3 Un nuovo modello

3.1 Il modello

Costruzione del modello [9]

L'idea
Passare al limite per $\boldsymbol{n \to \infty}$ nella distribuzione binomiale, nell'ipotesi che $\boldsymbol{\lambda} = np$ sia **costante**.
Osserviamo che ciò significa passare al limite per $p \to 0$.

3.1.1 Le ipotesi del modello

Iniziamo esplicitando le ipotesi che è opportuno assumere affinché abbia senso schematizzare mediante il "nuovo" modello il problema generale di conteggio proposto nel paragrafo 1.2.

1. Una data realizzazione[a] dell'evento ha la **stessa probabilità** di verificarsi in ogni intervallo di uguale ampiezza.
2. Ogni realizzazione non è influenzata dalle altre, ovvero le realizzazioni sono **indipendenti**.
3. Il numero medio di realizzazioni λ è "piccolo" rispetto al numero delle prove (evento "**raro**").

[a] Nell'esempio del centralino, già analizzato, tale evento è l'arrivo di una telefonata.

Possiamo sintetizzare le ipotesi 1. e 2. dicendo che il processo è "**casuale**" e **uniforme**[10].

3.1.2 La nuova distribuzione

Si può dimostrare[11] che passando al limite nel modo indicato si ottiene una "nuova" variabile aleatoria X che conta, come la variabile aleatoria S, il numero di realizzazioni dell'evento nell'intervallo e una distribuzione di probabilità per X. Precisamente, la **distribuzione di probabilità** di X si può ottenere *per passi* in questo modo:

- si inizia da
$$P(X = 0) = e^{-\lambda}$$
- per ogni naturale $k \geq 1$, il valore $P(X = k)$ si ottiene da $P(X = k - 1)$ moltiplicando per $\frac{\lambda}{k}$, ovvero
$$P(X = k) = \frac{\lambda}{k} \cdot P(X = k - 1)$$

[9] Preciseremo il significato dell'affermazione all'interno del riquadro "L'idea" nel paragrafo 3.2.
[10] Utilizzeremo più volte nel seguito questi due termini e nel farlo intenderemo riferirci proprio alle ipotesi 1. e 2.
[11] Lo faremo, mediante un'attività guidata, nel paragrafo 3.2.

Pertanto se si vuole esplicitare l'espressione analitica di $P(X = k)$ basta applicare più volte l'ultima uguaglianza a partire dal caso $k = 0$:

$$
\begin{array}{ll}
 & P(X = 0) = e^{-\lambda} \\
\boxed{\cdot\frac{\lambda}{1}} & P(X = 1) = \boxed{\frac{\lambda}{1}} \cdot P(X = 0) = \lambda \cdot e^{-\lambda} \\
\boxed{\cdot\frac{\lambda}{2}} & P(X = 2) = \boxed{\frac{\lambda}{2}} \cdot P(X = 1) = \frac{\lambda}{2} \cdot \lambda \cdot e^{-\lambda} = \frac{\lambda^2}{2} \cdot e^{-\lambda} \\
\boxed{\cdot\frac{\lambda}{3}} & P(X = 3) = \boxed{\frac{\lambda}{3}} \cdot P(X = 2) = \frac{\lambda}{3} \cdot \frac{\lambda^2}{2} \cdot e^{-\lambda} = \frac{\lambda^3}{3!} \cdot e^{-\lambda} \\
 & \vdots
\end{array}
$$

In generale, quindi, si ottiene:

$$\boldsymbol{P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}}$$

Si può dimostrare formalmente che questa è una distribuzione di probabilità; essa prende il nome di **distribuzione di Poisson**.

3.1.3 Un esempio di calcolo: ancora il centralino

Possiamo modellizzare mediante la nuova distribuzione il problema del centralino, discusso dal paragrafo 1.2.
Ricordando che, in tale situazione, il parametro che caratterizza la distribuzione è $\lambda = 2$ e che siamo interessati alla realizzazione di $k = 3$ telefonate, abbiamo:

$$
\begin{array}{ll}
 & P(X = 0) = e^{-2} \\
\boxed{\cdot\frac{2}{1}} & P(X = 1) = \boxed{\frac{2}{1}} \cdot P(X = 0) = 2\, e^{-2} \\
\boxed{\cdot\frac{2}{2}} & P(X = 2) = \boxed{\frac{2}{2}} \cdot P(X = 1) = 1 \cdot 2\, e^{-2} = 2\, e^{-2} \\
\boxed{\cdot\frac{2}{3}} & P(X = 3) = \boxed{\frac{2}{3}} \cdot P(X = 2) = \frac{2}{3} \cdot 2\, e^{-2} = \frac{4}{3}\, e^{-2} \approx 0{,}180
\end{array}
$$

Oppure si giunge allo stesso risultato applicando direttamente la formula che fornisce l'espressione analitica di $P(X = k)$:

$$P(X = 3) = \tfrac{2^3}{3!} e^{-2} \simeq \mathbf{0{,}180}.$$

Concludiamo dunque che la probabilità che si realizzino esattamente 3 telefonate in 10 minuti è quasi del 20%.

Osservazione

Torniamo a considerare il modello binomiale che avevamo adottato inizialmente[12] per schematizzare la situazione delle telefonate. Abbiamo scelto di suddividere l'intervallo $[0, 10]$ in $n = 100$ intervallini e dunque abbiamo considerato $n = 100$ prove.

Come sarebbe cambiata la probabilità richiesta se avessimo scelto un altro valore per n, ad esempio $n = 200$?

In tal caso il parametro $\lambda = 2$ (numero medio di telefonate nell'intervallo di 10 minuti) rimane lo stesso, mentre il valore di p cambia:

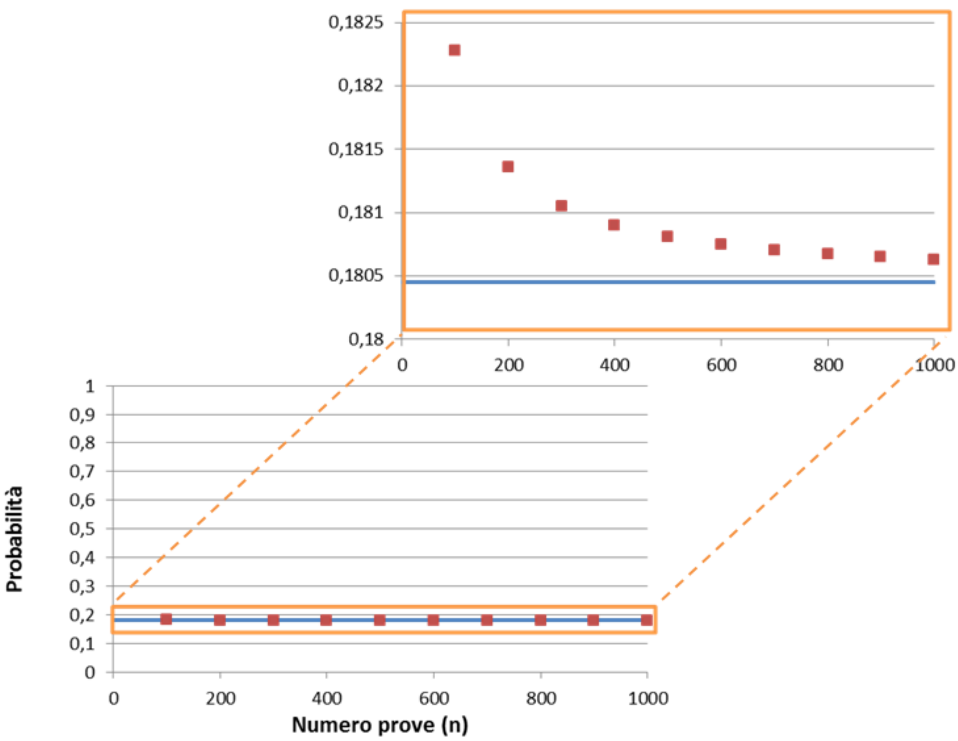
$$p = \frac{\#\ \textit{realizzazioni evento}}{\#\ \textit{prove}} = \frac{2}{200} = 0{,}01.$$

E mediante questo secondo modello binomiale (cioè con $n = 200$ e $p =$0,01), la probabilità richiesta vale:

$$\binom{200}{3} \cdot 0{,}01^3 \cdot 0{,}99^{197} \simeq \mathbf{0{,}1814}.$$

I seguenti grafici mostrano cosa accade all'aumentare del numero n di prove.

Le ordinate dei punti rossi rappresentano i valori di probabilità ottenuti mediante il modello binomiale, mentre la retta blu ha equazione $y = c$, dove $c \simeq$0,180 è il valore trovato utilizzando il nuovo modello, ossia la distribuzione di Poisson. Possiamo osservare che il valore ottenuto col nuovo modello si può considerare, in un senso opportuno[13], il limite a cui tende il valore ottenuto dal modello binomiale, al crescere di n.



Dal grafico in alto, sembra che i valori di probabilità ottenuti al variare di n siano molto diversi tra loro. Ciò, però, dipende fortemente dalla scala adottata: utilizzando come range, per la probabilità l'intero l'intervallo $[0, 1]$, come nel grafico in basso, i valori risultano praticamente indistinguibili.

[12] Vedi il paragrafo 1.2.

[13] Come vedremo nel paragrafo 3.2.