

# Skriftlig innlevering 3

## Oppgave 1 \*\*

En bedrift har kjøpt inn en ny kraftig datamaskin og vurderer behovet for vedlikeholdsarbeid på denne. De regner med at feilene som vil oppstå er av to typer, feil som de kan utbedre selv (type A-feil) og feil som krever assistanse fra merkeleverandør for å utbedre (type B-feil). De antar at feil av type A kommer ifølge en poissonprosess med intensitet  $\lambda_A$  (per døgn), og at feil av type B kommer ifølge en poissonprosess med intensitet  $\lambda_B$  (per døgn). Dessuten antar de at feil av type A og type B oppstår uavhengig av hverandre. La  $X_A$  og  $X_B$  være henholdsvis antall feil som vil oppstå av type A og av type B i løpet av et år (365 døgn).

### Deloppgave a) \*\*

Dersom  $\lambda_A = \frac{3}{73} = 0.0411$  og  $\lambda_B = \frac{1}{73} = 0.0137$ , finn

$$E[X_A], \quad P(X_A \leq 15), \quad P(X_A \leq 15, X_B \leq 10) \quad \text{og} \quad P(X_A \leq 15 | X_B \leq 10).$$

Her er deloppgave a) slutt.

La  $Z = X_A + X_B$  være totalt antall feil som oppstår i løpet av et år.

### Deloppgave b)

I forelesning ble det vist at summen av to uavhengige Poissonfordelte stokastiske variabler (med ulike rater) også er Poissonfordelt. Bruk nå momentgenererende funksjoner til å vise at summen av to uavhengige Poissonfordelte stokastiske variabler er Poissonfordelt.

### Deloppgave c) \*\*

Finn den betingede fordelingen for antall feil av type A i løpet av et år, gitt at totalt antall feil i løpet av dette året var lik et gitt tall  $z$ , dvs finn et uttrykk for

$$f_{X_A|Z}(x_A|z) = P(X_A = x_A | Z = z).$$

Kan du gjenkjenne denne betingede fordelingen som en fordeling vi har diskutert?

## Oppgave 2 \*\*

La  $X$  være fødselsvekt i gram til et tilfeldig barn i Norge. Anta at  $X$  er normalfordelt med forventningsverdi  $\mu = E[X] = 3500$  og standardavvik  $\sigma = SD[X] = 570$ .

## Deloppgave a)

Finn sannsynlighetene

$$P(X \geq 3000), \quad P(3200 \leq X < 4000), \\ P(X \geq 3500 | X \geq 3000) \quad \text{og} \quad P(X \leq 4000 | X \geq 3200).$$

### Her er deloppgave a) slutt

Et barn blir klassifisert som undervektig dersom fødselsvekta er under  $c$  gram, der  $c$  velges slik at

$$P(X < c) = 0.01.$$

La  $Y$  være antall av  $n = 100$  barn i Norge som var undervektige ved fødselen.

## Deloppgave b)

- Finn verdien til  $c$ .
- Hvilke antagelser må du gjøre (i tillegg til de som er spesifisert over) for at  $Y$  skal være binomisk fordelt?
- Forutsatt at  $Y$  er binomisk fordelt, finn  $P(Y \geq 1)$  og  $P(Y \geq 2 | Y \geq 1)$ .

## Oppgave 3 \*

Betrakt en poissonprosess med intensitet  $\lambda$ . La  $Z_1$  være tidspunkt for den første hendelsen i prosessen, la  $Z_2$  være tidspunkt for den andre hendelsen i prosessen, og tilsvarende videre slik at  $Z_k$  er tidspunkt for hendelse nummer  $k$  i prosessen.

La  $X_1 = Z_1$  og for  $k = 2, 3, \dots$  la

$$X_k = Z_k - Z_{k-1}.$$

Dermed har vi altså at  $X_1$  er tid til første hendelse i poissonprosessen, og for  $k = 2, 3, \dots$  er  $X_k$  tid mellom to etterfølgende hendelser i poissonprosessen. Vi vet dermed at  $X_1, X_2, \dots$  er uavhengige og ekseponensialfordelte, alle med samme parameterverdi  $\lambda$ .

## Deloppgave a)

Finn sannsynlighetene

$$P(X_1 \geq 2), \quad P(X_1 + X_2 \geq 4) \quad \text{og} \quad P(X_1 + X_2 \geq 4 | X_1 \geq 2).$$

Finn først formler for sannsynlighetene som funksjon av  $\lambda$ , og sett deretter inn  $\lambda = 0.5$  og regn ut en numerisk verdi.

*Merk at sannsynlighetene med to stokastiske variabler kan beregnes ved å sette opp dobbeltintegraler over egnet område eller ved å omskrive sannsynligheten til å omhandle en annen stokastisk variabel med en annen sannsynlighetsfordeling.*

**Her er deloppgave a) slutt.**

I de to neste deloppgavene skal du bevise ved hjelp av induksjon at  $Z_k$  er gammafordelt med parametre  $\alpha = k$  og  $\beta = \frac{1}{\lambda}$ . Vi starter med å vise at påstanden er korrekt for  $k = 1$ .

## Deloppgave b)

Sett  $\alpha = 1$  og  $\beta = \frac{1}{\lambda}$  i sannsynlighetstettheten til en gammafordeling og vis at sannsynlighetstettheten da blir identisk med en eksponensialfordeling med parameter  $\lambda$ . *Merk: Du har dermed bevist at  $Z_1$  er gammafordelt med parametre  $\alpha = 1$  og  $\beta = \frac{1}{\lambda}$ .*

## Deloppgave c)

Anta så at påstanden vi ønsker å bevise er korrekt for  $k = n - 1$ , dvs vi antar at  $Z_{n-1}$  er gammafordelt med parametre  $\alpha = n - 1$  og  $\beta = \frac{1}{\lambda}$ . Merk dessuten at vi vet at  $Z_n = Z_{n-1} + X_n$  og at  $Z_{n-1}$  og  $X_n$  er uavhengige stokastiske variabler.

Benytt setningen om total sannsynlighet (for kontinuerlige stokastiske variabler) til å vise at da blir  $Z_n$  gammafordelt med parametre  $\alpha = n$  og  $\beta = \frac{1}{\lambda}$ .

*Merk: Ved induksjon har du dermed bevist, for alle  $k = 1, 2, \dots$ , at  $Z_k$  er gammafordelt med parametre  $\alpha = k$  og  $\beta = \frac{1}{\lambda}$ .*

## Oppgave 4 \*

La  $X$  være normalfordelt med forventningsverdi  $\mu$  og varians lik  $\sigma^2$ , og la  $Y = aX + b$  for to konstanter  $a$  og  $b$ .

## Deloppgave a)

Finn  $E[Y]$  og  $\text{Var}[Y]$ .

Lag en pythonfunksjon som simulerer  $n$  realisasjoner av  $Y$  ved først å simulere  $n$  realisasjoner av  $X$  ved hjelp av pythonfunksjonen `numpy.random.normal`, og deretter regner ut tilsvarende verdier av  $Y$ . Funksjonen din skal ha fem input-parametre,  $n$ ,  $\mu$ ,  $\sigma$ ,  $a$  og  $b$ .

For  $\mu = 1$ ,  $\sigma = 2$ ,  $a = 2$  og  $b = 0.5$ , benytt pythonfunksjonen din til å generere  $n = 100\,000$  realisasjoner av  $Y$ . Lag et sannsynlighetshistogram av disse  $n$  verdiene

og spesifiser at sannsynlighetshistogrammet skal ha 100 intervaller (se eventuelt kode fra skriftlig innlevering 2).

Plott så sannsynlighetstettheten til en normalfordeling med forventningsverdi lik  $E[Y]$  og varians lik  $\text{Var}[Y]$  i samme plott som histogrammet. Observer (dersom du har gjort ting rett) at sannsynlighetstettheten sammenfaller meget godt med sannsynlighetshistogrammet, og konkluder dermed med at det kan se ut som om  $Y$  er normalfordelt.

Gjenta gjerne denne simuleringsprosedyren også for noen andre verdier av  $\mu$ ,  $\sigma$ ,  $a$  og  $b$  og observer at sannsynlighetshistogrammet også da sammenfaller med tilhørende sannsynlighetstetthet.

```
In [ ]: # Her kan du skrive phytonkoden din
```

## Deloppgave b)

Benytt transformasjonsformelen til å utlede sannsynlighetstettheten til  $Y$ , og vis at denne kan skrives på formen av en normalfordelingstetthet med forventningsverdi og varians lik formelene du fant for henholdsvis  $E[Y]$  og  $\text{Var}[Y]$  i punktet over.

*Merk: Du har nå vist at enhver **lineær** funksjon av en normalfordelt variabel er normalfordelt, slik simuleringsstudien du gjorde over antydte.*

## Oppgave 5

I denne oppgaven skal vi benytte momentgenererende funksjoner til å se på et spesialtilfelle av sentralgrenseteoremet.

Anta at vi har uavhengige og identisk fordelte stokastiske variabler  $X_1, X_2, \dots, X_n$ , det  $X_i \in \{-1, 1\}$  og

$$P(X_i = -1) = \frac{1}{2} \text{ og } P(X_i = 1) = \frac{1}{2}.$$

Som vanlig lar vi

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

være gjennomsnittet av  $X_i$ 'ene.

## Deloppgave a)

Utlede en formel for momentgenererende funksjon for  $X_i$ ,  $M_{X_i}(t)$ .

Benytt  $M_{X_i}(t)$  til å finne  $E[X_i]$  og  $E[X_i^2]$ . Bestem også  $\text{Var}[X_i]$ .

Bruk så regneregler for forventningsverdi og varians til å bestemme  $E[\bar{X}]$  og  $\text{Var}[\bar{X}]$ .

Her er deloppgave a) slutt.

La  $U$  betegne er standardisert versjon av  $\bar{X}$ , dvs la

$$U = \frac{\bar{X} - E[\bar{X}]}{\sqrt{\text{Var}[\bar{X}]}}.$$

## Deloppgave b)

Benytt regneregler for momentgenererende funksjoner til å finne et uttrykk for  $M_{\bar{X}}(t)$ .

Benytt så uttrykkene du fant for  $E[\bar{X}]$  og  $\text{Var}[\bar{X}]$  i deloppgave a), og igjen regneregler for momentgenererende funksjoner, til å finne et uttrykk for  $M_U(t)$ . Vis spesielt at  $\ln M_U(t)$  kan skrives på formen

$$\ln M_U(t) = n \ln \left( \frac{\exp\left\{\frac{t}{\sqrt{n}}\right\} + \exp\left\{-\frac{t}{\sqrt{n}}\right\}}{2} \right).$$

Her er deloppgave b) slutt.

Sentralgrenseteoremet sier som kjent at fordelingen til  $U$  vil konvergere mot en standard normalfordeling når  $n \rightarrow \infty$ . Dette betyr at  $M_U(t)$  må konvergere mot den momentgenererende funksjonen til en standard normalfordelt variabel  $Z \sim N(0, 1)$ , som er gitt ved

$$M_Z(t) = \exp\left\{\frac{t^2}{2}\right\}.$$

## Deloppgave c)

For å undersøke om det ser ut til å stemme at  $M_U(t)$  konvergerer mot  $M_Z(t)$  når  $n \rightarrow \infty$ , lag et plott av begge funksjonene  $M_U(t)$  og  $M_Z(t)$  (for eksempel) i intervallet  $t \in [0, 1]$ . Lag plottet for ulike verdier av  $n$ , for eksempel for  $n = 4$ ,  $n = 10$  og  $n = 100$ . Ser det ut til at sentralgrenseteoremet stemmer i dette tilfellet?

```
In [ ]: # Her kan du skrive koden for å plotte de to funksjonene.
```

Bevis så matematisk at  $M_U(t)$  konvergerer mot  $M_Z(t)$  ved å vise at

$$\lim_{n \rightarrow \infty} \ln M_U(t) = \frac{t^2}{2}.$$

*Hint: For å bestemme grensen trenger du å benytte L'Hopitals regel. For å komme frem trenger du faktisk å benytte L'Hoptals regel to ganger.*

## Fasit:

- Oppgave 1a): 15, 0.5681, 0.5603, 0.5681
- Oppgave 2a): 0.8106, 0.5125, 0.6168, 0.7302
- Oppgave 2b): 2174.2, 0.6340, 0.4168
- Oppgave 3a): 0.3679, 0.4060, 0.7358
- Oppgave 4a):  $E[Y] = a\mu + b$ ,  $\text{Var}[Y] = a^2\sigma^2$
- Oppgave 5a):  $E[\bar{X}] = 0$ ,  $\text{Var}[\bar{X}] = \frac{1}{n}$