STK1100 våren 2023

Gammafordelingen, eksponentialfordelingen og kjikvadratfordelingen

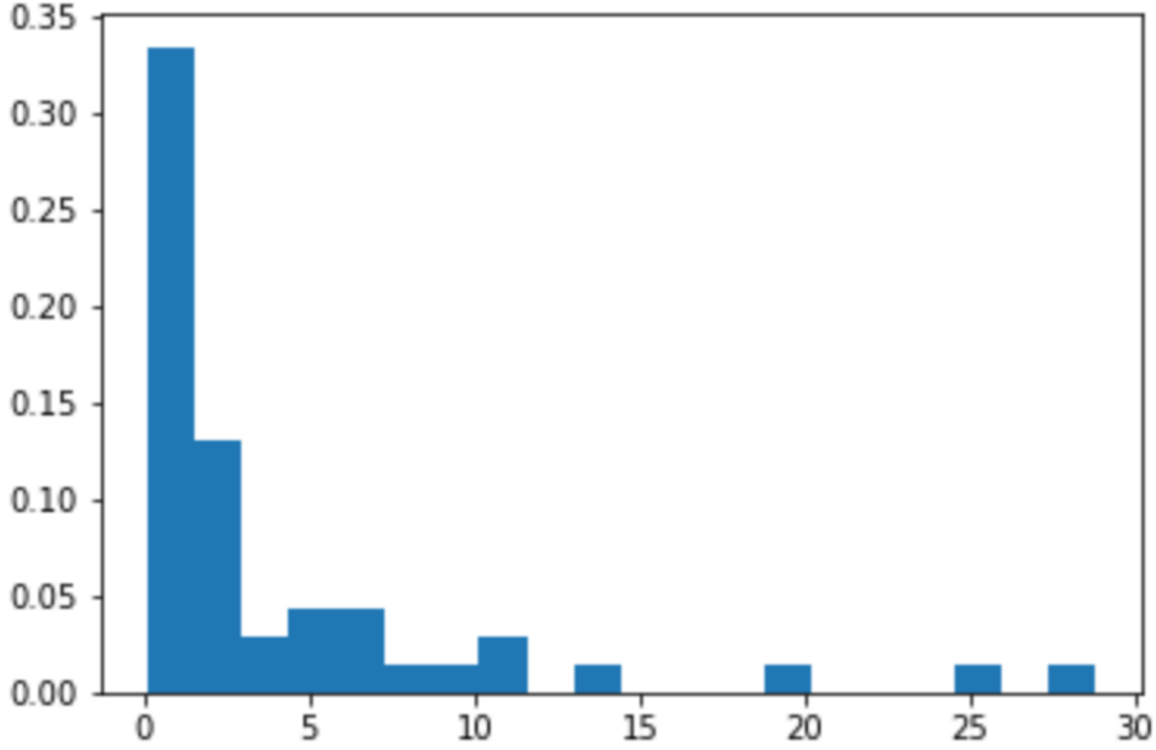
Svarer til avsnitt 4.4 i læreboka

Matematisk institutt
Universitetet i Oslo

Normalfordelingen er symmetrisk.

I flere situasjoner kan vi ikke bruke en symmetrisk fordeling til å beskrive variasjonen for en kontinuerlig stokastisk variabel.

Eksempel: Nedbør på Blindern per døgn mai - juli 2022 (de 48 dagene det regnet, data fra yr.no):



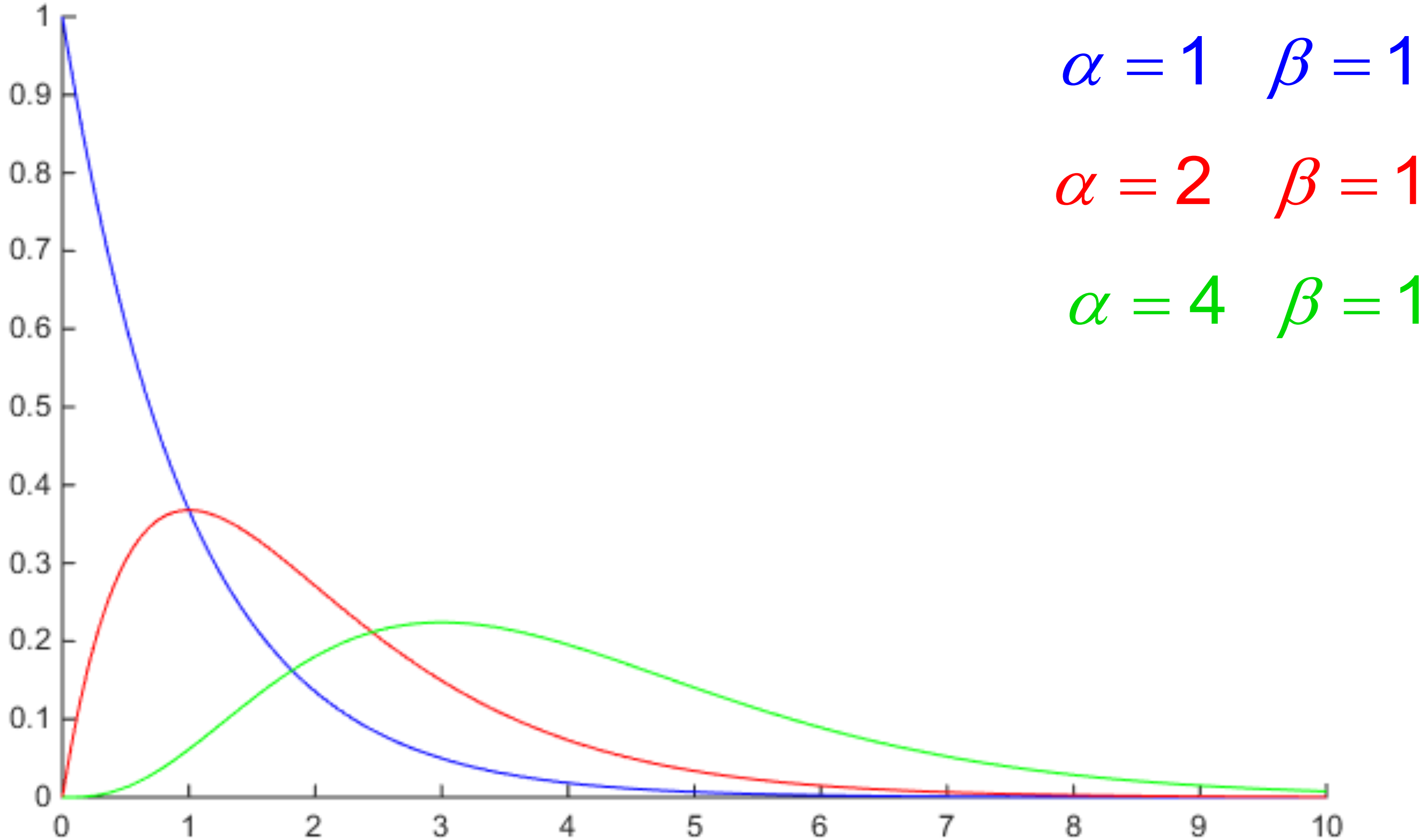
Det fins flere fordelinge som ikke er symmetriske. En av dem er gammafordelingen.

En kontinuerlig stokastisk variabel X er gamma-fordelt med formparameter $\alpha > 0$ og skalaparameter $\beta > 0$ hvis den har sannsynlighetstetthet

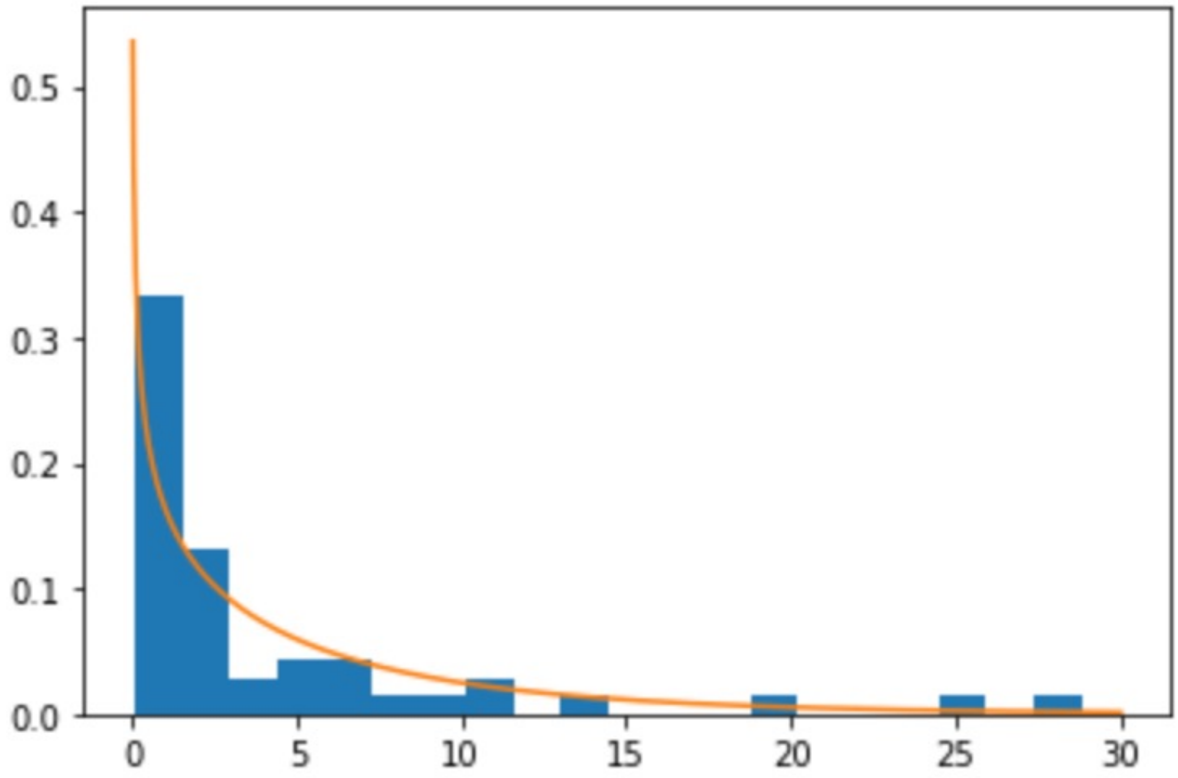
$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\,\Gamma(\alpha)}\, x^{\alpha-1} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

der $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du$ er gammafunksjonen.

Gammatettheter for ulike verdier av α og β



Gammafordelingen med $\alpha = 0.70$ og $\beta = 7.5$ gir en rimelig god beskrivelse av nedbørsdataene:



Gammafunksjonen

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u}\, du \qquad (\alpha > 0)$$

har følgende egenskaper:

1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$

2) $\Gamma(n) = (n-1)!$ når n er et positivt heltall

3) $\Gamma(1/2) = \sqrt{\pi}$

Vi viser disse resultatene på forelesningen

Momentgenerererende funksjon (detaljer på forelesningen):

$$M_X(t) = E(e^{tX}) \quad = \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \, dx$$

$$= \frac{1}{(1-\beta t)^\alpha} \quad \text{for } t < \frac{1}{\beta}$$

Dette gir

$$R_X(t) = \ln\{M_X(t)\} = -\alpha \ln(1-\beta t)$$

Dermed er

$$E(X) = R_X'(0) = \alpha\beta$$

$$V(X) = R_X''(0) = \alpha\beta^2$$

Kumulativ fordeling (for $x > 0$):

$$F(x;\alpha,\beta) = P(X \le x) \quad = \int_0^x \frac{1}{\beta^\alpha\,\Gamma(\alpha)}\, y^{\alpha-1} e^{-y/\beta}\, dy$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^{x/\beta} u^{\alpha-1} e^{-u}\, du$$

$$= F(x/\beta;\, \alpha)$$

der

$$F(x;\,\alpha) = \frac{\int_0^x u^{\alpha-1}\, e^{-u}\, du}{\Gamma(\alpha)}$$

er den ufullstendige gammafunksjonen (ofte blir telleren kalt den ufullstendige gammafunksjonen).

Tabell A.4 bak i boka gir $F(x;\alpha)$ for noen verdier av α og *x.* Kan glemmes!

Gammafordeling i Python

Sannsynlighetstetthet: $f(x;\alpha,\beta) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$ for $x > 0$

Python: import scipy.stats as stats
stats.gamma.pdf(x,*a*,scale=*b*)

Kumulativ fordeling:

$$F(x;\alpha,\beta) = \int_0^x \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}\,dy \quad \text{for } x > 0$$

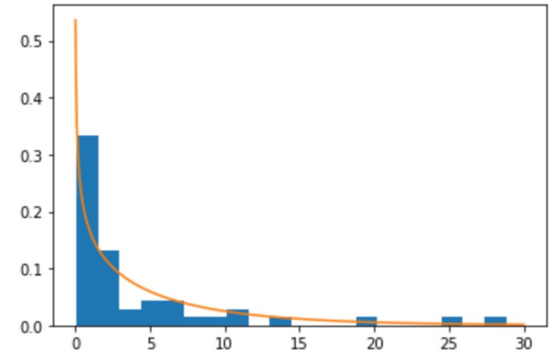
Python: stats.gamma.cdf(x,*a*,scale=*b*)

Persentiler: $\eta(p)$ er gitt ved at $F(\eta(p);\alpha,\beta) = p$

Python: stats.gamma.ppf(p,*a*,scale=*b*)

Eksempel – nedbør på Blindern

```
alpha=0.7
beta=7.5
xx = np.linspace(0,30,1000)
f = stats.gamma.pdf(xx,alpha,scale=beta)
plt.hist(x, bins=20, density=True)
plt.plot(xx,f)
plt.show()
```



X er mengden nedbør en dag det regner.

Anta at X er gammafordelt med $\alpha = 0.70$ og $\beta = 7.5$

Vi bruker Python og finner f.eks. at

$$P(X \leq 2) = 0.392$$

$$P(X > 10) = 0.162$$

$$P(5 \leq X \leq 10) = 0.196$$

Eksponentialfordelingen er et spesialtilfelle av gammafordelingen, der vi setter $\alpha = 1$ og $\beta = 1/\lambda$

En kontinuerlig stokastisk variabel X er eksponential-fordelt med parameter $\lambda > 0$ hvis den har sannsynlighetstetthet

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

Vi har at

$$E(X) = \alpha\beta = \frac{1}{\lambda}$$

$$V(X) = \alpha\beta^2 = \frac{1}{\lambda^2}$$

Eksponentialfordelingen

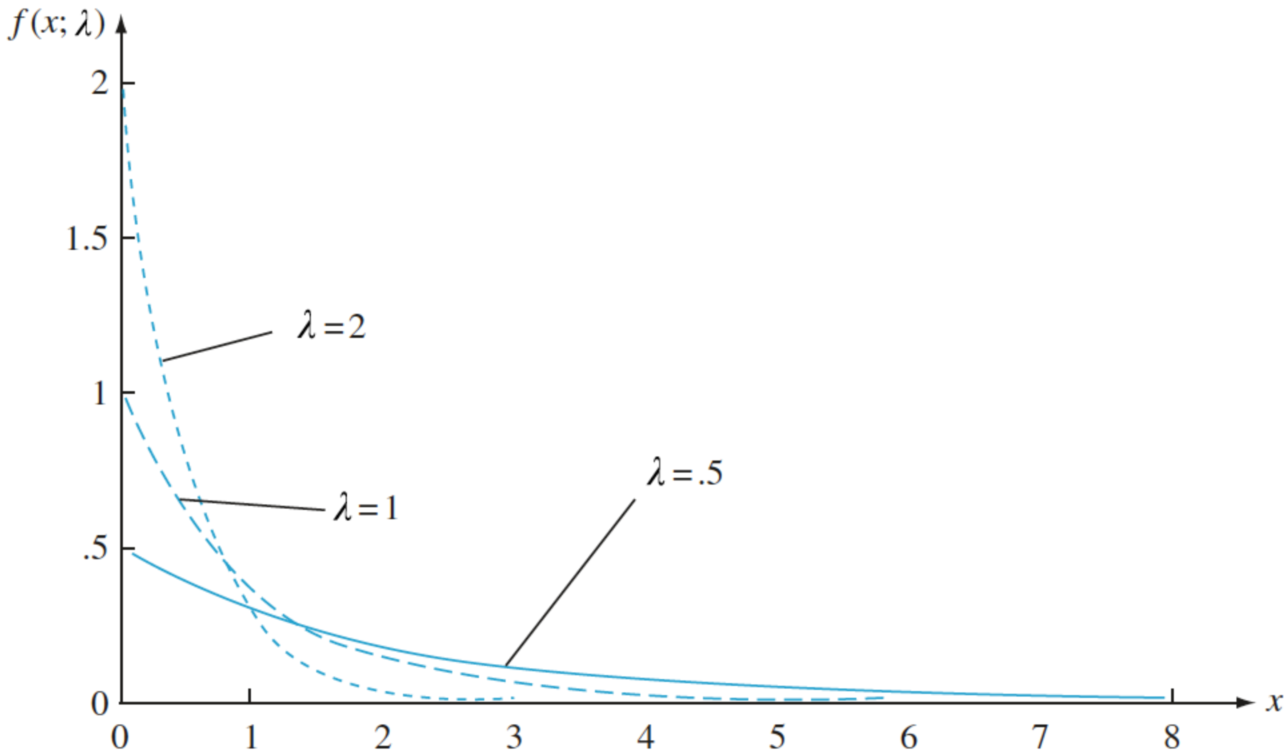


Figure 4.27 Exponential density curves

Kumulativ fordeling:

$$F(x;\lambda) = P(X \le x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\lambda x} & \text{for } x \ge 0 \end{cases}$$

Vi har at:

$$P(X \ge t + t_0 \mid X \ge t_0) = \frac{P\{(X \ge t + t_0) \cap (X \ge t_0)\}}{P(X \ge t_0)}$$

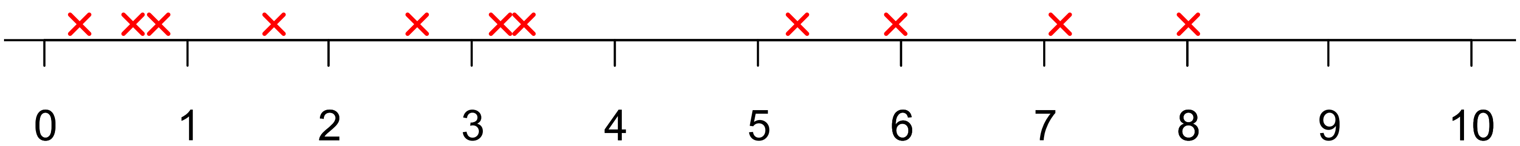
$$= \frac{P(X \ge t + t_0)}{P(X \ge t_0)} = \frac{1 - F(t + t_0;\lambda)}{1 - F(t_0;\lambda)}$$

$$= \frac{e^{-\lambda(t + t_0)}}{e^{-\lambda t_0}} = e^{-\lambda t} = P(X \ge t)$$

Eksponentialfordelingen har «ingen hukommelse»!

Vi minner om Poissonprosessen:

Vi observeret begivenheter (markert med x)
som hender over tid:



Vi antar at:

- sannsynligheten for at det inntreffer én begivenhet i et intervall av lengde Δt er $\alpha\Delta t + o(\Delta t)$
- sannsynligheten for at det inntreffer mere enn én begivenhet i et intervall av lengde Δt er $o(\Delta t)$
- antall begivenheter i et intervall er uavhengig av hvor mange begivenheter som har skjedd tidligere

La $P_k(t)$ være sannsynligheten for at det inntreffer *k* begivenheter i et intervall av lengde *t.*

Da har vi at $P_k(t) = \frac{(\alpha t)^k}{k!} e^{-\alpha t}$

Antall begivenheter i et intervall av lengde *t* er Poissonfordelt med parameter αt .

La *T* være tiden mellom to påfølgende begivenheter i en Poissonprosess.

Da er *T* eksponentialfordelt med parameter α .

Vi nøyer oss med å vise resultatet for tid til første begivenhet:

$$
\begin{aligned}
P(T \leq t) &= 1 - P(T > t) \\
&= 1 - P\{\text{ingen begivenhet i } (0,t)\} \\
&= 1 - P_0(t) \quad = 1 - \frac{(\alpha t)^0}{0!} e^{-\alpha t} \\
&= 1 - e^{-\alpha t}
\end{aligned}
$$

Det viser at T er eksponentialfordelt med parameter a .

Kjikvadratfordelingen er et spesialtilfelle av gammafordelingen, der vi har $\alpha = \nu/2$ og $\beta = 2$

En stokastisk variabel X er kjikvadratfordelt med ν frihetsgrader hvis den har sannsynlighetstetthet

$$f(x;\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

Vi har at

$$E(X) = \alpha\beta = \nu \qquad V(X) = \alpha\beta^2 = 2\nu$$

Kjikvadratfordelingen brukes til å beskrive funksjoner av stokastiske variabler

F.eks. vil vi senere vise at hvis $Z \sim N(0,1)$, så er Z^2 kjikvadratfordelt med ν=1