Ályktunarfræði

- Síðast byrjað á „inference" teoríu.
- Þegar hugtökin í ályktunarfræði eru kynnt þá er venjulega gengið út frá því að mælingarnar:

  $x_1, \ldots, x_n$
  hafi fengist með random-úrtaki (iid), þ.e. gildi á:
  $X_1, \ldots, X_n$

  Þetta er vegna þess að líkindafræðin fyrir margvíða hendingu er einfaldari fyrir iid en annars.

Grunneiginleikar punktmetla

Gengið er út frá að mælingar, $x_1, \ldots, x_n$ séu gildi á $X_1, \ldots, X_n$.

- Metill er vörpun, h, frá úrtaksrúmi $\mathcal{X}$ yfir í parameterrúm Θ.

  $h : \mathcal{X} \to \Theta$

  $\hat{\theta} = h(X_1, \ldots, X_n)$ er kallað, metill fyrir θ

  $\hat{\theta} = h(x_1, \ldots, x_n)$ er kallað, mat á θ

- Það er áhugavert að geta sagt eitthvað um dreifingu:

  $\hat{\theta} = h(X_1, \ldots, X_n)$

- Við skilgreinum hugtakið *bias* sem:

  $bias(\hat{\theta}) = E(\hat{\theta}) - \theta$

  Flestir telja það æskilegan eiginleika metils (matsaðferðar) að $bias(\hat{\theta})$ sé lítil tala.

- Einnig er æskilegt að:

  $$V(\hat{\theta})$$

  sé lítil tala.

- Þessar tvær kröfur eru stundum sameinaðar með því að krefjast þess að:

  $$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \quad MSE = \text{Mean-Square-Error}$$

  sé lítil tala.

- Metill er sagður *unbiased* ef $bias(\hat{\theta}) = 0$.

- Það getur verið erfitt eða ómögulegt að finna unbiased estimator.

- Ef bornir eru saman tveir unbiased metlar, $\hat{\theta}$ og $\hat{\upsilon}$ og

  $$V(\hat{\theta}) < V(\hat{\upsilon})$$

  þá er sagt að $\hat{\theta}$ nýti upplýsingar betur en $\hat{\upsilon}$. Þ.e. meira „efficient".

- Æskilegt er að nota aðferðir sem nýta gögn sem best. T.d. er verðugt markmið að leita metils, $\hat{\theta}$, sem er *minimum-mean-square-error.*

- Metill $\hat{\upsilon}_1$ sem er þannig að:

  $$MSE(\hat{\upsilon}_2) < MSE(\hat{\upsilon}_1) \quad \text{fyrir öll gildi í parameterrúmi}$$

  er sagður *inadmissible.*

Hvað gerist með vaxandi úrtaksstærð?

- Við segjum að $\hat{\theta}_n = h(X_1, \ldots, X_n)$ sé *consistent* ef

$$\hat{\theta}_n \xrightarrow[n\to\infty]{p} \theta \quad \text{þ.e.}$$
$$\lim_{n\to\infty} P(|\theta_n - \theta| < \varepsilon) = 1 \quad \text{fyrir öll } \varepsilon > 0$$

- Þetta er stundum kallað *weak-consistency*. E.t.v. er auðveldara að skilja hugtakið *strong-consistency* skilgreint sem:

$$MSE(\hat{\theta}_n) \xrightarrow[n\to\infty]{} 0 \quad \text{þ.e.}$$
$$bias(\hat{\theta}_n)^2 \xrightarrow[n\to\infty]{} 0 \text{ og } V(\hat{\theta}_n) \xrightarrow[n\to\infty]{} 0$$

- Sumir (Spanos) kalla strong consistency:

$$\hat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta$$

- Athugið að þegar talað eru um „weak" (t.d. weak-law-of-large-numbers) þá er átt við samleitni í líkum eða í dreifingu. Þegar talað eru um strong þá er átt við „almost-surely" eða „mean-square" samleitni. Sjá mynd í lemmu 5, bls. 509.

Cramer-Rao ójafnan. Hversu miklar upplýsingar er hægt að kreista út úr úrtaki?

- Fyrir gefna úrtakkstærð eru takmörk fyrir því hversu nákvæmlega er hægt að álykta um θ.
- *Fischer information*:

$$
\begin{aligned}
&L(\theta|x_1,\ldots,x_n) = f(x_1,\ldots,x_n;\theta) \quad \text{likelihood fall}\\
&l(\theta|x_1,\ldots,x_n) = log(L(\theta|x_1,\ldots,x_n))\\
&sc(\theta,\boldsymbol{X}) = \frac{\partial l(\theta|X_1,\ldots,X_n)}{\partial\theta} = \text{ score function}\\
&\text{ hægt er að sýna að } E(sc(\theta,\boldsymbol{X})) = 0\\
&I_n(\theta) = E(sc(\theta,X)^2)
\end{aligned}
$$

- Ef parameterinn er k-víður þá er:

$$
\begin{aligned}
&sc(\boldsymbol{\theta},\boldsymbol{X}) \quad \text{k-víður vektor og}\\
&I_n(\boldsymbol{\theta}) = E(sc(\boldsymbol{\theta},\boldsymbol{X})sc(\boldsymbol{\theta},\boldsymbol{X})') \quad \text{kxk vítt fylki}
\end{aligned}
$$

- Ef metill $\hat{\theta}$ (eða $\hat{\boldsymbol{\theta}}$) er unbiased þá gildir að:

$$
\begin{aligned}
&V(\hat{\theta}) > 1/I_n(\theta) \quad \text{og}\\
&V(\hat{\boldsymbol{\theta}}) - I_n^{-1}(\boldsymbol{\theta}) \quad \text{er pósítívt definite fylki}
\end{aligned}
$$

- Metill, $\hat{\theta}_n$, er sagður CAN (*consistent-asymptotically-normal*) ef:

$$
\begin{aligned}
&c_n(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} N(0, V_\infty(\theta))\\
&\frac{V_\infty(\theta)}{c_n} \quad \text{er kallaður asymptótískur varíans } \hat{\theta}_n\\
&\text{ athugið ritháttt í bók á bls. 618}
\end{aligned}
$$

- CAN metill er sagður *asymptotically efficient* ef:

$$
c_n(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{d} N(\mathbf{0}, I_n^{-1}(\boldsymbol{\theta}))
$$

þ.e. nær Cramer-Rao mörkunum stórum úrtökum.

- Það má sleppa, jacknife (athyglivert hugmynd).
- Lesa lauslega um sufficiency. Ef maður ætlar að lesa um sufficiency þá er factorization theorem grundvallar atriði. Góðir metlar eiga helst að vera föll af „sufficient-statistic”.

Hvernig eru metlar fundnir?

- Nokkrar meginaðferðir:
  1. Method of moments
  2. Rúmfræðilegar aðferðir (least-squares)
  3. Maximum-likelihood aðferðir
  4. Bayes aðferðir

  Aðferðir 3 og 4 byggja báðar á sennileikafallinu (likelihood-fallinu).

- Í Spanos er góð umræða um hvernig þessar hugmyndir þróuðust. Málfræðilegan misskilning ofl.

- MM (Method-of-moments) gengur út á að reikna fræðileg móment sem föll af óþekktum parameter og leysa síðan jöfnuna „fræðilegt-móment=úrtaks-móment" fyrir parameterinn θ. Dæmi:

  $$X_i, i = 1, \ldots, n, \quad \text{iid } E(X_i) = \theta.$$
  $$\text{úrtaksmómentið er } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
  $$\text{þar sem}$$
  $$E(\bar{X}) = \theta \quad \text{leysir maður jöfnuna}$$
  $$\theta = \bar{X}$$

  Sjá töflu 13.2 í Spanos.

- Lærið að nota litla og stóra o bls. 644. o(x) O(x).

- Það þarf ekki að kunna töflurnar í kafla 13.

- Takið eftir notkun á Taylor reglu (bls. 647) fyrir

  $$E(g(\hat{\mu})) \text{ og } V(g(\hat{\mu}))$$

- Least-squares aðferð gengur út á að finna punkt (punkta) sem nálgast mæld gildi í rúmfræðilegum skilningi.
- Dæmi:

  $x_i, i = 1, \ldots, n$ mælingar á iid $X_i, E(X_i) = \theta$

  finnum það gildi á θ sem lágmarkar

  $$\sum_{i=1}^{n} (x_i - \theta)^2$$

- Setning Gauss-Markov segir að LS aðferðir séu BLUE (best-linear-unbiased-estimator). Það verður heimadæmi um Gauss-Markov.

Maximum-likelihood

- Aðalaðferð nútímans.
- Hugmyndin er að finna það gildi á θ sem hefur mælda útkomum sem sennilegasta.
- Stærðfræðilega felst þetta í að leysa eftirfarandi hámörkunarvandamál:

  $$\max_{\theta \in \Theta} l(\boldsymbol{\theta}|x_1, \ldots, x_n)$$

  lausnin á þessu hámörkunarvandamáli er oft rituð

  $$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|x_1, \ldots, x_n)$$

- Þessi hámörkun getur verið erfið.
- Einfalt dæmi:

  $X_i, i = 1, \ldots, n$ eru iid Bernoulli $E(X_i) = \theta$

  sennileikafallið er:$L(\theta|x_1, \ldots, x_n) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{(1-x_i)}$

  $$l(\theta|x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \log(\theta) + (1-x_i)\log(1-\theta)$$

  hámarkað með diffrun

  $$\frac{\partial l(\theta|x_1, \ldots, x_n)}{\partial \theta} = \sum_{i=1}^{n} \frac{x_i}{\theta} - \frac{(1-x_i)}{1-\theta} \quad \text{o.s.frv. gefur } \hat{\theta}_{ML} = \bar{x}$$

Eiginleikar ML aðferða

- Ef ákveðin regularity skilyrði gilda (1-10 í bók) þá er ML-metill er CAN.
- Ef ákveðin regularity skilyrði gilda (1-10 í bók) þá er ML-metill asymptótísk nýtinn (þ.e. nær Cramer-Rao mörkum)