

Funksjoner av stokastiske variable.

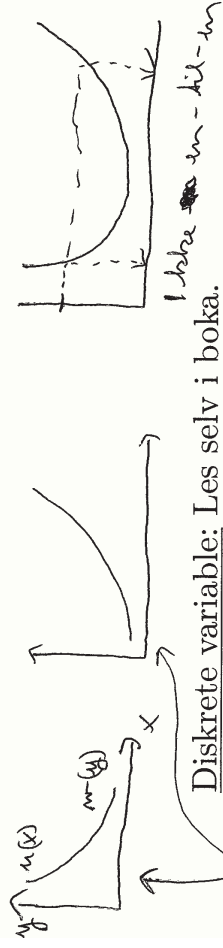
Dekkes av pensumssidene i kap. 7

I ulike sammenhenger, blant annet for å finne fordelingen til estimatorer, er vi interesserte i fordelingen til funksjoner av stokastiske variable, f.eks. fordelingen til $Y = u(X)$ når vi kjenner fordelingen til X , eller fordelingen til $W = g(X_1, \dots, X_n)$ når vi kjenner fordelingen til $X_1, \dots, X_n$.

Transformasjoner (7.2)

Vi kjenner fordelingen til X . Ønsker å finne fordelingen til $Y = u(X)$. (Boka omhandler også situasjonen når vi har to funksjoner av to variable - vi skal hovedsaklig se på situasjonen med en variabel.)

Eksempel: Vi kjenner fordelingen til hastigheten til molekylene i en gass, X , hva er fordelingen til energien $Y = \frac{1}{2}mX^2$? $\square$



Kontinuerlige variable:

La X være en kontinuerlig stok. var. med sanns. tetthet $f(x)$. Anta videre at $Y = u(X)$ definerer en en-til-en transformasjon fra X til Y . Vi ønsker å finne fordelingen til Y .

Siden $u(X)$ er en-til-en har vi også en entydig transformasjon fra Y til X , $X = w(Y) = u^{-1}(Y)$. Sannsynlighetstettheten til Y er da gitt ved **transformasjonsformelen**:

$$g(y) = f(w(y))|w'(y)|$$

(Se side 34 i formelsamlingen/teorem 7.3 i boka)

Bervis:

1. $u(X)$ voksende:

$$F_Y(y) = P(Y \leq y) = P(X \leq w(y)) = F_X(w(y))$$

$$g(y) = F'_Y(y) = \frac{d}{dy} F_X(w(y)) = f(w(y))w'(y)$$



2. $u(X)$ avtakende:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X \geq w(y)) \\ &= 1 - P(X \leq w(y)) = 1 - F_X(w(y)) \\ g(y) &= \frac{d}{dy}(1 - F_X(w(y))) = -f(w(y))w'(y) \\ &= f(w(y))|w'(y)| \text{ siden } w'(y) \text{ negativ.} \end{aligned}$$

Alternativt bevis: Se boka. $\square$

Eksempel: Fordelingen til hastigheten, X , til molekylene i en gass er gitt ved:

$$f(x) = 4x^3 e^{-x^4}, \quad x > 0$$

Hva blir fordelingen til kinetisk energi til molekylene, dvs fordelingen til $Y = \frac{1}{2}mX^2$? Siden $X > 0$ er transformasjonen en-til-en.

$X = \sqrt{2Y/m} = w(Y)$ og vi får:

$$\begin{aligned} g(y) &= 4\left(\frac{2y}{m}\right)^{(3/2)} e^{-(2y/m)^2} \underbrace{\frac{1}{2} \frac{2y}{m}^{-1/2}}_{f(w(y))} \frac{2}{m} \\ &= \frac{4}{m} \frac{2y}{m} e^{-(2y/m)^2} \underbrace{\frac{1}{2} \frac{2y}{m}^{-1/2}}_{w'(y)} \\ &= \frac{8y}{m^2} e^{-4y^2/m^2}, \quad y > 0 \quad \square \end{aligned}$$

$$y = \frac{1}{2}mx^2$$

$$\begin{aligned} \frac{2y}{m} &= x^2 \\ x &= \left(\frac{2y}{m}\right)^{1/2} = w(y) \quad w'(y) = \frac{1}{2} \left(\frac{2y}{m}\right)^{-1/2} \cdot \frac{2}{m} \end{aligned}$$

Eksempel: $X \sim N(\mu, \sigma^2)$. Hva er fordelingen til $Z = \frac{X-\mu}{\sigma}$?

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}, \quad -\infty < x < \infty$$

$X = \sigma Z + \mu = w(Z)$ er en-til-en og vi får

$$\begin{aligned} g(z) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(\sigma z + \mu - \mu)^2}{\sigma^2}} \sigma \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} \underbrace{f(w(z))}_{w'(z)} \quad -\infty < z < \infty \end{aligned}$$

Dvs $Z \sim N(0, 1)$. $\square$

Merk: Det er ikke alltid slik som i eksemplet over at en lineærkombinasjon har samme fordeling (bare med andre parametre) som den opprinnelige variabelen.

$$Z = \frac{X-\mu}{\sigma}$$

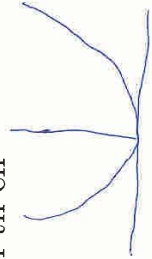
$$0Z = X - \mu$$

$$X = 0Z + \mu = w(Z)$$

$$w'(Z) = 0$$

Hva gjør vi dersom transformasjonen ikke er en-til-en?

Eksempel: $X \sim N(\mu, \sigma^2)$. Hva er fordelingen til $Y = (\frac{X-\mu}{\sigma})^2 = Z^2$? Ikke en-til-en for alle mulige Z , men kan deles opp i to en-til-en funksjoner:



- for $z < 0$: $z = -\sqrt{y} = w_1(y)$
- for $z \geq 0$: $z = \sqrt{y} = w_2(y)$

Kan vises (se boksa) at da er:

$$g(y) = f(w_1(y))|w'_1(y)| + f(w_2(y))|w'_2(y)|$$

I vårt eksempel gir dette:

$$\begin{aligned} g(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-\sqrt{y})^2} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \frac{1}{2\sqrt{y}} \\ &= \frac{1}{2^{1/2}\sqrt{\pi}} e^{-\frac{y}{2}} \frac{1}{\sqrt{y}} \\ &= \frac{1}{2^{1/2}\Gamma(1/2)} y^{-1/2} e^{-y/2}, \quad y \geq 0 \end{aligned}$$

Dette er en χ^2_1 -fordeling! $\square$

$$Z \sim N(0,1) \text{ da er } Z^2 \sim \chi^2_1$$

Eksempel Bølgefunksjon $f_X(x) = \frac{2x}{\theta} e^{-x^2/\theta}$

Observat: $X_1, \dots, X_n$

$$SME: \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i^2$$

Fordelingsfunkt?

Se først på fordelingen til $Y = X^2$

$$X = y^{1/2} = w(y), \quad w'(y) = \frac{1}{2} y^{-1/2}$$

$$f_Y(y) = \underbrace{\frac{2y^{1/2}}{\theta}}_{f_X(w(y))} \underbrace{e^{-y/\theta} \cdot \frac{1}{2} y^{-1/2}}_{w'(y)} = \frac{1}{\theta} e^{-y/\theta}$$

Der $Y = X^2 \sim \text{eksponensial}(\theta)$

$$\Rightarrow E(X^2) = \theta \text{ og } \text{Var}(X^2) = \theta^2$$

$$\Rightarrow E(\hat{\theta}) = E\left(\frac{1}{n} \sum X_i^2\right) = \frac{1}{n} \sum E(X_i^2) = \frac{1}{n} \sum \theta = \theta \quad \text{fordelingsfunkt}$$

$$\text{Var}(\hat{\theta}) = \text{Var}\left(\frac{1}{n} \sum X_i^2\right) = \frac{1}{n^2} \text{Var}\left(\sum X_i^2\right)$$

$$\begin{aligned} \text{var.} \quad & \Rightarrow \frac{1}{n^2} \sum \text{Var}(X_i^2) \\ &= \frac{1}{n^2} \sum \theta^2 = \frac{\theta^2}{n} \end{aligned}$$

Summer av stok. var.

Eksempel: Du kjøper to lyspærer til lampen din. La X_1 og X_2 være funksjonstiden til de to pærene. Når den første har feilet setter du i den neste slik at den samlede funksjonstiden blir $Y = X_1 + X_2$. Hva blir fordelingen til Y ? Anta at X_1 og X_2 er uavh. og eksponensialfordelte med parameter β .

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) = P(X_1 + X_2 \leq y) \\
 &= \iint_{x_1+x_2 \leq y} \frac{1}{\beta} e^{-x_1/\beta} \frac{1}{\beta} e^{-x_2/\beta} dx_2 dx_1 \\
 &= \int_0^y \int_0^{y-x_1} \frac{1}{\beta} e^{-x_1/\beta} \frac{1}{\beta} e^{-x_2/\beta} dx_2 dx_1 \\
 &= \int_0^y \frac{1}{\beta} e^{-x_1/\beta} \int_0^{y-x_1} \frac{1}{\beta} e^{-x_2/\beta} dx_2 dx_1 \\
 &= \int_0^y \left(\frac{1}{\beta} e^{-x_1/\beta} - \frac{1}{\beta} e^{-y/\beta} \right) dx_1 \\
 &= 1 - e^{-y/\beta} - \frac{y}{\beta} e^{-y/\beta}
 \end{aligned}$$

$$\begin{aligned}
 f_Y(y) &= F'_Y(y) \\
 &= \frac{1}{\beta} e^{-y/\beta} - \frac{1}{\beta} e^{-y/\beta} + \frac{y}{\beta^2} e^{-y/\beta} \\
 &= \frac{1}{\beta^2} y e^{-y/\beta}
 \end{aligned}$$

Dette er en gamma-fordeling med parametre $\alpha = 2$ og β . Videre kan det på samme måte vises at dersom $X_1, \dots, X_n$ er uavh. eksponensialfordelte med parameter β så er $Y = \sum_{i=1}^n X_i$ gamma-fordelt med parametre $\alpha = n$ og β . $\square$

La oss se på denne typen problemer mer generelt. La X_1 og X_2 være to uavhengige stok. var. med sanns. tettheter $f_{X_1}(x_1)$ og $f_{X_2}(x_2)$. Vi ønsker å finne fordelingen til $Y = X_1 + X_2$.

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) = P(X_1 + X_2 \leq y) \\
 &= \iint_{x_1+x_2 \leq y} f_{X_1}(x_1) f_{X_2}(x_2) dx_2 dx_1 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{y-x_1} f_{X_1}(x_1) f_{X_2}(x_2) dx_2 dx_1
 \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) \int_{-\infty}^{y-x_1} f_{X_2}(x_2) dx_2 dx_1 \\
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) F_{X_2}(y-x_1) dx_1 \\
f_Y(y) &= F'_Y(y) \\
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y-x_1) dx_1
\end{aligned}$$

Uttrykket i siste linje kalles konvolusjonsformelen. Tilsvarende kan vises for diskrete stok. var. (integral erstattes da av summer) slik at vi totalt får:

Konvolusjonsformelen: Fordelingen til $Y = X_1 + X_2$ der X_1 og X_2 er uavh. stok. var. finnes ved:

$$f_Y(y) = \begin{cases} \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(y-x) dx & , \text{ kont.} \\ \sum_x f_{X_1}(x) f_{X_2}(y-x) & , \text{ disk.} \end{cases}$$

Konvolusjonsformelen kan generaliseres til summer av flere variable.

Eksempel: Mellom to byer går det to veier. Antall trafikkulykker per år på den ene veien, X_1 , er Poisson-fordelt med forventning μ_1 , mens antall ulykker på den andre veien, X_2 , er Poisson-fordelt med forventning μ_2 . Hva er fordelingen til totalt antall ulykker på de to veiene, $Y = X_1 + X_2$?

$$\begin{aligned}
f_Y(y) &= \sum_x f_{X_1}(x) f_{X_2}(y-x) \\
&= \sum_{x=0}^y \frac{\mu_1^x}{x!} e^{-\mu_1} \frac{\mu_2^{y-x}}{(y-x)!} e^{-\mu_2} \\
&= \sum_{x=0}^y \frac{y!}{x!(y-x)!} \mu_1^x \mu_2^{y-x} e^{-(\mu_1+\mu_2)} \frac{1}{y!} \\
&= \frac{1}{y!} e^{-(\mu_1+\mu_2)} \sum_{x=0}^y \binom{y}{x} \mu_1^x \mu_2^{y-x} \\
&= \frac{1}{y!} e^{-(\mu_1+\mu_2)} (\mu_1 + \mu_2)^y \\
&= \frac{(\mu_1 + \mu_2)^y}{y!} e^{-(\mu_1+\mu_2)}
\end{aligned}$$

Y er altså Poisson-fordelt med forventningsverdi $\mu_1 + \mu_2$.

Noen fordelinger (f.eks. normal, binomisk, Poisson og χ^2) har den egenskapen, som vist for Poisson, at summen av uavhengige variable er i samme fordelingsklasse (dvs summen har samme fordeling bare med andre parameterverdier). Se side 34/35 i formelsamlingen for en oversikt.

Merk at det bare er noen fordelinger som har denne egenskapen - det er ikke en generell egenskap ved fordelinger. F.eks. så vi at summen av to identisk eksponensialfordelte variable ble gammafordelt.

NR!

La $x_1 = w_1(\lambda_1, \lambda_2)$ og $x_2 = w_2(\lambda_1, \lambda_2)$, der funksjonene w_1 og w_2 er strengt monotone og deriverbare for alle verdier av argumentene. La $X_1 = w_1(Y_1, Y_2)$ og $X_2 = w_2(Y_1, Y_2)$. Hvis (X_1, X_2) er kontinuerlig fordelt vil også (Y_1, Y_2) være det, og vi har

$$g(y_1, y_2) = f(w_1(y_1, y_2), w_2(y_1, y_2)) \cdot |J|$$

der J er determinanten til Jacobi-matrisen, $J = \begin{vmatrix} \partial w_1 / \partial y_1 & \partial w_1 / \partial y_2 \\ \partial w_2 / \partial y_1 & \partial w_2 / \partial y_2 \end{vmatrix}$.

Lineærkombinasjoner

La $Y = \sum_{i=1}^n a_i X_i + b$. Da er

$$E(Y) = \sum_{i=1}^n a_i E(X_i) + b,$$

$$\text{Var}(Y) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^n \sum_{j=1}^{i-1} a_i a_j \text{Cov}(X_i, X_j).$$

Fordeling til lineærkombinasjoner

La $X_1, \dots, X_n$ være uavhengige variable.

Dersom X_i er normalfordelt med forventning μ_i og varians σ_i^2 vil $Y = \sum_{i=1}^n a_i X_i$ være normalfordelt med forventning $\sum_{i=1}^n a_i \mu_i$ og varians $\sum_{i=1}^n a_i^2 \sigma_i^2$. Dersom X_i er binomisk fordelt med parameter m_i og p vil $Y = \sum_{i=1}^n X_i$ være binomisk fordelt med parameter $\sum_{i=1}^n m_i$ og p . Dersom X_i er Poissonfordelt med parameter μ_i vil $Y = \sum_{i=1}^n X_i$ være Poissonfordelt med parameter $\sum_{i=1}^n \mu_i$. Dersom X_i er χ^2 -fordelt med ν_i frihetsgrader vil $Y = \sum_{i=1}^n X_i$ være χ^2 -fordelt med $\sum_{i=1}^n \nu_i$ frihetsgrader.

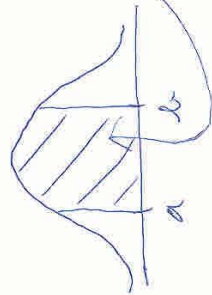
Regulering

X kontinuerlig

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du$$

↑
kumulativ fordelingsfunksjon

$$f(x) = F'(x)$$



$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

$$= F(b) - F(a)$$

$$P(X=a) = 0$$

X diskont

↑
 $f(x)$ sannsynlighetsfunksjon



$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u)$$

↑
kumulativ fordelingsfunksjon

$$P(a \leq X \leq b) = \sum_{x=a}^b f(x)$$

$$P(X=a) = f(a)$$

Estimering 2 -Konfidensintervall

Dekkes av kap. 9.4-9.5, 9.10, 9.12 og forelesingsnotatene.

En (punkt-)estimator $\hat{\theta}$ gir oss et anslag på en ukjent parameterverdi, men gir oss ikke noen direkte informasjon om usikkerheten i anslaget.

Et *konfidensintervall* er et intervall $[\hat{\theta}_L, \hat{\theta}_U]$ som med stor sannsynlighet vil komme til å inneholde den sanne parameterverdien θ .

Definisjon:

Intervall $[\hat{\theta}_L, \hat{\theta}_U]$ er et $(1 - \alpha)100\%$ konfidensintervall for θ dersom

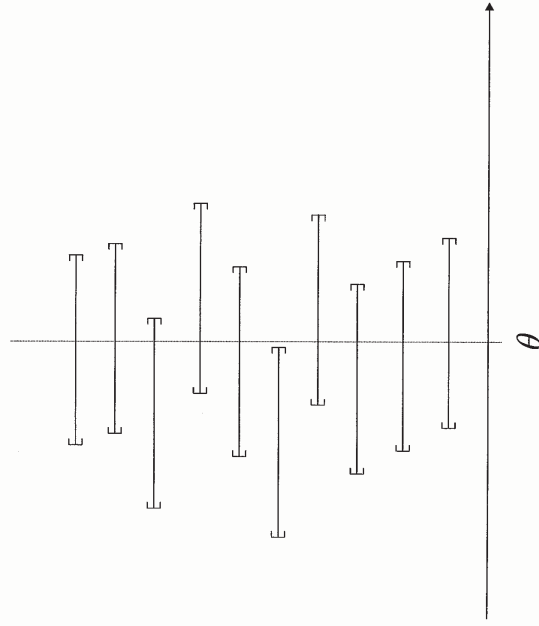
$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha \quad \square$$

$$\alpha = 0.01$$

$$\alpha = 0.05$$

$$\alpha = 0.10$$

Dersom forsøket gjentas mange ganger vil $(1 - \alpha)100\%$ av intervallene $[\hat{\theta}_L, \hat{\theta}_U]$ inneholde den ukjente parameteren θ (som er et fast men ukjent tall).



Eksempel: Bærebjelker. Bæreevnen (hvor stor belastning en bjelke tåler før den bryter sammen) til bærebjelker av en bestemt type antas normalfordelt med forventning μ og varians σ^2 . Målinger av bæreevnen til 14 bjelker gav resultatene: 5.9, 6.1, 5.4, 5.9, 5.7, 6.5, 5.8, 6.0, 5.3, 6.3, 5.8, 5.8, 5.8 og 6.0. Basert på denne informasjonen ønsker vi å lage konfidensintervall for μ og σ^2 . $\square$

Eksempel: Levetiden til en bestemt type krets-kort antas eksponensialfordelt med forventning β . Vi har observert at levetiden for seks slike kort ble hhv 612, 1009, 95, 1303, 599, og 780. Basert på dette ønsker vi å lage konfidensintervall for β . $\square$

Fordelingen til noen estimatorer

For å kunne lage konfidensintervall, og senere for å gjøre hypotesetester, må vi kjenne fordelingen til de ulike estimatorene.

Fordelingen til $\bar{X}$

$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ er en lineærkombinasjon av u.i.f. variable. Fordelingen til $\bar{X}$ vil avhenge av fordelingen til $X_1, \dots, X_n$.

$X_1, \dots, X_n$ normalfordelte og σ kjent:

Da vet vi at en lineærkombinasjon av uavh. normalfordelte variable er normalfordelt (f.saml. side 34), og vi får at

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

$X_1, \dots, X_n$ normalfordelte og σ ukjent:

Vi erstatter da σ^2 med estimatoren S^2 , og får:

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}$$

Fordelingen er ikke lenger en normalfordeling, men en (student) t -fordeling. t -fordeling ligner $N(0, 1)$ men har større varians. Se boka(8.7)/f.saml.(s. 27) for flere detaljer om t -fordelingen.

$X_1, \dots, X_n$ ikke normalfordelte og n stor:

Da gir SGT oss at:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \approx N(0, 1)$$

$X_1, \dots, X_n$ ikke normalfordelte og n liten:

Avhenge av fordelingen til $X_1, \dots, X_n$, må se på hver enkelt situasjon.