

## APPENDICE I

### Proprietà del coefficiente binomiale:

$$\binom{n}{r} = \binom{n}{n-r} = \frac{n!}{r!(n-r)!} = \frac{n^{(r)}}{r!} \quad \text{con} \begin{cases} n \in \mathbb{R} \\ r \in \mathbb{N}^+ \end{cases}$$

$$\binom{n}{r} = 0 \quad \text{se } n < r \text{ oppure } r < 0; \quad \binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}; \quad r \binom{n}{r} = n \binom{n-1}{r-1}$$

$$\text{Formula di cancellazione: } r^{(s)} \binom{n}{r} = n^{(s)} \binom{n-s}{r-s}$$

$$\text{Identità ipergeometrica: } \sum_{r=0}^{\infty} \binom{a}{r} \binom{b}{n-r} = \binom{a+b}{n}$$

$$\text{Binomio di Newton: } (a+b)^n = a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n-1} a b^{n-1} + b^n = \sum_{r=0}^n \binom{n}{r} a^{n-r} b^r$$

### Coefficienti multinomiali:

$$P(x_1 \dots x_k) = f(x_1 \dots x_k) = \binom{n}{x_1 x_2 \dots x_k} = \frac{n!}{x_1! x_2! \dots x_k!} \quad \text{con } x_1 + x_2 + \dots + x_k = n$$

### Funzione gamma:

$$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx = \left[ -e^{-x} x^{n-1} \right]_0^{\infty} + (n-1) \int_0^{\infty} x^{n-2} e^{-x} dx \Rightarrow \Gamma(n) = (n-1) \Gamma(n-1) \quad \text{con } n > 1$$

$$\Gamma(1) = 1; \quad \Gamma(n+1) = n!; \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}; \quad \Gamma\left(n + \frac{1}{2}\right) = \left(n - \frac{1}{2}\right) \left(n - \frac{3}{2}\right) \dots \left(\frac{1}{2}\right) \sqrt{\pi}$$

### Formula di Stirling:

$$n! \approx \sqrt{2\pi n} (n^n e^{-n})$$

## CAPITOLO I - Introduzione

**Distribuzione non uniforme o Geometrica:** esperimento relativo al tempo di vita di una lampadina. Supponiamo di avere inizialmente  $N$  lampadine e che, nell'intervallo di un'ora, la percentuale di lampadine che sopravvive sia pari ad  $s$  (con  $s$  costante col trascorrere delle ore), cioè:

dopo la prima ora	$sN$	lampadine sopravvivono
"	$N - sN = (1-s)N$	lampadine si guastano (hanno zero ore di vita)
dopo la seconda ora	$s \cdot sN = s^2 N$	lampadine sopravvivono
"	$sN - s^2 N = (1-s)sN$	lampadine si guastano (hanno un'ora di vita)
...dopo la $i$ -esima ora	$s^i N$	lampadine sopravvivono
"	$s^{i-1} N - s^i N = (1-s)s^{i-1} N$	lampadine si guastano (hanno $i-1$ ore di vita)

I possibili risultati dell'esperimento sono "0 ore", "1 ora", ... (di vita) e a questi risultati (eventi semplici) possiamo associare i numeri naturali. Lo spazio campionario è quindi  $S = \{0, 1, 2, \dots\}$ , e  $p_n = (1-s)s^n$  è la probabilità che una lampadina abbia  $n$  ore di vita.

## CAPITOLO II - Calcolo combinatorio

**Disposizioni semplici:** i gruppi che si possono formare con  $n$  oggetti dati in modo che ognuno contenga solo  $k$  oggetti distinti e che due gruppi differiscano tra loro o per qualche elemento o per l'ordine:

$$D_{n,k} = n(n-1)(n-2)\dots(n-k+2)(n-k+1) \equiv n^{(k)} \quad \text{con } k \leq n$$

**Permutazioni:** disposizioni semplici di  $n$  oggetti di classe  $n$  (differiscono solo per l'ordine):

$$P_n = D_{n,n} = n^{(n)} = n(n-1)(n-2)\dots 2 \cdot 1 = n!$$

**Combinazioni:** i gruppi di  $k$  oggetti che si possono formare con  $n$  oggetti, in modo che i gruppi

differiscano almeno per un oggetto; nelle disposizioni semplici i gruppi possono differire anche per l'ordine, nelle combinazioni differiscono tra loro solo quando differiscono per almeno un elemento:

$$C_{n,k} = \frac{D_{n,k}}{k!} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} = \frac{n^{(k)}}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k} = \binom{n}{k}$$

**Disposizioni con ripetizione:** i gruppi che si possono formare con  $n$  oggetti dati in modo che ognuno contenga solo  $k$  oggetti, con la possibilità di ripetere ciascuno di essi una o più volte all'interno di una stessa sequenza, e che due gruppi differiscano tra loro o per qualche elemento o per l'ordine:

$$D'_{n,k} = n^k$$

### CAPITOLO III – Risultati equiprobabili

**Distribuzione Ipergeometrica:** se una certa caratteristica divide una intera popolazione in due gruppi distinti  $a$  e  $b$  e un campione di dimensione  $n$  viene estratto a caso (random) e senza reimmissione, la probabilità che il campione contenga  $x$  membri di tipo 1 (e quindi  $n-x$  membri del secondo tipo), è:

	Tipo 1	Tipo 2	Totale
Popolazione	$a$	$b$	$a+b$
Campione	$x$	$n-x$	$n$

•Sequenze ordinate:

$$\left. \begin{aligned} N_S &= (a+b)^{(n)} \\ N_A &= \binom{n}{x} a^{(x)} b^{(n-x)} \end{aligned} \right\} \Leftrightarrow f(x) = P(A) = \frac{N_A}{N_S} = \frac{\binom{n}{x} a^{(x)} b^{(n-x)}}{(a+b)^{(n)}}$$

•Sequenze non ordinate:

$$\left. \begin{aligned} N_{S'} &= \binom{a+b}{n} \\ N_{A'} &= \binom{a}{x} \binom{b}{n-x} \end{aligned} \right\} \Leftrightarrow f(x) = P(A') = \frac{N_{A'}}{N_{S'}} = \frac{C_{a,x} C_{b,n-x}}{C_{a+b,n}} = \frac{\binom{a}{x} \binom{b}{n-x}}{\binom{a+b}{n}} \quad \text{con } x \in \mathbb{N}$$

Relazione di ricorrenza:  $f(x) = r(x) f(x-1) r(x) = \frac{a-x+1}{x} \cdot \frac{n-x+1}{b-n+x}$

**Campioni con tutti gli elementi distinti nel campionamento random con reimmissione:** probabilità che tutti gli elementi siano distinti:

$$P(A) = \frac{N_A}{N_S} = \frac{n^{(r)}}{n^r} = \frac{n!}{(n-r)! n^r} \equiv q_r$$

$$\text{per } n \rightarrow \infty \Rightarrow \frac{n!}{(n-r)! n^r} \cong \frac{\sqrt{2\pi n} n^n e^{-n}}{\sqrt{2\pi(n-r)} (n-r)^{n-r} e^{-n+r} n^r} = \underbrace{\sqrt{\frac{n}{n-r}}}_{\rightarrow 1} \underbrace{\left(\frac{n}{n-r}\right)^{n-r}}_{\substack{= \left(1+\frac{r}{n-r}\right)^{n-r} \\ \rightarrow e^r}} e^{-r} \Leftrightarrow q_r \rightarrow 1$$

**Distribuzione Binomiale:** come nell'ipergeometrica ma con reimmissione:

$$\left. \begin{aligned} p &= \frac{a}{a+b} \\ 1-p &= \frac{b}{a+b} \end{aligned} \right\} \Leftrightarrow f(x) = \binom{n}{x} p^x (1-p)^{n-x} \equiv b(n, p)$$

Relazione di ricorrenza:  $f(x) = r(x) f(x-1) \quad \text{con } r(x) = \frac{n-x+1}{x} \cdot \frac{a}{b}$

**Approssimazione dell'Ipergeometrica alla Binomiale:**

$$\binom{a}{x} \binom{b}{n-x} / \binom{a+b}{n} \approx \binom{n}{x} p^x (1-p)^{n-x} \quad \text{con} \begin{cases} n \ll (a+b) \\ x \in \mathbb{N} \\ p = \frac{a}{a+b} \end{cases}$$

### CAPITOLO IV - Spazi di probabilità e Calcolo delle probabilità

**Algebra booleana:** è l'insieme della parti  $\beta(S) \subseteq \wp(S)$  se:

$$\forall A \in \beta(S) \Rightarrow \bar{A} \in \beta(S); \quad \forall A, B \in \beta(S) \Rightarrow A \cup B \in \beta(S)$$

cioè è chiuso rispetto a  $\cup, \cap, ^c$ .

**Misura:** un qualsiasi valore reale non negativo associato ad ogni insieme con:

$$A \cap B = \emptyset \Rightarrow m(A \cup B) = m(A) + m(B) \quad B \subseteq A \Rightarrow m(B) \leq m(A)$$

**Spazi di probabilità:** si passa alla  $\sigma$ -algebra passando all'unione di infiniti sottoinsiemi. E' formato da:

- 1)  $S$  = insieme o spazio campionario (i possibili risultati);
- 2)  $\wp(S)$  = insieme delle parti di  $S$ ;
- 3)  $p$  = probabilità;

**Probabilità:** misura di massa totale pari a 1.

**Eventi indipendenti:**  $P(AB) = P(A) \cdot P(B)$  con  $A, B$  eventi (vedi "Variabili random independenti")

**Eventi disgiunti o mutuamente esclusivi:**  $P(S_n) = P(a_1) + P(a_2) + \dots + P(a_n)$  con  $S_n = a_1 + a_2 + \dots + a_n$

**Prove di Bernoulli:** ripetizioni indipendenti di un esperimento che dà luogo a due soli possibili risultati (con reimmissione)  $s$  (success) ed  $f$  (failure). La probabilità di ottenere esattamente  $x$  successi in  $n$  prove di Bernoulli è data dalla binomiale  $b(n, p)$ .

**Probabilità condizionata:** probabilità di  $A$  dato (che si è verificato)  $B$ :

$$P(A|B) = \frac{P(AB)}{P(B)} \quad \text{con } P(B) \neq 0$$

Regola di Bayes:  $P(AB) = P(A|B)P(B) = P(B|A)P(A)$  essendo  $AB = BA$

Ogni evento  $B$  può essere scritto come unione di  $k$  eventi mutuamente esclusivi con:

$$S = A_1 \cup A_2 \cup \dots \cup A_k; \quad A_i \cap A_j = \emptyset; \quad i \neq j;$$

$$B = (BA_1) \cup (BA_2) \cup \dots \cup (BA_k) \Leftrightarrow P(B) = \sum_{i=1}^k P(B|A_i)P(A_i)$$

$$\text{Teorema di Bayes: } P(A_i|B) = \frac{P(BA_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^k P(B|A_j)P(A_j)}$$

**Binary symmetric channel:** esperimenti dipendenti, poiché l'output dipende dall'input:

$$\text{Distorsione del segnale} \begin{cases} P(1_o | 0_i) = P(0_o | 1_i) = \varepsilon \\ P(1_o | 1_i) = P(0_o | 0_i) = 1 - \varepsilon \end{cases}$$

$$P(1_o) = \sum_{j \in I} P(1_o, j) = \sum_{j \in I} P(1_o | j)P(j) = P(1_o | 0_i)P(0_i) + P(1_o | 1_i)P(1_i) = \varepsilon(1-p) + (1-\varepsilon)p = \varepsilon + p - 2\varepsilon p$$

$$P(0_o) = \sum_{j \in I} P(0_o, j) = \sum_{j \in I} P(0_o | j)P(j) = P(0_o | 0_i)P(0_i) + P(0_o | 1_i)P(1_i) = (1-\varepsilon)(1-p) + \varepsilon p = 1 - \varepsilon - p + 2\varepsilon p$$

$$P(1_i | 1_o) = \frac{P(1_i, 1_o)}{P(1_o)} = \frac{P(1_o | 1_i)P(1_i)}{P(1_o)} = \frac{(1-\varepsilon)p}{\varepsilon + p - 2\varepsilon p}$$

## CAPITOLO V – Variabile aleatoria discreta

**Variabile aleatoria:** funzione.

**Funzione di probabilità:** gode delle proprietà delle probabilità:

$$f(x) = P(X = x) = \sum_{i \in (X=x)} p_i \quad \forall x \in R_x$$

**Funzione indicatrice di A:** è la v.a.  $X_A$  che assume valore 1 se l'evento  $A$  si avvera e 0 in caso contrario:

$$f(1) = P(X_A = 1) = P(A); \quad f(0) = P(X_A = 0) = P(\bar{A})$$

**Funzione Cumulativa o di ripartizione:** distribuzione di probabilità dell'evento  $x \leq t$ :

$$F(t) = P(X \leq t) = \sum_{x \leq t} f(x) \quad \text{con} \begin{cases} -\infty \leq t \leq \infty \\ F(-\infty) = 0 \\ F(+\infty) = 1 \end{cases}$$

$$f(x_i) = P(X = x_i) = P\{(X \leq x_i) - (X \leq x_{i-1})\} = P(X \leq x_i) - P(X \leq x_{i-1}) = F(x_i) - F(x_{i-1})$$

**Funzioni composte:** una funzione a valori reali  $y=h(X)$  di una v.a.  $X$  è una v.a. con distribuzione:

$$g(y) = \sum_{x:h(x)=y} f(x)$$

**Tempi di attesa nelle prove di Bernoulli:**

$$X_n \equiv \text{n}^\circ \text{ di successi nelle prime } n \text{ prove} \Rightarrow P(X_n = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{con } x = 0, 1, \dots, n$$

$$N_x \equiv \text{n}^\circ \text{ di prove per } x \text{ successi} \Rightarrow P(N_x = n) = \binom{n-1}{x-1} p^x (1-p)^{n-x} \quad \text{con } n = x, x+1, \dots$$

$$P(N_x = n) = \frac{x}{n} P(X_n = x); \quad P(N_x > n) = P(X_n < x)$$

**Mancanza di memoria nella legge geometrica:** se  $X$  è una v.a. che segue una legge geometrica di parametro  $p$  vale la relazione:

$$P(X \geq k) = \sum_{n=k}^{\infty} p(1-p)^n = (1-p)^k \sum_{n=k}^{\infty} p(1-p)^{n-k} = (1-p)^k \sum_{l=0}^{\infty} p(1-p)^l = (1-p)^k$$

$$P(X = k+m | X \geq k) = \frac{P(X = k+m, X \geq k)}{P(X \geq k)} = \frac{P(X = k+m)}{P(X \geq k)} = \frac{p(1-p)^{k+m}}{(1-p)^k} = p(1-p)^m = P(X = m)$$

$$P(N_1 = k+m | N_1 > k) = P(N_1 - 1 = k+m-1 | N_1 - 1 > k-1) = P(N_1 - 1 = k+m-1 | N_1 - 1 \geq k) = P(N_1 - 1 = m-1) = P(N_1 = m)$$

**Tempi di attesa nel semplice campionamento random:** estrazioni non indipendenti, senza reimmissione:

$$P(X_n = x) = \binom{n}{x} a^{(x)} b^{(n-x)} / (a+b)^{(n)} = \binom{a}{x} \binom{b}{n-x} / \binom{a+b}{n}$$

$$\text{Come prove Bernoulli: } P(N_x = n) = \frac{x}{n} P(X_n = x) = \binom{n-1}{x-1} a^{(x)} b^{(n-x)} / (a+b)^{(n)}; \quad P(N_x > n) = P(X_n < x)$$

**Distribuzione di Poisson:** o legge degli eventi rari:

$$f(x) = \frac{\mu^x}{x!} e^{-\mu} \quad \text{con} \begin{cases} x \in \mathbb{N} \\ \mu \in \mathbb{R}^+ = \text{media della distribuzione} \end{cases}$$

$$\text{Formula di ricorrenza: } \frac{f(x)}{f(x-1)} = \frac{\mu}{x}$$

**Approssimazione di Poisson alla Binomiale:** di parametro  $\mu=np$ :

$$n \gg p \Rightarrow n \gg np \Leftrightarrow \binom{n}{x} p^x (1-p)^{n-x} \approx \frac{\mu^x}{x!} e^{-\mu}$$

**Distribuzioni bivariate – Congiunta, Marginale e Condizionale:**

Funzione probabilità **congiunta** di  $X$  e  $Y$ :  $f(x, y) = P(X = x, Y = y)$  con  $x \in R_x, y \in R_y$

$$(X = x) = (X = x) \cap S = (X = x) \cap \left( \bigcup_i (Y = y_i) \right) = \bigcup_i (X = x) \cap (Y = y_i) \Leftrightarrow$$

$$\Leftrightarrow f_1(x) = P(X = x) = P\left\{ \bigcup_i (X = x) \cap (Y = y_i) \right\} = \sum_i P(X = x, Y = y_i) = \sum_{y \in R_y} f(x, y)$$

Funzione probabilità **marginale** di  $X$  e  $Y$ :  $f_1(x) = \sum_{y \in R_y} f(x, y); \quad f_2(y) = \sum_{x \in R_x} f(x, y)$

$$P(Y = y | X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{f(x, y)}{f_1(x)} \Leftrightarrow f_2(y | x) = P(Y = y | X = x) = \frac{f(x, y)}{f_1(x)}$$

Funzione probabilità **condizionale** (di  $Y$  condizionato ad  $X=x$ ) è:

$$f_2(y | x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{f(x, y)}{f_1(x)} \Leftrightarrow f(x, y) = f_2(y | x) f_1(x) \quad \text{con} \begin{cases} y \in R_y \\ f_1(x) \neq 0 \end{cases}$$

**N.B.:**  $\sum_{y \in R_y} f_2(y | x) = \frac{1}{f_1(x)} \sum_{y \in R_y} f(x, y) = \frac{f_1(x)}{f_1(x)} = 1$

**Variabili random indipendenti:** per definizione gli eventi  $(X=x)$  e  $(Y=y)$  sono indipendenti se e solo se:

$$P(Y = y, X = x) = P(X = x) \cdot P(Y = y) \Leftrightarrow \begin{cases} f(x, y) = f_1(x) \cdot f_2(y) \\ f_2(y | x) = f_2(y) \end{cases}$$

Due v.a.  $X$  e  $Y$  sono indipendenti se l'equazione di prima è vera  $\forall x \in R_x, \forall y \in R_y$

**Legge della somma di due variabili o Convoluzione:** date due v.a.  $X$  e  $Y$  con distribuzioni  $f_1(x)$  ed  $f_2(y)$ , si vuole trovare la legge  $Z = X + Y$ . La probabilità dell'evento si può scrivere come:

$$g(z) = P(Z = z) = P(X + Y = z) = P\left\{ \bigcup_i (X = i) \cap (Y = z - i) \right\} = \sum_i P(X = i, Y = z - i) = \sum_i f_1(i) f_2(z - i)$$

Se le v.a. sono indipendenti si ha:  $g(z) = \sum_i P(X = i) P(Y = z - i) = \sum_i f_1(i) f_2(z - i)$

**Variabili indipendenti e identicamente distribuite IID:** sono variabili random indipendenti con la stessa distribuzione di probabilità marginale; la distribuzione di probabilità congiunta è:

$$f(y_1, y_2, \dots, y_n) = \prod_{i=1}^n f_i(y_i)$$

**Somma di distribuzioni poissoniane:** date due v.a. indipendenti  $X$  e  $Y$  con distribuzione poissoniana di media  $\mu$  e  $\nu$  rispettivamente, si ha che la v.a.  $Z = X + Y$  è data da una poissoniana di media  $\mu + \nu$ :

$$\left. \begin{aligned} f_1(x) &= \mu^x e^{-\mu} / x! \\ f_2(y) &= \nu^y e^{-\nu} / y! \end{aligned} \right\} \Leftrightarrow g(z) = P(Z = z) = P(X + Y = z) = \sum_i f_1(i) f_2(z - i) = \sum_{i=0}^z \frac{\mu^i \nu^{z-i}}{i!(z-i)!} e^{-\mu-\nu} \Leftrightarrow$$

$$\Leftrightarrow \frac{1}{i!(z-i)!} = \frac{1}{z!} \frac{z!}{i!(z-i)!} = \frac{1}{z!} \binom{z}{i} \Leftrightarrow g(z) = \frac{e^{-\mu-\nu} \nu^z}{z!} \sum_{i=0}^z \binom{z}{i} \left( \frac{\mu}{\nu} \right)^i = \frac{e^{-\mu-\nu} \nu^z}{z!} \left( 1 + \frac{\mu}{\nu} \right)^z = \frac{(\mu + \nu)^z}{z!} e^{-(\mu+\nu)}$$

**Distribuzione Multinomiale:** caso generale della distribuzione binomiale:

Evento	$A_1$	$A_2$	...	$A_k$	tot
Probabilità	$p_1$	$p_2$	...	$p_k$	1
Frequenza	$X_1$	$X_2$	...	$X_k$	$n$

$$f(x_1, x_2, \dots, x_k) = \binom{n}{x_1 x_2 \dots x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} \quad \text{con il fattore multinomiale} \quad \binom{n}{x_1 x_2 \dots x_k} = \frac{n!}{x_1! x_2! \dots x_k!}$$

### CAPITOLO VI – Media, Varianza e Covarianza

**Aspettazione matematica o Media:** sia  $X$  una v.a. discreta con range  $R_x = \{x_1, x_2, \dots\}$  e funzione probabilità  $f$ . L'aspettazione matematica di  $X$  è il numero reale definito da:

$$E(X) = \sum_{x \in R_x} x f(x) = x_1 f(x_1) + x_2 f(x_2) + \dots \quad (\text{ammesso che questa serie converga assolutamente})$$

Quando le probabilità vengono interpretate in termini di frequenze relative,  $E(X)$  rappresenta la media aritmetica dei risultati ottenuti in una lunga serie di esperimenti ripetuti e da il centro della distribuzione.

**Aspettazione o Media di una funzione di  $X$ :** sia  $U \equiv h(X)$  una funzione della v.a.  $X$ , dove  $h$  è una funzione a valori reali definita su  $R_x$ . Sappiamo che anche  $U$  è una v.a.. Indichiamo con  $R_u$  l'insieme dei suoi possibili valori e con  $g$  la sua funzione probabilità:

$$E(U) = E\{h(X)\} = \sum_{x \in R_x} h(x) f(x)$$

$$E(X) = \sum \sum x f(x, y) = \sum_{x \in R_x} x \sum_{y \in R_y} f(x, y) = \sum_{x \in R_x} x f_1(x)$$

Il vantaggio di questo risultato è che per calcolare il valore atteso di  $U$  non è necessario ricavare la funzione probabilità  $g$ . Tale risultato si estende al caso di più variabili random: se  $U \equiv h(X, Y)$ , dove  $X$  e  $Y$  sono v.a. con funzione probabilità congiunta  $f$ , allora:

$$E(U) = E\{h(X, Y)\} = \sum_{x \in R_x} \sum_{y \in R_y} h(x, y) f(x, y)$$

**Aspettazione di funzioni lineari:**

$$E(aX + b) = aE(X) + b; \quad E(aX + bY) = aE(X) + bE(Y) = a\mu_x + b\mu_y$$

**Momento  $r$ -esimo:** sia  $X$  una v.a. discreta con range  $\bar{X}$  e distribuzione di probabilità  $n$ , e sia  $\bar{X}$  un intero non negativo. Il momento  $r$ -esimo di  $X$  (o della distribuzione  $f$ ) è definito come il valore medio della potenza  $r$ -esima di  $X$  e sono dei numeri reali che descrivono la distribuzione di probabilità:

$$m_r = E(X^r) = \sum_{x \in R_x} x^r f(x) \quad (\text{ammesso che la serie converga assolutamente})$$

**Media:** il primo momento è la media di  $X$  e specifica la posizione del centro della distribuzione:

$$m_1 = E(X)$$

$$E(X) = \mu \Rightarrow E(X - \mu) = E(X) - \mu = 0$$

**Varianza:** è il secondo momento di  $X - \mu$ , e da la larghezza della distribuzione:

$$\text{var}(X) = \sigma^2 = E\{(X - \mu)^2\} = \sum_{x \in R_x} (x - \mu)^2 f(x) \quad \text{con } \text{var}(X) \geq 0$$

$$\text{var}(X) = 0 \Rightarrow P(X = \mu) = 1$$

Un'espressione equivalente della varianza di  $X$  è data da:  $\sigma^2 = E(X^2) - E(X)^2 = E(X^2) - \mu^2$

**Deviazione standard:** la radice quadrata positiva della varianza è la deviazione standard di  $X$  che si indica con  $\sigma$  ed è una misura dell'incertezza sul valore che la v.a. assume in relazione ad un dato esperimento; essa descrive la larghezza o la dispersione della distribuzione di probabilità.

Con la disuguaglianza di Chebyshev: gli 8/9 della probabilità sono contenuti in:  $(\mu - 3\sigma, \mu + 3\sigma)$

**Varianza di una funzione lineare:**

$$\text{var}(aX + b) = a^2 \text{var}(X); \quad \sigma_{aX+b} = |a| \sigma_x$$

**Forma standard:** v.a. con media 0 e varianza 1:

$$X^* \equiv \frac{X - \mu}{\sigma}$$

$$E(X^*) = \frac{1}{\sigma} E(X - \mu) = \frac{1}{\sigma} [E(X) - \mu] = 0; \quad \text{var}(X^*) = \text{var}\left(\frac{1}{\sigma} X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2} \text{var}(X) = 1$$

**Disuguaglianza di Chebyshev:** sia  $X$  una v.a. con deviazione standard  $\sigma$  e media  $\mu$  finite:

$$\forall t > 0 \Rightarrow P\{|X - \mu| \geq t\sigma\} \leq \frac{1}{t^2}; \quad t = 3 \Rightarrow P\{|X - \mu| \geq 3\sigma\} \leq \frac{1}{9}; \quad (\mu - 3\sigma, \mu + 3\sigma)$$

**Momento fattoriale  $r$ -esimo:** momento fattoriale  $r$ -esimo  $g_r$  è definito con:

$$g_r = E\{X^{(r)}\} = E\{X(X-1)\dots(X-r+1)\} = \sum x^{(r)} f(x) \quad (\text{se la serie converga assolutamente})$$

$$g_1 = E\{X\} = m_1 = \mu; \quad g_2 = E\{X(X-1)\} = E(X^2) - E(X) = m_2 - \mu$$

$$\sigma^2 = m_2 - \mu^2 = g_2 + \mu - \mu^2$$

**Distribuzione di Poisson:**

$$g_r = \sum_{x=r}^{\infty} \frac{\mu^x e^{-\mu}}{(x-r)!} = \mu^r \sum_{x=r}^{\infty} \frac{\mu^{x-r} e^{-\mu}}{(x-r)!} = \mu^r \sum_{y=0}^{\infty} \frac{\mu^y e^{-\mu}}{y!} = \mu^r$$

$$E\{X\} = g_1 = \mu$$

$$\text{var}(X) = g_2 + \mu - \mu^2 = \mu$$

**Distribuzione Binomiale:**

$$g_r = \sum n^{(r)} \binom{n-r}{x-r} p^x (1-p)^{n-x} = n^{(r)} p^r \sum \binom{n-r}{x-r} p^{x-r} (1-p)^{n-x} \Leftrightarrow y = x - r \Rightarrow$$

$$\Rightarrow n^{(r)} p^r \underbrace{\sum \binom{n-r}{y} p^y (1-p)^{n-r-y}}_{\text{binomio di Newton } b(n-r, p)} \Leftrightarrow g_r = n^{(r)} p^r$$

$$E\{X\} = g_1 = np$$

$$\text{var}(X) = g_2 + \mu - \mu^2 = n(n-1)p^2 + np - n^2 p^2 = np(1-p)$$

**Distribuzione dei tempi di attesa nelle prove di Bernoulli:** il tempo di attesa medio per il primo successo cresce al diminuire della probabilità  $p$  di ottenere un successo nella singola prova:

$$E(N_1) = \sum_n n g_1(n) = \sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{p}{1-p} \sum_{n=0}^{\infty} n(1-p)^n = \frac{p}{1-p} \cdot \frac{1-p}{[1-(1-p)]^2} = \frac{1}{p}$$

$$E(N_1^2) = \sum_n n^2 g_1(n) = \frac{2-p}{p^2}$$

$$\text{var}(N_1) = E(N_1^2) - E(N_1)^2 = \frac{1-p}{p^2}$$

**Distribuzione Ipergeometrica:**

$$g_r = E\{X^{(r)}\} = \sum_{x=0}^{\infty} x^{(r)} \binom{a}{x} \binom{b}{n-x} / \binom{a+b}{n} \Leftrightarrow x^{(r)} \binom{a}{x} = a^{(r)} \binom{a-r}{x-r} \Leftrightarrow g_r = \sum_{x=0}^{\infty} a^{(r)} \binom{a-r}{x-r} \binom{b}{n-x} / \binom{a+b}{n}$$

$$\Leftrightarrow g_r = a^{(r)} \binom{a+b-r}{n-r} / \binom{a+b}{n} = a^{(r)} \frac{(a+b-r)!}{(n-r)!(a+b-n)!} \cdot \frac{n!(a+b-n)!}{(a+b)!} = \frac{a^{(r)} n^{(r)}}{(a+b)^{(r)}}$$

$$g_1 = \frac{an}{a+b}; \quad g_2 = \frac{a(a-1)n(n-1)}{(a+b)(a+b-1)}$$

$$\mu = np$$

$$\sigma^2 = np(1-p) \frac{a+b-n}{a+b-1}; \quad \left( p = \frac{a}{a+b} \right)$$

### Processo di Poisson:

- 1) Indipendenza = il numero di eventi in intervalli temporali che non sovrappongono sono indipendenti;
- 2) Individualità = gli eventi avvengono singolarmente e non a coppie o a gruppi;
- 3) Omogeneità = nell'intervallo di tempo totale considerato gli eventi si succedono ad un tasso uniforme;

Viene anche usato per esperimenti in cui gli eventi avvengono in modo casuale e uniforme nello spazio.

**Varianza di una somma e Covarianza:** indica la presenza di una certa interdipendenza, o correlazione, tra le due variabili:

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y); \quad \text{var}(aX+bY) = a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab \text{cov}(X,Y)$$

$$\text{Covarianza di } X \text{ e } Y: \text{cov}(X,Y) = E\{(X-\mu_x)(Y-\mu_y)\} = E(XY) - E(X)E(Y)$$

$$\text{cov}(X,Y) \begin{cases} > 0 \Leftrightarrow \text{le due variabili sono correlate} \\ = 0 \Leftrightarrow \text{le due variabili non sono correlate} \\ < 0 \Leftrightarrow \text{le due variabili sono anticorrelate} \end{cases}$$

**Coefficiente di Correlazione:** il coefficiente di correlazione  $\rho$  di due variabili random  $X$  e  $Y$  si definisce:

$$\rho(X,Y) = \text{cov}(X^*,Y^*) = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} \quad \text{con } -1 \leq \rho \leq 1$$

$$\rho(aX+b, cY+d) = \frac{ac \cdot \text{cov}(X,Y)}{|ac| \cdot \sigma_x \sigma_y} = \begin{cases} ac > 0 \Rightarrow \rho(X,Y) \\ ac < 0 \Rightarrow -\rho(X,Y) \end{cases} \quad \text{da questa relazione si ha che se la covarianza}$$

dipende dalla scala di misura usata, ciò non accade per il coefficiente di correlazione.

$$\left. \begin{aligned} \text{var}(X^* \pm Y^*) &= \text{var}(X^*) + \text{var}(Y^*) \pm 2\text{cov}(X^*, Y^*) \\ \text{var}(X^*) &= \text{var}(Y^*) = 1 \end{aligned} \right\} \Leftrightarrow \text{var}(X^* \pm Y^*) = 2(1 \pm \rho) \geq 0$$

$\rho = \pm 1 \Rightarrow$  una v.a. può essere espressa come una funzione lineare dell'altra:

$$X \equiv aY + b \quad \text{con} \begin{cases} \rho = 1 \Rightarrow a = \frac{\mu_x}{\sigma_y}; & \rho = -1 \Rightarrow a = -\frac{\mu_x}{\sigma_y} \\ b = \mu_x - a\mu_y \end{cases}$$

**Variabili random non correlate:** due variabili random t.c.  $\text{cov}(X,Y) = 0$  sono non correlate:

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$$

**N.B.:** indipendenti  $\Rightarrow$  non correlate  $\sim \text{cov}=0$  (non il viceversa; l'indipendenza è più "forte" della non correlazione)

**N.B.:** correlate  $\sim \text{cov} \neq 0 \Rightarrow$  dipendenti

### Media di una somma:

$$E(S_n) = \sum E(X_i) = \sum \mu_i; \quad E(X+Y) = E(X) + E(Y) \text{ sempre}$$

### Varianza di una somma:

$$\text{var}(S_n) = \sum \sigma_i^2 + 2 \sum_{i < j} \text{cov}(X_i, X_j)$$

Se tutte le covarianze sono nulle le variabili sono non correlate:

$$\text{cov}(X_i, X_j) = 0 \Rightarrow \text{var}(S_n) = \sum \sigma_i^2$$

**Varianza di una combinazione lineare:**

$$E(\sum a_i X_i) = \sum a_i E(X_i) \Rightarrow \text{var}(\sum a_i X_i) = \sum a_i^2 \text{var}(X_i) + 2 \sum_{i < j} a_i a_j \text{cov}(X_i, X_j)$$

$$\text{v.a. indipendenti} \Rightarrow \text{cov}(X_i, X_j) = 0 \Leftrightarrow \text{var}(\sum a_i X_i) = \sum a_i^2 \text{var}(X_i)$$

Proprietà notevole della varianza di  $\bar{X}$ : al crescere di  $n$  la varianza decresce e la distribuzione di  $\bar{X}$  diventa sempre più concentrata nell'intorno di  $\mu$ .

**Sample mean:** è la media aritmetica delle  $X_i$ , dati  $X_1, X_2, \dots, X_n$  variabili random indipendenti con la stessa media  $\mu$  e stessa varianza  $\sigma^2$ , cioè:

$$\bar{X} \equiv \frac{1}{n} \sum X_i$$

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

$$\text{var}(\bar{X}) = \sum \left(\frac{1}{n}\right)^2 \text{var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$\text{intervallo che contiene la quasi totalità della probabilità} \Rightarrow E(\bar{X}) \pm 3\sqrt{\text{var}(\bar{X})} = \mu \pm 3\sigma/\sqrt{n}$$

quindi la sua larghezza tende a zero quando  $n \rightarrow \infty$ . Con alta probabilità,  $\bar{X}$  sarà vicino a  $\mu$  per  $n$  grande. Questo risultato giustifica l'uso del valore osservato di  $\bar{X}$  come stima di  $\mu$ .

### CAPITOLO VII – Informazione ed Entropia

**Informazione:** come misura della sorpresa. Si considera una  $I$  che sia una funzione decrescente della probabilità  $P(E)$  dell'evento  $E$ , cioè richiediamo che, se  $E$  ed  $F$  sono eventi di  $S$  con  $P(E) \leq P(F) \Rightarrow I(E) \geq I(F)$ :

$$I(E) = -k \log_a [P(E)] \geq 0 \quad \text{con} \begin{cases} a > 1 \\ k > 0 \end{cases}; \quad \text{solitamente} \begin{cases} a = 2 \\ k = 1 \end{cases} \Leftrightarrow I(E) = -\log_2 [P(E)]$$

$$P(E) \rightarrow 0 \Leftrightarrow I(E) \rightarrow \infty$$

$$\text{Se eventi indipendenti: } I(E_1 E_2) = I(E_1) + I(E_2)$$

**Bit:** unità del contenuto di informazione; si guadagna un bit d'informazione quando si sceglie tra due alternative ugualmente probabili.

**Entropia:** il contenuto d'informazione di  $X$  si può considerare come una v.a., funzione di  $X$ , che si indica con  $I(X)$ . La media dell'informazione  $I(X)$  è detta entropia di  $X$  e che si indica con  $H(X)$ :

$$H(X) \equiv E\{I(X)\} = - \sum_{j=1}^n \overbrace{p_j}^{f(x)} \log(p_j) \quad p_j = 0 \Leftrightarrow H(X) = 0$$

Si può interpretare come l'informazione media che prevediamo di acquisire dal sistema, ed è massima quando massima è l'incertezza.

Per calcolare l'entropia di una variabile random  $X$  di Bernoulli di parametro  $p$  si ha che il range di  $X$  è  $R_x = \{x_1, x_2\}$  con probabilità  $p_1 = p$  e  $p_2 = 1 - p$ , mentre il range di  $I(X)$  è  $R_I = \{I(p), I(1-p)\}$ :

$$H(X) \equiv H_b(p) = E\{I(X)\} = - \sum_{j=1}^2 p_j \log(p_j) = -p \log(p) - (1-p) \log(1-p)$$

**Lemma 7.1**  $\ln(x) \leq x - 1$  e l'uguaglianza vale  $\Leftrightarrow x = 1$

**Teorema 7.1** Sia  $X$  una v.a. discreta, e sia  $n$  il numero dei suoi possibili valori, allora:

A.  $H(X) \geq 0$  e  $H(X) = 0$  se e solo se  $X$  assume uno dei suoi possibili valori con certezza;

B.  $H(X) \leq \log(n)$  e l'uguaglianza vale se e solo se  $X$  è uniformemente distribuita;

La parte (a) ci dice che l'entropia di  $X$  è nulla proprio quando siamo in presenza dell'evento certo. La parte (b) mostra che l'entropia ha un massimo quando si ha la massima incertezza sul fenomeno aleatorio, cioè quando tutte le opzioni sono ugualmente probabili. Denotando con  $H_n$  l'entropia di una v.a. che può assumere  $n$  valori, con distribuzione di probabilità uniforme:

$$H_n = \log(n); \quad m \leq n \Rightarrow \log(m) \leq \log(n) \Leftrightarrow H_m \leq H_n$$

ossia si ha maggiore incertezza (sul risultato) quando il numero dei possibili risultati (equiprobabili) è relativamente grande e si ha una maggiore entropia, cioè una aspettativa di maggiore informazione.

## CAPITOLO VIII - Variabili aleatorie continue

**Definizioni e notazione:** data la v.a.  $X$  a valori reali, la funzione di distribuzione cumulativa o integrale di probabilità  $F(t)$  è definita come la probabilità dell'evento  $(X \leq t)$ :

$$F(t) = P(X \leq t) \quad \text{con } -\infty \leq t \leq \infty; \quad F(-\infty) = 0; \quad F(\infty) = 1$$

**Variabile random continua:** è una variabile random a valori reali con funzione di ripartizione  $F(x)$  continua, la cui derivata esiste ed è continua eccetto, eventualmente, in un numero finito di punti:

$$f(x) = \frac{d}{dx} F(x) \geq 0 \quad \forall x$$

La funzione  $f$  è detta **funzione densità di probabilità** di  $X$ . Da notare che  $f$  può risultare non definita in un numero finito di punti.

Condizione di normalizzazione nel caso continuo:  $F(x) = \int_{-\infty}^x f(t) dt = 1$

**N.B.:** è la quantità  $f(t)dt$  che ha significato di probabilità.

$$P(a < X \leq b) = P\{(X \leq b) - (X \leq a)\} = P(X \leq b) - P(X \leq a) = F(b) - F(a) \Leftrightarrow P(a < X \leq b) = \int_a^b f(x) dx$$

**N.B.:**  $f(b)$  non dà la probabilità che  $X=b$ ; la probabilità che una variabile continua assuma un particolare valore reale è zero.

Se  $(a, b)$  è piccolo e se  $f(x)$  esiste ed è continua per  $a < x < b$  allora:  $P(a < X \leq b) = \int_a^b f(x) dx \approx (b - a) f(t)$

**Aspettazione matematica o Media:** di una v.a.  $X$  è definito da:

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

La media di una funzione è:  $E\{h(X)\} = \int_{-\infty}^{\infty} h(x) f(x) dx$ .

**Mediana  $m$  di  $X$  (o della distribuzione):** è il valore della variabile random per cui:

$$F(m) = P(X < m) = \frac{1}{2} \Leftrightarrow P(X > m) = \frac{1}{2}$$

La mediana indica il “centro” della distribuzione continua, e sarà minore della media per quelle distribuzioni che hanno una lunga coda a destra. Per una distribuzione con una lunga coda a sinistra si avrà  $m > \mu$ . Se una distribuzione ha un punto di simmetria centrale, tale punto sarà la media e la mediana.

**Quantile:** quantità che generalizza il concetto di mediana. Data la v.a.  $X$  continua con funzione cumulativa  $F(x)$ , il quantile di ordine  $\alpha$  di  $X$  (o della distribuzione), con  $0 < \alpha < 1$ , è definito come il valore  $q_\alpha$  di  $X$  tale che:

$$P(X \leq q_\alpha) = F(q_\alpha) = \alpha$$

**Distribuzione Uniforme o Rettangolare:**

$$X \sim U(a, b) \Leftrightarrow f(x) = \begin{cases} a < x < b \Leftrightarrow k \\ \text{altrove} \Leftrightarrow 0 \end{cases}$$

Poiché l'area sotto la curva della densità di probabilità deve essere 1, segue che:

$$k = \frac{1}{(b-a)} \Leftrightarrow \begin{cases} x \leq a \Leftrightarrow F(x) = P(X \leq x) = 0 \\ a < x < b \Leftrightarrow F(x) = \int_{-\infty}^x f(t) dt = \int_a^x k \cdot dt = \frac{x-a}{b-a} \\ x \geq b \Leftrightarrow F(x) = 1 \end{cases}$$

**Distribuzione Esponenziale:** risulta appropriata come distribuzione dei tempi di vita nei casi in cui non

c'è invecchiamento con l'età:

$$X \sim \Gamma(1, \lambda) \Leftrightarrow f(x) = \begin{cases} x > 0 \Leftrightarrow \lambda e^{-\lambda x} \\ x \leq 0 \Leftrightarrow 0 \end{cases} \quad \text{con } \lambda \in \mathbb{R}^+$$

$$F(x) = \begin{cases} x > 0 \Leftrightarrow \int_0^x \frac{1}{\theta} e^{-t/\theta} dt = [-e^{-t/\theta}]_0^x = 1 - e^{-x/\theta} \\ x \leq 0 \Leftrightarrow 0 \end{cases} \quad \text{con } \theta = \frac{1}{\lambda}$$

$$m_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx = \int_0^{\infty} \frac{1}{\theta} x^r e^{-x/\theta} dx \Leftrightarrow u = \frac{x}{\theta}, \quad dx = \theta du \Leftrightarrow m_r = \theta^r \int_0^{\infty} u^r e^{-u} du = \theta^r \Gamma(r+1) = r! \theta^r$$

$$E(X) = \mu = m_1 = \theta = \frac{1}{\lambda}$$

$$\text{var}(X) = \sigma^2 = m_2 - \mu^2 = 2\theta^2 - \theta^2 = \theta^2$$

Una caratteristica importante della distribuzione esponenziale è la proprietà di mancanza di memoria:  
 $P(X > b + c | X > b) = P(X > c) \quad \text{con } b, c \geq 0$

**Tempi di attesa in un processo di Poisson:**

$$X_t \equiv n^\circ \text{ di eventi in un intervallo di tempo } t \Rightarrow f_t(x) = P(X_t = x) = (\lambda t)^x \frac{e^{-\lambda t}}{x!}$$

$$T_x \equiv \text{tempo d'attesa per l}'x\text{-esimo evento} \Rightarrow g_x(t) = \frac{d}{dt} G_x(t) = \frac{\lambda^x}{\Gamma(x)} t^{x-1} e^{-\lambda t} \quad \text{con } t > 0$$

$$P(T_x > t) = P(X_t < x)$$

$$G_x(t) = P(T_x \leq t) = 1 - P(T_x > t) = 1 - P(X_t < x) = 1 - [f_t(0) + f_t(1) + \dots + f_t(x-1)] =$$

$$= 1 - \left[ 1 + \lambda t + \frac{(\lambda t)^2}{2!} + \dots + \frac{(\lambda t)^{x-1}}{(x-1)!} \right] e^{-\lambda t} \quad \text{con } t > 0$$

Derivando rispetto a  $t$  e semplificando, troviamo che la densità di probabilità per  $T_x$  è data da:

$$g_x(t) = \frac{d}{dt} G_x(t) = \frac{\lambda^x}{\Gamma(x)} t^{x-1} e^{-\lambda t} \quad \text{con } t > 0 \text{ è detta distribuzione gamma o di Erlang}$$

$$E(X) = \mu = \frac{\alpha}{\lambda}$$

$$\text{var}(X) = \sigma^2 = \frac{\alpha}{\lambda^2}$$

**Tempo di attesa per il primo evento:** il tempo di attesa tra eventi consecutivi in un processo di Poisson segue una distribuzione esponenziale con media  $\theta = 1/\lambda$ .

**Distribuzioni Gamma o di Erlang:** una v.a.  $X$  continua è distribuita secondo la legge gamma di parametri  $\alpha > 0$  e  $\lambda > 0$  se la relativa densità di probabilità è data da un'espressione del tipo:

$$X \sim \Gamma(\alpha, \lambda) \Leftrightarrow f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{con } \begin{cases} \alpha > 0 \\ \lambda > 0 \\ x > 0 \end{cases}$$

$$E(X) = \mu = \frac{\alpha}{\lambda}$$

$$\text{var}(X) = \sigma^2 = \frac{\alpha}{\lambda^2}$$

**Distribuzione Normale o Gaussiana:** una variabile random continua  $X$  con densità di probabilità:

$$X \sim N(\mu, \sigma^2) \Leftrightarrow f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{con} \begin{cases} \mu, \sigma \in \mathbb{R} \\ \sigma > 0 \\ -\infty < x < \infty \end{cases}$$

$$E(X) = \mu = x$$

$$\text{var}(X) = \sigma^2$$

La distribuzione gaussiana è una curva “a campana” centrata in  $x=\mu$  che è la media e la mediana della distribuzione. La larghezza dipende da  $\sigma$  e si può vedere che il 99.7% della probabilità è contenuto nell'intervallo  $(\mu - 3\sigma, \mu + 3\sigma)$ .

**Distribuzione Normale Standardizzata:** distribuzione normale con media  $\mu=0$  e varianza  $\sigma^2=1$ :

$$X \sim N(0,1) \Leftrightarrow f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{con} \quad -\infty < x < \infty$$

$$F(x) = \int_{-\infty}^x f(u) du = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Quest'integrale non può essere calcolato algebricamente e si ricorre all'integrazione numerica o alle tavole (Appendice B). Per l'uso di queste tavole è utile ricordare che, essendo la distribuzione normale standardizzata simmetrica rispetto a  $x=0$  si ha:

$$P(X \leq -x) = P(X \geq x) \Leftrightarrow F(-x) = 1 - F(x)$$

**Forma standard:** supponiamo che  $X \sim N(\mu, \sigma^2)$  e sia  $Z \equiv \frac{X-\mu}{\sigma}$  la forma standard di  $X$ . Poiché la trasformazione è monotona, la funzione densità di  $Z$  è:

$$g(z) = f(x) \left| \frac{dx}{dz} \right| = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Se  $X$  ha una distribuzione gaussiana, la forma standard di  $X$  ha una distribuzione gaussiana standardizzata. Questo risultato ci permette di ottenere le probabilità di ogni distribuzione normale  $X \sim N(\mu, \sigma^2)$  dalle tavole di  $N(0,1)$ . Con argomenti simili si dimostra che :

$$X \sim N(\mu, \sigma^2) \Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$$

**Distribuzione del Chi-quadrato:** si chiama legge del chi-quadrato a  $n$  gradi di libertà e si indica con

$$\chi_{(n)}^2 \text{ la distribuzione gamma } \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \quad \text{con} \begin{cases} \alpha = \frac{n}{2} \\ \lambda = \frac{1}{2} \end{cases}$$

$$X \sim \chi_{(n)}^2 \sim \chi_{(n)}^2 \equiv \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \Leftrightarrow f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

Il parametro  $n$  caratterizza la media e la varianza della stessa distribuzione:

$$m_r = \frac{2^r}{\Gamma(n/2)} \Gamma(n/2 + r)$$

$$E(X) = m_1 = n$$

$$\text{var}(X) = m_2 - m_1^2 = 2^2 (n/2 + 1)(n/2) - n^2 = 2n$$

Funzioni cumulative di  $\chi_{(n)}^2$  da 2 a 10 gradi di libertà:

$$\int_0^x \chi_{(1)}^2 dx = \frac{2}{\sqrt{\pi}} \int_0^{\frac{\sqrt{x}}{\sqrt{2}}} e^{-t^2} dt$$

$$\begin{aligned} \chi^2_{(2)} &= \frac{1}{2} \sqrt{\frac{1}{e^x}} & \chi^2_{(3)} &= \frac{\sqrt{\frac{1}{8}}}{0,8862} \sqrt{x} \sqrt{\frac{1}{e^x}} & \chi^2_{(4)} &= \frac{x}{4} \sqrt{\frac{1}{e^x}} \\ \chi^2_{(5)} &= \frac{\sqrt{\frac{1}{32}}}{1,329} \sqrt{x^3} \sqrt{\frac{1}{e^x}} & \chi^2_{(6)} &= \frac{x^2}{16} \sqrt{\frac{1}{e^x}} & \chi^2_{(7)} &= \frac{\sqrt{\frac{1}{128}}}{3,3234} \sqrt{x^5} \sqrt{\frac{1}{e^x}} \\ \chi^2_{(8)} &= \frac{x^3}{96} \sqrt{\frac{1}{e^x}} & \chi^2_{(9)} &= \frac{\sqrt{\frac{1}{512}}}{11,6317} \sqrt{x^7} \sqrt{\frac{1}{e^x}} & \chi^2_{(10)} &= \frac{x^4}{768} \sqrt{\frac{1}{e^x}} \end{aligned}$$

**Proprietà additive del Chi-quadrato:** supponiamo che le v.a.  $X_1, X_2, \dots, X_n$  siano indipendenti, e che  $X_i$  sia distribuita secondo la legge  $\chi^2$  con  $\nu_i$  gradi di libertà. Dalla proprietà di additività delle leggi gamma segue che la somma delle  $X_i$  ha una distribuzione  $\chi^2$ , con gradi di libertà pari alla somma  $\sum \nu_i$ :

$$X_1 + X_2 + \dots + X_n \sim \chi^2_{(\nu_1 + \nu_2 + \dots + \nu_n)}$$

Tale proprietà additiva è molto simile a quella di una distribuzione di Poisson.

**Connessione con la  $N(0,1)$ :**

$$Z \sim N(0,1) \Rightarrow Z^2 \sim \chi^2_{(1)}$$

$Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi^2_{(n)}$  la somma dei quadrati di  $n$  variabili random IID, distribuite secondo la  $N(0,1)$ , ha una distribuzione  $\chi^2$  con  $n$  gradi di libertà.

$$Z \sim N(0,1) \Leftrightarrow Z^2 \sim \chi^2_{(1)}$$

Poniamo  $X \equiv Z^2$ , e poiché tale trasformazione non è monotonica, usiamo il metodo della cumulativa. Sia  $G$  la funzione di ripartizione di  $X$ , allora:

$$G(x) = P(X \leq x) = P(Z^2 \leq x) = P(|Z| \leq \sqrt{x}) = P(-\sqrt{x} \leq Z \leq \sqrt{x}) = F(\sqrt{x}) - F(-\sqrt{x})$$

dove  $F$  è la cumulativa di  $Z$ . La densità  $g(x)$  di  $X$  è data quindi dall'espressione:

$$g(x) = \frac{d}{dx} G(x) = \frac{d}{dx} F(\sqrt{x}) - \frac{d}{dx} F(-\sqrt{x})$$

Notiamo che queste relazioni valgono in generale. Tenendo conto del fatto che la densità di  $Z$  è la  $N(0,1)$

$$\frac{d}{dz} F(z) = f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \Leftrightarrow \left\{ \frac{d}{dx} = \frac{d}{dz} \frac{dz}{dx} \right\} \Leftrightarrow g(x) = f(\sqrt{x}) \frac{d\sqrt{x}}{dx} - f(-\sqrt{x}) \frac{d(-\sqrt{x})}{dx} = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2} \quad x > 0$$

**Torema del limite centrale:** ogni combinazione lineare di variabili random normali indipendenti  $X_1, X_2, \dots, X_n$  ha una distribuzione normale. Si può mostrare che se:

$$X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow \sum a_i X_i \sim N(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2)$$

Considerando un insieme di variabili random indipendenti  $X_1, X_2, \dots, X_n$ , con la stessa media  $\mu$  e stessa varianza  $\sigma^2$ . La somma  $S_n$  e la media campione  $\bar{X}$  delle variabili  $\{X_i\}$  sono date da:

$$S_n = \sum X_i; \quad \bar{X} = \frac{1}{n} \sum X_i = \frac{1}{n} S_n$$

Dalla varianza di una combinazione lineare con v.a. indipendenti e dal sample mean si ha:

$$E(S_n) = n\mu; \quad \text{var}(S_n) = n\sigma^2$$

$$E(\bar{X}) = \mu; \quad \text{var}(\bar{X}) = \sigma^2/n$$

Se le  $X_i$  sono normalmente distribuite si ottiene:

$$S_n \sim N(n\mu, n\sigma^2)$$

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

La somma di variabili random indipendenti e gaussiane ha ancora una distribuzione gaussiana, con media  $\sum \mu_i$  e varianza  $\sum \sigma_i^2$ . Se  $X_1, X_2, \dots, X_n$  sono variabili random indipendenti con medie  $\mu_1, \mu_2, \dots, \mu_n$  e varianze  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$  allora la variabile somma  $S_n$  ha media  $\sum \mu_i$  e varianza  $\sum \sigma_i^2$ . La forma standard di  $S_n$ ,  $S_n^* = (S_n - \sum \mu_i) / \sqrt{\sum \sigma_i^2}$ , ha media 0 e varianza 1 per ogni  $n$ .

Il Teorema del limite centrale afferma che la forma standard di  $S_n$  è distribuita approssimativamente come la normale standardizzata  $S_n^* \approx N(0,1)$  per  $n$  sufficientemente grande; quindi la distribuzione limite di  $S_n^*$  è la stessa a prescindere dalle distribuzioni delle  $X_i$ .

**Approssimazione Normale della Binomiale:** considerando  $n$  prove di Bernoulli, definiamo la variabile random  $X_i$  che assume valore 1 se il risultato dell' $i$ -esima prova è un successo, 0 in caso contrario:

$$P(X_i = 1) = p; \quad P(X_i = 0) = 1 - p \quad \text{con } i = 1, 2, \dots, n$$

Considerando la somma  $S_n = X_1 + X_2 + \dots + X_n$ , ogni successo dà un contributo 1, ogni insuccesso dà contributo nullo; quindi  $S_n$  rappresenta il numero totale di successi in  $n$  prove. Conseguentemente,  $S_n$  ha una distribuzione Binomiale  $b(n, p)$ , con media  $np$  e varianza  $np(1-p)$ . Le  $X_i$  si riferiscono a prove indipendenti, quindi sono variabili indipendenti, e per il Teorema del limite centrale si ha che la distribuzione di  $S_n$  approssima quella Normale per  $n$  sufficientemente grande  $N(np, np(1-p))$ .

**Approssimazione Normale della Poissoniana:** prese  $X_1, X_2, \dots, X_n$  variabili random indipendenti poissoniane con media  $m$  e  $S_n = X_1 + X_2 + \dots + X_n$  la loro somma; questa ha ancora una distribuzione poissoniana di media  $\mu = nm$ . Poiché le v.a. sono indipendenti, il Teorema del limite centrale assicura che, per  $n$  grande,  $S_n$  ha una distribuzione Normale. Da queste considerazioni segue che una distribuzione di Poisson con media  $\mu$  grande può essere approssimata da una distribuzione Normale  $N(\mu, \mu)$ .

**Tempo di attesa per il primo evento:** si è visto che il tempo di attesa per il primo evento non dipende dall'origine dell'intervallo, cioè il tempo medio di attesa per il primo evento coincide con il tempo medio tra eventi consecutivi; questo risultato sembra apparentemente paradossale e va contro l'intuizione. Supponiamo di avere una collezione di aste di varia lunghezza. Ne selezioniamo una in modo random e la spezziamo, in modo random, in due parti di cui una viene conservata. Ripetiamo l'operazione molte volte. E' ovvio, e in questo caso è vero, che la lunghezza media delle aste spezzate sarà pari alla metà della lunghezza media delle aste madri.

Sembra legittimo, per analogia, estendere la stessa conclusione alla collezione di intervalli di tempo in un processo di Poisson: in questo caso gli intervalli di tempo tra eventi consecutivi prendono il posto delle aste. Il tempo di attesa per il primo evento, che è l'intervallo tra un'origine random e il primo evento, successivo a tale origine, dovrebbe essere, per analogia con l'esempio precedente, pari alla metà del tempo medio tra eventi consecutivi (l'origine random dovrebbe spezzare il generico intervallo in due parti di lunghezza pari, mediamente, alla metà della lunghezza media degli intervalli tra eventi consecutivi). In realtà, questi due tempi medi coincidono. Una spiegazione qualitativa può essere la seguente; la scelta dell'origine random è equivalente a porre stocasticamente un punto su un asse temporale. Quindi l'origine random cadrà preferibilmente sugli intervalli più lunghi, con probabilità proporzionale alla lunghezza, cosicché quella che appare essere una scelta casuale è effettivamente sbilanciata a favore degli intervalli di tempo più lunghi. Tornando all'esempio precedente, è come se selezionassimo le aste da spezzare in modo tale che la probabilità di beccare un'asta non fosse uniforme, ma proporzionale alla lunghezza delle aste. Rimane sorprendente, in qualche misura, che i due effetti si compensino esattamente. Se si vuole, questa è una prerogativa della poissoniana.

## CAPITOLO IX - Processi stocastici. Teoria delle code

**Processi:** si dividono in:

- 1) **stocastici** = funzione di 2 v.a. di cui una è il tempo  $t$ :  

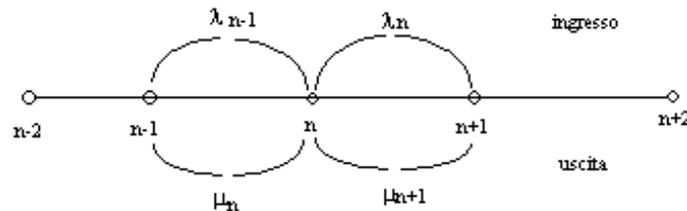
$$X(t) = f(Y, t)$$

2) **di Markov** = il passo  $n$  dipende soltanto da quello precedente:

$$f(x_n t_n | x_{n-1} t_{n-1}, \dots, x_1 t_1) = f(x_n t_n | x_{n-1} t_{n-1}) \quad \text{con } t_1 < t_2 < \dots < t_n$$

3) **ad un passo** = continui nel tempo con  $R_x \in \mathbb{N}$ :

$$\frac{d}{dt} p_n(t) = \mu_{n+1} p_{n+1}(t) + \lambda_{n-1} p_{n-1}(t) - (\mu_n + \lambda_n) p_n(t) \quad \text{con } p_n(t) \equiv P(X(t) = n)$$



4) **di Poisson** = ad un passo di puro ingresso con  $\mu_n = 0$  e con tasso di transizione costante  $\lambda_n = \lambda$ :

$$\frac{d}{dt} p_n(t) = \lambda [p_{n-1}(t) - p_n(t)]$$

Facendo la **derivazione** di quest'equazione differenziale si ha che, con  $(t, t + \varepsilon]$  con  $\varepsilon > 0$ , il numero aspettato di eventi in questo intervallo è  $\lambda \varepsilon$ , si può scrivere  $\lambda \varepsilon = 0 \cdot q_0 + 1 \cdot q_1 + 2 \cdot q_2 + \dots$  e segue che  $q_1 = \lambda \varepsilon$ ,  $q_0 = 1 - \lambda \varepsilon$  e  $q_2 = q_3 = \dots = 0$ . Considerando ora l'intervallo  $(0, t + \varepsilon]$  come l'unione dei due intervalli disgiunti,  $(0, t]$  e  $(t, t + \varepsilon]$ . Poiché  $n$  eventi in  $(0, t + \varepsilon]$  si possono suddividere nei due sottointervalli disgiunti in  $n+1$  modi, segue che (per l'indipendenza):

$$p_n(t + \varepsilon) \equiv P\{n \text{ eventi in } (0, t + \varepsilon]\} = \sum_{i=0}^n P\{n-i \text{ eventi in } (0, t] \text{ e } i \text{ eventi in } (t, t + \varepsilon]\} = \sum_{i=0}^n p_{n-i}(t) \cdot q_i$$

cioè:

$$p_n(t + \varepsilon) = p_n(t) \cdot q_0 + p_{n-1}(t) \cdot q_1 = p_n(t) \cdot (1 - \lambda \varepsilon) + p_{n-1}(t) \cdot \lambda \varepsilon \Leftrightarrow \frac{p_n(t + \varepsilon) - p_n(t)}{\varepsilon} = \lambda [p_{n-1}(t) - p_n(t)]$$

Per  $\varepsilon \rightarrow 0$  si ottiene infine l'equazione differenziale, con  $p_0(0) = 1$ ;  $p_n(0) = 0 \Rightarrow p_n(0) = \delta_{n,0}$  poiché in un intervallo temporale di lunghezza nulla non si possono avere eventi. Si può facilmente verificare che la soluzione dell'equazione differenziale con le condizioni dà la distribuzione di Poisson:

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

5) **Catene di Markov** = processi di Markov ad un passo con  $t \in \mathbb{N}$ . Una catena di Markov, vista come processo di Markov, è caratterizzata da:

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = k, \dots, X_1 = m\} = P\{X_{n+1} = j | X_n = i\} \equiv p_{ij}(n)$$

Una classe particolare di catene è quella in cui le probabilità  $p_{ij}(n)$  non dipendono da  $n$ ; in questo caso diciamo che la **catena è omogenea nel tempo** e si scrive  $P\{X_{n+1} = j | X_n = i\} = p_{ij}$

Se l'insieme degli stati della catena, cioè i possibili valori di  $X$ , è finito allora l'insieme dei numeri  $p_{ij}$  forma una matrice quadrata di ordine pari alla cardinalità di  $R_x$ , detta **matrice di transizione**, che ha le seguenti proprietà:

- tutti i suoi elementi sono  $\geq 0$ ;
- la somma degli elementi di ciascuna riga è pari a 1  $\Rightarrow \sum_{j \in R_x} p_{ij} = 1$ ;

**Teoria delle code: xxx**