

Sandsynlighedsregning

6. forelæsning

Bo Friis Nielsen

Matematik og Computer Science
Danmarks Tekniske Universitet
2800 Kgs. Lyngby – Danmark
Email: bfn@dtu.dk

Antag, at rosinboller fra Genbrugsbageren har gennemsnitligt 3 friske og 2 rådne rosiner pr. bolle.

Spørgsmål 1

Hvad er sandsynligheden for ikke at få nogen rosiner i en bid, der tager 20% af en bolle?

- 1 $\frac{3}{5}$
- 2 0.368
- 3 0.2514
- 4 $\frac{1}{5} \left(\frac{3}{5}\right)^5$
- 5 0.0067
- 6 Ved ikke

Dagens emner: Afsnit 4.2, 4.3 og 4.4



- Poissonprocessen/eksponentialfordelingen $f(x) = \lambda e^{-\lambda x}$
 - ◊ Rate λ Overlevelseshæderfunktion: $G(x) = e^{-\lambda x}$
 - ◊ Gammafordelingen $f(x) = \lambda \frac{(\lambda x)^{r-1}}{\Gamma(r)} e^{-\lambda x}$ hvor $\Gamma(r) = (r-1)!$ for r heltallig. For r heltallig $P(X \leq x) = 1 - \sum_{i=0}^{r-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x}$
- Hazard-funktionen (hazard rate)

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T \in (t, t + \Delta t) | T > t)}{\Delta t} = \frac{f(t)}{G(t)}$$

- Variabelskift, $P(X \in [x, x + dx]) = f_X(x)dx$, $Y = g(X)$

$$P(Y \in [y, y + dy]) = \frac{f_X(g^{-1}(y))}{\left| \frac{dy}{dx} \right|}$$

Punktprocesser på linien



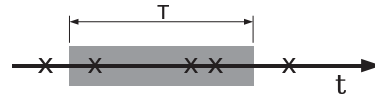
- Uheld i et vejkryds
- Radioaktivt henfald
- Opkald til internetudbyder
- Produktionsfejl på et kabel



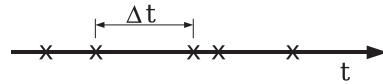
Fællestræk: Vi kan modellere ...



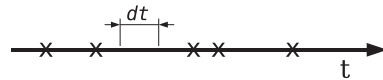
Antal i et tidsinterval



Tid mellem hændelser



Et lille tidsinterval



Een grundlæggende model:



Poissonprocessen

(side 289)

- Antal hændelser i et tidsinterval er Poissonfordelt (2. p.289)
- Tiden mellem hændelser er eksponential fordelt (3. p.289)
- Konstant intensitet (sandsynlighed for en hændelse i Δt er proportional med Δt) (1. p.289).

Eksponentialfordelingen side 279



En eksponentialfordelt (λ) variabel T har tæthed

$$f(t) = \lambda e^{-\lambda t}$$

overlevelsesfunktion

$$G(t) = P(T > t) = e^{-\lambda t}$$

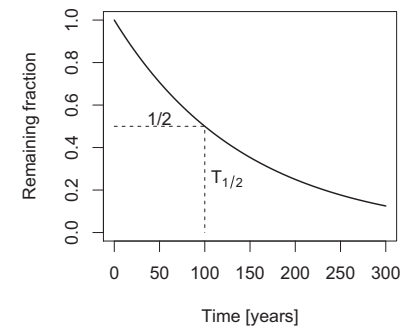
middelværdi og varians

$$E(T) = \frac{1}{\lambda} \quad \text{Var}(T) = \frac{1}{\lambda^2}$$

Parameteren λ er intensiteten eller (*døds*)raten (hazard rate).

Eksponentialfordelingen er *uden hukommelse* (side 279 nederst).

Halveringstid



Antag at en type af radioaktive atomer har halveringstid 1 år.

Spørgsmål 2



Sandsynligheden for, at et atom af denne type overlever 5 år, er

- 1 $\frac{1}{5}$
- 2 e^{-5}
- 3 $\frac{1}{32}$
- 4 $1 - \Phi\left(\frac{5-1}{1}\right)$
- 5 $1 - \Phi\left(\frac{5-1}{\sqrt{5}}\right)$
- 6 Ved ikke

Hvor Φ som sædvanlig betegner fordelingsfunktionen for en standardiseret normalfordelt variabel.

Opgave 4.2.6



En Geigertæller opsamler baggrundsstråling med en gennemsnitlig rate af en registrering pr. minut.

Lad T_3 betegne det tidspunkt hvor den tredje partikel registreres (målt i minutter).

Spm: Bestem $P(2 \leq T_3 \leq 4)$

Ventetid til r 'te hændelse: Gammafordelingen

Ofte vil vi beskrive tiden til den r 'te hændelse.

Sammenhæng mellem antal (N_t) og tid (T_r).

$$P(N_t \leq r - 1) = P(T_r > t)$$

For Poisson-processen fås:

$$P(T_r > t) = \sum_{i=0}^{r-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t}$$

Fra dette kan vi udlede tætheden for Gammafordelingen

$$f(t) = -\frac{d}{dt}P(T_r \leq t) = \lambda \frac{(\lambda t)^{r-1}}{(r-1)!} e^{-\lambda t}$$

Svar: T_3 antages Gamma (3,1) fordelt (hvorfor?)



Således er

$$P(T_3 > 4) = e^{-1 \cdot 4} \left(1 + 4 + \frac{4^2}{2}\right) = 0.238$$

$$P(T_3 \geq 2) = e^{-1 \cdot 2} \left(1 + 2 + \frac{2^2}{2}\right) = 0.677$$

... og dermed

$$P(2 \leq T_3 \leq 4) = P(T_3 \geq 2) - P(T_3 > 4) = 0.439$$

Gammafordeling med ikke-heltallig parameter

Definér Gammafunktionen (s. 291):



$$\Gamma(r) = \int_0^{\infty} t^{r-1} e^{-t} dt$$


For heltallig parameter r har vi $\Gamma(r) = (r-1)!$

Hvis X har tætheden

$$f(t) = \frac{1}{\Gamma(r)} \lambda^r t^{r-1} e^{-\lambda t}$$

så er X Gammafordelt med formparameter r og skalaparameter λ .

Gammafordelingen bruges ofte til at beskrivelse af levetider, etc. der ikke kan beskrives ved eksponentialfordelingen.

En kontinuert stokastisk variabel T har overlevelsesfunktionen 
 $G(t) = P(T > t)$.

Spørgsmål 3

Bestem $P(a < T \leq b)$

- 1 $\frac{G(b)-G(a)}{G(b)}$
- 2 $\frac{G(a)-G(b)}{G(a)}$
- 3 $G'(b) - G'(a)$
- 4 $G(a) - G(b)$
- 5 $G(b) - G(a)$
- 6 Ved ikke

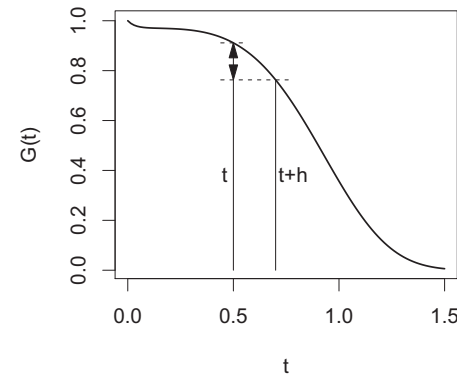
Hazard rate side 296



Under service af en bil overvejer man præventivt at skifte en komponent (f.eks. en tandrem). Komponenten har allerede alderen t , næste service er til tiden $t+h$, og komponentens levetid har overlevelsesfunktion $G(\cdot)$.

Spm: Hvad er sandsynligheden for, at komponenten svigter inden næste service?

Svar:



Det relevante er den **be-tingede** sandsynlighed:

$$P(T \leq t+h \mid T > t)$$

som bestemmes til

$$\frac{G(t) - G(t+h)}{G(t)}$$

Hazard rate

For korte tidsrum („serviceintervaller“) h har vi

$$\frac{G(t) - G(t+h)}{G(t)} = \frac{f(t)}{G(t)} \cdot h + o(h)$$

Vi definerer nu **hazard raten** $\lambda(t)$ for sandsynligheden for at hændelsen indtræffer kort efter t , **givet** at den ikke er indtruffet før.

$$\lambda(t) = \frac{f(t)}{G(t)}$$

Sandsynligheden for at overleve indtil næste „service“ findes som

$$P(T > t+h | T > t) = \exp\left(-\int_t^{t+h} \lambda(s) ds\right)$$

Hazardraten bestemmer sandsynligheden for at en hændelse indtræffer (fejl, død, ankomst) kort efter et givet tidspunkt t , **givet** at den ikke er indtruffet før.

Overlevelsesfunktionen fra hazard rate



$$\lambda(t) = \frac{f(t)}{G(t)} = -\frac{G'(t)}{G(t)}$$

D.v.s. $G(t)$ løser

$$G'(t) = -\lambda(t) \cdot G(t) \quad , \quad G(0) = 1$$

som har løsningen

$$G(t) = e^{-\int_0^t \lambda(u) du}$$

- Eksponentialfordelingen som eksempel:

$$G(t) = e^{-\int_0^t \lambda du} = e^{-\lambda t}$$

Hazard rate for eksponentialfordelingen



Vi har $f(t) = \lambda e^{-\lambda t}$. Så

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

... d.v.s. eksponentialfordelingen er **hukommelsesløs**.

Eksponentialfordelingen er *den eneste* fordeling med denne egenskab (opgave 4.3.1)

Fra tæthed til overlevelsesfunktion



$$G(t) = \int_t^\infty f(u) du \quad \Leftrightarrow \quad f(t) = -\frac{dG}{dt}(t) \quad , \quad G(0) = 1$$

Eksponentialfordeling (λ)	Gammafordeling (λ, r)
$f(t) = \lambda e^{-\lambda t}$	$f(t) = e^{-\lambda t} \frac{\lambda^r t^{r-1}}{(r-1)!}$
$G(t) = e^{-\lambda t}$	$G(t) = e^{-\lambda t} \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!}$

Forventningsværdien i en overlevelsesfordeling

Lad T være en ikke-negativ kontinuert stokastisk variabel.

T 's forventningsværdi er

$$E(T) = \int_0^{\infty} t f(t) dt = \int_0^{\infty} G(t) dt$$

hvor G er overlevelsesfunktionen og ligheden skyldes partiel integration. (Dette er den kontinuerte version af *tail sum formula*.)

For en eksponentialfordelt (λ) variabel finder vi

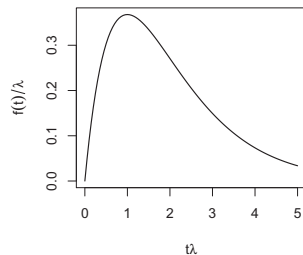
$$E(T) = \int_0^{\infty} \exp(-\lambda t) dt = \frac{1}{\lambda}$$

Vi beskriver komponenternes levetid ved stokastiske variable X_i .

Levetiden T af apparatet er da givet ved $T = X_1 + X_2$.

Da X_i er uafhængige og eksponentialfordelte (λ) er T Gamma($2, \lambda$) fordelt.

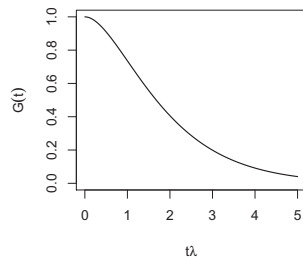
$$f(t) = \lambda^2 t e^{-\lambda t}$$



Spm: Bestem overlevelsesfunktionen

Svar:

$$G(t) = e^{-\lambda t} (1 + \lambda t)$$



Opgave R.4.20

Et apparat har en kritisk komponent med konstant fejlrate λ .

Apparatet har en tilsvarende komponent som backup. Når den første komponent fejler bliver dens funktion straks overtaget af den anden, som derefter har konstant fejlrate λ uafhængigt af, hvornår den første komponent fejlede.

Lad T være levetiden af apparatet bestemt ved det tidspunkt erstatningskomponenten fejler.

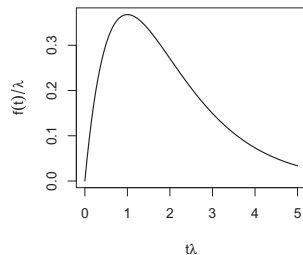
Spm: Bestem tæthedsfunktionen for T .

Vi beskriver komponenternes levetid ved stokastiske variable X_i .

Levetiden T af apparatet er da givet ved $T = X_1 + X_2$.

Da X_i er uafhængige og eksponentialfordelte (λ) er T Gamma($2, \lambda$) fordelt.

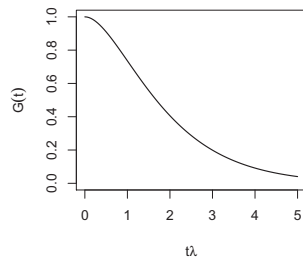
$$f(t) = \lambda^2 t e^{-\lambda t}$$



Spm: Bestem overlevelsesfunktionen

Svar:

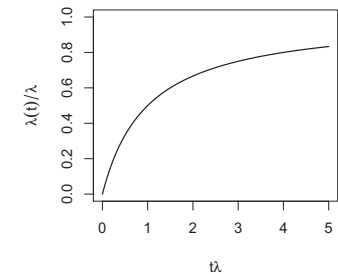
$$G(t) = e^{-\lambda t} (1 + \lambda t)$$



Spm: Bestem fejlfunktionen (hazard-rate)

Svar:

$$\lambda(t) = \frac{\lambda^2 t e^{-\lambda t}}{e^{-\lambda t} (1 + \lambda t)} = \frac{\lambda}{1 + \frac{1}{\lambda t}}$$



Med $\lambda = 1$ per time, hvad er da sandsynligheden for, at et apparat, der har virket i to timer, fejler i det næste minut

$$P\left(T \leq 2 + \frac{1}{60} \text{ timer} \mid T > 2 \text{ timer}\right) = \lambda(2) \frac{1}{60} = \frac{1 \cdot (1 \cdot 2)}{1 + 1 \cdot 2} \frac{1}{60} = \frac{1}{90}$$

Variabelskift



Givet en stokastisk variabel X med tæthed $f(x)$,

og en funktion $g : \mathbf{R} \mapsto \mathbf{R}$.

Det definerer en ny stokastisk variabel

$$Y = g(X)$$

Vi har for eksempel

$$E(Y) = E(g(X)) = \int g(x)f(x)dx$$

Hvad er fordelingen af Y ?



Bestem for eksempel overlevelsesfunktion

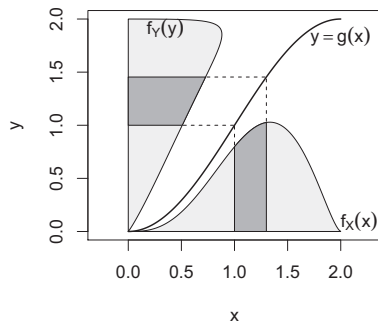
$$P(Y > y) = E(I_{Y>y}) = \int_{\{x:g(x)>y\}} f(x) dx$$

... men hvordan bestemmer vi tætheden af Y ud fra tætheden af X ?

Variabelskift



Givet X med tæthedsfunktion $f_X(x)$, og relationen $Y = g(X)$.



Hvis g er voksende:

$$P(y_1 < Y \leq y_2) =$$

$$P(g^{-1}(y_1) < X \leq g^{-1}(y_2))$$

Dermed er fordelingen af Y fastlagt!

Bestem tæthedsfunktionen for Y



$Y = g(X)$ med $g(x)$ voksende

Indfør $F_X(x) = \int_{-\infty}^x f_X(u)du$ og $F_Y(y) = \int_{-\infty}^y f_Y(v)dv$.

Tæthedsfunktionen $f_Y(y)$ kan findes ved at differentiere:

$$f_Y(y)h = P(y < Y \leq y+h) = F_Y(y+h) - F_Y(y)$$

$$= P(g^{-1}(y) \leq g^{-1}(Y) \leq g^{-1}(y+h)) = F_X(g^{-1}(y+h)) - F_X(g^{-1}(y))$$

og dermed

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|} \quad \text{hvor } x = g^{-1}(y)$$

Opgave 4.4.1, 4.4.2, 4.4.3, 4.4.6 og 4.4.10 er af denne type

Variabelskift teknisk: $Y = g(X)$



For en strengt monoton funktion $y = g(x)$ kan vi bestemme $f_Y(y)$ ud fra $f_X(x)$ ved

$$f_Y(y) = \frac{f_X(x)}{\left| \frac{dy}{dx}(x) \right|}$$

hvor vi erstatter x med y ved brug af den omvendte funktion $x = g^{-1}(y)$

Modificeret opgave 4.4.3




Antag, at X er ligefordelt på $] -1; 0[$.

Spm: Find tætheden for X^2 .

Svar: Vi indfører $Y = X^2$, således at $X = -\sqrt{Y}$, og $\frac{dy}{dx}(x) = 2x$.

$$f_Y(y) = \frac{f_X(x)}{\left| \frac{dy}{dx} \right|} = \frac{1}{|2x|} = \frac{1}{2\sqrt{y}}, 0 < y < 1$$

Antag at X er en eksponential (λ) fordelt stokastisk variabel. 

Spørgsmål 4

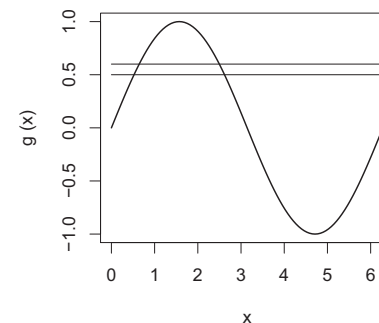
Tætheden $f(y)$ af $Y = cX$ for $c > 0$ er

- 1 $f(y) = \lambda e^{-\lambda(cy)}$
- 2 $f(y) = c\lambda e^{-c\lambda y}$
- 3 $f(y) = \frac{\lambda}{c} e^{-\frac{\lambda}{c}y}$
- 4 $f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-c\lambda)^2}$
- 5 $f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\frac{c}{\lambda})^2}$
- 6 Ved ikke

Mange til en funktion

Lad X være ligefordelt på $[0, 2\pi]$; lad $Y = \sin(X)$.

Spm: Find tætheden for Y .



Tætheden af X er $\frac{1}{2\pi}$.

For hvert skæringspunkt er bidraget

$$\frac{1}{2\pi} \frac{1}{|\cos x|}$$

med $x = \sin^{-1} y$. I alt fås

$$f_Y(y) = \frac{1}{\pi\sqrt{1-y^2}}$$

Mange-til-eeen funktioner, generelt



Hvis funktionen $g(x)$ er strengt monoton i delintervaller gælder:

$$f_Y(y) = \sum_{\{x:g(x)=y\}} \frac{f_X(x)}{\left| \frac{dy}{dx} \right|}$$

Opgave 4.4.5



- Antag at X er ligefordelt på $[-1; 2]$. Find tætheden af X^2 .
- For Y i intervallet $[0; 1]$ har vi to løsninger til $y = x^2$, medens vi har en løsning for Y i intervallet $]1; 2]$. Vi får således:

$$f_Y(y) = \begin{cases} \frac{1}{3\sqrt{y}}, & 0 < y \leq 1 \\ \frac{1}{6\sqrt{y}}, & 1 < y < 2 \end{cases}$$

Opgave 4.4.4



Antag at X er ligefordelt på $] - 1; 1[$. *Spm:* Find tætheden af $Y = X^2$.

Svar: Ligningen $x^2 = y$ har de to løsninger $x_1 = \sqrt{y}$ og $x_2 = -\sqrt{y}$.

Både for x_1 og for x_2 finder vi:

$$f_X(x_1) = f_X(x_2) = \frac{1}{2} \text{ og } \left| \frac{dy}{dx}(x_i) \right| = 2|x_i|$$

Dermed:

$$f_Y(y) = 2 \frac{f_X(x_i)}{\left| \frac{dy}{dx}(x_i) \right|} = 2 \frac{\frac{1}{2}}{|2x|} = \frac{1}{2\sqrt{y}}, 0 < y < 1$$

Nye begreber i dag



1. - hazard raten, og dens sammenhæng med
 - Overlevelsesfunktionen
 - Sandsynlighedstætheden for levetiden
2. Variabelskift
 - ... for monotone funktioner
 - ... for stykkevist monotone funktioner

Afsnit 4.2, 4.3 og 4.4



- Poissonprocessen/eksponentialfordelingen $f(x) = \lambda e^{-\lambda x}$
 - ◇ Rate λ Overlevelsesfunktion: $G(x) = e^{-\lambda x}$
 - ◇ Gammafordelingen $f(x) = \lambda \frac{(\lambda x)^{r-1}}{\Gamma(r)} e^{-\lambda x}$ hvor $\Gamma(r) = (r-1)!$ for r heltallig. For r heltallig $P(X \leq x) = 1 - \sum_{i=0}^{r-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x}$
- Hazard-funktionen (hazard rate)

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T \in (t, t + \Delta t) | T > t)}{\Delta t} = \frac{f(t)}{G(t)}$$

- Variabelskift, $P(X \in [x, x + dx]) = f_X(x)dx$, $Y = g(X)$

$$P(Y \in [y, y + dy]) = \frac{f_X(g^{-1}(y))}{\left| \frac{dy}{dx} \right|}$$