Konfidensintervall og stokastisk simulering

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

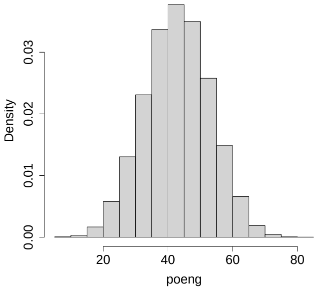
Sannsynligheten for bonus i tvungen yatzy

- ⋆ Simulerer yatzy $n = 10\,000$ ganger
  - – registrerer hver gang y_i, summen av poeng på enere, toere, ... og seksere
  - – registrerer hver gang

$$z_i = \mathbb{I}(y_i \geq 42) = \begin{cases} 1 & \text{hvis } y_i \geq 42, \\ 0 & \text{ellers} \end{cases}$$

- ⋆ Sannsynlighetshistogram over $y_i, i = 1, 2 \ldots, 10\,000$

- ⋆ Interessert i $p = P(Y_i \geq 42)$
  - – (forventningsrett) estimator for p:
    $\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(Y_i \geq 42)$
  - – estimat for p:
    $\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(y_i \geq 42) = 0.5724$



- ⋆ Konfidensintervall for p:
  - – bruker $X = \sum_{i=1}^{n} Z_i \sim \text{binomisk}(n, p)$
  - – lager konfidensintervall med utgangspunkt i

$$\frac{X - np}{\sqrt{n\widehat{p}(1-\widehat{p})}} \approx N(0,1)$$

  - – 95%-konfidensintervall for p blir

$$\left[\frac{x}{n} - z_{0.025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \frac{x}{n} + z_{0.025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right] = [0.5627, 0.5821]$$

  - – øker antall simuleringer, $n = 1\ 000\ 000$, $x = 575\ 858$: $[0.5749, 0.5768]$

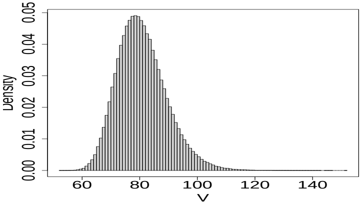
Approksimere en fordeling ved stokastisk simulering

- Anta: Vi er interessert i

$$V = \frac{S}{T},$$

hvor $S \sim N(\mu_S, \sigma_S^2)$ og $T \sim N(\mu_T, \sigma_T^2)$ er uavhengige stokastiske variabler

- Trenger tallverdier for μ_S, σ_S^2, μ_T og σ_T^2
- La $\mu_S = 40$, $\sigma_S^2 = 1^2$, $\mu_T = 0.5$, $\sigma_T^2 = 0.05^2$
- Sannsynlighetshistogram for V
- Interessert i $\mu = E[V]$
  - estimator og estimat for μ



$$\widehat{\mu} = \overline{V} = \frac{1}{n}\sum_{i=1}^{n} V_i, \quad \widehat{\mu} = \overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i = 80.83$$

- Konfidensintervall for μ:
  - bruker at $\overline{V} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad \frac{\overline{V} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx N(0,1)$
  - variansen σ^2 er ukjent, erstatter σ^2 med estimator S^2: når n er stor kan man vise at

$$\frac{\overline{V} - \mu}{\sqrt{\frac{S^2}{n}}} \approx N(0,1)$$

  - 95%-konfidensintervall for μ blir (med $n = 1\ 000\ 000$)

$$\left[\overline{v} - z_{0.025}\sqrt{\frac{s^2}{n}}, \overline{v} + z_{0.025}\sqrt{\frac{s^2}{n}}\right] = [80.813, 80.847]$$

Oppsummering

- Konfidensintervall kvantifiserer usikkerhet i forbindelse med stokastisk simulering
- Har sett på to eksempler
  - p i binomisk fordeling
  - forventningsverdi μ i en fordeling vi kan simulere fra
- For å øke nøyaktighet med en faktor 10 må vi øke antall simuleringer med en faktor 100
  - lengden av konfidensintervallet går som $1/\sqrt{n}$