DISTRIBUZIONE DI POISSON

La distribuzione di Poisson é una distribuzione di probabilità discreta, dipendente da un parametro (λ o λ_T), utilizzata per il calcolo della probabilità che un evento si verifichi un certo numero di volte in un intervallo temporale $[0, T]$, sapendo che mediamente esso si verifica λ volte in tale intervallo.
Si evince che essa si usa per rappresentare una vasta gamma di fenomeni casuali, in particolare quelli connessi ad un qualche "conteggio" di eventi in un prefissato periodo di tempo: errori in una pagina, telefonate che arrivano ad un centralino, incidenti che una compagnia deve risarcire, interruzioni di energia elettrica, infortuni sul lavoro, richieste di intervento per manutenzione, o nell'ambito idrologico il numero di perturbazioni o di piene di un fiume, o ancora in sismologia il numero di terremoti di una data intensità. In particolare le ultime due problematiche hanno un riflesso importante in ingegneria civile (costruzione di ponti, dighe, sbarramenti fluviali, edifici che rispettino le normative antisismiche).
L'intervallo temporale su cui calcoliamo la media del fenomeno varia di volta in volta e in alcuni casi può accadere che lo modifichiamo noi adattandolo alle nostre esigenze. Ad esempio, se vogliamo studiare la probabilità che ad un centralino arrivino un certo numero di telefonate nell'arco di un'ora e sappiamo che ne arrivano in media 40 dalle otto alle tredici, ossia ogni cinque ore ($\lambda_5 = 40$), allora conviene lavorare con $\lambda = 8$ e scegliere come intervallo temporale $[0, 1]$. In generale il dato λ si dà sull'unita temporale, ma non é una regola fissa.
A volte gli eventi anziché essere relativi a intervalli temporali, possono essere riferiti a unità spaziali (voluni o superfici), ma noi non tratteremo questi casi.
Tornando all'esempio delle telefonate al centralino con $\lambda = 8$ in $[0, 1h]$, ricaviamo che il tempo medio tra l'arrivo di due telefonate successive é $\frac{1}{8}h = 7,5$ minuti.
Questo vale in generale: se un evento si verifica λ volte nell'unità di tempo allora il tempo medio che intercorre tra due eventi successivi (o a partire dall'istante $t = 0$ fino al primo evento) é $\frac{1}{\lambda}$. $\frac{1}{\lambda}$ si chiama anche tempo medio di ritorno (in generale se l' evento si verifica λ_T volte in $[0, T]$, il tempo di ritorno é $\frac{T}{\lambda_T}$). Se λ_T é il numero medio di eventi in $[0, T]$, allora il numero medio di eventi in $[0, T_1]$ sarà $\lambda_{T_1} = \frac{\lambda_T \cdot T_1}{T}$ (da 40 telefonate in cinque ore ricaviamo ad esempio $\frac{40 \cdot 2}{5} = 16$ telefonate ogni due ore). Il tempo medio di ritorno non cambia, perché é sempre il tempo che intercorre tra due eventi successivi.
Consideriamo lo spazio dei possibili risultati dei fenomeni descritti su: $\Omega = \{0, 1, \ldots, n \ldots\}$ (possono arrivare 5 telefonate, 10, 0). La scrittura $X = k$ rappresenta l'evento "il fenomeno avviene k volte nell'intervallo considerato" (ad esempio $X = 6$ vuol dire arrivano sei telefonate in un'ora). Si prova che

$$(1) \qquad p(X = k) = p(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \ \forall k \in N, \, p(x) = 0 \text{ se } x \notin N, \ X \sim Pois(\lambda).$$

Dalla formula

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

ricaviamo che

$$\sum_{k=0}^{\infty} p(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1\,.$$

Inoltre

$$p(X \leq j) = e^{-\lambda} \cdot \sum_{k=0}^{j} \frac{\lambda^k}{k!}, \; p(X \geq j) = e^{-\lambda} \cdot \sum_{k=j}^{\infty} \frac{\lambda^k}{k!} = 1 - p(X < j) = 1 - e^{-\lambda} \cdot \sum_{k=0}^{j-1} \frac{\lambda^k}{k!}$$

ESERCIZI

1) Un centralino riceve in media tre telefonate ogni dieci minuti. Qual é la probabilità che nei prossimi dieci minuti riceva non più di due telefonate? E che ne riceva 8? Qual é la probabilità che nei prossimi cinque minuti non arrivino telefonate?
$\lambda = 3, \; T = 10$ minuti.

$$p(X \leq 2) = \sum_{k=0}^{2} e^{-3} \cdot \frac{3^k}{k!} = e^{-3} \cdot \left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!} \right) = e^{-3} \cdot 8,5 = 42\% \, .$$

$$p(X = 8) = e^{-3} \cdot \frac{3^8}{8!} \simeq 0,81\% \, .$$

Se i minuti sono 5 allora $\lambda = 1,5$ e

$$p(X = 0) = e^{-1,5} \cdot \frac{3^0}{0!} = e^{-1,5} \simeq 22\% \, .$$

2) 200 errori sono distrbuiti a caso in un libro di 500 pagine. Qual é la probabilità che una pagina abbia almeno due errori.
Il numero medio di errori per pagina é $\lambda = \frac{200}{500} = 0,4$.

$$p(X \geq 2) = 1 - p(X < 2) = 1 - \sum_{k=0}^{1} e^{-0,4} \cdot \frac{(0,4)^k}{k!} \simeq \qquad .$$

3) In una regione si verificano tre scosse sismiche di magnitudo $0,4$ al mese. Qual é la probabilità che nella prossima settimana se ne verifichino più di due?
$X \sim Pois(0,75)$ e $p(X > 2) = 1 - p(X \leq 2) = \qquad \%$.

Se λ é intero allora $p(X = \lambda - 1) = p(X = \lambda)$ e la probabilità più alta si ha in corrispondenza di $\lambda - 1$ e di λ. Se λ non é intero allora la probabilità più alta si ha in corrispondenza di $k = [\lambda]$. Nell' esercizio 1, ad esempio, con $\lambda = 3$ si ha $p(X = 2) = p(X = 3)$, mentre per $\lambda = 1,5$ la probabilità più alta é $p(X = 1)$.

Proprietà 0.1. *La somma di v.a. di Poisson indipendenti, é una v.a. di Poisson avente come media la somma delle medie delle v.a. di partenza. Inoltre*

$$E[Pois(\lambda)] = \lambda, \; \sigma^2(Pois(\lambda)) = \lambda \, .$$

ESERCIZI

1) La ricezione di telefonate a due centralini **indipendenti** nel tempo $[0, 1\,h]$ hanno legge di Poisson (X e Y) di parametri $\lambda = 2$ e $\mu = 4$. Qual é la probabilità che il primo riceva **più** di 6 telefonate in un'ora? Qual é la probabilità che il primo riceva più di tre telefonate in un'ora **e** che il secondo non ne riceva nessuna? Che il totale di telefonate ricevute in un'ora sia 5? Qual é la probabilità che il primo riceva una telefonata in un'ora, sapendo che in quell'ora in totale ne arrivano 8?
$X \sim Pois(2), \; Y \sim Pois(4)$.

$$p(X > 6) = P(X \geq 7) = 1 - p(X \leq 6) = \qquad ,$$

$$p(X>3,Y=0)=p(X>3)\cdot p(Y=0)=(1-p(X\leq 3))\cdot p(Y=0)= \qquad .$$

$$p(X+Y=5)=e^{-6}\cdot\frac{6^5}{5!}= \qquad .$$

$$p(X=1|X+Y=8)=\frac{p(X=1,X+Y=8)}{p(X+Y=8)}=\frac{p(X=1,Y=7)}{p(X+Y=8)}=\frac{p(X=1)\cdot p(Y=7)}{p(X+Y=8)}=$$

$$=\frac{e^{-2}\cdot\frac{2}{1!}\cdot e^{-4}\cdot\frac{4^7}{7!}}{e^{-6}\cdot\frac{6^8}{8!}}=\frac{2\cdot\frac{4^7}{7!}}{\frac{6^8}{8!}}=\frac{2\cdot 4^7}{\frac{6^8}{8}}= \qquad \%\,.$$

2) In una zona priva di strutture fognarie il rischio di infezione da colera é di 2 casi su 1000 persone. Valutare la probabilità che in un campione casuale di 300 persone vi siano: a) esattamente 2 casi di colera, b) non più di 1 caso di colera.
$\lambda=0,002$, $\lambda_{300}=0,002\cdot 300=0,6$.

$$P(X=2)=e^{-0,6}\cdot\frac{(0,6)^2}{2}= \ .$$

$$P(X\leq 1)=e^{-0,6}\cdot(1+0,6)= \ .$$

3) In media in un paracadute su 1000 il paracadute principale é difettoso e non si apre durante il lancio. Un paracadutista professionista compie 4000 lanci nella sua carriera; quanto vale la probabilità che il paracadute principale non si apra in almeno uno dei 4000 lanci?
$\lambda=4$

$$P(X\geq 1)=1-P(X=0)=1-e^{-4}=0,9816\,.$$

DISTRIBUZIONE DI POISSON ED EVENTI RARI

Consideriamo ora eventi dipendenti dal tempo. Abbiamo visto che se un evento si verifica in media λ volte nell'unità di tempo ($[0,1]$), allora, detto $N(t)$ il numero di volte in cui si verifica in $[0,t]$, la probabilità che si verifichi k volte in ($[0,t]$) é

$$P(N(t)=k)=e^{-\lambda t}\cdot\frac{(\lambda t)^k}{k!}\,. \tag{2}$$

Abbiamo scritto una formula generale che permette di determinare le probabilità al variare dell'intervallo temporale.
$P(N(t)=1)=e^{-\lambda t}\cdot\lambda t$ e se $t\approx 0$ allora $\frac{P(N(t)=1)}{t}\approx\lambda$.
Utilizzeremo tra poco le osservazioni fatte.
Consideriamo una serie di eventi di durata brevissima (istantanea) e sia $N(t)$ il numero di eventi avvenuti in $[0,t]$ (ad esempio numero di veicoli che arrivano in una certa località in un fissato periodo di tempo $[0,t]$). In tal caso non abbiamo una sola distribuzione di probabilità (quella che avremmo per un intervallo temporale fissato), ma delle variabili casuali dipendenti dal tempo. È quello che si dice un processo stocastico. In particolare $N(t)$ **si dice processo di Poisson di intensità $\lambda>0$ se verifica le seguenti proprietà:**

- $N(0)=0$, ossia iniziamo a contare gli eventi dell'istante iniziale;
- il numero di eventi che hanno luogo in intervalli di tempo disgiunti sono indipendenti (ad esempio il numero di eventi avvenuti in $[1,4]$ non influenza il numero di eventi che avverrano dopo il tempo 4, cosi come il numero di eventi avvenuti prima del tempo 1 non influenza il numero di eventi in $[1,4]$ o in un qualsiasi altro intervallo dopo il tempo 1);

- la distribuzione del numero di eventi che si verificano in un intervallo dipende solo dalla sua ampiezza e non dalla posizione temporale (la distribuzione degli eventi in $[1, 3]$ é uguale a quella in $[2, 4]$, $[5, 7]$ e cosi via): si chiama stazionarietá degli incrementi;
- $\lim_{h \to 0^+} \frac{P(N(h)=1)}{h} = \lambda$: in un intervallo di tempo molto piccolo di durata h vi é una probabilità circa λh che avvenga un solo evento e
- $\lim_{h \to 0^+} \frac{P(N(h) \geq 2)}{h} = 0$: in un intervallo di tempo molto piccolo di durata h vi é una probabilità quasi nulla che avvengano due o più eventi.

Tali condizioni permettono di affermare che

$$P(N(t) = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!} \,,$$

dove λ é il valore del limite su e rappresenta il numero medio di eventi che si verificano in un intervallo di ampiezza unitaria.
N.B. Se riguardate quanto fatto in precedenza, abbiamo già viato come si calcolano le probabilità in un processo di Poisson. Dobbiamo solo vedere come si comportano i tempi che separano due eventi in questo processo, ma tatteremo questo problema in seguito perché ci occorre la distribuzione esponenziale.