

MVE090 LP4 2021 Föreläsning 5-6

Kapitel 4. "Continuous Distributions"

Definition 4.1. En stokastisk variabel X är kontinuerlig om den har mer än uppräknligt oändligt många (olika) möjliga värden.

Definition 4.2. Frekvensfunktionen $f(x) = f_X(x)$ för en kontinuerlig stokastisk variabel X definieras $f_X(x) = F'_X(x) = \frac{d}{dx}F_X(x)$ [där $F_X(x)$ är fördelningsfunktionen enligt definition 3.4].

Sats 4.3. För en kontinuerlig stokastisk variabel X med frekvensfunktion $f_X(x)$ är

1. $f_X(x) \geq 0$ för $x \in \mathbb{R}$,
2. $F_X(x) = \int_{-\infty}^x f_X(y) dy$ för $x \in \mathbb{R}$,
3. $\int_{-\infty}^{\infty} f_X(y) dy = 1$,
4. $\mathbf{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(y) dy$ för $-\infty \leq a < b \leq \infty$,
5. $\mathbf{P}(X \in C) = \int_{x \in C} f_X(y) dy$ för $C \subseteq \mathbb{R}$,
6. $\mathbf{P}(X = x) = 0$ för $x \in \mathbb{R}$.

Bevis. 1. Följer av att $f_X(x)$ är derivatan av en växande funktion: $F_X(y) = \mathbf{P}(X \leq y) \leq \mathbf{P}(X \leq z) = F_X(z)$ för $y \leq z$.

2. Följer av ta $a = -\infty$ och $b = x$ i 4 eftersom $F_X(-\infty) = \mathbf{P}(X \leq -\infty) = 0$.

3. Följer av ta $a = -\infty$ och $b = \infty$ i 4 eftersom $\mathbf{P}(-\infty < X \leq \infty) = 1$.

4. Eftersom $\{X \leq b\} \cap \{X \leq a\} = \{X \leq a\}$ för $a \leq b$ ger 4 i sats 2.2 att

$$\mathbf{P}(a < X \leq b) = \mathbf{P}(\{X \leq b\} \setminus \{X \leq a\}) = \mathbf{P}(X \leq b) - \mathbf{P}(\{X \leq b\} \cap \{X \leq a\}) = F_X(b) - F_X(a).$$

Den andra likheten följer direkt av analysens huvudsats och definitionen av $f_X(x)$. \square

Definition 4.4. Väntevärdet av en kontinuerlig stokastisk variabel X definieras

$$\mu = \mu_X = \mathbf{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Sats 4.5. För en kontinuerlig stokastisk variabel X och funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ är

$$\mathbf{E}(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x) dx.$$

Sats 4.6. För två kontinuerliga stokastiska variabler X och Y är

1. $\mathbf{E}(cX) = c\mathbf{E}(X)$ för en konstant $c \in \mathbb{R}$,
2. $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$.

Definition 3.9 av varians gäller även för kontinuerliga stokastiska variabler.

Sats 4.7. För två kontinuerliga stokastiska variabler X och Y är

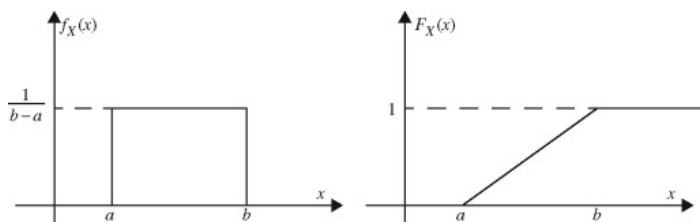
1. $\mathbf{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$,
2. $\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \mathbf{E}(X^2) - \mu_X^2$,
3. $\mathbf{Var}(cX) = c^2 \mathbf{Var}(X)$ för en konstant $c \in \mathbb{R}$,
4. $\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$ om X och Y är oberoende.

Definition 4.8. En stokastisk variabel X med möjliga värden $[a, b]$ och frekvensfunktion $f_X(x) = 1/(b-a)$ för $x \in [a, b]$ är likformigt fördelad Uni $[a, b]$ över $[a, b]$.

När som ovan värdet för $f_X(x)$ endast anges i en viss region av x -värden underförstås att $f_X(x) = 0$ för övrigt. "Uni" kommer från engelskans "uniform".

Exempel 4.1. För en Uni $[a, b]$ stokastisk variabel X är enligt 2 i sats 4.3

$$F_X(x) = \int_{-\infty}^x f_Y(y) dy = \begin{cases} 0 & \text{för } x < a \\ \int_a^x \frac{dy}{b-a} = \frac{x-a}{b-a} & \text{för } x \in [a, b] \\ 1 & \text{för } x > b \end{cases} .$$



Vidare måste $\mathbf{E}(X) = \frac{a+b}{2}$ vilket också kan räknas ut:

$$\int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \left[\frac{x^2}{2(b-a)} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} .$$

Nästan samma räkning ger att $\mathbf{E}(X^2) = \frac{b^3 - a^3}{3(b-a)}$ så att slutligen enligt 2 i sats 4.7

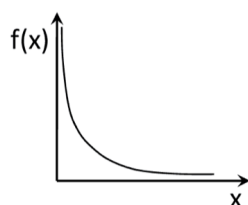
$$\mathbf{Var}(X) = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12} .$$

Definition 4.9. En stokastisk variabel X med möjliga värden $[0, \infty)$ och frekvensfunktion $f_X(x) = \frac{1}{\beta} e^{-x/\beta}$ för $x \geq 0$ är exponentialfördelad $\text{Exp}(\beta)$.

Exempel 4.2. För en $\text{Exp}(\beta)$ stokastisk variabel X är enligt 2 i sats 4.3

$$F_X(x) = \int_{-\infty}^x f_Y(y) dy = \begin{cases} 0 & \text{för } x < 0 \\ \int_0^x \frac{1}{\beta} e^{-y/\beta} dy = 1 - e^{-x/\beta} & \text{för } x \geq 0 \end{cases} .$$

Grafen av de flesta kontinuerliga fördelningsfunktioner ser ungefär likadana ut för ögat så det är ofta liten mening visa dem utan man plottar frekvensfunktionen:



Den momentgenererande funktionen räknas enkelt ut mha. 1 i sats 4.7:

$$m_X(t) = \mathbf{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \int_0^{\infty} e^{tx} \frac{1}{\beta} e^{-x/\beta} dx = \left[\frac{e^{x(t-1/\beta)}}{\beta t - 1} \right]_0^{\infty} = 0 - \frac{1}{\beta t - 1}$$

för $t \in [0, 1/\beta)$. Här konvergerar integralen för $m_X(t)$ bara för vissa t -värden men det viktiga är att 0 är med bland dem och så är det ju ovan.

Sats 3.12 om beräkna moment genom derivera $m_X(t)$ är sann även för kontinuerliga stokastiska variabler (med oförändrat bevis) vilket ger $\mathbf{E}(X) = m'_X(0)$

```
In[1] := Simplify[D[(1/(1-beta*t)), t] /. {t->0}]
```

```
Out[1]= beta
```

och $\mathbf{E}(X^2) = m''_X(0)$

```
In[2] := Simplify[D[D[1/(1-beta*t), t], t] /. {t->0}]
```

```
Out[2]= 2 beta^2
```

så att enligt 2 i sats 4.7

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \beta^2.$$

Definition 4.10. En stokastisk variabel X med möjliga värden $(-\infty, \infty)$ och frekvensfunktion $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$ för $x \in \mathbb{R}$ är normalfördelad $N(\mu, \sigma^2)$.

Normalfördelning finns naturligt pga. följande i statistikteori fundamentala resultat som säger att summan av många små oberoende slumpmässiga bidrag är normalt:

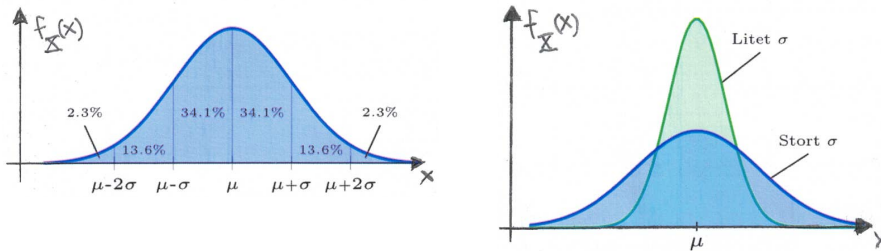
Sats 4.11. Om X_1, X_2, \dots är oberoende stokastiska variabler med gemensamt väntevärde μ och varians σ^2 är $\sum_{i=1}^n (X_i - \mu) / \sqrt{n\sigma^2} \approx N(0, 1)$ då $n \rightarrow \infty$.

Exempel 4.3. För en $N(\mu, \sigma^2)$ stokastisk variabel X är enligt 2 i sats 4.3

$$F_X(x) = \int_{-\infty}^x f_X(y) dy = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(y-\mu)^2/\sigma^2} dy \quad \text{för } x \in \mathbb{R}.$$

Numerisk beräkning av $F_X(x)$ mha. tabell V i M&A diskuteras efter exemplet.

Frekvensfunktionen för en normalfördelad stokastisk variabel är Gauss klockan:



Mha. kvadratkomplettering i exponenten i andra integralen ses att

$$\begin{aligned} m_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx \\ &= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\sigma^2 t - \mu)^2/\sigma^2} dx = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \end{aligned}$$

där sista integralen är 1 ty arean under en $N(\sigma^2 t + \mu, \sigma^2)$ frekvensfunktion.

Nu beräknas $\mathbf{E}(X) = m'_X(0)$

```
In[3] := Simplify[D[Exp[mu*t+sigma^2*t^2/2],t]/.{t->0}]
```

```
Out[3]= mu
```

och $\mathbf{E}(X^2) = m''_X(0)$

```
In[4] := Simplify[D[D[Exp[mu*t+sigma^2*t^2/2],t],t]/.{t->0}]
```

```
Out[4]= mu^2 + sigma^2
```

så att

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \sigma^2.$$

För lära räkna normalsannolikheter numeriskt mha. tabell V i M&A behövs följande:

Definition 4.12. En $N(0, 1)$ stokastisk variabel kallas standard normalfördelad.

Sats 4.13. Om X är normal $N(\mu, \sigma^2)$ är $(X - \mu)/\sigma$ standard normal $N(0, 1)$.

Bevis. Enligt exempel 4.3 är

$$m_{(X-\mu)/\sigma}(t) = \mathbf{E}(e^{t(X-\mu)/\sigma}) = \mathbf{E}(e^{(t/\sigma)X}) e^{-\mu t/\sigma} = m_X(t/\sigma) e^{-\mu t/\sigma} = \dots = e^{t^2/2}$$

som (enligt exempel 4.3) är momentgenererande funktionen för $N(0, 1)$. \square

Exempel 4.4. Om X är $N(1, 0.25)$ kan mha. sats 4.12 och tabell V beräknas

$$\begin{aligned} \mathbf{P}(0.9 < X \leq 1.54) &= \mathbf{P}\left(\frac{0.9-1}{0.5} < \frac{X-1}{0.5} \leq \frac{1.54-1}{0.5}\right) \\ &= \mathbf{P}\left(N(0, 1) \leq \frac{1.54-1}{0.5}\right) - \mathbf{P}\left(N(0, 1) \leq \frac{0.9-1}{0.5}\right) = 0.8599 - 0.4207. \end{aligned}$$

Definition 4.14. En stokastisk variabel X med möjliga värden $(0, \infty)$ och frekvensfunktion $f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$ för $x > 0$ är gamma fördelad.

Exempel 4.5. En gamma stokastisk variabel har momentgenererande funktion

$$\begin{aligned} m_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \int_0^{\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= (1-\beta t)^{-\alpha} \int_0^{\infty} \frac{(1/\beta-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/(1/\beta-t)^{-1}} dx = (1-\beta t)^{-\alpha} \end{aligned}$$

där sista integralen är 1 eftersom arean under en gamma frekvensfunktion. Det går nu enkelt räkna ut $\mathbf{E}(X) = m'_X(0) = \alpha\beta$ och $\mathbf{E}(X^2) = m''_X(0) = \alpha(\alpha+1)\beta^2$.

Definition 4.15. En stokastisk variabel X med möjliga värden $(0, \infty)$ och frekvensfunktion $f_X(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}$ för $x > 0$ är Weibull fördelad.

Exempel 4.6. För en Weibull stokastisk variabel X är $\mathbf{E}(X)$

```
In[5] := FullSimplify[Integrate[x*alpha*beta*x^(beta-1)*
  Exp[-alpha*x^beta], {x, 0, Infinity}, Assumptions->alpha>0&&beta>0]]
Out[5] = alpha^-1/beta Gamma[1 + 1/beta]
```

och $\mathbf{E}(X^2)$

```
In[6] := FullSimplify[Integrate[x^2*alpha*beta*x^(beta-1)*
  Exp[-alpha*x^beta], {x, 0, Infinity}, Assumptions->alpha>0&&beta>0]]
Out[6] = alpha^-2/beta Gamma[1 + 2+beta/beta].
```

I matte är ofta olikheter viktigare än likheter och en viktig sannolikhetsolikhet är

Sats 4.16. (TJEBYSJEVS OLIKHET) $\mathbf{P}(|X - \mu_X| > k \sigma_X) \leq \frac{1}{k^2}$ för $k > 0$.

Transformation av kontinuerlig stokastisk variabel. Om X är en kontinuerlig stokastisk

variabel med frekvensfunktion $f_X(x)$ och $g : \mathbb{R} \rightarrow \mathbb{R}$ en deriverbar och strikt växande funktion, vilken frekvensfunktion får då $Y = g(X)$? Svar:

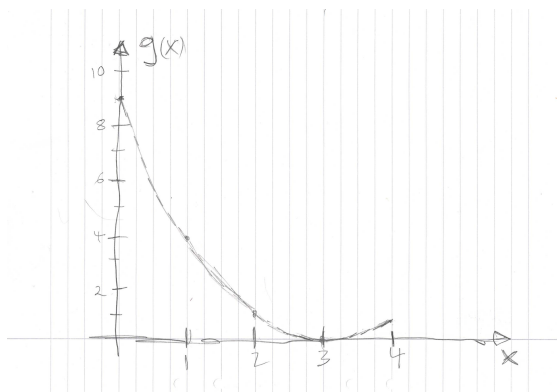
Sats 4.17. $f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$.

Bevis. Eftersom $F'_X(X) = f_X(x)$ är mha. vanlig inre derivata

$$f_Y(y) = \frac{d}{dy} \mathbf{P}(g(X) \leq y) = \frac{d}{dy} \mathbf{P}(X \leq g^{-1}(y)) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y). \quad \square$$

Exempel 4.7. Om X är $\text{Uni}[0, 4]$ och $g(x) = (x-3)^2$ vad blir frekvensfunktionen $f_Y(y)$ för $Y = g(X)$?

Lösning. Notera att grafen av $g(x)$ är en parabel



så g är ej monoton och sats 4.17 kan ej användas. Istället beräknar vi

$$\begin{aligned} \mathbf{P}(Y \leq y) &= \mathbf{P}((X-3)^2 \leq y) = \mathbf{P}(-\sqrt{y} \leq X-3 \leq \sqrt{y}) \\ &= \mathbf{P}(3-\sqrt{y} \leq X \leq 3+\sqrt{y}) \\ &= (3+\sqrt{y})/4 - (3-\sqrt{y})/4 = \sqrt{y}/2 \end{aligned}$$

för $y \in [0, 1]$ så $f_Y(y) = 1/(4\sqrt{y})$ för $y \in [0, 1]$, medan

$$\mathbf{P}(Y \leq y) = \frac{1}{4} + \mathbf{P}(-\sqrt{y} \leq X-3 \leq 0) = \frac{1}{4} + \mathbf{P}(3-\sqrt{y} \leq X \leq 3) = \frac{1}{4} + \sqrt{y}/4$$

för $y \in (1, 9]$ så $f_Y(y) = 1/(8\sqrt{y})$ för $y \in (1, 9]$.

Simulering av kontinuerlig stokastisk variabel. Hur gör man en observation av en kontinuerlig stokastisk variabel X med fördelningsfunktion $F_X(x)$ i dator? Svar:

Sats 4.18. Tag en stokastisk variabel ξ som är $\text{Uni}[0, 1]$ och sätt $X = F_X^{-1}(\xi)$.

Bevis. Enligt exempel 4.1 är

$$\mathbf{P}(X \leq x) = \mathbf{P}(F_X^{-1}(\xi) \leq x) = \mathbf{P}(\xi \leq F_X(x)) = F_\xi(F_X(x)) = F_X(x). \quad \square$$

Exempel 4.8. Hur simulerar man en stokastisk variabel X som är $\text{Exp}(\beta)$?

Lösning. Eftersom enligt exempel 4.2 $F_X(x) = 1 - e^{-x/\beta}$ som är lika med y för $x = -\beta \ln(1 - y) = F_X^{-1}(y)$ blir receptet $X = F_X^{-1}(\xi) = -\beta \ln(1 - \xi)$.

Tillförlitlighetsteori. En systemkomponent håller en kontinuerlig stokastisk tid $T > 0$.

Definition 4.19. Överlevnadsfunktionen ("reliability function" i M&A) är

$$R_T(t) = \mathbf{P}(T > t) = 1 - F_T(t) \quad \text{för } t \geq 0.$$

Definition 4.20. Felintensiteten (engelska "hazard rate") är

$$\rho_T(t) = f_T(t)/R_T(t) = f_T(t)/(1 - F_T(t)) = -\frac{d}{dt} \ln[1 - F_T(t)] \quad \text{för } t \geq 0.$$

Sats 4.21. $\rho_T(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbf{P}(T \in [t, t + \Delta t] | T > t)$.

Bevis. Eftersom $f_T(t)$ är derivatan av $F_T(t)$ ger 4 i sats 4.3 tillsammans med definitionen av betingad sannolikhet att

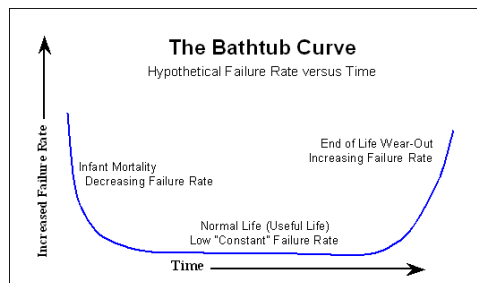
$$\begin{aligned} \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbf{P}(T \in [t, t + \Delta t] | T > t) &= \lim_{\Delta t \downarrow 0} \frac{\mathbf{P}(T \in [t, t + \Delta t], T > t)}{\Delta t \mathbf{P}(T > t)} \\ &= \lim_{\Delta t \downarrow 0} \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t (1 - F_T(t))} = \frac{f_T(t)}{1 - F_T(t)} = \rho_T(t). \quad \square \end{aligned}$$

Sats 4.22. $R_T(t) = \exp(-\int_0^t \rho_T(s) ds)$.

Bevis. Eftersom $T > 0$ är $F_T(0) = 0$ så att

$$\exp\left(-\int_0^t \rho_T(s) ds\right) = \exp\left(-\int_0^t \frac{f_T(s)}{1 - F_T(s)} ds\right) = \exp([\ln(1 - F_T(s))]_0^t) = 1 - F_T(t). \quad \square$$

Exempel 4.9. Ett klassiskt exempel på en felintensitet använd i teknik mm. är



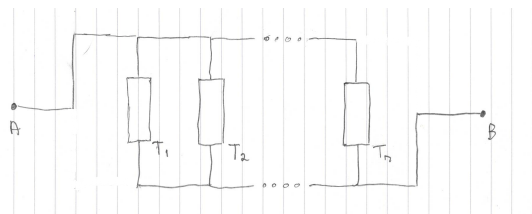
Weibull fördelning användes mycket i tillförlitlighet och för den är

$$F_T(t) = \int_{-\infty}^t f_T(x) dx = \int_0^t \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} dx = 1 - e^{-\alpha t^\beta} \Rightarrow \rho_T(t) = \dots = \alpha \beta t^{\beta-1}.$$

Vi studerar tillförlitlighetssystem byggda av två grundkopplingar: seriekoppling



som fungerar då alla ingående komponenterna fungerar och parallellkoppling



som fungerar då minst en av komponenterna fungerar. I dessa kopplingar förutsättes ingående komponenternas livslängder T_1, \dots, T_n vara oberoende (stokastiska variabler).

Sats 4.23. Seriekoppling har överlevnadsfunktion $R_{T_1}(t) \dots R_{T_n}(t)$.

Bevis. $\mathbf{P}(T > t) = \mathbf{P}(\min(T_1, \dots, T_n) > t) = \mathbf{P}(T_1 > t, \dots, T_n > t) = R_{T_1}(t) \dots R_{T_n}(t)$. \square

Sats 4.24. Parallellkoppling har överlevnadsfunktion $1 - (1 - R_{T_1}(t)) \dots (1 - R_{T_n}(t))$.

Bevis. Om parallellkopplingens livslängd betecknas T är pga. oberoende för T_1, \dots, T_n

$$\begin{aligned} \mathbf{P}(T > t) &= \mathbf{P}(\max(T_1, \dots, T_n) > t) \\ &= 1 - \mathbf{P}(\max(T_1, \dots, T_n) \leq t) \\ &= 1 - \mathbf{P}(T_1 \leq t) \dots \mathbf{P}(T_n \leq t) = 1 - F_{T_1}(t) \dots F_{T_n}(t). \end{aligned} \quad \square$$

Exempel 4.10. Tillförlitlighetssystemet



med livslängd T kan analyseras som seriekopplingen



som enligt sats 4.23 har tillförlitlighetsfunktion

$$R_T(t) = R_{T_A}(t) R_{T_B}(t)$$

där T_A och T_B är livslängder för parallellkopplingar för vilka enligt sats 4.24

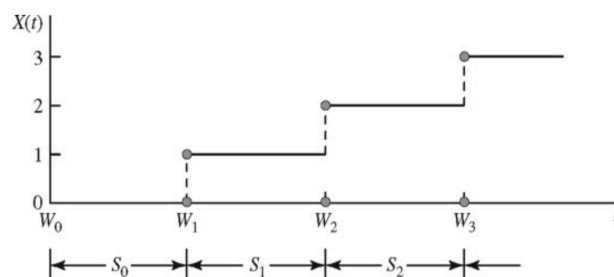
$$R_{T_A}(t) = 1 - F_{T_1}(t)F_{T_2}(t) \quad \text{och} \quad R_{T_B}(t) = 1 - F_{T_3}(t)F_{T_4}(t)F_{T_5}(t).$$

Varje bygge av serie- och parallellkopplingar kan analyseras analogt exempel 4.10.

Poisson process. En Poisson process med parameter eller intensitet (engelska “intensity” eller “rate”) $\lambda > 0$ är ett stokastiskt tidsförlopp $X(t)$, $t \geq 0$, dvs. för varje reell tid $t \geq 0$ finns ett stokastiskt Poisson process värde $X(t)$. Vidare gäller att:

1. $X(0) = 0$,
2. $X(t)$ ökar värde med en enhet i taget, dvs. det finns stokastiska tider $0 = W_0 < W_1 < W_2 < \dots$ sådana att $X(t) = i$ för $t \in [W_i, W_{i+1})$ för $i = 0, 1, 2, \dots$,
3. tiden mellan (enhets-) “hopp” $S_0 = W_1, S_1 = W_2 - W_1, S_2 = W_3 - W_2, \dots$ [dvs. de tider $X(t)$ befinner sig på sina olika möjliga värden $i = 0, 1, 2, \dots$ innan den byter till nästa värde $i + 1$] är oberoende exponentialfördelade med parameter $\beta = 1/\lambda$.

Egenskaperna ovan är definierande eftersom de är tillräcklig information för kunna tillverka observationer av processen i dator. Och de får följande principiella utseende:



När man ritar sådan bild menas att man först utför alla slumpexperiment och registrerar deras utfall så alla inblandade stokastiska variablers värde kan bestämmas och sedan använder man dem till att fabricera och plotta processen enligt receptet ovan.

Sats 4.25. $X(t) - X(s)$ är Poisson fördelad med parameter $\lambda(t - s)$ för $0 \leq s \leq t$ så speciellt $X(t)$ är Poisson fördelad med parameter λt .

Exempel 4.11. Om man med början tiden $t = 0$ registrerar det (stokastiska) totala antalet bilar $X(t)$ som har passerat på en väg till tiden t för $t \geq 0$ är det allmänt accepterat att $X(t)$ väl överensstämmer med en Poisson process.

Samma sak om man vid tiden $t = 0$ börjar registrera det totala antalet radioaktiva sönderfall $X(t)$ till tiden t för $t \geq 0$ för ett radioaktivt preparat.