

Sammendrag

MAT12

Christian Braathen

Innhold

	Side
1 Viktig viktig	1
2 Typiske antakelser og forutsetninger	4
3 Sammenheng mellom flere variabler	5
3.1 Mer utfyllende om særdeles relevante sammenhenger	6
4 Diskrete fordelinger	8
4.1 Binomisk fordeling	9
4.2 Poissonfordelingen	9
4.3 Geometrisk fordeling	10
4.4 Ikke-navngitt diskret	10
5 Kontinuerlige fordelinger	11
5.1 Normalfordeling	12
5.2 Gammafunksjonen	12
5.3 Kvikvadratfordelingen	12
5.4 T-fordeling	12
6 Statistisk inferens	13
6.1 Forventning og varians	13
6.2 Diverse om fordelinger	14
6.3 Max og min av andre variabler	15
6.3.1 Stokastisk variabel er minimum	15
6.3.2 Estimator blir minimum	16
6.3.3 Vise at en transformasjon blir ...fordelt	16
6.4 Sentralgrensesetningen	16
6.5 Transformasjon	18
6.6 Momentgenererende funksjon	19
6.7 Generelt om estimasjon	19
6.8 Momentestimator til θ (method of moments)	19
6.9 Maximum likelihood-estimator til θ	20
6.10 Asymptotisk approksimasjon	20
6.11 Konfidensintervall	21
6.11.1 Utledning av et CI	21
6.12 Hypotesetesting	22

6.12.1 Eksempel	22
6.13 Styrke til en hypotesetest	23
6.14 Likelihood ratio test	23
6.15 Betingede sannsynligheter	24
6.16 Bayesiansk inferens	24
6.17 Regresjon	26
7 Diverse	27

Del 1

Viktig viktig

- Første gangen jeg skal regne på et uekte integral, si, "Jeg behandler det uekte integralet som at det er ekte: det vil jeg gjøre gjennom resten av eksamensbesvarelsen."
- Under variabelskifter, husk å få med de nye integralgrensene.
- Pass på α_α versus $\alpha_{1-\alpha}$
- For asymptotisk varians, pass på å få med at det er opphøyd i -1 . Det har blitt gjort rett hvis n kommer i nevneren.
- Virker noe fremmed, er det smart å tegne det vi studerer.
- Utled fordelingene med $F_X(x) = P(X \leq x)$. De fleste vis at-oppgaver starter gjerne med denne, hvor X er den ukjente.
- Der en stokastisk variabel er en sum av en annen, og at variablene i summen er uavhengige av hverandre, utled fordelingen med MGF'er.
- Har vi uendelig i øvre integralgrense, faktoreriser så eksponenten til e er $-x(\dots)$. Da går den mot null.
- Si at vi har fått vite tetthetsfunksjon og \bar{x} , og at vi skal estimere θ . Finn uttrykket til $E[X]$ og finn den θ som gjør at likheten holder.
- $\sum_{x=1}^{\infty} \dots^{x-1} = \sum_{x=0}^{\infty} \dots^x$
- Uendelig geometrisk rekke: $\sum_{x=0}^{\infty} ay^x = \frac{a}{1-y}$
- $p = \frac{\text{bar}x}{\lambda}$. Vi har estimert λ allerede og kan sette inn. Derfor får vi $\hat{p} = \frac{\bar{x}}{\hat{\lambda}_{ML}}$

- Spør den om fordeling vil den gjerne ha $f_X(x) = \begin{cases} \text{noe} & , x \in [a; b] \\ 0 & , \text{ellers} \end{cases}$.
- viktig å få satt inn hat'ene på riktig plass. Se hva oppgaven spør etter!
- For SGS, vær oppmerksom og gjerne kritisk på om man faktisk kan anta uavhengighet eller ikke!
- Se over om det er noe info jeg kan dytte inn. Ha et ark med sårne, perhaps?
- Husk å argumentere hvorfor man skal bruke ensidet eller tosidet test. Referer til Jarle Møens argument om at ensidet brukes kun hvis det er umulig å oppnå disse verdiene eller at det ikke er beslutningsrelevant.
- På oppgaver hvor en stokastisk variabel er en rank av en annen, prøv *både* \leq og $>$ samt \geq og $<$. Får jeg ikke det jeg søker, prøv bare å snu ulikheten i $F_{X_{(1)}} = P(\min(X_i) \leq x)$ isteden.
- Har vi flere estimatorer og skal sammenligne disse, se på forventning og varians + tegn opp dartboard'et for å illustrere det tydelig. Hvis bias er en konstant, si at man kan finjustere siktet ved å trekke fra denne konstanten.
- Deriverer fra F til f . Husk kjerneregelen!
- Må bruke betinget variabel hvis, for eksempel, Y er binomisk og X avgjør hvor mange Bernoullivariabler som er i Y . Da må vi skrive $(Y|X = x)$.
- $P(X^2 \leq Y) = P(-\sqrt{Y} \leq X \leq \sqrt{Y}) = P(X \leq \sqrt{Y}) - P(X \leq -\sqrt{Y})$
- Skal man avgjøre om fordelingen er en god modell for dataene, studer noe eller alle av følgende: 1) $E[X]$ med \bar{x} , 2) $Var[X]$ med s^2 , og/eller 3) om domenet er realistisk til caset vi studerer (for eksempel om den kan ta negative verdier).
- Skal vi motivere valg av estimasjonsmetode, må vi finne den som er minst biased i de forskjellige parameterne. Vi må altså utlede så mange ligninger som vi har åparametre i fordelingene.
- Skal vi gjennomføre en hypotesetest om observasjonene er ...fordelt, så er nøkkelen å finne for hvilke parameterverdier sannsynlighetstettheten blir lik den fordelingen vi lurer på. Sett H_0 lik den fordelingen.
- Bayes lov: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$
- "Om jeg skriver opp et integral som dere vet er en sannsynlighetsfordeling, så trenger dere ikke regne den ut siden dere vet at den blir 1".

- Geometriske rekker: se om det er noen ledd for en i -verdi som eliminerer noen ledd for $i - 1$ -verdien slik at vi ender opp med noe fra første og siste ledd.
- Husk å se etter avtakende eller voksende funksjon i transformasjoner.
- En stikkprøve er en uavhengig sekvens av stokastiske variabler.
- $(a + b)^m = \sum_{x=0}^m \binom{m}{x} a^x b^{m-x}$
- L'Hôpital: $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}$
- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- Taylor's formel: $f(x) = \sum_{i=0}^{\infty} f^{(i)}(a) \frac{(x-a)^i}{i!}$
- Generelt: $E[g(X)] \neq g(E[X])$

Del 2

Typiske antakelser og forutsetninger

- i.i.d.: uavhengige og identisk fordelte stokastiske variabler. Uavhengighet betyr at én hendelse ikke påvirker sannsynligheten for at den andre skjer.
- Husk å skrive opp hvilke verdier (domenene) de forskjellige variablene kan ta!

Del 3

Sammenheng mellom flere variabler

- $\text{bin}(n, p), n = 1 \Rightarrow \text{bernoulli}(p)$
- $\Gamma(\alpha, \theta), \alpha = 1 \Rightarrow \exp(\theta)$
- $\Gamma(\alpha, \theta), \alpha = \frac{v}{2}, \theta = 2 \Rightarrow \chi^2(v)$
- $\text{Weibull}(\alpha, \beta), \alpha = 1 \Rightarrow \exp(\beta)$

Videre har vi:

- F_X^{-1} gir $U(0, 1)$.
- $X \sim N(\mu, \sigma^2) \Rightarrow e^X = Y \sim \text{lognormal}(\mu, \sigma^2)$
- $X \sim N(0, 1) \Rightarrow X^2 = Y \sim \chi^2(1)$
- $\forall i. \quad X_i \sim \text{Poi}(\mu_i) \Rightarrow \Sigma X_i = Y \sim \text{Poi}(\Sigma \mu_i)$
- $\forall i. \quad X_i \sim \Gamma(\alpha_i, \beta) \Rightarrow \Sigma X_i = Y \sim \Gamma(\Sigma \alpha_i, \beta)$
- $\forall i. \quad X_i \sim \chi^2(v_i) \Rightarrow \Sigma X_i = Y \sim \chi^2(\Sigma v_i)$
- $\forall i. \quad X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow \Sigma X_i = Y \sim N(\Sigma \mu_i, \Sigma \sigma_i^2)$
- $\forall i. \quad X_i \sim \text{bernoulli}(p) \Rightarrow \Sigma X_i = Y \sim \text{bin}(n, p)$
- $\forall i. \quad X_i \sim \exp(\theta) \Rightarrow \Sigma X_i = Y \sim \Gamma(\alpha, \theta)$
- $\forall i. \quad X_i \sim \exp(\theta_i) \Rightarrow \min(X_i) = Y \sim \exp(\Sigma \theta_i)$ hvis uavhengige.
- $X \sim N(0, 1), U \sim \chi^2(v) \Rightarrow \frac{X}{\sqrt{U/v}} = Y \sim t(v)$

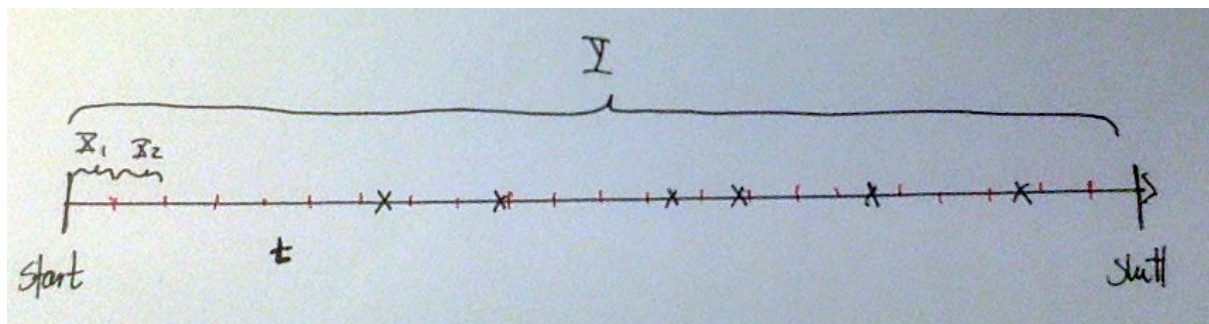
Grenser. Forutsetter uavhengige og identisk fordelte.

- $X \sim \text{bin}(n, p) \Rightarrow X \sim \text{Poi}(np)$ hvis n er stor og np er liten.
- $X \sim \text{Poi}(n, p) \Rightarrow X \sim N(\lambda, \lambda)$ hvis n er stor.
- $X \sim \text{bin}(n, p) \Rightarrow X \sim N(np, np(1-p))$ hvis n er stor.
- $X \sim \Gamma(\alpha, \theta) \Rightarrow X \sim N(\alpha\theta, \alpha\theta^2)$ hvis $\alpha \gg \beta$.
- $X \sim t(v) \Rightarrow X \sim N(\mu, s^2)$ hvis v er stor.

Og en siste, aktuelt for betingede sannsynligheter:

- $(X|N) \sim \text{bin}(N, p), N \sim \text{Poi}(\mu) \Rightarrow X \sim \text{Poi}(\mu p)$

3.1 Mer utfyllende om særdeles relevante sammenhenger



Figur 3.1

- Vi har et intervall. Vi stykker det opp så man får max én begivenhet i hvert intervall (i.e. $P(\text{flere begivenheter i et intervall})=0$)
- Hver bit er nå en Bernoullivariabel med $X_i = \begin{cases} 1 & p \text{ er sannsynligheten for å få en} \\ 0 & \text{begivenhet i én bit.} \end{cases}$
- Summen av alle Bernoullivariablene er binomisk fordelt. $Y = \sum_{i=1}^n X_i \sim \text{bin}(n, p)$
- Men hva er p egentlig? $P(\text{begivenhet})$ i én bit, ja. Men det kan vi finne ved å ta forventet antall begivenheter (λ) delt på altall biter (n) også. Så $p = \frac{\lambda}{n}$.
- Når $n \rightarrow \infty$, vil $Y \rightarrow \text{Poi}(\lambda)$. Som sagt er λ forventet antall begivenheter i hele intervallet.
- Det kan også være interessant å se på tiden til første begivenhet. $P(W \leq t)$ finner sannsynligheten for at tiden til første begivenhet er mindre enn t . Tid til første begivenhet er

eksponentialfordelt. Problemet er at vi har kanskje ikke begivenheter for t . Så vi flipper og studerer $1 - P(W > t)$ istedet. Kan bruke denne til å studere tid *mellom* to begivenheter også.

- Om vi heller ønsker å studere tid til α begivenheter inntreffer, kan vi bruke gammafordelingen. Kan bruke denne til å studere tid *mellom* begivenheter også (for eksempel fra tid fra begivenhet 3 til begivenhet $3 + \alpha$).

Del 4

Diskrete fordelinger

Hvordan skrive de ulike fordelingene:

- Bernoulli: $X \sim \text{Bernoulli}(p)$, eventuelt $X \sim \text{bin}(1, p)$
- Binomisk: $X \sim \text{bin}(n, p)$
- Geometrisk: $X \sim \text{Geo}(p)$
- Hypergeometrisk
- Negative binomisk
- Poisson $X \sim \text{Poi}(\lambda)$
- Uniform: $X \sim U[a, b]$

Kort beskrivelse:

- Bernoulli: takes value 1 with probability p and value 0 with probability $1 - p$.
- Binomisk: describes the number of successes in a series of independent Bernoulli experiments all with the same probability of success.
- Geometrisk: a discrete distribution which describes the number of attempts needed to get the first success in a series of independent Bernoulli trials, or alternatively only the number of losses before the first success (i.e. one less)
- Hypergeometrisk: which describes the number of successes in the first m of a series of n consecutive Yes/No experiments, if the total number of successes is known. This distribution arises when there is no replacement.

- Negative binomisk (*Pascal* is a special case of this) distribution as a generalization of the geometric distribution to the n th success.
- Poisson describes the number of successes in a series of independent Yes/No experiments with different success probabilities.
- Diskret uniform: This is the theoretical distribution model for a balanced coin, an unbiased die, a casino roulette, or the first card of a well-shuffled deck.

$$h(x) = \sum_{\theta} g(\theta) f(x|\theta) \quad (4.1)$$

Dette er bare et helt vanlig sannsynlighetstre der θ er i første gren og x er i andre gren. En variant av denne blir dermed (fra V16): $P(Y = y) = \sum_{x=y}^{\infty} P(Y = y|X = x) \cdot P(X = x)$

4.1 Binomisk fordeling

Tolkning: *antall gunstige utfall av n uavhengige forsøk.*

Merk at hvis $(n = x)_{X=x}$, så er $X = x$ en (kjent) stokastisk variabel. Siden Y summeres over $n = x$, vil dens verdi derfor avhenge av X . Vi må derfor betinge Y på X : $Y = (Y|X = x)$.

4.2 Poissonfordelingen

Tolkning: *antall begivenheter i et tidsintervall.*

Spesielt nyttig for å modellere hendelser, især sjeldne hendelser.

$$\lambda = \frac{\text{antall hendelser } k}{\text{antall målinger i dataene } n}$$

Summen av uavhengige Poissonfordelte variabler er Poissonfordelt.

Min tolkning fra eksamensjobbingen: La $\lambda = np$, der n er antall oppstykkinger av et intervall og p er sannsynligheten for suksess i hvert intervall. Da vil λ være det antallet suksesser vi forventer over hele intervallet vi studerer. \therefore sannsynligheten for suksess i en oppstykkning er forventet antall suksesser delt på antall oppstykkinger, $\frac{\lambda}{n}$. Hvorfor stykker vi opp det store intervallet? For å få at sannsynligheten for mer enn én hendelse i hvert av de små intervallene skal være 0.

Tolkning av λ : gjennomsnittlig/forventet antall (suksesser) for et gitt intervall. For eksempel, om vi studerer toårsintervaller og finner at λ er 2.4, for så å redusere intervallengden til ettårig, da blir den nye λ 1.2.

$$\left(1 + \frac{x}{n}\right)^n = e^x \text{ når } n \rightarrow \infty \quad (4.2)$$

4.3 Geometrisk fordeling

Hvis X er sekunder til første transaksjon, $X \sim Geo(p)$, da er p sannsynligheten for at en transaksjon skjer i løpet av ett sekund.

4.4 Ikke-navngitt diskret

$$\text{Vi har } X = \begin{cases} -1 & p \\ 0 & 1 - 2p \\ 1 & p \end{cases}$$

Da er $L(p) = p^{y_1} \cdot (1 - 2p)^{y_2} \cdot p^{y_3}$, hvor y_i representerer antall ganger x_i inntreffer.

Del 5

Kontinuerlige fordelinger

Hvordan skrive de ulike fordelingene:

- Beta:
- Chi-square: $X \sim \chi^2(r)$
- Exponential: $X \sim \text{Exp}(\lambda)$
- Gamma: $X \sim \text{Gamma}(\alpha, \frac{1}{\theta})$
- Normal: $X \sim N(\mu, \sigma^2)$
- Uniform: $X \sim U[a, b]$

Kort beskrivelse av dem:

- Beta: on $[0,1]$, a family of two-parameter distributions with one mode, of which the uniform distribution is a special case, and which is useful in estimating success probabilities.
- Chi-square: which is the sum of the squares of n independent Gaussian random variables. It is a special case of the Gamma distribution
- Exponential: **describes the time between consecutive rare random events in a process with no memory.**
- Gamma: **describes the time until n consecutive rare random events occur in a process with no memory.**
- Normal: It is ubiquitous in nature and statistics due to the central limit theorem: every variable that can be modelled as a sum of many small independent, identically distributed variables with finite mean and variance is approximately normal.

- Continuous uniform: on $[a,b]$, where all points in a finite interval are equally likely.
- Lognormal: describing variables which can be modelled as the product of many small independent positive variables
- Weibull: is used to model the lifetime of technical devices and is used to describe the particle size distribution of particles generated by grinding, milling and crushing operations.
- Pareto: or power law distribution, used in the analysis of financial data and critical behavior.

5.1 Normalfordeling

Sum av normalfordelte variabler er normalfordelt.

5.2 Gammafunksjonen

$$\int_S y^{\alpha-1} \cdot \exp[-y] dy = \Gamma(\alpha) \quad (5.1)$$

5.3 Kvikvadratfordelingen

$$Z^2 = \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1) \quad (5.2)$$

Hvis $Y = \sum_{i=1}^n Z_i^2$, hvor $Z_i \sim N(0, 1)$, da er $Z_i^2 \sim \chi^2(1)$ og $Y \sim \chi^2(n)$.

5.4 T-fordeling

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (5.3)$$

Del 6

Statistisk inferens

"Alt i statistisk inferens handler om å sammenligne datamodeller."

6.1 Forventning og varians

Generelt gjelder *ikke* at $E[g(X)] = g(E[X])$.

$$\text{Var}[X + a] = \text{Var}[X] \quad (6.1)$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}} = E[(X - \mu_X)(Y - \mu_Y)] \quad (6.2)$$

$$\text{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y] \quad (6.3)$$

$$E[XY] = E[Z] = \int_S z \cdot f(z) dz = \int_{S_X} \int_{S_Y} xy \cdot f(x, y) dy dx \quad (6.4)$$

$$E[X^m] = \int_S x^m \cdot f(x; \theta) dx \quad (6.5)$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 \quad (6.6)$$

Summen av to normalfordelte variabler ($Z = X + Y$) er normalfordelt, og følgende gjelder:

$$\begin{aligned} E[Z] &= \mu_X + \mu_Y \\ \text{Var}[Z] &= \sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X, Y) \\ Z &\sim N(\mu_X + \mu_Y; \sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X, Y)) \end{aligned} \quad (6.7)$$

$$\text{Bias}_\theta[\hat{\theta}] = E[\hat{\theta}] - \theta \quad (6.8)$$

Når vi viser om estimatoren er forventningrett, sett $\text{Bias}_\theta[\hat{\theta}] = 0$, og finn forventningen til estimatoren. Her blir det nyttig å bruke følgende:

$$E[X_i] = \mu = \frac{d}{dt} [M(t)]_{t=0} \quad (6.9)$$

Det er mange måter å måle risiko på: varians er én av dem.

6.2 Diverse om fordelinger

Definisjon av *sannsynlighetstetthet*:

- $\forall x \in S. \quad f_X(x; \theta) \geq 0$
- $\int_S f_X(x; \theta) = 1$

På oppgaver der vi skal finne sannsynlighetstettheten, skriv svaret som $f_X(x) = \begin{cases} \text{noe} & , x \in [a; b] \\ 0 & , \text{ellers} \end{cases}$.

Fordelingsfunksjon: $F_X(x) = P(X \leq x)$, og sannsynlighetstettheten til X er dermed den deriverte av $F(x)$, hvilket er .

Sammenheng mellom sannsynlighetstetthet og fordelingsfunksjon: $f(x) = \frac{d}{dx}[F(x)]$

Tips: skal man utlede en fordeling, for eksempel i transformasjoner, begynn med $F_Y(y) = P(Y \leq y)$. Denne er svært sentral og svært viktig å starte med. Obs på at vi må avgjøre om noe vokser eller avtar så vi setter riktig ulikhet ved transformasjonen.

Navn på fordelinger:

- Hvis $X \sim N(0, 1)$ & $F_Y(y) = F_X(\ln y)$, da er $Y \sim \text{lognormal}(0, 1)$

Mediangevinst: $P(Y \leq M) = 0.5$

Tre måter å vurdere en estimator:

1. $\text{Bias}_\theta(\hat{\theta}) = E[\hat{\theta}] - \theta$
2. $\text{Var}[\hat{\theta}]$
3. $\text{MSE}[\hat{\theta}] = \text{Var}[\hat{\theta}] + (\text{Bias}_\theta(\hat{\theta}))^2$

6.3 Max og min av andre variabler

6.3.1 Stokastisk variabel er minimum

$$Y = \min(X_1, \dots, X_n) = X_{(1)}$$

- vi har $F_Y(y) = P(Y \leq y)$
- Vi vet at $\forall i. \quad y \leq x_i$
- M.a.o. $\forall i. \quad x_i \geq y$
- Så vi bør snu ulikhetstegnet.
- Det gir $F_Y(y) = 1 - P(Y \geq y)$
- Sannsynligheten for at $Y = \min(x_1, \dots, x_n) > y$ er den samme som at $x_1 > y, \dots, x_n > y$.
- Setter inn og får $F_Y(y) = 1 - P(x_1 > y, \dots, x_n > y)$, der kommaene signaliserer "og/snitt".
Bruk at x 'ene er uavhengige slik at $P(A \cap B) = P(A) \cdot P(B)$ og utled ferdig derfra.

6.3.2 Estimator blir minimum

$$\hat{\theta}_{ML} = X_{(1)}$$

1. $L(\theta) = \prod_{i=1}^n \exp(-(x_i - \theta))$
2. $\ln L(\theta) = n\theta - \sum_{i=1}^n x_i$
3. Vi ser at $\ln L(\theta)$ vokser lineært med θ .
4. Vi ønsker en størst mulig (log)likelihood.
5. Det oppnås med størst mulig θ .
6. $f(x; \theta) = \exp(-(x - \theta))$ hvis $x \geq \theta$. Denne restriksjonen gjelder for alle x .
7. Så siden vi ønsker en størst mulig θ gitt $x_i \geq \theta \forall i = 1, 2, \dots, n$, vil vi velge $\theta = \min(x_i) = x_{(1)}$.
8. Dermed har vi $\hat{\theta}_{ML} = \min(X_i) = X_{(1)}$.

6.3.3 Vise at en transformasjon blir ...fordelt

På oppgaver hvor en stokastisk variabel er en rank av en annen, prøv *både* \leq og $>$ samt \geq og $<$. Får jeg ikke det jeg søker, prøv bare å snu ulikheten i $F_{X_{(1)}} = P(\min(X_i) \leq x)$ isteden.

6.4 Sentralgrensesetningen

Kort fortalt kan vi bruke sentralgrensesetningen til å vise at $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ og at $\sum X_i = Y \sim N(n\mu, n\sigma^2)$

Den **klassiske** sentralgrensesetningen sier at *summen* av *uavhengige* og *identisk fordelte* (i.i.d) stokastiske variabler er tilnærmet normalfordelt.

«Bruk den sentrale grensesetningen (SGS) for å argumentere for at normalfordelingen kan brukes som approksimasjon til summen av en stokastisk variabel.» Fremgangsmåte:

- Skriv at sentralgrensesetningen sier at X_1, X_2, \dots, X_n uavhengige stokastiske variabler trukket fra en fordeling med $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2 < \infty \forall i$. Da gjelder at $\frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \xrightarrow{d} N(0, 1)$ når $n \rightarrow \infty$

- Definer $Y = \sum_{i=1}^n X_i$. Det er denne som er nøkkelen: vi ønsker å vise at summen er tilnærmet normalfordelt. Erstatt \bar{X} med $\frac{1}{n}Y$. Vi har da $\frac{\frac{1}{n}Y - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$
- Sett $E\left[\frac{\frac{1}{n}Y - \mu}{\sigma/\sqrt{n}}\right] = 0$ og finn $E[Y]$.
- Sett $Var\left[\frac{\frac{1}{n}Y - \mu}{\sigma/\sqrt{n}}\right] = 1$ og finn $Var[Y]$.
- $\therefore Y \sim N(n\mu_X, n\sigma_X)$
- Har vi et utvalg vi skal finne sannsynligheter til, finn estimatene $\widehat{E[Y]}$ og $\widehat{Var[Y]}$, sett $P(Y \leq a) = P\left(\frac{Y - \widehat{E[Y]}}{\widehat{Var[Y]}} \leq \frac{a - \widehat{E[Y]}}{\widehat{Var[Y]}}\right) = \Phi\left(\frac{a - \widehat{E[Y]}}{\widehat{Var[Y]}}\right)$ og regn ut.

Vær obs på følgende:

- For binomisk: Y er summen av uavhengige Bernoullivariabler X_i , så det er Y som er binomisk fordelt.

Eksempel med Poisson

1. $Y = \sum_{i=1}^n W_i, W_i \sim Poi(\theta)$
2. $\therefore Y \sim Poi(n\theta)$
3. $\therefore Poi(n\theta) \approx N(n\theta, n\theta)$ når n er stor. Setter vi $n\theta = \lambda$, får vi
4. $\therefore Poi(\lambda) \approx N(\lambda, \lambda)$.

En alternativ måte å fremstille sentralgrensesetningen på, som er nyttig hvis variablene ikke er identisk fordelte, er,

$$\frac{\sum_{i=1}^n x_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \sim N(0, 1) \quad (6.10)$$

6.5 Transformasjon

Noen ganger er det lettere med å bruke CDF, noen ganger er det bedre å bruke formelen under.

$$\int_a^b f_X(x)dx = \int_{y(a)}^{y(b)} f_Y(x(y)) \cdot \left| \frac{dx}{dy} \right| \cdot dy \quad (6.11)$$

1. Definer Y
2. List grensene på begge sider av likhetstegnet
3. Finn X
4. Deriver og sjekk om den er voksende eller avtakende (eventuelt begge deler gjennom intervallet).
5. Sett inn.

Si at vi har en fordeling – for eksempel Paretofordelingen – og vi har funnet en transformasjon til den uniforme fordelingen. Hvis vi ønsker å simulere tall fra Paretofordelingen, finnes det en veldig enkel måte:

1. Genererer noen tall fra $U \sim U(0, 1)$
2. Sett de genererte tallene inn i transformasjonen – ligningen som viser sammenhengen mellom Paretofordelingen og den uniforme fordelingen – og vi har dermed simulert tall fra Paretofordelingen.

Enhver transformasjon trenger:

- sannsynlighetstettheten eller ekvivalent
- sammenhengen mellom de to variablene
- domenet til én av dem.

To type transformasjoner:

1. får vite sammenhengen
2. må finne sammenhengen

6.6 Momentgenererende funksjon

MGFen bestemmer unikt distribusjonen til en stokastisk variabel: har to stokastiske variabler samme MGF, da har de samme distribusjon.

$$M(t) = E[e^{tX}] \quad (6.12)$$

$$M(t) = \int_S e^{tx} f(x; \theta) dx \quad (6.13)$$

$$E[X^i] = M^{(i)}(0) \quad (6.14)$$

En variant av denne blir: $X \sim Poi(\lambda) \Leftrightarrow E[X^2] = \lambda + \lambda^2$

6.7 Generelt om estimasjon

Gjør rede for hvorfor en parameterestimator er en stokastisk variabel."

- $\hat{\theta}$ er en stokastisk variabel fordi den er en funksjon av stokastiske variabler (i.e. stikkprøven). \therefore trekker vi på nytt, altså får en ny stikkprøve, får vi en ny $\hat{\theta}$.

6.8 Momentestimator til θ (method of moments)

Nøkkel er at teoretisk moment i skal være lik empirisk moment i . Vi trenger like mange momenter som vi har ukjente som vi skal finne. Typisk skal vi bare finne én. Fremgangsmåten for momentmetoden rundt origo er (kalles **moments**:

1. Teoretisk moment: $\forall i \in \mathbb{N}. \quad \mu_i = M^{(i)}(0) = E[X^i]$
2. Empirisk moment: $\forall i \in \mathbb{N}. \quad m_i = \sum_{j=1}^n x_j^i$

3. $\forall i \in \mathbb{N}$. teoretisk moment i = empirisk moment i .

Vi har også momentmetoden rundt gjennomsnittet. Denne gir **central moments**. Noen ganger kan denne være greiere å bruke.

- $E[X] = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ for den første.
- $(E[X] - \mu)^k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$ for resten.

Vi må fortsette å regne momenter helt til vi har alle estimatorene vi ønsker.

Gjør rede for forskjellen mellom teoretisk og empirisk moment."

- Teoretisk moment: en *egenskap i en fordeling* med tetthet f .
- Empirisk moment: en *egenskap i en stikkprøve*.

6.9 Maximum likelihood-estimator til θ

Nøkkel er at vi ønsker å finne den θ som maksimerer sannsynligheten for at vi observerer utvalget vi faktisk har observert under en sannsynlighetsfordeling som er spesifisert av $f(x; \theta)$. Merk at $;$ er det samme som $|$, og at vi dermed sier fordelingsfunksjonen til x gitt θ ($f(x|\theta)$). Merk at $\forall i = 1, 2, \dots, n. x_i$ er utvalget vårt. Disse behandler vi som faste parametre", imens θ er ML-funksjonens variabel og kan variere fritt.

1. $L(\theta) = \prod_{i=1}^n f(x_i; \theta)_{X=x}$
2. Som regel (alltid?) må $\ln[VS] = \ln[HS]$ til. Her brukes $\ln(ab) = \ln a + \ln b$: logaritmen til et produkt er lik summen av logaritmene.
3. Finn $\frac{\partial}{\partial \theta} [\ln L(\theta)] = 0$ og $\frac{\partial^2}{\partial \theta^2} [\ln L(\theta)] < 0$

Forutsett at jeg ikke behøver finne annenderiverte for å verifisere at de er maksimum.

Kan maksimere L ved max L eller min $-L$.

6.10 Asymptotisk approksimasjon

Når n er stor—viktig å påpeke dette—og vi ikke kjenner den eksakte fordelingen til en ML-estimator, kan vi bruke den asymptotiske fordelingen.

Obs obs at det er tetthetsfunksjonen og ikke likelihoodfunksjonen som brukes her.

$$aVar[\hat{\theta}_{ML}] = \left(-nE \left[\frac{\partial^2 \ln f(X; \theta)}{\partial \theta^2} \right] \right)^{-1} \quad (6.15)$$

$$\hat{\theta}_{ML} \sim N \left(\theta; \left(-nE \left[\frac{\partial^2 \ln f(X; \theta)}{\partial \theta^2} \right] \right)^{-1} \right) \text{ for store } n \quad (6.16)$$

Obs obs: skal ikke substituere θ med ML-estimatoren før vi beveger oss inn i konfidensintervallene.

6.11 Konfidensintervall

Etter at vi har demonstrert at $\hat{\theta}$ er estimatoren vår – enten med MM- eller ML-metoden – og hvis vi vet at fordelingen er approksimalt normalfordelt (for eksempel via asymptotisk approksimasjon), da kan vi *substituere* θ med vår observerte $\hat{\theta}_{ML}$ i standardavviket. Da får vi,

Skriv at "I praksis må θ erstattes med $\hat{\theta}_{ML}$."

$$P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha, \quad Z = \frac{\hat{\theta} - \theta}{\sigma[\hat{\theta}]} \quad (6.17)$$

Innsikt: Konfidensintervallet har en $1 - \alpha$ sannsynlighet for å dekke den sanne parameteren θ . Merk også at intervallet vil bli større jo færre observasjoner vi har: få observasjoner betyr mye usikkerhet om hvor θ ligger.

Merk at vi ikke vil uttale oss om sannsynligheten for θ . θ ligger fast, stikkprøvene gir CI's som varierer fra gang til gang. Med andre ord, det er den øvre og nedre grensen som er stokastiske variabler, ikke θ .

6.11.1 Utledning av et CI

- $P(\hat{\theta}_{ML} \leq c) = 1 - \alpha/2$

- $P(\frac{\hat{\theta}_{ML} - \theta}{\theta/\sqrt{n}} \leq \frac{c - \theta}{\theta/\sqrt{n}}) = 1 - \alpha/2$
- $\Phi(\frac{c - \theta}{\theta/\sqrt{n}}) = 1 - \alpha/2$
- $\frac{c - \theta}{\theta/\sqrt{n}} = \Phi^{-1}(1 - \alpha/2)$
- $c = \theta + \frac{\theta}{\sqrt{n}} \cdot 1,96$
- $c = \hat{\theta} + \frac{\hat{\theta}}{\sqrt{n}} \cdot 1,96$

6.12 Hypotesetesting

Under H_0 gjelder at:

$$z_{obs} = \frac{\hat{\theta}_{obs} - E[\theta^{H_0}]}{\sigma[\theta^{H_0}]} \sim N(0, 1) \quad (6.18)$$

Behold H_0 om $P(\hat{\theta}_{obs} \in \theta^{H_0} \pm z_{1-\alpha/2} \cdot \sigma[\theta^{H_0}]) = 1 - \alpha$

Forkast H_0 om $P(\hat{\theta}_{obs} \notin \theta^{H_0} \pm z_{1-\alpha/2} \cdot \sigma[\theta^{H_0}]) = 1 - \alpha$

Formuler, via θ , en test om at prøven kommer fra en uniform fordeling". Nøkkelen er at vi må studere for hvilken θ -verdi tetthetsfunksjonen blir uniform fordelt. Etter å ha funnet hvilket verdi det er, er det den vi bruker som nullhypotese.

6.12.1 Eksempel

- $z_{obs} = (\frac{\mu_{obs} - \mu^{H_0}}{\sigma^{H_0}/\sqrt{n}}) \sim N(0, 1)$ under H_0
- Setter inn testobservatoren vår og verdiene til μ^{H_0} , σ^{H_0} , og n .
- Kritisk grense ved $P(Z \leq z_{obs}) = 1 - \alpha$
- \therefore Kritisk grense ved $\Phi(z_{obs}) = 1 - \alpha$
- \therefore Kritisk grense ved $\Phi^{-1}(\Phi(z_{obs})) = \Phi^{-1}(1 - \alpha)$
- \therefore Kritisk grense ved $z_{obs} = \Phi^{-1}(1 - \alpha)$. Sett inn og løs for \bar{x} .

6.13 Styrke til en hypotesetest

Vi har to stykker å bruke:

1. $K(\mu^{H_0}) = P(\bar{X} > c; \mu = \mu^{H_0}) = \alpha$
2. $K(\mu^{H_1}) = P(\bar{X} > c; \mu = \mu^{H_1}) = 1 - \beta$

Disse kan vi bruke til å finne det vi ønsker, for eksempel β og n .

$P(\text{type I-feil}) = \alpha$

$P(\text{type II-feil}) = \beta$

6.14 Likelihood ratio test

LR-testen sammenligner *max likelihood*-verdiene under H_0 og H_1 . Er verdien betydelig større under alternativhypotesen enn under nullhypotesen, forkaster vi.

Mer utdypende: LR ser på forholdet mellom *max likelihood*'er i to forskjellige scenarioer: H_0 og H_1 . Er H_0 sann, vet vi den ene parameterverdien: ingen vits i å bruke ML-estimatoren til den verdien. Men for alle andre, både i L_0 og L_1 , må vi bruke ML-estimatorene. k er differansen mellom antall "avhengige" variabler, som altså blir antallet parametre vi hypotesetester over.

Den enkle og anvendte metoden i faget er:

1. Finn $\ln L(\theta)$
2. For $\ln L_0$, sett inn $\theta = \theta^{H_0}$ fordi det er vår *max likelihood* under H_0 .
For $\ln L_1$, sett inn $\theta = \hat{\theta}_{ML}$ fordi den er vår *max likelihood* under H_1 .
3. Finn $LR = 2(\ln L_1 - \ln L_0)$
4. $LR \sim \chi^2(k)$ med k som er forskjell i antall parametre i hypotesen. Er LR stor, er forkaster vi.

Den fundamentale metoden er:

1. List H_0 og H_1 .
2. $\hat{\omega}$ er den mengden som H_0 utspenner. $\hat{\Omega}$ er den mengden som H_0 og H_1 samlet utspenner.
3. $\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$. Bruk $L(\theta) = \prod_{i=1}^n f(x_i; \theta)_{X=x}$ og finn λ . Hvis parameteren vår er

4. Forkastningsområdet er der $\lambda \leq k$. Løs for parameteren vi har en hypotesetest for.

6.15 Betingede sannsynligheter

- Hvis uavhengige: $f(x, y) = f_X(x)f_Y(y)$
- Marginalsannsynlighet: $f_X(x) = \int_{\text{nedre til } y}^x f(x, y)dy = \int_{\text{nedre til } y} f(x|y)f(y)dy$
- Merk at marginalsannsynligheten til X skal være en funksjon av x .

$$P(Y = y) = \sum_{x=0}^{\infty} P(Y|X = x)P(X = x) \quad (6.19)$$

Fra null til uendelig er fordi vi må summere over alle x -grenene, og det kan være inntil uendelig mange av disse.

Hvis X er kjent (men ingen konstant!), er

$$E[E[XY|X]] = E[X \cdot E[Y|X]] \quad (6.20)$$

- Finn $f(x, y)$. Med dette menes finn snittetthetsfunksjonen. $f(x, y) = f(y|x) \cdot f_X(x)$
- Finn $f_Y(y)$. Med dette menes finn marginaltettheten til Y .
- Merk at *forventningen* til Y gitt x skal være en funksjon av x .
- $E[g(X, Y)] = \int \int_{S_X S_Y} g(x, y)f(x, y)dydx$

6.16 Bayesiansk inferens

Fremgangsmåten:

1. Vi vet noen verdier. $\begin{cases} \lambda_1 \\ \lambda_2 \end{cases}$
2. Apriorifordelingen er $\begin{cases} P(\lambda_1) \\ P(\lambda_2) \end{cases}$

3. Dataene vi samler inn gir oss $\begin{cases} P(X = x|\lambda_1) \\ P(X = x|\lambda_2) \end{cases}$
4. Finner aposteriorfordelingen, som er vår oppdaterte tro. Den er et vektet gjennomsnitt av apriori og data. $\begin{cases} P(\lambda_1|X = x) \\ P(\lambda_2|X = x) \end{cases}$

Får god nytte for følgende to formler:

$$P(X = x, Y = y) = P(X = x|Y = y) \cdot P(Y = y) \quad (6.21)$$

$$P(X = x) = \sum_y P(X = x, Y = y) \quad (6.22)$$

Bayes lov:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (6.23)$$

Gjør rede for forskjellen mellom frekventistisk og Bayesianisk definisjon av sannsynlighet."

- Frekventistisk: gjentar vi et forsøk mange ganger, vil andelen forsøk der utfallet er A være $P(A)$. $\therefore P(krone) = \frac{\mathcal{N}(krone)}{\mathcal{N}(kast)}$ når $kast \rightarrow \infty$. Vi tildeler altså en begivenhet en sannsynlighet basert på et eksperiment.
- Bayesianisk: $P(A)$ er vår tro på at A skal skje.

Gjør rede for forskjell mellom parameterestimator i klassisk og Bayesianisk inferens."

- Frekventistisk: *parameterverdien er et fast men ukjent tall*. Vi bruker stikkprøven—stokastisk variabel—til å estimere dette tallet.
- Bayesianisk: *parameterverdien er en stokastisk variabel*. Fordelingen til denne *før* stikkprøven representerer vår tro på forskjellige parameterverdier. Vi bruker deretter stikkprøver til å *oppdatere* vår oppfatning.

Konfidensintervaller og kredibelintervaller:

- Frekventistisk: parameterverdi er fast, intervallgrenser er tilfeldig variabel.
- Bayesianisk: intervallgrenser er fast, estimert parameter er en tilfeldig variabel.

6.17 Regresjon

Her vi må bruke betinget. $Y = \alpha + \beta X + \varepsilon \Rightarrow (Y|X = x)$

1. Finn først normalfordelingen til Y_t . Sett inn disse verdiene i tetthetsfunksjonen til normalfordelingen.
2. Erstatt x i normalfordelingen med y_t .
3. Finn, for eksempel, ML-estimatorene. Disse er:
4. $\hat{\beta}_{ML} = \frac{n\Sigma_t y_t - \Sigma_t \Sigma y_t}{n\Sigma t^2 - (\Sigma t)^2}$
5. $\hat{\alpha}_{ML} = \bar{y} - \hat{\beta}_{ML} \cdot t$
6. $\hat{\sigma}_{ML}^2 = \frac{1}{n} \Sigma (y_t - \hat{\alpha}_{ML} - \hat{\beta}_{ML} \cdot t)^2$

Del 7

Diverse

- $f(x)$ til max an n uavhengige stokastiske variabler med fordelingsfunksjon F og tetthetsfunksjon f er $f(y = y_{(n)}) = n[F(y)]^{n-1} \cdot f(y)$

Maximum av n uavhengige stokastiske variabler er lik $n[F(y)]^{n-1}f(y)$.