Normalfordeling som tilnærming til binomisk fordeling

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Sentralgrenseteoremet og binomisk fordeling

Teorem (Sentralgranseteoremet)

La $X_1, X_2, \ldots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $Var[X_i] = \sigma^2$. La

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

der $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ er gjennomsnittet av X_i-ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \to \infty$.

- Praktisk konsekvens av teoremet: Når n er stor (nok) har vi

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx N(0,1) \quad \Leftrightarrow \quad \sum_{i=1}^{n} X_i \approx N(n\mu, n\sigma^2)$$

- La $Y \sim \text{binomisk}(n, p)$
  - skriver $Y = \sum_{i=1}^{n} X_i$, der $X_i = \begin{cases} 1 & \text{hvis suksess i } i\text{-te forsøk,} \\ 0 & \text{ellers.} \end{cases}$
  - $\mu = E[X_i] = 0 \cdot (1-p) + 1 \cdot p = p$
  - $\sigma^2 = \text{Var}[X_i] = E[X_i^2] - E[X_i]^2 = \ldots = p(1-p)$
  - hvis n er stor (nok) gir sentralgrenseteoremet at

$$Y = \sum_{i=1}^{n} X_i \approx N(np, np(1-p))$$

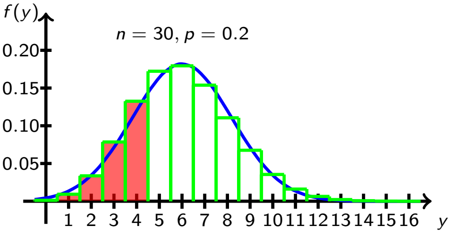
  - tommelfingerregel: god approksimasjon hvis $np \geq 5$ og $n(1-p) \geq 5$

Kontinuitetskorreksjon

Normalfordeling som tilnærming til binomisk:

$$Y \sim \text{binomisk}(n, p)$$
$$Y \approx N(np, np(1-p))$$

- ⋆ La $Y \sim \text{binomisk}(n = 30, p = 0.2)$
- ⋆ Tilnærming: $Y \approx N(6, 4.8)$
- ⋆ La $U \sim N(6, 4.8)$ og $Z \sim N(0, 1)$



- ⋆ Ønsker nå $P(Y \leq 4)$

$$\begin{aligned} P(Y \leq 4) &\approx P(U \leq 4 + 0.5) \\ &= P\left(\frac{U-6}{\sqrt{4.8}} \leq \frac{4.5-6}{\sqrt{4.8}}\right) \\ &= P(Z \leq -0.68) \\ &= \Phi(-0.68) \\ &= 0.2483 \end{aligned}$$

- ⋆ Vi har altså at

$$P(Y \leq 4) \approx 0.2483$$

Kontinuitetskorreksjon

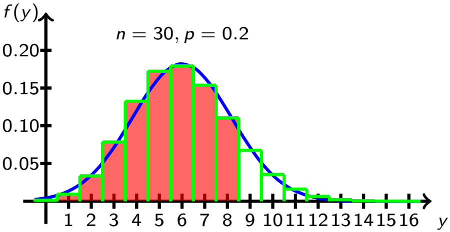
Normalfordeling som tilnærming til binomisk:

$Y \sim \text{binomisk}(n, p)$

$Y \approx N(np, np(1-p))$

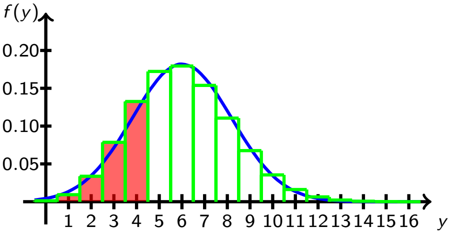
- ⋆ La $Y \sim \text{binomisk}(n = 30, p = 0.2)$
- ⋆ Tilnærming: $Y \approx N(6, 4.8)$
- ⋆ La $U \sim N(6, 4.8)$ og $Z \sim N(0, 1)$
- ⋆ Ønsker nå $P(Y < 9)$

$P(Y < 9) \approx P(U \leq 9 - 0.5)$

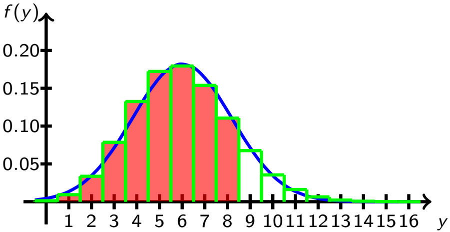


Oppsummering

- ⋆ Har sett hvordan sentralgrenseteoremet gir at
  - binomisk fordeling kan tilnærmes med normalfordeling
- ⋆ Tommelfingerregel: God approksimasjon hvis $np \geq 5$ og $n(1-p) \geq 5$
- ⋆ Bedre approksimasjon ved å benytte kontinuitetskorreksjon for å regne ut sannsynligheter



$P(Y \leq 4)$



$P(Y < 9)$