Viðskipta- og Hagfræðideild
Háskóli Íslands

Tölfræði II, fyrirlestur 7

Helgi Tómasson

Misserið hálfnað, fyrirlestraáætlun fyrir fyrirlestar 7-12

- 16. október Spanos, k12, k13, k14.
- 23. október Spanos k14, k15, Thomas k12
- 30. október Thomas k13-k15, sbr Spanos K8.
- 6. nóvember Thomas K13-15.
- 13. nóvember Upprifjun
- 20. nóvember, meiri upprifjun, síðasti reglulegi fyrirlestur.

Meiri ályktunarfræði

- Gerður er greinarmunur á „finite-sample" eiginleikum metla og „asymptótískum" eiginleikum (þ.e. þegar n er stórt).
- Í finite-sample vill maður lítinn mean-square-error, helst unbiased.
- Í stórum úrtökum vill maður að estimator sé, consistent og efficient.
- Ef ákveðin skilyrði eru uppfyllt (regularity conditions) þá er maximum-likelihood metill „bestur" í þeim skilningi að hann er consistent og varíansinn nær Cramer-Rao mörkunum í stórum úrtökum. Í bók er asymptótísku mörkin táknuð $CR_\infty(\theta)$ (Ég er ekki alveg sáttur við táknmál bókar hér).

  $$c_n(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} N(0, V_\infty(\theta))$$

  $$\frac{V_\infty(\theta)}{c_n} \quad \text{er kallaður asymptótískur varíans } \hat{\theta}_n$$

  Þ.e.

  $$\frac{V_\infty(\theta)}{c_n} = I_\infty(\theta) = \lim_{n\to\infty}(\frac{1}{c_n^2} I_n(\theta))$$

- Aðalatriði er að skilja að fyrir normal líkan, $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma)'$, og iid úrtak $X_1, \ldots, X_n$ þá er::

$$I_n(\theta) = n \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^4} \end{bmatrix}$$

- Takið vel eftir mynd 13.3.
- Í kafla 13 er rakin saga ályktunartölfræði í vísindasögulegu tilliti. Áður fyrir var gekk leit að líkani út á að finna líkan sem lýsti gögnum vel. Nú er hin ríkjandi menning að skilgreina líkanið fyrirfram, meta það og framkvæma kenningaprófanir. Berið saman myndir 13.2 og 13.3.
- Í kafla 13 á að kunna vel allt um MLE. Einnig á að skilja „least-squares" og hvernig moment er notuð.
- Skoðið dæmi um method-of-moment útleiðslur.

Um kenningaprófanir

- Kafli 14 hefst á sögulegu yfirliti.
- Nútímanálganir er þróaðar af Neyman,Pearson og Fischer.

Grunngangur í kenningaprófun, gefið úrtak, $\boldsymbol{X} = (X_1, \ldots, X_n)'$:

1. Set fram núllkenningu, H_0, um óþekktan parameter.
2. Finn prófstærð $\tau(\boldsymbol{X})$, reikna dreifingu hennar gefið H_0 rétt.
3. Reikna p-gildi$=P(\tau(\boldsymbol{X}) > \tau(\boldsymbol{x}))$ gefið H_0 rétt.

- Hvað svo? Hafna kenningu ef p er lítið. Þ.e. ef ótrúlega útkoma hefur komið upp. Neyman-Pearson leggja til að leita skuli að „góðu" prófi. Þ.e. einhverju sem stendur sig vel gegn áhugaverðum valkostum.
- Einföld (simple) kenning er kenning sem skilgreinir dreifingu algerlega, t.d. $\mu = 0$, σ þekkt í normaldreifingu. Samsett (composite) kenning leyfir mörg gildi, t.d. $\mu > 0$.
- Villa I=hafna réttri H_0. Villa II=hafna ekki rangri H_0.
- Significance-level=marktæknistig=α,skilgreinir reglu um hvenær skuli hafna H_0. H_0 er hafnað ef $p < \alpha$.
- Power-fall=$\mathcal{P}(\theta)$=líkur á að hafna H_0 sem fall af θ.
- Próf, τ^* er UMP=„Uniformly-Most-Powerful" ef

  $$\mathcal{P}(\tau^*, \theta), \theta \in \Theta_1 > \mathcal{P}(\tau, \theta), \theta \in \Theta_1$$
  $$\text{fyrir öll } \tau$$

  þ.e. ef ekki er að hægt að finna próf sem líklegra er til að hafna H_0 gefið H_1 rétt.

- Próf er *consistent* ef power-fall stefnir á 1.

- Hvernig á að finna UMP próf?
- Likelihood ratio:

$$\Lambda(\boldsymbol{X}) = \frac{\max_{\theta \in \Theta} L(\theta, \boldsymbol{X})}{\max_{\theta \in \Theta_0} L(\theta, \boldsymbol{X})}$$

- Likelihood-ratio-próf: Hafna H_0 ef:

$$\Lambda(\boldsymbol{x}) > k$$

  Athugið að í sumum bókum er hlutfallinu víxlað.
- Neyman-pearson lemma: UMP-próf byggir á likelihood ratio.
- Skoðum dæmi bls. 712 í bók.

Asymptótísk próf

- Likelihood-Ratio (LR) próf, ef n er stórt og H_0 rétt þá gildir (asymptótískt):

  $$-2(log(L(\hat{\theta}_0)) - log(L(\hat{\theta}_1))) \sim \chi^2(r)$$

  þar sem $\hat{\theta}_0$ er ML-mat á θ gefið H_0 rétt, $\hat{\theta}_1$ er ML mat á θ gefið H_1 rétt og
  r=fjöldi parametra í H_1 - fjöldi parametra í H_0(= fjöldi hliðarskilyrða)
- LR próf eru best í ýmsum skilningi (sbr. ML-metla).
- Hvers vegna ekki að nota þau alltaf?
- Ýmsar tæknilegar ástæður. Meta þarf líkanið bæði undir H_0 og H_1. Þ.e. það þarf að framkvæma tvær hámarkanir, eina undir H_0 (með hliðarskilyrðum) og aðra undir H_1 (án hliðarskilyrða).
- Stundum er erfitt að hámarka með hliðarskilyrðum. Þá er ein hugmynd (Wald-próf) að reikna:

  $$(\theta_0 - \hat{\theta}_1)^2 I_n(\theta_0) \sim \chi^2(r)$$

  þar sem $I_n(\theta_0)$ er Fischer information gefið $\theta = \theta_0$.
- Ef auðvelt er að framkvæma skilyrtu hámörkunina en óskilyrta hámörkun erfið má reikna score-test (Lagrange-multiplier-test):

  $$s(\boldsymbol{x}) = sc(\theta_0, \boldsymbol{x})^2 * I_n^{-1}(\theta_0) \sim \chi^2(r)$$

- Ef báðar hámarkanir erfiðar eru stundum reynd próf sem eru kölluð Hausman próf.

Lesefni í Spanos

- Kafli 1 og 2.1-2.2 almenn atriði ca. 40 bls.
- Kafli 3.4-3.7+appendix, Grunnatriði um hendingar ca. 45 bls.
- Kafli 4.1-4.7 Meira um líkindafræði, random úrtak ca. 35 bls.

- Kafli 5.1-5.7. Líkindafræði, gögn og grafísk tól, ca. 65 bls.
- Kafli 6.1-6.4. Um non-random úrtök, ca. 20 bls.
- Kafli 7.1-7.2 um regression (líkindafræðilegt), ca. 20.bls.
- Kafli 8.1-8.11 Um slembiferli (stochastic processes), ca. 60. bls.
- Kafli 9.6.1 + 9.9 Markgildi. Aðallega CLS og samleitnihugtök, ca. 10. bls.
- Kafli 10.1-10.2 Túlkanir á líkum. ca. 8 bls.
- Kafli 11.1-11.3 grunnhugtök í ályktunarfræði og 11.6-11.7 úrtaksdreifingar, ca. 25 bls.
- Kafli 12.1-12.5 Eiginleikar punktmetla ca. 25 bls.
- Kafli 13.1-13.5 Matsaðferðir, ca. 40. bls.
- Kafli 14.1-14.5 Kenningaprófanir, ca. 45 bls.
- Kafli 15 Ýmis próf um forsendur líkans. 40