Sentralgrenseteoremet

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Gjennomsnitt av uavhengige normalfordelte variabler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim N(\mu, \sigma^2)$
- Hva kan vi da si om gjennomsnittet?

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- $\overline{X}$ er en lineær funksjon av $X_1, X_2, \ldots, X_n$, så $\overline{X}$ er normalfordelt

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\cdot n\mu = \mu$$

$$\mathrm{Var}[\overline{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \left(\frac{1}{n}\right)^2 \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

- Har dermed

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

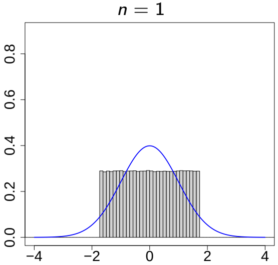
- Standardiserer

$$Z = \frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \sqrt{n}\cdot\frac{\overline{X}-\mu}{\sigma} \sim N(0,1)$$

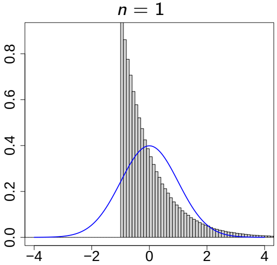
- Hva endres hvis X_i har en annen fordeling enn normalfordeling?
  - kan vise at i grensen når $n \to \infty$ er Z fremdeles standard normalfordelt

Eksempler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0,1]$
- Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering

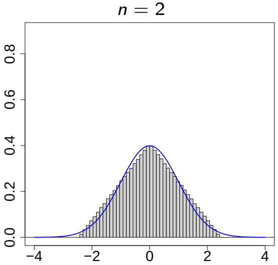


- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering

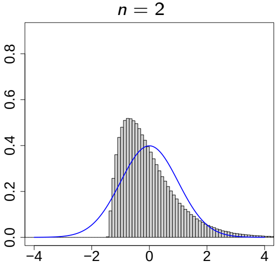


Eksempler

- ⋆ Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0, 1]$
- ⋆ Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- ⋆ La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- ⋆ Approksimerer fordelingen til Z ved stokastisk simulering

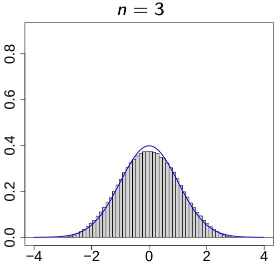


- ⋆ Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- ⋆ Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- ⋆ La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- ⋆ Approksimerer fordelingen til Z ved stokastisk simulering

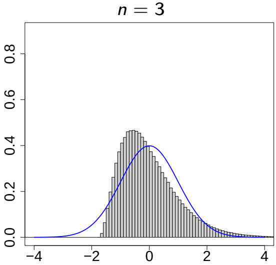


Eksempler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0,1]$
- Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



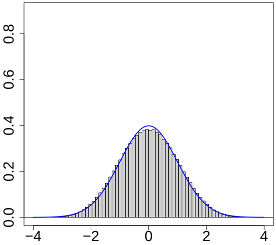
Eksempler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0, 1]$
- Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- La

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

- Approksimerer fordelingen til Z ved stokastisk simulering

$n = 4$

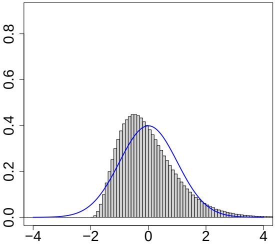


- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- La

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

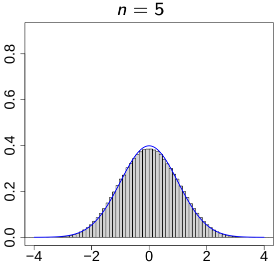
- Approksimerer fordelingen til Z ved stokastisk simulering

$n = 4$

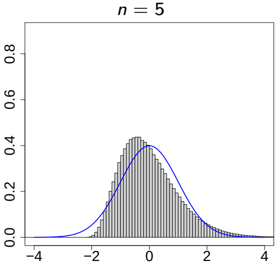


Eksempler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0, 1]$
- Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



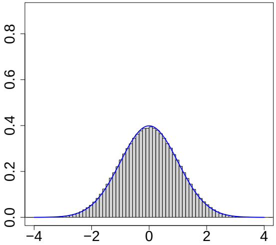
- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



Eksempler

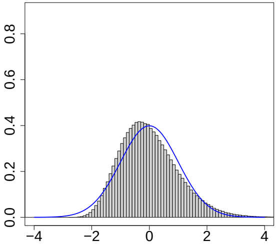
* Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0,1]$
* Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
* La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
* Approksimerer fordelingen til Z ved stokastisk simulering

$n = 10$



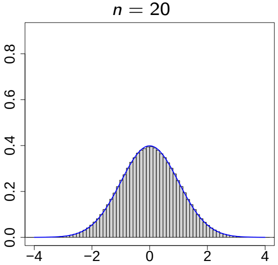
* Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
* Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
* La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
* Approksimerer fordelingen til Z ved stokastisk simulering

$n = 10$

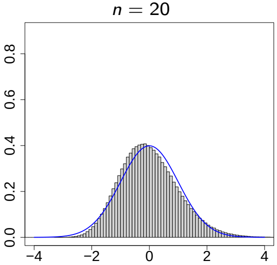


Eksempler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0, 1]$
- Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



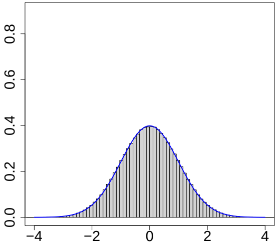
- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



Eksempler

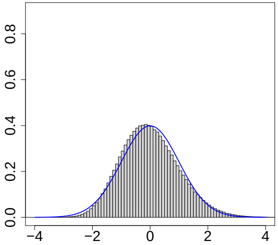
- ⋆ Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0,1]$
- ⋆ Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- ⋆ La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- ⋆ Approksimerer fordelingen til Z ved stokastisk simulering

$n = 30$



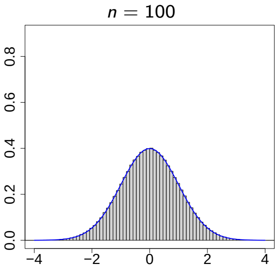
- ⋆ Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- ⋆ Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- ⋆ La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- ⋆ Approksimerer fordelingen til Z ved stokastisk simulering

$n = 30$

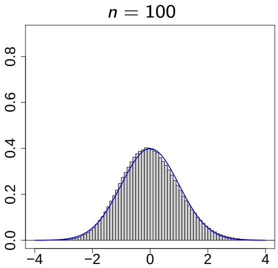


Eksempler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0, 1]$
- Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



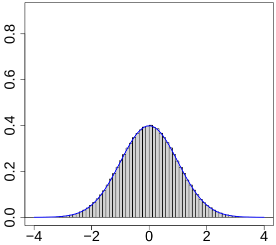
- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering



Eksempler

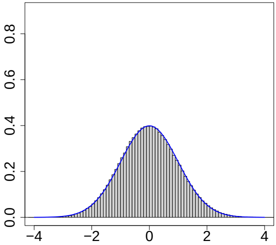
- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Unif}[0,1]$
- Da er $\mu = E[X_i] = \frac{1}{2}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{12}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering

$n = 1000$



- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt, $X_i \sim \text{Exponential}(3)$
- Da er $\mu = E[X_i] = \frac{1}{3}$ og $\sigma^2 = \text{Var}[X_i] = \frac{1}{9}$
- La
$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$
- Approksimerer fordelingen til Z ved stokastisk simulering

$n = 1000$



Sentralgrenseteoremet

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 $n \to \infty$.

- ⋆ Hva er den praktiske konsekvensen av teoremet?
- ⋆ Når n er stor (nok) kan vi med god approksimasjon regne som om

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$
$$\Updownarrow$$
$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$
$$\Updownarrow$$
$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

- ⋆ Mye brukt tommelfingerregel: God approksimasjon hvis $n \geq 30$
- ⋆ Bevis: Bevises ved å finne $M_Z(t)$ og så la $n \to \infty$.

Oppsummering

- ⋆ Har formulert sentralgrenseteoremet
  - har sett på eksempler relatert til teoremet
  - har diskutert praktiske konsekvenser av teoremet