Viðskipta- og Hagfræðideild
Háskóli Íslands

Tölfræði II, fyrirlestur 1

Helgi Tómasson

Gróf áætlun

Vika 1-4: Líkindafræði (probability theory), Spanos, K, 1-9.

Vika 5-8: Ályktunarfræði (inference theory), Spanos, K, 11-14.

Vika 9-12: Hagnýtingar, leit að líkani, einföld atriði um tímaraðir, Thomas K, 12-16.

- Kynnið ykkur nokkur hugtök úr kafla 1. Þið eigið eftir að rekast á þau víða.
  - Empírísk líkansmíði (empirical modeling)
  - Slembimekanismi þess sem rannsaka (stochastic nature)
  - Mæld gögn (observed data)
  - Tölfræðilegt líkan (statistical model)
  - Líkansmíð=skilgreining (specification), mat(estimation), prófanir (misspecification, respecification)
  - Skiljið gagnformin, cross-section-data, time-series-data, panel-data.
  - Takið eftir mæliformunum, ratio-scale, interval-scale, ordinal-scale, nominal-scale.
  - Takið eftir meðmælunum 12 varðandi líkanasmíði á bls. 13-29.
- Lesið kafla 1 um empíríska líkanagerð. Hugleiðið hvort hagfræði sé empírísk vísindagrein.

- Líkindafræði er grunnur að tölfræði. Líkindafræði er stærðfræðigrein.
- Tölfræði er ályktunarfræði, þ.e. kerfi sem segir hvernig skuli alhæfa út frá mælingum.
- Tölfræðilegt líkan er grundvöllur ályktunar út frá mælingum.
- Parametrískt líkan

  Líkindafræðilegt líkan $\Phi = \{f(x, \theta), \theta \in \Theta, x \in \mathbb{R}_X\}$
  Úrtakslíkan: $\boldsymbol{X} = (X_1, \ldots, X_n)$ er random úrtak

Stiklað úr grundvelli líkindafræðinnar (kafli 2 og 3 í Spanos)

- Random tilraun: Tilraun þar sem útkoma er óviss.
- S=mengi af mögulegum útkomum úr tilraun (oft táknað Ω).
- S, getur verið endanlegt,teljanlegt, óteljanlegt.
- Atburður=mengi af útkomum.
- Líkindamál P, er fall frá mengi atburða yfir í $[0, 1]$ Ef A er atburður og B er atburður þá er

$$P(A \cup B) = P(A) + P(B) - P(AB), \quad AB = A \cap B$$
$$P(S) = 1$$
$$P(\varnothing) = 0$$

- Atburðir A og B eru sagðir óháðir ef :

$$P(AB) = P(A)P(B)$$

- Mengi leyfilegra atburða $\mathfrak{F}$, verður að vera sigma-algebra (σ-algebra). Þetta er stærðfræðilegt atriði sem tryggir að hægt er að skilgreina líkindamál á útkomumengi með óteljanlegum útkomum. Áhugasamir geta lesið nánar í Spanos og/eða Billingsley.
- Líkur á atburðnum $A|B$, atburðurinn A gefið B er táknaður:

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

- Hending=Random-variable er fall sem úthlutar útkomu úr tilraun tölu:

  $$X : S \to \mathbb{R}$$

  (X þarf að vera mælanlegt, þ.e. $\{s : X(s) < a\} \in \mathfrak{F}$. Stærðfræðilegt atriði sem við hugsum ekki um hér)

- Það er miklu þægilegra að vinna með tölur en útkomur úr tilraun. Stærðfræðilega hefur

  formlega líkindarúminu $(S, \mathfrak{F}, \mathbb{P})$ verið varpað í $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$

  - Hér er $\mathfrak{B}(\mathbb{R})$ er Borel-algebra yfir rauntölurnar=minnsta σ-algebra sem inniheldur bil, $]-\infty, a]$.

- Dreififall X er fallið:

  $$F(x) = \mathbb{P}(-\infty < X(s) < x) = P(X < x)$$

- Hendingin X er sögð sundurslitin (discrete) ef hún getur aðeins tekið teljanlega mörg gildi.

- Hendingin X er sögð samfelld (continuous) ef hún getur tekið öll gildi á einhverju rauntölubili.

- Ef hending X er discrete þá er líkindamassafall X:

  $$f(x) = P(X = x)$$

- Ef hending X er samfelld þá er $f(x)$, þéttifall X:

$$F_X(x) = \int_{-\infty}^{x} f_x(u)du$$
$$f_X(x) = F'_X(x) \quad \text{ef } f \text{ er samfellt í } x$$

- Athugið

$$P(a < X < b) = \int_a^b f_X(u)du = F(b) - F(a)$$

- Ef við þekkjum föllin f_X eða F_X þá segjum við að við þekkjum dreifingu X.

Lýsitölur fyrir dreifingar

- Við viljum gjarnan lýsa dreifingu með tölu
- Við köllum, μ_X, meðaltal hendingarinnar X.

$$\mu_X = E(X) = \int_{-\infty}^{\infty} xf(x)dx \quad \text{fyrir samfelldar hendingar og}$$
$$\mu_X = E(X) = \sum_{-\infty}^{\infty} xf(x) \quad \text{fyrir discrete hendingar}$$

- Við köllum x_q q-kvantíl X ef:

$$F(x_q) = q \quad \text{sér í lagi}$$
$$F(x_{0.5}) = 0.5 \quad \text{þá er } x_{0.5} \text{ median=miðtala } X$$

- Við köllum μ'_r, r-ta móment (vægi?) hendingarinnar X og μ_r r-ta central móment.

$$\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)dx$$
$$\mu_r = E(X - \mu_X)^r$$

- Varíans=dreifni (variance) hendingarinnar X, σ_X^2 er skilgreind:

$$\sigma_X^2 = E(X^2) - \mu_X^2 = E(X - \mu_X)^2$$

- Fleiri hugtök, skewness, kurtosis, leptokurtic, platycurtic.
- Ýmsar dreifingar. Hvers vegna hafa þær nafn?