1. Variabili casuali discrete

1.1 definizioni

1.1.1 variabile casuale

def Una **variabile casuale** è una grandezza X che può assumere valori discreti indicati con x_i e ciascuno di essi è associato ad un corrispondente valore di probabilità indicato con p_i.
Il valore x_i può essere il risultato di una misura di una grandezza che può avere solo valori discreti (per esempio l'altezza misurata in cm in una popolazione) oppure essere semplicemente il modo per descrivere un evento. Per esempio potremmo individuare una variabile che descrive gli esiti di un doppio lancio di una moneta con una variabile che descrive il numero di teste ottenute con una variabile che prende i valori 0, 1, 2 e che rappresenta i risultati CC, TC (o CT), TT.

Una variabile casuale sarà dunque caratterizzata da una matrice del tipo: X = $\begin{pmatrix} x_1, x_2, x_3, \dots \\ p_1, p_2, p_3, \dots \end{pmatrix}$ con le ulteriori condizioni

$$\forall i, p_i \geq 0 \text{ e } \sum_{i=1}^{n} p_i = 1 \qquad (1.1)$$

L'estremo superiore n può anche essere infinito. I valori x_i per comodità si suppongono ordinati.

def Si chiama **funzione di ripartizione** della variabile X la funzione F(X) tale che

$$F(x_k) = p(x \leq x_k) \qquad (1.2)$$

es. data la variabile casuale X così definita X = $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0.2 & 0.3 & 0.1 & 0.15 & 0.05 & 0.2 \end{pmatrix}$ determinare la $p(x\leq3)$.
$p(x\leq3) = F(3) = 1 - p(x>3) = 1 - (p(x=4) + p(x= 5) = 1 - 0.25 = 0.75$

1.1.2 Valori di sintesi

def **Valore medio**, o *valore atteso* o *speranza matematica* viene indicato con il simbolo μ e corrisponde alla media ponderata della variabile casuale:

$$\mu_x = \sum_{i=1}^{n} x_i p_i \qquad (1.3)$$

es. Per la variabile dell'esercizio precedente $\mu_X = 0\cdot0.2 + 1\cdot0.3 + 2\cdot0.1 + 3\cdot0.15 + 4\cdot0.05 + 5\cdot0.2 = 2.15$

def **Valore modale,** o *moda* è il valore x_i per il quale si ha p_i = Max

es. Per la variabile dell'esercizio precedente il valore modale è x = 1

def **Valore mediano,** o *mediana* è il valore x_i (che può anche non esistere visto che la variabile si muove a scatti) per il quale si ha $F(x) = ½$.

es. Per la variabile dell'esercizio precedente la mediana è x = 1 e coincide con la moda.

Il valor medio, la moda e la mediana, quando è assegnata la legge di costruzione delle probabilità risultano calcolabili e sono fortemente correlate ma il tipo di correlazione dipende dalla legge medesima.
In fisica, in tutte le elaborazioni statistiche riguardanti sistemi complessi si fa un largo uso di tutti e tre questi indicatori (vita media di un nucleo radioattivo, valore modale nella distribuzione delle velocità molecolari di un gas di particelle, valore modale nello spettro energetico del corpo nero o dei raggi X).

def **Varianza**, è la media degli scarti quadratici e si indica con σ^2

$$\sigma^2_x = \mu_{(x-\mu)^2} = \sum_{i=1}^{n} (x_i - \mu_x)^2 p_i \qquad (1.4)$$

La radice quadrata della varianza indicata con σ è detta *scarto quadratico medio* o *deviazione standard*. Come si vedrà la varianza ha un ruolo centrale nell'indicare la dispersione della variabile intorno al suo valore medio e dunque gioca un ruolo essenziale nella teoria degli errori.

es. Per la variabile dell'esercizio precedente si ha:

$\sigma^2_x = (0 - 2.15)^2\cdot 0.2 + (1 - 2.15)^2\cdot 0.3 + (2 - 2.15)^2\cdot 0.1 + (3 - 2.15)^2\cdot 0.15 + (4 - 2.15)^2\cdot 0.05 + (5 - 2.15)^2\cdot 0.2 = 3.2275$

1.2 Proprietà notevoli di media e varianza

1.2.1 Linearità della media

T $Z = \alpha X + \beta \Rightarrow \mu_Z = \alpha\mu_X + \beta$ (1.5)

dim

$$\mu_Z = \sum_{i=1}^{n}(\alpha x_i + \beta)p_i = \sum_{i=1}^{n}(\alpha x_i)p_i + \sum_{i=1}^{n}(\beta)p_i = \alpha\sum_{i=1}^{n}x_i p_i + \beta\sum_{i=1}^{n}p_i = \alpha\mu_X + \beta$$

☺

T $Z = X + Y \Rightarrow \mu_Z = \mu_X + \mu_Y$ (1.6)

dim

$$\mu_Z = \sum_{i,j}(x_i + y_j)\, p(X = x_i \wedge Y = y_j) = \sum_{i,j} x_i\, p(X = x_i \wedge Y = y_j) + \sum_{i,j} y_j\, p(X = x_i \wedge Y = y_j)$$

ma sommando le probabilità congiunte si ottengono le probabilità marginali e pertanto

$$\mu_Z = \sum_{i} x_i\, p(X = x_i) + \sum_{j} y_j\, p(Y = y_j) = \mu_X + \mu_Y$$

☺

Dunque la media è un operatore lineare e prese due variabili X e Y se indichiamo con Z una terza variabile che sia la combinazione lineare delle altre due $Z = \alpha X + \beta Y$

$\mu_z = \alpha\mu_x + \beta\mu_y$ (1.7)

In particolare se si considerano n variabili $X_1, X_2, \ldots X_n$ tutte dotate dello stesso valore medio μ e si considera la variabile somma:

T $S_n = \sum_{i=1}^{n} X_i \Rightarrow \mu_{S_n} = n\,\mu$ (1.8)

Di questa proprietà si fa un notevole utilizzo nel seguito.

1.2.2 Media di una costante

La media di una variabile casuale costante è la costante stessa.

$\mu_k = k$ (1.6)

dim In effetti se $x_i = k$ si ha $p_i = \text{cost} = 1/n$ e dunque $\mu = \sum_{i=1}^{n} k\frac{1}{n} = k\sum_{i=1}^{n}\frac{1}{n} = k$

☺

$\mu_{(x-\mu)} = 0$ (1.7)

dim Basta applicare la linearità e la proprietà precedente.

1.2.3 Un metodo per calcolare contestualmente media e varianza

Se si osservano con attenzione i tasti statistici delle calcolatrici scientifiche si osserva che esse sono in grado di calcolare contestualmente sia la media sia la varianza. In effetti esse contengono due registri di memoria che tengono aggiornate $\sum x$ e $\sum x^2$ e queste due sommatorie consentono il calcolo diretto della varianza senza doversi riferire a tutti gli scarti quadratici (la cui valutazione richiede il calcolo preventivo del valor medio).

$\boxed{\text{T}}$ $\sigma^2 = \mu_{x^2} - (\mu_x)^2$ (1.8)

$\boxed{\text{dim}}$ La dimostrazione è di tipo diretto: $\sigma^2_x = \sum_{i=1}^{n} (x_i - \mu_x)^2 p_i = \sum_{i=1}^{n} (x_i^2 p_i - 2\mu_x x_i p_i + \mu_x^2 p_i) = \mu_{x^2} - 2\mu_x^2 + \mu_x^2 \sum p_i = \mu_{x^2} - (\mu_x)^2$ visto che $\sum p_i = 1$

☺

$\boxed{\text{es.}}$ calcoliamo la varianza della variabile casuale già usata con la nuova formula

X	0	1	2	3	4	5
p	0.2	0.3	0.1	0.15	0.05	0.2
X^2	0	1	4	9	16	25
pX^2	0	0.3	0.4	1.35	0.8	5.0

e dunque $\sum pX^2 = 7.85$, $(\sum pX)^2 = 2.15^2$ e infine $\sigma^2 = 10.35 - 2.15^2 = 3.23$

1.2.4 La varianza e le costanti moltiplicative o additive

$\boxed{\text{T}}$ $Z = aX + b \Rightarrow \sigma^2_Z = a^2\sigma^2_X$ (1.9)

$\boxed{\text{dim}}$ $\sigma^2_{aX+b} = \sum_{i=1}^{n}[(ax_i + b) - (a\mu + b)]^2 p_i = \sum_{i=1}^{n}(ax_i - a\mu)^2 p_i = \sum_{i=1}^{n} a^2(x_i - \mu)^2 p_i = a^2\sum_{i=1}^{n} (x_i - \mu)^2 p_i = a^2\sigma^2_X$

☺

Dunque le costanti moltiplicative incidono sulla varianza al quadrato mentre quelle addittive non hanno rilevanza.

Ne consegue che se si considera la variabile

$$Z = \frac{X - \mu_X}{\sigma_X} \Rightarrow \mu_Z = 0 \wedge \sigma^2_Z = 1 \qquad (1.10)$$

La trasformazione di variabile appena realizzata si chiama *standardizzazione di una variabile* e consente di poter confrontare tra loro variabili dotate di valori medi e varianze diverse attraverso la costruzione di variabili ausiliarie che, per definizione, hanno valor medio 0 e varianza 1.

1.2.5 Per variabili indipendenti la varianza della somma è pari alla somma delle varianze

Quando si opera sulla somma la varianza *conserva la proprietà di linearità solo in presenza di variabili casuali indipendenti*, cioè di variabili per le quali la probabilità congiunta è pari al prodotto delle probabilità.

$\boxed{\text{T}}$ $Z = X + Y \wedge p(X=x_i \wedge Y = y_j) = p(X=x_i)\, p(Y = y_j) \Rightarrow \sigma^2_Z = \sigma^2_X + \sigma^2_Y$ (1.11)

$\boxed{\text{dim}}$ La dimostrazione richiede di ricordare che $\mu_Z = \mu_X + \mu_Y$ e procede direttamente anche se in maniera un po' faticosa a causa dei doppi indici coinvolti:

$$\sigma^2_Z = \sum_{i,j}[(x_i + y_j) - (\mu_x + \mu_Y)]^2 p(X = x_i \wedge Y = y_j) = \sum_{i,j}[(x_i - \mu_x) + (y_j - \mu_Y)]^2 p(X = x_i \wedge Y = y_j) =$$

$$= \sum_{i,j}[(x_i - \mu_x)^2 p(X = x_i \wedge Y = y_j) + \sum_{i,j}(y_j - \mu_Y)^2 p(X = x_i \wedge Y = y_j) + 2\sum_{i,j}(x_i - \mu_x)(y_j - \mu_Y)\, p(X = x_i \wedge Y = y_j) =$$

$$= \sum_{i}[(x_i - \mu_x)^2 p(X = x_i) + \sum_{j}(y_j - \mu_Y)^2 p(Y = y_j) + 2\sum_{i,j}(x_i - \mu_x)(y_j - \mu_Y)\, p(X = x_i)\, p(Y = y_j) =$$

$$= \sigma^2_X + \sigma^2_Y + 2\sum_{i}(x_i - \mu_x)\, p(X = x_i) \cdot \sum_{j}(y_j - \mu_Y)\, p(Y = y_j) = \sigma^2_X + \sigma^2_Y + 2\cdot 0\cdot 0 = \sigma^2_X + \sigma^2_Y$$

☺

dove si è sfruttata la decomposizione del prodotto e il fatto che il valor medio degli scarti è nullo.

In particolare se si considerano n variabili $X_1, X_2, \dots X_n$ tutte dotate dello stesso valore medio μ e della stessa varianza σ^2 la variabile somma ha varianza pari a n volte la varianza di ciascuna

$\boxed{\text{T}}$ $S_n = \sum_{i=1}^{n} X_i \Rightarrow \sigma^2_{S_n} = n\,\sigma^2$ (1.12)

$$\boxed{\text{T}}\ \frac{S_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n} \Rightarrow \sigma^2_{S_n/n} = n\frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n} \qquad (1.13)$$

Il secondo dei due corollari appena visti è particolarmente interessante: ci dice che se consideriamo la media aritmetica dei valori assunti dalle n variabili casuali (tutte con caratteristiche simili) la sua varianza (cioè il suo grado di dispersione intorno alla media) tende a diminuire al crescere di n. La questione sarà ripresa in statistica inferenziale.

2. Alcune distribuzioni importanti

2.1 La distribuzione geometrica

2.1.1 Di che problema si tratta?

La distribuzione geometrica si presenta quando si cerca di rispondere alla seguente domanda: *se un evento ha probabilità p quanto vale la probabilità che nella ripetizione dell'esperimento esso abbia luogo per la prima volta all'ennesimo tentativo?*

2.1.2 Il teorema

Se indichiamo con X la variabile casuale che descrive il tentativo a cui si ha il primo successo e con I e S l'insuccesso e il successo, si richiede che accada l'evento $\boxed{\text{I}}\boxed{\text{I}}\boxed{\text{I}}\ldots\boxed{\text{S}}$ e dunque la corrispondente probabilità, visto che si tratta di eventi indipendenti, è:

$\boxed{\text{T}}$ $p(X = n) = (1 - p)^{n-1}p$ (2.1)

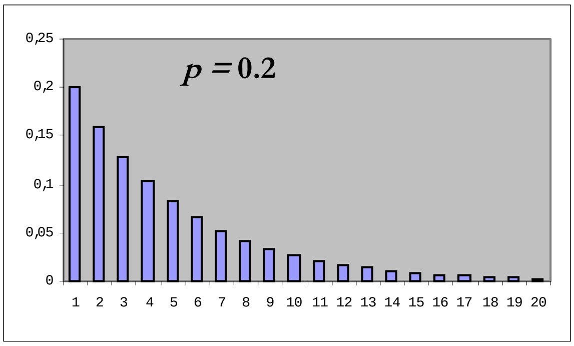
Non si tratta di una questione di testa o croce ma del dare risposta a questioni come la seguente.

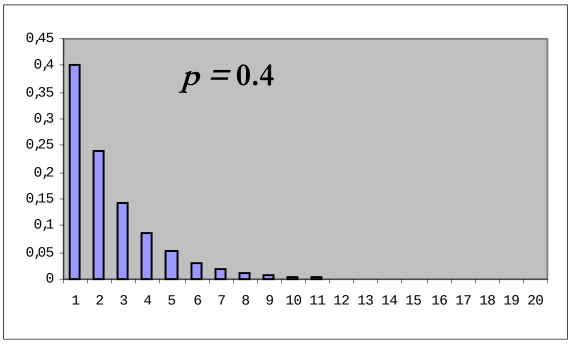
$\boxed{\text{es.}}$ Una coppia si sottopone alla inseminazione artificiale e la probabilità di una gravidanza è di 0.15 per ogni tentativo. Quanto vale la probabilità che si abbia una gravidanza al quinto tentativo?

$p(X = 5) = (1 - 0.15)^4 \cdot 0.15 = 0.0783$

N.B. attenzione ciò non significa che ad ogni tentativo la probabilità sia sempre più bassa; essa ogni volta è 0.15 (stiamo ragionando su eventi indipendenti). Quella che abbiamo valutato è, prescindendo da ogni conoscenza precedente, la probabilità che il successo si abbia al quinto tentativo e cioè che si abbiano quattro insuccessi seguiti da un successo.

La distribuzione ha un andamento tanto più rapidamente decrescente quanto è grande p come si può osservare dai due istogrammi seguenti.





2.1.3 La funzione di ripartizione

La probabilità che l'evento avvenga entro l'ennesimo tentativo cioè $p(X \le n)$ corrisponde ad una somma che rimanda alle progressioni geometriche e questo ne spiega il nome.

In effetti, visto che incrementare n di una unità equivale a richiedere un insuccesso in più, $p_{n+1} = p_n (1 - p)$ e dunque le probabilità costituiscono una progressione di ragione $1 - p$. La somma di una progressione geometrica di ragione q è:

$$S_n = p_1 + p_2 + \ldots + p_n = p_1(1 + q^1 + q^2 + \ldots q^{n-1}) = p_1 \frac{1 - q^n}{1 - q}$$

Nel nostro caso $p_1 = p$ mentre $q = 1 - p$ e dunque si ottiene

$$S_n = p\,\frac{1 - (1 - p)^n}{p} = 1 - (1 - p)^n$$

$\boxed{\text{T}}$ $F(n) = 1 - (1 - p)^n$ (2.2)

La conoscenza della funzione di ripartizione ci consente di risolvere qualunque problema relativo alla probabilità su un intervallo.

Per esempio se si vuol valutare la probabilità che l'evento si verifichi per la prima volta tra l'emmesimo e l'ennesimo tentativo con n > m avremo:

$p(m \le X \le n) = p(X \le n) - p(X \le m-1) = -(1-p)^n + (1-p)^{m-1}$

es. Una coppia si sottopone alla inseminazione artificiale e la probabilità di una gravidanza è di 0.15 per ogni tentativo. Quanto vale la probabilità di avere un esito positivo tra il secondo e il quarto tentativo?

$p(2 \le X \le 4) = -(1-p)^4 + (1-p)^1 = -0.85^4 + 0.85 = 0.328$

2.2 La distribuzione binomiale

2.2.1 Il teorema

Consideriamo n variabili casuali $X_1, X_2, \dots, X_n$ con le seguenti caratteristiche: $\forall i,\ X_i = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

Indichiamo ora, come è già stato fatto nei paragrafi precedenti, con S_n la somma dei valori delle n variabili casuali $S_n = \sum_{i=1}^{n} X_i$. Otterremo così una nuova variabile casuale che può assumere uno qualsiasi degli n+1 valori da 0 a n. La distribuzione binomiale, o distribuzione normale o distribuzione di Bernoulli [1] risponde al problema *quanto valgono le probabilità della variabile casuale S_n?*

T $p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ (2.1)

dim Poiché la variabile assume I valori 0 o 1 la variabile S_n rappresenta il numero di eventi che hanno dato risultato 1 tra le n prove. Per questa ragione la distribuzione binomiale prende anche il nome di distribuzione delle *prove ripetute*. E' come se si eseguisse un esperimento con due possibili esiti (favorevole o sfavorevole) e si stesse misurando la probabilità di avere k esiti favorevoli.

Per avere k esiti favorevoli l'evento favorevole deve avvenire esattamente k volte mentre n – k volte si deve verificare quello sfavorevole. Poiché si tratta di eventi del tutto indipendenti la probabilità che si verifichi una delle possibili configurazioni con k esiti favorevoli è dunque $p^k(1-p)^{n-k}$.

Ma gli esiti favorevoli corrispondono a dare valore 1 a k caselle scelte tra n cioè alle combinazioni di n oggetti in classe k. Si tratta di combinazioni e non disposizioni perché le nostre variabili casuali sono ordinate (la X_1 è sempre al primo posto e così via). Dunque la probabilità richiesta è la somma di $\binom{n}{k}$ probabilità tutte uguali $p^k(1-p)^{n-k}$ cioè il risultato.

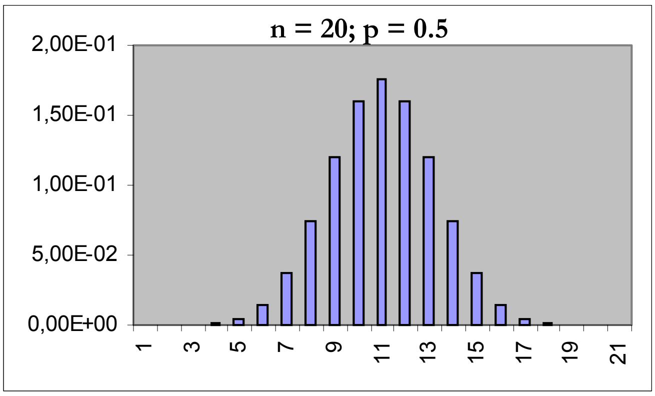
☺

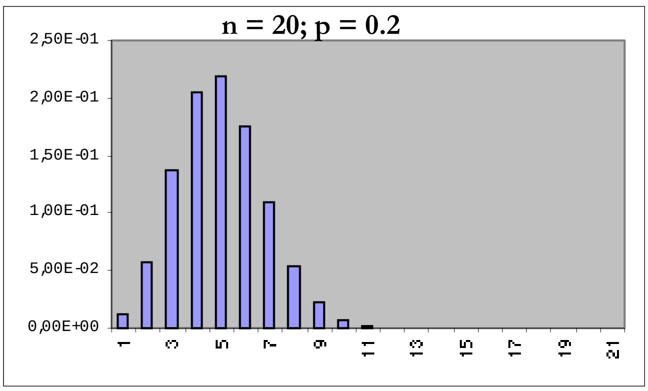
2.2.2 Visualizziamo il risultato

Supponiamo che sia n= 20. I corrispondenti coefficienti binomiali valgono rispettivamente

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\binom{20}{k}$	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756	167960	125970	77520	38760	15504	4845	1140	190	20	1

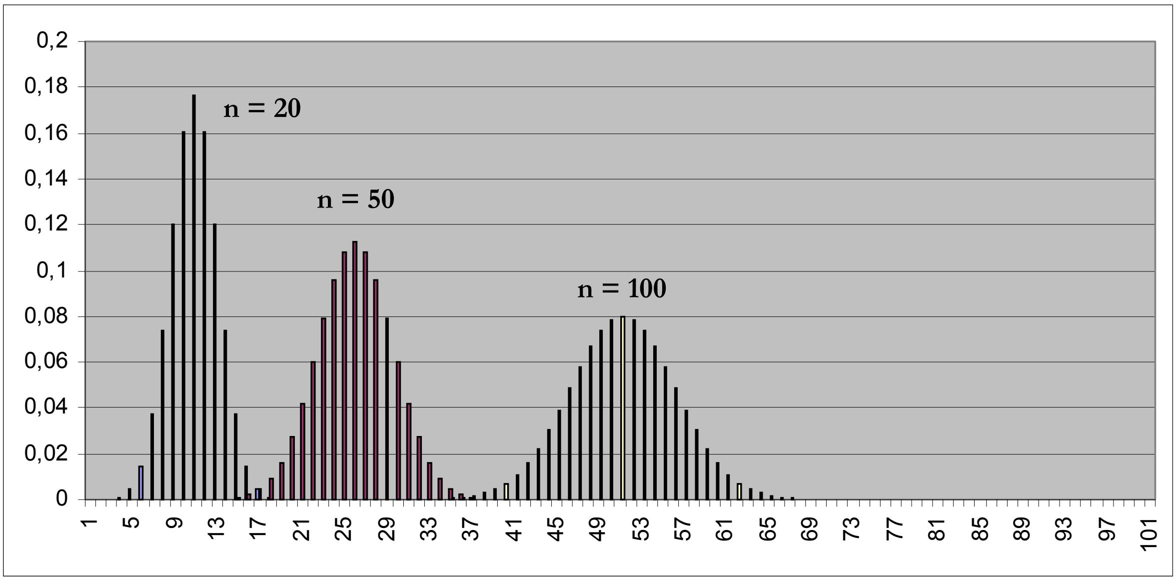
mentre i coefficienti binomiali sono simmetrici perché $\binom{n}{k} = \binom{n}{n-k}$ l'altro termine del prodotto $p^k(1-p)^{n-k}$ lo è solo per $p = ½$ e per questa ragione le curve della binomiali, in generale non sono simmetriche. [2]



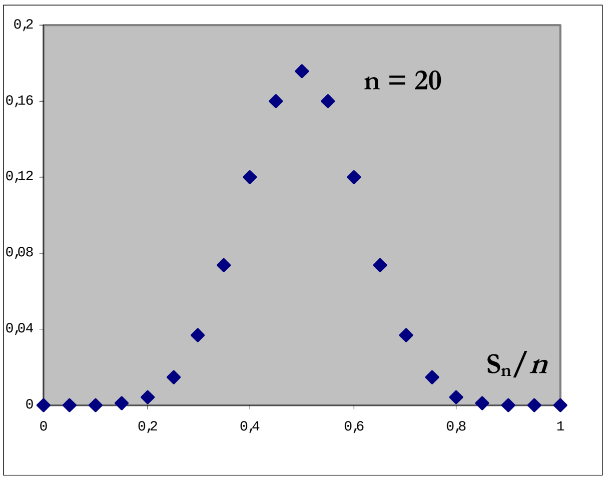


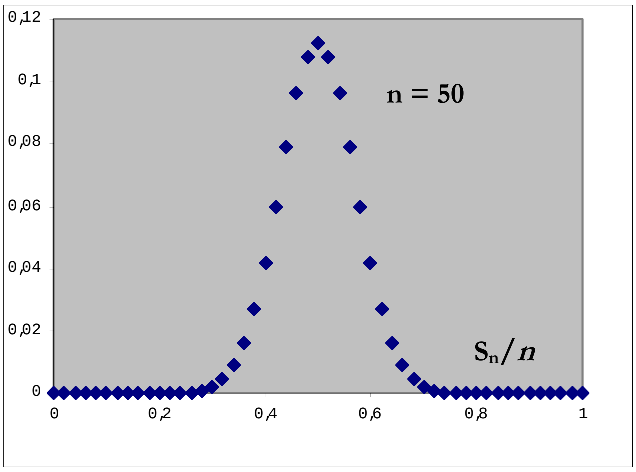
[1] Pubblicata nel 1713 da Nicola Bernoulli e dovuta allo zio Giacomo.

Cosa accade alla distribuzione binomiali al crescere di *n*? Ci aspettiamo che la curva si regolarizzi e cioè che i salti tra un valore di probabilità e quelli vicini diminuiscano. Ci aspettiamo anche che si presentino con sempre maggiore probabilità i risultati che si staccano poco dal valor medio. Ma, al crescere di *n*, aumenta il numero di risultati possibili e dunque, mentre il valore modale si sposta verso destra, osserveremo anche una attenuazione dei corrispondenti valori di probabilità visto che la somma di tutte le ordinate del diagramma deve comunque fare 1.
Nella figura qui sotto sono state rappresentate le binomiali per $n = 20, 50, 100$ ed è possibile osservare quanto previsto: maggiore regolarità, spostamento del valore modale ed attenuazione dei valori di probabilità.



Per poter confrontare le distribuzioni relative ad uno stesso valore di *p* con valori di *n* diversi è necessario scegliere come variabile, invece di *k* il suo valore relativo *k*/*n* che rappresenta la frequenza relativa di eventi favorevoli.
Le prossime due figure rappresentano i due istogrammi relative ai casi $n = 20$ e $n = 50$ con in ascissa la frequenza relativa che consente di rendere comparabili le due curve ed osservare come, al crescere di *n*, la curva tenda, in realtà, a stringersi intorno al valore modale.





2.2.3 Gli indicatori statistici della curva binomiale

$\boxed{\text{T}}$ $\mu_{S_n} = n\,p$ (2.2)

$\boxed{\text{dim}}$ La variabile X_i ha valore medio pari a *p*.

---

[2] Questi diagrammi e i successivi sono stati tutti ottenuti utilizzando le capacità statistiche e di rappresentazione di Excel.

In effetti $\mu_X = 1 \cdot p + 0 \cdot (1 - p) = p$. Poiché le variabili X_i hanno tutte lo stesso valore medio e sono n per la linearità si ha il risultato.
☺

T $\sigma^2_{S_n} = n\,p\,(1 - p)$ (2.3)

dim La variabile X_i ha varianza $\sigma^2_{X_i} = p\,(1 - p)^2 + (1 - p)(0 - p)^2 = p\,(1 - p)(1 - p + p) = p\,(1 - p)$

Poiché le variabili hanno tutte la stessa varianza e sono indipendenti le varianze si sommano e si ha la tesi.
☺

Se, invece di S_n, si considera la frequenza relativa S_n/n il valor medio risulta p mentre la varianza risulta pari a $\frac{n\,p\,(1 - p)}{n^2} = \frac{p\,(1 - p)}{n}$ e cioè la probabilità delle frequenze relative tende, al crescere di n, a stringersi intorno al valore modale.

Dove si trova il valore modale della distribuzione binomiale quando $p \neq ½$ cioè quando la curva non è simmetrica.?

Un modo per delinearne l'andamento consiste nello studiare il crescere della probabilità al variare di k e per farlo, visto che operiamo nel discreto, basterà studiare il rapporto delle probabilità per valori consecutivi. Quando il rapporto è maggiore di 1 la curva sarà crescente e diventerà decrescente viceversa.

Per ragioni di comodità di scrittura indichiamo $p(S_n = k)$ con $p_n(k)$.

$$\frac{p_n(k+1)}{p_n(k)} = \frac{n!}{(k+1)!\,(n - k - 1)!}p^{k+1}(1 - p)^{n-k-1}\frac{k!\,(n - k)!}{n!\,p^k\,(1 - p)^{n-k}} = \frac{n - k}{k + 1}\frac{p}{1 - p}$$

Basta ora risolvere la disequazione in k, $\frac{n - k}{k + 1}\frac{p}{1 - p} > 1$ per determinare la condizione di crescenza e viceversa.

2.2.4 Esercizi sulla distribuzione binomiale

1. Si consideri il seguente cammino casuale di n passi. Ad ogni passo, con probabilità p = ½, si può andare avanti di 1 o indietro di 1. Si parte da 0 e si vuole trovare la probabilità che la posizione Y sia pari ad un valore k preassegnata dopo n passi. Si esegua il calcolo per n = 10 e k = 5 oppure k = – 2.

Rispetto alla variabile S_n più volte considerata questa volta le X_i possono valere +1 o – 1, ma possiamo aggirare l'ostacolo che ci costringerebbe a costruire una nuova funzione di distribuzione, osservando che, indicato con h il numero di successi la posizione corrispondente è data dalla somma algebrica di successi ed insuccessi (pari a $n - h$) e dunque la posizione viene descritta da una nuova variabile casuale $Y_n = h - (n - h) = 2h - n$ dove h rappresenta la variabile casuale S_n. Abbiamo dunque $Y_n = 2\,S_n - n$ e per trovare la probabilità che Y_n = k basta trovare la probabilità del corrispondente valore di $S_n = ½\,(Y_n + n)$

$p(Y_{10} = 5) = p(S_{10} = 15/2) = 0$. In un numero pari di passi non si può compiere uno spostamento dispari.

$p(Y_{10} = -2) = p(S_{10} = 4) = \binom{10}{4}(½)^4(½)^6 = \binom{10}{4}(½)^{10} = 0.205$

2. Con riferimento al problema precedente si supponga che sia n = 5 e che la probabilità del passo in avanti sia p = 0.6. Determinare la distribuzione di probabilità per –5≤k≤5.

Visto che il numero di passi è dispari possiamo affermare che $p_{-4} = p_{-2} = p_0 = p_2 = p_4 = 0$.

$p(Y_5 = -5) = p(S_5 = 0) = \binom{5}{0}(0.6)^0(0.4)^5 = (0.4)^5 = 0.01024$

$p(Y_5 = -3) = p(S_5 = 1) = \binom{5}{1}(0.6)^1(0.4)^4 = (0.6)^5 = 0.0768$

$p(Y_5 = -1) = p(S_5 = 2) = \binom{5}{2}(0.6)^2(0.4)^3 = (0.6)^5 = 0.2304$

e così via.

2.2.5 Analisi del DNA e distribuzione binomiale

Una delle applicazioni delle analisi del DNA è quella sul riconoscimento di paternità. Gli attori sono la madre, il figlio e il presunto padre. In laboratorio vengono preparate sequenze di DNA confrontabili ed esse vengono poi allineate a formare una specie di *codice a barre* formato da n = 30 barre. Una barra del figlio deve essere riconducibile alla madre o al padre salvo il caso di mutazioni genetiche (ad esse si associa una probabilità di 1/300).

Dunque se dalla sequenza del figlio emerge un certo numero di frammenti non presenti tra quelli della madre e del presunto padre si deve concludere, con un certo grado di probabilità, che quest'ultimo non sia il padre.

Se indichiamo con $k = 1, 2, 3$ il numero di frammenti non coincidenti possiamo calcolare la probabilità di mutazioni spontanee in:

$$p_1 = \binom{30}{1}(1/300)^1(299/300)^{29} = 0.0908$$

$$p_2 = \binom{30}{2}(1/300)^2(299/300)^{28} = 0.0440$$

$$p_3 = \binom{30}{3}(1/300)^3(299/300)^{27} = 0.00014$$

$$p_3 = \binom{30}{4}(1/300)^4(299/300)^{26} = 0.000003$$

Dunque mentre la probabilità di una fluttuazione è del 9% la probabilità di 2 è del 4.4% e quella di 3 è solo di 1 parte su 10'000. Insomma se ci sono 3 frammenti fuori posto si tenderà a considerare il presunto padre come non tale.

Si può anche ragionare con un secondo criterio: supporre che il figlio erediti in ugual misura dal padre e dalla madre. Se si costruisce (con Excel) la funzione di ripartizione per n = 30 e si guarda l'andamento al variare di k da 0 a 15 si ha:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
9,31E-10	2,89E-08	4,34E-07	4,22E-06	2,97E-05	0,000162	0,000715	0,002611	0,008062	0,021387	0,049369	0,100244	0,180797	0,292332	0,427768	0,572232

Dunque si può fissare tra 6 e 7 il limite di sequenze comuni al di sotto del quale si può escludere la paternità.

Discutiamo ora il problema inverso; come possiamo garantire affidabilità ad un riconoscimento?

Supponiamo che 16 frammenti siano stati identificati con quelli della madre quante sequenze deve avere coincidenti il padre affinché possa essere riconosciuto tale?

Se si escludono incroci tra consanguinei la probabilità di sequenza coincidente è $(½)^2 = ¼$. Costruiamo la funzione di ripartizione per $n = 14$ e $p = ¼$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0,017818	0,100968	0,281128	0,52134	0,741535	0,888331	0,961729	0,98969	0,997846	0,999658	0,99996	0,999997	1	1	1

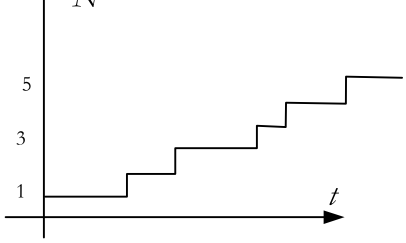
e questa volta concluderemo che se almeno 9 o 10 sequenze sono coincidenti si tratta del padre.

2.3 La distribuzione di Poisson

2.3.1 Introduzione

La distribuzione di Poisson pubblicata da Poisson nel 1837 riguarda i cosiddetti eventi rari, cioè gli eventi che hanno probabilità di accadimento molto basse e che, venendo osservati su tempi lunghi, possono avvenire, un numero finito (ma abbastanza piccolo) di volte.

Questa distribuzione cerca di mettere un po' d'ordine in fenomeni descritti, nel tempo o nello spazio, da una variabile con andamento simile a quello nel diagramma in figura caratterizzato dal fatto che la variabile scatta sempre di una unità alla volta e che l'intervallo tra un evento e l'altro è del tutto imprevedibile.



Fenomeni del genere nelle scienze naturali e sociali hanno a che fare con eventi quali accadimento di reati, code agli sportelli, concentrazione di eventi rari quali sviluppo di microrganismi in zone di una piastra di coltura, distribuzione dell'inquinamento su un'area geografica, decadimento radioattivo.

es. Una notevole applicazione della distribuzione di Poisson si ha nell'articolo di Geiger e Rutherford del 1910 in cui vengono osservate le scintillazioni prodotte dalle particelle alfa su uno schermo fluorescente. Vengono effettuati conteggi su intervalli di durata pari a 1/8 di minuto (7.5 s) e si effettua una statistica in frequenza degli intervalli a seconda del numero di particelle osservate in quell'intervallo.

Si ha così la seguente tabella:

N. α	0	1	2	3	4	5	6	7	8	9	10	11	12+
N. Δt	57	203	383	525	532	408	273	139	45	27	10	4	2

In totale vengono rilevati 2'608 intervalli e vengono conteggiate 10'094 particelle per una media $<\nu> = \frac{10'094}{2'608}$ = 3.87 particelle/intervallo = 0.516 particelle/s.

Osserviamo che la scelta dell'intervallo di osservazione Δt non è libera. L'intervallo deve essere abbastanza piccolo da consentire che non arrivi alcuna particella (ce ne sono Ben 57) e non troppo grande da consentire che oltre un valore di k non troppo elevato non si osservino intervalli favorevoli.

Il valore medio indicato non smentisce il fatto che nonostante una media di 3.87 particelle ogni ottavo di minuto trascorre un totale di 57 intervalli (pari a 427 s non consecutivi) in cui non arriva nulla.

Se nella tabella calcoliamo il rapporto tra il numero di intervalli e il totale degli intervalli (2'608) avremo una nuova riga con le frequenze relative.

k	0	1	2	3	4	5	6	7	8	9	10	11	12+
N_k	57	203	383	525	532	408	273	139	45	27	10	4	2
ν_k	0,021856	0,077837	0,146856	0,201304	0,203988	0,156442	0,104678	0,053298	0,017255	0,010353	0,003834	0,001534	0,000767

Possiamo ritenere che le frequenze relative siano 8identificabili con la probabilità? Possiamo cioè dire che la probabilità che nel tempo di 7.5 s arrivino sullo schermo 5 particelle sia 0.156? A questa domanda risponde la distribuzione di Poisson che ci apprestiamo a descrivere. Essa valuta la probabilità di un *evento raro* quando siano noti alcuni *parametri medi* del fenomeno.

Nella distribuzione di Poisson ragioneremo su fenomeni che avvengono nel tempo ma, al posto della variabile temporale, si può egualmente assumere una variabile spaziale, una superficie o un volume.

2.3.2 Gli assiomi di Poisson

Un fenomeno può essere descritto dalla distribuzione di Poisson quando soddisfa a tre assiomi che vengono esplicitamente utilizzati nella sua dimostrazione. La variabile casuale $N(\Delta t)$ si riferisce all'intervallo temporale Δt e può assumere valori $k = 0, 1, 2, \ldots$. Indicheremo la corrispondente probabilità con $p_k = p(N(\Delta t) = k)$

- Assioma di **indipendenza**: richiede che il fenomeno *non abbia memoria* e cioè che si considerano due intervalli temporali Δt_1 e Δt_2 disgiunti la probabilità dell'evento congiunto sia data dal prodotto delle probabilità: $p(N(\Delta t_1) = k_1 \wedge N(\Delta t_2) = k_2) = p(N(\Delta t_1) = k_1) \cdot p(N(\Delta t_2) = k_2)$.
- Assioma di **omogeneità**: richiede che la scelta della origine nella osservazione del fenomeno non abbia importanza; ovvero il fenomeno non sa *dove avviene l'esperimento* o *che ore sono*: $p(N(\Delta t) = k$ non dipende da t.
- Assioma di **regolarità**: è quello che definisce il concetto di *evento raro*. Afferma che esiste un intervallo temporale δt sufficientemente piccolo per il quale valgano le seguenti proprietà:

  $p(N(\delta t) \geq 2) = 0$

  $p(N(\delta t) = 1) = \alpha\delta t$ le tre richieste definiscono rispettivamente il concetto di *evento raro* (posso

  $p(N(\delta t) = 0) = 1 - \alpha\delta t$

  trovare un δt che elimina la simultaneità), la *intensità del fenomeno* espressa da α che garantisce una crescita lineare della probabilità per piccoli intervalli di tempo, il fatto che 0 è l'evento contrario di 1 (già implicito nella prima richiesta)

La *intensità del fenomeno* quando si opera nel tempo si misura in s^{-1} e il suo significato verrà pienamente compreso solo dopo aver costruito la poissoniana.

2.3.3 Il teorema di Poisson (1837)

$\boxed{T}$ Si consideri una variabile casuale N(t) che può assumere i valori 0, 1, 2, … e che soddisfa i 3 assiomi di indipendenza, omogeneità e regolarità, con intensità α allora si ha:

$$p(N(t) = k) = \frac{(\alpha t)^k}{k!}\, exp(-\alpha t) \tag{2.4}$$

N.B. Nei calcoli con la poissoniana è del tutto indifferente utilizzare come variabile t o Δt in virtù dell'assioma di omogeneità che rende i processi indipendenti dall'origine.

dim La dimostrazione consiste nel far vedere che, grazie ai tre assiomi, la poissoniana si riconduce ad una opportuna bernoulliana.

L'intervallo $[0,t]$ viene diviso in n intervallini $\delta t = \frac{\alpha t}{n}$ ai quali si possa applicare l'assioma di regolarità (cioè per i quali sia definito α e in ciascun intervallino la probabilità di successo (un arrivo) sia $p = \alpha\ \delta t$.

Allora la probabilità che negli $n\ \delta t$ si abbiano k successi segue la distribuzione binomiale

$$p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} (\alpha\ \delta t)^k (1 - \alpha\ \delta t)^{n-k} = \binom{n}{k}\left(\frac{\alpha\ t}{n}\right)^k \left(1 - \frac{\alpha\ t}{n}\right)^{n-k}$$

Poiché δt è un infinitesimo dobbiamo eseguire il passaggio al limite per $n \to \infty$ e per farlo conviene precedentemente scrivere in maniera opportuna la nostra relazione:

$$\binom{n}{k} \cdot \left(\frac{\alpha\ t}{n}\right)^k \cdot \left(1 - \frac{\alpha\ t}{n}\right)^{n-k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k} \cdot \frac{(\alpha\ t)^k}{k!} \cdot \frac{\left(1 - \frac{\alpha\ t}{n}\right)^n}{\left(1 - \frac{\alpha\ t}{n}\right)^k} = (1) \cdot (2) \cdot \frac{(3)}{(4)}$$

(1) $= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \ldots \frac{n-k+1}{n} \to 1 \cdot 1 \cdot 1 \ldots 1 = 1$

(2) è costante

(3) come è noto $\left(1 + \frac{1}{n}\right)^n \to e$ pertanto $\left(1 + \frac{h}{n}\right)^n = \left[\left(1 + \frac{h}{n}\right)^{n/h}\right]^h \to e^h$

ne consegue che $\left(1 - \frac{\alpha\ t}{n}\right)^n \to exp(-\alpha t)$

(4) $\left(1 - \frac{\alpha\ t}{n}\right)^k \to 1^k = 1$ perché k è fissato

Dunque $p(N(t) = k) \to \frac{(\alpha\ t)^k}{k!} \cdot exp(-\alpha t)$

☺

Nei calcoli in cui si utilizza la Poissoniana la quantità $\alpha t = \lambda$ è assegnata una volta per tutte e per questa ragione si parla di distribuzione poissoniana di parametro λ e si scrive in maniera più compatta:

$$p(N(t) = k) = \frac{\lambda^k}{k!}\, exp(-\lambda) \tag{2.5}$$

N.B. La distribuzione di Poisson, oltre che come strumento di analisi degli eventi rari si presta bene a fare da approssimazione della bernoulliana in alcuni dei casi in cui essa sia di difficile calcolo.

Supponiamo di avere un esperimento bernoulliana caratterizzato da un numero n molto elevato e che a sua volta p sia un infinitesimo in maniera che il prodotto dei due sia una quantità finita λ.

Allora se riflettiamo sulla tecnica di dimostrazione utilizzata

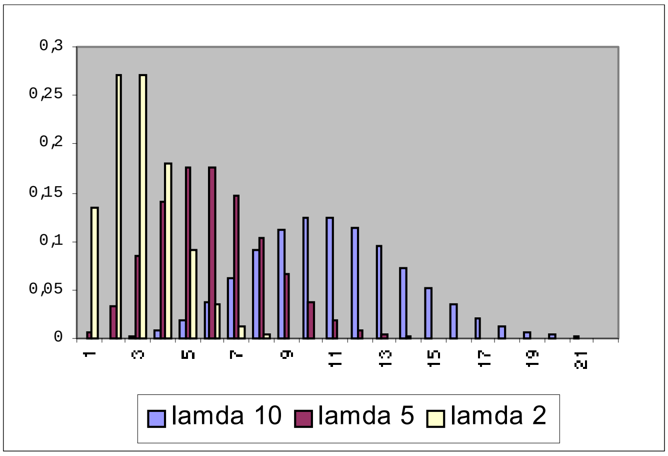
$\lambda = \delta p\ n = \delta p\ t / \delta t = \alpha t$ e dunque la bernoulliana può essere approssimata dalla poissoniana il cui calcolo è molto più agevole.

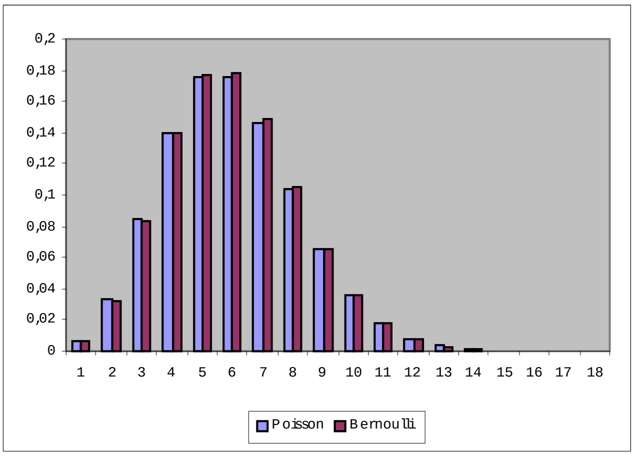
Questa approssimazione è solitamente considerata accettabile per valori di $\lambda \in [0,10]$

2.3.4 La rappresentazione grafica e il confronto con la bernoulliana

Al crescere del valore di λ la distribuzione di Poisson tende ad assumere un andamento abbastanza regolare come si può osservare nei diagrammi seguenti. La curva non è simmetrica (per la presenza del termine moltiplicativo esponenziale).

Nel diagramma a fianco sono state rappresentate la bernoulliana e la poissoniana affiancate, per un valore di $\lambda = 5$ e $n = 200$ (con $p = 0.025$) in modo di farsi una idea visuale della capacità della seconda di dare una buona approssimazione della prima.





2.3.5 Ritorniamo all'esperimento di Rutherford

Assumiamo che sia $\alpha = 0.516$ particelle/s e costruiamo la tabella delle probabilità previste dalla poissoniana per $\lambda = \alpha t = 0.516 \cdot 7.5 = 3.87$ particelle/intervallo.
Otterremo la tabella sottostante nella quale possiamo verificare il buon adattamento delle probabilità alle frequenze relative.

k	0	1	2	3	4	5	6	7	8	9	10	11	12+
ν_k	0,021856	0,077837	0,146856	0,201304	0,203988	0,156442	0,104678	0,053298	0,017255	0,010353	0,003834	0,001534	0,000767
p_k	0,020858	0,080722	0,156197	0,201494	0,194945	0,150888	0,097323	0,053805	0,026028	0,011192	0,004331	0,001524	0,000491

Per ragionare sulle frequenze dei diversi intervalli basta moltiplicare le probabilità per il numero di intervalli osservati (2608) e si ha:

k	0	1	2	3	4	5	6	7	8	9	10	11	12+
N Poiss.	54	211	407	525	508	394	254	140	68	29	11	4	1
N Ruth.	57	203	383	525	532	408	273	139	45	27	10	4	2

Ovvero, i dati sperimentali dell'esperimento di Rutherford si adattano bene ad una poissoniana di parametro $\lambda = \alpha t = 3.87$ particelle/intervallo.
Nel corso del loro esperimento Rutherford e Geiger , consci del fatto che la sorgente di Polonio diminuiva nel tempo la sua attività, introdussero un fattore correttivo nelle misure che consistette nello spostare progressivamente la sorgente verso lo schermo.

2.3.6 I parametri statistici e le proprietà della poissoniana

Stiamo per rispondere alla domanda relativa alla scelta del valore da assegnare alla *intensità del fenomeno*. Il valor medio di una variabile casuale poissoniana è il parametro λ e dunque, in un esperimento concreto si può determinare il valore di α calcolando il numero medio di arrivi.

$\boxed{\text{T.}}$ Se N(t) è una variabile poissoniana di intensità α (o di parametro λ) allora $\mu_{N(t)} = \alpha t = \lambda$ (2.6)

$\boxed{\text{dim}}$ La dimostrazione sfrutta lo sviluppo in serie di Taylor della funzione esponenziale e cioè

$$e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

$$\mu_{N(t)} = \sum_{k=0}^{\infty} k\, p(N(t) = k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} exp(-\lambda) = exp(-\lambda)\, \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = exp(-\lambda)\, \lambda\, exp(\lambda) = \lambda$$

☺

T. Se N(t) è una variabile poissoniana di intensità α (o di parametro λ) allora $\sigma^2_{N(t)} = \lambda$ (2.7)

dim Basta ricordare che per la bernoulliana $\sigma^2 = n\,p\,(1 - p)$ e che in questo caso $n\,p = \lambda$ mentre se $p \to 0 \Rightarrow 1 - p \to 1$
☺

Dunque αt si può trovare calcolando il numero medio della variabile casuale ; ecco perché nel caso dell'esperimento di Rutherford si aveva $\alpha t = \dfrac{10'094}{2'608} = 3.87$ particelle/intervallo

Abbiamo visto che la poissoniana viene spesso utilizzata per approssimare la bernoulliana e questa approssimazione si rivela particolarmente utile per calcolare distribuzioni relative ad intervalli (funzione di ripartizione).

In effetti la funzione di ripartizione non richiede di calcolare ripetutamente la probabilità perché il termine esponenziale è costante e si può mettere in evidenza:

T. $$p(k_1 \le N(t) \le k_1) = exp(-\lambda) \sum_{k=k_1}^{k=k_2} \frac{\lambda^k}{k!} \quad (2.7)$$

dunque in un esercizio da svolgere manualmente conviene calcolare i diversi termini $\dfrac{\lambda^k}{k!}$ e poi procedere ordinatamente alla somma.

2.3.7 Utilizzo della poissoniana

es. La distribuzione degli errori in un testo dattiloscritto si può ritenere con buona approssimazione una variabile poissoniana. Supponendo che in un libro si riscontri la presenza, in media, di 1.4 errori per pagina determinare:

a) la probabilità di trovare 0 errori in una pagina
Si ha $\lambda = 1.4$ e pertanto $\exp(-1.4) = 0.246597$
Dunque $p_0 = 1\cdot 0.246597 = 0.246597$

b) La probabilità di avere due errori
Dunque $p_2 = \dfrac{1.4^2}{2!}\cdot 0.246597 = 0.241665$

c) La probabilità di avere non più di 2 errori per pagina
$F(2) = 0.246597\ (1 + \dfrac{1.4}{1!} + \dfrac{1.4^2}{2!}) = 0.833498$

d) La probabilità di avere più di 4 errori per pagina
$p(N > 4) = 1 - p(N \le 4) = 1 - F(4) = 1 - 0.246597\ (1 + \dfrac{1.4}{1!} + \dfrac{1.4^2}{2!} + \dfrac{1.4^3}{3!} + \dfrac{1.4^4}{4!}) = 1 - 0.946275 = 0{,}053725$

e) La probabilità di avere 6 errori su 3 pagine consecutive
Rispetto ai casi precedenti abbiamo una novità: è cambiato il parametro (non cambia α ma cambia $\lambda = \alpha t$) che diventa ora $\lambda = 3\cdot 1.4 = 4.2$
$p(N = 6) = \dfrac{4.2^6}{6!}\, exp(-4.2) = 7{,}623655\cdot 0{,}014996 = 0.114321$

f) La probabilità che, considerato un insieme di 10 pagine se ne trovino 2 senza errori?
Siamo usciti dal terreno della poissoniana e siamo entrati in quello della bernoulliana. In effetti noi conosciamo la probabilità che una pagina non contenga errori (punto a); essa vale 0.246597.
$p(S_{10} = 2) = \dbinom{10}{2} 0.246597^2(1 - 0.246597)^8 = 0.28406$

g) Qual è la probabilità che considerando 5 pagine consecutive la prima, la terza e la quinta non abbiano errori mentre nelle altre due pagine ci sia un errore per pagina?
Non disponiamo più di eventi descrivibili con le distribuzioni finora considerate e dovremo analizzare l'evento. Indichiamo con X_i la variabile casuale che descrive il numero di errori della pagina i-esima.

Cerchiamo dunque $p(X_1 = 0 \wedge X_2 = 1 \wedge X_3 = 0 \wedge X_4 = 1 \wedge X_5 = 0) = p(X_1 = 0) \cdot p(X_2 = 1) \cdot p(X_3 = 0) \cdot p(X_4 = 1) \cdot p(X_5 = 0)$ dove si è sfruttata l'indipendenza nella distribuzione di errori tra le diverse pagine.
$p_0 = 0.246597$ mentre p_1 deve essere determinata con la poissoniana e risulta 0.345236
Dunque $p = 0.246597^3 0.345236^2 = 0001787$

$\boxed{\text{es.}}$ Calcolare la probabilità che in un gruppo di 1000 presone scelte a caso ci siano più di 5 diabetici sapendo che, in media, soffrono di diabete il 3.5 ‰ degli Italiani. Eseguire il calcolo in forma esatta con la bernoulliana e ripetere poi il calcolo approssimando la bernoulliana con la poissoniana.

$p(S_n \geq 5) = 1 - p(S_n \leq 4) = 1 - [p(S_n = 0) + p(S_n = 1) + p(S_n = 2) + p(S_n = 3) + p(S_n = 4)] = 1 - F(4)$
Nel nostro caso n = 1000 e $p = 0.0035$.
con $p(S_{1000} = k) = \binom{1000}{k} 0.0035^k (0.9965)^{n-k}$
Utilizzando le funzioni di Excel o con la calcolatrice si ottiene:

k	p_k	$F(k)$
0	$3,0013 \cdot 10^{-02}$	$3,0013 \times 10^{-02}$
1	$1,0541 \times 10^{-01}$	$1,3543 \times 10^{-01}$
2	$1,8494 \times 10^{-01}$	$3,2036 \times 10^{-01}$
3	$2,1608 \times 10^{-01}$	$5,3644 \times 10^{-01}$
4	$1,8917 \times 10^{-01}$	$7,2561 \times 10^{-01}$

e dunque $p(S_n \geq 5) = 1 - 0.72561 = 0.27439$
Se si utilizza la poissoniana sarà $\lambda = n\ p = 3.5$
$p_k = \frac{3.5^k}{k!} exp(-3.5)$ e in questo caso si ottiene

k	p_k	$F(k)$
0	0,030197383	0,030197383
1	0,105690842	0,135888225
2	0,184958973	0,320847199
3	0,215785469	0,536632668
4	0,188812285	0,725444953

e dunque $p(S_n \geq 5) = 1 - 0.725445 = 0.274555$
e l'errore incide sulla quarta cifra decimale.

2.4 Le distribuzioni in Excel

2.4.1 La binomiale

La funzione ha come sintassi
=distrib.binom(k;n;p;logico) (2.8)
dove k rappresenta l'argomento della funzione, n è il numero di prove, p la probabilità dell'evento favorevole e logico un campo che può essere impostato a falso se si vuole la distribuzione di probabilità o a vero se si vuole la funzione di ripartizione.

2.4.2 La poissoniana

La funzione ha come sintassi
=poisson(k;λ;logico) (2.9)
dove k rappresenta l'argomento della funzione, λ è il parametro e logico un campo che può essere impostato a falso se si vuole la distribuzione di probabilità o a vero se si vuole la funzione di ripartizione.

3. Disuguaglianze notevoli

3.1 La disuguaglianza di Cebyçev (1867)

3.1.1 Il teorema

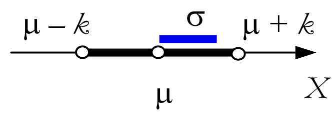
La disuguaglianza di Cebyçev [3] serve a stabilire la probabilità che una variabile casuale si allontani dal suo valore medio misurando tale distanza in unità di deviazione standard.

T. Considerata una variabile casuale X di valore medio μ e varianza σ^2 valgono sempre le due relazioni seguenti:

$$p(|X-\mu| \geq k) \leq \left(\frac{\sigma}{k}\right)^2 \qquad p(|X-\mu| \leq k) \geq 1-\left(\frac{\sigma}{k}\right)^2 \tag{3.1}$$

N.B. Si tratta di una disuguaglianza a *maglie larghe* che si utilizza solo quando non si hanno informazioni sul tipo di distribuzione della variabile casuale ma è comunque noto σ.

Ovviamente la disuguaglianza risulta interessante solo quando $\frac{\sigma}{k} < 1$ cioè quando $k > \sigma$ perché negli altri casi afferma che p è minore di un numero maggiore di 1 e questo, come è noto, è sempre vero, oltre che banale.



Nella prima forma essa stabilisce un limite superiore alla probabilità per l'ipotesi che la variabile casuale si allontani troppo dal valor medio e ci dice che, al crescere di k tale probabilità è minore di una quantità che tende a zero al crescere di k.

Nella seconda forma viene preso in esame l'evento contrario e si ha ovviamente una maggiorazione.

dim. La dimostrazione è di tipo diretto e si basa sulla definizione di varianza e sulla possibilità, dopo aver ordinato la variabile, di dividere gli scarti assoluti in due categorie, quelli minori e quelli maggiori o eguali a k. Indichiamo con A e con B i due insiemi.

$$\sigma^2 = \sum_{i=1}^{n}(x_i-\mu)^2 p_i = \sum_{A}(x_i-\mu)^2 p_i + \sum_{B}(x_i-\mu)^2 p_i \geq \sum_{B}(x_i-\mu)^2 p_i > \sum_{B} k^2 p_i = k^2\sum_{B} p_i = k^2 p(|X-\mu| \geq k)$$

e dunque $p(|X-\mu| \geq k) \leq \left(\frac{\sigma}{k}\right)^2$

Basta considerare l'evento contrario per ottenere l'altra disuguaglianza.

☺

Dalla disuguaglianza di Cebyçev possiamo estrarre delle limitazioni generali sulla dispersione di X intorno a μ se misuriamo tali dispersioni in unità di σ.

Per $k = 2\sigma$ si ha $p \geq 1 - ¼ = 0.75$

Per $k = 3\sigma$ si ha $p \geq 1 - 1/9 = 0.89$

3.1.2 Esempio di applicazione

Consideriamo un esempio per il quale, attraverso la funzione di ripartizione, siamo in grado di valutare la probabilità entro un intervallo in modo di cogliere l'*aspetto debole* della disuguaglianza di Cebyçev.

Consideriamo un esperimento con n = 1000 lanci di una moneta equa (p = 0.5) e confrontiamo la previsione della disuguaglianza di Cebyçev con la valutazione che ci fornisce la funzione di ripartizione della bernoulliana per l'intervallo di ampiezza 30 intorno al valore medio.

Il valore medio $\mu = n\,p = 1000 \cdot 0.5 = 500$ mentre $\sigma^2 = 1000 \cdot ½ \cdot ½ = 250$

$$p\,(|S_n - 500|) \leq 30) \geq 1 - \frac{250}{900} = 0.722$$

Ma dalla conoscenza della bernoulliana si ha:

[3] Pahnuty Lvovich Chebyshev russo, nato nel 1821 e morto nel 1894, il cui cognome è ovviamente scritto in cirillico; qui è scritto secondo la pronuncia inglese, in altri testi è scritto Tchebycheff in tedesco e Cebycev secondo la pronuncia francese

$$p\,(|S_n - 500|) \le 30) = \sum_{i=470}^{530} \binom{1000}{i} 0.5^i\, 0.5^{1000-i} \text{ e cioè } F(530) - F(469)$$

Si tratta di numeri che non si riescono a maneggiare con le calcolatrici ed impareremo ad approssimarli con le distribuzioni continue ma Excel può venire in nostro aiuto ed in effetti: F(530) – F(469) = 0.973161075 – 0.026838925 = 0.946322

Come si vede la disuguaglianza di Cebyçev con la sua previsione di una probabilità maggiore di 0.722 ci ha detto molto poco rispetto alla situazione reale visto che la probabilità è del 95%. Il risultato ci consente di riflettere sulla origine di questo limite. Essa va cercata direttamente nella dimostrazione nella quale abbiamo trascurato tutti i contributi che cadevano all'esterno ed abbiamo inoltre eseguito una minorazione con la quale tutti gli scarti venivano eguagliati a *k*.

3.2 La legge dei grandi numeri (legge empirica del caso) (1713)

3.2.1 Il teorema

E' detta anche disuguaglianza di Bernoulli (Giacomo) ed è di molto precedente alla disuguaglianza di Cebyçev anche se noi la dimostreremo da essa utilizzando proprietà generali della varianza.

T. Si considerino *n variabili casuali indipendenti* $X_1, X_2, \ldots, X_n$ tutte dotate dello stesso valor medio μ e della stessa varianza σ^2. Si indichi con $S_n = \sum_{i=1}^{n} X_i$ una nuova variabile casuale. Allora anche la media aritmetica $Y_n = \frac{S_n}{n}$ è una variabile casuale.

$$p\left(\left|\frac{S_n}{n} - \mu\right| \ge k\right) \le \frac{1}{n}\left(\frac{\sigma}{k}\right)^2 \quad \text{o anche } p\left(\left|\frac{S_n}{n} - \mu\right| < k\right) \ge 1 - \frac{1}{n}\left(\frac{\sigma}{k}\right)^2 \qquad (3.2)$$

N.B. $\frac{S_n}{n}$ può essere pensata come la media aritmetica dei risultati di un esperimento. Il teorema dice che per $n \to\infty$ la probabilità che la media aritmetica differisca dal valor medio di una quantità prefissata tende comunque a zero e dunque ci autorizza ad utilizzare il principio della ripetizione delle misure per inferire il *valore vero* di una grandezza sconosciuta.

dim. Per le proprietà di linearità della media si ha: $\mu_{Y_n} = \frac{1}{n} n\mu = \mu$

Per le proprietà della varianza $\sigma^2_{Y_n} = \frac{1}{n^2}\sigma^2_{S_n} = \frac{1}{n^2} n\,\sigma^2 = \frac{\sigma^2}{n}$

Basta ora applicare alla variabile Y_n la disuguaglianza di Cebyçev e la (3.2) è dimostrata.

☺

La legge dei grandi numeri sarà utilizzata in statistica inferenziale quando la variabile Y_n verrà a corrispondere alla cosiddetta *media campionaria*.

3.2.2 Una applicazione

es. Determinare il numero *n* di lanci di una moneta equa in modo che la probabilità che la frequenza relativa di teste differisca da ½ per meno di 1/100 sia maggiore di 0.9.

Le variabili X_i hanno tutte valore 0 o 1 con probabilità ½ e pertanto $\mu = ½$ e $\sigma^2 = (0 - ½)^2 ½ + (1 - ½)^2 ½ = ¼$ e $\sigma = ½$

Per la legge dei grandi numeri $p\left(|\nu_n - \mu| < 0.01\right) \ge 1 - \frac{1}{n}\left(\frac{½}{0.01}\right)^2$ e dunque dovrà essere

$1 - \frac{1}{n}\cdot 50^2 = 0.9 \Leftrightarrow \frac{2500}{n} = 0.1 \Leftrightarrow n = 25'000$. La probabilità è maggiore di 0.9 se $n \ge 25'000$.

Attenzione: la previsione è di tipo probabilistico. Ciò significa che la frequenza relativa può ancora essere maggiore di 0.51 o minore di 0.49 ma ciò accadrà con probabilità inferiore a 0.9.