

Upprifjun úr Spanos

- Af hverju tölfræði? Svar: Til að alhæfa út frá mælingum.
 1. Grunnurinn byggir á líkindafræði
 2. Gengið er út frá því að gögn sé mæld gildi á random-breytum (hendingum).
 3. Ályktun byggir alltaf á líkani.
 4. Líkan er byggt á einhverjum fræðum, það þarf að para saman vandmál og líkan.
 5. Líkan byggir á líkindadreifingu.
 6. Til að hægt sé að hanna vitrænt líkan þarf að vera fyrir hendi skilningur á líkindadreifingum.
 7. Venjulega er sett fram líkan sem er þannig að formið er algerlega þekkt, en nokkrar stýritölur (parametrar, stíkar) eru óþekktar og þarf að meta.
 8. Það er því nauðsynlegt þróa matsaðferðir. Matsaðferðir hafa ólíka eiginleika. Við þurfum skilgreina æskilega eiginleika.
 9. Viðmiðun er nauðsynleg. Við þurfum að geta tjáð okkur um fengna útkomu í samanburði við eitthvað annað.
 10. Gott líkan þarf að vera, túlkanlegt, mögulegt (hugsanlegt), get skýrt eitthvað og að vera hrekjanlegt.

- Kafli 2.1-2.2 (9 bls.). The primary aim, why do we care? Hér er lauslegur inngangur að hugsunarhætti. Efni þessa kafla á að vera vel þekkt úr tölfræði I og hagrannsóknnum I. Það þarf að þekkja grunnreglur líkindafræði, óháðir atburðir, skilyrtar líkur, regla Bayes, hending (random-variable), samfelld/discrete hending.
- Kafli 3.4-3.7 (41 bls.) Hér er verið að dýpka skilning á hugtakinu dreifing. Hugtök eins og cdf, pdf, pmf lýsa dreifingu.
- Það má reikna ýmsar lýsitölur til að lýsa dreifingu.
- Moment eru integröl sem lýsa eiginleikum hendinga. Meðaltal og varíans eru þekkt úr fyrri námskeiðum.
- Nýtt (held ég) hér eru mgf og cf. Ef mgf er til eru bein tengsl á milli mgf og cf.
- Takið eftir inversion-theorem á bls. 113. Það segir að vitneskjan um cf sé jafngild vitneskju um cdf/pdf/pmf (í öllum punktum þar sem cdf er samfelld).
- Það þarf að þekkja kurtosis og skewness.
- Einnig þarf að þekkja mode, quantile. CV=coefficient-of-variation má t.d. bera saman við Gini-stuðul, sbr. dæmi á dæmablöðum.
- Ágætt að kíkja yfir dreifingarnar í lok kaflans.

- Kafi 4.1-4.7 (36 bls.) Úrtak er safn hendinga. Gögn=gildi sem úrtak tekur.
- Til að skilja dreifingu úrtaks þarf að skilja hvað er margvið hending og vita hvað er átt við þegar er talað um dreifingu margviðrar hendingar. Við þurfum sérstakt tölfræðilegt líkan við hliðina á fræðilegu líkani sem lýsir dreifingu úrtaks.
- Random úrtak (iid) er sérstaklega meðfærilegt.
- Það þarf að kunna reiknireglur um væntanlegt gildi og varíans.
- Það þarf að þekkja hugtökin conditional-distribution og marginal-distribution.
- Ágætt að átta sig á dæminu um skilyrta normal-dreifingu á blaðsíðu 164, bera það saman við regression og jöfnur 1 og 2 í fyrirlestranótum.
- Hazard-fall er hugtak sem er gagnlegt að skilja.
- Efst á bls. 180 er áhugavert dæmi um fjölskyldu líkindadreifinga þar sem cf hefur þægilegt form en þéttifallið ekki (þarf ekki að kunna).
- Af fjölskyldunum, exponential-family, Pearson-family, Johnson-family er exponential-family mikilvægust. Þetta þarf ekki að kunna.
- Kafi 5 (60 bls.) Mjög léttur. Þetta er um grafísk framsetningu á gögnum með hjálp líkindafræðilegra hugtaka.
- Takið eftir því hvernig má búa til pseudo-random tölur úr ýmsum dreifingum, sbr. töflu á bls. 256.

- Kafi 6.1-6.4 (22 bls.) Tengsl hendinga.
- Hér þarf að átta sig á hvað gerist þegar slakað er á kröfu um að hendingar séu iid.
- Takið eftir trikkinu, „sequential conditioning” á bls.265.
- Takið eftir formúlum á bls. 266 og berið saman við formúlur 1 og 2 í fyrirlestrarnótum.
- Mest notaða tala til að lýsa tengslum hendinga er fylgnistuðullinn.
- Ef hendingar eru óháðar þá er fylgni þeirra 0.
- Ef tvær normaldreifðar hendingar hafa fylgni 0 þá eru þær óháðar.
- Hægt er að búa til hendingar sem eru háðar en hafa fylgni 0, sbr. dæmablað.

- Kafi 7.1-7.2. (20 bls.) Regression-hugtakið
- Þetta gengur út á að reikna:

$$E(Y|X = x) \quad V(Y|X = x)$$

- Takið eftir formúlum á bls. 340, 342 og berið saman við formúlur 1 og 2 í fyrirlestrarnótum.
- Hér þarf að þekka homoskedastic/heteroskedastic
- Dæmin neðst á blaðsíðu 347-356 má reikna með því að nota formúlur á bls. 185-189. Þetta er spurning um að æfa sig í integrationin. Hér dugar að leysa eitt dæmi til að skilja hvað er í gangi.

- Kafi 8.1-8.11 (62 bls.) Stochastic Processes. Mikið efni. Mikið notað í fjármálum og hagrannsóknum.
- Hér er verið að fjalla um safn hendinga sem er raðað í tíma.
- Stochastic processes er grundvallar byggingarefni í empírískri haglíkanagerð.
- Tengsl hendinganna hafa eitthvað með tímann að gera.
- Það þarf að kunna og skilja formúlu 8.2.
- Það þarf að kunna nothæfa skilgreiningu á stochastic process.
- Það þarf að kunna mynd 8.4.
- Það þarf að vita af Brown-hreyfingu (Wiener-ferli), Poisson-ferli, random-walk, white-noise.
- Martingale-eiginleikinn er mikið notaður við haglíkanagerð. Það er til mikil fræði um martingala. Hér dugar að vita grunnskilgreiningu. Það má lesa kafla 8.9 lauslega.
- Takið eftir dependency hugtökunum á bls. 420-421.
- 8.4.3 má sleppa.
- 8.4.4. má lesa lauslega. Það þarf að vita að hugtakið ergodic er til og kunna óformlega skilgreiningu.
- Stationary hugtök á að kunna.
- Exchangeability má sleppa.
- Lesið vel um ARMA.

- Kafli 9 (ca. 10-15). Um markgildi
- Það má lesa 9.2-9.5 lauslega (sleppa?)
- Lesa vel 9.6.1, CLT
- Lesið 9.9 og kunnið myndina á bls. 509.
- Á dæmablaði voru æfingadæmi sem æfa hugtökin.
- Kafli 10 (8 bls.). Brú frá líkindafræði yfir í ályktunarfræði.
- Hér á að átta sig á þeim heimspekilega vanda hvernig á að gefa líkum merkingu.
- Kafli 11 (34 bls.) Inngangur að ályktunarfræði
- Þekkja til lauslega til sögunnar um Fischer og Pearson
- Það þarf að þekkja hugtakið parameterrúm.
- Við höfum nokkrar tegundir ályktana, punktmát, bilmat, kenningprófanir (um óþekkta parameter) og spár (um mælingu sem ekki hefur átt sér stað).
- Varðandi punktmát (point-estimation), þarf að þekkja estimator (metill), estimate(mate).
- Það á að skilja myndina á bls. 570 um hvernig Bayes aðferðafræði virkar.
- Það þarf að skilja vel hugtakið, „sampling distribution”.
- Það þarf að kunna aðferð til að reikna dreifingu falls af hendingu.
- Það má lesa lauslega (sleppa?) kafla 11.8.

- Kafi 12 (24 bls.) Eiginleikar punktmats í endanlegur úrtaki.
- Eiginleikar metla, bias, variance, mean-square-error, efficiency.
- Það þarf að huga að því hvað gerist þegar n stækkar og þekkja consistency hugtakið (strong/weak).
- Það þarf að vita hvað er Fischer-information og Cramer-Rao lower bound.
- Það má lesa lauslega 12.5.3
- Kafi 13 (40 bls.) Meira um punktmat
- Nokkrar aðferðir við að leiða út estimatara
- Moment-aðferðir
- Rúmfræði-aðferðir (least-squares)
- Maximum-likelihood aðferðir
- Bayes aðferðir
- Hér þarf að kunna likelihood-teoríauna.
- Það er gott (en ekki mikilvægt í þessu námskeiði) að átta sig á númerísku aðferðafræðinni á bls. 667.
- Það þarf ekki að kunna Kullback-Leibler.

- Kafi 14 (47 bls.). Kenningaprófanir
- Hér þarf að sjálfsgöðu að þekkja grunnatriðin, H_0 , H_1 , einföld og samsett kenning, villa I, villa II, significance-level, power.
- Samkvæmt likelihood-teoríu gefur ML í ákveðnum skilningi besta mat. Það er ekki til hliðstæða fyrir próf.
- Gott próf er próf sem hefur hátt power. Einn slíkur optimalitetseiginleiki er UMP.
- Neyman-Pearson lemma segir að ef UMP er til þá næst það með LR-prófi.
- Í hagrannsóknnum er mikið notast við asymptótísk próf. LR, Wald og LM.
- Kafi 15 (55 bls.) Ýmis próf um forsendur líkans
- Flestir eiginleikar líkans þurfa að vera prófanlegir
- Við höfum próf um dependence, einsdreifni, form dreifingar, gildi/formerki á parametrum o.s.frv.

- Kafi 12 (30 bls.) í Thomas Val á líkani
- Það er gott að bera þetta saman við kafla 15 í Spanos.
- Það er margt sem getur verið að líkani, vantað breytur, ekki rétt form, ekki rétt dreifing, dependency strúktúr o.s.frv.
- Omitted variable bias á að þekkja og geta leitt út í einföldu regressi-
onlíkani með 2 skýristærðum.
- RESET próf er einfalt próf fyrir formi líkans.
- Það þarf að skilja muninn á general-to-specific og specific-to-general.
- Data-mining hugtakið þarf að þekkja.
- Hvenær á að nota encompassing test.
- Kafi 13 (25 bls.) Meðhöndlun non-stationary raða
- Spurious regression = delluaðhvarf
- ECM=error-correction-model. Fyrst skilja einföld ECM með 2 breyt-
um. Síðan að átta sig á hvernig þetta breytist þegar breytum fjölgar.
- Skilja hvers vegna ECM-formið er vinsælt.

- Kafi 14 (15 bls.) Prófanir á stationarity
- Það þarf að vita hvað „correlogram” (fræðilegt og úrtaks) er.
- Það er gott að þekkja mynstur í correlogrammi fyrir einföld AR, MA og ARMA ferli.
- I(d) formið af non-stationary röðum þarf að þekkja.
- Það þarf að átta sig á hvað difference-stationary (DS) og trend-stationary (TS) þýðir.
- Það þarf að geta notað Dickey-Fuller (DF) próf og ADF.
- Það er hollt að skoða myndir í appendix og reyna að herma þær í eigin tölvu.
- Kafi 15 (25 bls.) Um cointegration og mat á ECM líkönum.
- Granger representation theorem: Raðir cointegreraðar er jafngilt tilvist ECM.
- Lesið um Engle-Granger two-step á bls. 432.
- Eins og alltaf á að skilja vel tveggja breytu vandamálið og athuga síðan hvað gerist ef breytum er fjölgað.
- Kafi 16 Ýmislegt (má lesa lauslega).
- Hollt að lesa appendix um fylkjaalgebru.