

Pravděpodobnost a matematická statistika

Doc. RNDr. Gejza Dohnal, CSc.

dohnal@nipax.cz

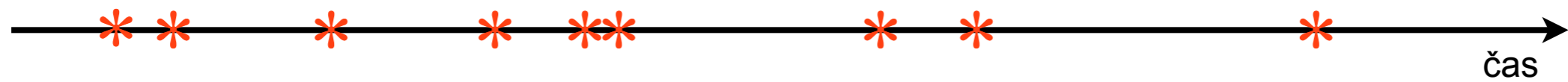


Pravděpodobnost a matematická statistika

2010

- 1.týden (20.09.-24.09.) Data, typy dat, variabilita, frekvenční analýza (histogramy, četnosti absolutní, relativní, prosté, kumulativní), základní statistické charakteristiky (průměr, výběr.rozptyl, minimum, maximum, medián, kvartily, boxplot), sešikmenná rozdělení (vzájemná poloha mediánu a střední hodnoty), chvosty, kvantily
2. týden (27.09.-01.10.) Princip statistické indukce, výběr, vlastnosti výběru, experiment. Náhodná veličina, rozdělení pravděpodobnosti a jeho souvislost s histogramem. Pravděpodobnost, pravidla pro počítání s pravděpodobnostmi, podmíněná pravděpodobnost, závislost náhodných veličin.
- 3.týden (04.10.-08.10.) Využití závislosti při stanovení pravděpodobnosti - věta o úplné pravděpodobnosti a Bayesova věta
- 4.týden (11.10.-15.10.) Rozdělení chyb měření - normální rozdělení a počítání s ním. Odhady parametrů normálního rozdělení. Intervaly spolehlivosti pro normální data. Jednovýběrové testy o střední hodnotě
- 5.týden (18.10.-24.10.) Výběrový poměr jako odhad pravděpodobnosti sledovaného jevu. Alternativní rozdělení, binomické rozdělení. Intervalový odhad výběrového poměru. Výběry s vracením a bez vracení (binomické a hypergeometrické rozdělení)
- 6.týden (25.10.-29.10.) odpadá
- 7.týden (01.11.-05.11.) Poruchy v čase (Poissonův proces). Poissonovo rozdělení, exponenciální rozdělení, jeho výhody a nevýhody, modelování doby do poruchy pomocí Weibullova rozdělení, lognormálního rozdělení, případně useknuté normální rozdělení.**
- 8.týden (08.11.-12.11.) Testy dobré shody, Q-Q graf (pouze vysvětlení), testy normality. Některé neparametrické testy
- 9.týden (15.11.-19.11.) Dvě náhodné veličiny - srovnání dvou výběrů (dvouvýběrové testy)
10. týden (22.11.-26.11.) Dvě náhodné veličiny. Dvourozměrné četnosti jako odhad dvourozměrného rozdělení, frekvenční tabulka. Marginální rozdělení (vše pouze diskrétně s tabulkou)
11. týden (29.11.-03.12.) Závislost náhodných veličin, míry závislosti (kovariance, korelace), test významnosti korelačního koeficientu
12. týden (06.12.-10.12.) Regrese, lineární regresní model (přímková, kvadratická, polynomická regrese), analýza reziduí, pásy spolehlivosti
13. týden (13.12.-17.12.) Více výběrů, jednoduché třídění, ANOVA.
14. týden (20.12.-22.12.) Rezerva, opakování, testy normality (náhrada za 28.10.)

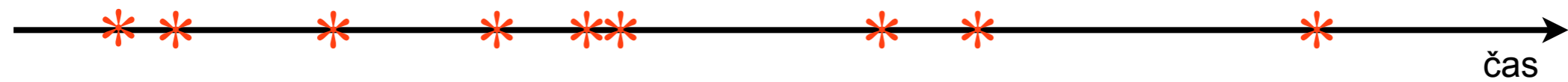
Náhodné události v čase



Náhodné události v čase

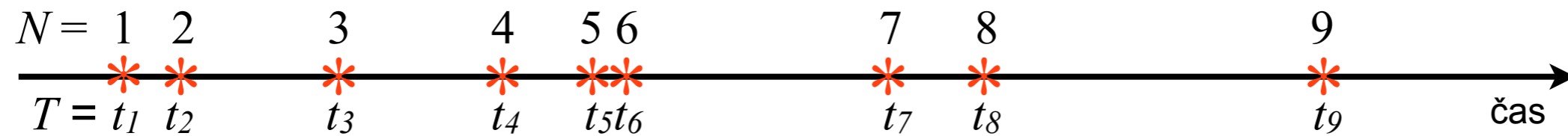


CYCLE OF BIRTH & DEATH



Náhodné události v čase

počet událostí v čase t : $N(t)$ ← diskrétní náhodná veličina



čas výskytu i -té události: T_i ← spojitá náhodná veličina

Náhodné události v čase

počet událostí v čase t : $N(t)$ ← diskrétní náhodná veličina



doby mezi událostmi: $\tau_i = t_i - t_{i-1}$

← spojitá náhodná veličina

$$\lambda(t) = \frac{N(t)}{t}$$

$$\xrightarrow{t \rightarrow \infty}$$

$$\lambda = \lim_{t \rightarrow \infty} \frac{N(t)}{t}$$

← střední počet událostí za jednotku času

$$\bar{T}_i = \frac{t_i}{i}$$

$$\xrightarrow{t \rightarrow \infty}$$

$$\bar{T} = \lim_{t \rightarrow \infty} \frac{t}{N(t)} = \frac{1}{\lambda}$$

← střední doba mezi událostmi

λ je intenzita poruch, $1/\lambda$ je střední doba mezi poruchami

Poissonovo rozdělení

počet událostí v čase t : $N(t)$



Jaká je pravděpodobnost, že za jednotku času nastane k událostí? $P(N = k) = ?$

$$P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

$$E(N) = \lambda, \quad Var(N) = \lambda$$

Jaká je pravděpodobnost, že za dobu t nastane k událostí? $P(N(t) = k) = ?$

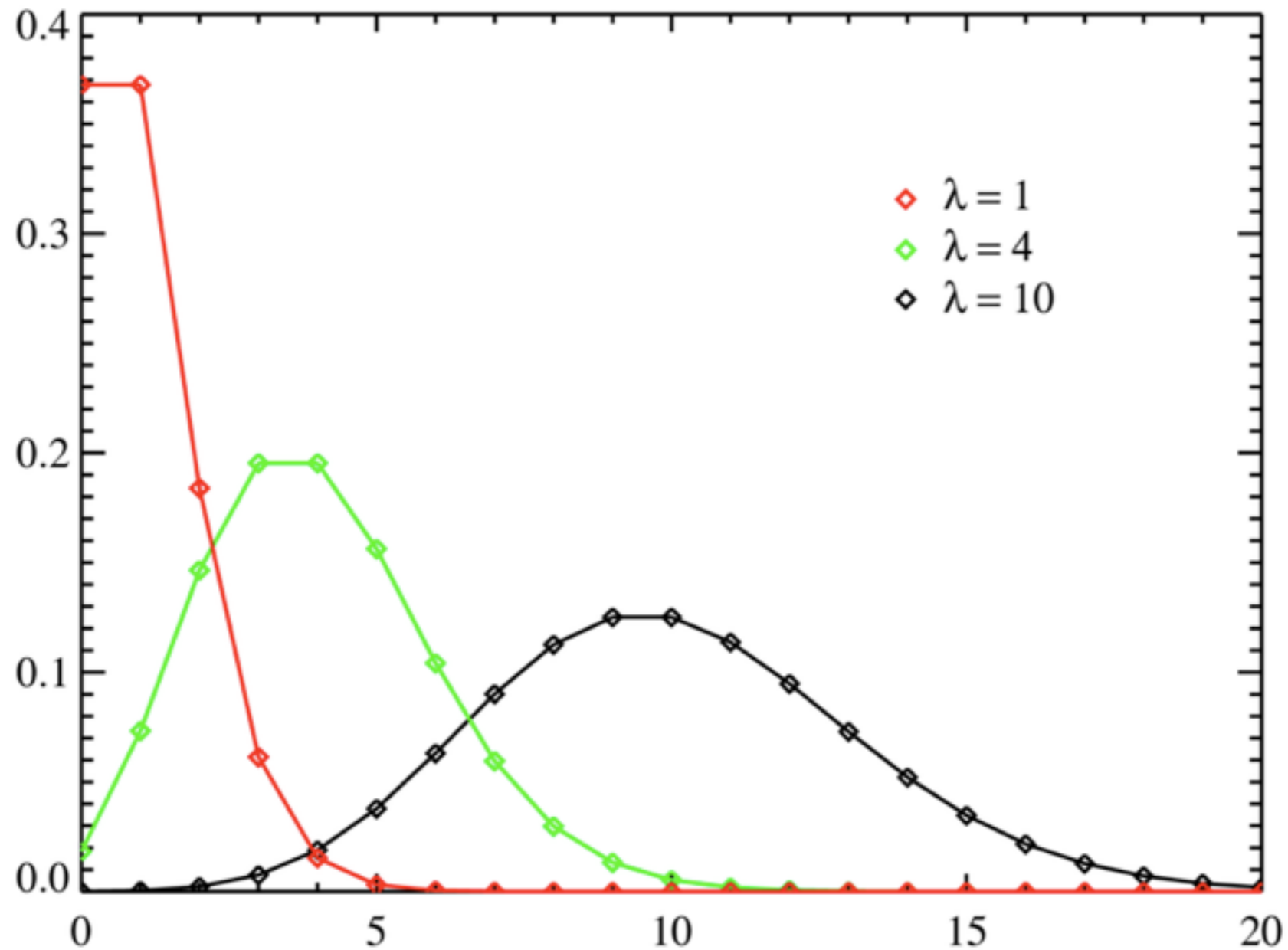
$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots$$

$$E(N(t)) = Var(N(t)) = \lambda t$$

Poissonovo rozdělení

$$P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

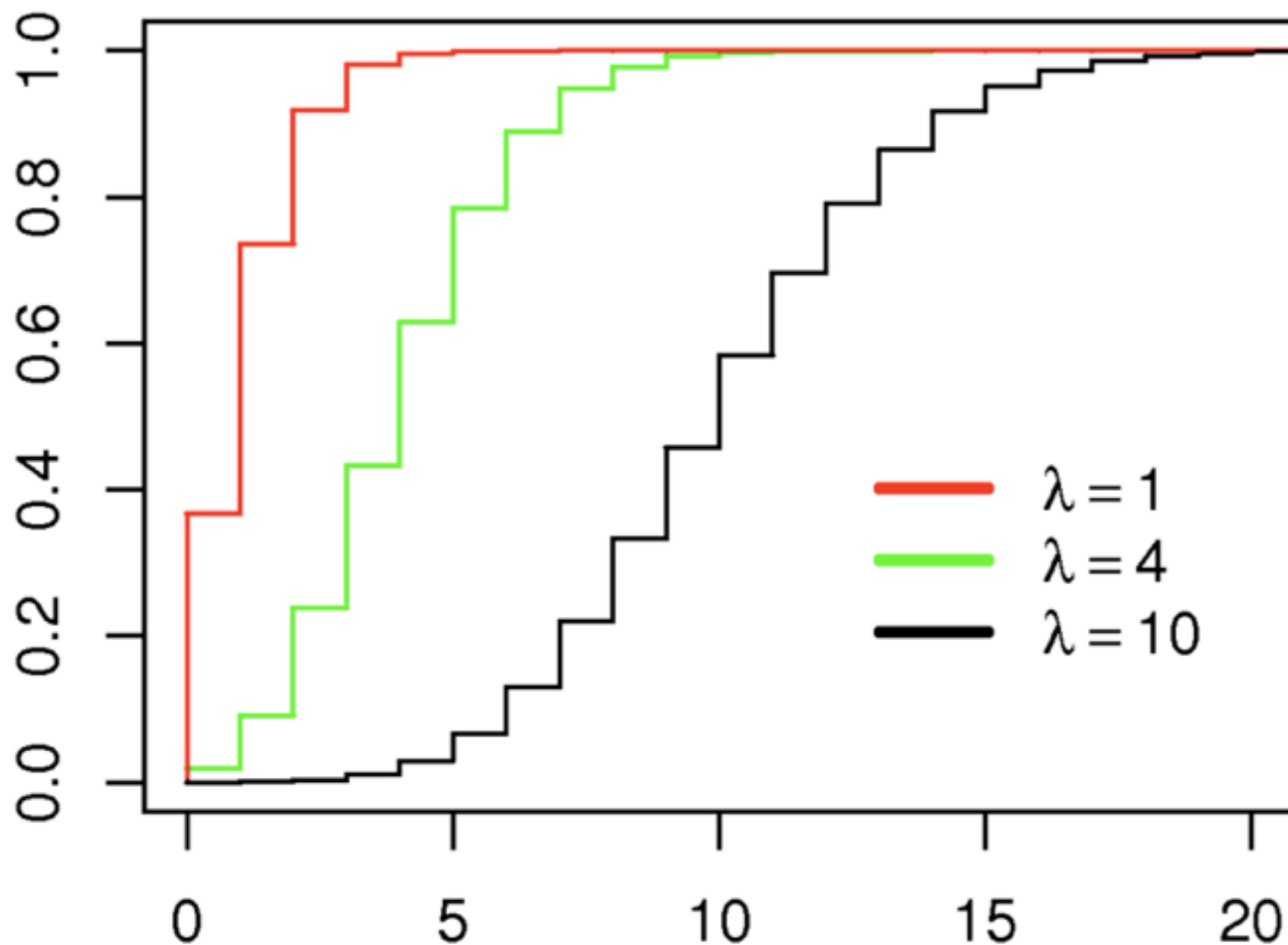
$$E(N) = \lambda, \quad \text{Var}(N) = \lambda$$



Poissonovo rozdělení

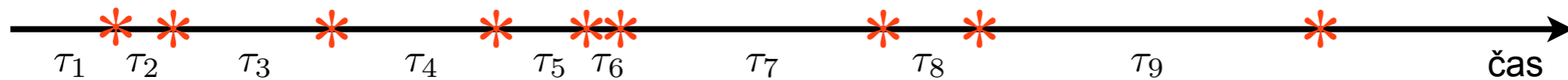
$$P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

$$E(N) = \lambda, \quad \text{Var}(N) = \lambda$$



Exponenciální rozdělení

Poissonův proces $\{N(t), t \geq 0\}$



doby mezi událostmi: $\tau_i = t_i - t_{i-1}$

Jaká je pravděpodobnost, že doba mezi událostmi nepřekročí hodnotu t ? $P(\tau \leq t) = ?$

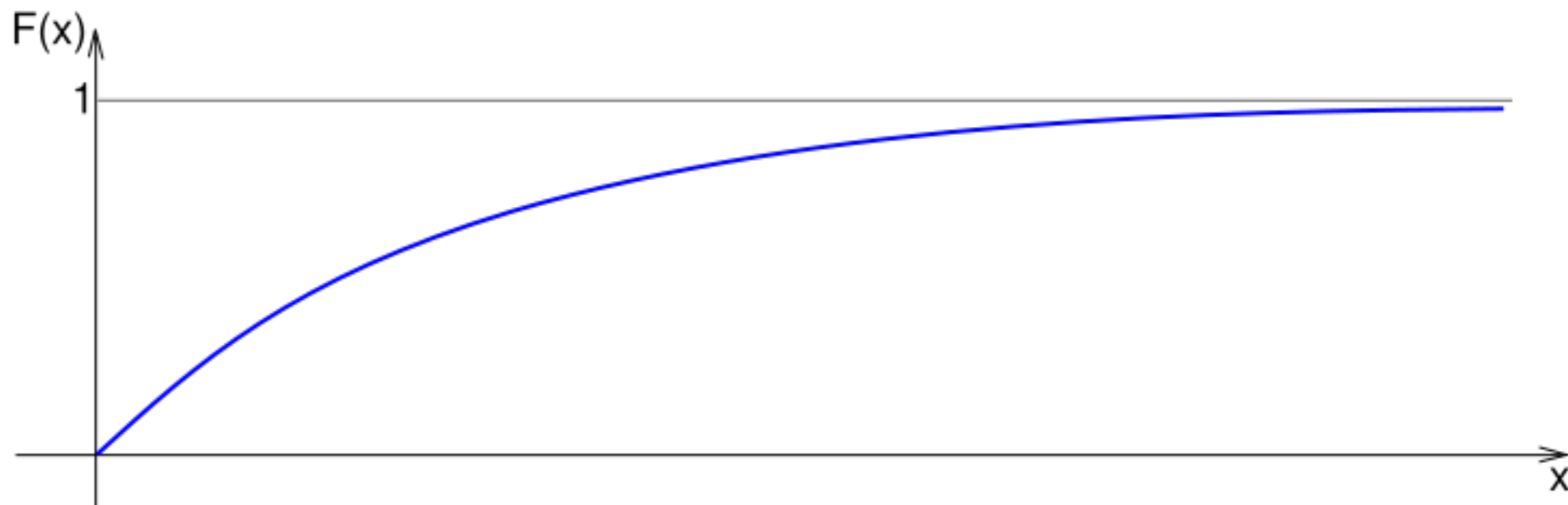
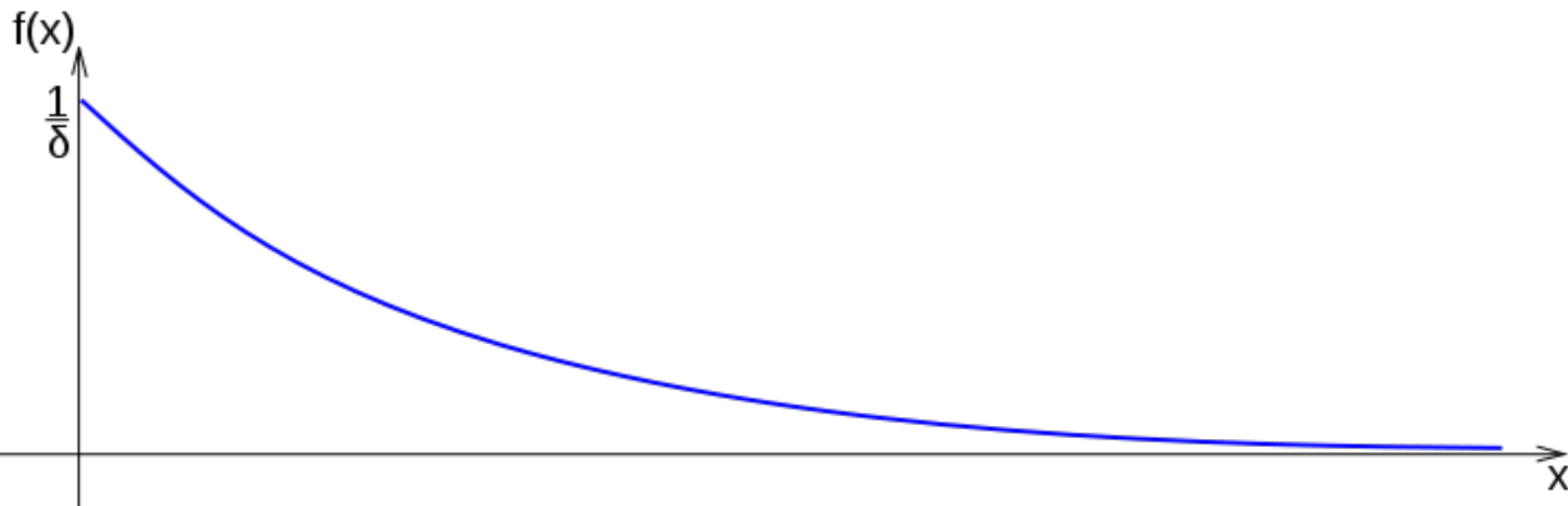
$$P(\tau \leq t) = 1 - e^{-\frac{t}{\delta}}, \quad t \geq 0$$

$$E(\tau) = \delta, \quad Var(\tau) = \delta^2$$

$$F(t) = 1 - e^{-\frac{t}{\delta}} \Rightarrow f(t) = \frac{dF}{dt} = \frac{1}{\delta} e^{-\frac{t}{\delta}}, \quad t \geq 0$$

Exponenciální rozdělení

$$f(t) = \begin{cases} 0, & t < 0 \\ \frac{1}{\delta} e^{-\frac{t}{\delta}}, & t \geq 0 \end{cases} \quad F(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\frac{t}{\delta}}, & t \geq 0 \end{cases}$$



Exponenciální rozdělení

$$f(t) = \begin{cases} 0, & t < 0 \\ \frac{1}{\delta} e^{-\frac{t}{\delta}}, & t \geq 0 \end{cases} \quad F(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\frac{t}{\delta}}, & t \geq 0 \end{cases}$$

$$\frac{1}{\delta} = \lambda$$

$$f(t) = \begin{cases} 0, & t < 0 \\ \lambda e^{-\lambda t}, & t \geq 0 \end{cases} \quad F(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\lambda t}, & t \geq 0 \end{cases}$$

Exponenciální rozdělení je rozdělení pravděpodobnosti náhodné veličiny, která popisuje dobu mezi nezávislými, náhodně se vyskytujícími událostmi v čase.

Přitom střední doba mezi těmito událostmi je rovna δ a střední počet těchto událostí za jednotku času je λ .

Exponenciální rozdělení

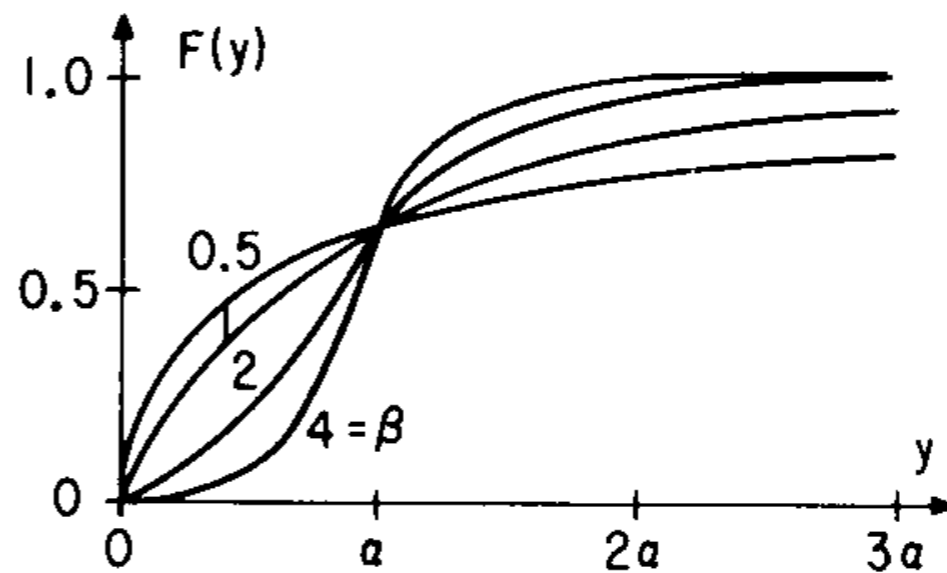
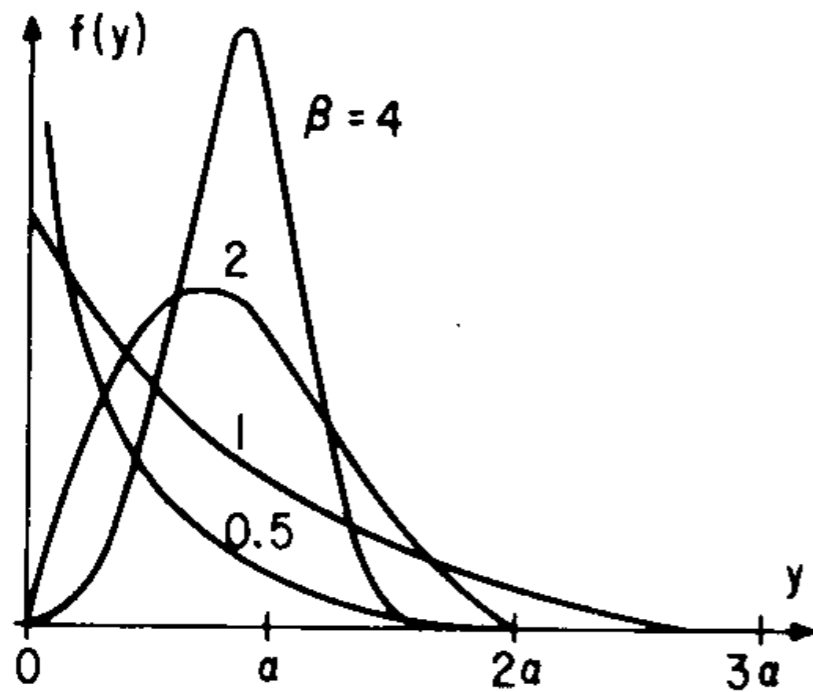
$$\begin{aligned} P(\tau \leq t + s | \tau \geq s) &= \frac{P((\tau \leq t + s) \& (\tau \geq s))}{P(\tau \geq s)} \\ &= \frac{P(\tau \in \langle s, t + s \rangle)}{P(\tau \geq s)} = \frac{F(t + s) - F(s)}{1 - F(s)} \\ &= \frac{1 - e^{-\frac{t+s}{\delta}} - 1 + e^{-\frac{s}{\delta}}}{e^{-\frac{s}{\delta}}} = \frac{e^{-\frac{s}{\delta}} - e^{-\frac{t}{\delta}} e^{-\frac{s}{\delta}}}{e^{-\frac{s}{\delta}}} \\ &= 1 - e^{-\frac{t}{\delta}} = F(t) = P(\tau \leq t) \end{aligned}$$

Exponenciální rozdělení nemá paměť!

Weibullovo rozdělení

$$F(t) = 1 - e^{-\left(\frac{t}{\delta}\right)^\beta}, \quad t \geq 0$$

$$f(t) = \left(\frac{t}{\delta}\right)^{\beta-1} \frac{\beta}{\delta} e^{-\left(\frac{t}{\delta}\right)^\beta}, \quad t \geq 0$$

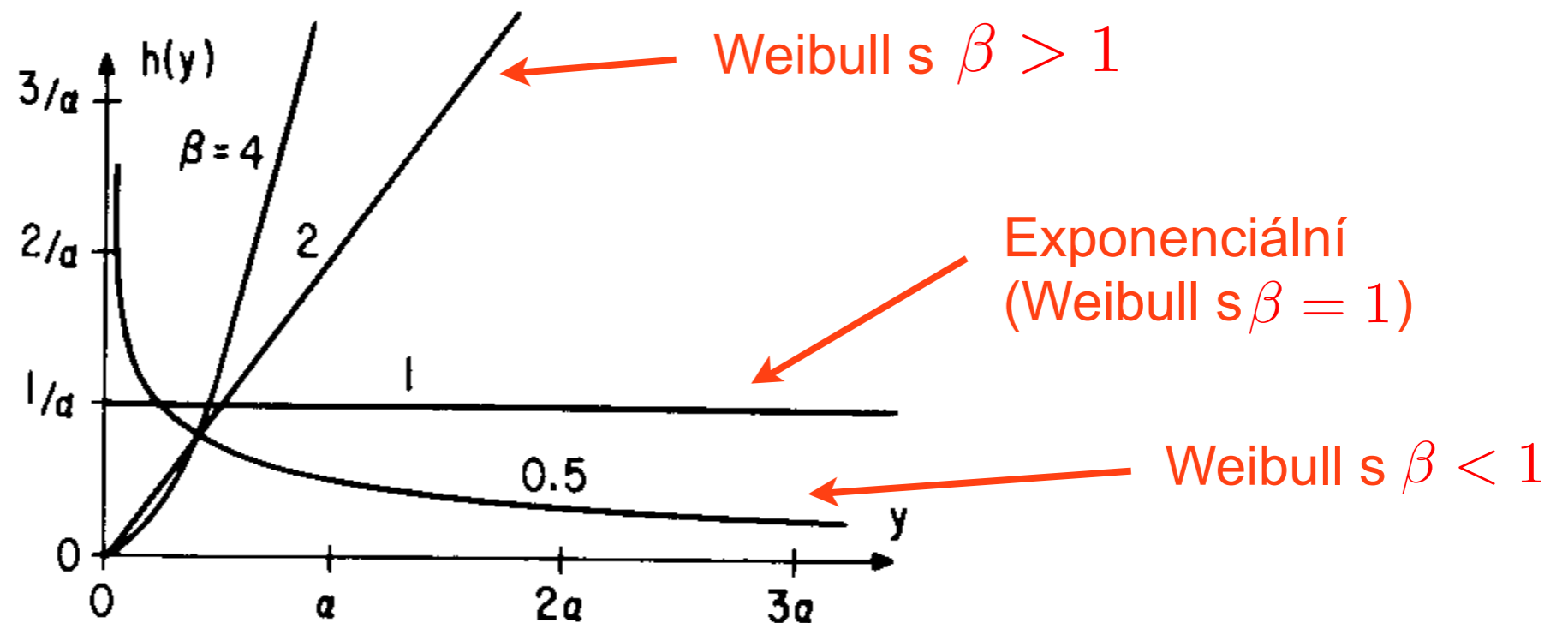


$$E(T) = \delta \Gamma\left(\frac{1}{\beta} + 1\right), \quad Var(T) = \delta^2 \left\{ \Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right) \right]^2 \right\}$$

Intenzita poruchy

$$\lambda(t) = \frac{f(t)}{1 - F(t)}, \quad t \in R$$

Pravděpodobnost, že se zařízení porouchá v čase $t + \Delta$, když víme, že se do doby t neporouchalo (pro hodně malá Δ) je rovna přibližně $\Delta \lambda(t)$.

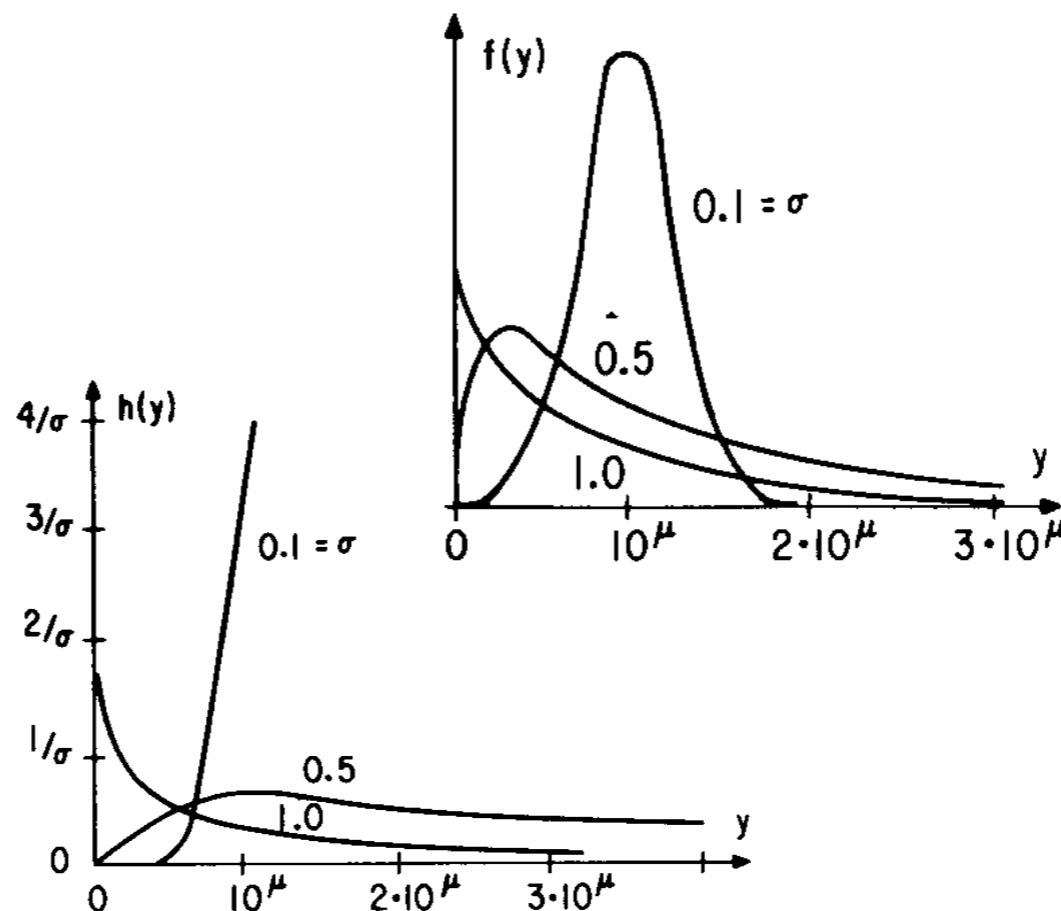
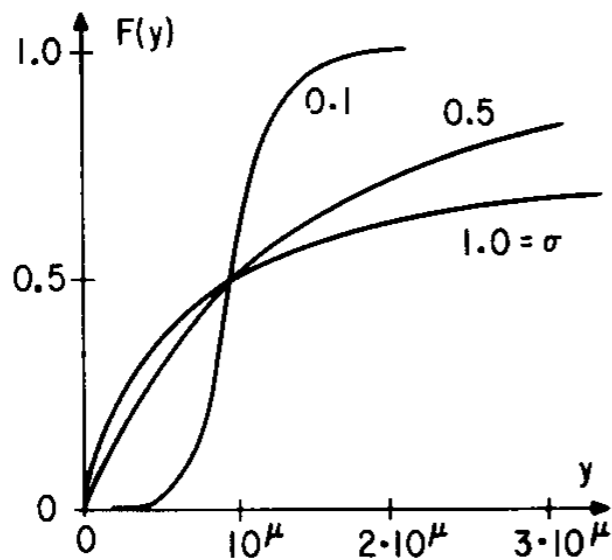


Logaritmicko-normální rozdělení

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), & x \geq 0 \end{cases}$$

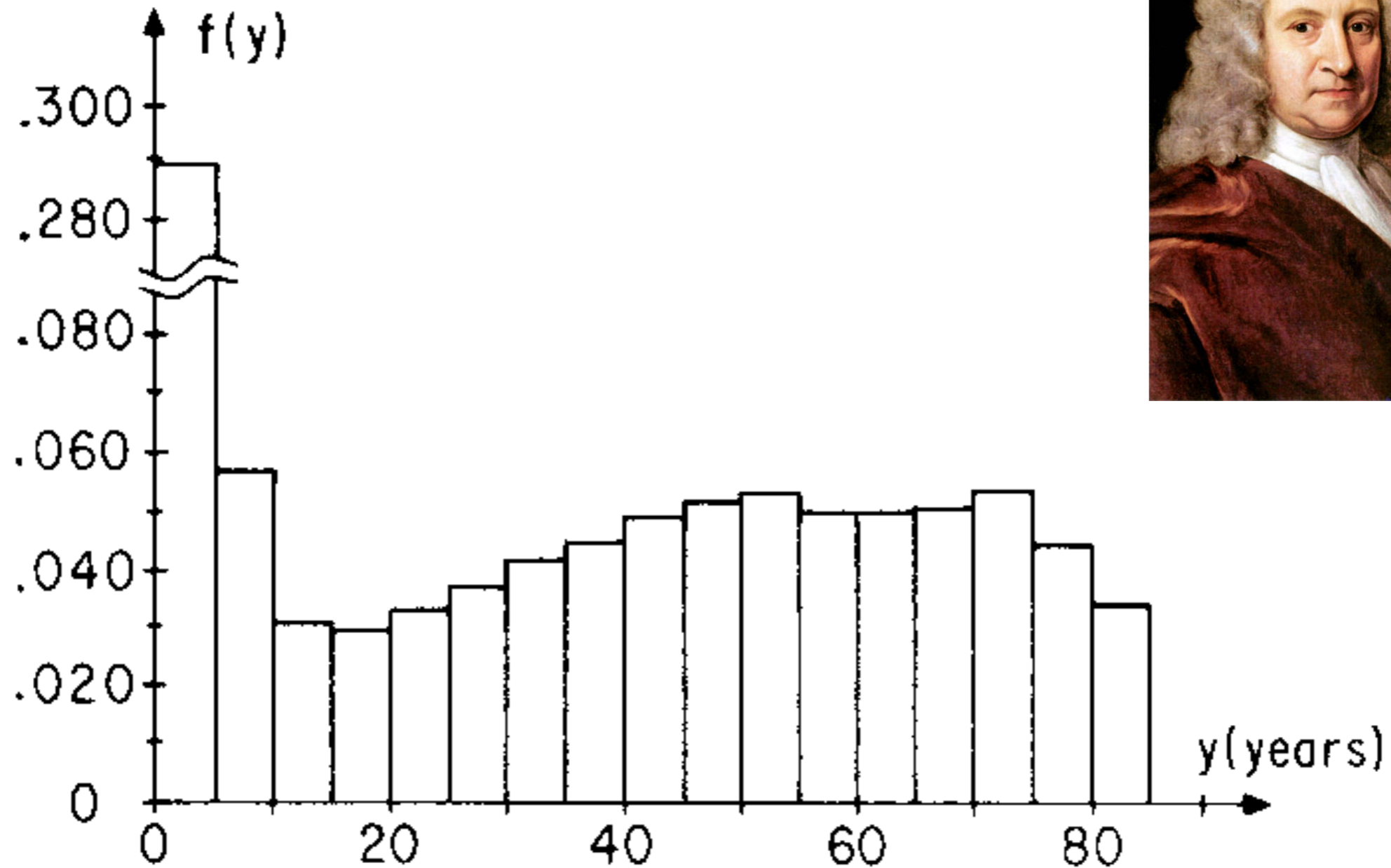
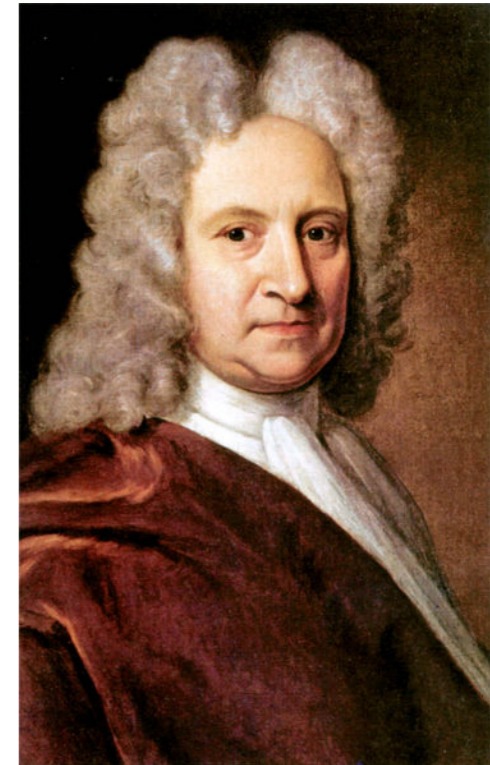
$$\begin{aligned} Y &\sim N(\mu, \sigma^2) \\ X &= \exp(Y) \\ X &\sim LN(\mu, \sigma^2) \end{aligned}$$

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad \text{Var}(X) = \exp(2\mu + \sigma^2) \cdot (\exp \sigma^2 - 1)$$



Charakteristiky doby života

Edmont Halley
(1656 - 1742)



Charakteristiky doby života

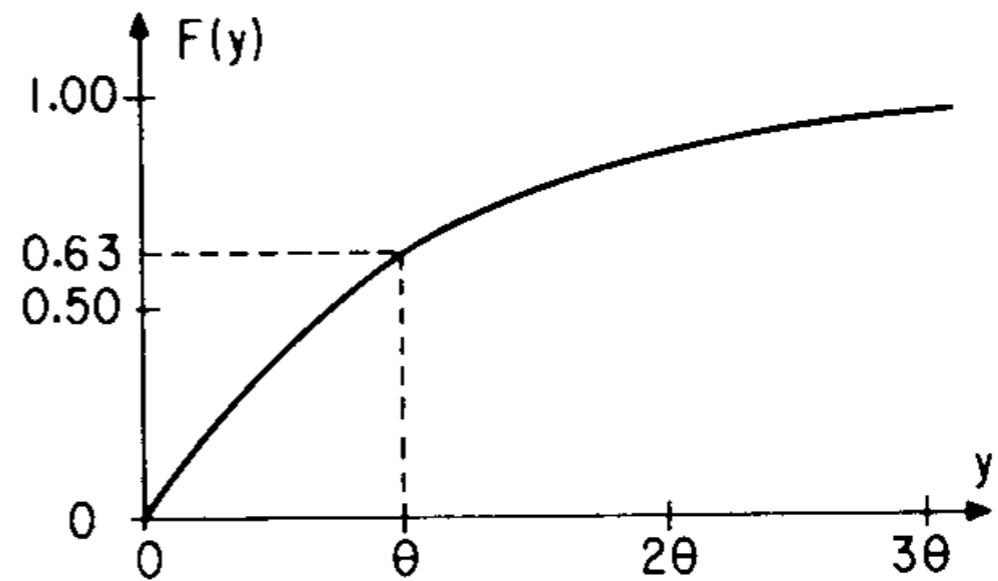
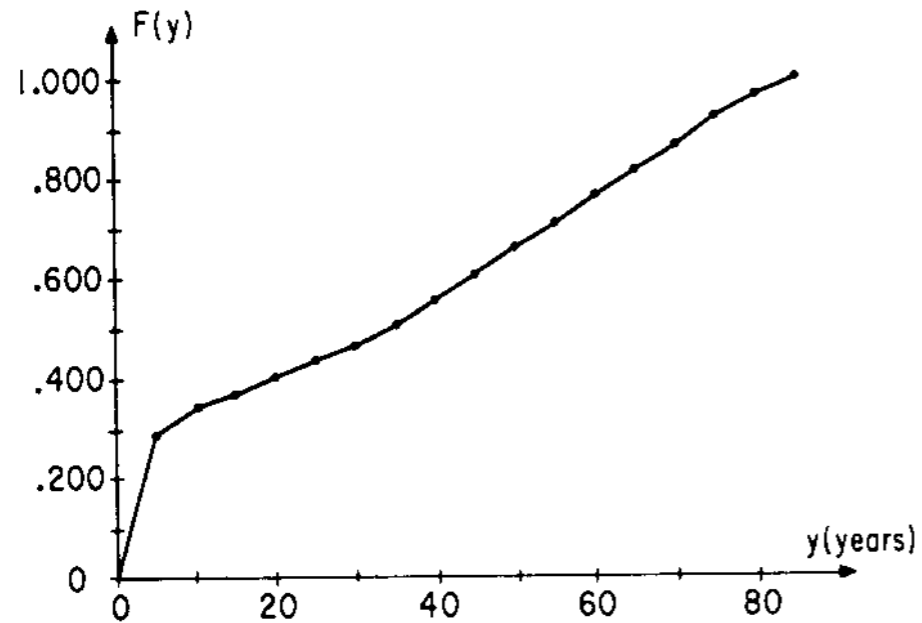
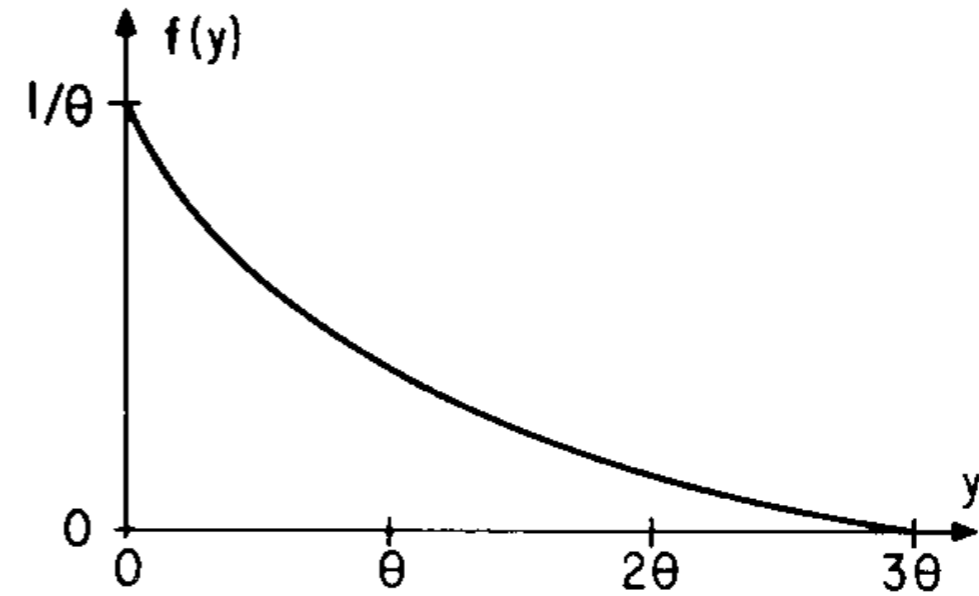
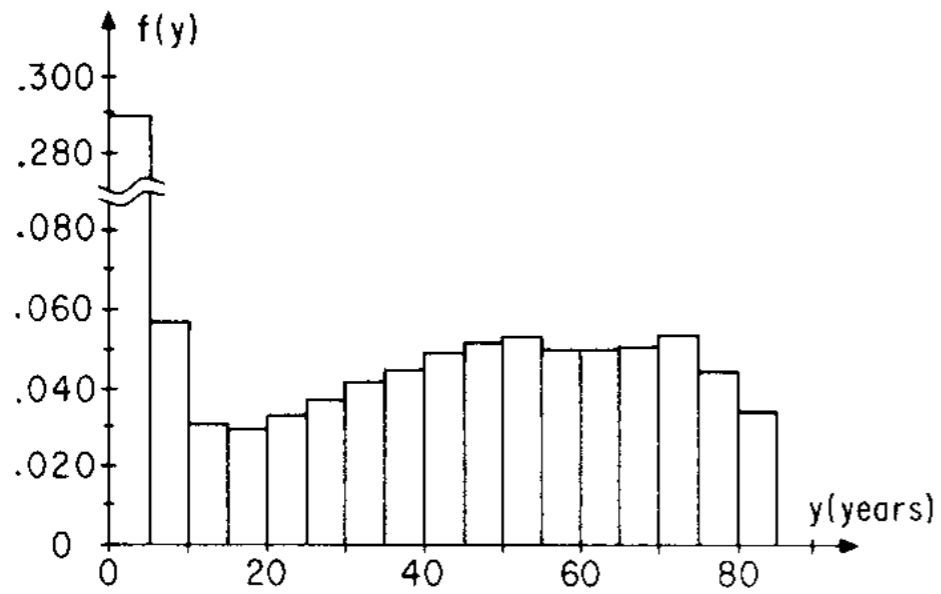
<u>y</u>	<u>f(y)</u>	<u>F(y)</u>	<u>R(y)</u>	<u>h(y)</u>
0	-	0	1.000	
0-5	.290	.290	.710	.058
5-10	.057	.347	.653	.016
10-15	.031	.378	.622	.010
15-20	.030	.408	.592	.010
20-25	.032	.440	.560	.011
25-30	.037	.477	.523	.013
30-35	.042	.519	.481	.016
35-40	.045	.564	.436	.019
40-45	.049	.613	.387	.022
45-50	.052	.665	.335	.027
50-55	.053	.718	.282	.032
55-60	.050	.768	.232	.035
60-65	.050	.818	.182	.043
65-70	.051	.869	.131	.056
70-75	.053	.922	.078	.081
75-80	.044	.966	.034	.113
80-85	.034	1.000	0	.200

Tabulka délky života lidí (Halley, 1693)

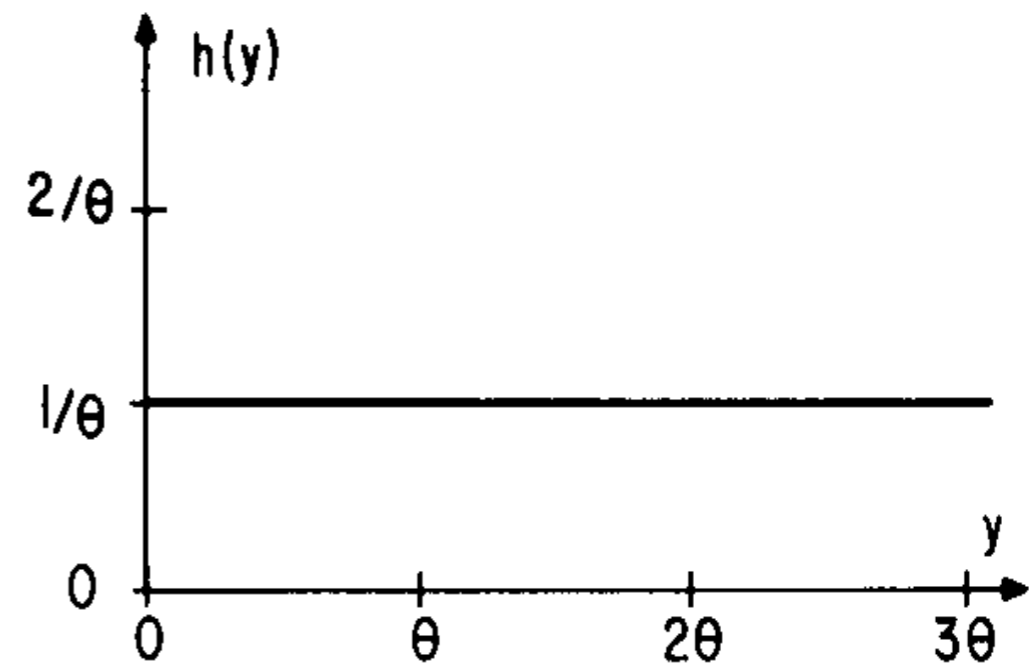
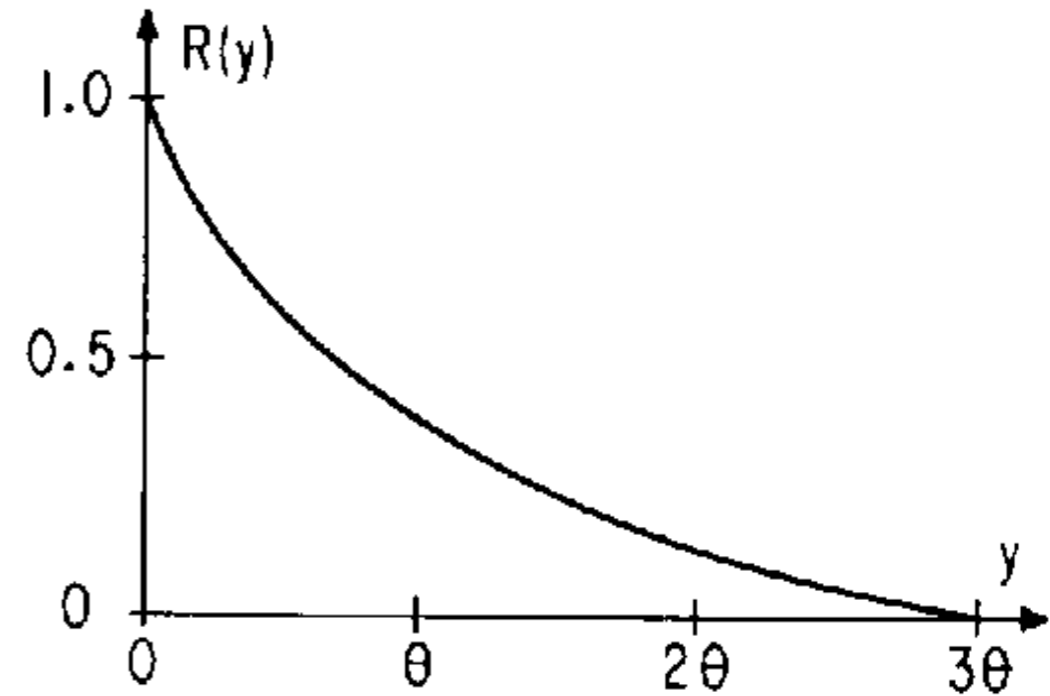
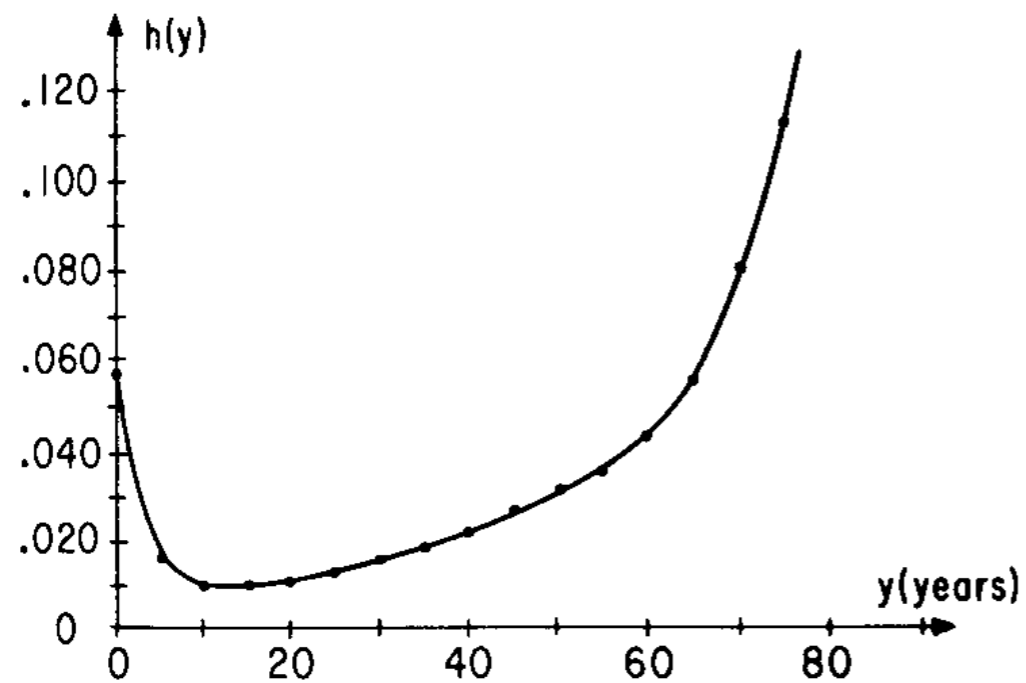
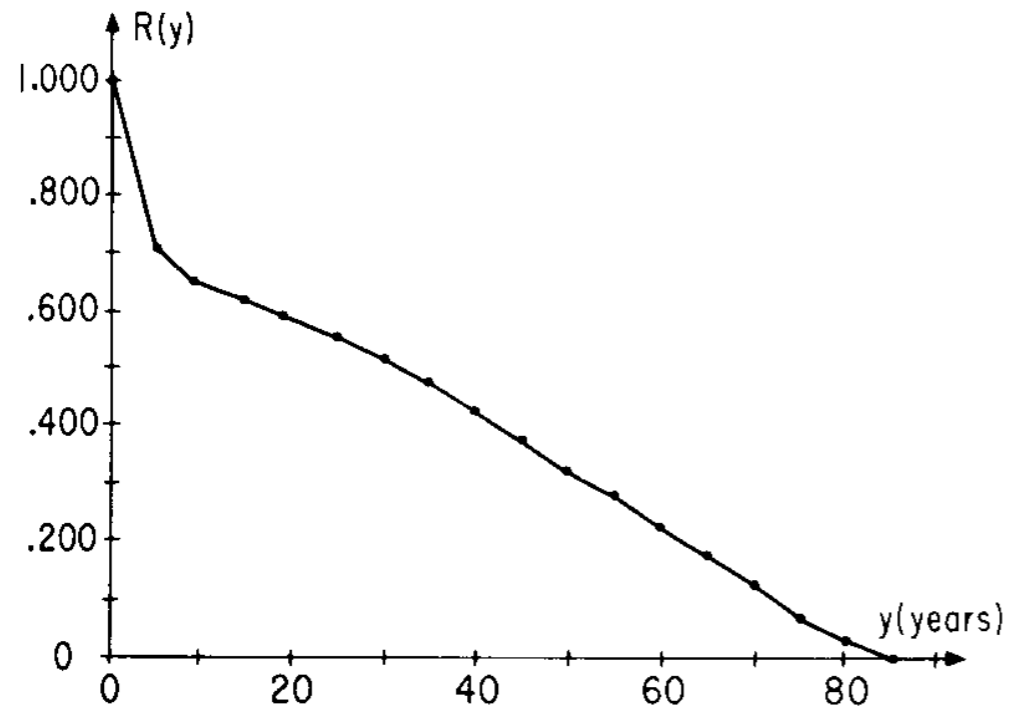


Edmont Halley
(1656 - 1742)

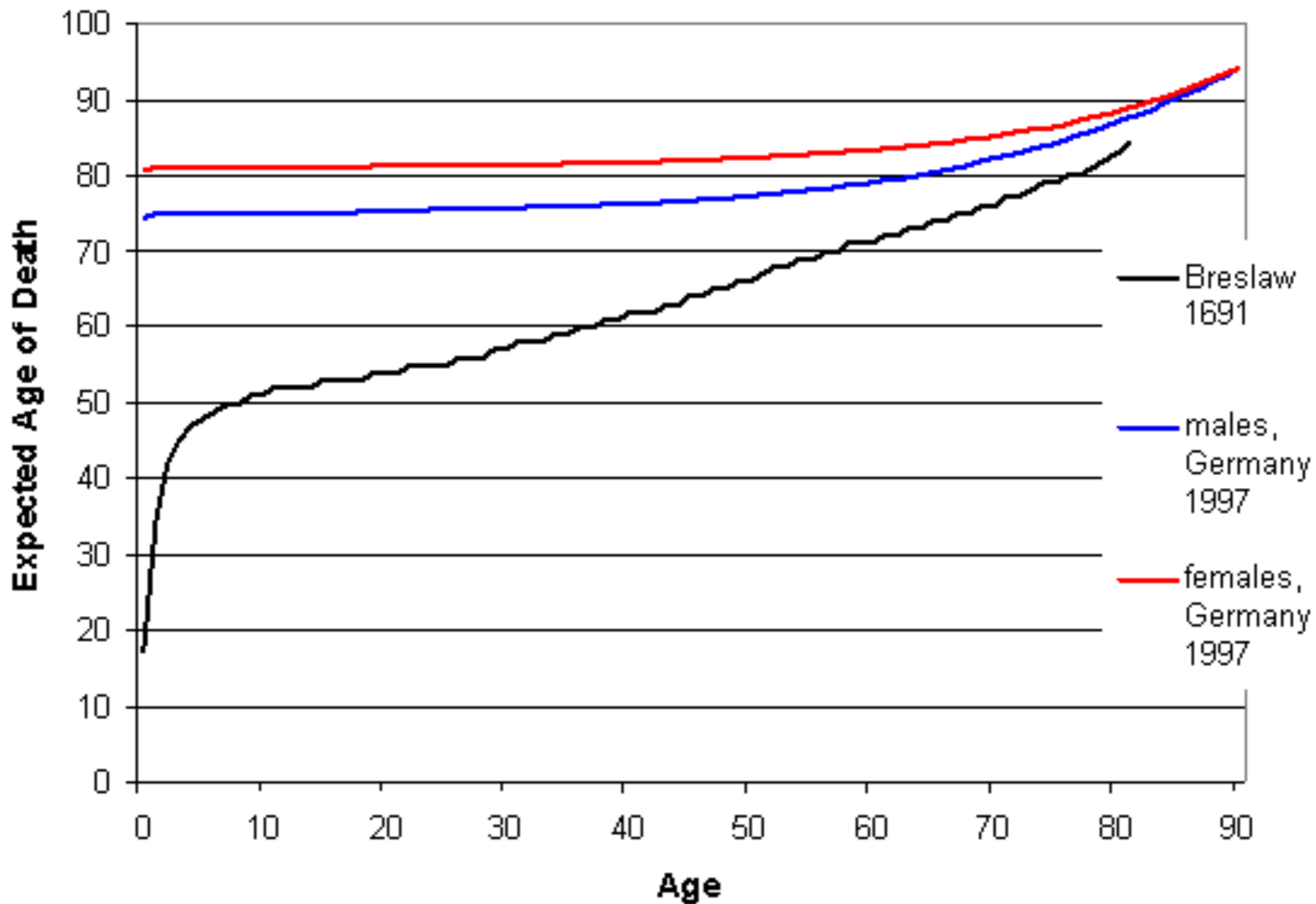
Charakteristiky doby života



Charakteristiky doby života



Charakteristiky doby života



Data from Edmond Halley's table p.600 and "Durchschnittliche Lebenserwartung im Alter x in Jahren (ex) Deutschland, nach Geschlecht, Sterbetafel 1997/99" made available by [Statistisches Informationssystem GeroStat](#).

Charakteristiky doby života

Příklad: Životnost ventilátoru dieselového motoru můžeme popsat pomocí exponenciálního rozdělení se střední dobou $\theta = 28.700$ hod.

1. S jakou pravděpodobností se ventilátor porouchá během prvních 100 hodin?

$$P(X \leq 100) = \int_0^{100} \frac{1}{28700} \exp\left(-\frac{x}{28700}\right) dx = 1 - e^{-\frac{100}{28700}} = 0,003478$$

2. Jaká je pravděpodobnost, že ventilátor vydrží bez poruchy záruční dobu 8.000 hodin?

$$P(X > 8000) = \int_{8000}^{\infty} \frac{1}{28700} \exp\left(-\frac{x}{28700}\right) dx = e^{-\frac{8000}{28700}} = 0,76$$

3. Jaká je intenzita poruchy?

Intenzita poruchy je potom $\lambda = 1/\theta = 1/28.700 = 34,8$ ppm.

$$P(X \leq 100) \cong \lambda \cdot 100 = 100 \cdot 0,00003484 = 0,003484$$

Charakteristiky doby života

Příklad: Životnost ventilátoru dieselového motoru můžeme popsat pomocí exponenciálního rozdělení se střední dobou $\theta = 28.700$ hod.

4. Jaká je „typická“ délka života ventilátoru?

$$\text{median} = \tilde{x}_{0,5} = -28.700 \ln(1 - 0.5) = 19.900$$

5. Jaká je střední doba života ventilátoru?

$$E(X) = \theta = 28.700$$

6. Do jaké doby se porouchá v průměru 90% všech ventilátorů?

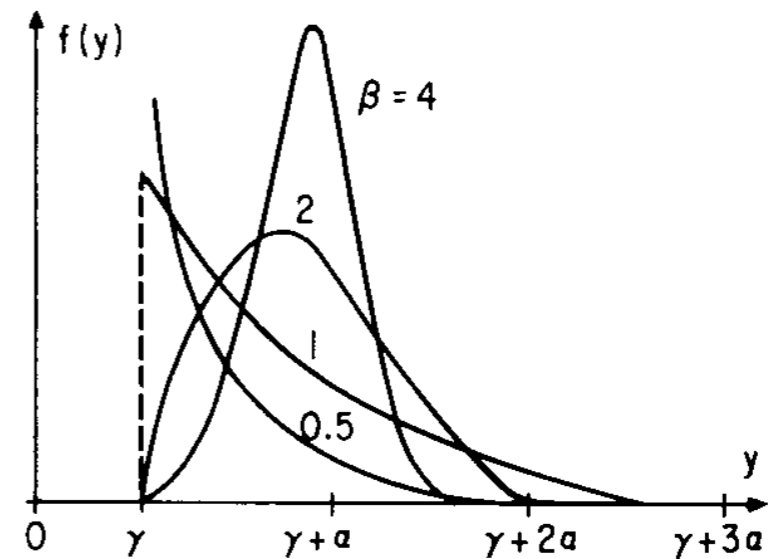
$$\tilde{x}_{0,9} = -28.700 \ln(1 - 0.9) = 66.084$$

Charakteristiky doby života

Posunutá rozdělení

Weibullovo rozdělení

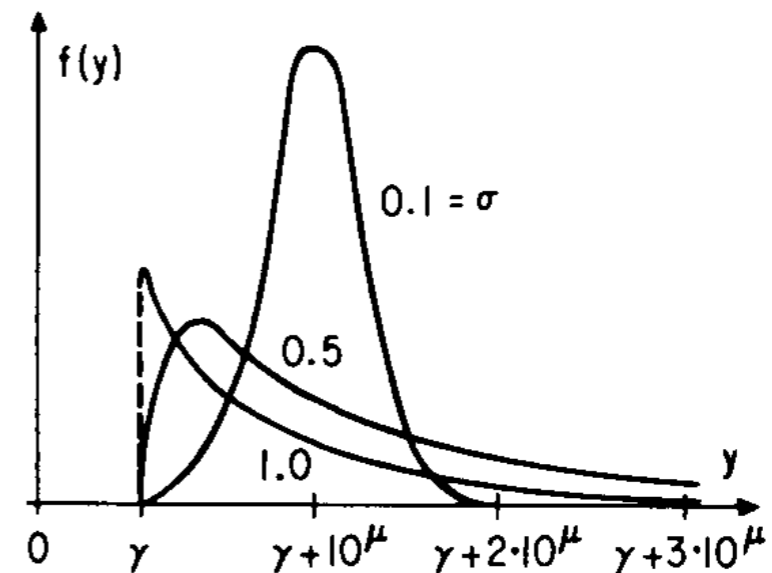
$$f(x) = \left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} \frac{\beta}{\alpha} \exp\left(-\left[\frac{x-\gamma}{\alpha}\right]^{\beta}\right), \quad \gamma \leq x$$



Logaritmicko-normální rozdělení

$$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x-\gamma)-\mu)^2}{2\sigma^2}\right),$$

$$\gamma \leq x$$



Charakteristiky doby života

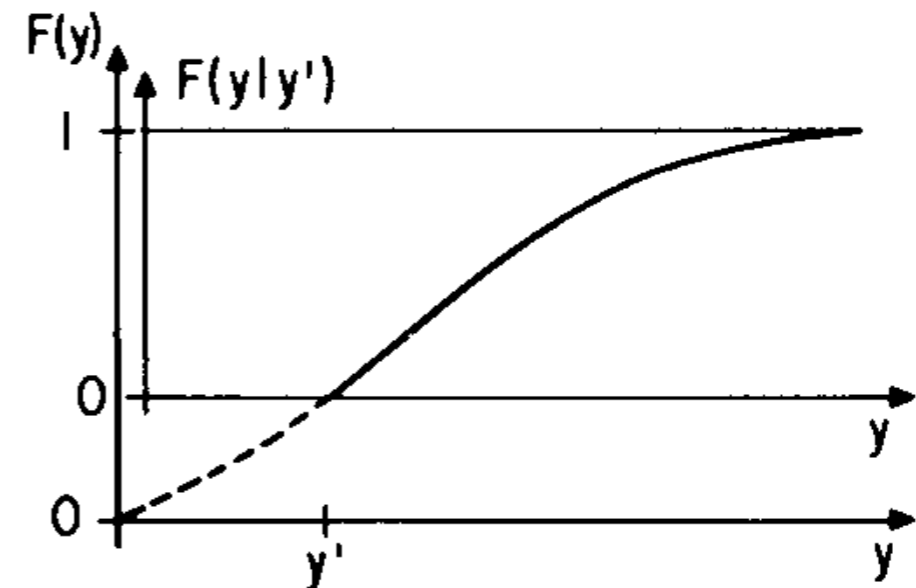
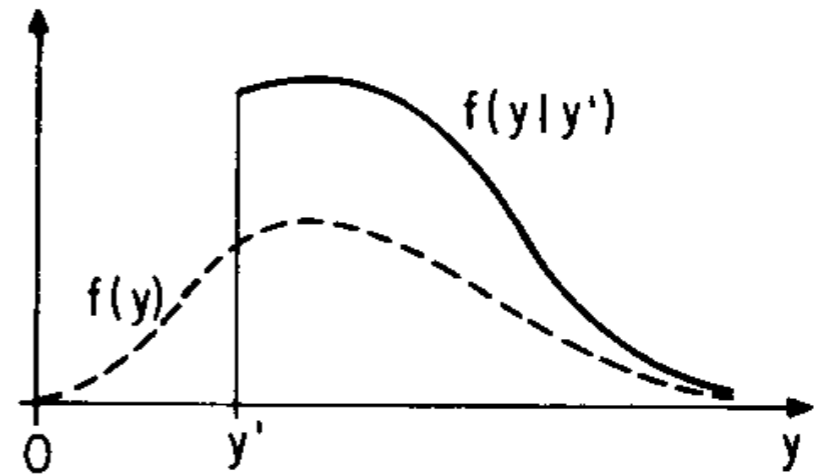
Podmíněná rozdělení

rozdělení délky života víme-li, že se zařízení dožilo doby y'

$$f(x | y') = \frac{f(x)}{R(y')}, \quad y' \leq x$$

$$F(x | y') = \frac{F(x) - F(y')}{R(y')}, \quad y' \leq x$$

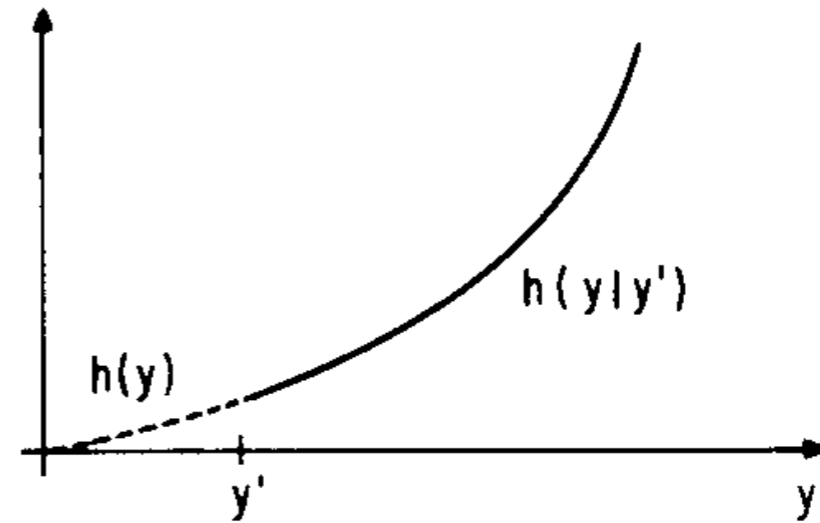
$$R(x | y') = \frac{R(x)}{R(y')}, \quad y' \leq x$$



Charakteristiky doby života

$$h(x | y') = h(x), \quad y' \leq x$$

$$E(X | y') = \frac{1}{R(y')} \int_{y'}^{\infty} x f(x) dx, \quad y' \leq x$$



Zbývající doba života Z

$$G(z | y') = \frac{F(z + y') - F(y')}{R(y')}, \quad 0 \leq z$$

$$g(z | y') = \frac{f(z + y')}{R(y')}, \quad 0 \leq z \quad h(z | y') = h(z + y'), \quad 0 \leq z$$