"FASIT" laget på tavleøv. 29.10.2010 ML

TMA4240/TMA4245 Statistikk

Quiz til Kapittel 9: Estimering.

1) HVA ER EN ESTIMATOR?

1. En funksjon av en eller flere stokastiske variabler.
2. En stokastisk variabel.
3. Et uttrykk som gir et anslag for en parameter.
4. Et tall. Est. mat

2) EGENSKAPER TIL ESTIMATOR Hvilke egenskaper ønsker vi at en estimator skal ha? Vi ser på $\hat{\theta}$ som er en estimator for θ.

1. Forventingsrett, dvs. $E(\hat{\theta}) = \theta$.
2. Minst mulig forventning, dvs. $\min E(\hat{\theta})$.
3. Minst mulig skjevhet, dvs. $\min(E(\hat{\theta}) - \theta)$.
4. Vi vil at hvis vi repeterer forsøket vårt m ganger (der m er stor) og hver gang lager $\hat{\theta}_i$, $i = 1, ..., m$, så skal gjennomsnittet over estimatorene vi har laget, $\frac{1}{m}\sum_{i=1}^{m}\hat{\theta}_i$ være nært den sanne parameteren θ.
5. Størst mulig varians, dvs. stor $\text{Var}(\hat{\theta})$. Nei, minst mulig.
6. Vi vil at variansen til estimatoren skal være uavhengig av antall observasjoner som estimatoren er basert på. Nei, variens bør minke når antall obs. øker.

3) FORVENTNINGSRETTHET La X_1 og X_2 være uavhengige stokastiske variable slik at X_1 er $N(\mu, \sigma^2)$ og X_2 er $N(2\mu, \sigma^2)$. Hvilke(n) av de følgende estimatorer er forventningsrett(e) estimator(er) for μ?

1. $\frac{1}{2}(X_1 + X_2)$ $E(\cdot) = \frac{3}{2}\mu$ Hva med Var?
2. $\frac{1}{3}(X_1 + X_2)$ $\frac{1}{3}(\mu + 2\mu)$
3. $\frac{1}{5}(X_1 + 2X_2)$ $\frac{1}{5}(\mu + 4\mu)$
4. $\frac{1}{4}(2X_1 + X_2)$ $\frac{1}{4}(2\mu + 2\mu)$

4) SAMMENLIGNING AV TO ESTIMATORER Vi ser på $\hat{\theta}_1$ og $\hat{\theta}_2$ som er to estimatorer for θ. Hvilken estimator vi du foretrekke hvis $E(\hat{\theta}_1) = E(\hat{\theta}_2)$ og $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$?

1. $\hat{\theta}_1$
2. $\hat{\theta}_2$

5) SANNSYNLIGHETSMAKSIMERING Vi har uavhengige observasjoner $x_1, x_2, \ldots, x_n$ fra en sannsynlighetfordeling $f(x;\theta)$. I hvilken rekkefølge skal følgende operasjoner gjøres?

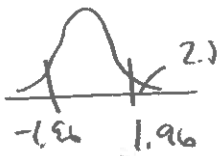
- 2) Under regningen tar vi gjerne log av rimelighetsfunksjonen kun for å lette regningen.
- 1) Rimelighetsfunksjonen $L(x_1, x_2, \ldots, x_n, \theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$.
- 5) Vi kan sjekke at den andre-deriverte til log-rimelighetsfunksjonen er negativ for å se at vi har funnet et maksimum og ikke et minimum.
- 3) Vi deriverer logaritmen til rimelighetsfunksjonen mhp θ og setter lik 0.
- 4) Vi løser ut $\frac{\partial \log L}{\partial \theta} = 0$ og får $\hat{\theta}$ som sannsynlighetsmaksimeringsestimatoren.

Hvilke av følgede påstander er korrekte? sannsynlig → distret

1. $L(x_1, x_2, \ldots, x_n, \theta)$ forteller oss hvor rimelig det vi har observert er, for en gitt verdi av θ.
2. Sannsynlighetsmaksimeringsestimatoren finner vi ved å minimere rimelighetsfunksjonen mhp θ. max
3. Sannsynlighetsmaksimeringsestimatoren vil alltid være forventningsrett. ← asympt. forv.rett
4. I normalfordelingen er $\bar{X}$ sannsynlighetsestimatoren (SME) for μ og $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ sannsynlighetsestimatoren (SME, max) for σ^2.
   $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ er SME
5. Vi kan ikke finne sannsynlighetsmaksimeringsestimatoren til en parameter hvis vi ikke kjenner den parametriske formen på fordelingen $f(x;\theta)$.

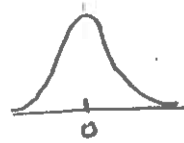
6) KONFIDENSINTERVALL

1. Et 95% konfidensintervall for en parameter er et intervall der vi har 95% tillit til at den sanne verdien for parameteren ligger.
2. For et konkret datasett vil den sanne parameteren ligge innenfor eller utenfor intervallet.
3. Utfører vi forsøket 100 ganger og hver gang lager vi et 95% konfidensintervall for en parameter, vil vi i gjennomsnitt finne den sanne parameteren i 95 av konfidensintervallene vi har laget.
4. Det er 95% sannsynlighet for at den sanne verdien for parameteren ligger i et 95% konfidensintervall.
5. Et 95% konfindensintervall er ~~bredere~~ SMALERE enn et 99% konfidensintervall.



6. Kvantiler i normalfordelingen kaller vi z_α, der $P(Z \geq z_\alpha) = \alpha$. ✓
7. Siden χ^2-fordelingen er symmetrisk vil $\chi^2_{(1-\alpha)} \neq \chi^2_\alpha$.

8. Siden t-fordelingen er ~~ikke er~~ symmetrisk vil $t_{(1-\alpha)} \neq -t_\alpha$.



9. Vi finner et konfidensintervall for en parameter ved å starte med en testobservator med kjent fordeling (der testobservatoren inneholder parameteren) og setter testobservatoren mellom to kvantiler i denne fordelingen og løser ut slik at parameteren havner alene i midten.

7) KONFIDENSINTERVALL FOR μ MED KJENT σ^2 La $X_1, X_2, \ldots, X_n$ være uavhengige og normalfordelte, $N(\mu, \sigma^2)$, der σ^2 er et kjent tall. Hva er konfidenskoeffisientene for følgende konfidensintervaller for μ:

1. $[\hat{\mu} - \frac{\sigma}{\sqrt{n}}, \hat{\mu} + \frac{\sigma}{\sqrt{n}}]$ .. 68%
2. $[\hat{\mu} - 2\frac{\sigma}{\sqrt{n}}, \hat{\mu} + 2\frac{\sigma}{\sqrt{n}}]$ .. 95%
3. $[\hat{\mu} - 3\frac{\sigma}{\sqrt{n}}, \hat{\mu} + 3\frac{\sigma}{\sqrt{n}}]$ .. 99.7%

8) PREDIKSJONSINTERVALL

1. Et 95 % prediksjonsintervall for en ny observasjon er et intervall som med 95% sannsynlighet vil inneholde den nye observasjonen.
2. Et prediksjonsintervall er ~~smalere enn et~~ BREDERE konfidensintervall.
3. Et prediksjonintervall kan brukes til å sjekke om en observasjon er en "outlier".

9) INFERENS OM μ MED KJENT σ $X_1, X_2, \ldots, X_n$ u.i.f. normalfordelte variabler med $E(X) = \mu$ og $\text{Var}(X) = \sigma^2$. Vi ser på testobservatoren $U = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$, der σ^2 er kjent. Ring rundt korrekt(e) alternativ i []-parenteser for hvert delpunkt.

1. Vi jobber med [ett]..[to] utvalg.
2. U er [normal]...[kjikvadrat]...[t]-fordelt.
3. ~~Et $(1-\alpha)100\%$ konfiden~~sintervall for μ er $[\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$..$[\bar{X}-t_{\alpha/2,n-1}\frac{\sigma}{\sqrt{n}}, \bar{X}+t_{\alpha/2,n-1}\frac{\sigma}{\sqrt{n}}]$...$[\bar{X}-\chi^2_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}+\chi^2_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}]$

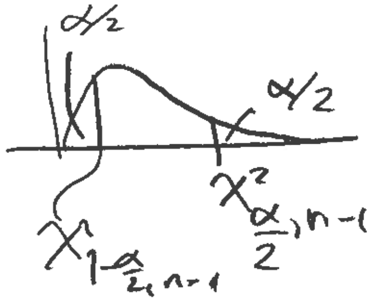
10) INFERENS OM μ MED UKJENT σ $X_1, X_2, \ldots, X_n$ u.i.f. normalfordelte variabler med $E(X) = \mu$ og $\text{Var}(X) = \sigma^2$. Vi ser på testobservatoren $U = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$, der er $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Ring rundt korrekt(e) alternativ i []-parenteser for hvert delpunkt.

1. Vi jobber med [ett]..[to] utvalg.
2. U er [normal]...[kjikvadrat]...[t]-fordelt.
3. Et $(1-\alpha)100\%$ konfidensintervall for μ er $[\bar{X}-z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X}+z_{\alpha/2}\frac{S}{\sqrt{n}}]$..$[\bar{X}-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X}+t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}]$..$[\bar{X}-\chi^2_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X}+\chi^2_{1-\alpha/2}\frac{S}{\sqrt{n}}]$

11) INFERENS OM σ^2 . $X_1, X_2, \ldots, X_n$ u.i.f. normalfordelte variabler med $E(X) = \mu$ og $\text{Var}(X) = \sigma^2$. Vi ser på testobservatoren $U = \frac{(n-1)S^2}{\sigma^2}$, der er $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Ring rundt korrekt(e) alternativ i []-parenteser for hvert delpunkt.

1. Vi jobber med [ett]..[to] utvalg.

$\frac{(n-1)S^2}{\sigma^2}$ $(n-1)$



2. U er [normal]...[kjikvadrat]...[t]-fordelt.
3. Et $(1-\alpha)100\%$ ~~konfidensintervall for~~ σ^2 er
   $[\frac{(n-1)S^2}{z_{\alpha/2}}, \frac{(n-1)S^2}{-z_{\alpha/2}}]$..$[\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}]$..$[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}}]$

12) TO UTVALG, μ_1 OG μ_2 Vi baserer oss på to utvalg. For det første utvalget: $X_1^{(1)}, X_2^{(1)}, ..., X_{n_1}^{(1)}$ u.i.f. $N(\mu_1,\sigma_1)$, med estimatorer $\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ og $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i^{(1)} - \overline{X}_1)^2$ For det andre utvalget: $X_1^{(2)}, X_2^{(2)}, ..., X_{n_2}^{(2)}$ u.i.f. $N(\mu_2,\sigma_2)$, med estimatorer $\overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_j$ og $S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(X_j^{(2)} - \overline{X}_2)^2$ Ring rundt korrekt(e) alternativ i []-parenteser for hvert delpunkt.

1. Vi ser på testobservatoren $U = \frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ når σ_1 og σ_2 er kjente.
   Testobservatoren U er [normalfordelt]...[kjikvadratfordelt med n_1+n_2-2 frihetsgrader]...[t-fordelt med n_1+n_2-2 frihetsgrader].
2. Vi ser på testobservatoren $U = \frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{1/n_1+1/n_2}}$ når $\sigma_1 = \sigma_2$, men ukjente.
   S_p er $[\frac{S_1^2+S_2^2}{2}][\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}]$
   Testobservatoren U er
   [normalfordelt]...[kjikvadratfordelt med n_1+n_2-2 frihetsgrader]...[t-fordelt med n_1+n_2-2 frihetsgrader].
3. Vi ser på testobservatoren $U = \frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{S_1^2/n_1+S_2^2/n_2}}$ når $\sigma_1 \neq \sigma_2$ og ukjente. Testobservatoren U er tilnærmet
   [normalfordelt]...[kjikvadratfordelt med ν frihetsgrader]...[t-fordelt med ν frihetsgrader]. Her er ν avhengig av n_1, n_2, S_1^2, S_2^2.
4. Vi ser på testobservatoren $U = \frac{\overline{D}-(\mu_1-\mu_2)}{S_d/\sqrt{n}}$, der $D_i = X_i^{(1)} - X_i^{(2)}$, $\overline{D} = \frac{1}{n_1}\sum_{i=1}^{n} D_i$ og $n_1 = n_2 = n$ og $S_d = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$
   Denne testobservatoren bruker vi når $[n_1 = n_2]$...$[S_1^2 = S_2^2]$...[observasjonene $X_i^{(1)}$ og $X_i^{(2)}$ er gjort på samme individ (pasient, kornåker) og dermed er korrelerte]

13) EN ANDEL p X er binomisk (n,p). Ring rundt korrekt(e) alternativ i []-parenteser for hvert delpunkt.

1. I en binomisk fordeling er $E(X) = np$. Vi lar $\hat{p} = \frac{X}{n}$. Hva er da $E(\hat{p})$?
   $[p]$..$[\frac{p}{n}]$..$[np]$
2. I en binomisk fordeling er $\text{Var}(X) = np(1-p)$. Vi lar $\hat{p} = \frac{X}{n}$. Hva er da $\text{Var}(\hat{p})$?
   $[p(1-p)]$..$[\frac{p(1-p)}{n}]$..$[np(1-p)]$ $\text{Var}(\hat{p}) = \frac{1}{n^2}\cdot\text{Var}(X) =$
3. Tilnærming til normalfordeling kan benyttes hvis
   [n er stor]..$[np \geq 5]$...$[n(1-p) \geq 5]$...[p ikke er for nært 0]..[p ikke er for nært 1]..[p ikke er for nært 0.5]

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

X bin $\approx N$