

Hovedpointer fra SaSt

Martin Nørgaard Petersen*

9. januar 2019

Følgende gennemgår udvalgte begreber fra 'En Introduktion til Sandsynlighedsregning' af M. Sørensen (9. udgave), 'Introduction to Likelihood-based Estimation and Inference' af H.B. Nielsen. Der tages forbehold for fejl og mangler.

Undervisningen i Sandsynlighedsteori og Statistik (SaSt) dækker i høj grad over indlæring af diverse regneregler. Nærværende note har ikke til formål at gengive alle disse regneregler – her henvises til grundbøger, slides, mv. Noten vil opridse de vigtigste resultater med henblik på en grundlæggende forståelse.

1 Begreber

1.1 Overordnet

1.1.1 Sandsynlighedsfunktion

også p.f. (probability function). Er en funktion p fra en endelig mængde M ind i intervallet $[0, 1]$. Der skal gælde at summen af sandsynlighederne på mængden M giver 1.

1.1.2 Sandsynlighedsmål

eller fordelingsfunktion. Til enhver sandsynlighedsfunktion findes et sandsynlighedsmål $P(A)$ for en delmængde $A \subseteq M$. Den er givet som

$$P(A) = \sum_{x \in A} p(x)$$

For diskrete fordelinger på en delmængde af \mathbb{N}_0 gælder, at sandsynlighedsmålet vokser i spring, samt at:

$$P(X = i) = F(i) - F(i - 1)$$

1.1.3 Tæthedsfunktion

også p.d.f. (probability density function). En funktion $p : I \subseteq \mathbb{R} \rightarrow [0, \infty[$, som opfylder:

$$\int_I p(x) dx = 1,$$

kaldes en sandsynlighedstæthed på I .

*Tak til Sebastian Ake Aaen for tilføjelser.

Angives med små bogstaver $f(x) = P(X = x)$ og vi bemærker, at sandsynligheden i et enkelt punkt for en kontinuert fordeling er 0. Kaldes undertiden sandsynlighedstæthed eller blot tæthed.

Der gælder for en stokastisk variabel X med tæthed $p(x)$, at sandsynligheden på A er

$$P(X \in A) = P(A) = \int_{-\infty}^{\infty} 1_A(x)p(x)dx,$$

hvor $1_A(x)$ er en indikatorfunktion, der antager værdien 1 på A .

1.1.4 Fordelingsfunktion

også c.d.f. (cumulative density function). Angives ved store bogstaver $F_X(x) = P(X < x)$ og er sandsynligheden for at den *kontinuerte* stokastiske variabel X antager en værdi mindre end x . Der gælder at $F'_X(x) = f_X(x)$, hvor $f_X(x)$ er tæthedsfunktionen.¹

Fordelingsfunktionen for en stokastisk variabel X med tæthed $p(x)$ er givet ved:

$$F(x) = P(X \leq x) = \int_{-\infty}^x p(y)dy$$

Bemærk endvidere, at $F(x) = 0,5$ betyder, at halvdelen af sandsynlighedsmassen ligger i mængden $] -\infty, x]$ og halvdelen ligger i $]x, \infty[$. Tallet x er således medianen. (Dette kan anvendes til at finde kvartiler, mm.)

1.1.5 Stokastisk variabel

Lad E være et udfaldsrum og P et sandsynligheds mål. En stokastisk variabel X er da en funktion fra udfaldsrummet ind i \mathbb{R} . Stokastiske variable angives typisk med kapitaler.

Vi kan skrive en tilfældig variation af X for enhver delmængde A af \mathbb{R} som

$$P_X(A) = P(X \in A) = P(\{e \in E | X(e) \in A\}), \quad (1)$$

hvor sidste skrivemåde sjældent benyttes.

Fler-dimensionel stokastisk variabel Ser vi på flere stokastiske på en gang kan disse sammenfattes til

$$X = (X_1, X_2, \dots, X_n), \quad (2)$$

hvor X_i er den en-dimensionale stokastiske variable. X kaldes for en n -dimensional stokastisk variabel.

¹Der gælder følgende for en fordelingsfunktion: Den er defineret på \mathbb{R} ind i $[0, 1]$, den er ikke-aftagende og $\lim_{n \rightarrow \infty} F(n) = 1$ og $\lim_{n \rightarrow \infty} F(-n) = 0$

1.2 Simultan og marginal sandsynlighed

1.2.1 Simultan fordeling

Fordelingen for den stokastiske variable $X = (X_1, X_2, \dots, X_n)$ på \mathbb{R}^n kaldes den *simultane fordeling* og er givet ved:

$$P_X(A) = P((X_1, X_2, \dots, X_n) \in A), \quad (3)$$

hvor $A \subseteq \mathbb{R}^n$

Simultan fordelingsfunktion For fordelingen af X kan vi skrive fordelingsfunktionen som:

$$F_X(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

Notation Når vi beskæftiger os med hændelser skriver vi den simultane sandsynlighed som $P(A \cap B)$. Med stokastiske variable noterer vi samme $P(X_1, x_2)$.

1.2.2 Marginal fordeling

Når de n -fordelinger af de enkelte koordinater X_i af $X = (X_1, X_2, \dots, X_n)$ kaldes de for *marginale fordelinger*.

Marginal fordelingsfunktion Betragt en to-dimensionel stokastisk variabel (X_1, X_2) , hvor X_i er koncentreret på mængden T_i for $i = 1, 2$. Lad være givet en sandsynlighedsfunktion $p : T_1 \times T_2 \rightarrow [0, 1]$ for den simultane fordeling af (X_1, X_2) . Da er de sandsynlighedsfunktionerne for fordeling af hhv. X_1 og X_2 , $p_1 : T_1 \rightarrow [0, 1]$ og $p_2 : T_2 \rightarrow [0, 1]$ givet ved (se evt. sætning 4.2.1 i Sørensen):

$$p_1 = \sum_{x_2 \in T_2} p(x_1, x_2) \text{ og } p_2 = \sum_{x_1 \in T_1} p(x_1, x_2) \quad (4)$$

1.3 Transformation

Antag at X er en stokastisk variabel på \mathbb{R} og t er en funktion fra \mathbb{R} ind i \mathbb{R} . Lad $Y = t(X)$. Da kan fordelingen for Y nu bestemmes ved $A \subseteq \mathbb{R}$:

$$P_Y(A) = P(Y \in A) = P(t(X) \in A) = P(X \in t^{-1}(A)) = P_X(t^{-1}(A)),$$

hvor $t^{-1}(A) = \{x \in \mathbb{R} | t(x) \in A\}$ kaldes originalmængden af A .

Transformationssætningen (se sætning 5.4.1 i Sørensen). Antag X er en kontinuert stokastisk variabel koncentreret på et interval I . Lad t være en kontinuert differentiabel funktion på et interval (a, b) hvor $t'(x) \neq 0$ på intervallet (a, b)

$$q(y) = \begin{cases} p(t^{-1}(y)) \left| \frac{d}{dy} t^{-1}(y) \right|, & y \in t(I) \\ 0 & \text{ellers} \end{cases}$$

Transformation af fordelingsfunktion Hvis t er strengt voksende er

$$F_Y(y) = \begin{cases} 0 & \text{hvis } y < v \\ F_X(t^{-1}(y)) & \text{hvis } y \in J \\ 1 & \text{hvis } y > h \end{cases}$$

Hvis t er strengt aftagende er

$$F_Y(y) = \begin{cases} 0 & \text{hvis } y < v \\ 1 - F_X(t^{-1}(y)) + P(X = t^{-1}(y)) & \text{hvis } y \in J \\ 1 & \text{hvis } y > h \end{cases}$$

2 Regneregler

2.1 Sandsynligheder

For et sandsynlighedsmål, hvor A og B betegner vilkårlige hændelser gælder og E er den hele mængde:

$$\begin{aligned} P(A) &= 1 - P(E \setminus A) \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

For betinget sandsynlighed gælder følgende:

$$\begin{aligned} P(A|B) &= \frac{P(A, B)}{P(B)} \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)}, \end{aligned}$$

hvor sidste resultat er Bayes' formel og gælder for positive sandsynligheder $P(A), P(B) > 0$. Endelig kan vi opskrive loven om total sandsynlighed (sætning 1.4.4. i Sørensen)

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

2.2 Uafhængighed

Vi betragter to *uafhængige* hændelser A og B . Da gælder, at den simultane sandsynlighed er lig produktet af de marginale sandsynligheder:

$$P(A \cap B) = P(A)P(B), \text{ derudover gælder at } P(A|B) = P(A)$$

Første resultat bruges ofte til at afgøre, hvorvidt to stokastiske variable er uafhængige. Derudover gælder der, at to uafhængige stokastiske variable nødvendigvis må være en produktmængde $T_1 \times T_2$, hvis dette er tilfældet, da er de ikke uafhængige.

Betinget uafhængighed Vi indfører nu en tredje hændelse C . A og B siges at være betinget uafhængige givet² C , hvis

$$P(A \cap B|C) = P(A|C)P(B|C)$$

2.3 Middelværdi

Middelværdien for en diskret stokastisk variabel X koncentreret på et interval³ I er givet ved:

$$E(X) = \sum_{x \in I} xp(x)$$

Middelværdien for en ikke-diskret stokastisk variabel Y er givet ved:

$$E(Y) = \int_I yp(y)dy$$

Vi lader X og Y være stokastiske variable. Der gælder generelt, at:

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) \text{ og} \\ E(aX + b) &= aE(X) + b \end{aligned} \tag{2.3.1}$$

Middelværdien for et produkt af to stokastiske variable er givet som:

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j P(X = x_i, Y = y_j)$$

Hvis X og Y desuden er uafhængige gælder, at

$$E(XY) = E(X)E(Y)$$

For en afbildning $t : \mathbb{R} \rightarrow \mathbb{R}$ gælder at

$$E(t(X)) = \sum_{x_i \in T} t(x_i)p(x_i)$$

2.3.1 Betinget middelværdi

Den betingede middelværdi $E(X|Y = y)$ for en fordelingsfunktion $P(Y = y) > 0$ er defineret som:

$$E(X|Y = y) = \sum_{i=1}^{\infty} x_i P(X = x_i|Y = y) = \sum_{i=1}^{\infty} x_i \frac{p(x_i, y)}{p(y)}$$

Et andet vigtigt resultat er:

$$E(X|X \in A) = \int_A x \frac{f_X(x)}{P(X \in A)} dx = \frac{\int_A x f_X(x) dx}{P(X \in A)}$$

²Kan også noteres

$$P(X_1, X_2|X_3) = P(X_1|X_3)P(X_2|X_3)$$

³Hvis vi i stedet betragtede en diskret stokastisk variabel på en uendelig tællelig mængde, da skulle vi være opmærksom på om middelværdien er veldefineret.

Derudover gælder, at:

$$E(X|X) = X,$$

og hvis X og Y er uafhængige gælder, at

$$E(X|Y) = E(X)$$

Bemærk, at (2.3.1) også gælder for betingede middelværdier. Se note om betingede middelværdier af A. Rahbek for mere.

Loven om itereret middelværdier Der gælder følgende:

$$E[Y] = E[E[Y|X]]$$

2.4 Varians og spredning

Lad være givet en stokastisk variabel X på en endelig mængde⁴. Da er variansen givet som:

$$\text{Var}(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2$$

Der gælder for summen af to stokastiske variable X og Y , at:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{cov}(X, Y),$$

hvor vi bemærker, at $\text{cov}(X, Y) = 0$, hvis X og Y er uafhængige, hvilket medfører at variansen af summen af vilkårligt mange uafhængige stokastiske variable blot er summen af de enkelte varianser.

Spredningen defineres som:

$$\text{spredning: } \sigma = \sqrt{\text{Var}(X)}$$

For lineære transformationer gælder desuden, at

$$\text{Var}(a + bX) = b^2 \cdot \text{Var}(X)$$

For en stokastisk variabel gælder, at $\text{Var}(X) = 0$, hvis og kun hvis der findes et $c \in \mathbb{R}$, så $P(X = c) = 1$.

2.4.1 Betinget varians

2.5 Kovarians

Kovariansen mellem to stokastiske variable X og Y er et udtryk for om de to stokastiske variable samvarierer. Lad a, b, c og d reelle tal, da gælder følgende:

$$\text{Cov}(X, Y) = E[(X - E[X]) \cdot (Y - E[Y])] = E(XY) - E(X)E(Y)$$

$$\text{Cov}(X, X) = \text{Var}(X)$$

$$\text{Cov}(a + bX, c + dY) = bd\text{Cov}(X, Y)$$

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

$$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$$

⁴Vi kan anvende samme udtryk for variansen defineret på en tællelig mængde $T = \{x_i | i \in \mathbb{N}\}$, dersom vi bemærker, at variansen skal være veldefineret:

$$\sum_{i=1}^{\infty} x_i^2 p(x_i) < \infty$$

- Hvis $Cov(X, Y) = 0$ er der ingen samvariation.
- Hvis $Cov(X, Y) < 0$ er der en 'tendens' til negativ samvariation. Eksempelvis, at store værdier af X og små værdier af Y er mere sandsynlige.
- Hvis $Cov(X, Y) > 0$ er der en 'tendens' til positiv samvariation.

Der gælder, at **hvis X og Y er uafhængige. Da er $Cov(X, Y) = 0$.** Bemærk, at kovariansen kan være 0, selvom de to variable ikke er uafhængige. At kovariansen er 0 er således en *tilstrækkelig* betingelse for uafhængig mellem de stokastiske variable, men ikke en *nødvendig* betingelse.

Dog gælder det for normalfordelingen – og kun denne – at er kovariansen 0, da er de to stokastiske variable også uafhængige.

Korrelation Korrelationen er et udtryk for samvariationen, standardiseret, så $\text{corr}(X, Y) \in [-1, 1]$. korrelationen mellem X og Y givet ved:

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

2.6 Diverse

Indikatorfunktion En indikator beskriver en afbildning, hvor om det gælder:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{hvis } x \in A \\ 0 & \text{hvis } x \notin A \end{cases}$$

Der gælder, at

$$\int_{-\infty}^{\infty} \mathbf{1}_A(x)f(x)dx = \int_{x \in A} f(x)dx$$

Binomialkoefficient er givet som:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n^{(k)}}{k!},$$

hvor $n^{(k)} = n \cdot (n-1) \cdot \dots \cdot (n-k+1)$.

3 Diskrete fordelinger

En stokastisk variabel siges at være diskret, hvis den antager et endeligt antal værdier eller et uendeligt antal tællelige værdier. For diskrete fordelinger gælder at for en stokastisk variabel X at sandsynligheden for hændelsen A er summen af punktsandsynlighederne hørende til A :

$$P(A) = \sum_{x \in A} p(x)$$

3.1 Bernoullifordeling

Bernoullifordelingen er den eneste fordeling af en stokastisk variabel med to-punkts udfaldsrum og kan betragtes som et særtilfælde af binomialfordelingen for $n = 1$. Vi skriver $X \sim \text{Ber}(p)$, hvor sandsynlighedsparameteren $p \in]0, 1[$. Der gælder, at:

$$P(X = 1) = p \text{ og } P(X = 0) = 1 - p,$$

Den tilhørende sandsynlighedsfunktion er givet som:

$$p(x) = \begin{cases} 1 - p & \text{hvis } x = 0 \\ p & \text{hvis } x = 1 \end{cases} = p^x(1 - p)^{1-x} \text{ for } x \in \{0, 1\}$$

Middelværdi $E(X) = p$

Varians $V(X) = p(1 - p)$

3.2 Binomialfordeling

Binomialfordelingen er en fordeling på $\{0, 1, \dots, n\}$ udfald. Lad X_1, X_2, \dots, X_n være uafhængige stokastiske variable på mængden $\{0, 1\}$, der alle er bernoullifordelt, jf. ovenfor, med samme sandsynlighedsparameter $p \in]0, 1[$.

Fordelingen af summen $S = X_1 + X_2 + \dots + X_n$ kaldes binomialfordelingen med antalsparameter n og sandsynlighedsparameter p (definition 3.2.1). Vi skriver $X \sim \text{Bin}(n, p)$

Sandsynlighedsfunktionen er, jf. sætning 3.2.3:

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \text{ for } x \in \{0, 1, 2, \dots, n\}$$

Middelværdi $E(X) = np$

Varians $V(X) = np(1 - p)$

3.2.1 Sum af to binomialfordelinger

Lad S_1, S_2 være uafhængige stokastiske variable, der er binomialfordelte med sandsynlighedsparameter p og antalsparameter n_i (for $i = 1, 2$). Da er $S = S_1 + S_2$ binomialfordelt med sandsynlighedsparameter p og antalsparameter $n = n_1 + n_2$

3.3 Polynomialfordeling

Polynomialfordelingen er en fordeling på $\{0, 1, \dots, n\}^k$ udfald og er således en generalisering af binomialfordelingen, hvis man har mere end to udfald. Sandsynlighedsfunktionen er givet som:

$$p(x) = \binom{n}{s_1 \ s_2 \ \dots \ s_k} \cdot p_1^{x_1} \cdot \dots \cdot p_k^{x_k}$$

hvor polynomialkoefficienten er givet ved:

$$\binom{n}{s_1 \ s_2 \ \dots \ s_k} = \frac{n!}{s_1! \cdot s_2! \cdot \dots \cdot s_k!}$$

Middelværdi $E(X) = np_i$
Varians $V(X) = np_i(1 - p_i)$

3.4 Poissonfordeling

Er en fordeling på \mathbb{N}_0 og er i modsætning til de tre ovenstående således en fordeling på et tælleligt, men ikke endeligt udfaldsrum. Vi skriver $X \sim \text{Poi}(\lambda)$, $\lambda > 0$

Sandsynlighedsfunktionen til en poissonfordelinger er

$$p(x) = P(X = x) = \frac{\lambda^x}{x!} \exp(-\lambda), \quad x \in \mathbb{N}_0$$

hvor $p(x) \in]0, 1]$

Middelværdi $E(X) = \lambda$
Varians $V(X) = \lambda$.

3.4.1 Sum af poissonfordelinger

Jf. sætning 4.3.4 i Sørensen gælder, at for X_1, \dots, X_n uafhængige stokastiske variable, der er poissonfordelte med parametre $\lambda_1, \dots, \lambda_n$, at $X_1 + \dots + X_n$ er poissonfordelt med parameter $\lambda_1 + \dots + \lambda_n$.

3.5 Geometrisk fordeling

Kaldes undertiden en diskret ventetidsfordeling. Har sandsynlighedsfunktionen

$$p(x) = (1 - \theta)^x \theta, \quad x \in \mathbb{N}_0,$$

hvor $\theta \in]0, 1[$ er sandsynlighedsparemeteren.

Middelværdi $E(X) = \frac{1-\theta}{\theta}$
Varians $V(X) = \frac{1-\theta}{\theta^2}$.

4 Kontinuerte fordelinger

En stokastisk variabel siges at være kontinuert, hvis den antager værdier på \mathbb{R} eller intervaller. For kontinuerte sandsynligheder, hvor en stokastisk variabel X er koncentreret på mængden I , og hvor $A \in I$ gælder at sandsynligheden for hændelsen A er givet som:

$$P(X) = \int_I 1_A(x)p(x)dx = \int_A p(x)dx$$

For en kontinuert fordeling gælder, at sandsynligheden i et punkt a er 0 (man kan dog betragte sandsynligheden på stykket a til $a + \delta$ for et lille $\delta > 0$.)

4.1 Uniform fordeling (ligefordeling)

Fordelingen ($X \sim \text{uni}(a, b)$) på intervallet a til b , hvor $b > a$ har følgende tæthed:

$$p(x) = 1_{(a,b)}(x) \frac{1}{b-a},$$

hvoraf følger at fordelingsfunktionen er givet ved:

$$F(x) = \int_{-\infty}^x \frac{1_{(a,b)}(y)}{b-a} dy = \begin{cases} 0 & \text{for } x \leq a \\ \frac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x \geq b \end{cases}$$

Middelværdi $E(X) = \frac{a+b}{2}$

Varians $V(X) = \frac{(b-a)^2}{12}$

4.2 Eksponentialfordeling

Eksponentialfordelingen med $\lambda > 0$ (også skrevet $X \sim \text{Exp}(\lambda)$) har tæthedsfunktion en

$$p(x) = \lambda \exp(-\lambda x), x > 0$$

Fordelingsfunktionen er

$$F(x) \begin{cases} 1 - \exp(-\lambda x), & \text{hvis } x > 0 \\ 0, & \text{hvis } x \leq 0 \end{cases}$$

Middelværdi $E(X) = \frac{1}{\lambda}$

Varians $V(X) = \frac{1}{\lambda^2}$

4.3 Normalfordeling

Denne fordeling har følgende tæthed:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Den tilhørende fordelingsfunktion er:

$$\Phi(x) = F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx$$

Middelværdi $E(X) = \mu$

Varians $V(X) = \sigma^2$

Standard normalfordelingen, ofte noteret Z , har middelværdi $\mu = 0$ og varians $\sigma^2 = 1$ og er symmetrisk om 2. akse, så $\phi(x) = \phi(-x)$. Endvidere er $E(X^3) = 0$ og kurtosis $E(X^4) = 3 = \kappa$

Sum af variable Vi betragter X_1, X_2, \dots, X_n indbyrdes uafhængige normalfordelte stokastiske variable. Da gælder for en linearkombination $S = a_1X_1 + a_2X_2 + \dots + a_nX_n$, at:

$$E(S) = a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n) \text{ og}$$

$$V(S) = a_1^2V(X_1) + a_2^2V(X_2) + \dots + a_n^2V(X_n)$$

Bemærk, at første resultat er generelt, mens sidste resultat kun gælder for *uafhængige* variable. Er $Cov(X, Y) \neq 0$ da må man bruge, at

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$$

Uafhængighed Der gælder, at *hvis, og kun hvis* kovariansen for en multivariat normalfordeling (herunder bivariat) er 0, er de stokastiske variable uafhængige. For multivariate normalfordelinger gælder dette begge veje.

Positions- og skalaparameter Den generelle normalfordeling kan opskrives ved at indføre positionsparameter $\mu \in \mathbb{R}$ og skalaparameter $\sigma > 0$. Hvis X er standard normalfordelt, er

$$Y = \mu + \sigma X$$

normalfordelt med middelværdi μ og varians σ^2 . Dette er et resultat af det generelle udtryk for positions- og skalaparametre, Eksempel 5.4.2 i Sørensen.

5 Statistik

5.1 Deskriptiv statistik

5.1.1 Momenter

Det teoretiske moment af orden k er defineret som:

$$E[Y^k]$$

Første moment er middelværdien. For simple fordelinger finder vi ofte, at den bedste estimator for parameteren (se likelihood-funktion) er middelværdien. Dette tjener bl.a. til argument for, hvorfor middelværdien er et relevant begreb. Kender vi middelværdien kan vi definere centrale momenter:

$$E([Y - E(Y)]^k)$$

Variansen er det andet centrale moment. Variansen er et udtryk for sandsynligheden for at observere stor variation fra middelværdien og et udtryk for bredden af fordelingen.

Tredje moment er *skewness* eller skævhed. Hvis fordelingen er symmetrisk er skævheden 0 (normalfordeling). En negativ skævhed indikerer at fordelingen er venstreskæv og vice versa.

Endelig findes fjerde moment, som kaldes kurtosis, der er et mål for sandsynlighedsmassen i fordelings 'haler'. Normalfordelingen har en kurtosis på 3. En højere kurtosis betyder mere sandsynlighedsmasse i halerne.

5.1.2 Fraktiler

Fraktiler er defineret som den værdi $q_Y(\alpha)$, så den andel α af sandsynlighedsmassen ligger under denne værdi:

$$P(Y \leq q_Y(\alpha) = F_Y(q_Y(\alpha)) = \alpha,$$

hvor F_Y er c.d.f.'en. Altså er fraktil-funktionen den inverse af c.d.f.'en: $q_Y(\alpha) = F_Y^{-1}(\alpha)$

5.2 Inferens-baseret statistik

Det bør bemærkes, at estimatet (et tal, der afhænger af data) $\hat{\theta}(y_1, \dots, y_n)$ er en realisation af estimatoren (en stokastisk variabel, der afhænger af en række andre stokastiske variable) $\hat{\theta}(Y_1, \dots, Y_n)$.

Antagelse I vores arbejde, antager vi, at modellen er *korrekt specificeret*, dvs. at vores antagelser (3.1) holder for en given sand værdi θ_0 , så de stokastiske variable er karakteriseret ved en og samme tæthedsfunktion.

5.2.1 Egenskaber på en endelig mængde

Unbiased Vi siger at vores estimator er unbiased, hvis $E(\hat{\theta}_n) = \theta_0$, altså at fordelingen af vores estimator er allokeret, så den er lig den sande værdi.

Vi kan bruge variansen af vores estimator som udtryk for usikkerheden ved vores estimator. Denne vil være givet som en funktion af den sande værdi, men da den sande værdi er os ukendt estimerer vi variansen ved at indsætte vores estimator.

Standardfejl Enheden af variansen er givet i samme enhed som θ . Derfor anvendes standardfejl, givet som:

$$se(\hat{\theta}_n) = \sqrt{V(\hat{\theta}_n)}$$

5.2.2 Asymptotiske egenskaber

Det er muligt at vise, at normalfordelingen ikke er automatisk *unbiased*. Når $n \rightarrow \infty$ ser vi blandt andre resultater, at normalfordeling nærer sig unbiased. I praksis har vi ikke uendelige datamængder, men vi antager, at resultater tilnærmelsesvis gælder for store værdier af n .

Teorem 4.1 Under antagelse om at de stokastiske variable kan karakteriseres med ved en tæthedsfunktion (p.d.f) eller en \square (p.f.), at de er identisk fordelt, så $\theta_1 = \theta_2 = \theta$ og at de er indbyrdes uafhængige (i.i.d antagelse, se p. 50 i Nielsen) gælder følgende om Maksimum Likelihood Estimatoren (MLE).

1. MLE er konsistent⁵: $\hat{\theta}_n \xrightarrow{p} \theta_0$ for $n \rightarrow \infty$. For et vilkårligt stort n vil estimatoren således nærme sig den sande værdi. Vi taler om *konvergens i sandsynlighed*.

⁵Dette kan også noteres som $\forall \delta > 0 : P(|\hat{\theta}_n - \theta_0| > \delta) \rightarrow 0$

2. MLE er asymptotisk normalfordelt for $n \rightarrow \infty$

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Omega_\theta).$$

Dette er altså et udtryk for at for et vilkårligt stort n opfører fordelingen af estimatoren sig som en normalfordelt stokastisk variabel. Vi taler om *konvergens i fordeling*. Resultatet kan opskrives til:

$$\hat{\theta}_n \stackrel{a}{\sim} N(\theta_0, n^{-1}\Omega_\theta)$$

3. Den asymptotiske varians er givet ved $\Omega_\theta = I(\theta_0)^{-1}$, hvor informationsmatricen er

$$I(\theta_0) = -E(H_i(\theta_0)),$$

og

$$H_i(\theta_0) = \frac{\partial \log l(\theta|Y_i)}{\partial \theta \partial \theta},$$

evalueret for en bestemt værdi $\theta = \theta_0$.

4. MLE er asymptotisk efficient: alle andre konsistente og asymptotiske normale estimators har en asymptotisk varians, der er større end eller lig $I(\theta)^{-1}$

Vi bemærker, at det er den sande værdi θ_0 , der indgår i grænseresultaterne fordi egenskaberne for estimatoren er udledt under antagelsen om *korrekt specifikation*. Da θ_0 i praksis er ukendt indsættes estimatoren $\hat{\beta}_n$, jf. rettevejledning til eksamenssæt til ordinær eksamen vinter 2015.

5.2.3 Konfidensinterval

Vi har ovenfor noteret os, at en af den approksimerede fordeling af $\hat{\theta}_n$ er:

$$\hat{\theta}_n \stackrel{a}{\sim} N(\theta_0, n^{-1}\Omega_\theta), \quad (5.2.3)$$

hvilket kan anvendes til at finde intervaller, hvori det er sandsynligt at den sande værdi af parameteren θ_0 ligger. Vi kan isolere i ovenstående udtryk og finde:

$$\frac{\hat{\theta}_n - \theta_0}{\text{se}(\hat{\theta}_n)} \stackrel{a}{\sim} N(0, 1)$$

For en standard normalfordeling gælder det, at 95 pct. af sandsynlighedsmassen ligger indenfor

$$P(\Phi^{-1}(0,025) \leq Z \leq \Phi^{-1}(0,975)) = P(-1,96 \leq Z \leq 1,96) = 0,95,$$

hvor $Z = \frac{\hat{\theta}_n - \theta_0}{\text{se}(\hat{\theta}_n)}$. Indsætter vi dette og isolerer finder vi, at:

$$\{\underline{\theta} \leq \theta_0 \leq \bar{\theta}\} = \{\hat{\theta}_n - 1,96 \cdot \text{se}(\hat{\theta}_n) \leq \theta_0 \leq \hat{\theta}_n + 1,96 \cdot \text{se}(\hat{\theta}_n)\}$$

Der er således 95 pct. sandsynlighed for at intervallet rummer den sande værdi θ_0 , eller omvendt 5 pct. sandsynlighed for at intervallet ikke rummer den sande værdi.

5.2.4 P-værdi

P-værdien er et udtryk for sandsynligheden for at få en realisation, der er mere ekstrem end den numeriske værdi af vores test-værdi. Jo tættere estimatet er på vores (sande) værdi, desto mindre er z-værdien for en z-test. Vi kan udregne den tilhørende p-værdi som:

$$p = P(X \leq x) + P(X \geq x) = 2P(X \leq x) = 2 \cdot (1 - P(X \geq x)) = 2 \cdot (1 - \Phi(x)),$$

hvor $\Phi(x)$ er c.d.f. for normalfordelingen.

Hvis p-værdien er mindre end vores signifikansniveau, $p < \alpha$ forkastes nulhypotesen.

5.3 Hypotese-test

For at undersøge, hvorvidt den sande værdi af en given parameter antager en bestemt værdi, anvender vi hypotese-test. Til dette formål opstiller vi en nulhypotese, samt en alternativ hypotese:

$$\mathcal{H}_0 : \theta_0 = a \text{ og } \mathcal{H}_A : \theta_0 \neq a$$

5.3.1 Z-test (Wald-test)

Denne test er baseret på afstanden mellem estimatoren og den foreslåede værdi a samt fordelingen givet i ligning (5.2.3).

Vi kan omdanne dette udtryk for fordelingen til (se p. 101 i Nielsen for udregninger)

$$z_n(\theta_0 = a) = \frac{\hat{\theta}_n - a}{\text{se}(\hat{\theta}_n)} \stackrel{a}{\sim} N(0, 1)$$

Hvis nulhypotesen er sand, altså $\theta_0 = a$, da vil udtrykket konvergere i fordeling mod en standard normalfordeling, men er det ikke sandt vil den gå mod $\pm\infty$.

Hvis $z_n(\theta_0 = a)$ ligger i den kritiske mængde af vores fordeling (udenfor konfidensintervallet, se ovenfor) vil vi afvise \mathcal{H}_0 og argumentere for at \mathcal{H}_A er mere sandsynlig, og vice versa.

Foruden Z-test, der er mest anvendt i eksamensopgaver, findes også kvadreret Z-test og LR-test. Her henvises til anden litteratur.

Appendix

5.4 Regneregler for summer

$$\begin{aligned}\sum_n f(n) \pm \sum_n g(k) &= \sum_n (f(n) \pm g(k)) \\ \sum_n (f(n) \cdot c) &= c \cdot \sum_n f(n) \\ \sum_n (f(n) + c) &= c \cdot n + \sum_n f(n) \\ \sum_n f(n) \cdot \sum_k g(k) &= \sum_n \sum_k f(n)g(k) \\ \sum_n \sum_k f(n)g(k) &= \sum_k \sum_n f(n)g(k)\end{aligned}$$

5.4.1 Regel til inversion af matricer

$$A' = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$