Sannolikhet och statistik VII

Johan Jonasson

April 2019

Exponentialfördelning. En kont sv som har täthet

$$f(x) = \lambda e^{-\lambda x},\ x > 0$$

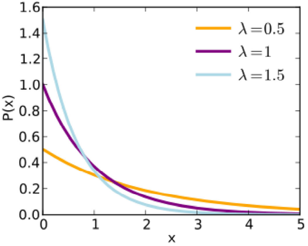
sägs vara *exponentialfördelad med parameter (intensitet)* λ.
Kortform $X \sim exp(\lambda)$.
Uppstår som livslängd av saker som inte åldras, t.ex.

- Livslängden av en LED-lampa.
- Väntetiden till nästa bil som passerar på en enslig skogsväg.
- Tiden till nästa jordbävning.

Vi ser att $F(x) = 1 - e^{-\lambda x}$, dvs $\mathbb{P}(X > x) = e^{-\lambda x}$.

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Ex på frekvensfuntion för exponentialfördelningen:



Figur: Eponentialfördelning.

Det gäller också att

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

så

$$\mathbb{V}\mathrm{ar}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Exempel: En LED-lampa sägs ha livslängd tio år. Hur stor är sannol att den håller minst tio år? Rimligt (?) att anta att lampan inte åldras och därmed att dess livslängd, X, är $exp(\lambda)$-fördelad, där det verkar rimligt att tolka uppgiften som att $\mathbb{E}[X] = 10$, dvs $\lambda = 0.1$. Alltså

$$\mathbb{P}(X > 10) = e^{-0.1 \cdot 10} = e^{-1} \approx 0.37.$$

$\square$

Allmän observation: $X \sim exp(\lambda) \Rightarrow \mathbb{P}(X > \mathbb{E}[X]) = e^{-1}$.

Exponentialfördelningen uppfyller *glömskeegenskapen* (dvs inget åldrande):

$$\mathbb{P}(X > t + x | X > t) = \mathbb{P}(X > x)$$

ty

$$\begin{aligned} \mathbb{P}(X > t + x | X > t) &= \frac{\mathbb{P}(X > t + x, X > t)}{\mathbb{P}(X > t)} \\ &= \frac{\mathbb{P}(X > t + x)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} \\ &= e^{-\lambda x} = \mathbb{P}(X > x). \end{aligned}$$

Vice versa: antag att $\mathbb{P}(X > t + x | X > t) = \mathbb{P}(X > x)$ för alla x och t. Skriv $G(x) = \mathbb{P}(X > x)$. Precis som ovan

$$\mathbb{P}(X > t + x | X > t) = \frac{\mathbb{P}(X > t + x)}{\mathbb{P}(X > t)}$$

och detta antogs vara lika med $\mathbb{P}(X > x)$, vilket ger

$$G(x + t) = G(t)G(x)$$

för alla x och t. Subtrahera $G(x)$ från båda sidor, dela med t och låt $t \to 0$ och få

$$G'(x) = G'(0)G(x)$$

för alla x. Detta har den unika lösningen

$$G(x) = e^{-\lambda x}$$

där $\lambda = -G'(0) > 0$ eftersom G är avtagande. Alltså är X exponentialfördelad.

Vi har visat att X har glömskeegenskapen *om och endast om* X är exponentialfördelad.

Exempel: Exponentialfördelning och geometrisk fördelning.

Låt $X \sim exp(\lambda)$ och $Y = \lceil X \rceil$. Då gäller

$$\begin{aligned}
\mathbb{P}(Y = k) &= \mathbb{P}(X \in (k-1, k]) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k) \\
&= e^{-\lambda(k-1)} - e^{-\lambda k} = e^{-\lambda(k-1)}(1 - e^{-\lambda}) \\
&= p(1-p)^{k-1}
\end{aligned}$$

med $p = 1 - e^{-\lambda}$.

Kortfattat: $X \sim exp(\lambda) \Rightarrow \lceil X \rceil \sim Geo(1 - e^{-\lambda})$. $\square$

Proposition: Om $X_1 \sim exp(\lambda_1)$, $X_2 \sim exp(\lambda_2)$ och X_1 och X_2 är oberoende, så gäller $\min(X_1, X_2) \sim exp(\lambda_1 + \lambda_2)$.

Bevis:

$$\begin{aligned}
\mathbb{P}(\min(X_1, X_2) > x) &= \mathbb{P}(X_1 > x, X_2 > x) \\
&= \mathbb{P}(X_1 > x)\mathbb{P}(X_2 > x) \\
&= e^{-\lambda_1 x} e^{-\lambda_2 x} \\
&= e^{-(\lambda_1 + \lambda_2)x}.
\end{aligned}$$

$\square$

Låt $T_1, T_2, \ldots, T_n$ vara oberoende och $exp(\lambda)$-fördelade (obs. samma λ för alla). Summan

$$X = \sum_{k=1}^{n} T_k$$

sägs vara *gammafördelad med parametrar n och λ*. Kortform $X \sim \Gamma(n, \lambda)$.

Proposition:

$$\mathbb{P}(X > x) = e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}.$$

Bevis: Vi återkommer. $\square$

Poissonprocessen. Låt $0 < \tau_1 < \tau_2 < \tau_3 < \ldots$ vara en följd av tidpunkter då en viss typ av händelser (impulser) inträffar (t.ex. jordbävningar, benbrott, mål i en fotbollsmatch, bilar som passerar på en enslig väg). Skriv

$$T_1 = \tau_1,\ T_2 = \tau_2 - \tau_1,\ T_3 = \tau_3 - \tau_2, \ldots$$

för tidsmellanrummen mellan impulser. Om $T_1, T_2, T_3, \ldots$ är oberoende och $\exp(\lambda)$-fördelade, kallas följden $\{\tau_1, \tau_2, \ldots\}$ för en *Poissonprocess med intensitet* λ.

Obs: För en Poi-process med int λ gäller $\tau_n \sim \Gamma(n, \lambda)$.

Låt $X(t) = \max\{n : \tau_n \leq t\}$ = antal impulser som kommit vid tid t. Enlig prop ovan

$$\begin{aligned}
\mathbb{P}(X(t) = n) &= \mathbb{P}(\tau_n \leq t, \tau_{n+1} > t) \\
&= \mathbb{P}(\tau_{n+1} > t) - \mathbb{P}(\tau_n > t) \\
&= e^{-\lambda t} \frac{(\lambda t)^n}{n!}.
\end{aligned}$$

Alltså: $X(t) \sim Poi(\lambda t)$. Tack vare glömskegenskapen gäller $X(s) - X(t) \sim Poi(\lambda(s-t))$ för alla $t < s$.

Exempel: På en given lågtrafikerad väg kommer i genomsnitt fem bilar per timme. Vad är sannolikheten att det passerar minst två bilar under en kvart?

Rimligt att anta att bilar kommer enligt en Poissonprocess. Det måste då gälla att $\lambda = 5$ bilar/timme. Vi är intr av $X(1/4)$, som är $Poi(5/4)$-fördelad.

$$\mathbb{P}(X(1/4) \geq 2) = 1 - e^{-5/4}\left(\frac{(5/4)^0}{0!} + \frac{(5/4)^1}{1!}\right) \approx 0.36.$$

$\square$

Betrakta två oberoende Poissonprocesser, kalla dem P1 och P2, med intensiteter λ_1 och λ_2. Då gäller enligt glömskeegenskapen vid varje tidpunkt

- Tiden T_{P1} till nästa impuls i P1 är $exp(\lambda_1)$-förd.
- Tiden T_{P2} till nästa impuls i P2 är $exp(\lambda_2)$-förd.
- Dessa tider är oberoende.

Tiden till nästa impuls i processen som räknar impulser i både P1 och P2 är då $\min(T_{P1}, T_{P2}) \sim exp(\lambda_1 + \lambda_2)$.
Alltså: den sammanvägda processen är en Poissonprocess med intensitet $\lambda_1 + \lambda_2$.
Konsekvens: $X_1(t) + X_2(t) \sim Poi(\lambda_1 + \lambda_2)$. M.a.o.

$$Y_1 \sim Poi(\lambda_1),\ Y_2 \sim Poi(\lambda_2),\ Y_1, Y_2 \text{ oberoende} \Rightarrow Y_1 + Y_2 \sim Poi(\lambda_1 + \lambda_2)$$

Direktbevis av detta:

$$
\begin{aligned}
\mathbb{P}(Y_1 + Y_2 = k) &= \sum_{j=0}^{k} \mathbb{P}(Y_1 = j)\mathbb{P}(Y_2 = k-j) \\
&= \sum_{j=0}^{k} e^{-\lambda_1} \frac{\lambda_1^j}{j!} e^{-\lambda_2} \frac{\lambda_2^{k-j}}{(k-j)!} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{\sum_{j=0}^{k} \binom{k}{j} \lambda_1^j \lambda_2^{k-j}}{k!} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}
\end{aligned}
$$

där sista likheten är binomialsatsen. $\square$

Exempel: På en lågtrafikerad väg kommer i genomsnitt fem bilar per timme. Det kommer också i genomsnitt två mc. Vad är sannol att det passerar exakt tre fordon på en halvtimme?

Rimligt att anta två ober Poi-processer med int 5 resp 2 så bägge sorters fordon sammanvägda ger en Poi-process med int 7. Låt $X(t)$ vara antal fordon på tid t. $X(1/2) \sim Poi(7/2)$. Alltså

$$\mathbb{P}(X(1/2) = 3) = e^{-7/2}\frac{(7/2)^3}{3!} \approx 0.22.$$

Om det passerar exakt n fordon på tid t, vad är den betingade sannol att k av dessa är mc? Skriv $X_b(t)$ resp $X_m(t)$ för antal bilar resp mc.

$$\begin{aligned}
\mathbb{P}(X_m(t) = k | X(t) = n) &= \frac{\mathbb{P}(X_m(t) = k, X(t) = n)}{\mathbb{P}(X(t) = n)} \\
&= \frac{\mathbb{P}(X_m(t) = k)\mathbb{P}(X_b(t) = n - k)}{\mathbb{P}(X(t) = n)} \\
&= \frac{e^{-2t\frac{(2t)^k}{k!}} e^{-5t\frac{(5t)^{n-k}}{(n-k)!}}}{e^{-7t\frac{(7t)^n}{n!}}} \\
&= \binom{n}{k} \left(\frac{2}{7}\right)^k \left(\frac{5}{7}\right)^{n-k}
\end{aligned}$$

dvs en $Bin(n, 2/7)$-fördelning. $\square$