Innhold

2 Kapittel 2—Sannsynlighet

2.3 Sannsynlighetsfunksjon

Teorem 2.3.1 $P(A^C) = 1 - P(A)$.

Teorem 2.3.2 $P(\emptyset) = 0$.

Teorem 2.3.3 Om $A \subset B$, er $P(A) \leq P(B)$.

Teorem 2.3.4 For uansett hendelse er $P(A) \leq 1$.

Teorem 2.3.5 La $A_1, \ldots, A_n$ være hendelser definert over S. Om $A_i \cap A_j = \emptyset$ for $i \neq j$, er

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Teorem 2.3.6 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

2.4 Diskrete og kontinuerlige sannsynlighetsfunksjoner

Krav for **diskrete** sannsynlighetsfunksjoner er at

- $P(s) \geq 0$ for hver $s \in S$
- $\sum_{alle\ s \in S} P(s) = 1$

For **kontinuerlige** sannsynlighetsfunksjoner har man at

- $f(x) \geq 0$ for alle $x \in S$
- $\int_S f(x)\, dx = 1$
- $P(A) = \int_A f(x)\, dx$

2.6 Betinget sannsynlighet

Om A og B er to hendelser, hvor man antar at B allerede har skjedd, kan man regne ut sannsynligheten for A skjer, $P(A|B)$ gitt ved:

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

Bemerk at $P(A \cap B) = P(A \,|\, B)P(B)$.

Teorem 2.6.1 La $\{A_i\}_{i=1}^n$ være et sett av hendelser definert over S slik at $S = \bigcup_{i=1}^n A_i$, $A_i \cap A_j = \emptyset$ for $i \neq j$. Da er

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

Teorem 2.6.2 *(Bayes)* La $\{A_i\}_{i=1}^n$ være et sett av hendelser definert over S slik at $S = \bigcup_{i=1}^n A_i$, $A_i \cap A_j = \emptyset$ for $i \neq j$. Da er

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

2.7 Uavhengig sannsynlighet

Def. 2.7.1 To hendelser A og B sies å væ re *uavhengige* hvis

$$P(A \cap B) = P(A) \cdot P(B)$$

Def. 2.7.2 Hendelser $A_1, \ldots, A_n$ er *uavhengige* hvis hvert sett av $i_1, i_2, \ldots, i_k$ mellom 1 og n,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

2.8 Gjentatte uavhengige forsøk

Uavhengige forsøk kalles *trials*. Det vil si at det j'te forsøket er upåvirket av de j-1 tidligere. Eksempelvis vil $P(A)$ = "summen av alle gunstige utfall som utgjør A". Et nyttig tips i denne sammenhengen er den uendelige summen av geometriske rekker, hvor $p \in [0, 1]$

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$$

2.9 Kombinatorikk

Multiplikasjonregelen Om en operasjon A kan utføres på m forskjellige måter og B sine n måter, kan sekvensen (operasjon A, operasjon B), utføres på $m \cdot n$ forskjellige måter. Dette kan generaliseres, slik at det gjelder for n operasjoner.

Teorem 2.9.1 Antall permutasjoner av lengde k som kan formes av n distinkte elementer uten gjentakelse, er

$$n(n-1)(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

Av dette får man et korollar som sier at antall måter å permutere n objekter på er $n!$

Binomialformelen Antall måter for å forme kombinasjoner av størrelse k fra et sett på n distinkte objekter hvor gjentakelse ikke er tillatt, er

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

2.10 Kombinatorisk sannsynlighet

Det som går igjen her er P("hendelse") $= \frac{\text{antall gunstige}}{\text{antall mulige}}$.

2.11 Kombinatorisk–hypergeometrisk og binomalfordeling

Hypergeometrisk Anta en urne inneholder r røde og w hvite brikker ($r + w = N$). Om n brikker trekkes vilkårlig uten gjentakelse og Y forteller totalt antall røde brikke valgt, er

$$P(Y = k) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

Binomialfordeling Anta en serie med n uavhengige forsøk, hvert av dem resulterende i "suksess" eller "feil". La $p = P$(suksess i gitt forsøk) konstant. Y forteller antall suksesser i n forsøk. $E(X) = p$ og $Var(X) = np(1-p)$.

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots, n$$

3 Kapittel 3—Tilfeldige variabler

3.2 Tetthet og fordelinger

Def. 3.2.1 En reell funksjon som har utfallsrommet S som definisjonsmengde, kalles en *tilfeldig variabel*. Disse gjengis ved store bokstaver, f.eks. X, Y, Z. Oftest teller man opp antall utfall som oppfyller betingelsen og deler på antall mulige.

pdf $f_Y(y)$ Forklarer sannsynlighetsstrukturen indusert på den reelle linja av Y. Eksempel er $f_Y(3) = P(\text{sum lik } 3) = P((1,2),(2,1)) = \frac{2}{36}$. For kontinuerlige funksjoner sier man generelt

$$P(a \leq Y \leq b) = \int_a^b f_Y(y)\, dy$$

Medianen, y^* Dette er verdien y^* hvor $P(Y = y^*) = \frac{1}{2}$. Dette løser man ved å finne integralverdien fra 0 til y^*, og dermed løse den ukjente y^* ved å sette uttrykket lik $\frac{1}{2}$.

Def.3.2.2 La Y være tilfeldig variabel på S med sannsynlighetsfunksjon P. Den *kumulative fordelingsfunksjonen av* Y, $F_Y(y)$, er sannsynligheten for at verdiene fra Y mappes til verdier på den reelle linja *mindre eller lik* y.

$$F_Y(y) = P(\{s \in S | Y(s) \leq y\}) = P(Y \leq y)$$

Regler forbundet med den kumulative fordelingsfunksjonen, *cdf*:

- $P(Y > y) = 1 - F_Y(y)$
- $P(a < Y \leq b) = F_Y(b) - F_Y(a)$
- $P(Y = t) = F_Y(t) - \lim_{y \to t^-} F_Y(y)$

Teorem 3.2.1 Om Y er kontinuerlig variabel med pdf $f_Y(y)$ og cdf $F_Y(y)$, da er $f_Y(y) = F_Y'(y)$

3.3 Simultane tettheter

Def. 3.3.1 Diskrete X og Y er to diskrete variabler på S. Den *simultane pdf* $f_{X,Y}(x,y) = P(X = x, Y = y)$.

Def. 3.3.1 Kontinuerlig X og Y er to kontinuerlige variabler på S. Den *simultane pdf* $F_{X,Y}(x,y)$ er området R i xy-planet,

$$P((X,Y) \in R) = P(\{s \in S \mid (X(s), Y(s)) \in R\}) = \int_R \int f_{X,Y}(x,y)\, dx\, dy$$

Def. 3.3.2 La X og Y være to tilfeldige variabler. Den *simultane kumulative fordelingsfunksjonen* (eller simultan cdf) av X og Y, er

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

Domene til $F_{X,Y}(x,y)$ er settet med alle par av reelle tall.

Teorem 3.3.1 La X og Y være kontinuerlige tilfeldige variabler og $F_{X,Y}(x,y)$ være simultane cdf. Da er:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x,y)$$

Teorem 3.3.2 La X og Y være diskrete variabler med simultan pdf $f_{X,Y}(x,y)$. De *individuelle* pdf's for X og Y—$f_X(x)$ og $f_Y(y)$, respektivt, kan utregnes ved:

$$\text{(a) } f_X(x) = \sum_{alle\ y} f_{X,Y}(x,y) \quad \text{(b)} f_Y(y) = \sum_{alle\ x} f_{X,Y}(x,y)$$

Liknende er det for kontinuerlige variabler.

$$\text{(a) } f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{dy} \quad \text{(b)} f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{dx}$$

3.4 Uavhengige tilfeldige variabler

Def. 3.4.1 Tilfeldige variabler X og Y er *uavhengige* hvis for hvert intervall A og B

$$P(X \in A, Y \in B) = P((X,Y) \in A \otimes B) = P(X \in A) \cdot P(Y \in B)$$

Teorem 3.4.1 To tilfeldige variabler X og Y er uavhengige hvis og bare hvis $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ for alle x og y.

Def. 3.4.2 n tilfeldige variabler $X_1, \ldots, X_n$ er *uavhengige* for alle $x_1, \ldots, x_n$ hvis $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$.

Def. 3.4.3 Om alle $X_1, \ldots, X_n$ har samme pdf, er settet *ett tilfeldig utvalg av størrelse n*.

3.5 Kombinasjon og transformasjon av variabler

Teorem 3.5.1 La X være tilfeldig variabel med pdf $f_X(x)$, $a \neq 0$ og b være konstanter og definer $Y = aX + b$.

(a) Om X er diskrete,

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)$$

(b) Om X er kontinuerlig,

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Teorem 3.5.2 La X være en kontinuerlig tilfeldig variabel med pdf $f_X(x)$. La $Y = X^2$. For $y > 0$,

$$f_Y(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y}))$$

Teorem 3.5.3 La X og Y være uavhengige tilfeldige variabler med pdf's $f_X(x)$ og $f_Y(y)$, respektivt. La $Z = X + Y$.

(a) Om X og Y er diskrete,

$$f_Z(z) = \sum_{alle\ x} f_X(x) f_Y(z-x)$$

(b) Om X og Y er kontinuerlige,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$$

Sistnevnte integral kalles ofte *konvolusjonen* f_X og f_Y.

3.6 Ordningsobservatører

En **observator** er en funksjon av en tilfeldig variabel. Eksempelvis kan gjennomsnittet være en slik observatør. I mange tilfeller kan det også være en tilfeldig variabel. F.eks. $X_1, \ldots, X_n$ er et tilfeldig utvalg (stikkprøver).
Når alle X_i er observert kan vi ordne dem fra minste til største verdi. Dette medfører en ny notasjon $X_1^{'}, \ldots, X_n^{'}$. Dette er **ordningsobservatøren**. Ofte er $X_{min} = X_1^{'}$ og $X_{max} = X_n^{'}$.

Teorem 3.6.1 La X være kontinuerlig med pdf $f_X(x)$. Om et utvalg av størrelse n trekkes fra $f_X(x)$, så er den marginale pdf for i'te orden statistikk gitt:

$$f_{X_i}(y) = \frac{n!}{(i-1)!(n-i)!}(F_X(y))^{i-1}(1-F_X(y))^{n-i}f_X(y)$$

Korollar X_{min} og X_{max} er henholdsvis minste og største orden statistikk.

(a) $f_{X_{\min}}(y) = n \cdot f_X(y) \cdot (1-F_X(y))^{n-1}$
(b) $f_{X_{\max}}(y) = n \cdot f_X(y) \cdot (F_X(y))^{n-1}$

3.7 Betingede tetthetsfunksjoner

Def. 3.7.1 Den betingede pdf av Y gitt x, dvs. sannsynligheten for at Y tar verdi y gitt at $X = x$, skrives $f_{Y|x}(y)$ og gis ved:

$$f_{Y|x}(y) = P(Y = y|X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Denne definisjonen kan generaliseres for å involvere flere enn to tilfeldige variabler.

Kontinuerlig tilfelle Om X er kontinuerlig, tenker man på $P(Y \leq y|X = x)$ som grense:

$$P(Y \leq y|X = x) = \lim_{h \to 0} \frac{\int_x^{x+h} \int_{-\infty}^{y} f_{X,Y}(t,u)\,du\,dt}{\int_x^{x+h} f_X(t)\,dt}$$

3.8 Forventede verdier

Def. 3.8.1 La X være tilfeldig variabel med pdf $f_X(x)$. Forventet verdi for X skrives $E(X)$ eller μ .
(a) $E(X) = \sum_{\text{alle x}} x \cdot f_X(x)$ om X er diskret.
(b) $E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx$ om X er kontinuerlig.

Medianen Medianen er nesten den samme som forventet verdi. La m være medianen. Da er $P(X < m = 0.5)$, som løses med hensyn på m.

3.9 Egenskaper til forventede verdier

Teorem 3.9.1 La X og Y være t.v. og a, b reelle tall.

$$E(aX + bY) = aE(X) + bE(Y)$$

Dette teoremet kan generaliseres slik at det involverer flere variabler og konstanter. Man snakker om at E er en *lineær transformasjon*.

Teorem 3.9.2 Om X er **diskret** med pdf $f_X(x)$ og hvis $g(x)$ er funksjon av X, er:

$$E(g(X)) = \sum_{\text{alle x}} g(x) \cdot f_X(x)$$

Om X er **kontinuerlig** så er:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx$$

Teorem 3.9.3 Anta X og Y er diskrete t.v. med simultan pdf $f_{X,Y}(x,y)$. $g(x,y)$ er en funksjon av X og Y.

$$E(g(X,Y)) = \sum_{\text{alle x}} \sum_{\text{alle y}} g(x,y) \cdot f_{X,Y}(x,y)$$

For kontinuerlige t.v. har man:

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y}(x,y)\, dx\, dy$$

Teorem 3.9.4 La X og Y være uavhengige t.v. Da er:

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

3.10 Variansen

Def. 3.10.1 Variansen, Var(X) er forventet verdi av dens kvadrerte deviation fra $\mu = E(X)$.

$$Var(X) = \sigma^2 = E((X-\mu)^2)$$

Teorem 3.10.1 La μ = mean = E(X). Da er:

$$Var(X) = E(X^2) - \mu^2$$

Teorem 3.10.2 La X være t.v. og a, b konstanter. Definer $Y = aX + b$.

$$Var(Y) = a^2 Var(X)$$

Teorem 3.10.3 La $X_1, \ldots, X_n$ være **uavhengige** variabler og $Y = X_1 + \cdots + X_n$.

$$Var(Y) = Var(X_1) + \cdots + Var(X_n)$$

Tips Husk at E(snittet av X) = μ og Var(snittet av X) = $\frac{\sigma^2}{n}$.

3.12 Momentgenererende funksjon

Def. 3.12.1 Den "momentgenererende funksjonen" (MGF) for X er gitt ved

$$M_X(t) = E(e^{tX})$$

Teorem 3.12.1 La X være en t.v. med pdf $f_X(x)$ og $M_X(t)$ være MGF for X.

$$M_X^{(r)}(0) = E(X^r)$$

Teorem 3.12.2 Anta X og Y er t.v. hvor $M_X(t) = M_Y(t)$ for et intervall av t inneholdende 0. Da er $f_X = f_Y$, dvs. at de har samme pdf.

Teorem 3.12.3 a) La X være t.v. med MGF $M_X(t)$ og definer $Y = aX + b$. Så:

$$M_Y(t) = e^{bt} M_X(at)$$

Teorem 3.12.3 b) La $X_1, \ldots, X_n$ være *uavhengige* t.v. med MGF'er $M_{X_i}(t)$. Definer $Y = X_1 + \cdots + X_n$. Dermed er

$$M_Y(t) = M_{X_1} \cdot M_{X_2} \cdots M_{X_n}$$

3.13 Chebyshev's ulikhet

Teorem 3.13.1 Chebyshev's ulikhet La X være t.v. med mean μ og varians σ^2. For $\epsilon > 0$ er den nedre grensa:

$$P(|X - \mu| < \epsilon) > 1 - \frac{\sigma^2}{\epsilon^2}$$

mens den øvre grensa er:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

4 Kapittel 4—Spesielle fordelinger

Det er 5 familier med sannsynlighetsfunksjoner i kapittel 4—poisson, normal, geometrisk, negativ binomial og gamma.

4.2 Poisson-fordelinga

Teorem 4.2.1 La λ være fiksert og n et vilkårlig heltall. For $x \geq 0$ er

$$lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}$$

Teorem 4.2.2 La $p(x, \lambda)$ være poisson-pdf'n.

$$p(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Vet også at $E(X) = \lambda$ og $Var(X) = \lambda$.

Poisson-fordelinga kan blant annet brukes til radioutstråling, krigsutbrudd, telefonsamtaler, trafikkulykker og skrivefeil i bøker. Når antall n blir stor i Bernoulliforsøk (binomialfordelinga) og p er liten, er omtrentlig bin(n, p) $\approx$ Po(λ). Dessuten har man ved binomisk situasjon at $p = \frac{\lambda}{n}$.

4.3 Normal-fordelinga

Teorem 4.3.1 DeMoivre-Laplace La X være binomial t.v. med n uavhengige forsøk med suksess p.

$$lim_{n\to\infty} P\left(c < \frac{X - np}{\sqrt{npq}} < d\right) = \frac{1}{\sqrt{2\pi}} \int_c^d e^{-x^2/2}\, dx$$

Tips Når n blir stor ved binomialfordelinga og skal finne $P(X \geq k)$, regner man ut c og eventuelt d for innsetting rett inn i dette ovenstående integralet. Dermed har man en god tilnærming på sannsynligheten.

Teorem 4.3.2 $P(a < X < b) = P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \int_v^h e^{-x^2/2}\, dx$.

Def. 4.3.1 En t.v. Z har *standard normalfordelinga* dersom pdf'n gis ved

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

X sies være *normalfordelt med parametre μ og σ*, $N(\mu, \sigma^2)$, dersom

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-0.5((x-\mu)/\sigma)^2} \quad -\infty < x < \infty$$

Teorem 4.3.3 Når X har normalfordelinga, $N(\mu, \sigma^2)$, er $E(X) = \mu$ og $Var(X) = \sigma^2$.

4.4 Geometrisk fordeling

Den geometriske fordelinga brukes for å finne ut hvor mange forsøk som trengs for å oppnå første suksess. Fordelinga blir

$$f_N(n) = p(1-p)^{n-1} = pq^{n-1}$$

Denne har bare en parameter.

Teorem 4.4.1 La N være geometrisk fordelt med p (= suksess på forsøk). Da er $E(N) = \frac{1}{p}$ og $Var(N) = \frac{1-p}{p^2}$.

4.5 Den negative binomial-fordelinga

Denne familien brukes når man skal finne ut sannsynligheten for r'te suksess i et antall Bernoulli-forsøk. Definer X som antall forsøk for den r'te suksessen. Videre er Y = antall fiaskoer før suksess er oppnådd. Dermed er $Y = X - r$. To måter for negative binomialfordelinga (har to parametre, r og p):

$$f_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad x = r, r+1, r+2, \ldots \text{ suksesser}$$

og

$$f_Y(n) = \binom{n+r-1}{n} p^r (1-p)^n \quad n = 0, 1, 2, \ldots \text{ fiaskoer}$$

Teorem 4.5.2 Om Y er t.v. og har negativ binomialfordeling med r og p, så er

$$E(Y) = \frac{r}{p} \text{ og } Var(Y) = \frac{r(1-p)}{p^2}$$

4.6 Gamma-fordelinga

Gamma-fordelinga brukes til å måle *tida det tar før r hendelser skjer.*

Def. 4.6.1 For et reelt tall $r > 0$ gis gamma-funksjonen av r slik:

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\, dx$$

Teorem 4.6.1 La $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}\,dx$

(a) $\Gamma(1) = 1$

(b) $\Gamma(0.5) = \sqrt{\pi}$

(c) $\Gamma(r+1) = r\Gamma(r)$ for postive reelle r

(d) $\Gamma(r+1) = r!$ dersom r er ikkenegativt heltall

(e) $\binom{n+r-1}{n} = \frac{\Gamma(n+r)}{\Gamma(n+1)\Gamma(r)}$

(f) $\frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^\infty u^{r-1}(1-u)^{s-1}\,du$

Def. 4.6.2 La X være t.v. med fordeling gitt nedenfor. Da er X gamma-fordelt med parametre r og λ, hvor r, $\lambda > 0$.

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x} \quad x > 0$$

Teorem 4.6.2 La X være gamma-fordelt med parametre r og λ. Da er

$$E(X) = \frac{r}{\lambda} \quad \text{og} \quad Var(X) = \frac{r}{\lambda^2}$$

Tricks Summen av eksponensialfordelte variabler er **gammafordelt**.

5 Kapittel 5—Estimering

5.2 Definisjoner

Om en gitt parameter θ_i er ukjent, må vi tilnærme den. Til dette bruker vi en funksjon kjent som *statistisk*, $W = h(Y_1, \ldots, Y_n)$, fra et tilfeldig utvalg av størrelse n. Eksempler: $mean = \bar{Y} = (1/n)\sum_{i=1}^n Y_i$, standardavviket $= S = \sqrt{(1/(n-1))\sum_{i=1}^n (Y_i - \bar{Y})^2}$.

Vi kaller formen av uttrykket for estimering av ukjent parameter for en *estimator*. Den resulterende numeriske verdien fra det, kalles *estimat*.

Det er to typer estimering, *punktestimat* og *intervallestimat*. Sistnevnte forteller litt mer om dataene—lengda av intervallet forteller estimatorens nøyaktighet.

5.3 Sentrering og effisient av punktestimater

To krav til en god estimator er at den skal være sentrert rundt den virkelige verdien (θ), dvs. at den er *unbiased*. Videre må estimatoren være nøyaktig (precision)—dvs. effisienten har liten varians. Dette tas nøyere opp i de to neste underkapitlene.

5.4 Sentrering (unbiasedness)

Def. 5.4.1 La $Y_1, Y_2, \ldots, Y_n$ være tilfeldig utvalg fra $f_Y(y;\theta)$. Estimatoren $W = h(Y_1, \ldots, Y_n)$ er sentrert for θ om $E(W) = \theta$ for alle θ.

Merk at det finnes flere sentrerte estimatorer for en undersøkelse. Ellers hender det at vi må estimere σ^2 uten å kjenne μ. Statistikken ofte brukt her er *sample variansen* $S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$.

5.5 Effisient

Effisient brukes til å bestemme hvilken estimator som er *bedre* enn andre.

Def. 5.5.1 La W_1 og W_2 være to sentrerte estimatorer. W_1 er mer effisient enn W_2 dersom

$$Var(W_1) < Var(W_2)$$

Relativ effisient av W_1 i forhold til W_2 er $Var(W_2)/Var(W_1)$.

5.6 Minimum-varians — Cramer-Rao nedre grense

Teorem 5.6.1 Cramer-Rao ulikheten La $Y_1, \ldots, Y_n$ være tilfeldig utvalg fra $f_Y(y;\theta)$. Anta settet hvor $f_Y(y;\theta) \neq 0$ ikke avhenger av θ. La $W = h(Y_1, \ldots, Y_n)$ være sentrert estimator, som i kontinuerlig tilfelle har 1. og 2. ordens partiell deriverte:

$$\textbf{a)}\ Var(W) \geq \left\{ nE\left[\frac{\partial \ln f_Y(Y;\theta)}{\partial\theta}^2 \right] \right\}^{-1}$$

$$\textbf{b)}\ \text{Likhet hvis: } \sum_{i=1}^{n} \frac{\partial \ln f_Y(Y;\theta)}{\partial\theta} = A(\theta)[h(y_1, \ldots, y_n) - \theta]$$

for alle $y_1, \ldots, y_n$, hvor $A(\theta)$ ikke avhenger av y_i. I det tilfellet, $Var(W) = [A(\theta)]^{-1}$.

$$\textbf{c)}\ \text{Kontinuerlig: } Var(W) \geq \left\{ -nE\left[\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial\theta^2} \right] \right\}^{-1}$$

Def. 5.6.1 W^* er *beste* estimator om den er med i klassen av sentrerte estimatorer (W) og $Var(W^*) \leq Var(W)$, dvs. sammenliknet med estimatorer som oppnår Cramer-Raos nedre grense.

Def. 5.6.2 En estimator $W = h(Y_1, \ldots, Y_n)$ er *effisient* hvis og bare hvis

$$Var(W) = \frac{1}{nE\left(\left[\frac{\partial \ln f_Y(Y;\theta)}{\partial\theta} \right]^2 \right)}$$

Effisienten av en sentrert W defineres som "utregnet varians" / "virkelige varians". **NB!** Om variansen til W er lik Cramer-Rao, så er den *beste estimator*. Det motsatte er ikke alltid sant—ingen trenger møte Cramer-Rao grensa. Dermed er ingen *effisient*, men minst en av dem er den *beste estimatoren*. Det kan eksistere varianser mindre enn Cramer-Rao, men da er ikke hypotesene i Teorem 5.6.1 tilfredsstilt.

5.7 Konsistens av estimator

En estimator er *konsistent* når P(W_n ligger vilkårlig nær parameteren som skal estimeres) = 1 om n blir stor. W_n er da asymptotisk sentrert og $Var(W_n)$ konvergerer mot 0.

Def. 5.7.1 $W_n = h(Y_1, \ldots, Y_n)$ er konsistent om den konvergerer i sannsynligheten til θ—dvs. for $\epsilon > 0$ og $\delta > 0$ eksisterer det $n(\epsilon, \delta)$ s.a.:

$$P(|W_n - \theta| < \epsilon) > 1 - \delta \qquad \text{for } n > n(\epsilon, \delta)$$

Merk! Fra Chebychevs ulikhet kan vi skrive $P(|W_n - \theta| < \epsilon) > 1 - \frac{Var(W)}{\epsilon^2}$.

5.8 Maximum Likelihood (MLE) og momentprinsippet

Her diskuteres to metoder for å finne estimatorer, MLE og momentprinsippet. Sistnevnte er ofte mye enklere å regne ut.

Def. 5.8.1 La $Y_1, \ldots, Y_n$ være utvalg fra $f_Y(y;\theta)$. Om simulane pdf for alle Y_i, L, tenkes på som funksjon av θ og hvis y_i er fiksert, da er "Likelihood-funksjonen"

$$L = L(\theta) = \Pi_{i=1}^{n} f_Y(y_i;\theta)$$

Def. 5.8.2 (Maximum-Likelihood-Estimering) La $L(\theta)$ være likelihood-funksjonen. For å konstruere estimatorer, leter vi etter den w som maksimerer $L(\theta)$, dvs. $L(w) \geq L(\theta)$. Dette er maximum-likelihood-estimatet (MLE) for θ. Ofte skrevet som $w = \hat{\theta}$.

NB! Ofte er det enklere å maksimere $\ln L(\theta)$ istedenfor $L(\theta)$.
NB! Om funksjonen har k ukjente parametre, lønner det seg å løse de k likningene av den deriverte $L(\theta_1, \ldots, \theta_k)$ lik 0.
NB! Det kan hende at derivasjon ikke fører frem. Plott grafen!

Definer $\mu_{(j)} = E(Y^j) = \int_{-\infty}^{\infty} y^j f_Y(y;\theta_1, \ldots, \theta_k)\, dy$. $\mu_{(j)}$ blir funksjoner av $\theta_1, \ldots, \theta_k$—slik $\mu_{(i)} = \mu_{(i)}(\theta_1, \ldots, \theta_k)$. Videre er $m_{(j)} = \frac{1}{n}\sum_{i=1}^{n} y_i^j \qquad j = 1, \ldots, k$. Momentprinsippet går ut på å sette $\mu_{(j)} = m_{(j)}$.

Def. 5.8.3 (Momentprinsippet) La $Y_1, \ldots, Y_n$ være tilfeldig utvalg fra $f_Y(y;\theta_1, \ldots, \theta_k)$. La $\mu_{(j)}$ og $m_{(j)}$ være teoretisk og virkelige momenter, henholdsvis. For å finne momentestimat for $\theta_1, \ldots, \theta_k$, løses likningene med hensyn på $\theta_1, \ldots, \theta_k$:

$$\mu_{(j)} = \mu_j(\theta_1, \ldots, \theta_k) = m_{(j)} \qquad j = 1, 2, \ldots, k$$

Løsningene, $w_1, \ldots, w_k$ kalles momentestimater. Ofte er w_i skrevet som $\hat{\theta}$.

5.9 Intervall-estimering

Punktestimater sier ikke noe om hvor nært parameteren estimatet ligger. Derfor skal vi bruke *konfidens-intervall* for (1) å finne sannsynligheten for at W ligger mellom to tall og (2) finne tallene som inneholder W med gitt sannsynlighet. Vi ønsker å finne to grenser, a og b, slik at $P(a < W < b) = 0.90$ og $P(W \leq a) = P(W \geq b) = 0.05$. Disse finner man ved å løse (for uniform fordeling)

$$\int_0^a f_W(w)\, dw = 0.05 \qquad \text{og} \qquad \int_b^{((n+1)/n)\theta} f_W(w)\, dw = 0.05$$

For uniform-fordelinga har man derfor følgende *90% konfidens-intervall*

$$\left(\frac{w}{\sqrt[n]{0.95}\,(\frac{n+1}{n})}, \frac{w}{\sqrt[n]{0.05}\,(\frac{n+1}{n})} \right)$$

Dette betyr at ved 100 forsøk, vil cirka 90 av intervallene inneholde parameteren θ. Konfidens-intervallet for **normalfordelinga** er

$$P(\bar{Y} - u_{0.05}\frac{\sigma}{\sqrt{n}} < \mu < \bar{Y} + u_{0.05}\frac{\sigma}{\sqrt{n}}) = 0.90$$

5.10 Konfidens-intervall for "p" i binomisk fordeling

La Y være bin(n, p). Når $n \to \infty$ er $\frac{Y}{n} \to N(p, \frac{p(1-p)}{n})$. Dette og DeMoivre-Laplace gir en parabel med to nullpunkter, p_1 og p_2. Området mellom disse er konfidens-intervallet vi ønsker. Definer $z_{\alpha/2}$ som $\frac{\alpha}{2}$-kvantilen for $Z \sim N(0,1)$, dvs. skillepunktet mellom konfidensintervallet. Om $z_{\alpha/2} \gg n$ får man $100(1-\alpha)\%$ konfidens-intervall for p

$$\left(\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{\frac{Y}{n}(1-\frac{Y}{n})}{n}}, \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{\frac{Y}{n}(1-\frac{Y}{n})}{n}}\right)$$

Ofte er det interessant å finne **minste** n slik at $P(|\frac{Y}{n} - p| < d) = 1-\alpha$. Ved å bruke identiteten for $N(0,1)$ ved stor n, får man først et uttrykk for n inneholdende ukjent p. For å være sikker på at valgte n tilfredsstiller kravene, velger vi den verdien som gir størst mulig verdi av $p(1-p)$ (binomial), dvs. $p = 0.5$. Dermed blir

$$n = \frac{z^2_{\alpha/2}}{4d^2}$$

6 Kapittel 6—Hypotesetesting

6.2 Desisjonsregelen

Hypotesetesting er å bruke sannsynlighetsteori for å velge mellom to alternativer, H_0 (den originale, nullhypotesen) og H_1 (det nye alternativet). H_0 antas være korrekt til det motsatte er bevist (Murder 1).

Def. 6.2.1 Enhver funksjon av observerte data avgjørende om vi skal godta eller forkaste H_0, kalles `test statistikk`. Etter utregning med utfallsdata har man w, en *observert test statistikk*. Området som resulterer i forkastelse av H_0, kalles *kritisk region* og skrive som C.

Forskere definerer ofte *herskende tvil* dersom W sier at H_0 er korrekt i mindre enn 5% av tida. Med andre ord, y^* burde velges slik at $P(Y \geq y^* \mid H_0 \text{ korrekt}) = 0.05$. Den utregnede verdien av likninga bør være større eller lik den observerte, for at H_0 skal forkastes.

Teorem 6.2.1 Anta y observerte suksesser blant n Bernoulli-tester. La $p = P(suksess)$ og α være definisjonen på tvilen. For å sjekke

$$H_0 : p = p_0 \qquad \text{mot} \qquad H_1 : p \neq p_0$$

bør vi forkaste H_0 om $\frac{y-np_0}{\sqrt{np_0(1-p_0)}}$ enten er $\leq -z_{\alpha/2}$ eller $\geq +z_{\alpha/2}$, hvor $P(Z \geq z_{\alpha/2}) = P(Z \leq -z_{\alpha/2}) = \alpha/2$. For å sjekke om

$$H_0 : p = p_0 \qquad \text{mot} \qquad H_1 : p > p_0$$

bør vi forkaste H_0 om $\frac{y-np_0}{\sqrt{np_0(1-p_0)}} \geq +z_\alpha$ For å sjekke om

$$H_0 : p = p_0 \qquad \text{mot} \qquad H_1 : p < p_0$$

bør vi forkaste H_0 om $\frac{y-np_0}{\sqrt{np_0(1-p_0)}} \leq -z_\alpha$

Fremgangsmåten ved hypotesetesting er

- Formuler nullhypotese og alternativ hypotese.

- Velg testobservator (X).
- Velg signifikansnivå α. Ofte er $\alpha = 0.05$.
- Bestem forkastningsområdet for testen.
- Undersøk om observert verdi for testobservator faller i forkastningsområdet eller ikke.

Tosidig test Merk at når vi tester $H_0 : p = k$ mot $H_1 : p \neq k$, får man to kritiske regioner. Vi løser begge likninger

$$P(Y \leq y_1^* \mid H_0 \text{ er sann}) \leq 0.025 \qquad P(Y \geq y_2^* \mid H_0 \text{ er sann}) \leq 0.025$$

hvor y_1^* og y_2^* er de to kristiske skillepunktene.

6.3 Type I og type II feil

Type 1 feil Vi forkaster H_0 når H_0 er korrekt. Slike unngås ved å kreve at signifikansnivået er lavt (vektleggende bevis), f.eks. $\alpha = 0.01$.

Type 2 feil Vi aksepterer H_0 når H_0 er feil. Symbolet for P(Type II feil oppstår) er $\beta(p)$. Den forteller hvordan det nye systemet går "uoppdaget" etterhvert som p varierer. Spesifikt

$$\beta(p) = P(\text{forkaste } H_0 \mid p)$$

Styrkefunksjonen, $1 - \beta = P(W \in C \mid H_1 \text{ sann})$, brukes for å angi *kvaliteten* av en test. Ved sammenlikning av tester, bør man velge den med "brattest" kurve.

6.4 Optimalitet—generalisert likelihood ratio

Generalisert Likelihood Ratio, GLR, foreslår faktisk testprosedyrer—analogt med maximum likelihood ved estimering.

Def. 6.4.1 La $Y_1, \ldots, Y_n$ være tilfeldig utvalg fra $f_Y(y; \theta_1, \ldots, \theta_k)$. GLR, skrevet med Λ, defineres som

$$\Lambda = \frac{\max_{\omega} L(\theta_1, \ldots, \theta_k)}{\max_{\Omega} L(\theta_1, \ldots, \theta_k)} = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

NB! Λ er en tilfeldig variabel som en funksjon av det tilfeldige utvalget.

Def. 6.4.2 En generalisert-likelihood-ratio test (GLRT) er en som forkaster H_0 når

$$0 < \lambda \leq \lambda^*$$

hvor y^* velges slik at $P(0 < \Lambda \leq \lambda^* \mid H_0 \text{ er korrekt}) = \alpha$

Dersom Λ var nær 0, ville det være naturlig å forkaste H_0. Om distribusjonen under H_0 var kjent, $f_\Lambda(\lambda \mid H_0)$, kan man finne y^* ved å løse

$$\alpha = \int_0^{y^*} f_\Lambda(\lambda \mid H_0)\, d\lambda$$

Som oftest er ikke dette kjent. Da må vi vise at Λ er en monoton funksjon av størrelse W, hvor fordelinga av W *er* kjent. Når vi har funnet en slik statistikk, enhver test basert på W er ekvivalent med en basert på Λ.

7 Kapittel 7—Normalfordelinga

7.2 Punktestimater for μ og σ^2

Teorem 7.2.1 La $Y_1, \ldots, Y_n$ være t.v. fra $\mathrm{N}(\mu, \sigma^2)$. MLE for parametrene er:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Teorem 7.2.2 Gitt $Y_1, \ldots, Y_n$ være t.v. fra $\mathrm{N}(\mu, \sigma^2)$, $\bar{Y}$, MLE estimatorene er sentrerte, effisiente og konsistente.

Teorem 7.2.3 Gitt $Y_1, \ldots, Y_n$ være t.v. fra $\mathrm{N}(\mu, \sigma^2)$, er følgende sentrert og konsistent (brukes til tilnærme σ^2)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

7.3 Lineære kombinasjoner av normalfordelte variabler

Lemma Om $Y_1 \sim N(0, \sigma^2)$ og $Y_2 \sim N(0, 1)$, hvor Y_1 og Y_2 er uavhengige, er:

$$Y_1 + Y_2 \sim N(0, \sigma^2 + 1)$$

Teorem 7.3.1 Om $Y_1 \sim N(\mu_1, \sigma_1^2)$ og $Y_2 \sim N(\mu_2, \sigma_2^2)$, hvor Y_1 og Y_2 er uavhengige, er:

$$Y_1 + Y_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Korollar Om $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ er snittet av n uavhengige $N(\mu, \sigma^2)$ t.v., så er $\bar{Y} \sim N(\mu,\ \sigma^2\ /\ n)$.

7.4 Sentral-grense-teoremet

En sum av mange "ikke-normale" variabler, er tilnærmet *normalfordelinga*.

Sentral-grense-teoremet La $Y_1, Y_2, \ldots$ være uendelig sekvens av t.v. med samme fordeling (uansett type). Anta at snittet μ og variansen σ^2 til $f_Y(y)$ er endelige. For c og d, er:

$$\lim_{n\to\infty} P\left(c < \frac{Y_1 + \cdots + Y_n - n\mu}{\sqrt{n}\sigma} < d\right) = \frac{1}{\sqrt{2\pi}}\int_c^d e^{-(1/2)y^2}\,dy$$

Lemma Anta at $Y, Y_1, Y_2, \ldots$ er t.v. med $\lim_{n\to\infty} M_{Y_n}(t) = M_Y(t)$ for alle t i et intervall rundt 0. Da er $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$ for alle reelle verdier y.

8 Hot Vault

- $\displaystyle\int_0^\infty y^k \frac{1}{\theta} e^{-y/\theta}\,dy = k!\theta^k$
- $\displaystyle\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda}$

- Summen av n bernoulli-forsøk er binomialfordelt.

- $\sum_{n=0}^{\infty} a_0 r^n = \frac{a_0}{1-r}$

- $\int uv' = uv - \int vu'$

- Konfidensintervall med uniform fordeling $W = ((n+1)/n) \cdot Y_{\max}$. W går fra 0 til $((n+1)/n)\theta$. $f_W(w) = \frac{n^{n+1} w^{n-1}}{(n+1)^n \theta^n}$. Metode: Finn grenser for W. Integrer for å finne a og b. Deretter, løs ut slik at man estimatoren θ i midten av ulikheten.

$$P(a < W < b) = 0.90.$$
$$\int_0^a f_W(w)\,dw = 0.05 \iff a = \sqrt[n]{0.05}\left(\tfrac{n+1}{n}\right)\theta.$$
$$\text{b finnes ved } \int_b^{((n+1)/n)\theta} f_W(w)\,dw = 0.05.$$
$$P(\sqrt[n]{0.05}\left(\tfrac{n+1}{n}\right)\theta < W < \sqrt[n]{0.95}\left(\tfrac{n+1}{n}\right)\theta) = 0.90 \iff$$
$$P(\frac{W}{\sqrt[n]{0.95}\left(\frac{n+1}{n}\right)\theta} < \theta < \frac{W}{\sqrt[n]{0.05}\left(\frac{n+1}{n}\right)\theta}) = 0.90 \rightarrow$$

$$\left(\frac{w}{\sqrt[n]{0.95}\left(\frac{n+1}{n}\right)}, \frac{w}{\sqrt[n]{0.05}\left(\frac{n+1}{n}\right)}\right)$$

- Uniformfordelinga med parameter θ innenfor intervall (a, b), har: $f_U(x) = \frac{1}{\theta}, F_U(x) = \frac{x}{\theta}$ og $E(U) = \frac{a+b}{2}$.

- Oversikt over noen fordelinger.

Fordeling	**Forventning**	**Varians**
Binomisk	np	np(1-p)
Poisson	λ	λ
Eksponensiell	β	β^2
Normal	μ	σ^2
Geometrisk	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Neg.binom	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Normal uniform	1/2	1/12
Gamma	$\frac{r}{\theta}$	$\frac{r}{\theta^2}$