NTNU
Kunnskap for ei betre verd

Observatorar og utvalsfordeling

Torstein Fjeldstad
Institutt for matematiske fag, NTNU
09.10.2018

I dag

- Til no i emnet
- Observatorar
- Utvalsfordelingar
- Sentralgrenseteoremet

Til no i emnet

— definisjon av hendingar, sannsyn og stokastiske variable

Til no i emnet

- definisjon av hendingar, sannsyn og stokastiske variable
- definisjon sannsynsfordeling, forventning, varians, osb.

Til no i emnet

— definisjon av hendingar, sannsyn og stokastiske variable

— definisjon sannsynsfordeling, forventning, varians, osb.

— viktige fordelingar

- $b(x; n, p)$ binomisk
- $p(x; \lambda)$ Poisson
- $n(x; \mu, \sigma)$ normal
- $f(x; \lambda)$ eksponential

Til no i emnet

- definisjon av hendingar, sannsyn og stokastiske variable
- definisjon sannsynsfordeling, forventning, varians, osb.
- viktige fordelingar
  - $b(x; n, p)$ binomisk
  - $p(x; \lambda)$ Poisson
  - $n(x; \mu, \sigma)$ normal
  - $f(x; \lambda)$ eksponential

Typiske antakingar i spel (kast terning, kortspel, osb.)

Eksempel 1

- Industriell prosess som produserer eit produkt
- Vil anslå andelen defekte artikler, p
- Kontrollerer n artikler og lar X vere talet på defekte

Merknad 1: me vel n, men p er ukjend
Merknad 2: $X \sim b(x; n, p)$ (om n er "stor")

Eksempel 2

- Undersøk levetida X til elektroniske komponentar
- Anta at levetida er eksponentialfordelt ($E(X) = 1/\lambda$, λ ukjend)
- Test n komponenter

$$
\begin{aligned}
X_1 &\sim f(x_1;\lambda) \\
X_2 &\sim f(x_2;\lambda) \\
&\vdots \\
X_n &\sim f(x_n;\lambda)
\end{aligned}
$$

Innhald resten av emnet

1. Punktestimator: eit "godt" gjett på verdien til den ukjende parameteren
2. Intervallestimat/konfidensintervall: finn intervall $[\hat{p}_{\text{lower}}, \hat{p}_{\text{upper}}]$ som dekker den sanne verdien med eit gitt sannsyn[1]

[1] Dette er ei grov forenkling som me vil sjå seinare

Statstisk inferens I

Populasjon

Ein populasjon består av alle moglege observasjonar me kan gjere.

Utval

Eit utval er ei delmengde av ein populasjon.

Tilfeldig utval

Tilfeldig utval (eng: random sample): einingane som trekkes frå populasjonen velges tilfeldig og uavhengig av kvarandre.

Statstisk inferens II

Mål: Trekke konklusjonar om eigenskapar til ein populasjon.

$$\begin{tikzcd}[column sep=huge] \text{Populasjon} \arrow[r, bend left, "\text{sannsyn}"] & \text{Stokastisk utval} \arrow[l, bend left, "\text{statistisk inferens}"] \end{tikzcd}$$

Statstisk inferens III

Obervator

Ein **observator** (eng: statistic) er ein funksjon av stokastiske variablar i eit tilfeldig utval.

Utvalsfordeling

Fordelinga til ein observator kallast utvalsfordelinga (eng: sampling distribution).

Lineærkombinasjon av normalfordelte stokastiske variablar

Teorem

La $X_1, \ldots, X_n$ vere normalfordelte stokastiske variablar med forventning $E(X_i) = \mu_i$ og varians $Var(X_i) = \sigma_i^2$. La

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i,$$

der a_i er konstantar. Då har Y forventning

$$\mu_Y = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n = \sum_{i=1}^{n} a_i \mu_i$$

og varians

$$\sigma_Y^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Sentralgrenseteoremet I

Teorem

La X_i vere uavhengig identisk fordelte (u.i.f.) stokastiske variablar med $E(X_i) = \mu_X$ og $Var(X_i) = \sigma_X^2 < \infty$ for $i = 1, \ldots, n$. La $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Då vil fordelinga til

$$Z = \frac{\bar{X}_n - E(\bar{X}_n)}{SD(\bar{X}_n)} = \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} = \sqrt{n}\left(\frac{\bar{X}_n - \mu_X}{\sigma_X}\right)$$

gå mot ei standard normalfordeling når $n \to \infty$.

Eksempel sentralgrenseteoremet

1. Anta $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ er eit tilfeldig utval frå ei kjend fordeling $g(x)$
2. Rekn ut $\bar{x} = \sum_{i=1}^{n} x_i$
3. Gjenta steg 1 og 2 $k(= 5,000,000)$ gonger. Sjå på utvalsfordelinga til $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$

Sentralgrenseteoremet for normalfordelinga ($n(x; 0, 1)$)

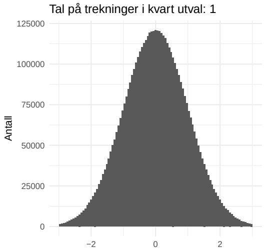

Figure: $n = 1$

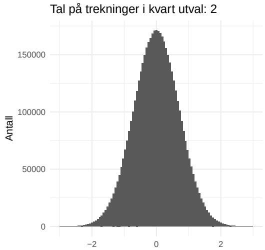

Figure: $n = 2$

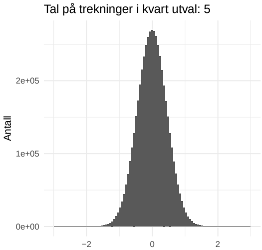

Figure: $n = 5$

Sentralgrenseteoremet for uniformfordelinga ($f(x; 0, 1)$)

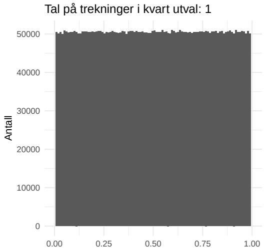

Figure: $n = 1$

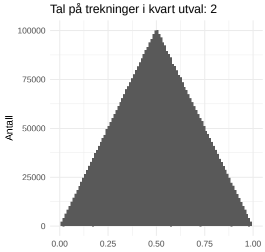

Figure: $n = 2$

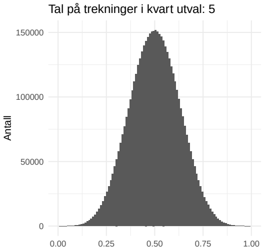

Figure: $n = 5$

Sentralgrenseteoremet for eksponentialfordelinga ($f(x; 1)$)

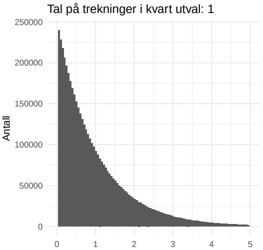

Figure: $n = 1$

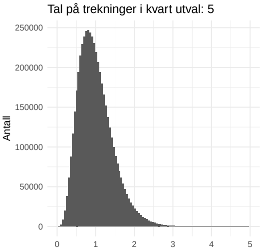

Figure: $n = 5$

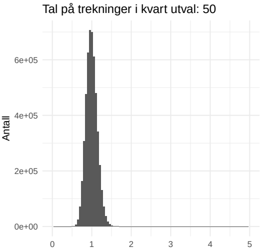

Figure: $n = 50$

Torsdag

- Estimatorar
- Kahoot