

## LØSNING, noen oppgaver fra kap. 5 i Løvås: Statistikk

### Oppgave 2

Dette er en binomisk fordeling,  $X \sim \text{Bin}(8, 0.2)$ , slik at

$$P(X = 2) = \binom{8}{2} 0.2^2 (1 - 0.2)^{8-2} = \frac{8 \cdot 7}{1 \cdot 2} 0.2^2 0.8^6 = 0.294$$

Da er

$$\begin{aligned} P(2 \leq X \leq 5) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\ &= 0.294 + \binom{8}{3} 0.2^3 0.8^5 + \binom{8}{4} 0.2^4 0.8^4 + \binom{8}{5} 0.2^5 0.8^3 \\ &= 0.495 \end{aligned}$$

Et alternativ er å bruke  $P(2 \leq X \leq 5) = P(X \leq 5) - P(X \leq 1)$  og slå opp de kumulative binomiske sannsynligheter  $P(X \leq k)$  i tabellen på side 575 i Løvås. Forventingen regnes ut ved setningen

$$E(X) = np = 8 * 0.2 = 1.6$$

### Oppgave 6

Vi skal regne ut sannsynligheten  $P(115 \leq X \leq 150) = P(X \leq 150) - P(X \leq 114)$ , hvor  $X \sim \text{Bin}(900, 0.15)$ . Det er upraktisk å gjøre for hånd, men ved hjelp av Python kan vi regne det ut:

```
stats.binom.cdf(150, 900, 0.15) - stats.binom.cdf(114, 900, 0.15)
```

som gir svaret 0.899. (Vi skal se senere at et tilnærmet svar kan regnes ut for hånd ved hjelp av sentralgrenseteoremet)

### Oppgave 8

En litt mer krevende oppgave. Kursen etter  $n$  dager er en stokastisk variabel  $Y = 25 + X_1 + X_2 + \dots + X_n$ , hvor  $X_i$  er stokastiske variabler som er enten 1 eller  $-1$  med sannsynligheten 0.5.

Legg merke til at  $X_i = 2Z_i - 1$ , hvor  $Z_i$  er binomiske forsøk med  $p = 0.5$ , dvs  $Z_i \sim \text{Bin}(1, 0.5)$ . Derfor er

$$Y = 25 + (2Z_1 - 1) + \dots + (2Z_n - 1) = 25 + 2(Z_1 + \dots + Z_n) - n = (25 - n) + 2Z,$$

hvor  $Z = Z_1 + \dots + Z_n \sim \text{Bin}(n, 0.5)$ .

Kursen etter 3 dager er  $Y = 22 + 2Z$ , hvor  $Z \sim \text{Bin}(3, 0.5)$ .

Etter 5 dager er  $P(Y = 28) = P(20 + 2Z = 28) = P(Z = 4)$ , og vi regner ut

$$P(Z = 4) = \binom{5}{4} 0.5^4 0.5^1 = 5 * 0.5^5 = 0.156$$

Etter 8 dager er  $P(Y = 23) = P(17 + 2Z = 23) = P(Z = 3)$ , og

$$P(Z = 3) = \binom{8}{3} 0.5^6 0.5^2 = 56 * 0.5^8 = 0.219$$

## Oppgave 15

Sannsynlighetsfordelingen er bestemt av at det er  $x - 1$  dager på rad der hun ikke lykkes,  $p = 1 - 0.001 = 0.999$ , og 1 dag der hun lykkes,  $p = 0.001$ :

$$P(X = x) = (1 - 0.001)^{x-1} 0.001$$

Dette er den geometriske fordelingen med forventningsverdi  $E(X) = 1/0.001$  og varians  $\text{Var}(X) = (0.999)/0.0012$ :

$$\mu = 10^3 \text{ dager} \quad \sigma = 999.5 \text{ dager}$$

For å summere opp sannsynlighetene for å lykkes i løpet av et år kan du bruke summen til geometrisk rekke:

$$P(X \leq 365) = 0.001(1 + 0.999 + 0.999^2 + \dots) = 0.001 \frac{1 - 0.999^{365}}{1 - 0.999} = 1 - 0.999^{365} = 0.306$$

Alternativt kan du se på sannsynlighet for ikke å lykkes i løpet av et år og bruke komplementet:

$$P(X \leq 365) = 1 - 0.999^{365} = 0.306$$

*Obs: setningen  $P(X \leq x) = 1 - (1 - p)^x$  for sannsynlighetsfunksjon til den geometriske fordelingen er å finne i formelarket, slik at utledningene ovenfor er ikke nødvendig.*

## Oppgave 16

$X$  er Poissonfordelt med  $\lambda t = 0.25 * 12 = 3$ . Fra formelarket har vi  $E(X) = \lambda t = 3$ , og

$$P(X = 4) = \frac{(\lambda t)^4}{4!} e^{-\lambda t} = \frac{3^4}{4!} e^{-3} = 0.168$$

For å regne ut  $P(X \leq 4)$  kan vi enten ta summen  $P(X = 0) + P(X = 1) + \dots + P(X = 4)$  eller bruke tabell E2 (på side 576 av Løvås) eller Python. Vi får isåfall  $P(X \leq 4) = 0.815$

## Oppgave 17

Antall fisker  $X$  i 4 timer er også Poissonfordelt, denne gangen med  $\lambda t = 5 * 4 = 20$ . Da er

$$P(X = 15) = \frac{20^{15}}{15!} e^{-20} = 0.052,$$

mens tabellen viser (det siste tallet nederst til høyre!)  $P(X \leq 15) = 0.157$ .

## Oppgave 18

Se fasit i boka.

## Oppgave 21

$X$  er Poissonfordelt med  $\lambda t = 4 * 2 = 8$ , mens  $T$  er eksponentialfordelt med  $\lambda = 4$ . Derfor er

$$P(X = 6) = \frac{8^6}{6!} e^{-8} = 0.122$$

og  $P(4 \leq X \leq 10) = P(X \leq 10) - P(X \leq 3) = 0.816 - 0.042 = 0.774$ . I tillegg har vi

$$P(T > 0.5) = 1 - P(T \leq 0.5) = 1 - (1 - e^{-0.5 * \lambda}) = e^{-2} = 0.135$$

### Oppgave 25

Vi finner ved å bruke tabell E3:

$$\begin{aligned}P(Z < 0) &= 0.5, & P(Z < 1.22) &= \Phi(1.22) = 0.889 \\P(1.13 < Z < 2.45) &= \Phi(2.45) - \Phi(1.13) = 0.122 \\P(-0.84 < Z < 1.11) &= \Phi(1.11) - \Phi(-0.87) = \Phi(1.11) - (1 - \Phi(0.87)) = 0.674\end{aligned}$$

I den siste utregningen har vi brukt normalfordelingens symmetri-egenskap  $\Phi(-z) = 1 - \Phi(z)$ .

### Oppgave 27

Her må vi overføre den variable  $X$  til standard form,  $Z = \frac{X-\mu}{\sigma}$ , før vi kan bruke tabellen:

$$\begin{aligned}P(X < 167) &= \Phi\left(\frac{167 - 180}{8}\right) = \Phi(-1.625) = 1 - \Phi(1.625) = 0.052 \\P(X > 195) &= \Phi\left(\frac{195 - 180}{8}\right) = \Phi(1.875) = 0.970 \\P(175 < X < 182) &= P(X < 182) - P(X < 175) = \Phi\left(\frac{182 - 180}{8}\right) - \Phi\left(\frac{175 - 180}{8}\right) = 0.333\end{aligned}$$

### Oppgave 29

For et 99% sentrert spredningsintervall har vi en øvre og nedre “hale” lik  $(1-0.99)/2 = 0.005$  utenfor intervallet. Vi skal dermed finne kvantilverdiene  $z$  for  $\Phi(z_1) = 1 - 0.005 = 0.995$  og  $\Phi(z_2) = 0.005$ .

Tabellen E4 oppgir den øvre kvantilen til standard normalfordelingen:  $z_{0.005} = 2.576$ . For den nedre kvantilen bruker vi symmetriegenskapene til Normalfordelingen: Den nedre blir lik den øvre men med motsatt fortegn, her  $-2.576$ .

Det som gjenstår er å gå fra standardiserte  $z$ -verdier til observerte verdier  $x$  ved å bruke forventningsverdien og standardavviket:

$$\begin{aligned}x_1 &= \sigma z_1 + \mu = 16000 * 2.576 + 190000 = 231200 \\x_2 &= \sigma z_2 + \mu = 16000 * -2.576 + 190000 = 148800\end{aligned}$$

### Oppgave 30

Gitt bølgehøyden  $X \sim N(12, 2)$ , kravet i oppgaven formuleres slik: finn  $x$  slik at

$$P(X \leq x) = 0.95 \Rightarrow P(X > x) = 1 - 0.95 = 0.05$$

Bruker kvantiltabellen E4 etter at vi har normalisert  $X$ :

$$P(Z > z) = 0.05 \Rightarrow z = 0.1645$$

dvs fra Normalfordelingen har vi fant antall standardavvik  $z$  som tilsvarer den oppgitte sannsynligheten. Da regnes ut  $x$ -verdien ved  $x = \mu + z\sigma$ , slik at

$$x = 12 + 1.645 * 2 = 15.3$$

### Oppgave 32

Oppgitt de 4 normalfordelingene  $X_1 \sim N(4, 1)$ ,  $X_2 \sim N(5, 1)$ ,  $X_3 \sim N(7, 1)$  og  $X_4 \sim N(9, 1)$ . Summen av de 4,  $Y$ , er også normalfordelt, med forventning og varians gitt av regnereglene:

$$\begin{aligned}\mu_Y &= \mu_{X_1} + \mu_{X_2} + \mu_{X_3} + \mu_{X_4} = 4 + 5 + 7 + 9 = 25 \\ \sigma_Y^2 &= \sigma_{X_1}^2 + \sigma_{X_2}^2 + \sigma_{X_3}^2 + \sigma_{X_4}^2 = 4 \Rightarrow \sigma_Y = 2,\end{aligned}$$

dvs  $Y \sim N(25, 2)$ . Da er sannsynligheten for at prosjektet varer høyst 20 dager

$$P(Y \leq 20) = \Phi\left(\frac{20 - 25}{2}\right) = \Phi(-2.5) = 0.0062$$

### Oppgave 33

Ifølge sentralgrenseteoremet vil middelveiden av 50 variable med lik sanns.fordeling ha en sanns.fordeling som er svært nær normalfordelingen. Vi kan bruke reglene for forventnings-verdi og standardavvik til uavhengige variable:

$$\begin{aligned}\bar{X} &= \frac{X_1 + \dots + X_{50}}{50} \\ E(\bar{X}) &= \frac{1}{50} \sum_{i=1}^{50} E(X_i) = \frac{50 * 1500}{50} = 1500 \\ \sigma_{\bar{X}}^2 &= \frac{1^2}{50^2} \sum_{i=1}^{50} \sigma_{X_i}^2 = \frac{50 * 1500^2}{50^2} = \frac{1500^2}{50} \\ \sigma_{\bar{X}} &= \sqrt{\sigma_{\bar{X}}^2} = \frac{1500}{\sqrt{50}} \approx 212\end{aligned}$$

Sannsynligheten for at den gjennomsnittlige levetiden er større enn 1540 timer kan da tilnærmes med:

$$P(\bar{X} < 1540) = 1 - P(\bar{X} \leq 1540) = 1 - \Phi\left(\frac{1540 - 1500}{212}\right) = 0.425.$$