Slučajne varijable i modeli doživljenja

Bojan Basrak

2008

Uvod

Tokovi novca koje smo do sada promatrali bili su deterministički. Poopćene tokove novca tretiramo vrlo slično no sada se teorija snažno oslanja na vjerojatnost.

Slučajna varijabla X na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ je izmjerivo preslikavanje $X : \Omega \to \mathbb{R}$, tj. preslikavanje sa svojstvom

$$\{\omega : X(\omega) \leq b\} \in \mathcal{F}, \qquad \text{za sve } b \in \mathbb{R}.$$

Funkcija distribucije sl. var. X je $F : \mathbb{R} \to [0, 1]$ dana sa

$$F(x) = F_X(x) = P(X \leq x)\,.$$

Ako je razdioba slučajne varijable X koncentrirana na prebrojivom ili konačnom skupu kažemo da je X diskretna sl.var. Tada, za neki prebrojiv ili čak konačan $D \subseteq \mathbb{R}$, vrijedi $P(X \in D) = 1$ (UViS).

Funkciju $f(x) = f_X(x) = P(X = x)$ zovemo fja. gustoće diskretne slučajne varijable X. Preko nje obično zadajemo razdiobu od X.

Primjeri (tablica razdiobe diskretne sl. var.)

$$X \sim \begin{pmatrix} a_1 & a_2 & a_3 & \ldots \\ p_1 & p_2 & p_3 & \ldots \end{pmatrix},$$

gdje je naravno $p_i = f(a_i)$. Brojevi a_i u prvom retku su međusobno različiti, a brojevi p_i moraju zadovoljavati $\sum p_i = 1$, $p_i \geq 0$.

Diskretne razdiobe su npr.: binomna, Poissonova, geometrijska, hipergeometrijska itd. ⌟

Slučajna varijabla je neprekidna ako postoji funkcija $f : \mathbb{R} \to [0, \infty)$ t.d. je za svaki $x \in \mathbb{R}$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\mathrm{d}s.$$

Funkciju f zovemo funkcija gustoće od X. Nužno je $\int_{-\infty}^{\infty} f(s)\mathrm{d}s = 1$. Ako je f neprekidna u točki x, za "male" vrijednosti $\triangle x$ vrijedi

$$P(x < X \leq x + \triangle x) = \int_{x}^{x+\triangle x} f(s)\mathrm{d}s \approx f(x)\triangle x\,.$$

Napomena Studenti sa znanjem I&M ili TV će razmišljati o Lebesgue-integrabilnoj f i govoriti da je X apsolutno neprekidna u odnosu na Lebesgueovu mjeru. Mi ćemo tipično pretpostavljati da je f i sama neprekidna osim u eventualno konačno mnogo točaka, tada je ona i Riemann integrabilna na svakom konačnom intervalu.

Propozicija 1 *Ako je X neprekidna sl. var. tada je neprekidna i njena funkcija distribucije F.*

D. jasan. $\square$

Napomena Posebno ako je X neprekidna tada $P(X \in D) = 0$, za svaki prebrojiv $D \subseteq \mathbb{R}$.

Ako je fja distribucije F neprekidna i neprekidno derivabilna (osim u eventualno konačno mnogo točaka) tada ona ima gustoću, koju možemo definirati kao $F'(\cdot)$ na skupu gdje je F derivabilna, a 0 izvan tog skupa (v. 2. osnovni teorem calculusa).

Primjeri (neprekidne sl. var.)

i) Uniformna razdioba, npr. $U(0,1)$

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{inače} \end{cases} \quad , \quad F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \geq x < 1 \\ 1 & x > 1 \end{cases} \quad ,$$

ii) Eksponencijalna razdioba, oznaka Exp(λ), $\lambda > 0$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & 0 < x \\ 0 & \text{inače} \end{cases} \quad , \quad F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases} \quad ,$$

iii) Cauchyeva razdioba (simetrična)

$$f_X(x) = \frac{1}{\pi(1+x^2)}.$$

Pokažite da je f_X zaista gustoća.

iv) Gaussova ili normalna razdioba, $N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Za $\mu = 0, \ \sigma = 1$ razdioba se zove i standardna normalna, za koju koristimo standardne oznake

$$\varphi(x) = f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt.$$

v) Gama razdioba, $\Gamma(a, b)$, $a, b > 0$

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(a)} b^a x^{a-1} e^{-bx} & 0 < x \\ 0 & x \leq 0 \end{cases}$$

$\lrcorner$

Teorem 2 *(o transformaciji neprekidne sl. varijable) Pretpostavimo da je* X *neprekidna sl. var. s funkcijom gustoće* f_X*, te da je* ψ *striktno rastuća diferencijabilna funkcija, a* $Z = \psi^{-1}(X)$*, tada je i* Z *neprekidna sl. varijabla s funkcijom gustoće*

$$f_Z(z) = f_X(\psi(z))\psi'(z)$$

ili preciznije

$$P(Z \leq z) = \int_{-\infty}^{z} f_X(\psi(s))\psi'(s)\mathrm{d}s.$$

Matematičko očekivanje nenegativne diskretne sl. var. X možemo definirati kao

$$EX = \sum_{i \geq 1} a_i p_i = \sum_x x f(x)$$

uz dogovor da ako ovaj red divergira smatramo $EX = +\infty$. Posebno, za slučajan dogadjaj A, i sl. varijablu $\mathbf{1}_A$ očito vrijedi da je $E\mathbf{1}_A = P(A)$.

Matematičko očekivanje općenite diskretne sl. var. X definiramo kao

$$EX = EX^+ - EX^-,$$

gdje je $X^+ = \max(X, 0)$, a $X^- = -\min(X, 0)$, ako je bar jedno od očekivanja na desnoj strani konačno. Prema dogovoru za svaki realan broj a vrijedi $\pm\infty + a = \pm\infty$.

Ako je funkcija $xf(x),\ x > 0$ Riemann (Lebesgue) integrabilna definiramo matematičko očekivanje nenegativne neprekidne slučajne varijable kao

$$EX = \int_0^\infty xf(x)\mathrm{d}x\,.$$

Kao i gore za općenite neprekidne sl. varijable stavljamo

$$EX = EX^+ - EX^-.$$

Za (izmjerivu) realnu funkciju u može se pokazati da je

$$E\,u(X) = \int_{-\infty}^\infty u(s)f(s)\mathrm{d}s\,,$$

ako $Eu(X)$ postoji.

Varijancu sl. var. X definiramo kao $E(X - EX)^2$ gdje je taj izraz definiran u gornjem smislu.

Napomena Uočite: ponekad se kaže da očekivanje postoji ako i samo ako je $E|X| < \infty$ (v. npr. NS 1.dio), dakle tada čekivanje odn. varijanca ne postoje općenito. Ako je $X \geq 0$, a EX divergira k $+\infty$, mi uvijek pišemo $EX = +\infty$.

Teorem 3 *Za neprekidnu i nenegativnu sl. var. X s funkcijom distribucije F i konačnim očekivanjem vrijedi*

$$EX = \int_0^\infty (1 - F(x))\mathrm{d}x.$$

O raznim interpretacijama matem. očekivanja

Očekivanje kao limes u jakom zakonu velikih brojeva

Teorem 4 *(Kolmogorov) Za niz njd sl. varijabli* (X_i) *vrijedi: očekivanje* EX_1 *postoji i konačno je ako i samo ako*

$$\frac{X_1 + \cdots X_n}{n} \xrightarrow{\text{g.s.}} EX_1, \quad n \to \infty.$$

Kako konvergencija g.s. povlači konvergenciju po vjerojatnosti, uz uvjet $E|X| < \infty$ vrijedi i

$$\frac{X_1 + \cdots X_n}{n} \xrightarrow{P} EX_1, \quad n \to \infty.$$

Očekivanje kao najbolja aproksimacija

Pretpostavimo da sl. varijabla X ima konačnu varijancu $\mathrm{var} X < \infty$ i da želimo odrediti $c \in \mathbb{R}$ tako da je

$$E(X-c)^2$$

minimalno, tada je $c = EX$, jer

$$E(X-z)^2 = E(X-EX)^2 + (EX-z)^2 \geq \mathrm{var} X.$$

Očekivanje kao fair cijena odn. premija za igru sa slučajnim ishodom

Ako promatramo igru koja kada se dogodi sl. dogadjaj A donosi dobitak $a > 0$, odn. gubitak iznosa $b > 0$ ako se A ne dogodi, i ako je $P(A) = p, \; q = 1 - P(A)$, mogli bismo reći da je fair cijena ove igre

$$-bq + ap$$

što je dakako očekivanje sl. varijable koja ima razdiobu dobitka u ovoj igri

$$Y \sim \begin{pmatrix} -b & a \\ q & p \end{pmatrix}.$$

Ispostavlja se da postoji puno situacija u kojima bi ovakvo odredjivanje fair cijene imalo nerazumne ili neočekivane posljedice: Evropska call opcija, Sankt Petersburg paradox, problem dvije omotnice.

Primjeri (neprekidne sl. var. bez konačnog očekivanja)

i) Pareto razdioba, oznaka Par(α, κ) za neke parametre $\alpha, \kappa > 0$.

$$\bar{F}_X(x) = 1 - F_X(x) = \begin{cases} \left(\frac{\kappa}{\kappa+x}\right)^\alpha & 0 < x \\ 1 & \text{inače} \end{cases},$$

i stoga

$$f_X(x) = \begin{cases} \frac{\kappa^\alpha}{(\kappa+x)^(\alpha+1)} & 0 < x \\ 0 & \text{inače} \end{cases}.$$

Provjerite: $EX = +\infty$ akko $\alpha \leq 1$.

ii) Cauchyeva razdioba (simetrična)

$$f_X(x) = \frac{1}{\pi(1+x^2)}.$$

Pokažite da je $EX^+ = EX^- = \infty$, dakle očekivanje ne postoji.

Teorem 5 *i) linearnost - ako su X, Y sl. varijable s konačnim očekivanjima, a i b proizvoljni realni brojevi, tada*

$$E(aX + bY) = aEX + bEY.$$

ii) Čebiševljeva nejednakost - ako X ima konačnu varijancu, tada je za svaki $\varepsilon > 0$

$$P(|X - EX| > \varepsilon) \leq \frac{\mathrm{var} X}{\varepsilon^2}.$$

iii) Markovljeva nejednakost - ako X ima konačno očekivanje, tada je za svaki $\varepsilon > 0$

$$P(|X| > \varepsilon) \leq \frac{E|X|}{\varepsilon}.$$

iv) Jensenova nejednakost - ako X ima konačno očekivanje , a funkcija g je konveksna, tada vrijedi

$$E(g(X)) \geq g(EX).$$

Slučajni vektor (X, Y) ima zajedničku funkciju distribucije $F = F_{(X,Y)} : \mathbb{R}^2 \to [0, 1]$ zadanu sa

$$F_{(X,Y)}(u, v) = F(u, v) = P(X \leq u, Y \leq v)\,.$$

Slučajni vektor (X, Y) je neprekidan ako postoji Riemann integrabilna funkcija $f : \mathbb{R}^2 \to [0, \infty)$ t.d. je

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{(-\infty,x]\times(-\infty,y]} f(s, t)\mathrm{d}s\mathrm{d}t. \tag{1}$$

Funkciju f zovemo (zajednička) funkcija gustoće od (X, Y) (alternativno i od F). Nužno je $\int_{\mathbb{R}^2} f(s, t)\mathrm{d}s\mathrm{d}t = 1$.

Ako postoji zajednička gustoća, postoje i *marginalne gustoće* varijabli X i Y, npr. X ima gustoću

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y.$$

Ako je $u : \mathbb{R}^2 \to \mathbb{R}$ izmjeriva funkcija tada je

$$E\,u(X,Y) = \int_{\mathbb{R}^2} u(s,t)f(s,t)\mathrm{d}s\mathrm{d}t\,,$$

čim postoji $Eu(X,Y)$.

Slučajne varijable $X_1, \dots, X_n$ su nezavisne ako

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

za sve izmjerive skupove $A_i \subseteq \mathbb{R},\ i = 1, \dots, n$.

Prisjetimo se: diskretne slučajne varijable su nezavisne akko im se zajedničke funkcije gustoće faktoriziraju. Za neprekidne varijable vrijedi

Teorem 6 *Ako su X i Y nezavisne sl. var. s marginalnim gustoćama f_X i f_Y, tada je i sl. vektor (X, Y) neprekidan sa gustoćom*

$$f(x, y) = f_X(x) f_Y(y)\,.$$

Obrnuto, ako se zajednička fja gustoće može faktorizirati kao $f(x, y) = \varphi(x)\psi(y)$, tada su X i Y nezavisne.

Teorem 7 *Ako su X i Y neprekidne i nezavisne sl. var. s gustoćama F_X odn. f_Y tada je neprekidna i $Z = X + Y$, te ima gustoću*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(u) f_Y(z-u) \mathrm{d}u.$$

Gustoća f_Z iz gornjeg teorema se cesto naziva i konvolucija gustoća sl. varijabli X i Y.

Za nezavisne sl. var. X i Y s konačnim očekivanjem

$$E(XY) = E(X)E(Y) .$$

Uvjetnu vjerojatnost dogadjaja A uz uvjet B, gdje $P(B) > 0$, definiramo kao

$$P(A|B) = P(A \cap B)/P(B)\,.$$

Ako je B slučajan dogadjaj takav da $P(B) > 0$, a X diskretna sl. var., s

$$f(x|B) = P(X = x|B)$$

zadana je uvjetna razdioba (točnije gustoća) sl. var. X uz uvjet B.

Kako je $f(x|B)$ ponovo fja. gustoće (zašto?) možemo izračunati uvjetno očekivanje od X uz uvjet B kao

$$E(X|B) = \sum_x x^+ f(x|B) - \sum_x x^- f(x|B)\,,$$

kada bar jedan red na d.s. konvergira.

Uvjetnu razdiobu odn. očekivanje možemo definirati i za neprekidne sl. var. Neka je X neprekidna sl. var. i neka je dan sl. dogadjaj A, $P(A) > 0$, tada je dobro definirana uvjetna funkcija distribucije

$$F_{X|A}(x) = P(X \leq x|A),$$

pa i za nju možemo tražiti gustoću ako postoji.

Nas će posebno zanimati uvjeti oblika

$$\{\omega : a < X(\omega) \leq b\},$$

za $-\infty \leq a < b \leq \infty$. Uvjetnu gustoću sl. var. X uz uvjet A je tada posebno lako naći kao

$$f_{X|A}(x) = \begin{cases} f(x)/(F(b) - F(a)) & \text{za } x \in (a, b] \\ 0 & \text{inače}. \end{cases}$$

Tada uvjetno očekivanje definiramo kao

$$E(X|A) = \frac{E(X\mathbf{1}_A)}{P(A)} = \int_a^b v \frac{f(v)}{F(b) - F(a)} \mathrm{d}v\,.$$

Primjer Neka je T duljina života sl. odabrane osobe, katkad će nam trebati uvjetna fja. distribucije i uvjetno očekivanje

$$F(t|x) := F_{T|T>x}(t+x) = P(T \le t + x | T > x) \quad \text{i} \quad E(T - x | T > x)\,.$$

Neka je X proizvoljna sl. var. s konačnim očekivanjem, te neka je Y diskretna sl. var. na istom vj. prostoru, koristeći funkciju

$$\phi(y) = P(X \leq x | Y = y) = \frac{E(X\mathbf{1}_{\{Y=y\}})}{P(Y = y)},$$

definiramo uvjetno očekivanje $E(X|Y)$ kao sl. varijablu

$$E(X|Y) = \phi(Y).$$

Uvjetno očekivanje $E(X|Y)$ možemo definirati i puno općenitije. Mogli bismo uvjetovati na općenitu sl. varijablu Y (ili čak na σ–algebru). Tada je dobro intuitivno razumijeti ovo očekivanje kao očekivanje od X uz informaciju koju nam o njoj prenosi sl. varijabla Y.

Formalno govorimo da je sl. varijabla Z jednaka $E(X|Y)$ ako se može prikazati kao $\phi(Y)$ za neku izmjerivu funkciju ϕ, te za za svaki Borelov podskup A od $\mathbb{R}$ vrijedi

$$E(Z\mathbf{1}_{\{Y\in A\}}) = E(X\mathbf{1}_{\{Y\in A\}}).$$

Takvih sl. varijabli može biti više, no one su nužno jednake g.s.

Ukoliko (X,Y) imaju zajedničku gustoću $f_{X,Y}$ tada funkciju ϕ možemo naći kao

$$\phi(y) =: E(X|Y\in \mathrm{d}y) = \int x\frac{f_{X,Y}(x,y)}{f_Y(y)}\mathrm{d}x,$$

te postaviti $E(X|Y) = \phi(Y)$ i na taj način izbjeći mukotrpnu diskusiju poput ove gore.

Intenzitet hazarda

Duljinu trajanja života slučajno odabrane osobe promatramo kao sl. var. T s vrijednostima u $[0, \infty)$. Osim funkcije distribucije sl. var. T nas će često zanimati i *funkcija doživljenja* od T tj.

$$\bar{F}(t) = P(T > t) = 1 - F(t)\,.$$

Mi u nastavku pretpostavljamo da je T neprekidna sl. var., tj da postoji njena gustoća i da je za $t \geq 0$ možemo izraziti kao

$$f(t) = \frac{\mathrm{d}}{\mathrm{d}t}F(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\bar{F}(t)\,.$$

Intenzitet hazarda ili *smrtnosti* [force of mortality/hazard rate] neprekidne sl. var. T definiramo kao

$$\mu(t) = \frac{f(t)}{\bar{F}(t)}$$

za sve t za koje je $\bar{F}(t) > 0$. Uočite da je

$$\mu(t) = \frac{f(t)}{\bar{F}(t)} = \frac{\mathrm{d}}{\mathrm{d}t}(-\log \bar{F}(t)) \tag{2}$$

za sve t za koje je $\bar{F}(t) > 0$ i u kojima je $\bar{F}$ derivabilna.

U literaturi se katkad funkcija $-\log \bar{F}(\cdot)$ naziva funkcijom hazarda, mi ćemo je zvati kumulativnom funkcijom hazarda.

Intuitivno, za malo $\mathrm{d}t$ je

$$\mu(t)\mathrm{d}t = \frac{f(t)\mathrm{d}t}{\bar{F}(t)} \approx \frac{P(t < T < t + \mathrm{d}t)}{P(T > t)} = P(T \leq t + \mathrm{d}t | T > t)\,.$$

Tj. vjerojatnosti smrti osobe starosti t u predstojećem intervalu duljine $\mathrm{d}t$ je proporcionalna $\mathrm{d}t$. Faktor proporcionalnosti je upravo intenzitet smrtnosti.

Pokažite da za $\varepsilon \to 0$, u točkama neprekidnosti od f:

$$\frac{P(T \leq t + \varepsilon | T > t)}{\varepsilon} \to \mu(t)\,.$$

Integrirajući (2) dobijemo

$$\bar{F}(t) = e^{-\int_0^t \mu(s)\mathrm{d}s},$$

kao i

$$f(t) = \mu(t)\bar{F}(t) = e^{-\int_0^t \mu(s)\mathrm{d}s}\mu(t)\,.$$

Napomena Očito su funkcije F, $\bar{F}$, f i μ ekvivalentni načini zadavanja zakona smrtnosti. Zbog $\bar{F}(\infty) = 0$ mora biti $\int_0^\infty \mu(s)\mathrm{d}s = \infty$. Funkciju $\Lambda(t) = \int_0^t \mu(s)\mathrm{d}s$ zovemo *kumulativna funkcija hazarda*.

Ako postoji najveća moguća životna dob u populaciji, recimo $w \geq 0$, (tj. $\bar{F}(w) = 0$ i $\bar{F}(t) > 0$ za $t < w$) tada

$$\int_0^t \mu(s)\mathrm{d}s \to \infty$$

za $t \to w$. Posebno ako pretpostavimo da je μ monotono rastuća, tada mora biti $\mu(t) \to \infty$, za $t \to w$.

Tipično promatramo osobu koja je već doživjela dob $x > 0$ (*life aged* x) i zanima nas njena preostala duljina života T_x. Razdioba od T_x je uvjetna razdioba od $T - x$ uz uvjet $T > x$. Njena je funkcija distribucije

$$F_{T_x}(t) = F(t|x) := P(T \leq t + x|T > x) = \frac{F(x+t) - F(x)}{1 - F(x)}$$

za sve $t \geq 0$, a funkcija doživljenja

$$\bar{F}(t|x) := P(T > t + x|T > x) = \frac{\bar{F}(x+t)}{\bar{F}(x)},$$

za sve $t \geq 0$. Gornji izrazi su dobro definirani samo za x t.d. $\bar{F}(x) > 0$. Gustoća ove uvjetne razdiobe je

$$f(t|x) = \frac{f(x+t)}{\bar{F}(x)}\,,$$

Pokažite da je uvjetni intenzitet smrtnosti $\mu(t|x) = \frac{f(t+x)}{\bar{F}(t+x)} = \mu(t+x)$.

Medjunarodne aktuarske oznake

The IAA propisuje sljedeće oznake

$$\begin{aligned}
{}_tq_x &= F(t|x) = P(T_x \le t), \\
{}_tp_x &= \bar{F}(t|x) = P(T_x > t) = 1 - {}_tq_x, \\
\mu_{x+t} &= \mu(x+t).
\end{aligned}$$

Posebno

$$\begin{aligned}
{}_tq_0 &= F(t), \\
{}_tp_0 &= \bar{F}(t), .
\end{aligned}$$

Ako je $t = 1$ može se izostaviti $q_x = {}_1q_x$ i $p_x = {}_1p_x$.
Dok očekivanu preostalu duljinu života označavamo sa

$$\mathring{e} = \bar{e}_x = E(T_x) = \int_0^\infty t f(t|x)\mathrm{d}t = \int_0^\infty {}_tp_x \mathrm{d}t\,.$$

Primjer (zaboravljivost eksponencijalne razdiobe) Ako je duljina života u nekoj populaciji eksponencijalno distribuirana, tj. $T \sim \text{Exp}(\lambda)$, $\lambda > 0$, tada

$$\mu(x) = \lambda \quad \text{i} \quad \bar{F}(t|x) = \bar{F}(t),\ x, t > 0$$

Kako se mnogi tokovi novca odvijaju u diskretnim vremenskim trenucima ponekad je korisno diskretizirati duljinu života. Pretpostavljamo dakle da je T neprekidna sl. var., ali promatramo sl. var. $K = \lfloor T \rfloor$ s vrijednostima u skupu $\mathbb{N}_0$.

Ako je $T \sim \text{Exp}(\lambda)$ kao gore, tada će K imati geometrijsku razdiobu s parametrom $1 - p = 1 - e^{-\lambda}$, tj. za $k \in \mathbb{N}_0$

$$P(K = k) = P(k \le T < k + 1) = e^{-k\lambda} - e^{-(k+1)\lambda} = p^k(1 - p)\,.$$

Općenito za preostalu duljinu života u godinama $K_x = \lfloor T_x \rfloor$, uz uvjet neprekidnosti razdiobe od T dobijemo

$$P(K_x = k) = P(T_x \in [k, k+1)) = {}_k p_x \, q_{x+k} \, .$$

Tako da

$$e_x = E(K_x) = \sum_{k=0}^{\infty} k P(K_x = k) = \sum_{k=0}^{\infty} k \, {}_k p_x \, q_{x+k} = \sum_{k=0}^{\infty} {}_k p_x \, .$$

Uvedimo $T_x = K_x + S$, katkad koristimo aproksimaciju $\bar{e}_x \approx e_x + 1/2$.

Primjetite

$$P(S \le u | K_x = k) = \frac{{}_u q_{x+k}}{q_{x+k}} \, .$$

Ako su S i K_x nezavisne sl. var. tada je gornji izraz neovisan o k, tj. $P(S \le u | K_x = k)$ je isključivo funkcija od u, recimo $H(u)$.

Primjer $H(u) = u$ za $u \in [0,1]$ akko je $S \sim \mathsf{U}(0,1)$.

Slično definiramo $S^{(m)}$ kao duljinu života u posljednjoj godini zaokruženu na prvi veći od brojeva $1/m, 2/m, \ldots, 1$, tj.

$$S^{(m)} = \frac{1}{m}\lfloor mS + 1 \rfloor.$$

Ako su K i S nezavisni, a $S \sim \mathsf{U}(0,1)$ tada $S^{(m)}$ ima diskretnu uniformnu razdiobu na $\{1/m, 2/m, \ldots, 1\}$.

Zakoni smrtnosti

Razdiobe sl. var. T možemo dakle zadati preko $f, \ F, \ \bar{F}$ ili μ

- de Moivreov zakon (1724)

  Za $w > 0$, neka je

$$f(x) = \frac{1}{w} \qquad \text{na } (0, w)$$

- Gompertz (1824) Za realne parametre $B > 0$ i $c > 0$

$$\mu(x) = Bc^x \qquad \text{za } x > 0.$$

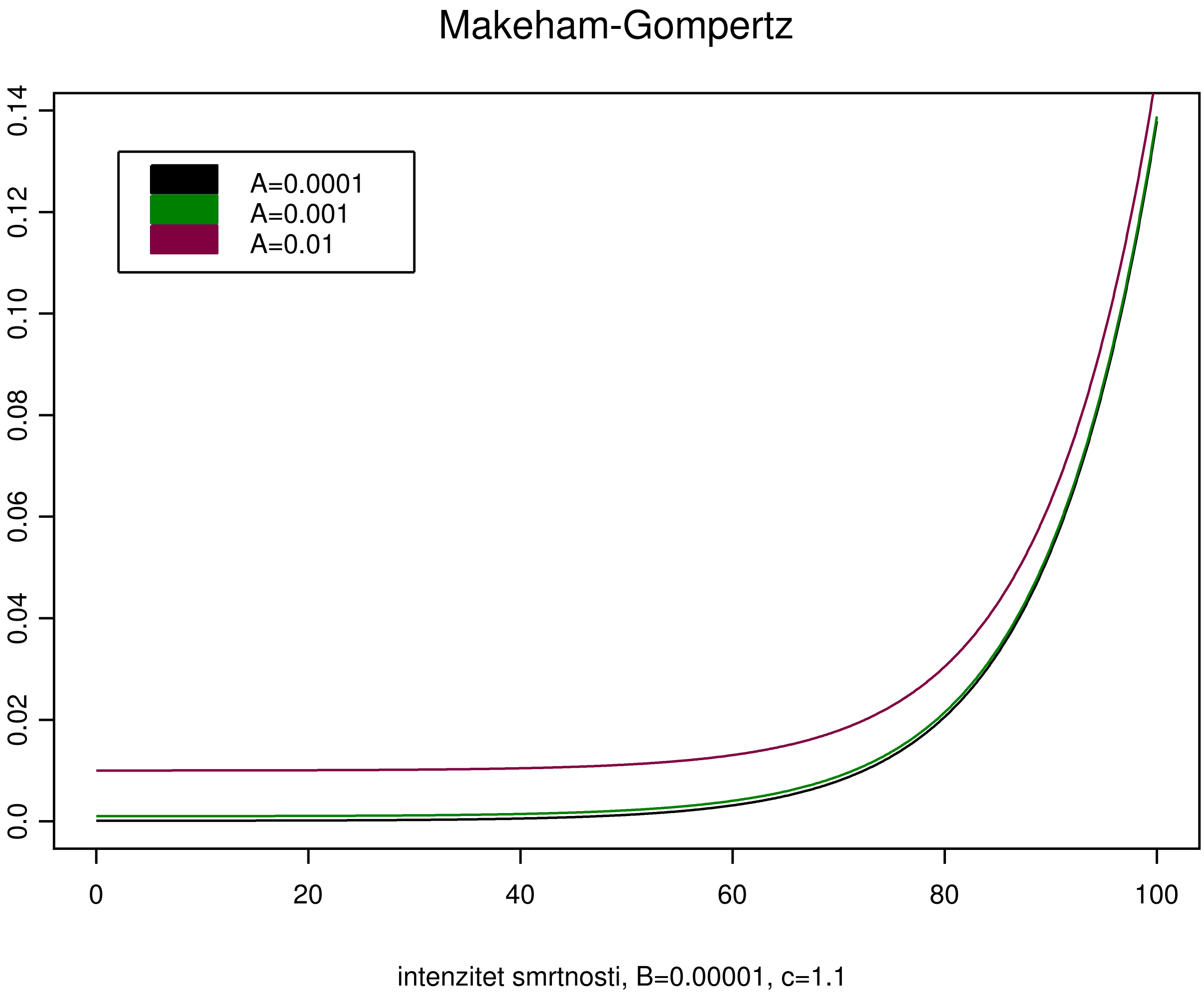
- Makeham (1860) Za realne parametre $A > 0,\ B \geq 0$ i $c > 0$

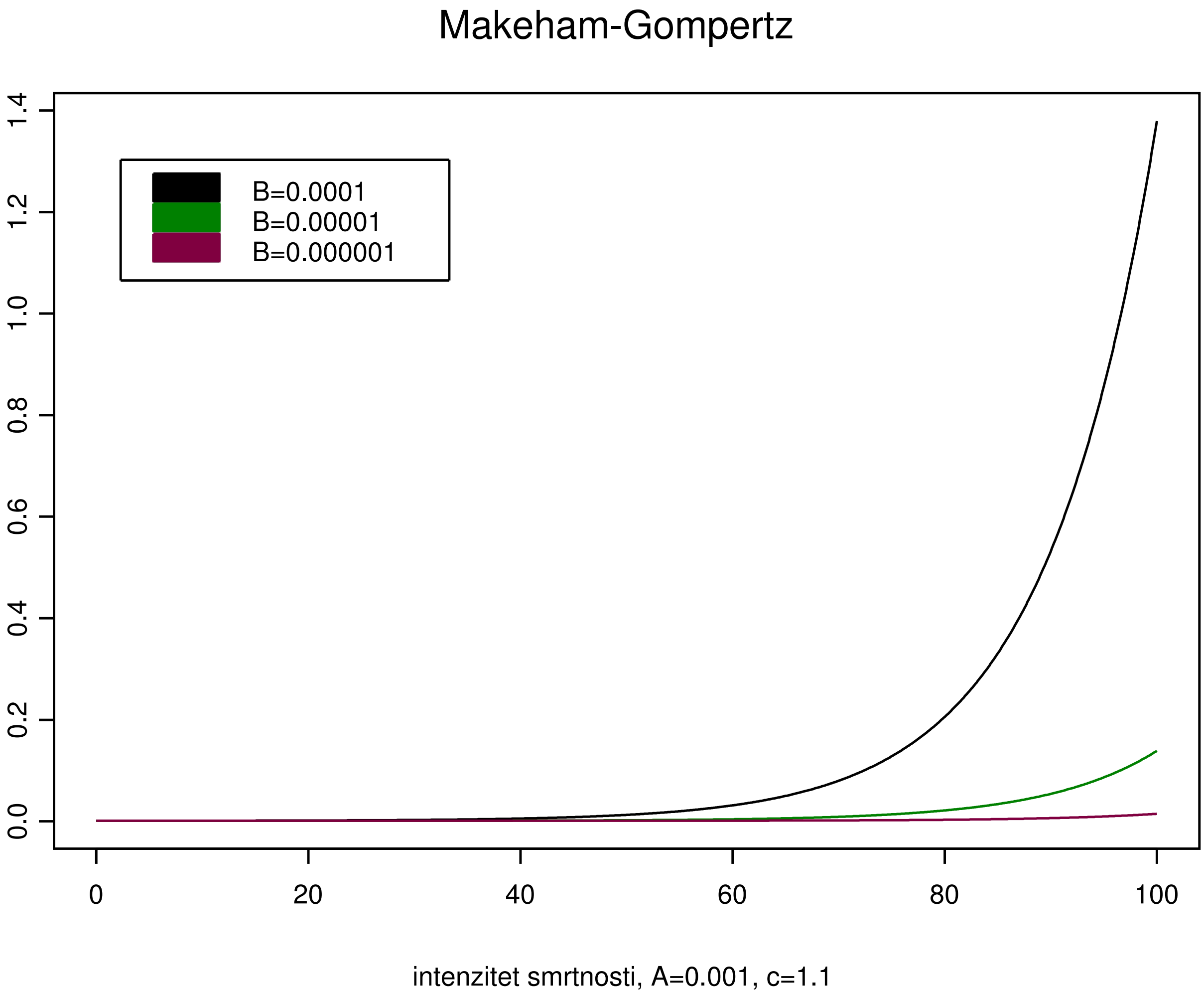
$$\mu(x) = A + Bc^x \qquad \text{za } x > 0.$$

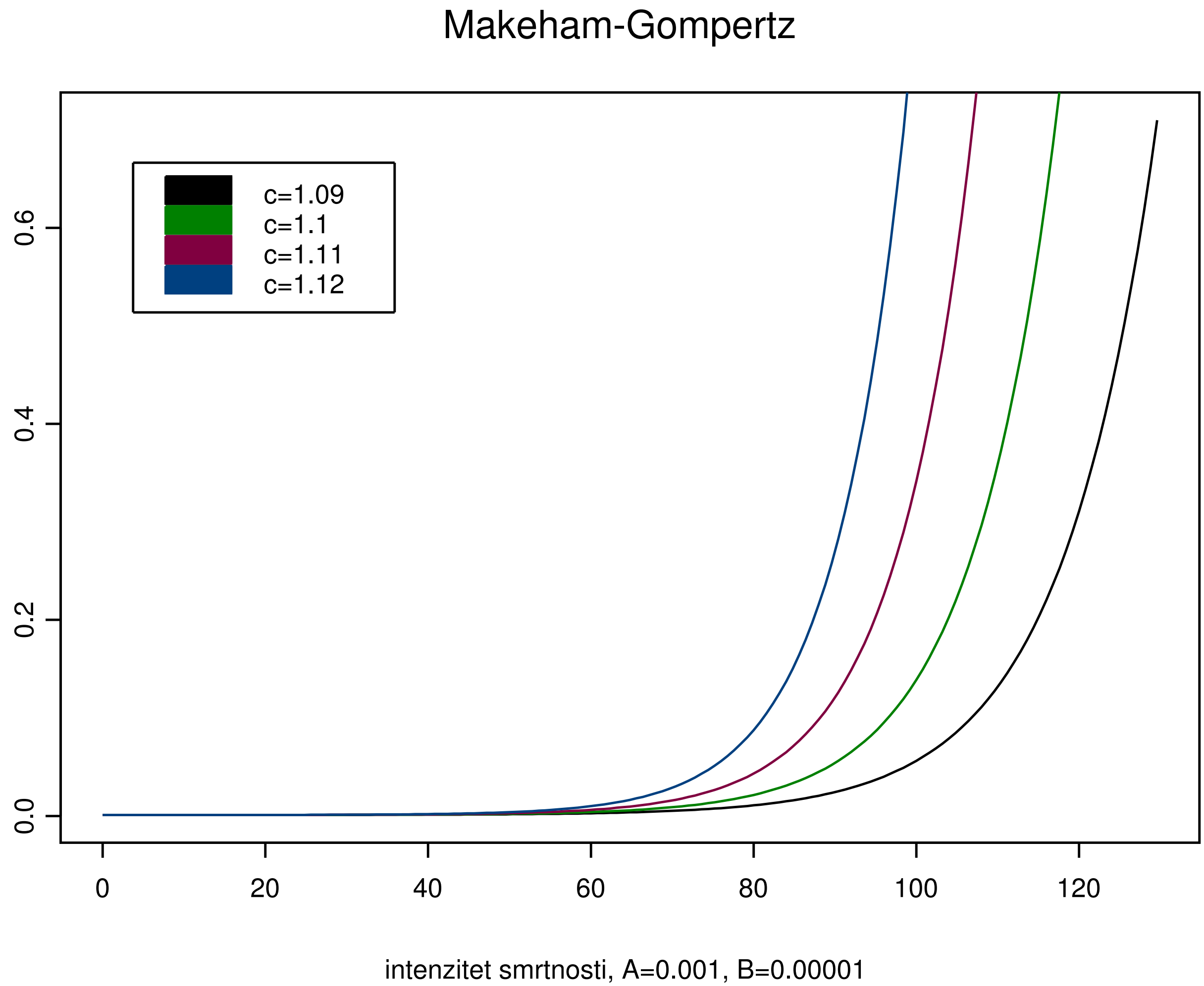
Pokažite da tada

$$\bar{F}(t|x) = \exp\left(-At - mc^x(c^t - 1)\right),$$

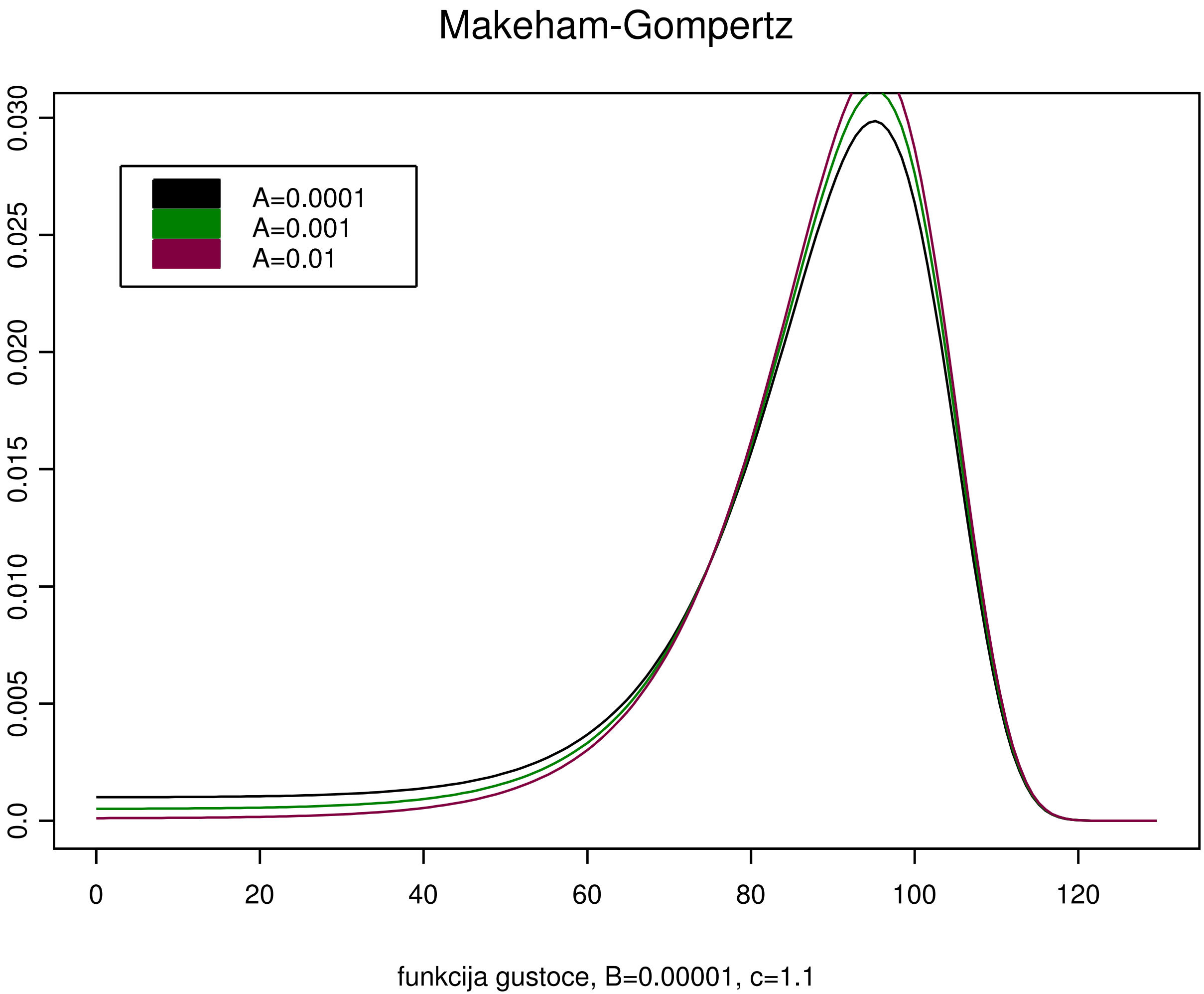
gdje je $m = B/\log c$.

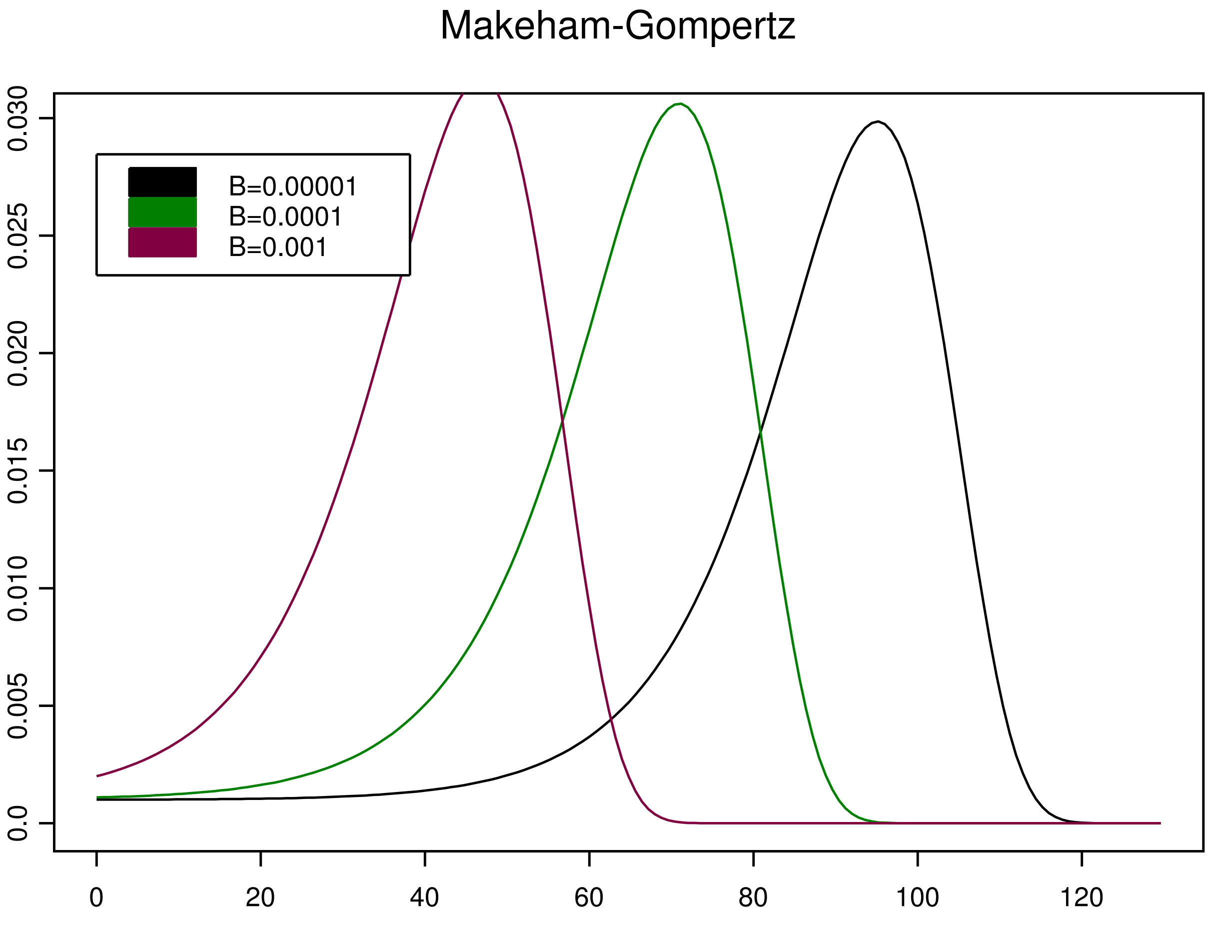

Makeham-Gompertz
A=0.0001
A=0.001
A=0.01
0.0
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0
20
40
60
80
100
intenzitet smrtnosti, B=0.00001, c=1.1

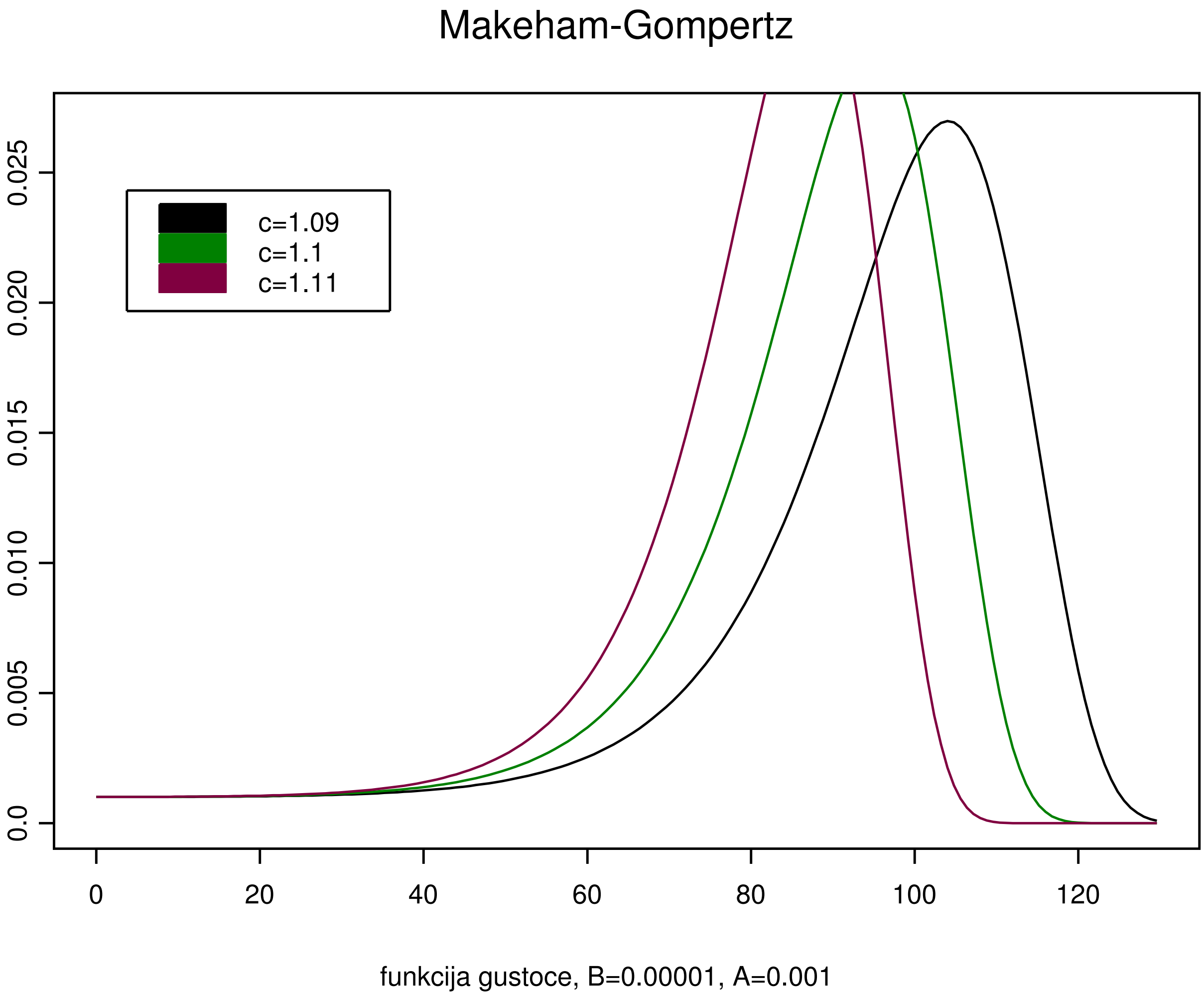

Makeham-Gompertz
B=0.0001
B=0.00001
B=0.000001
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
20
40
60
80
100
intenzitet smrtnosti, A=0.001, c=1.1



Makeham-Gompertz
c=1.09
c=1.1
c=1.11
c=1.12
0.0
0.2
0.4
0.6
0
20
40
60
80
100
120
intenzitet smrtnosti, A=0.001, B=0.00001


Makeham-Gompertz
A=0.0001
A=0.001
A=0.01
0.0
0.005
0.010
0.015
0.020
0.025
0.030
0
20
40
60
80
100
120
funkcija gustoce, B=0.00001, c=1.1


Makeham-Gompertz
B=0.00001
B=0.0001
B=0.001
0.0
0.005
0.010
0.015
0.020
0.025
0.030
0
20
40
60
80
100
120
funkcija gustoce, A=0.001, c=1.1


Makeham-Gompertz
c=1.09
c=1.1
c=1.11
0.0
0.005
0.010
0.015
0.020
0.025
0
20
40
60
80
100
120
funkcija gustoce, B=0.00001, A=0.001

- Weibull (1939) Za realne parametre $\kappa > 0,\ \alpha \geq 0$

$$\mu(x) = \kappa\, x^{\alpha} \qquad \text{za } x > 0.$$

ako je $\alpha = 0$, razdioba je zapravo eksponencijalna.

- Dvostruki geometrijski zakon (1867)

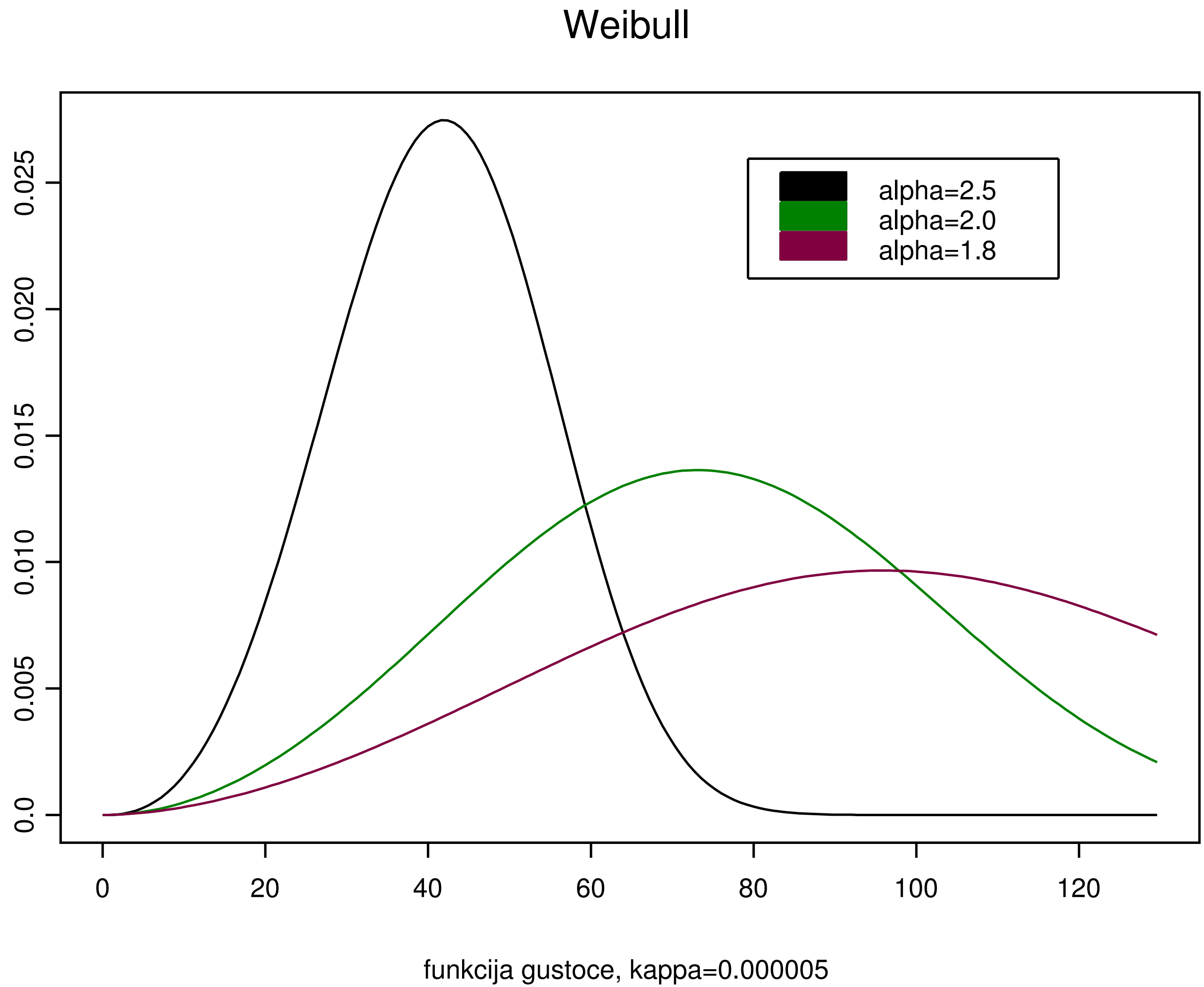
$$\mu(x) = Bc^{x} + Mn^{x} \qquad \text{za } x > 0.$$

- Makeham II (1889)

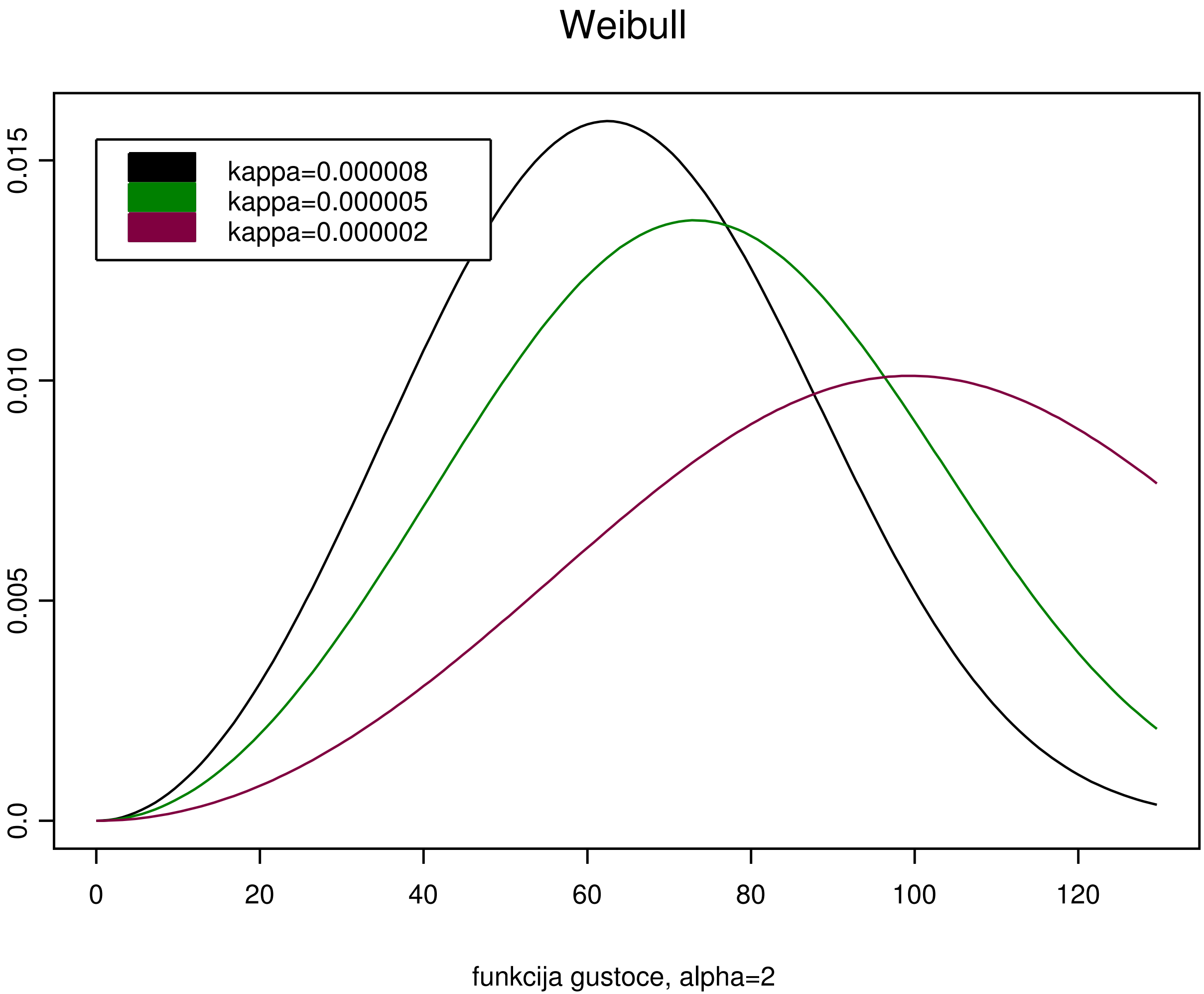
$$\mu(x) = A + Hx + Bc^{x} \qquad \text{za } x > 0.$$

- Perk (1931)

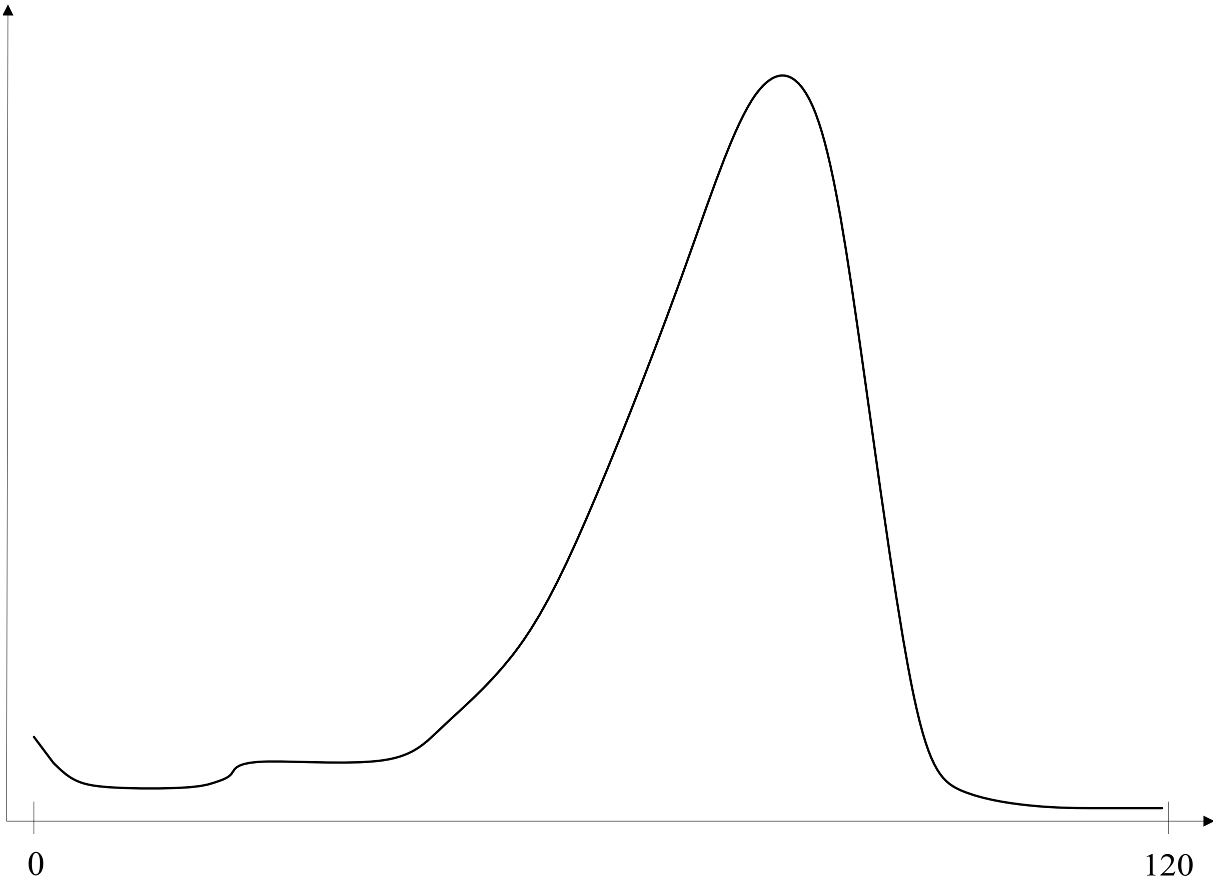
$$\mu(x) = \frac{A + Bc^{x}}{Kc^{-x} + 1 + Dc^{x}} \qquad \text{za } x > 0.$$



Weibull
alpha=2.5
alpha=2.0
alpha=1.8
0.0
0.005
0.010
0.015
0.020
0.025
0
20
40
60
80
100
120
funkcija gustoce, kappa=0.000005



Weibull
kappa=0.000008
kappa=0.000005
kappa=0.000002
0.0
0.005
0.010
0.015
0
20
40
60
80
100
120
funkcija gustoce, alpha=2



Empirijska procjena fukcije gustoće za duljinu života

u tipičnom suvremenom zapadnoevr. društvu

Razdiobe s monotonim intenzitetom smrtnosti

Promotrimo prvo diskretne razdiobe s vrijednostima u $\mathbb{N}_0$. Za njih možemo definirati (diskretni) intenzitet smrtnosti ili hazarda kao

$$m(n) = P(K = n | K \geq n), \quad n \in \mathbb{N}.$$

Intenzitet smrtnosti je često samo lokalno monotona funkcija. Za diskretne razdiobe pak, funkcija m je nerijetko monotona. Ako je duljina života T neprekidna sl. var., tada je za $K = \lfloor T \rfloor$, intenzitet smrtnosti $m(n) = q_n$, $n \in \mathbb{N}$.

I funkciju doživljenja za K možemo izraziti preko m, npr. za $k \in \mathbb{N}_0$ imamo

$$\bar{F}(k) = \prod_{i=0}^{k} (1 - m(i)),$$

a gustoću sl. var. K nadjemo kao

$$P(K = k) = m(k)\bar{F}(k-1), \ k \in \mathbb{N}.$$

Primjer Niz $(m(n))$ može biti i rastući i padajući ali i konstantan, to je naime slučaj za geometrijsku razdiobu. Naime u tom je primjeru $m(n) = 1-p$, za sve $n \in \mathbb{N}_0$, za parametar $p \in (0,1)$. Lako se vidi i da geometrijska razdioba ima svojstvo (diskretne) zaboravljivosti, tj. ako je K geometrijski distribuirana

$$P(K \geq i+j|K \geq j) = P(K \geq i), \qquad \text{za sve } i,j \in \mathbb{N}_0.$$

Kažemo da je razdioba diskretne sl. var. K, razdioba rastućeg hazarda ako je niz $(m(n))$ monotono rastući. Za niz realnih brojeva (p_k) kažemo da je log-konkavan ako $p_{k+1}^2 \geq p_{k+2}p_k$, za svaki $k \in \mathbb{N}$ (odn. log-konveksan ako vrijedi obrnuta nejednakost).

Propozicija 8 *Ako je niz $p_k = P(K = k)$, $k \in \mathbb{N}_0$, strogo pozitivan i log-konkavan tada je razdioba sl. var. K monotono rastućeg hazarda.*

Napomena Svaka Poissonova razdioba je razdioba rastućeg hazarda. Negativna binomna razdioba je razdioba monotono rastućeg hazarda ako $\alpha \geq 1$.

Ako je $T \sim \text{Exp}(\lambda)$, znamo da za sve $s, t \geq 0$ vrijedi

$$P(T > s + t | T > s) = P(T > t).$$

Pokažite da gornja tvrdnja vrijedi i ako s zamijenimo bilo kakvom nenegativnom i neprekidnom sl. varijablom nezavisnom s T.

Promotrimo sada funkciju hazarda neprekidnih razdioba. Mi smo već uveli

$$F_{T_x} = F(t|x) = P(T - t \leq x | T > x), \qquad t > 0,$$

uvedim još i *funkciju očekivanja ostatka* (odn. očekivane preostale duljine života)

$$\bar{e}_x := E(T - x | T > x), \qquad x \geq 0.$$

Parcijalnom integracijom pokažite da za neprekidnu i nenegativnu sl. var. T vrijedi

$$\bar{e}_x = \frac{\displaystyle\int_x^{\infty} \bar{F}(t)\mathrm{d}t}{1 - F(x)}\,.$$

Uvedimo parcijalni uredjaj na skup funkcija distribucije (tj. razdioba), naime kažemo da je dist. F stohastički manja od dist. G. i pišemo

$$F \overset{st}{\leq} G$$

ako

$$\bar{F}(x) \leq \bar{G}(x), \qquad x \in \mathbb{R}\,.$$

Za funkciju hazarda neprekidnih razdioba vrijedi da su monotono rastuće akko je familija razdioba $F(\,|x)$ stohastički padajuća. Tada kažemo da je T, odn. njena razdioba, rastućeg hazarda.

Propozicija 9 *Neprekidna razdioba s fjom distribucije* F *je monotono rastućeg hazarda akko je za sve* $0 \leq x \leq y < \infty$

$$F_{T_x} \overset{st}{\geq} F_{T_y}\,.$$

Razdioba od T_x nije važna samo za osiguranje života, već i u općoj teoriji osiguranja. Npr. ako je X ukupna šteta za osig. kompaniju u danom periodu,

$$\bar{e}_x := E(X - x | X > x)$$

je očekivana vrijednost štete kod tzv. stop-loss reosiguranja.

U teoriji pouzdanosti uvode se i pojmovi NBU [new better than used] i NWU [new worse than used] razdioba ako

$$F_{T_x} \stackrel{st}{\leq} F \qquad x \geq 0,$$

odn.

$$F_{T_x} \stackrel{st}{\geq} F \qquad x \geq 0.$$

Očito je razdioba rastućeg hazarda i NBU.

Tablice smrtnosti

Parametarski modeli nam nisu potrebni ako postoji dovoljno empirijskih podataka o zakonu smrtnosti.

Prve tablice smrtnosti su sastavili J. Graunt (1662) i Sir E. Halley (1693). Nedugo zatim proučavali su ih poznati matematičari tog vremena, medju ostalima J. Bernoulli i C. Huygens. Danas su sveprisutne, a sastavljaju se i u Hrvatskoj.

Obično, tablica smrtnosti za $k \in \mathbb{N}_0$ sadrži brojeve

$$l_k = \text{ (očekivani) broj osoba u danoj populaciji koji je živ u dobi } x\,.$$

Posebno, broj

$$l_0 = \text{ (očekivani) broj novorodjenih osoba u danoj populaciji}\,,$$

se još zove i korijen tablice. Ovo je zapravo samo korisna interpretacija brojeva u tablici, no uvijek možemo staviti da je npr. $l_0 = 10^6$ ili $l_0 = 10^3$.

Upravo se funkcija l_x, zove tablica smrtnosti. U tablicama se tipično nalaze i brojevi

$$d_k = \text{ broj smrti u intervalu } [k, k+1) = l_k - l_{k+1}.$$

Napomena Tablicama je moguće dati i deterministčku interpretaciju, tj. pretpostavimo da je l_x upravo broj osoba koje će biti žive u dobi x. I takva prepostavka omogućuje da korektno izračunamo premije, no ona predstavlja problem ako moramo koristiti bilo koju metodu statističkog zaključivanja, zato je mi nećemo koristiti.

Iz tablica je lako *odrediti* razne vjerojatnosti, npr.

$$\bar{F}_T(k) = P(T > k) = P(T \geq k) = {}_kp_0 = l_k/l_0\,.$$

Uočite da gore pišemo znak jednakosti iako se u praksi zapravo uvijek radi o procjeni. Na intervalima $[k, k{+}1)$, l_x odn. ${}_xp_0$ se može dobiti interpolacijom.

Nadjite izraze i za ${}_kp_n$ te ${}_kq_n$ preko funkcije l_x. Npr.

$$q_k = d_k/l_l \;\; k \geq 0.$$

Napomena Mi pretpostavljamo da je ukupna duljina života neprekidna sl. var., posebno za sve $t > 0$, $P(T_x = t) = P(T = t) = 0$. Zato je za $k \in \mathbb{N}_0$

$$P(K_x = k) = P(k \leq T_x < k+1) = P(T_x \geq k) - P(T_x \geq k+1)\,.$$

No za $k \geq 1$

$$\begin{aligned}
P(T_x \geq k) &= P(T_x \geq k-1)\frac{P(T_x \geq k)}{P(T_x \geq k-1)} \\
&= P(T_x \geq k-1)\frac{P(T - x \geq k | T > x)}{P(T - x \geq k-1 | T > x)} \\
&= P(T_x \geq k-1)\frac{P(T \geq k+x)}{P(T \geq x+k-1)} \\
&= P(T_x > k-1)\frac{P(T > k+x)}{P(T > x+k-1)} \\
&= {}_{k-1}p_x \; p_{x+k-1} = \cdots = \prod_{l=0}^{k-1}(1 - q_{x+l})
\end{aligned}$$

Tablice s odabirom

Osiguravajuća društva obično dijele potencijalne klijente po spolu, ganeraciji i sl. No često imaju razloga pretpostaviti i da se potpisnici ugovora o osiguranju razlikuju od ostalog dijela populacije. Te osobe npr.
- prolaze medicinski pregled prije sklapanja ugovora ili
- vjeruju da su dobroga zdravlja te na toj osnovi sklapaju ugovor o životnoj renti.

Tablice za ovakve osobe se obično zovu *select* tj. tablice s odabirom. U njima se vjerojatnosti nešto drugačije označavaju. Npr. vjerojatnost da će takva osoba koja je pristupila osiguranju u dobi x, a sada je u dobi $x+t$ umrijeti u sljedećih godinu dana je sa

$$q_{[x]+t}$$

Nakon izvjesnog perioda ovakve osobe imaju jednaku funkciju doživljenja kao i ostatak populacije. *Period odabira* obično traje $r \in \mathbb{N}$ godina. Nakon toga se koriste tzv. *ultimate (krajnje)* tablice. Dakle

$$q_{[x]+t} = q_{x+t}, \qquad \text{za sve } t \geq r \; .$$

Dakle ako tražimo razdiobu sl. var. $K = K_x = \lfloor T_x \rfloor$ i ako je period odabira 3 godine imamo

$$P(K_x = k) = q'_{x+k} \prod_{l=0}^{k-1} (1 - q'_{x+l})$$

gdje je $(q'_x, q'_{x+1}, q'_{x+2}, q'_{x+3}, q'_{x+4}, \ldots) = (q_{[x]}, q_{[x]+1}, q_{[x]+2}, q_{x+3}, q_{x+4}, \ldots)$.
Tablice koje ne razlikuju odabranu populaciju zovu se *aggregate*, tj sveukupne ili agregatne.

Vjerojatnost smrti u dijelovima godine

Životne tablice zadaju samo zakon razdiobe sl. var. $K = \lfloor T_x \rfloor$, tj. duljine života u godinama

$$P(K \geq k) = {}_k p_x = p_x p_{x+1} \cdots p_{x+k+1}, \qquad k = 1, 2, \ldots$$

Razdioba sl. var. T može se dobiti interpolacijom uz neke dodatne pretpostavke o funkcijama

$${}_u q_x \;\text{ odn. }\; \mu_{x+u} \qquad \text{za } u \in (0,1)\,.$$

Česte prepostavke su, za $x \in \mathbb{N}_0,\ u \in (0,1)$:

a) linearnost funkcije ${}_u q_x$, tj.

$${}_u q_x = u \cdot q_x\ ,$$

što je ekvivalentno tome da $\mu_{x+u} = q_x/(1 - u q_x)$ odn. K i S su nezavisne, te

$$T - K = S \sim U(0,1).$$

b) intenzitet smrtnosti tokom godine je konstantan i jednak nekoj konstanti $\mu_{(x+1/2)}$. Tada mora biti

$$\mu_{(x+1/2)} = -\ln p_x \quad \text{i} \quad {}_u p_x = e^{-u\mu_{(x+1/2)}} = p_x^u \,,$$

a S ima odrezanu eksponencijalnu razdiobu. Možemo reći dakle da T ima po dijelovima eksponencijalnu razdiobu

c) linearna je funkcija ${}_{1-u}q_{x+u}$ (Balduccijeva pretpostavka) tj.

$${}_{1-u}q_{x+u} = (1-u)q_x \,.$$

Napomena Za ove prepostavke možemo reći da su "poluparametarske". Nijedna od njih nije vrlo realistična, intenzitet smrtnosti u sva tri slučaja ima skokove, a u slučaju c) je čak padajuća funkcija na intervalima $[k, k+1)$, $k \in \mathbb{N}$.