Plenumsregning — TMA4245 — Uke 15

Repetisjon 1

a)

1. $f(x,y) \geq 0 \quad \forall x,y.$ OK!
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_0^1\int_0^1 (x+y)\,dx\,dy$

$$= \int_0^1 \left[\frac{x^2}{2} + xy\right]_{x=0}^{x=1} dy$$

$$= \int_0^1 \left(\frac{1}{2} + y\right) dy$$

$$= \left[\frac{y}{2} + \frac{y^2}{2}\right]_{y=0}^{y=1}$$

$$= 1. \qquad \text{OK!}$$

Ja, gyldig.

b)

$$P(Y \leq 0.5) = \int_0^{0.5}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

$\Downarrow$ som over

$$= \int_0^{0.5} \left(\frac{1}{2} + y\right) dy$$

$$= \left[\frac{y}{2} + \frac{y^2}{2}\right]_0^{0.5}$$

$$= \frac{1}{4} + \frac{1}{8} = \underline{3/8}$$

$$P(X \leq 0.5 \cap Y \leq 0.5) = \int_0^{0.5}\int_0^{0.5} (x+y)\,dx\,dy$$

$$= \int_0^{0.5} \left[\frac{x^2}{2} + xy\right]_{x=0}^{0.5} dy$$

$$= \int_0^{0.5} \left(\frac{1}{8} + \frac{1}{2}y\right) dy$$

$$= \left[\frac{y}{8} + \frac{1}{4}y^2\right]_0^{0.5}$$

$$= \frac{1}{16} + \frac{1}{16} = \underline{1/8}$$

$$P(X \leq 0.5 \cup Y \leq 0.5) = P(X \leq 0.5) + P(Y \leq 0.5) - P(X \leq 0.5 \cap Y \leq 0.5)$$

$\Downarrow$ Symmetry: $P(X \leq 0.5) = P(Y \leq 0.5) = 3/8$

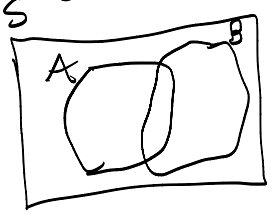
$$= \frac{3}{8} + \frac{3}{8} - \frac{1}{8} = \underline{5/8}$$

c) $P(A \cap B) > 0$. Ikke disjunkte!

$$P(A)P(B) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64} \neq \frac{1}{8} = P(A \cap B).$$

Ikke uavhengige!

Venndiagram



(Overlapper med areal $\approx 1/8$
Hver har areal $\approx 3/8$)

---

d) $$f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 (x+y)\,dx$$
$$= \left[\frac{x^2}{2} + xy\right]_{x=0}^{x=1}$$
$$= \frac{1}{2} + y, \quad 0<y<1$$

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{(x+y)}{\frac{1}{2}+y} \quad \text{For } 0<x<1$$

Repetisjon 2

1a)

B ser uavhengig ut. C er definitivt avhengige.
A ser uavhengig ut fordi X har samme fordeling for alle $Y=y$.

1b)

A ser ukorrelert ut. Virker ikke å være en lineær avhengighet mellom X og Y.

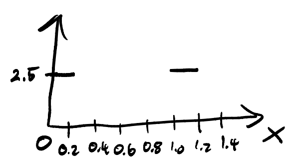
B ser ukorrelert ut siden de ser uavhengige ut.

C ser korrelerte ut. Virker å være positiv korrelasjon mellom X og Y.

2a)

A ser ca. uniformt ut i områder $[0,0.2]^2$, $[0,0.2]\times[1,1.2]$, $[1,1.2]\times[0,0.2]$ og $[1,1.2]^2$.

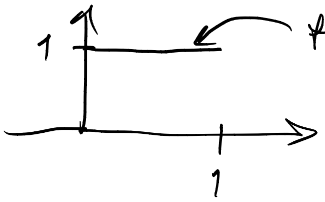
$\Rightarrow$ Skisse



$$f(x) = \begin{cases} 2.5, & 0<x<0.2 \text{ og } 1<x<1.2, \\ 0, & \text{ellers.} \end{cases}$$

B ser ut som uniform fordeling på $[0,1]$.

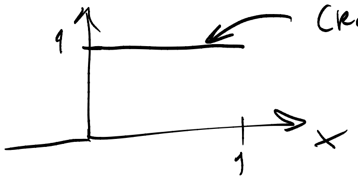
$\Rightarrow$ Skisse



$f(x) = 1, \quad 0<x<1.$

C er vanskelig å se eksakt. Men kan virke som tettheten av er konstant innenfor $0<x<0.1$, $0.1<x<0.2$, osv.

$\Rightarrow$ Skisse



(kanskje) $f(x) = 1, \quad 0<x<1.$

2b)

Basert på skisser bør det være.

A: $E[X] \approx 0.6$

B: $E[X] \approx 0.5$

C: $E[X] \approx 0.5$

Repetisjon 3

a)

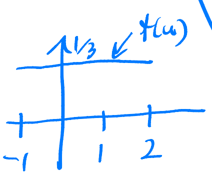
Merk at $f(u) = \frac{1}{3}, \quad -1<u<2.$

$$F_X(x) = P(X \le x) = P(U^2 \le x) = P(-\sqrt{x} \le U \le \sqrt{x}) \quad (\text{NB: } x \ge 0)$$

$$= \int_{-\sqrt{x}}^{\sqrt{x}} f(u)\,du$$

$$= \begin{cases} \frac{2\sqrt{x}}{3}, & 0 \le x < 1 \\ \frac{1}{3} + \frac{\sqrt{x}}{3}, & 1 \le x < 4 \end{cases}$$

Tegn integralet for å se dette



Det vil si

$$F_X(x) = \begin{cases} 0, & x<0 \\ \frac{2}{3}\sqrt{x}, & 0 \le x < 1 \\ \frac{1}{3} + \frac{\sqrt{x}}{3}, & 1 \le x < 4 \\ 1, & 4 \le x \end{cases}$$

b)

$$f_X(x) = F_X'(x) = \begin{cases} 0, & x<0 \\ \frac{1}{3\sqrt{x}}, & 0<x<1 \\ \frac{1}{6\sqrt{x}}, & 1<x<4 \\ 0, & 4>x \end{cases}$$

[NB: Ikke definert for $x=0, 1, 4$. Dette er Ok.]

c)

Skal løse $F_X(\tilde{x}) = 0.5$. Fra a) ser vi at det må skje for $0<x<1$. Altså må vi løse

$$\frac{2}{3}\sqrt{\tilde{x}} = 1/2$$

$$\sqrt{\tilde{x}} = 3/4$$

$$\tilde{x} = 9/16$$

Repetisjon 4

a)

Foreslår $\hat{p} = \bar{X}$.

$n\hat{p} = X_1 + \dots + X_n$ er en sum av iid stokastiske variable og Sentralgrenseteoremet gir $\hat{p} \overset{[\approx]}{\sim} N(E[\hat{p}], Var[\hat{p}])$. Vi ser at $E[\hat{p}] = E[\bar{X}] = p$ og

$$Var[\hat{p}] = Var[\bar{X}] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2}\cdot n \cdot p(1-p) = \frac{p(1-p)}{n}$$

Vi får dermed tilnærmet fordeling

$$\hat{p} \overset{[\approx]}{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

Vi tilnærmer i tillegg variansen $\frac{p(1-p)}{n} \approx \frac{\hat{p}(1-\hat{p})}{n}$ slik at vi får

$$\hat{p} \overset{[\approx]}{\sim} N\left(\underset{\text{ukjent}}{p}, \underset{\text{kjent}}{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

---

b)

Standard 90% konfidensintervall basert på

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

er $\hat{p} \pm z_{0.05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Konsekvensen av å bruke dette konfidensintervallet er at vi ikke kan garantere 90% konfidens. Det kan være lavere (eller høyere) konfidens avhengig av hvor gode de to tilnærmingene 1) normalfordelt og 2) $SD[\hat{p}] = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ er.

c)

Steg 1:

$Y \overset{[\approx]}{\sim} N(mp, mp(1-p))$ fra sentralgrenseteoremet (hvis mp og $m(1-p)$ er store nok).

Steg 2:

Tilnærmer variansen med $Var[Y] \approx m\hat{p}(1-\hat{p})$

NB: $\hat{p} = \frac{x_1 + \dots + x_n}{n}$. Vi har ikke sett Y!

$$\Rightarrow Y \overset{[\approx]}{\sim} N(mp, m\hat{p}(1-\hat{p}))$$

Steg 3

Ønsker $Z = \frac{Y - mp}{\sqrt{m\hat{p}(1-\hat{p})}} \sim N(0,1)$, men p ukjent! Må bytte med $\hat{p}$

$$\Rightarrow Z = \frac{Y - m\hat{p}}{\sqrt{m\hat{p}(1-\hat{p}) + \frac{m^2}{n}\hat{p}(1-\hat{p})}} \sim N(0,1)$$

[Merk at $E[Y - m\hat{p}] = mp - mp$ og $Var[Y - m\hat{p}] = Var[Y] + m^2 Var[\hat{p}] = m\hat{p}(1-\hat{p}) + \frac{m^2}{n}\hat{p}(1-\hat{p})$]

Steg 4:

Standard Fremgangsmåte gir 90% prediksjonsintervall for y

$$m\hat{p} \pm z_{0.05} \cdot m\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\hat{p}\,(1-\hat{p})}$$

Repetisjon 5

a)

Ja. Standard med σ kjent.

b)

Nei. $z_{0.05}$ må byttes med $t_{0.05,\,n-1}$.

c)

Ja. Konstruksjon av forkastningsområde for tosidig test med $\alpha=0.1$ identisk med konstruksjon av konfidensintervall med konfidens $100(1-\alpha)\% = 90\%$.

d)

Nei. Forventer at Z er bedre siden vi må estimere varians når vi bruker T.

e)

Ja. Følger av sentralgrenseteoremet

f)

Nei!

g)

Ja. Med stor nok n kan vi bli vilkårlig sikre.

Repetisjon 6

1.

$T_1 \sim \chi^2_n$ [Siden $\frac{(Y_i - \alpha - \beta x_i)^2}{\sigma^2} \sim \chi^2_1$ og sum av n uavhengige χ^2_1-fordelte variabler gir χ^2_n]

$T_2 \sim \chi^2_{n-1}$ [Mister én frihetsgrad på å bytte α med $\hat{A}$]

$T_3 \sim \chi^2_{n-1}$ [Mister én frihetsgrad på å bytte β med $\hat{B}$]

$T_4 \sim \chi^2_{n-2}$ [Mister to frihetsgrader på å bytte α & β med $\hat{A}$ og $\hat{B}$]

2.

$$S_1^2 = \frac{\sigma^2}{n} T_1 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2 \quad (\text{Siden } E[T_1] = n)$$

$$S_2^2 = \frac{\sigma^2}{n-1} T_2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \hat{A} - \beta x_i)^2 \quad (\text{Siden } E[T_2] = n-1)$$

$$S_3^2 = \frac{\sigma^2}{n-1} T_3 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \alpha - \hat{B} x_i)^2 \quad (\text{Siden } E[T_3] = n-1)$$

$$S_4^2 = \frac{\sigma^2}{n-2} T_4 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{A} - \hat{B} x_i)^2 \quad (\text{Siden } E[T_4] = n-2)$$

3.

Bare S_4^2 er uavhengig av både $\hat{A}$ og $\hat{B}$.

Dette klarer dere ikke å vise basert på det dere har lært. S_4^2 uavhengig av $\hat{A}$ og $\hat{B}$ er gitt i formelsamlingen og kan brukes uten bevis.

Intuisjon for at S_1^2 er avhengig av $\hat{A}$ og $\hat{B}$:

Hvis $|\hat{A}-\alpha|$ eller $|\hat{B}-\beta|$ er store forventer vi større kvadratavvik fra $\alpha+\beta x$, og dermed større verdier av S_1^2. S_1^2 må dermed være avhengig av $\hat{A}$ og $\hat{B}$.

(NB: Dette er ikke et matematisk bevis)