

1. Antalet snittar, X , som en godtycklig gäst äter beskrivas av följande sannolikhetsfunktion

k	0	1	2	3
$p_X(k)$	0.2	0.1	0.3	0.4

Där vi använt att $1 - (0.1 + 0.3 + 0.4) = 1 - 0.8 = 0.2$ av gästerna inte äter några snittar.

- (a) Väntevärdet ges av

$$\mathbf{E}(X) = \sum_{k=0}^3 k \cdot p_X(k) = 0 \cdot 0.2 + 1 \cdot 0.1 + 2 \cdot 0.3 + 3 \cdot 0.4 = 1.9.$$

Och för variansen har vi $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$ där det första väntevärdet ges av

$$\mathbf{E}(X^2) = \sum_{k=0}^3 k^2 \cdot p_X(k) = 0^2 \cdot 0.2 + 1^2 \cdot 0.1 + 2^2 \cdot 0.3 + 3^2 \cdot 0.4 = 4.9,$$

vilket ger

$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = 4.9 - 1.9^2 = 1.29$$

- (b) Eftersom antalet snittar som de 100 gästerna äter är oberoende av varandra så beskrivs det totala antalet ätna snittar av $Y = \sum_{i=1}^n X_i$ där $n = 100$ och

$$\mathbf{E}(Y) = \mathbf{E}\left(\sum_{i=1}^n X_i\right) = n \cdot \mathbf{E}(X) = 100 \cdot 1.9 = 190,$$

$$\mathbf{V}(Y) = \mathbf{V}\left(\sum_{i=1}^n X_i\right) = n \cdot \mathbf{V}(X) = 100 \cdot 1.29 = 129.$$

Eftersom Y är **en summa av många likafördelade oberoende** stokastiska variabler så ger **centrala gränsvärdessatsen** att

$$Y \underset{\sim}{\in} \mathbf{N}\left(190, \sqrt{129}\right).$$

Vi söker nu ett antal snittar, y , så att $\mathbf{P}(Y \leq y) = 0.99$. Värdet på y ges av

$$\mathbf{P}\left(\underbrace{\frac{Y - 190}{\sqrt{129}}}_{\underset{\sim}{\in} \mathbf{N}(0,1)} \leq \frac{y - 190}{\sqrt{129}}\right) = 0.99$$

$$\frac{y - 190}{\sqrt{129}} = \lambda_{0.01} \approx 2.3263$$

$$y = 190 + \lambda_{0.01} \cdot \sqrt{129} \approx 216.422$$

Vi behöver alltså 217 snittar (lämpligt att avrunda uppåt eftersom vi vill ha ett antal snittar som räcker i 99% av fallen).

Med halvkorrektur har vi istället

$$P(Y \leq y + 0.5) = P\left(\underbrace{\frac{Y - 190}{\sqrt{129}}}_{\substack{\in \mathbf{N}(0,1) \\ \approx}} \leq \frac{y + 0.5 - 190}{\sqrt{129}}\right) = 0.99$$

$$y = 189.5 + \lambda_{0.01} \cdot \sqrt{129} \approx 215.922$$

2. Eftersom väntevärdet för en exponentialfördelning är $1/\lambda = 10$ har vi att

$$X \in \mathbf{Exp}(1/10), \quad f_X(x) = \frac{1}{10}e^{-x/10}, \quad 0 \leq x \quad \text{och} \quad Y \in \mathbf{R}(0, 2), \quad f_Y(y) = \frac{1}{2}, \quad 0 \leq y \leq 2$$

(a) Eftersom X och Y är oberoende så ges den bivariata täthetsfunktionen, $f_{XY}(x, y)$, av

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y) = \frac{1}{20}e^{-x/10}, \quad 0 \leq x, \quad 0 \leq y \leq 2$$

(b) Tätheten för en summa av oberoende stokastiska variabler ges av en faltningintegral

$$f_T(t) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(t - x) dx.$$

Lämpliga gränser för integralen fås genom att studera definitionsområdet

$$\begin{cases} 0 \leq x \\ 0 \leq t - x \leq 2 \end{cases} \Leftrightarrow \begin{cases} 0 \leq x \\ t - 2 \leq x \leq t \end{cases} \Leftrightarrow \begin{cases} 0 \leq x \leq t, & t \leq 2 \\ t - 2 \leq x \leq t, & t > 2 \end{cases}$$

Vilket ger följande

$$\begin{aligned} f_T(t) &= \int_a^b f_X(x) \cdot f_Y(t - x) dx = \left[-\frac{1}{2}e^{-x/10}\right]_a^b = \frac{1}{2} \left(e^{-a/10} - e^{-b/10}\right) = \\ &= \begin{cases} \frac{1}{2} (e^{-0/10} - e^{-t/10}), & 0 < t \leq 2 \\ \frac{1}{2} (e^{-(t-2)/10} - e^{-t/10}), & 2 < t \end{cases} = \begin{cases} \frac{1}{2} (1 - e^{-t/10}), & 0 < t \leq 2 \\ \frac{(e^{0.2} - 1)}{2} e^{-t/10}, & 2 < t \end{cases} \end{aligned}$$

Vilket ger tätheten för summan som

$$f_T(t) = \begin{cases} \frac{1}{2} (1 - e^{-t/10}), & 0 < t \leq 2 \\ \frac{(e^{0.2} - 1)}{2} e^{-t/10}, & 2 < t \end{cases}$$

(c) Sannolikheten att $T \leq 10$ kan nu beräknas antingen som:

Alt 1. En integral över tätheten för T

$$P(T \leq 10) = \int_{-\infty}^{10} f_T(t) dt = \int_0^2 \frac{1}{2} (1 - e^{-t/10}) dt + \int_2^{10} \frac{(e^{0.2} - 1)}{2} e^{-t/10} dt$$

En enklare integral fås genom att använda komplement sannolikheten

$$\begin{aligned} P(T \leq 10) &= 1 - P(T > 10) = 1 - \int_{10}^{\infty} \frac{(e^{0.2} - 1)}{2} e^{-t/10} dt = \\ &= 1 - \left[-5 (e^{0.2} - 1) e^{-t/10}\right]_{10}^{\infty} = 1 - (0 + 5 (e^{0.2} - 1) e^{-1}) = \\ &= 1 - 5 (e^{0.2} - 1) e^{-1} \approx 0.59275 \end{aligned}$$

Alt 2. En dubbelintegral över tätheten för X, Y .

$$\begin{aligned} \mathbf{P}(T \leq 10) &= \mathbf{P}(X + Y \leq 10) = \int_{x+y \leq 10} f_{XY}(x, y) \, dx \, dy = \\ &= \int_0^2 \int_0^{10-y} \frac{1}{20} e^{-x/10} \, dx \, dy = \int_0^2 \left[-\frac{1}{2} e^{-x/10} \right]_0^{10-y} \, dy = \\ &= \int_0^2 \frac{1}{2} \left(1 - e^{-1+y/10} \right) \, dy = \left[\frac{1}{2} (y - 10e^{-1+y/10}) \right]_0^2 = \\ &= \frac{1}{2} ((2 - 10e^{-0.8}) + 10e^{-1}) = 1 - 5e^{-0.8} + 5e^{-1} \approx 0.59275 \end{aligned}$$

3. En lämplig modell är att antalet korrekta gissningar, X , följer en binomialfördelning med $X \in \mathbf{Bin}(n, p)$. Vi vill testa en noll-hypotes om rena gissningar mot att Smaug är bättre än slumpen, dvs

$$H_0 : p = 0.5 \quad \text{mot} \quad H_1 : p > 0.5$$

med $n = 16$ försök har vi att $np_0(1 - p_0) = 4 < 10$ och vi kan **inte** använda CGS.

Ett lämpligt test är därmed att använda direktmetoden och vi beräknar p-värdet som

$$\begin{aligned} \mathbf{P}(\text{få det vi fick eller längre från } H_0 | H_0) &= \mathbf{P}(X \geq 13) = 1 - \mathbf{P}(X \leq 12) = \\ &= 1 - \sum_{k=0}^{12} \binom{16}{k} 0.5^k 0.5^{16-k} = [\text{tabell 6}] = 1 - 0.98936 = 0.01064 \end{aligned}$$

Eftersom p-värdet är mindre än 5% så **förkastar** vi H_0 och det verkar ligga något i legenden.

4. För en Poissonprocess med intensitet λ händelser per minut gäller att tiden (i minuter) mellan händelser är exponentialfördelad, $T \in \mathbf{Exp}(\lambda)$, och att antalet händelser efter t minuter följer en Poissonfördelning med $X(t) \in \mathbf{Po}(\lambda \cdot t)$.

- (a) Tiden mellan två gäster är exponentialfördelad $T \in \mathbf{Exp}(\lambda)$ med $\lambda = 2$ och vi söker sannolikheten

$$\mathbf{P}(T > 1) = \int_1^\infty f_T(t) \, dt = \int_1^\infty 2e^{-2t} \, dt = [-e^{-2t}]_1^\infty = 0 + e^{-2} \approx 0.1353$$

- (b) Antalet gäster efter 5 minuter är Poissonfördelat med $X(5) \in \mathbf{Po}(\lambda \cdot 5) = \mathbf{Po}(10)$ och den sökta sannolikheten är

$$\mathbf{P}(X(5) \geq 15) = 1 - \mathbf{P}(X(5) \leq 14) = 1 - \sum_{k=0}^{14} \frac{10^k}{k!} e^{-10} = [\text{tabell 5}] = 1 - 0.91654 = 0.08346$$

- (c) En Poissonprocess har oberoende ökningar där antalet händelser i ett intervall följer en Poissonfördelning med väntevärde som beror av intervallens längd, $X(t) - X(s) \in \mathbf{Po}(\lambda \cdot (t - s))$. Vi tar $n = 50$ och skriver om den betingade sannolikheten som

$$\begin{aligned} \mathbf{P}(X(15) = k | X(25) = n) &= \frac{\mathbf{P}(X(15) = k \cap X(25) = n)}{\mathbf{P}(X(25) = n)} = \\ &= \frac{\mathbf{P}(X(15) = k \cap X(25) - X(15) = n - k)}{\mathbf{P}(X(25) = n)} = \frac{\mathbf{P}(X(15) = k) \cdot \mathbf{P}(X(25) - X(15) = n - k)}{\mathbf{P}(X(25) = n)} = \\ &= \frac{\frac{(15\lambda)^k}{k!} e^{-15\lambda} \cdot \frac{(10\lambda)^{n-k}}{(n-k)!} e^{-10\lambda}}{\frac{(25\lambda)^n}{n!} e^{-25\lambda}} = \frac{n!}{k! \cdot (n-k)!} \cdot \frac{15^k \lambda^k 10^{n-k} \lambda^{n-k}}{25^n \lambda^n} \cdot \frac{e^{15\lambda} e^{10\lambda}}{e^{25\lambda}} = \\ &= \binom{n}{k} \frac{15^k 10^{n-k}}{25^n} = \binom{50}{k} \frac{15^k 10^{50-k}}{25^{50}}. \end{aligned}$$

Vilket ger

$$\mathbf{P}(X(15) = k | X(25) = 50) = \binom{50}{k} \frac{15^k 10^{50-k}}{25^{50}}.$$

(d) Omskrivning av svaret i 4c ger

$$\begin{aligned} P(k \text{ gäster i tid}) &= \binom{50}{k} \frac{15^k 10^{50-k}}{25^{50}} = \binom{50}{k} \left(\frac{15}{25}\right)^k \left(\frac{10}{25}\right)^{50-k} = \\ &= \binom{50}{k} \left(\frac{15}{25}\right)^k \left(1 - \frac{15}{25}\right)^{50-k} = \binom{50}{k} 0.6^k (1 - 0.6)^{50-k}. \end{aligned}$$

Vi känner nu igen sannolikhetsfunktionen för en **Bin** (50, 0.6)-fördelning.

5. (a) β -koefficienterna i Modell 2 ger

β_0 : Modellens intercept, dvs "Grundpriset" för en lägenhet i Lund som ligger precis vid järnvägsstationen och är på 0 kvm.

β_1 : Ökningen i pris per kvadratmeter i tusentals kronor, dvs om det skiljer 1 kvm mellan två, i övrigt identiska, lägenheter skulle den större lägenheten i medel kosta 37 700 kr mer.

β_2 : Hur snabbt priset avtar med avståndet från järnvägsstationen. För varje kilometer från stationen (eller möjligen från centrum) faller priset med 34 100 kronor.

(b) För att konstruera 95%-iga konfidensintervall för β_1 och β_2 behöver vi första skatta variansen för ε_i ,

$$s^2 = \frac{Q_0}{n - p - 1} = \frac{4.3632 \cdot 10^7}{516 - 3} \approx 85\,053 \quad s = \sqrt{\frac{4.3632 \cdot 10^7}{516 - 3}} \approx 291.6378$$

Konfidensintervall för β_i ges nu av

$$\begin{aligned} I_{\beta_1} &= \beta_1^* \pm t_{0.025}(513) \cdot s \cdot \sqrt{1.61 \cdot 10^{-6}} = [36.9747 \quad 38.4253] \\ I_{\beta_2} &= \beta_2^* \pm t_{0.025}(513) \cdot s \cdot \sqrt{9.26 \cdot 10^{-4}} = [-51.4939 \quad -16.7061] \end{aligned}$$

där vi använt skattningen av σ , de två sista diagonal elementen i $(X^T X)^{-1}$ och det faktum att $t_{0.025}(513) \approx \lambda_{0.025} = 1.96$.

Tolkning av intervallen blir att ökningen i priset per kvm för lägenheterna i Lund är **signifikant högre** än motsvarande pris ökning för lägenheter i Skåne eftersom 26.756 inte ingår i intervallet I_{β_1} . Vidare så ser vi att effekten av avståndet till järnvägsstationen är **signifikant** eftersom 0 inte ingår i intervallet I_{β_2} .

(c) Eftersom båda koefficienterna i modell 2 är signifikanta finns det ingen anledning att använda den enklare modell 1. Ett lämpligt pris för en lägenhet på 67 kvm som ligger 2.5 km från stationen borde alltså vara

$$\mu_0^* = x_0 \cdot \beta^* = \beta_0^* + \beta_1^* \cdot 67 + \beta_2^* \cdot 2.5 = 161.9 + 37.7 \cdot 67 - 34.1 \cdot 2.5 = 2\,602.6.$$

Eller 2 602 600 kr.

(d) Priset för en enskild lägenhet ges av ett prediktionsintervall

$$I_{y_0} = \mu_0^* \pm t_{\alpha/2}(513) \cdot s \cdot \sqrt{1 + x_0(X^T X)^{-1}x_0^T}.$$

Eftersom vi är intresserade av hur ovanligt bra lägenheten är så studerar vi övre gränsen i ett ensidigt prediktionsintervallet

$$2\,980 = \mu_0^* + t_{\alpha}(513) \cdot s \cdot \sqrt{1 + x_0(X^T X)^{-1}x_0^T}.$$

Om vi löser ut kvantilen har vi att

$$\frac{2\,980 - \mu_0^*}{s \cdot \sqrt{1 + x_0(X^T X)^{-1}x_0^T}} = t_{\alpha}(513) \approx \lambda_{\alpha}.$$

Med relevanta storheter har vi

$$\begin{aligned} x_0 &= [1 \quad 67 \quad 2.5] & x_0(X^\top X)^{-1}x_0^\top &\approx 0.0028 \\ x_0 \cdot \beta^* &= 2\,602.6 & s \cdot \sqrt{1 + x_0(X^\top X)^{-1}x_0^\top} &\approx 292.0440 \end{aligned}$$

och

$$\lambda_\alpha = \frac{2\,800 - \mu_0^*}{s \cdot \sqrt{1 + x_0(X^\top X)^{-1}x_0^\top}} \approx \frac{2980 - 2602.6}{292.0440} = 1.29$$

Vilket svarar mot en 10% kvantil eftersom $t_{0.1}(\infty) = \lambda_{0.1} = 1.28$

6. (a) Eftersom X är den kraft där glaset går sönder så svarar en övre kvantil, $\mathbf{P}(X \leq x) = 0.999$, mot ett värde där 99.9% av glasskivorna har gått sönder. Vi söker alltså ett värde x så att $\mathbf{P}(X \leq x) = 1 - 0.999 = 0.001$, dvs att endast 0.1% av skivorna har gått sönder.

$$\begin{aligned} 0.001 &= \mathbf{P}(X \leq x) = F_X(x) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) \\ \log(0.999) &= -\left(\frac{x}{\lambda}\right)^k & (-\log(0.999))^{1/k} &= \frac{x}{\lambda} \\ x &= \lambda \cdot (-\log(0.999))^{1/k} = 75 \cdot (-\log(0.999))^{1/10} \approx 37.59 \text{ MPa} \end{aligned}$$

- (b) För ML-skattningen baserat på n observationer, $x_1, x_2, \dots, x_n$, är likelihooden

$$L(\lambda) = \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \frac{k x_i^{k-1}}{\lambda^k} \exp\left(-\left(\frac{x_i}{\lambda}\right)^k\right)$$

eller, med $b = \lambda^k$,

$$L(b) = \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \frac{k x_i^{k-1}}{b} \exp\left(-\frac{x_i^k}{b}\right)$$

vilket ger log-likelihooden som

$$\begin{aligned} l(b) &= \log L(b) = \sum_{i=1}^n \log f_X(x_i) = \sum_{i=1}^n \left(\log(k) + (k-1) \log(x_i) - \log(b) - \frac{x_i^k}{b} \right) = \\ &= n \log(k) + (k-1) \left(\sum_{i=1}^n \log(x_i) \right) - n \log(b) - \frac{1}{b} \left(\sum_{i=1}^n x_i^k \right). \end{aligned}$$

Derivering med avseende på b ger nu (eftersom $b = \lambda^k$ är en monoton funktion så kommer maximum med avseende på b och λ att sammanfalla)

$$l'(b) = -\frac{n}{b} + \frac{1}{b^2} \sum_{i=1}^n x_i^k.$$

Vilket ger skattningen som

$$\begin{aligned} 0 &= -\frac{n}{b} + \frac{1}{b^2} \sum_{i=1}^n x_i^k & \iff & \frac{n}{b} = \frac{1}{b^2} \sum_{i=1}^n x_i^k \\ \lambda^k &= b = \frac{1}{n} \sum_{i=1}^n x_i^k & \iff & \lambda^* = \left(\frac{1}{n} \sum_{i=1}^n x_i^k \right)^{1/k} \end{aligned}$$

Alt: Utan variabelbytet har vi

$$\begin{aligned} l(\lambda) &= \log L(\lambda) = \sum_{i=1}^n \log f_X(x_i) = \sum_{i=1}^n \left(\log(k) + (k-1) \log(x_i) - k \log(\lambda) - \frac{x_i^k}{\lambda^k} \right) = \\ &= n \log(k) + (k-1) \left(\sum_{i=1}^n \log(x_i) \right) - nk \log(\lambda) - \frac{1}{\lambda^k} \left(\sum_{i=1}^n x_i^k \right), \end{aligned}$$

med derivata

$$\begin{aligned} l'(\lambda) &= -\frac{nk}{\lambda} + \frac{k}{\lambda^{k+1}} \sum_{i=1}^n x_i^k & 0 &= -\frac{nk}{\lambda} + \frac{k}{\lambda^{k+1}} \sum_{i=1}^n x_i^k \\ \frac{\lambda^{k+1}}{\lambda} &= \lambda^k = \frac{1}{n} \sum_{i=1}^n x_i^k & \lambda^* &= \left(\frac{1}{n} \sum_{i=1}^n x_i^k \right)^{1/k} \end{aligned}$$

(c) För ett minimum har vi att

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \leq y) = \mathbf{P}(\min(X_1, X_2, \dots, X_n) \leq y) = 1 - \mathbf{P}(\min(X_1, X_2, \dots, X_n) > y) = \\ &= 1 - \mathbf{P}(X_1 > y \cap X_2 > y \cap \dots \cap X_n > y) = [\text{oberoende}] = \\ &= 1 - \mathbf{P}(X_1 > y) \cdot \mathbf{P}(X_2 > y) \cdot \dots \cdot \mathbf{P}(X_n > y) = 1 - \prod_{i=1}^n (1 - \mathbf{P}(X_i \leq y)) = \\ &= 1 - \prod_{i=1}^n (1 - F_X(y)) = 1 - \prod_{i=1}^n \exp\left(-\left(\frac{y}{\lambda}\right)^k\right) = 1 - \exp\left(-n\left(\frac{y}{\lambda}\right)^k\right) = \\ &= 1 - \exp\left(-(n^{1/k})^k \left(\frac{y}{\lambda}\right)^k\right) = 1 - \exp\left(-\left(\frac{y}{\lambda n^{-1/k}}\right)^k\right) \end{aligned}$$

Vilket vi känner igen som fördelningsfunktionen för en Weibullfördelning med parametrar k och $\lambda_{\min} = \lambda n^{-1/k}$. Derivering ger nu tätheten som

$$\begin{aligned} f_Y(y) &= F'_Y(y) = -\exp\left(-\left(\frac{y}{\lambda_{\min}}\right)^k\right) \cdot \left(-\left(\frac{y}{\lambda_{\min}}\right)^{k-1}\right) \cdot \frac{k}{\lambda_{\min}} = \\ &= \frac{k}{\lambda_{\min}} \cdot \left(\frac{y}{\lambda_{\min}}\right)^{k-1} \cdot \exp\left(-\left(\frac{y}{\lambda_{\min}}\right)^k\right) = \frac{k y^{k-1}}{\lambda_{\min}^k} \exp\left(-\left(\frac{y}{\lambda_{\min}}\right)^k\right) \end{aligned}$$