

# Manuale dell'ingegnere intrippato con la statistica

What are the odds?

30 agosto 2019

## 1 Statistica descrittiva

### 1.1 Le grandezze che sintetizzano i dati

#### 1.1.1 Media

Dato un insieme  $x_1, x_2, \dots, x_n$  di dati, si dice media campionaria la media aritmetica di questi valori.

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$$

#### 1.1.2 Mediana

Dato un insieme di dati di ampiezza  $n$ , lo si ordina dal minore al maggiore. La mediana è il valore che occupa la posizione  $\frac{n+1}{2}$  in caso di un insieme dispari, o la media tra  $\frac{n}{2}$  e  $\frac{n}{2} + 1$  se pari.

#### 1.1.3 Moda

La moda campionaria di un insieme di dati, se esiste, è l'unico valore che ha frequenza massima.

#### 1.1.4 Varianza e deviazione standard campionarie

Dato un insieme di dati  $x_1, x_2, \dots, x_n$ , si dice varianza campionaria ( $s^2$ ), la quantità

$$s^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Una comodità per il calcolo è che

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

Si dice **deviazione standard campionaria** e si denota con  $s$ , la quantità

$$s := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

(la radice quadrata di  $s^2$ )

### 1.1.5 Percentili campionari e box plot

Sia  $k$  un numero intero  $0 \leq k \leq 100$ . Dato un campione di dati, esiste sempre un dato che è contemporaneamente maggiore del  $k$  percento dei dati, e minore del  $100 - k$  percento. Per trovare questo dato, dati  $n$  e  $p = \frac{k}{100}$ :

1. Disponiamo i dati in ordine crescente
2. Calcoliamo  $np$
3. Il numero cercato è quello in posizione  $np$ , arrotondato per eccesso se non intero.

Il 25-esimo percentile si dice *primo quartile*, il 50-esimo *secondo* (ed è pari alla mediana), il 75-esimo *terzo*. Il box plot è un grafica con un quadrato sulla linea dei dati, con i lati sul primo e terzo quartile, e un segno sul secondo.

### 1.2 Disuguaglianza di Chebyshev

Siano  $\bar{x}$  e  $s$  media e deviazione standard campionarie di un insieme di dati. Nell'ipotesi che  $s > 0$ , la disuguaglianza di Chebyshev afferma che per ogni reale  $k \geq 1$ , almeno una frazione  $(1 - 1/k^2)$  dei dati cade nell'intervallo che va da  $\bar{x} - ks$  a  $\bar{x} + ks$ . Usando il ~~pessimo~~ *fantastico* linguaggio da statista: sia assegnato un insieme di dati  $x_1, \dots, x_n$  con media campionaria  $\bar{x}$  e deviazione standard campionaria  $s > 0$ . Denotiamo con  $S_k$  l'insieme degli indici corrispondenti a dati compresi tra  $\bar{x} - ks$  e  $\bar{x} + ks$ . Sia  $\#S_k$  il numero dei suddetti. Allora abbiamo che

$$\frac{\#S_k}{n} \geq 1 - \frac{n-1}{nk^2} > 1 - \frac{1}{k^2}$$

### 1.3 Insiemi di dati bivariati e coefficiente di correlazione campionaria

A volte non abbiamo a che fare con dati singoli, ma con coppie di numeri, tra i quali sospettiamo l'esistenza di relazioni. Dati di questa forma prendono il nome di *campione bivariato*. Uno strumento utile è il diagramma di dispersione. Una questione interessante è capire se vi sia correlazione tra i dati accoppiati. Parleremo di correlazione positiva quando abbiamo una proporzionalità diretta tra i due, di correlazione negativa quando abbiamo una proporzionalità inversa.

#### 1.3.1 Coefficiente di correlazione campionaria

Dato un campione bivariato  $(x_i, y_i)$ , sono definite le medie  $\bar{x}$  e  $\bar{y}$ . Possiamo senz'altro dire che se un valore  $x_i$  è grande rispetto alla media, la differenza  $x_i - \bar{x}$  sarà positiva, mentre se  $x_i$  è piccolo, la differenza sarà negativa. Quindi, considerando il prodotto  $(x_i - \bar{x})(y_i - \bar{y})$ , sarà positivo per correlazioni positive, negativo per correlazioni negative. Se l'intero campione mostra quindi un'elevata correlazione, ci aspettiamo che la somma di tutti i prodotti  $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  darà una buona stima della correlazione. Normalizziamola dividendo per  $(n-1)$  e per il prodotto delle deviazioni standard campionarie, e otteniamo il **coefficiente di correlazione campionaria**

$$r := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

con  $s_x$  e  $s_y$  deviazioni standard campionarie di  $x$  e  $y$ .

### 1.3.2 Proprietà del coefficiente di correlazione campionaria

Sebbene parleremo meglio di questo bastardo nella sezione sulla regressione, elenchiamo qui alcune proprietà:

1.  $-1 \leq r \leq 1$
2. Se per opportune costanti  $a$  e  $b$ , con  $b > 0$  sussiste la relazione lineare  $y_i = a + b x_i$ , allora  $r = 1$ .
3. Se per opportune costanti  $a$  e  $b$ , con  $b < 0$  sussiste la relazione lineare  $y_i = a + b x_i$ , allora  $r = -1$ .
4. Se  $r$  è il coefficiente di correlazione del campione  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , allora lo è anche per il campione  $(a + b x_i, c + d y_i)$ , purché le costanti  $a$  e  $b$  abbiano lo stesso segno.

## 2 Elementi di probabilità

### 2.1 Spazio degli esiti ed eventi

Si dice spazio degli esiti l'insieme di tutti gli esiti possibili di un esperimento. Se ad esempio l'esito dell'esperimento fosse il sesso di un neonato, lo spazio degli esiti sarebbe

$$S = \{f, m\}$$

I sottoinsiemi dello spazio degli esiti si dicono **eventi**, quindi un evento  $E$  è un insieme i cui elementi sono esiti possibili. Si dice  $E^c$  l'opposto dell'evento, quindi  $P(E^c) = 1 - P(E)$ . Risulta ovvio che  $1 = P(E^c) + P(E)$ . Se abbiamo due eventi qualsiasi, la loro unione  $P(E \cup F) = P(E) + P(F) - P(E \cap F)$ .

### 2.2 Spazi di esiti equiprobabili

Per tanti esperimenti è naturale assumere che ognuno degli esiti abbia la stessa probabilità di accadere. Abbiamo quindi che la probabilità che  $E$  accada è pari a  $P(E) = \frac{\#E}{N}$ .

#### 2.2.1 Principio di enumerazione

Consideriamo la realizzazione di due diversi esperimenti che possono avere rispettivamente  $m$  ed  $n$  esiti. Allora complessivamente avremo  $mn$  risultati.

### 2.3 Coefficiente binomiale

Vogliamo ora determinare il numero di diversi gruppi di  $r$  oggetti che si possono formare scegliendoli da un insieme di  $n$ . Ad esempio, quanti gruppi di 3 lettere possono formarsi dal gruppo  $\{A, B, C, D, E\}$ . In generale, poiché il numero di modi diversi di scegliere  $r$  oggetti su  $n$  tenendo conto dell'ordine è dato da  $n(n-1)\dots(n-r+1)$ , e poiché ogni gruppo di lettere viene contato  $r!$  volte (uno per permutazione), il numero di gruppi di  $r$  elementi su  $n$  totali è dato da

$$\frac{n(n-1)\dots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}$$

## 2.4 Probabilità condizionata

Vogliamo ora calcolare la probabilità che un evento accada, appurato che ne è accaduto un altro. Ad esempio, lanciamo due dadi. L'evento E cercato è che il risultato sia 8. L'evento F già accaduto è che il primo dato risulta in un 3. Si dice probabilità condizionata di E dato F

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

## 2.5 Fattorizzazione di un evento e formula di Bayes

Siano E ed F due eventi qualsiasi. È possibile esprimere E come

$$P(E) = P(E \cap F) + P(E \cap F^c)$$

Visto inoltre che i due sono eventi disgiunti, si ha che

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$
$$P(E|F)P(F) + P(E|F^c)[1 - P(F)]$$

Questa *orribile* equazione, ci mostra che la probabilità dell'evento E si può ricavare come media pesata delle probabilità condizionali di E sapendo che: F si è verificato e non si è verificato. I pesi sono ovviamente le probabilità degli eventi a cui si condiziona.

## 2.6 Eventi indipendenti

Due eventi si dicono indipendenti quando il risultato di uno non influenza l'altro. In altre parole, significa che avendo due eventi E ed F, se so che F è accaduto, la probabilità che accada E non cambia.

$$P(E \cap F) = P(E)P(F)$$

## 3 Variabili aleatorie e valore atteso

Quando realizziamo un esperimento casuale, non sempre siamo interessati a tutti i risultati del suddetto. Se ad esempio lanciassimo due dadi, potrebbe interessarci la sola somma e non i singoli risultati. Queste quantità di interesse sono dette **variabili aleatorie**. Siccome il valore di questa variabile è dato dal risultato dell'esperimento, possiamo assegnare delle probabilità a queste. Queste variabili aleatorie hanno una *funzione indicatrice* definita, ad esempio, così:

$$I := \begin{cases} 1 & \text{se } X = 1 \text{ o } 2 \\ 0 & \text{se } X = 0 \end{cases}$$

Variabili aleatorie con un numero finito o numerabile di valori possibili sono dette **discrete**. Esistono anche variabili aleatorie **continue**.

### 3.0.1 Funzione di ripartizione

La funzione di ripartizione F di una variabile aleatoria X, è definita, per ogni numero reale x, tramite

$$F(x) := P(X \leq x)$$

Quindi  $F(x)$  esprime la probabilità che la variabile aleatoria X assuma un valore *minore o uguale* a x. Tutte le questioni di probabilità che si possano sollevare su una variabile aleatoria, ammettono una risposta in termini della sua funzione di ripartizione.

### 3.1 Variabili aleatorie discrete e continue

Se  $X$  è una variabile aleatoria discreta, la sua funzione di massa di probabilità, o funzione di massa, si definisce nel modo seguente:

$$p(a) := P(X = a)$$

La funzione  $p(a)$  è non nulla su un insieme al più numerabile di valori. Infatti, se  $x_1, x_2, \dots, x_n$  sono i possibili valori di  $X$ , allora

$$\begin{aligned} p(x_i) &> 0 \quad i = 1, 2, \dots \\ p(x) &= 0 \quad \text{tutti gli altri valori di } x \end{aligned}$$

Siccome  $X$  deve assumere i suddetti valori, necessariamente deve essere vero che

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

Una variabile aleatoria che possa assumere un'infinità non numerabile di valori, non potrà essere discreta. Si dirà **continua** se esiste una funzione non negativa  $f$ , definita su tutto  $\mathbb{R}$ , avente la proprietà che per ogni insieme  $B$  di numeri reali,

$$P(X \in B) = \int_B f(x) dx$$

Questa funzione è detta **funzione di densità di probabilità**. L'equazione dice che la probabilità che una variabile aleatoria continua  $X$  appartenga a un insieme  $B$  si può trovare integrando la sua densità su tale insieme. Pare ovvio che

$$1 = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x) dx$$

Tutte le probabilità che riguardano una variabile aleatoria continua possono essere espresse in funzione della sua densità di probabilità:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Se poniamo  $b = a$ , notiamo che la probabilità che una variabile aleatoria continua assuma un valore particolare  $a$  è nulla:

$$P(X = a) = \int_a^a f(x) dx = 0$$

Leghiamo la funzione di ripartizione  $F$  alla densità  $f$  così:

$$F(a) := P(X \in (-\infty, a]) = \int_{-\infty}^a f(x) dx$$

Derivando entrambi otteniamo la relazione fondamentale:

$$\frac{d}{da} F(a) = f(a)$$

La densità è quindi la derivata della funzione di ripartizione. Notiamo che quando conosciamo la funzione di massa di probabilità di una variabile aleatoria discreta, o la funzione di densità di probabilità di una continua, abbiamo abbastanza informazioni per poter calcolare le probabilità di ogni evento che dipenda dalla sola variabile aleatoria.

## 3.2 Coppie e vettori di variabili aleatorie

Ci sono situazioni in cui abbiamo necessità di studiare le **relazioni** tra variabili aleatorie multiple. Per specificare la relazione tra due variabili aleatorie  $X$  e  $Y$ , il primo passo è estendere il concetto di funzione di ripartizione. Siano quindi  $X$  e  $Y$  due variabili aleatorie che riguardano lo stesso esperimento casuale. Si dice *funzione di ripartizione congiunta* di  $X$  e  $Y$  la funzione di due variabili seguente:

$$F(x, y) := P(X \leq x, Y \leq y)$$

dove la virgola denota l'intersezione tra gli eventi. La conoscenza di questa funzione permette, almeno in teoria, di calcolare le probabilità di tutti gli eventi che dipendono, singolarmente o congiuntamente, da  $X$  e  $Y$ .

### 3.2.1 Distribuzione congiunta per variabili aleatorie discrete

Se sappiamo che un vettore aleatorio è di tipo discreto, possiamo definire e utilizzare la funzione di massa di probabilità. Se  $X$  e  $Y$  sono variabili aleatorie discrete che assumono i valori  $x_1, x_2, \dots$  e  $y_1, y_2, \dots$ , la funzione

$$p(x_i, y_j) := P(X = x_i, Y = y_j), \quad i = 1, 2, \dots \quad j = 1, 2, \dots$$

è la loro funzione di massa di probabilità congiunta. Le funzioni di massa individuali si possono ricavare da questa, notando che, siccome  $Y$  deve assumere uno dei valori  $y_j$ , l'evento  $\{X = x_i\}$ , può essere visto come l'unione al variare di  $j$  degli eventi  $\{X = x_i, Y = y_j\}$ , che sono mutuamente esclusivi. Da qui:

$$p_X(x_i) := P(X = x_i) = \sum_j p(x_i, y_j)$$

Anche se le individuali possono essere ricavate dalla congiunta, la congiunta non può essere ricavata dalle condizionali.

### 3.2.2 Distribuzione congiunta per variabili aleatorie continue

Due variabili aleatorie  $X$  e  $Y$  sono congiuntamente continue se esiste una funzione non negativa  $f(x, y)$  tale che, per ogni sottoinsieme  $C$  del piano cartesiano,

$$P((X, Y) \in C) = \int \int_{(x, y) \in C} f(x, y) dx dy$$

questa è detta **densità congiunta** delle variabili aleatorie  $X$  e  $Y$ . Otteniamo inoltre che

$$P(X \in A, Y \in B) = \int_B \int_A f(x, y) dx dy$$

E, in conclusione,

$$P(X \in A) = \int_A f_X(x) dx$$

Per ricavare le individuali, otteniamo

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \end{aligned}$$

### 3.2.3 Variabili aleatorie indipendenti

Due variabili aleatorie sono indipendenti se tutti gli eventi relativi alla prima sono indipendenti da tutti quelli relativi alla seconda. La definizione è che se, per ogni coppia di insiemi di numeri reali  $A$  e  $B$  è soddisfatta

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

le due V.A. sono indipendenti. Se le V.A. sono discrete, l'equazione equivale a dire che la funzione di massa congiunta è il prodotto delle marginali:

$$p(x, y) = p_X(x)p_Y(y)$$

Possiamo generalizzare le osservazioni suddette anche per vettori di variabili aleatorie. *Lo faremo? Non credo proprio.*

### 3.2.4 Distribuzioni condizionali

Le relazioni esistenti tra due variabili aleatorie possono essere chiarite dallo studio della distribuzione condizionale di una delle due, dato il valore dell'altra. Si ricorda che che presi comunque due eventi  $E$  e  $F$  con  $P(F) > 0$ , la probabilità di  $E$  condizionata a  $F$  è data da

$$P(E|F) := \frac{P(E \cap F)}{P(F)}$$

è naturale applicare questo schema anche alle variabili aleatorie discrete. Siano  $X$  e  $Y$  due variabili aleatorie discrete con funzione di massa congiunta  $p(\cdot, \cdot)$ , diciamo funzione di massa di probabilità condizionata di  $X$  dato  $Y$  e si indica con  $p_{X|Y}(\cdot|\cdot)$ , la funzione di due variabili così definita:

$$p_{X|Y}(x|y) := P(X = x|Y = y) \\ \frac{p(x, y)}{p_Y(y)}, \quad \forall x \forall y \text{ con } p_Y(y) > 0$$

Se  $y$  non è un valore possibile di  $Y$ , ovvero se  $P(Y = y) = 0$ , la quantità  $p_{X|Y}(x|y)$  non è definita.

### 3.3 Valore atteso

Uno dei concetti più importanti di tutta la teoria della probabilità (*ziocane*) è quello di valore atteso. Esso è definito come il numero

$$E[X] := \sum_i x_i P(X = x_i)$$

In altri termini, si tratta della media pesata dei valori possibili di  $X$ , usando come pesi le probabilità che vengano assunti. È ovvio che nel caso di V.A. continue, il giochino non funziona. Definiamo quindi il valore atteso di una V.A. continua con funzione di densità  $f$ , come

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

### 3.4 Proprietà del valore atteso

Consideriamo una V.A. di cui conosciamo la distribuzione. *What if*, se anziché calcolare il valore atteso di  $X$ , volessimo calcolare quello di una funzione  $g(X)$ ? Notiamo che  $g(X)$  è comunque una variabile aleatoria. Ricaviamo quindi la sua distribuzione, e ne calcoliamo il valore atteso. Ponendo le cose in maniera rigorosa:

$$E[g(X)] = \sum_x g(x)p(x) \text{ per V.A. discrete}$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \text{ per V.A. continue}$$

Per ogni coppia di costanti reali  $a$  e  $b$ , abbiamo anche che

$$E[aX + b] = aE[X] + b \quad \text{e quindi} \quad E[aX] = aE[X]$$

#### 3.4.1 Valore atteso della somma di variabili aleatorie

Con *complessi calcoli tendenzialmente inutili*, otteniamo che

$$E[X + Y] = E[X] + E[Y]$$

Tale risultato vale sia nel caso discreto, che in quello continuo.

### 3.5 Varianza

A volte, conoscere la media di una distribuzione non basta. *Se inserissimo l'autore di questi riassunti con la testa in un freezer e i piedi in un forno, la temperatura media sarebbe abbastanza ok, l'autore no.* Per questo, è utile conoscere quanto i valori si allontanano dalla media. Questo è proprio il compito della **varianza**. Sia  $X$  una variabile aleatoria con media  $\mu$ , la varianza di  $x$ , che denotiamo con  $Var(X)$  è la quantità

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

o, in alternativa (*questa è molto più comoda*)

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

### 3.6 La covarianza e la varianza della somma di V.A.

Come sappiamo, la media della somma di V.A. coincide con la somma delle loro medie. Per la varianza, in generale, questo non è vero. **In un caso sì: quando le V.A. sono indipendenti.** Prima di tutto, però, definiamo il concetto di Covarianza: date due V.A.  $X$  e  $Y$  di media  $\mu_X$  e  $\mu_Y$ , essa vale

$$Cov(X, Y) := E[(X - \mu_X)(Y - \mu_Y)]$$

o, in alternativa

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

Derivano anche alcune semplici proprietà

$$Cov(X, Y) = Cov(Y, X)$$
$$Cov(X, X) = Var(X)$$
$$Cov(aX, Y) = aCov(X, Y) = Cov(X, aY)$$



E se avessimo 3 V.A.?

$$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$$

Inoltre, generalizzando i concetti, se avessimo  $n$  V.A.  $X_1, \dots, X_n$  e  $n$   $Y_1, \dots, Y_n$

$$\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$$

### 3.6.1 Variabili aleatorie indipendenti

Se abbiamo due V.A.  $X$  e  $Y$  indipendenti, sappiamo che

$$E[XY] = E[X]E[Y]$$

Questo implica inoltre che

$$\text{Cov}(X, Y) = 0$$

e quindi, se abbiamo  $n$  V.A., la varianza della somma è la somma delle varianze.

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

## 4 La funzione generatrice dei momenti

La *funzione generatrice dei momenti*, o più semplicemente, la funzione generatrice  $\phi$  di una V.A.  $X$ , è definita, per tutti i  $t$  reali per i quali il valore atteso di  $e^{tX}$  ha senso, dall'espressione

$$\phi(t) := E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & \text{se } X \text{ è discreta} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{se } X \text{ è continua} \end{cases}$$

Il nome deriva dal fatto che tutti i momenti di cui è dotata  $X$  possono essere ottenuti derivando più volte nell'origine la funzione  $\phi(t)$ . Ad esempio,

$$\phi'(t) = \frac{d}{dt} E[e^{tX}] = E[Xe^{tX}]$$

Quindi,  $\phi'(0) = E[X]$ , e, più in generale,  $\phi^n(0) = E[X^n]$ . Se  $X$  e  $Y$  sono variabili indipendenti con funzioni generatrici  $\phi_X$  e  $\phi_Y$ , e se  $\phi_{X+Y}$  è la funzione generatrice dei momenti di  $X + Y$ , allora

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

Un'osservazione interessante sulla generatrice dei momenti, è che essa *determina la distribuzione*, ossia se due V.A. hanno identica generatrice, hanno identica legge (quindi funzione di ripartizione e funzione di massa).

## 5 La legge debole dei grandi numeri

Per introdurre la suddetta, prima enunciamo la **disuguaglianza di Markov**: se  $X$  è una variabile aleatoria che non è mai negativa. allora per ogni  $a > 0$

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Come corollario, ricaviamo la disuguaglianza di Chebyshev: data una V.A.  $X$  con media  $\mu$  e varianza  $\sigma^2$ , allora per ogni  $r > 0$

$$P(|X - \mu| \geq r) \leq \frac{\sigma^2}{r^2}$$

Otteniamo infine la **legge debole dei grandi numeri**. *No, non quella con cui giustificate il vostro provarci con ogni essere vivente femminile.* Sia  $X_1, X_2, \dots$  una successione di variabili aleatorie i.i.d (*indipendenti, identicamente distribuite*), tutte con media  $\mu$ . Allora, per ogni  $\epsilon > 0$ :

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \epsilon\right) \rightarrow 0 \quad \text{quand } n \rightarrow \infty$$

*Ora, vi chiederete: cosa me ne faccio? Posso spiegarla alle tipe in discoteca?* **No.** Una applicazione interessante è la seguente: supponiamo di ripetere in successione molte copie indipendenti di un esperimento, in ciascuna delle quali può verificarsi un certo evento  $E$ :

$$X_i := \begin{cases} 1 & \text{se } E \text{ si realizza nell'esperimento } i\text{-esimo} \\ 0 & \text{se } E \text{ non si realizza} \end{cases}$$

la sommatoria  $X_1 + \dots + X_n$  rappresenta il numero di prove - tra le prime  $n$  - in cui si è verificato l'evento  $E$ . Poiché

$$E[X_i] = P(X_i = 1) = P(E)$$

si deduce che la frazione delle  $n$  prove nelle quali si realizza  $E$ , tende alla probabilità  $P(E)$ .