Si considera un esperimento casuale che può dar luogo a due possibili risultati S: successo $\bar{\text{S}}$: insuccesso e sia π la probabilità di S.

La variabile casuale indicatore I assume valore uno se si verifica S e zero altrimenti.

Funzione di probabilità:

Valore	Probabilità
1	π
0	$1-\pi$

VALORE ATTESO E VARIANZA

$\mu = \pi \qquad \sigma^2 = \pi(1-\pi)$

VARIABILE CASUALE BINOMIALE

numero di successi in n prove indipendenti

Si ripete n **volte** un esperimento casuale che può dar luogo a due possibili risultati S: successo $\bar{\text{S}}$: insuccesso. Sia π la probabilità di S.

L'esperimento è ripetuto in modo che:

- le n prove sono indipendenti
- la probabilità π di successo non cambia di prova in prova

  La variabile casuale discreta semplice X numero di ripetizioni dell'esperimento che danno luogo ad un successo è chiamata variabile casuale Binomiale. Le possibili determinazioni della Binomiale sono : $0, 1, 2, ..., n$.

La funzione di probabilità della Binomiale deriva dalle tre semplici considerazioni:

- una realizzazione delle n ripetizioni è una stringa 00110101...01

  che contiene 1 nella i-esima posizione se nella i-esima prova si è verificato un successo e 0 altrimenti.

- tutte le stringhe che contengono x volte il numero 1 (il successo si è verificato x volte su n) hanno probabilità $\pi^x(1-\pi)^{n-x}$.

- le stringhe di cui al punto precedente sono $\binom{n}{x}$

Dalle precedenti considerazioni e dall'assioma di additività deriva che:

$$p(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

VALORE ATTESO E VARIANZA della BINOMIALE

$$\mu = \sum_{x=0}^{n} x \binom{n}{x} \pi^{x} (1-\pi)^{n-x} = n\pi$$

$$\begin{aligned} \sigma^2 &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} \pi^{x} (1-\pi)^{n-x} + n\pi - n^2\pi^2 = \\ &= n\pi(1-\pi) \end{aligned}$$

PROCESSI POISSONIANI E VARIABILE CASUALE DI POISSON

Si considera un evento che ricorre nel tempo in modo casuale

(es: interruzioni di energia elettrica, chiamate a un centralino di pronto intervento, infortuni sul lavoro, incidenti stradali, richieste di intervento per manutenzione ecc.)

in modo che:

- Le variabili casuali $N(t, t + \Delta t)$ numero di ricorrenze nell'intervallo

  $(t, t+\Delta t)$ hanno funzione di probabilità che dipende dall'ampiezza dell'intervallo Δt ma non dalla origine t (assunzione di stazionarietà)

- Le variabili casuali $N(t_1, t_2)$ e $N(t'_1, t'_2)$ sono indipendenti se si riferiscono ad intervalli disgiunti cioè:

$$p\Big(N(t_1, t_2) = k \cap N(t'_1, t'_2) = h\Big) =$$
$$= p\Big(N(t_1, t_2) = k) \cdot p(N(t'_1, t'_2) = h\Big)$$

- $lim_{dt \to 0} \frac{pN(t,t+dt)=1)}{dt} = \lambda$ il parametro λ è detto intensità di frequenza

  per unità di tempo e in base a quanto appena detto si ha che

  $p(N(t, t + dt) = 1) \approx \lambda dt$ per ampiezze di intervallo infinitesimali

- $lim_{dt \to 0} \frac{pN(t,t+dt)>1)}{dt} = 0$ (principio degli eventi rari, formalizza l'idea che gli eventi ricorrano isolatamente e non a grappoli)

Sotto le precedenti assunzioni la variabile casuale $N(\Delta t)$, numero di eventi in un intervallo di ampezza Δt (si noti che l'origine dell'intervallo è irrilevante in base alla assunzione di stazionarietà), è una variabile casuale discreta con funzione di probabilità:

$$p(x) = \frac{(\lambda \Delta t)^x}{x!} e^{-\lambda \Delta t}$$

VALORE ATTESO E VARIANZA

$$\mu = \sum_{x=0}^{\infty} x \frac{(\lambda \Delta t)^x}{x!} e^{-\lambda \Delta t} = \lambda \Delta t$$

$$\sigma^2 = \sum_{x=0}^{\infty} x(x-1) \frac{(\lambda \Delta t)^x}{x!} e^{-\lambda \Delta t} + \lambda \Delta t - (\lambda \Delta t)^2 = \lambda \Delta t$$

RELAZIONE BINOMIALE POISSON

facile da dimostrare e da commentare

$$\lim_{\substack{n\to\infty \\ n\pi=\lambda}} \binom{n}{x}\pi^x(1-\pi)^{n-x} = \frac{\lambda^x}{x!}e^{-\lambda}$$

Dal punto di vista pratico se X è una binomiale con $n = 50000$ e $\pi = \frac{1}{10000}$ è un problema calcolare $p(X > 5)$ ma in base al precedente risultato tale probabilità può essere approssimata usando la f.d.p. di una poisson con parametro $\lambda = n\pi = 50000\frac{1}{10000} = 5$

VARIABILE CASUALE BINOMIALE NEGATIVA

Sotto le stesse assunzioni descritte a proposito della Binomiale

si considera la variabile casuale : numero di prove X necessarie per ottenere un numero r di successi.

X è detta variabile casuale Binomiale negativa.

Si noti che nel caso della binomiale il numero delle prove è una costante ed il numero di successi una variabile casuale

mentre nel caso della binomiale negativa il numero di successi è una costante nota mentre ilnumero di prove è una variabile casuale.

FUNZIONE DI PROBABILITA'

$$p(x) = \binom{x-1}{r-1}\pi^r(1-\pi)^{x-r}, x = r, r+1, r+2, ...$$

VALORE ATTESO VARIANZA

$$\mu = \frac{r}{\pi}, \sigma^2 = r\frac{1-\pi}{\pi^2}$$

La funzione di probabilità della binomiale negativa deriva da semplici considerazioni probabilistiche:

1-ogni realizzazione sperimentale è una successione 0001011......1 di lunghezza x contenente r elementi 1 (i successi) e x-r elementi 0 (gli insuccessi).

2- ogni stringa ha necessariamente un 1 in ultima posizione (l'erresimo successo). Ogniuna di queste successioni di lunghezza x ha probabilità $\pi^r(1-\pi)^{x-r}$

3-vi sono $\binom{x-1}{r-1}$ successioni di questo tipo (tante quante sono i modi con cui si possono disporre r-1 1 nelle prime $x-1$ posizioni.

Il resto ovvio (assioma di additività).

INTEGRALE FONDAMENTALE

(Funzione Gamma)

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

RISULTATI CHE NON VAN SCORDATI

$$\Gamma(1) = 1$$

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

$$\Gamma(n) = (n - 1)!$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

V.C. NORMALE STANDARD Z

densità di probabilità

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

funzione di ripartizione

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

valore atteso

$$E(z) = \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 0$$

varianza

$$Var(Z) = \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz =$$

$$2\int_{0}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz =$$

$$ponendo \qquad t = \frac{z^2}{2}$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\infty} t^{\frac{3}{2}-1} e^{-t} dt = \frac{2}{\sqrt{\pi}} \Gamma(\frac{3}{2}) =$$

$$= \frac{2}{\sqrt{\pi}} \frac{1}{2} \Gamma(\frac{1}{2}) = 1$$

V.C. NORMALE O DI GAUSS

La normale standard ha uno dei suoi più importanti impieghi nel modello ERRORE DI MISURA:

$$X = \mu + \sigma Z$$

dove X rappresenta la misura fornita dalla strumentazione di una quantità non nota μ, il multiplo della normale standard σZ rappresenta l'errore della strumentazione e σ è una costante non negativa tanto più grande quanto più imprecisa è la strumentazione.

La variabile casuale X precedentemente definita è chiamata variabile casuale normale o di Gauss.

Funzione di ripartizione e densità della variabile casuale normale

$$
\begin{aligned}
F(x) &= P(\mu + \sigma Z \leq x) = P(Z \leq \frac{x-\mu}{\sigma}) = \\
&= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \Phi(\frac{x-\mu}{\sigma})
\end{aligned}
$$

derivando si ottiene la densità di probabilità:

$$
f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}
$$

VALORE ATTESO

$$
E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu
$$

VARIANZA

$$
Var(X) = Var(\mu + \sigma Z) = E(\sigma^2 Z^2) = \sigma^2
$$

VARIABILE CASUALE GAMMA

Variabile casuale importante per le sue relazioni con la Poisson e la Normale e per i casi particolari noti come variabile casuale esponenziale e variabile casuale chi-quadro

La variabile casuale Gamma ha densità di probabilità:

$$f(t) = \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}t^{\alpha-1}e^{-\lambda t} \qquad \alpha, \lambda > 0$$

$$F(x) = \int_0^x \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}t^{\alpha-1}e^{-\lambda t}dt =$$

$$ponendo\ \lambda t = z$$

$$= \int_0^{\lambda z} \frac{1}{\Gamma(\alpha)}z^{\alpha-1}e^{-z}dt = G(\lambda z)$$

$$E(X) = \frac{\alpha}{\lambda}, Var(X) = \frac{\alpha}{\lambda^2}$$

CALCOLO VALORE ATTESO E VARIANZA

$$E(X) = \int_0^\infty x \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx =$$

$$= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\lambda} \int_0^\infty \frac{1}{\Gamma(\alpha+1)} \lambda^{\alpha+1} x^{(\alpha+1)-1} e^{-\lambda x} dx =$$

$$= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\lambda} = \frac{\alpha}{\lambda}$$

$$E(X^2) = \int_0^\infty x^2 \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx =$$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{1}{\lambda^2} \int_0^\infty \frac{1}{\Gamma(\alpha+2)} \lambda^{\alpha+2} x^{(\alpha+2)-1} e^{-\lambda x} dx =$$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{1}{\lambda^2} = \frac{\alpha(\alpha+1)}{\lambda^2}$$

quindi

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

V.C. ESPONENZIALE

Ponendo $\alpha = 1$ nelle relazioni precedenti si ottiene la densità di probabilità e la funzione di ripartizione **della variabile casuale esponenziale** importatantissima insieme alla Weibull nei problemi di affidabilità (per modellare vita utensili e componenti).

$$f(x) = \lambda e^{-\lambda x}$$

$$F(x) = 1 - e^{-\lambda x}$$

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

VARIABILE CASUALE CHI QUADRO

Per il ruolo fondamentale che gioca nell'inferenza statistica va ricordato il caso di $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$ dove n è un intero.

La variabile casuale che così si ottiene è chiamata variabile casuale chi quadro con n gradi di libertà che ha densità di probabilità:

$$f(x) = \frac{1}{\Gamma(\frac{n}{2})} \frac{1}{2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}$$

RELAZIONI IMPORTANTI

Se un evento ricorre nel tempo secondo la legge di Poisson di parametro λ allora il tempo intercorrente tra un accadimento ed il successivo è una variabile casuale esponenziale di parametro λ mentre il tempo intercorrente tra un evento e l' r-esimo successivo è una variabile casuale gamma di parametri $\alpha = r$ e λ.

Se Z è una normale standard Z^2 è una chi quadro con un grado di libertà ($\alpha = \frac{1}{2}$)

Se $Z_i, i = 1, 2, ..., r$, sono normali standard indipendenti in probabilità $\sum_{i=1}^{r} Z_i^2$ è una chi quadro con r gradi di libertà ($\alpha = \frac{r}{2}$).

Definizione: le variabili casuali $Z_i, i = 1, 2, ..., r$, aventi funzione di ripartizione $F(Z_i), i = 1, 2, ..., r$, sono indipendenti in probabilità se

$$F(z_1, z_2, ..., z_r) =$$

$$= P(Z_1 \leq z_1, Z_2 \leq z_2, ..., Z_r \leq z_r) = \prod_{i=1}^{r} F(z_i)$$

V.C. WEIBULL

Come caso particolare comprende la variabile casuale esponenziale, è importante nel modellare la durata di molti sistemi.

Consideriamo la variabile casuale X=T$^{1/\beta}$ dove T è una esponenziale di parametro λ.

Si ha la seguente funzione di ripartizione di X:

$$\begin{aligned} F(x) &= P(T^{1/\beta} \leq x) = P(T \leq x^{\beta}) = 1 - e^{-\lambda x^{\beta}}, \\ \lambda &> 0, \beta > 0 \end{aligned}$$

derivando rispetto x si ottiene la densità di probabilità di X

$$f(x) = \beta \lambda x^{\beta - 1} e^{-\lambda x^{\beta}}, \lambda > 0, \beta > 0$$

Una semplice applicazione della definizione della funzione gamma porta a:

$$E(X) = \lambda^{-1/\beta}\Gamma(1 + \frac{1}{\beta})$$

$$Var(X) = \lambda^{-2/\beta}\left[\Gamma(1 + \frac{2}{\beta}) - \Gamma^2(1 + \frac{1}{\beta})\right]$$

Calcolo varianza e valore atteso

$$\begin{aligned}
E(X) &= \int_0^\infty x\beta\lambda x^{\beta-1}e^{-\lambda x^\beta}dx \\
pongo\ t &= \lambda x^\beta \\
&= \lambda^{-1/\beta}\int_0^\infty t^{\left(1+\frac{1}{\beta}\right)-1}e^{-t} = \\
&= \lambda^{-1/\beta}\Gamma(1+\frac{1}{\beta})
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= \int_0^\infty x^2\beta\lambda x^{\beta-1}e^{-\lambda x^\beta}dx \\
pongo\ t &= \lambda x^\beta \\
&= \lambda^{-2/\beta}\int_0^\infty t^{\left(1+\frac{2}{\beta}\right)-1}e^{-t} = \\
&= \lambda^{-2/\beta}\Gamma(1+\frac{2}{\beta})
\end{aligned}$$

La formula della varianza segue immediatamente.

VARIABILI CASUALI DI INTERESSE STATISTICO

Le seguenti variabili casuali **T di Student** e **F di Fisher** insieme alla già introdotta chi quadro sono importantissime per le loro applicazioni statistiche.

Qui se ne riporta solo la definizione rinunciando ad esporre le espressioni delle funzioni di densità e omettendo il calcolo di valore atteso e varianza.

Data l'enorme importanza applicativa di queste variabili casuali si ricorda però che l'uso consapevole delle tavole delle rispettive funzioni di ripartizione è assolutamente necessario.

Si ricordi infine che le variabili casuali Z_1, Z_2 aventi funzione di ripartizione $F(Z_i)$sono indipendenti in probabilità se:

$$P(Z_1 \leq z_1, Z_2 \leq z_2) = \prod_{i=1}^{2} F(z_i)$$

T di Student

Sia Z una variabile casuale normale standard e X una chi quadro con k gradi di libertà. Se Z e X sono indipendenti in probabilità allora il rapporto

$$T = \frac{Z}{\sqrt{\frac{V}{k}}}$$

è una variabile casuale di Student con k gradi di libertà. Per curiosità notiamo che questa variabile csuale ha una densità che ad occhio è difficilmente distinguibile da quella della normale standard sopratutto per $k > 30$. La V.C. T di Student ha valore atteso nullo e la varianza per $k > 2$ è pari a $\frac{k}{k-2}$

F di FiSHER

Se X e Y sono due variabili casuali chi quadro indipendenti aventi rispettivamente g_x e g_y gradi di libertà allora il rapporto

$$F = \frac{X/g_x}{Y/g_y}$$

è una variabile casuale F di Fisher con g_x gradi di libertà per il numeratore e g_y gradi di lbertà per il denominatore.

Lo studente curioso può trovare ulteriori dettagli in Montgomery -Runger (1999) Applied Statistics and Probability for Engineers - Wiley-New York.

SOMME DI VARIABILI CASUALI INDIPENDENTI

Se $X_i, i = 1, 2, ..., r$, sono binomiali indipendenti in probabilità con parametro numero di prove diverso $n_i, i = 1, 2, ..., r$,ma con la stessa probabilità di successo π,allora la somma $\sum_{i=1}^{r} Z_i$ è una binomiale di parametri $\sum_{i=1}^{r} n_i$ e π.

Se $X_i, i = 1, 2, ..., r$, sono Poisson indipendenti in probabilità con parametri $\lambda_i, i = 1, 2, ..., r$, allora la somma $\sum_{i=1}^{r} Z_i$ è una Poisson di parametro $\sum_{i=1}^{r} \lambda_i$.

Se $X_i, i = 1, 2, ..., r$, sono Normali indipendenti in probabilità con parametri $\mu_i, \sigma_i^2, i = 1, 2, ..., r$, allora la somma $\sum_{i=1}^r Z_i$ è una Normale di parametri $\sum_{i=1}^r \mu_i, \sum_{i=1}^r \sigma_i^2$.

Se $X_i, i = 1, 2, ..., r$, sono Gamma indipendenti in probabilità con parametri $\alpha_i, i = 1, 2, ..., r$, possibilmente diversi ma con lo stesso λ allora la somma $\sum_{i=1}^r Z_i$ è una Gamma di parametri $\sum_{i=1}^r \alpha_i, \lambda$.

Esercizio: Che variabile casuale è una somma di variabili chi quadro indipendenti in probabilità.

CONSIDERAZIONI SULLA MEDIA

CAMPIONARIA

DEFINIZIONE: Siano $X_i, i = 1, 2, ...n$, n variabili casuali aventi lo stesso valore atteso μ e la stessa varianza σ^2. La variabile casuale

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

è chiamata media campionaria e la variabile casuale:

$$\bar{Z}_n = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}$$

è detta media campionaria standardizzata.

Osservazioni:

- $E(\bar{X}_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = n\frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$

- se le X_i sono indipendenti in probabilità

$$var(\bar{X}_n) = var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}$$

- $E(\bar{Z}_n) = 0, \qquad var(\bar{Z}_n) = 1$

TEOREMA DI NORMALITA' ASINTOTICA

Siano $X_i, i = 1, 2, ..., n$, variabili casuali indipendenti in probabilità tutte con valore atteso μ e varianza σ^2. Sia

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{r} X_i$$

la variabile casuale media campionaria e

$$\bar{Z}_n = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}$$

sia la variabile casuale media campionaria standardizzata. Si indichi con $F_n(z)$ la funzione di ripartizione di $\bar{Z}_n$.

Si ha che:

$$\lim_{n \to \infty} F_n(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

Applicazioni: approssimazione Binomiale e Poisson con la normale, test statistici asintotici, intervalli di confidenza asintotici.

Applicazione alla media campionaria di variabili casuali di Poisson.

Siano $X_i, i = 1, 2, ..., n$, n variabili casuali di Poisson indipendenti in probabilità e con lo stesso valore atteso λ. In applicazione del teorema di normalità asintotica si hanno le seguenti relazioni molto utili nelle applicazioni:

$$P(\sum_{i=1}^{r} X_i \leq nm) = P(\bar{X}_n \leq m) =$$

$$= P(\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}}\sqrt{n} \leq \frac{m - \lambda}{\sqrt{\lambda}}\sqrt{n}) \approx$$

$$\approx \int_{-\infty}^{\frac{m-\lambda}{\sqrt{\lambda}}\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

Applicazione alla media campionaria di variabili casuali Binomiali

Siano $X_i, i = 1, 2, ..., n$, n variabili casuali Binomiali indipendenti in probabilità di parametri $n = 1$ e π. In applicazione del teorema di normalità asintotica si hanno le seguenti relazioni molto utili nelle applicazioni:

$$P(\sum_{i=1}^{r} X_i \leq nm) = P(\bar{X}_n \leq m) =$$

$$= P(\frac{\bar{X}_n - \pi}{\sqrt{\pi(1-\pi)}}\sqrt{n} \leq \frac{m-\pi}{\sqrt{\pi(1-\pi)}}\sqrt{n}) \approx$$

$$\approx \int_{-\infty}^{\frac{m-\pi}{\sqrt{\pi(1-\pi)}}\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

La definizione seguente sarà utile in Statistica quando parleremo di valori critici per i Test Statistici e di estremi degli Intervalli di Confidenza

PERCENTILI

Da ta una v.c. X discreta con funzione di ripartizione F(x) si definisce p-percentile il valore x(p) tale che:

$$x(p) = \inf \{x : F(x) = p\}$$

Data una v.c. X continua con funzione di ripartizione F(x) assolutamente continua si definisce p-percentile il valore x(p) soluzione della equazione:

$$F(x) = p$$

Casi particolari:

per p=0.5 si ha la mediana

per p=0.25 si ha il primo quartile

per p=0.75 si ha il terzo quartile