

KONZISTENCE NEPARAMETRICKÉHO BAYESOVSKÉHO ODHADU

Michal Friesl

*Katedra matematiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni*

Obsah

Pokračování příspěvků o neparametrickém bayesovském odhadu v Koziolově-Greenově modelu náhodného cenzorování

Na minulých Robustech

- odhad za zjednodušených předpokladů
- porovnání se standardními odhady
- model s cenzorováním zleva

Dnešní program

- připomenutí modelu
- konzistence neparametrických bayesovských odhadů obecně
- konzistence odhadu v Koziolově-Greenově modelu

Model náhodného cenzorování

Doba života X (do poruchy, události, ...)

- nezáporná n.v.
- s funkcí spolehlivosti $S(t) = 1 - F(t) = P(X > t)$, $t \geq 0$

Pozorování může být zprava cenzorováno časem Y

- pozorujeme jen $Z = X \wedge Y$ a $I = I_{[X \leq Y]}$
- Y je nezávislá s X

Cíl

- z náhodného výběru dvojic (Z, I)
- odhadnout funkci spolehlivosti S doby života X

Proporcionalita cenzorování

Dodatečný předpoklad (Koziolův-Greenův)

- rozdělení cenzoru Y souvisí s rozdělením X
- pro nějaké $\gamma > 0$

$$S_Y(t) = (S(t))^\gamma, \quad t \geq 0,$$

- neboli $\Lambda_Y(t) = \Lambda(t) \cdot \gamma$ pro kumulativní intenzity

$$\Lambda(t) = -\ln S(t) = \int_0^t f(x)/S(x) \, dx$$

Ve spojitém případě

- zda cenzorováno, tj. I , nezávisí na pozorovaném čase Z
- pravděpodobnost cenzorování $p = P(I = 1) = 1/(1 + \gamma)$

Neparametricky bayesovsky

Tvar $S(t)$ není předem dán

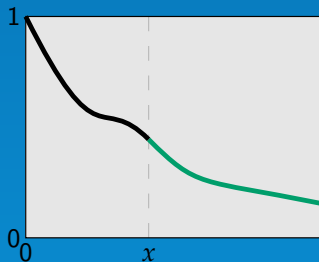
- parametrem nikoli jedna hodnota (dvě, tři, ...)
- nýbrž celá $(S(t), t \geq 0)$
- apriorní rozdělení pro $S = (S(t), t \geq 0)$

Zprava neutrální proces

- $\wedge$ neklesající s nezávislými přírůstky
- tj. pro každé x je $(S(t), t \leq x)$
nezávislé s $(S(t)/S(x), t > x)$

Má-li $\wedge$ jen náhodnou složku

- s. j. je $S \leftrightarrow$ diskrétní rozdělení



Apriorní rozdělení

Pro intenzitu Λ

- gama proces $\Lambda(s, t) = \Lambda(t) - \Lambda(s) \sim G(n_0, n_0 \Lambda_0(s, t))$
- centrován kolem $E \Lambda(s, t) = \Lambda_0(s, t)$
- resp. $E S(t) = E e^{-\Lambda(t)} = (1 + 1/n_0)^{-n_0 \Lambda_0(t)} \approx e^{-\Lambda_0(t)}$
- n_0 řídí rozptyl $\text{var } \Lambda(s, t) = \Lambda_0(s, t)/n_0$

Koeficient γ

- hustota $\pi(\gamma)$ vzhledem k nějaké míře na $(0, \infty)$
- nezávislý s Λ

Přeznačení

Pozorované časy

- $Z_1, \dots, Z_n$ uspořádáme ...
- navzájem různé časy $0 = T_0 < T_1 < \dots < T_N < T_{N+1} = \infty$

V čase T_j

- $U_j = \#$ necenzorovaných pozorování se $Z_k = T_j$
- $C_j = \#$ cenzorovaných pozorování se $Z_k = T_j$
- $N_j = \#$ pozorování s $Z_k > T_j$
 $= n - \sum_{i \leq j} (U_i + C_i)$

Pro časy pocházející ze spojitého rozdělení je $U_j + C_j = 1$.

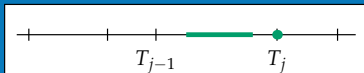
Dle apriorního modelu se ale (i při spojitě Λ_0) časy řídí diskrétním.

Aposteriorní rozdělení Λ

Při daném γ nezávislé přírůstky:

Na intervalech mezi pozorovanými časy

- změna měřítka



$$(\Lambda(s, t) \mid \text{data}, \gamma) \sim G(M_{j-1}(\gamma), n_0 \Lambda_0(s, t))$$

kde $M_j^m(\gamma) = n_0 + N_j(1 + \gamma) + m \quad j = 0, \dots, N$

V okamžicích ukončení pozorování

- vždy skok s hustotou (při daném γ)

$$x^{-1} e^{-(M_j(\gamma) + C_j)x} (1 - e^{-x})^{U_j} (1 - e^{-\gamma x})^{C_j} / c_j(\gamma)$$

kde $c_j(\gamma)$ je normovací konstanta

Aposteriorní rozdělení γ

Hustota

$$\pi(\gamma \mid \text{data}) \propto \left(\prod_{j=1}^N q_j(\gamma) \right) \pi(\gamma)$$

kde

$$q_j^m(\gamma) = (M_{j-1}^m(\gamma))^{-n_0 \Lambda_0(T_{j-1}, T_j)} c_j^m(\gamma), \quad j = 1, \dots, N$$

$$c_j^m(\gamma) = \sum_{k=1}^{U_j} \sum_{\ell=1}^{C_j} (-1)^{k+\ell} \binom{U_j}{k} \binom{C_j}{\ell} \ln \frac{M_j^m(\gamma) + C_j}{M_j^m(\gamma) + C_j + k + \ell \gamma}$$

Odhad

Aposteriorní střední hodnota S

- při daném γ jako součin přes intervaly a skoky

$$\begin{aligned} E(S(t) \mid \text{data}, \gamma) &= E(e^{-\Lambda(t)} \mid \text{data}, \gamma) \\ &= \prod E(e^{-\Lambda(\cdot, \cdot)} \mid \text{data}, \gamma) \end{aligned}$$

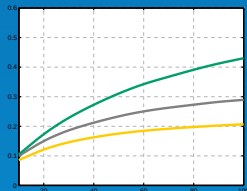
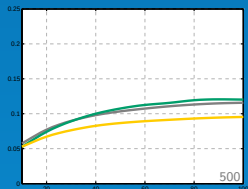
- zprůměrování přes $\pi(\gamma \mid \text{data})$
- pro $t \in \langle T_{i-1}, T_i \rangle$ dostáváme

$$\hat{S}(t) = \frac{\int \left(\prod_{j < i} q_j^1(\gamma) \right) \left(\prod_{j \geq i} q_j(\gamma) \right) \left(\frac{M_{i-1}(\gamma)}{M_{i-1}^1(\gamma)} \right)^{n_0 \Lambda_0(T_{i-1}, s)} \times \pi(\gamma) d\mu(\gamma)}{\int \left(\prod_{j \geq i} q_j(\gamma) \right) \times \pi(\gamma) d\mu(\gamma)}$$

Odhadujeme dobře?

Bayesovské riziko

- $\text{BR } \hat{S}(t) = E(\hat{S}(t) - S(t))^2 \rightarrow 0$
- např. pro $\Lambda_0 \leftrightarrow \text{Exp}(1)$ — $n \text{ BR } \hat{S}(t)$ pro $t_{0,2}$ a $t_{0,5}$



Pro konkrétní S_*

- bude $\hat{S}(t) \rightarrow S_*(t)$ (s.j.)?

Konzistence obecně

Situace

- náhodný výběr $X_1, \dots, X_n$ z $S = S_\theta$
- parametr θ neznáme, apriorní rozdělení $\pi(\theta)$
- aposteriorní rozdělení $\pi_n = \pi(\theta \mid X_1, \dots, X_n)$... odhad $\hat{S}_n$

Konzistence odhadu $\hat{S}_n$, resp. aposteriorního rozdělení π_n

- pokud ve skutečnosti $\theta = \theta_*$, bude $\hat{S}_n \rightarrow S_*$ s.j.
- případně $\pi(S \mid \text{data}) \rightarrow \delta_{S_*}$ s.j.

Pro která θ_* jsou konzistentní

- Doob (1949): pro π -s.v. θ_*
- rádi bychom pro všechna θ_*

Při odhadování rozdělení

Parametrem θ je $S \leftrightarrow \Lambda \leftrightarrow P$

Konečněrozměrný případ

- veličina X s daným konečným oborem hodnot
- S závisí na konečně parametrech, $P \leftrightarrow p_1, \dots, p_k$
- konzistence pro všechna P_* (z nosiče apriorního rozdělení)

Nekonečněrozměrný případ

- problémy už při odhadu diskrétních rozdělení ($P \leftrightarrow p_1, p_2, \dots$)
- Freedman (1963, 1965)

Nekonečněrozměrný diskrétní případ

Pro libovolné rozdělení $\tilde{P} \neq P_*$

- existuje apriorní rozdělení π , že jeho nosič obsahuje P_*
- ale přitom s.j. $\pi_n \rightarrow \delta_{\tilde{P}}$, tedy ne k δ_{P_*}

Př.: při „vhodném“ apriorním rozdělení pro neznámé P

- pozorování generovaná z $P_* = \text{Geom}(1/4)$
- vedou k odhadu $\hat{P}_n$ konvergujícímu k $\tilde{P} = \text{Geom}(3/4)$

Navíc

- dvojic θ_*, π , pro které je θ_n konzistentní, je málo
- ... volbu apriorního rozdělení nutno dobře uvážit

Prakticky

Při praktickém použití požadujeme, aby

- konzistence nastávala sice pro všechna θ_* ,
- ale stačí pro konkrétní zvolené apriorní rozdělení

Pro apriorní rozdělení

- Dirichletovo, tail-free – ano
- konečné směsi konzistentních – ano
- nekonečné směsi konzistentních – nemusí

Spojitý případ

Konzistence pro apriorní rozdělení

- procesy neutrální zprava – nemusí (Kim a Lee, 2001)
- Dirichletův proces – ano
- beta proces – ano
- gama proces – ano
- zobecněný beta – ne ($\hat{S}_n \rightarrow S_*^\alpha$)

V modelu s cenzorováním

- v případě zprava neutrálních procesů právě tehdy, když v modelu bez cenzorování (Dey, Erickson a Ramamoorthi, 2003)
- týká se pouze neinformativního cenzorování

Konzistence odhadu v K.-G. modelu

Předpokládejme, že skutečné parametry modelu jsou γ_* a S_* — funkce spolehlivosti spojitého rozdělení

Pak pozorování nastávají v navzájem různých časech a při daném γ je $E(S(t) \mid \text{data}, \gamma) = [1] \cdot [2]$, kde

$$[1] = \prod_{j < i} \left(1 + \frac{1}{n_{j-1}}\right)^{-n_0 \Lambda_0(T_{j-1}, T_j)} \cdot \left(1 + \frac{1}{n_{i-1}}\right)^{-n_0 \Lambda_0(T_{i-1}, t)} \quad (\text{intervaly})$$

$$[2] = \prod_{j < i} \frac{\ln\left(1 + \frac{U_j + \gamma C_j}{n_j + C_j + 1}\right)}{\ln\left(1 + \frac{U_j + \gamma C_j}{n_j + C_j}\right)} \quad (\text{skoky})$$

$$\text{a } n_j = M_{j-1}(\gamma) = n_0 + N_{j-1}(1 + \gamma)$$

Konzistence odhadu v K.-G. modelu

Předpokládejme, že skutečné parametry modelu jsou γ_* a S_* — funkce spolehlivosti spojitého rozdělení

Pak pozorování nastávají v navzájem různých časech a při daném γ je $E(S(t) \mid \text{data}, \gamma) = [1] \cdot [2]$, kde při $n \rightarrow \infty$

$$[1] = \prod_{j < i} \left(1 + \frac{1}{n_{j-1}}\right)^{-n_0 \Lambda_0(T_{j-1}, T_j)} \cdot \left(1 + \frac{1}{n_{i-1}}\right)^{-n_0 \Lambda_0(T_{i-1}, t)} \rightarrow 1$$

$$[2] = \prod_{j < i} \frac{\ln\left(1 + \frac{U_j + \gamma C_j}{n_j + C_j + 1}\right)}{\ln\left(1 + \frac{U_j + \gamma C_j}{n_j + C_j}\right)} \approx \prod_{j < i} \left(1 - \frac{1}{(1 + \gamma)N_j}\right) \rightarrow (S_*(t))^{\frac{1+\gamma}{1+\gamma_*}}$$

$$\text{a } n_j = M_{j-1}(\gamma) = n_0 + N_{j-1}(1 + \gamma)$$

Chování γ

Aposteriorní hustota γ se soustřeďuje kolem γ_*

Pro $\gamma \neq \gamma_*$ je

$$\frac{\pi(\gamma \mid \text{data})}{\pi(\gamma_* \mid \text{data})} \leq \text{konst} \left(\frac{(1+x)^c}{1+cx} \right)^n \frac{\pi(\gamma)}{\pi(\gamma_*)} \rightarrow 0,$$

$$\text{kde } x = \frac{\gamma - \gamma_*}{\gamma_*} \quad \text{a} \quad c = \frac{\gamma_*}{1 + \gamma_*} = 1 - p_*$$

Příklad 1 — Exp

Nasimulovaná data z

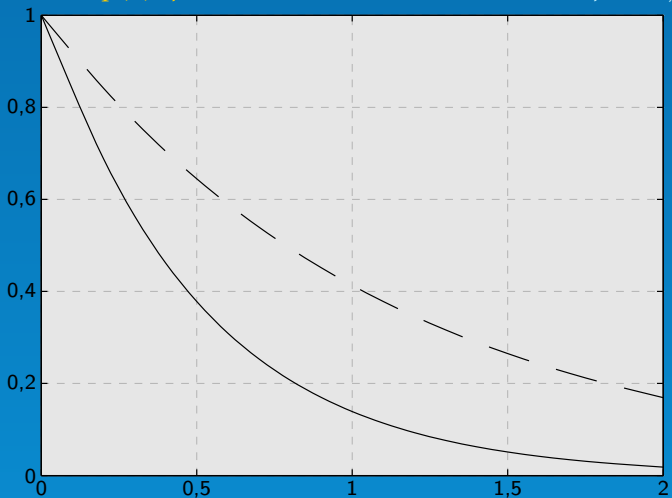
- $S_* \sim \text{Exp}(1)$
- $\gamma_* = 1/p_* - 1$ pro podíl necenzorovaných $p_* = 0,9$
- $n = 10, 20, 50, 100, 200$

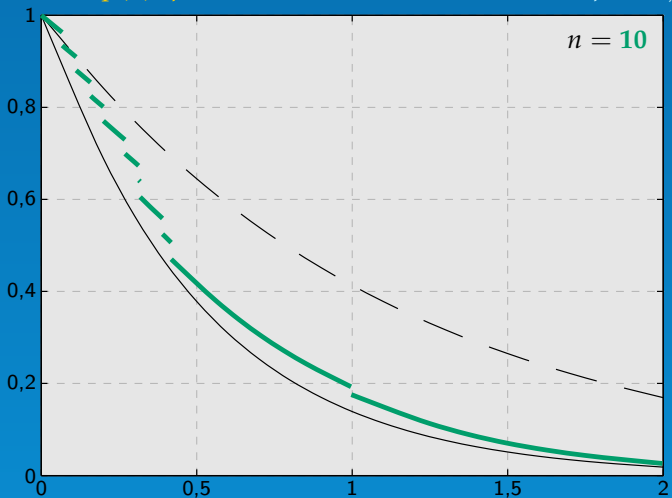
Jako apriorní rozdělení (i v dalších příkladech)

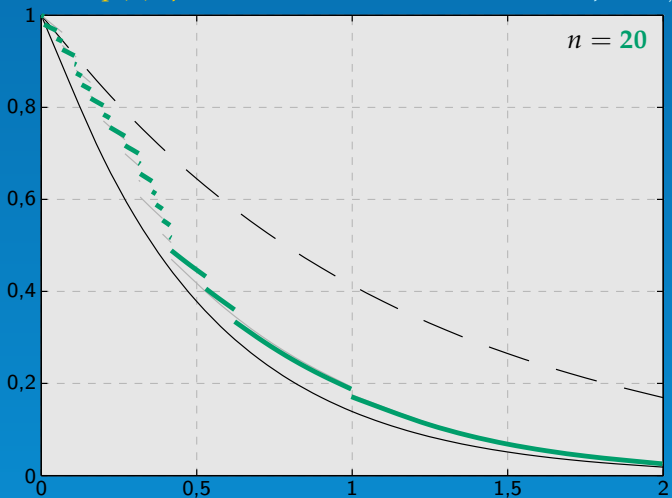
- pro Λ : $\Lambda_0(t) = t, t > 0$ ($\text{Exp}(1)$), $n_0 = 10$
- pro γ : rovnoměrné rozdělení na $1/p - 1$, $p = 0,1; 0,2; \dots; 0,9$

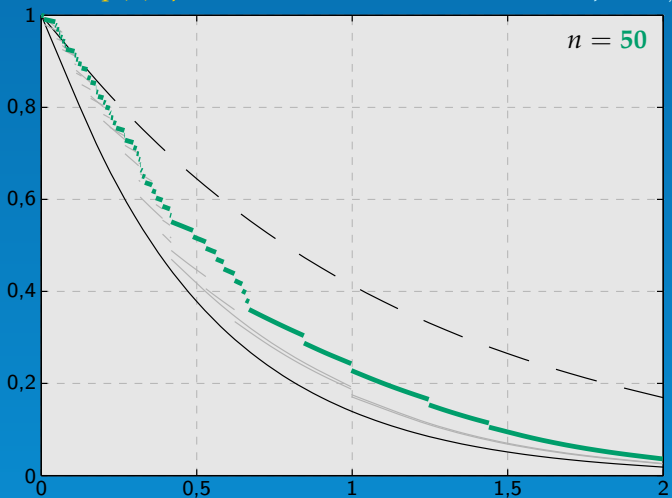
Následují grafy

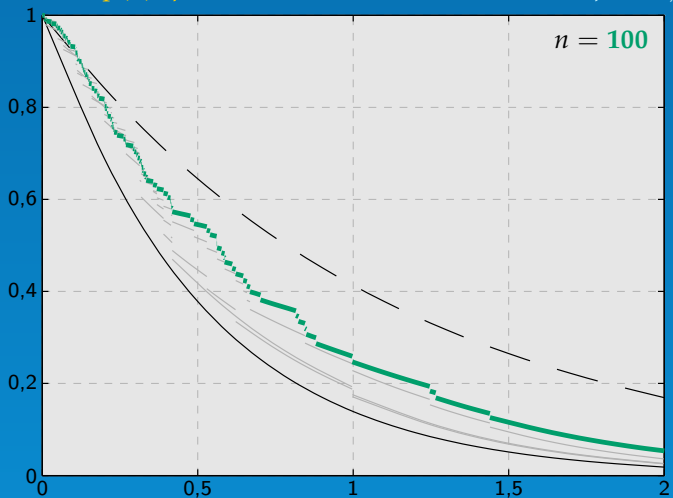
- podmíněný odhad $E(S \mid \text{data}, \gamma)$ pro $\gamma = 1/0,4 - 1$
- aposteriorní $\pi(p \mid \text{data})$

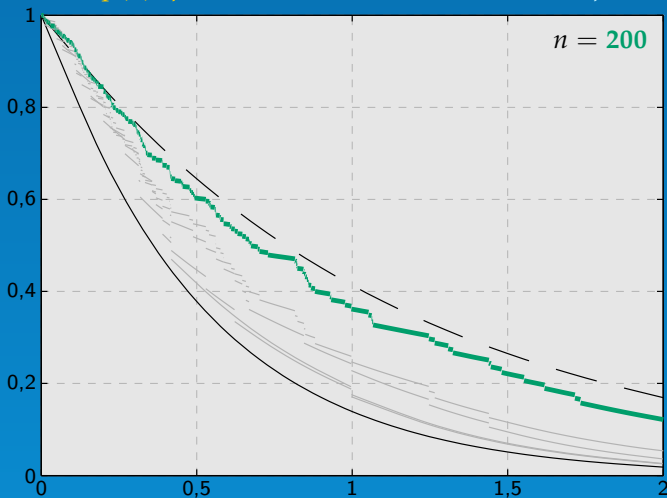
Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$, $p \sim \delta_{0,4}$ 

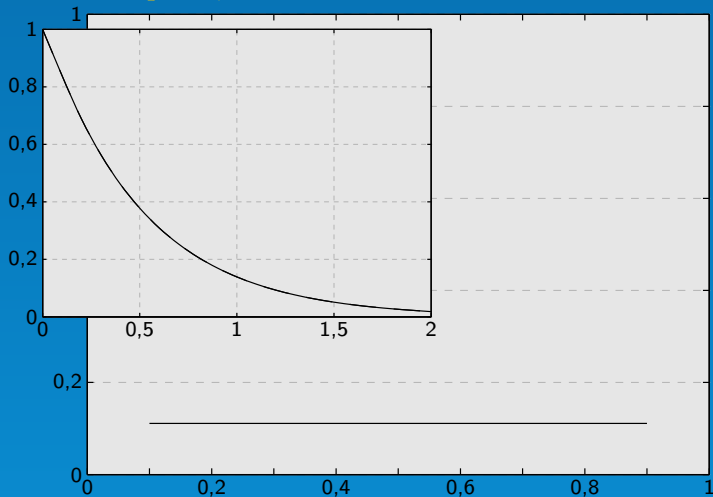
Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$, $p \sim \delta_{0,4}$ 

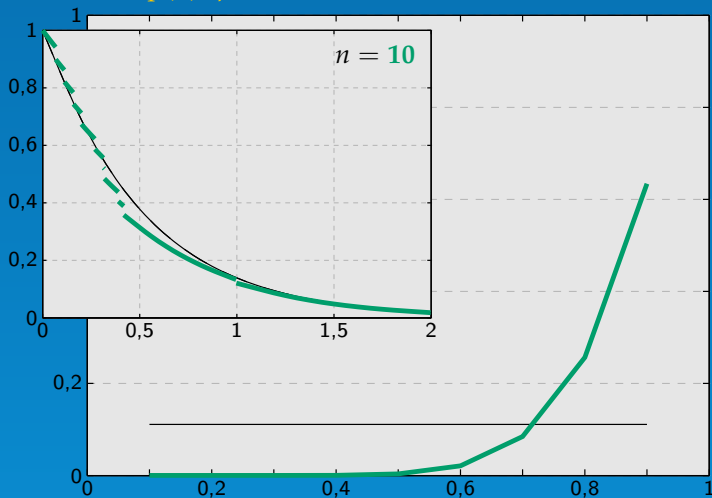
Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$, $p \sim \delta_{0,4}$ 

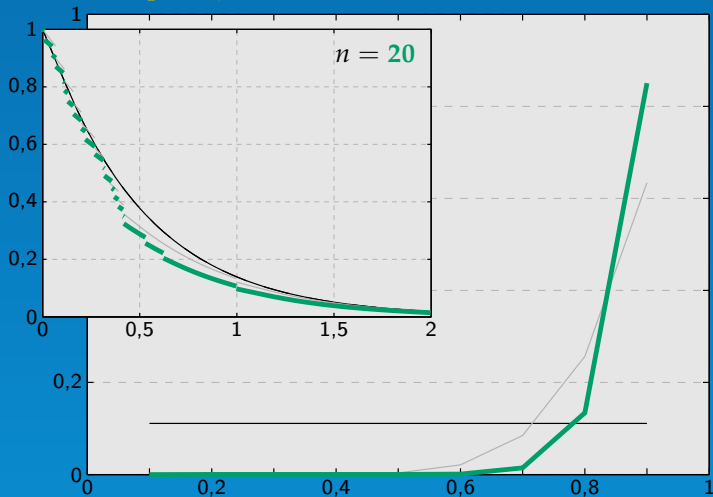
Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$, $p \sim \delta_{0,4}$ 

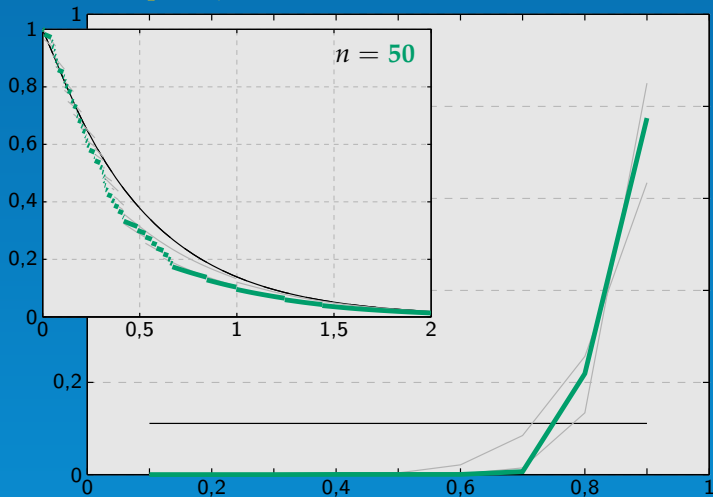
Data: Exp(1), $p_* = 0,9$ $n_0 = 10, p \sim \delta_{0,4}$ 

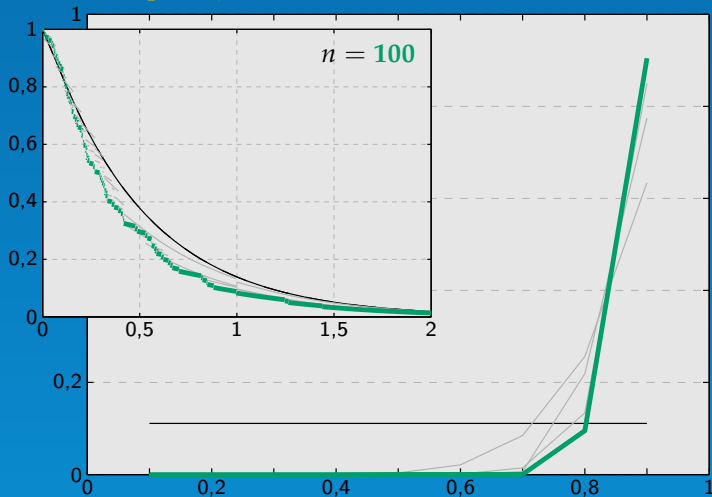
Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$, $p \sim \delta_{0,4}$ 

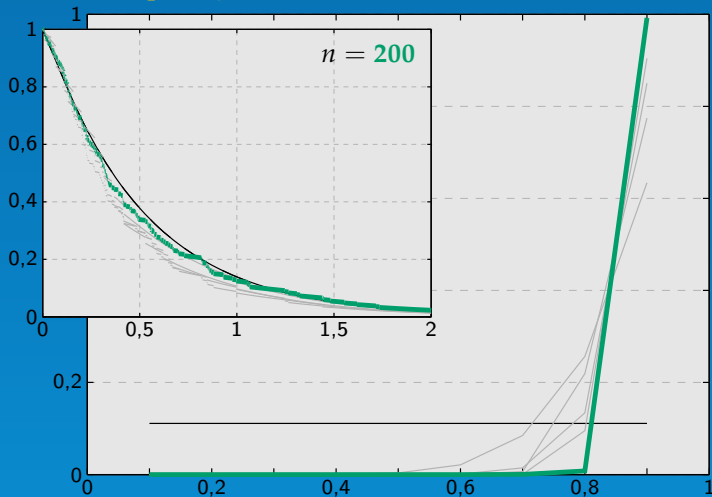
Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$ 

Data: Exp(1), $p_* = 0,9$ $n_0 = 10$ 

Data: Exp(1), $p_* = 0,9$ $n_0 = 10$ 

Data: Exp(1), $p_* = 0,9$ $n_0 = 10$ 

Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$ 

Data: $\text{Exp}(1)$, $p_* = 0,9$ $n_0 = 10$ 

Příklad 2 — Weib(5)

Skutečné rozdělení

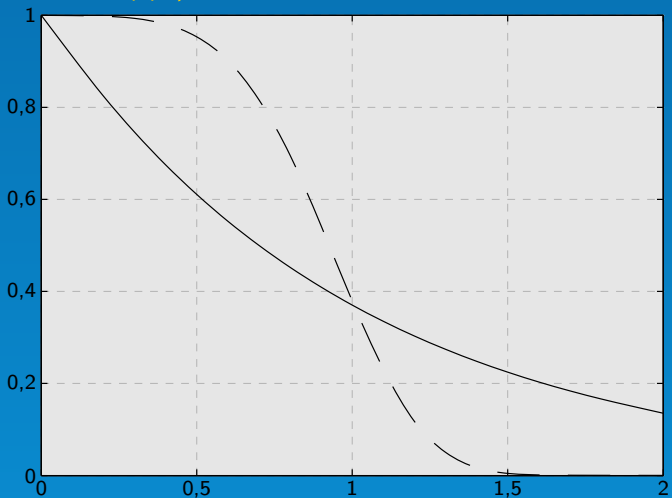
- Weibullovo $S_*(t) = \exp(-t^5)$, $t > 0$
- skutečný koeficient γ_* odpovídající $p_* = 0,8$

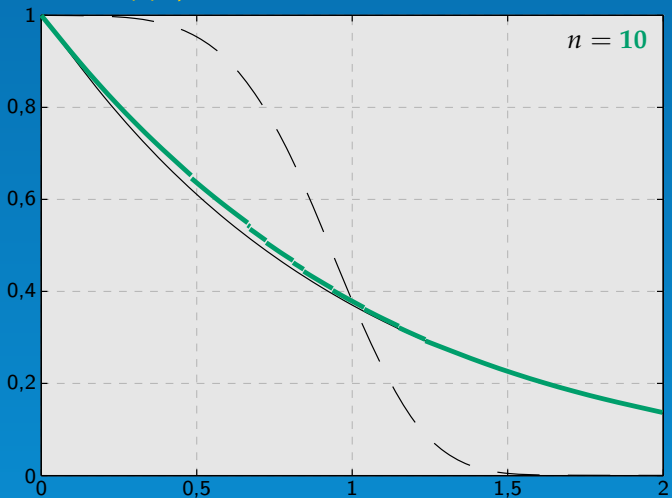
Počty pozorování

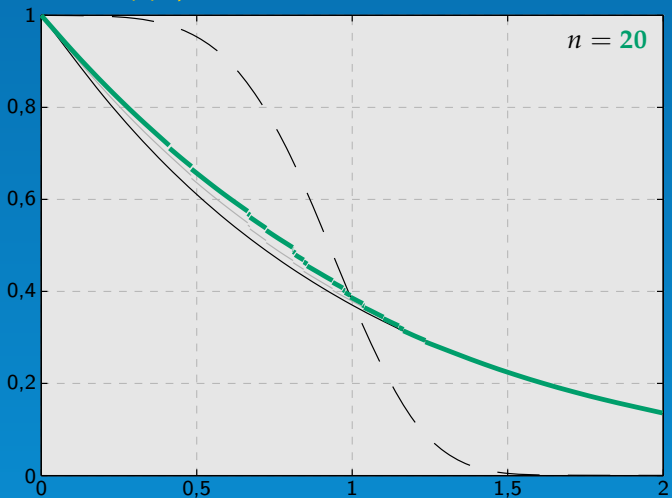
- $n = 10, 20, 50, 100, 200$
- $n_0 = 100$

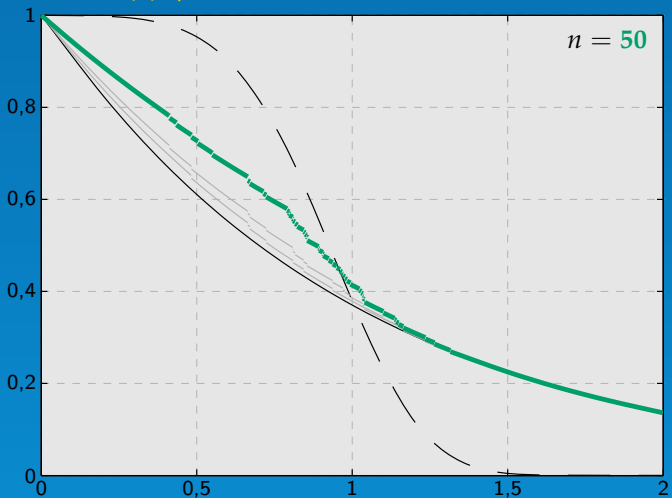
Následují grafy

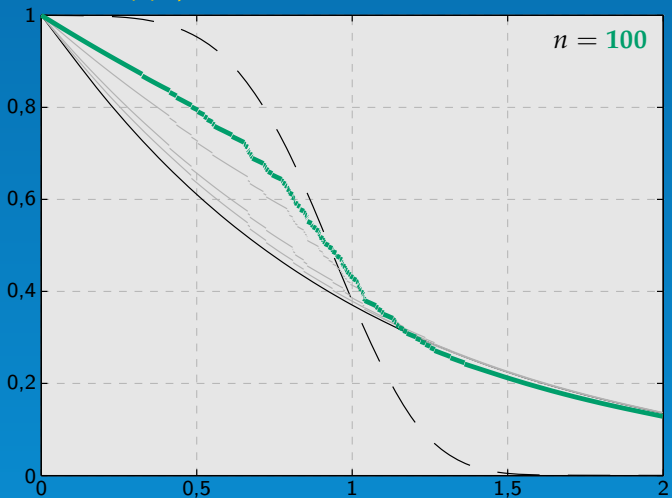
- odhad $\hat{S}(t)$
- aposteriorní $\pi(p \mid \text{data})$

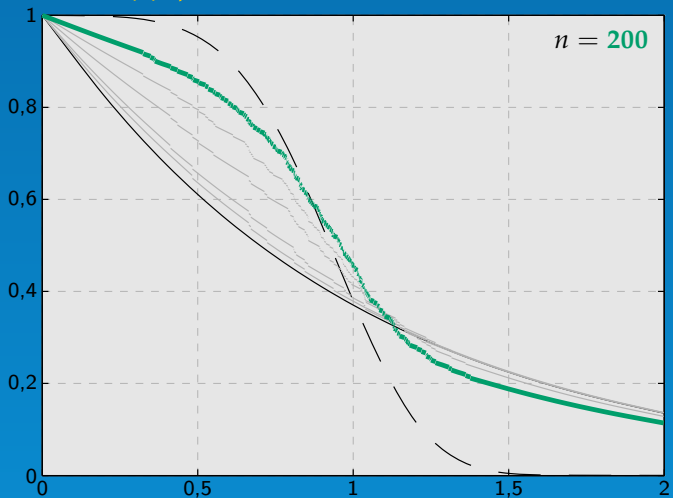
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

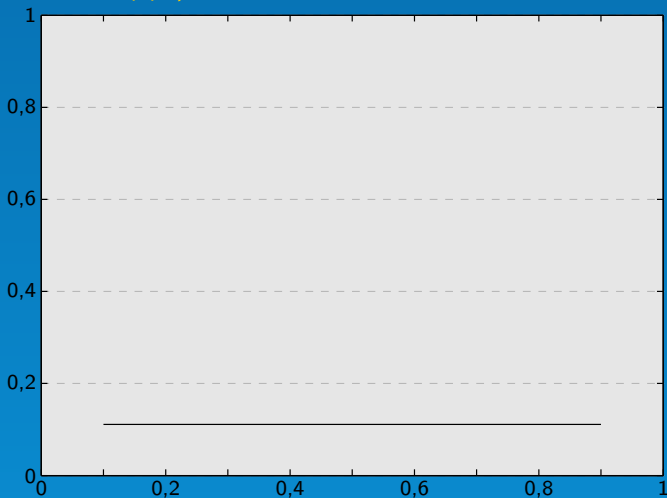
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

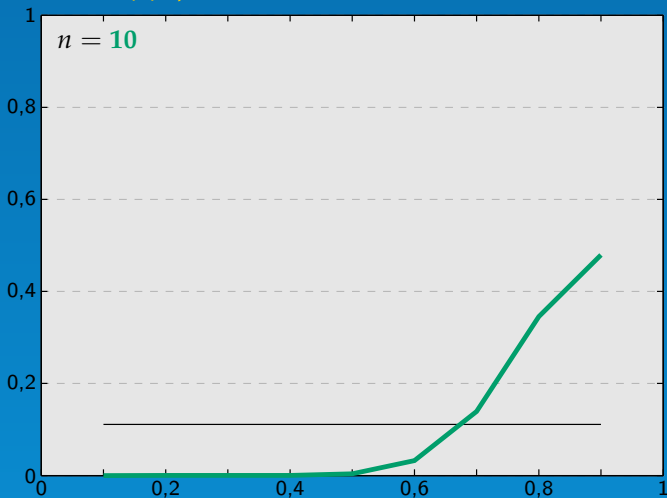
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

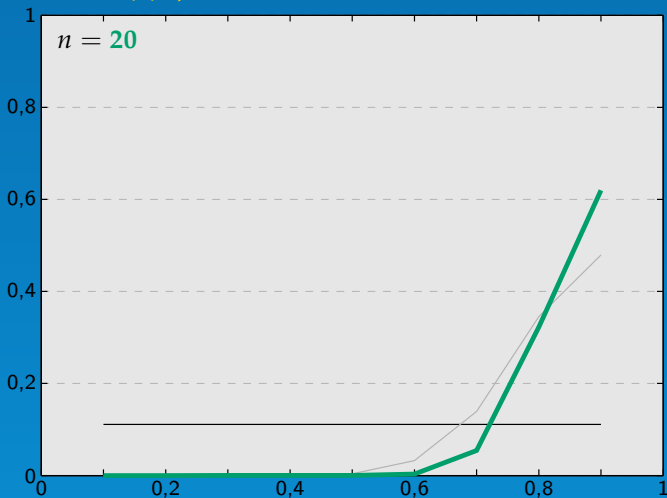
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

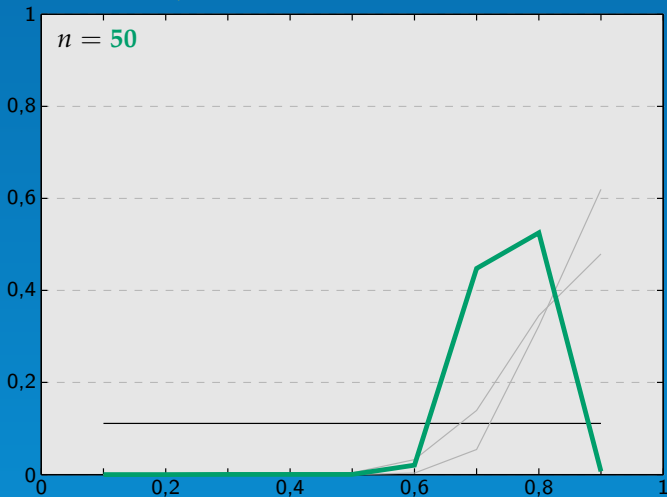
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

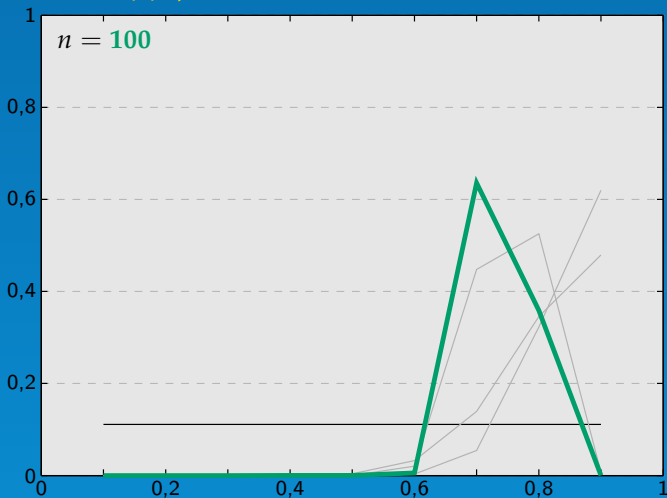
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

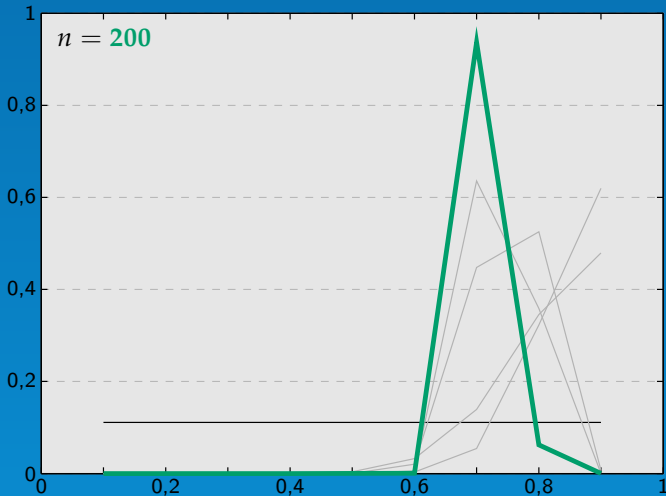
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 100$ 

Vliv některých parametrů

Větší γ_*

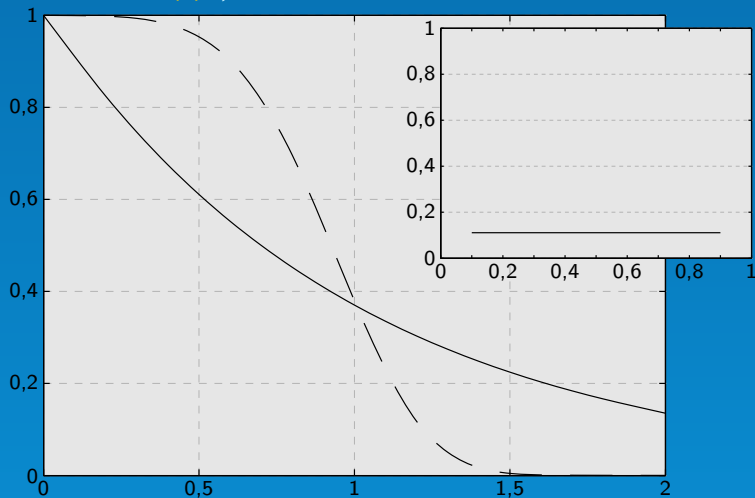
- místo $\gamma_* = 1/0,8 - 1 = 1/4$
generujeme s $\gamma_* = 1/0,2 - 1 = 4$
- méně necenzorovaných pozorování
- více menších časů

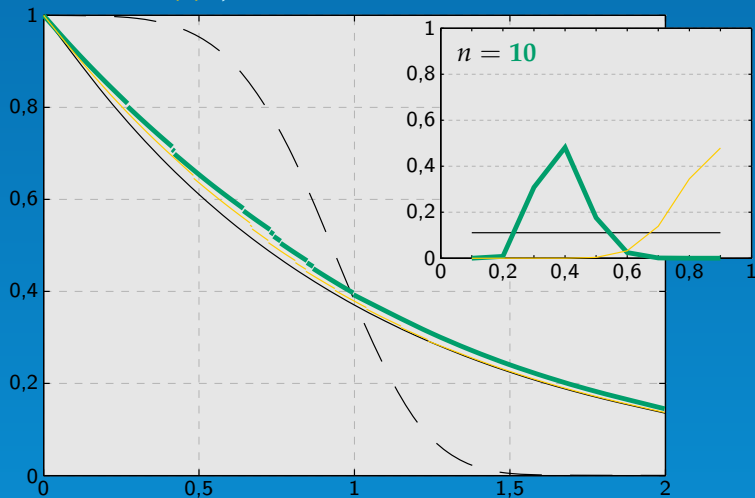
Větší apriorní váhu malým p

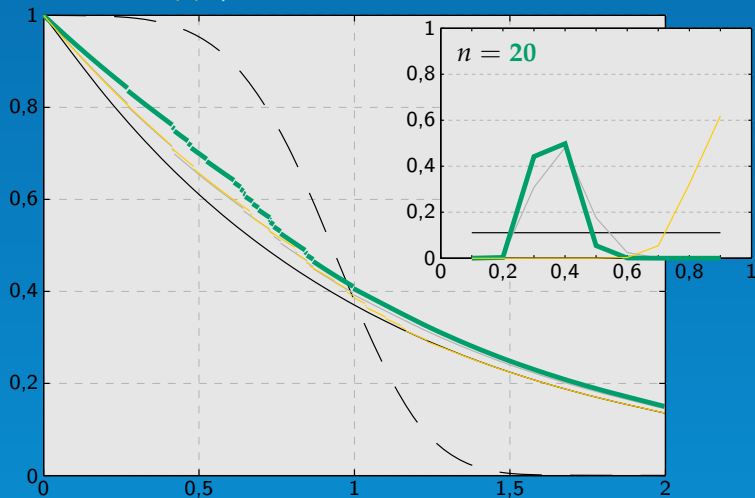
- apriorní rozdělení $\pi(p) \propto (1 - p)^2$

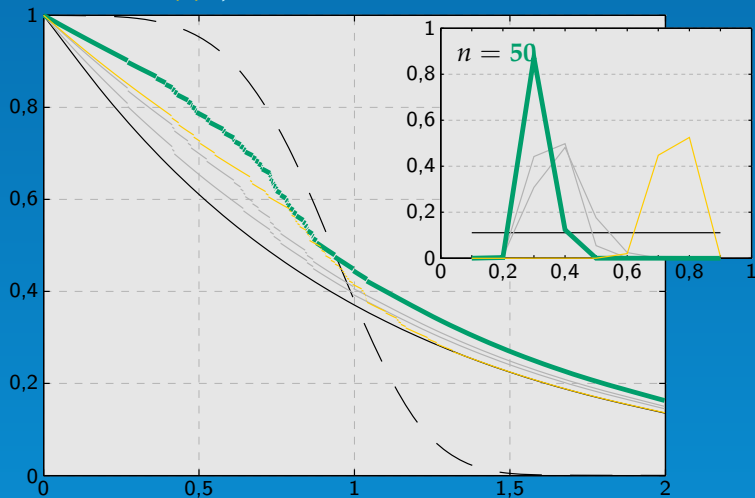
Zmenšení vlivu apriorního rozdělení Λ

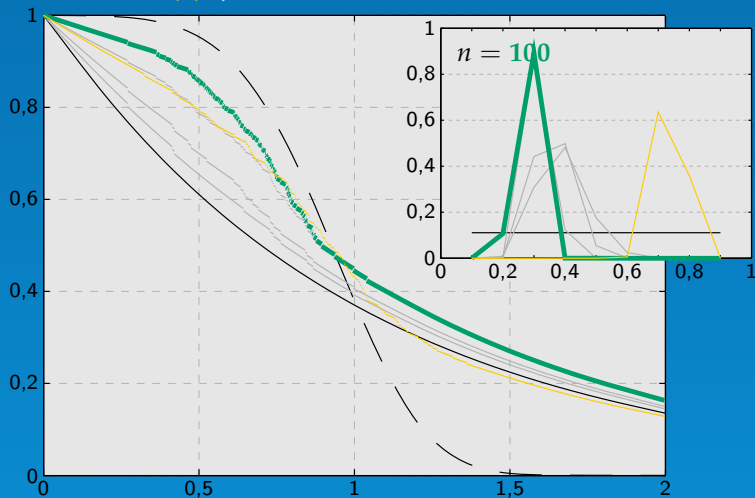
- místo $n_0 = 100$ vezmeme $n_0 = 10$

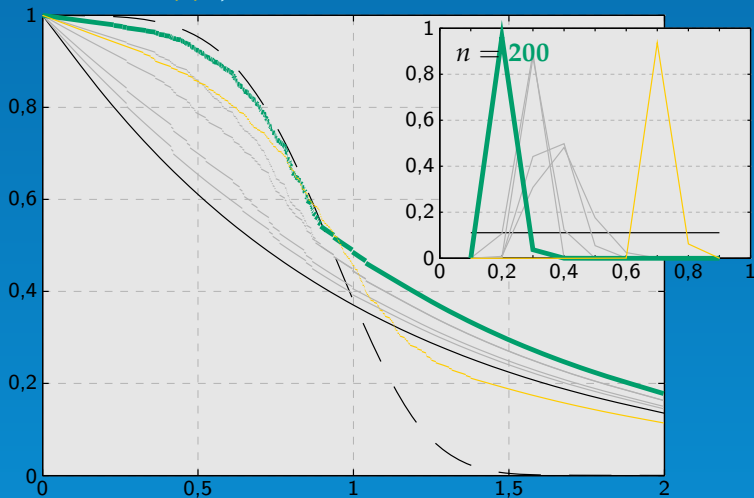
Data: Weib(5), $p_* = 0,2$ $n_0 = 100$ 

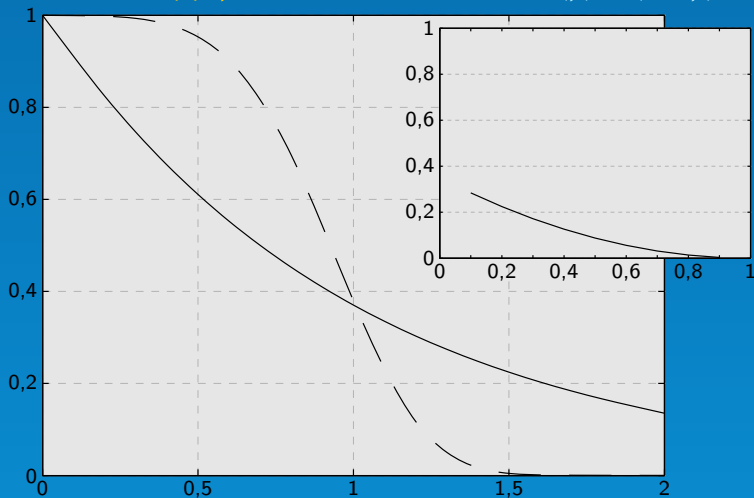
Data: Weib(5), $p_* = 0,2$ $n_0 = 100$ 

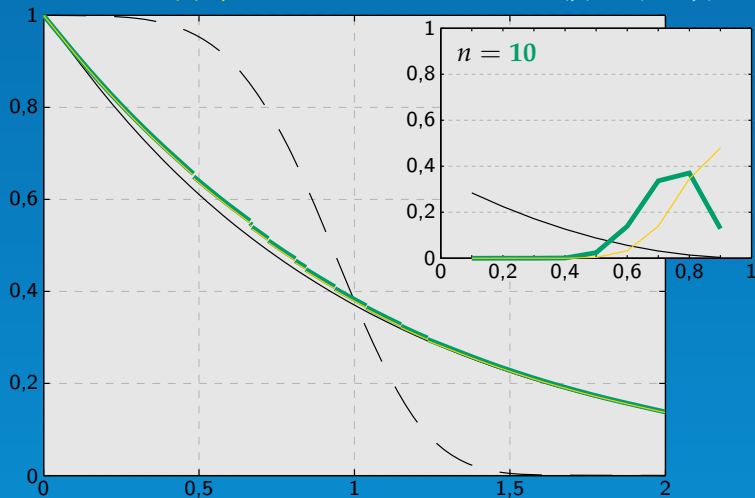
Data: Weib(5), $p_* = 0,2$ $n_0 = 100$ 

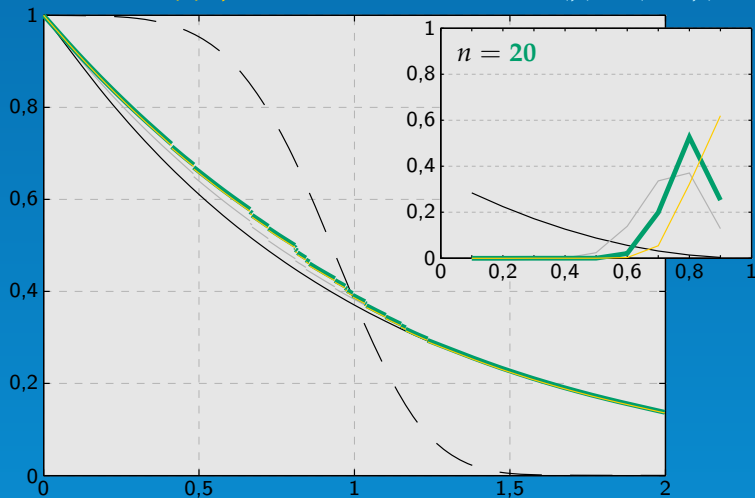
Data: Weib(5), $p_* = 0,2$ $n_0 = 100$ 

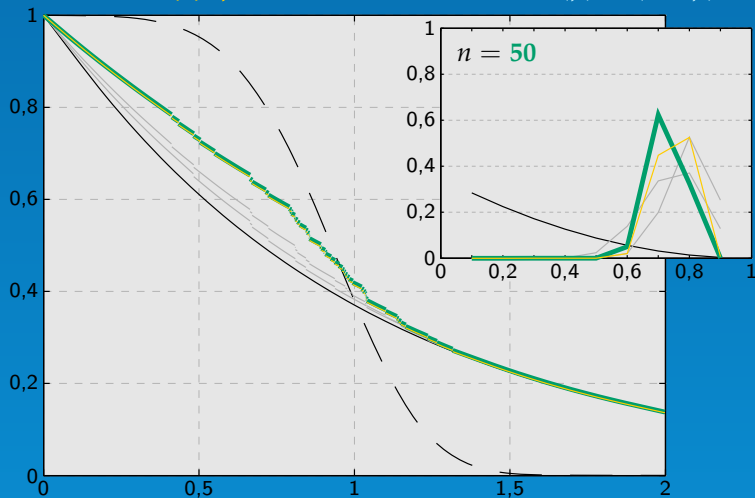
Data: Weib(5), $p_* = 0,2$ $n_0 = 100$ 

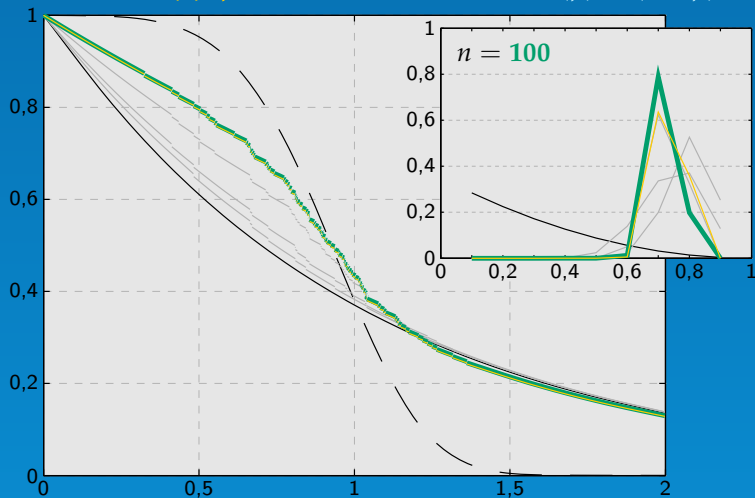
Data: Weib(5), $p_* = 0,2$ $n_0 = 100$ 

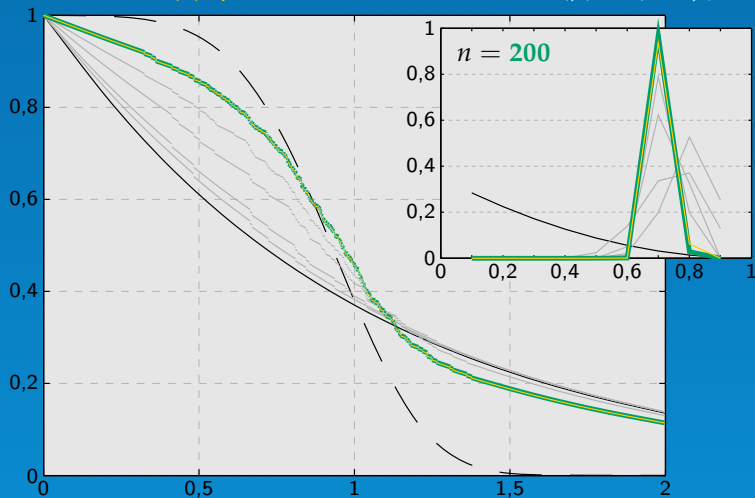
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$, $\pi(p) \propto (1 - p)^2$ 

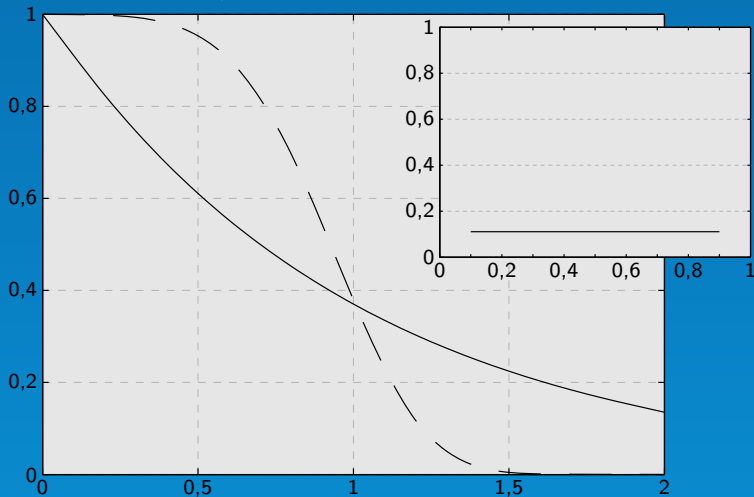
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$, $\pi(p) \propto (1-p)^2$ 

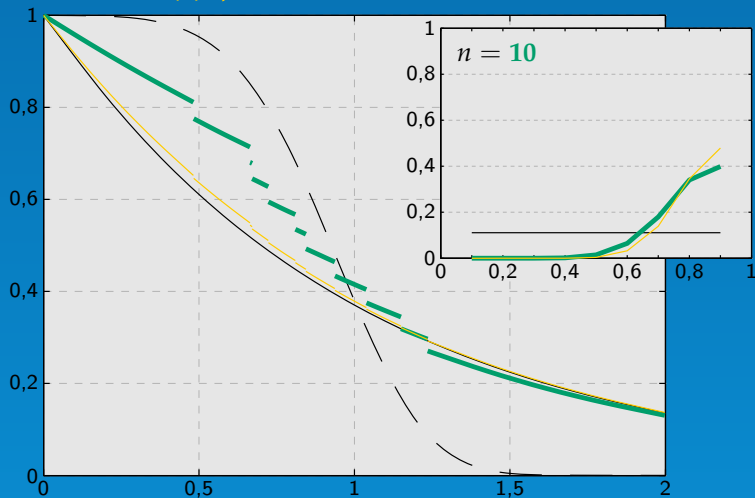
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$, $\pi(p) \propto (1-p)^2$ 

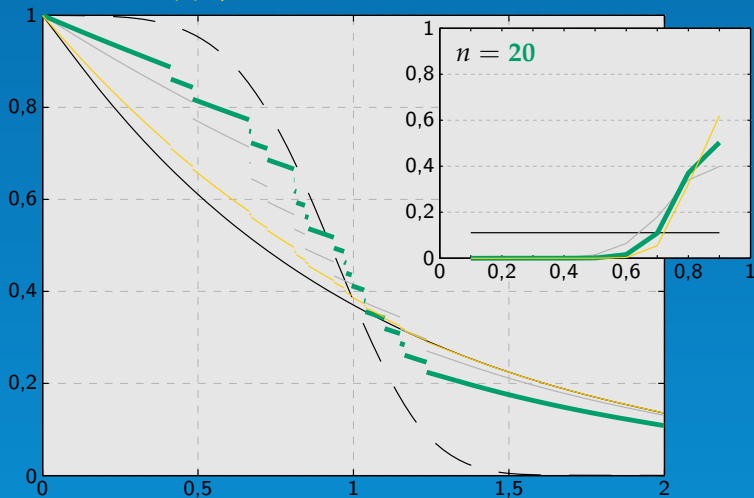
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$, $\pi(p) \propto (1-p)^2$ 

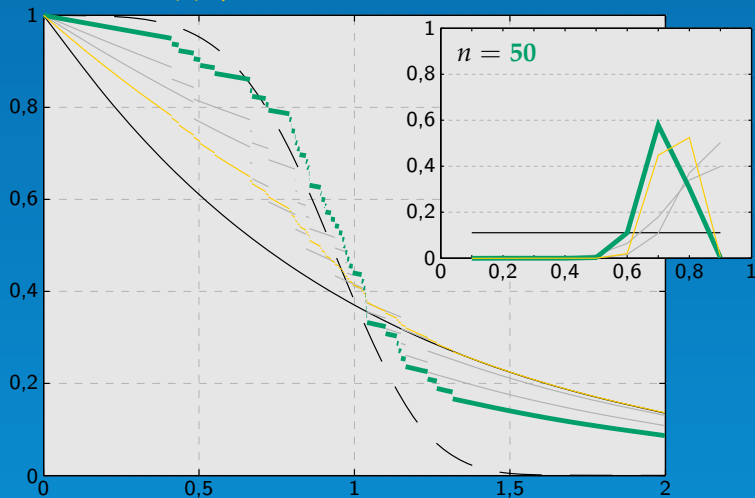
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$, $\pi(p) \propto (1-p)^2$ 

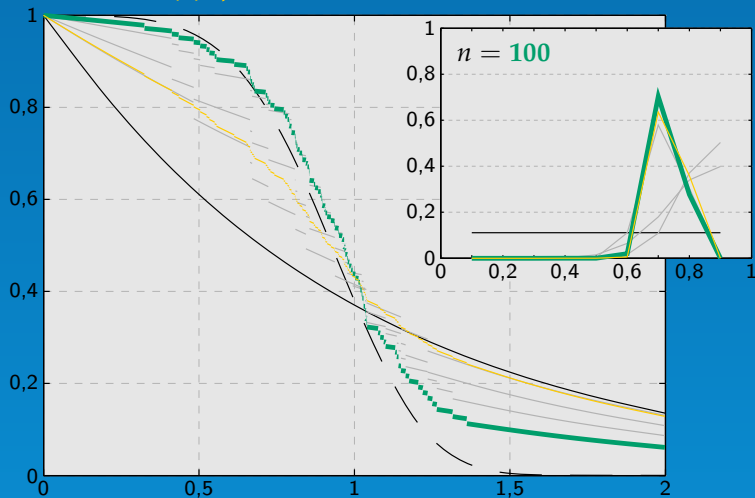
Data: Weib(5), $p_* = 0,8$ $n_0 = 100$, $\pi(p) \propto (1-p)^2$ 

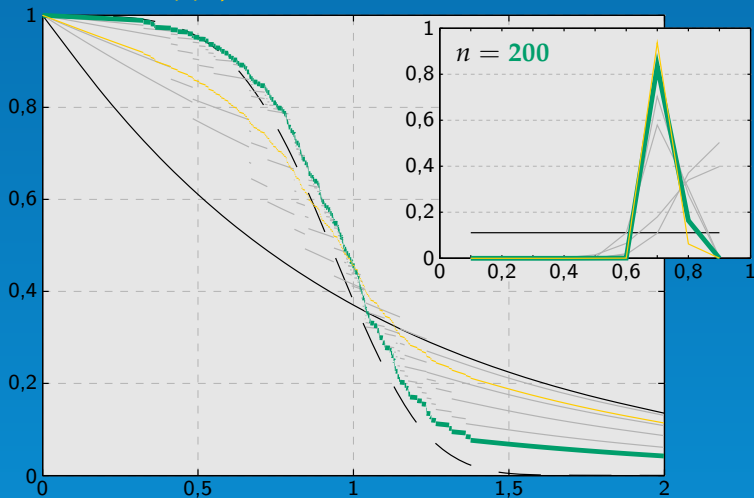
Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

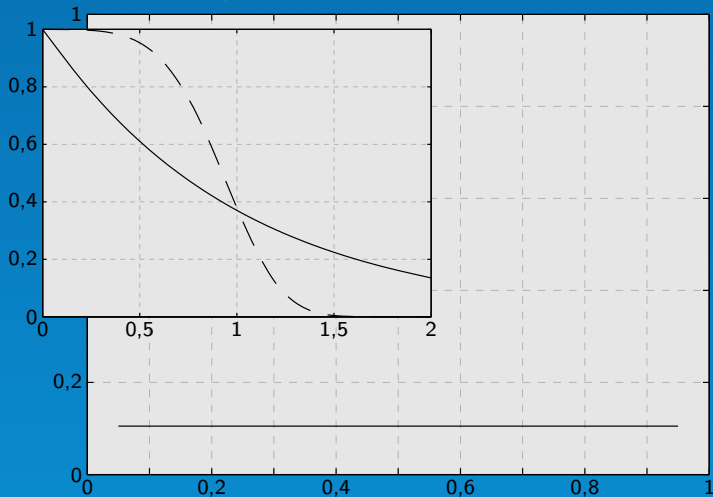
Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

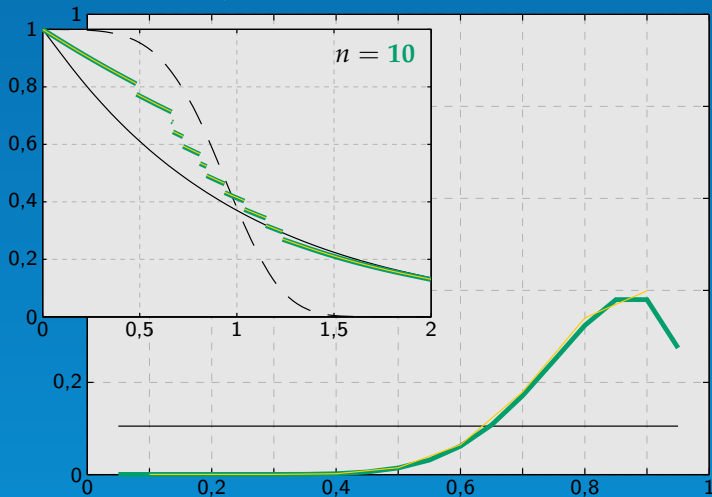
Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

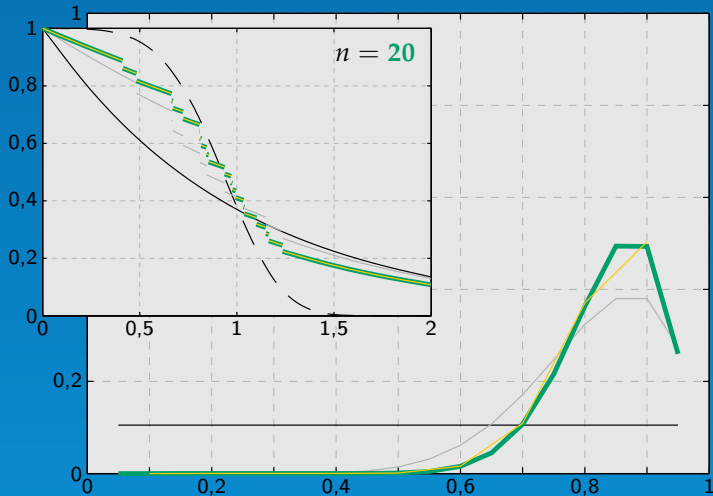
Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

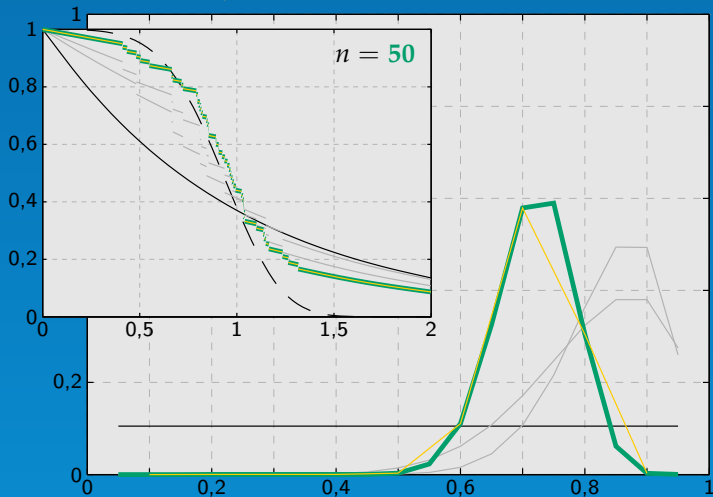
Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

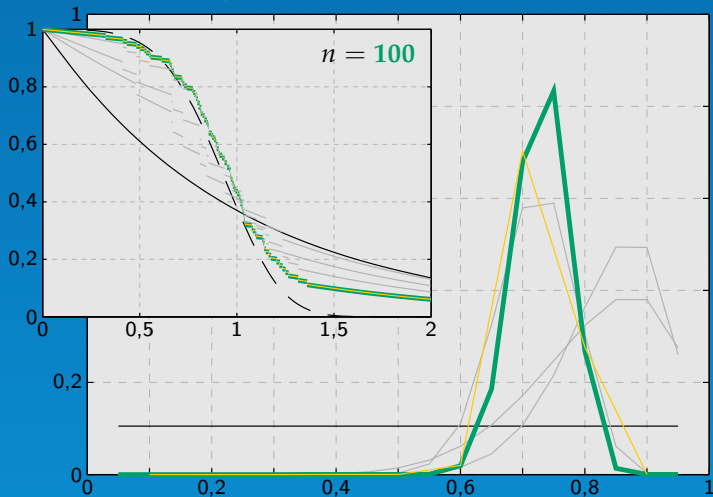
Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

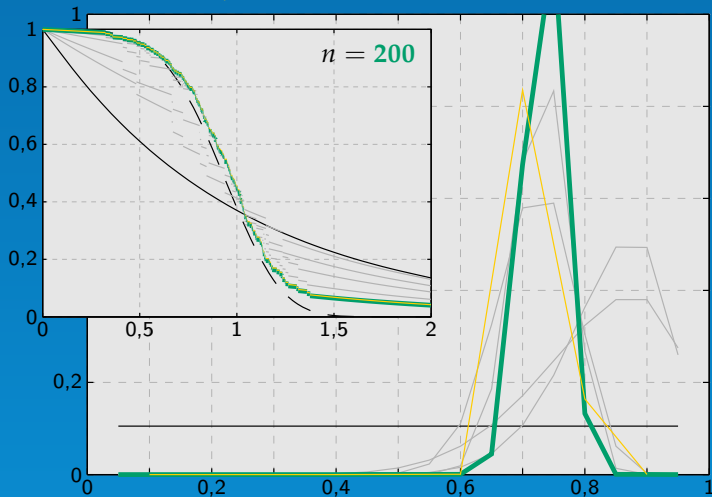
Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

Data: Weib(5), $p_* = 0,8$

$$n_0 = 10$$


Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

Data: Weib(5), $p_* = 0,8$ $n_0 = 10$ 

Příklad 3 — Weib(0,5)

Data

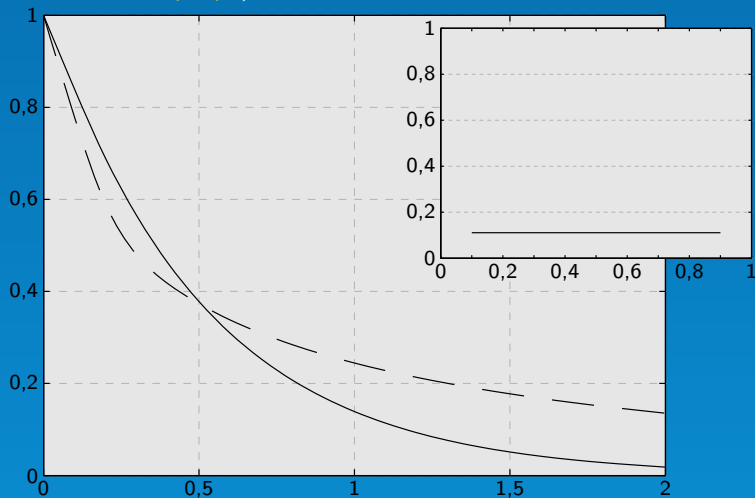
- z Weibullova $S_*(t) = \exp(-t^{0,5})$, $t > 0$

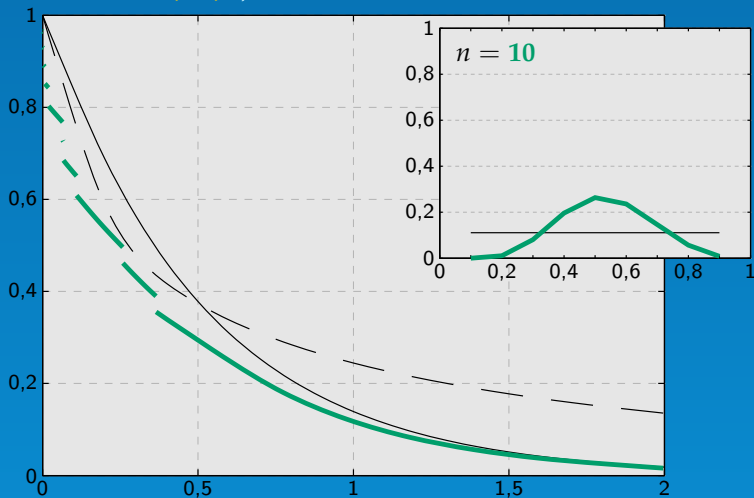
Apriorní rozdělení

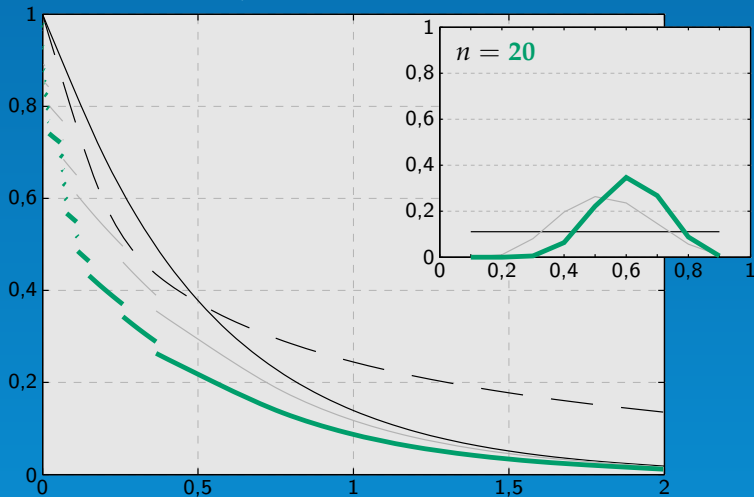
- $n_0 = 10$ a stále s $\Lambda_0(t) = t$
- $\pi(\gamma)$ stále „rovnoměrná“

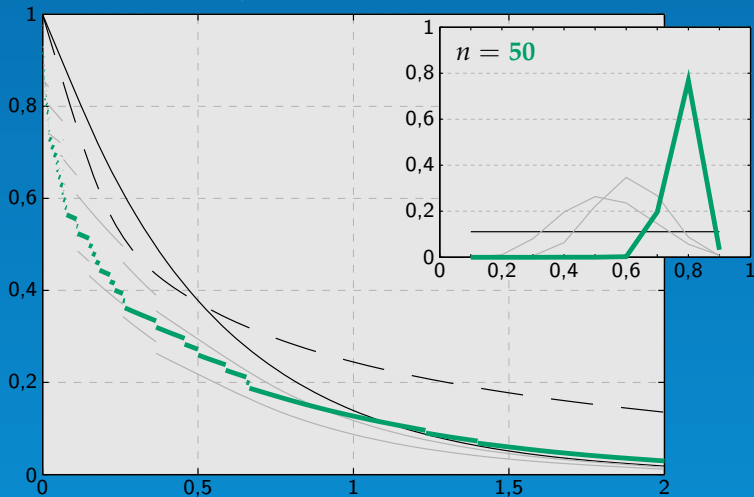
Následují grafy pro simulaci s

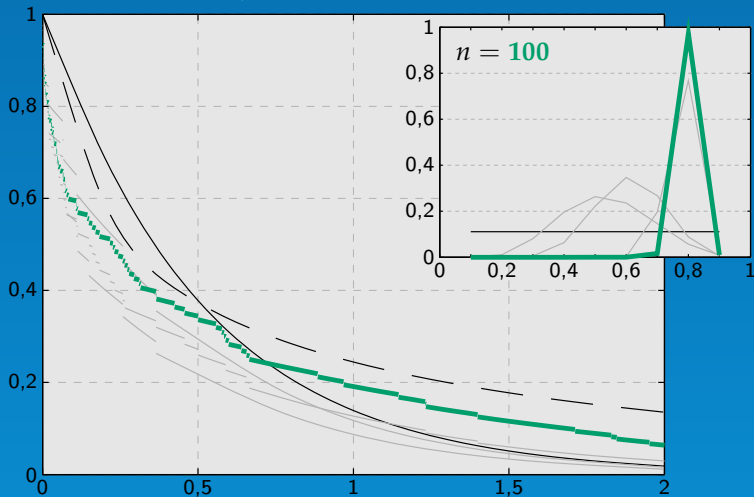
- $\gamma_* = 1/4$ ($p_* = 0,8$)
- $\gamma_* = 1$ ($p_* = 0,5$)

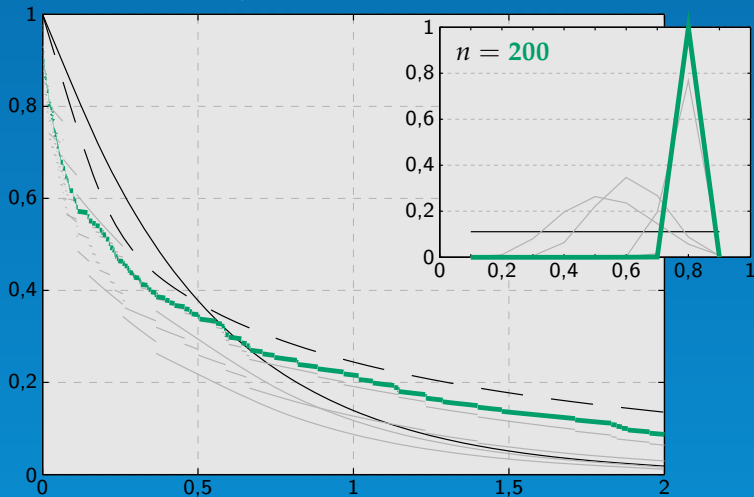
Data: Weib(0,5), $p_* = 0,8$ $n_0 = 10$ 

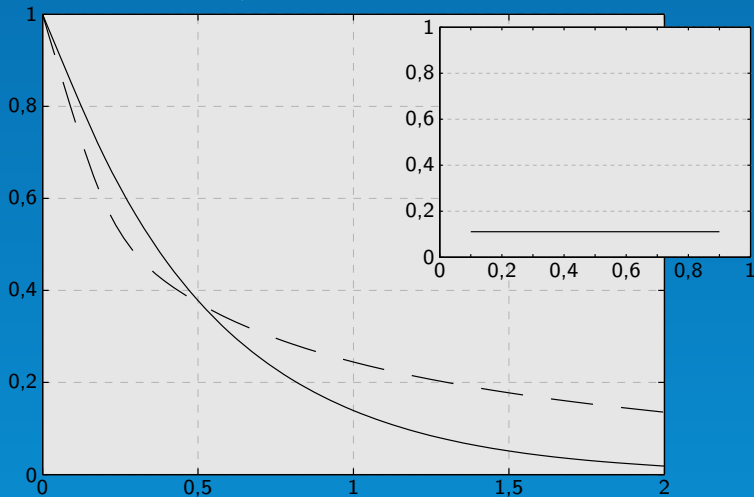
Data: Weib(0,5), $p_* = 0,8$ $n_0 = 10$ 

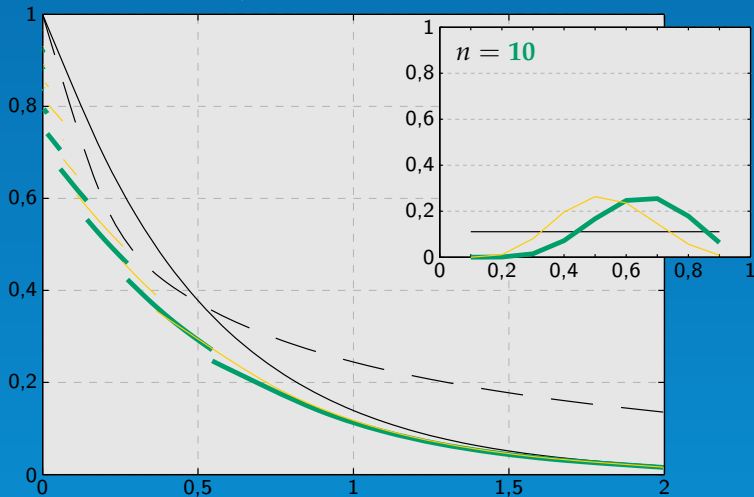
Data: Weib(0,5), $p_* = 0,8$ $n_0 = 10$ 

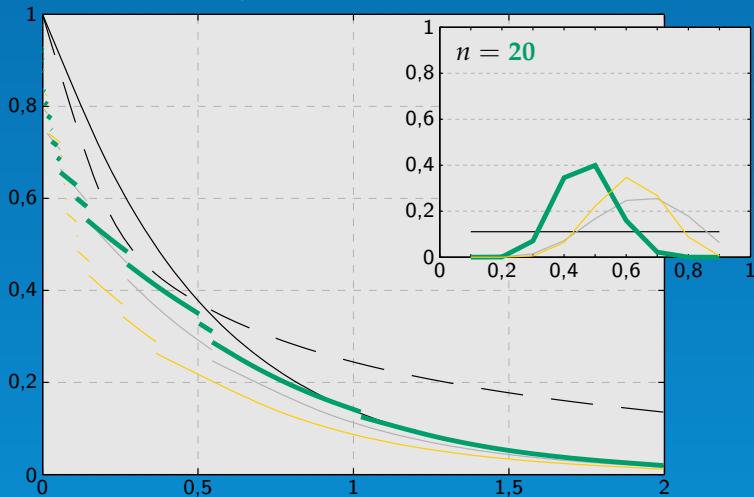
Data: Weib(0,5), $p_* = 0,8$ $n_0 = 10$ 

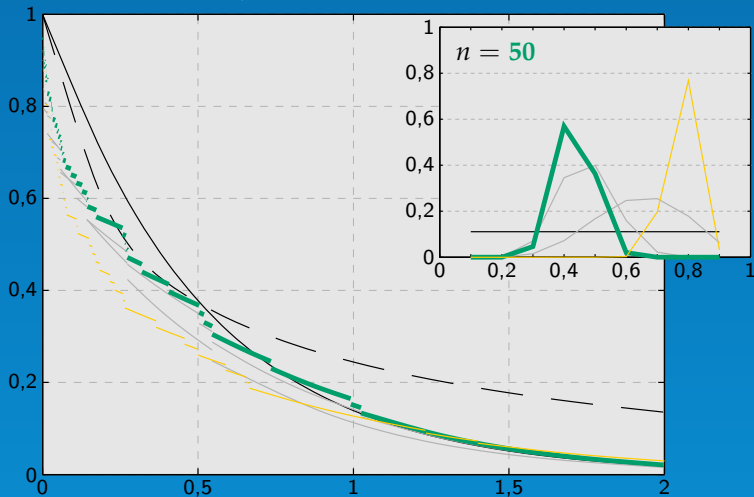
Data: Weib(0,5), $p_* = 0,8$ $n_0 = 10$ 

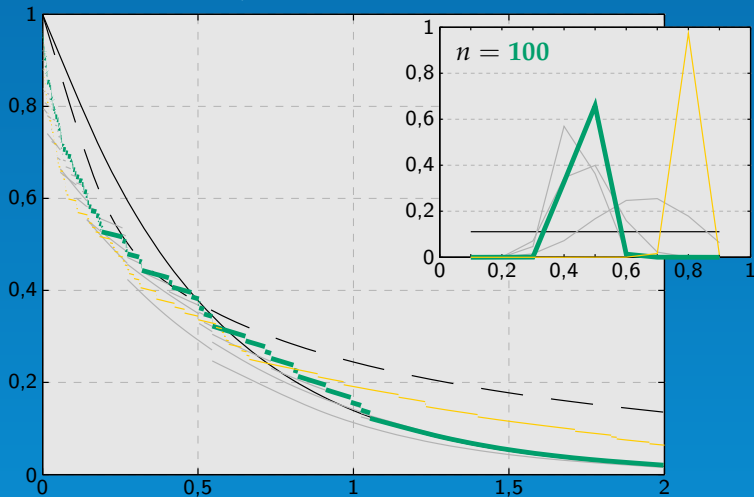
Data: Weib(0,5), $p_* = 0,8$ $n_0 = 10$ 

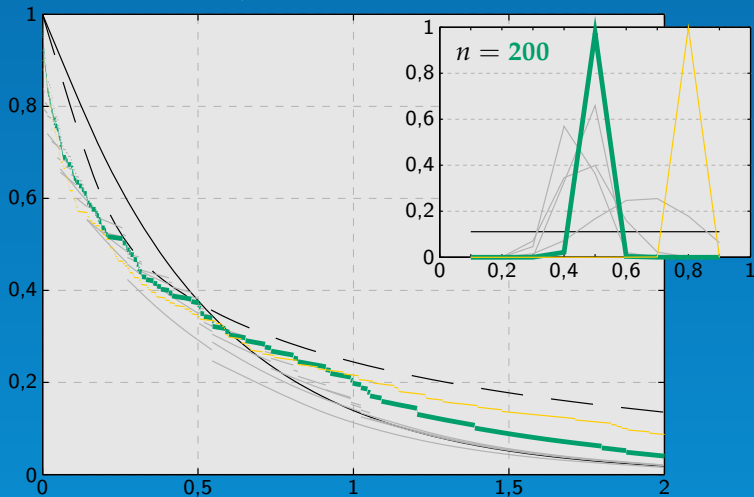
Data: Weib(0,5), $p_* = 0,5$ $n_0 = 10$ 

Data: Weib(0,5), $p_* = 0,5$ $n_0 = 10$ 

Data: Weib(0,5), $p_* = 0,5$ $n_0 = 10$ 

Data: Weib(0,5), $p_* = 0,5$ $n_0 = 10$ 

Data: Weib(0,5), $p_* = 0,5$ $n_0 = 10$ 

Data: Weib(0,5), $p_* = 0,5$ $n_0 = 10$ 

KONZISTENCE NEPARAMETRICKÉHO BAYESOVSKÉHO ODHADU

Michal Friesl

*Katedra matematiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni*

- <http://home.zcu.cz/~friesl/Archiv/Rob10.html>