

Stokastisk simulering og pseudo-tilfeldige tall

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

Tvungen yatzy

- ★ Hva er sannsynligheten for å få bonus i tvungen yatzy?
- ★ Spilleregler:
 - spiller med 5 terninger
 - kaster først for å få enere, så toere og så videre
 - for hver verdi, kan kaste hver terning inntil tre ganger
 - får poeng lik antall øyne på terningene som viser riktig antall øyne
 - hvis man får yatzy kan man ta dette utenfor tur
 - får bonus dersom man tilsammen får minst 42 poeng
- ★ Antar strategi:
 - når verdi v , tar vare på terninger hvor man får v øyne, kaster resten på nytt
 - hvis man får yatzy tar man dette utenfor tur, men kaster ikke for å få yatzy
- ★ Hvordan kan man regne ut sannsynligheten?
 - identifisere alle mulige kast-serier (veldig mange!!)
 - finne sannsynligheten for hver kast-serie
 - bestemme hvilke kast-serier som gir bonus
 - summere sannsynlighetene for kast-serier som gir bonus
- ★ Kan vi enklere finne en tilnærming til sannsynligheten?
 - husk tolkning av sannsynlighet: gjenta forsøket mange ganger

$$\frac{s(n)}{n} \rightarrow P(A)$$

- vi kan spille n ganger og regne ut

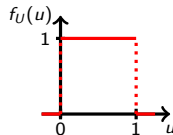
$$P(A) \approx \frac{s(n)}{n}$$

- når n er stor er det en god approksimasjon
- kan vi få en datamaskin til å spille spillene for oss?

Pseudo-tilfeldige tall

- ★ Kan vi generere tilfeldige tall på en datamaskin?
 - datamaskiner er deterministiske — kan ikke generere tilfeldige tall
 - er utgitt bøker med tilfeldige tall, f.eks. «A million random digits with 100,000 normal deviates»
- ★ I dag: Genererer på datamaskin en deterministisk følge av tall som «ser tilfeldige ut»
 - viktig at vi har kontroll på hvilken fordeling tallene kommer fra
 - viktig at tallene er «uavhengige»
 - slike «random-generatorer» må initialiseres med et eller flere frø (seed)
- ★ Det finnes mange programmer/programmeringsspråk med funksjoner for å generere pseudo-tilfeldige tall fra flere ulike sannsynlighetsfordelinger
 - den viktigste fordelingen er uniformfordelingen på intervallet $[0, 1]$

$$f_U(u) = \begin{cases} 1 & \text{for } x \in [0, 1], \\ 0 & \text{ellers.} \end{cases}$$



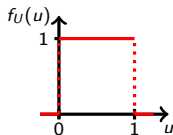
- kan bruke slike tall til å simulere (generere tilfeldig tall) fra andre fordelinger
- hvordan kan vi bruke $U \sim f_U(u)$ til å generere X med fordeling

$$f(x) = \frac{1}{6} \quad \text{for } x = 1, 2, 3, 4, 5, 6$$

Simulering av terningkast

- ★ Anta $U \sim f_U(u)$ der

$$f_U(u) = \begin{cases} 1 & \text{for } u \in [0, 1], \\ 0 & \text{ellers.} \end{cases}$$



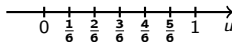
- ★ La

$$X = \lfloor U \cdot 6 \rfloor + 1$$

- ★ Påstand: Da er $f_X(x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$

- ★ Bevis: Merk at for $x = 1, 2, 3, 4, 5, 6$ har vi at

$$\frac{x-1}{6} \leq U < \frac{x}{6} \Leftrightarrow X = x$$



- dermed har vi at

$$\begin{aligned} f_X(x) &= P(X = x) = P\left(\frac{x-1}{6} \leq U < \frac{x}{6}\right) = \int_{\frac{x-1}{6}}^{\frac{x}{6}} f_U(u) du \\ &= \int_{\frac{x-1}{6}}^{\frac{x}{6}} 1 du = \left[u \right]_{\frac{x-1}{6}}^{\frac{x}{6}} = \frac{x}{6} - \frac{x-1}{6} = \frac{1}{6} \end{aligned}$$

- ★ MEN: hva skjer dersom $U = 1$?
- ★ I stedet for å la X være gitt som over kan vi la

$$X = \min(6, \lfloor U \cdot 6 \rfloor + 1)$$

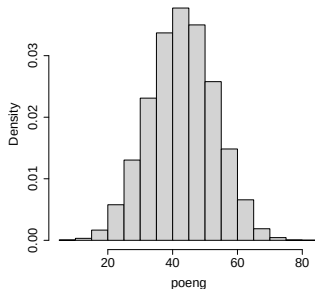
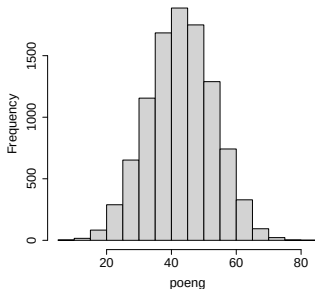
Sannsynligheten for bonus i tvungen yatzy

★ Simulerer yatzy $n = 10\,000$ ganger

- registrerer hver gang y_i , summen av poeng på enere, toere, ... og seksere
- registrerer hver gang

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

★ Histogram over $y_i, i = 1, 2, \dots, 10\,000$



★ Sannsynligheten for bonus: $P(\text{bonus}) \approx \frac{5724}{10\,000} = 0.5724$

★ Sannsynligheten for yatzy (i disse første seks rundene): $P(\text{yatzi}) \approx \frac{864}{10\,000} = 0.0864$

★ Hvor gode approximasjoner er dette? Bør vi ha $n > 10\,000$?

Oppsummering

- ★ Har diskutert hvordan pseudo-tilfeldige tall kan benyttes
 - til å approksimere sannsynligheter
 - til å approksimere fordelinger
 - merk: Slik bruk av pseudo-tilfeldige tall kalles stokastisk (Monte Carlo) simulering
- ★ Vi har sett hvordan en $U \sim \text{Unif}[0, 1]$ kan benyttes
 - til å trekke et tilfeldig tall blant 1, 2, 3, 4, 5, 6
- ★ Merk: Vi kan nå også simulere fra en urnemodell
- ★ Gjenstår å diskutere:
 - hvordan simulere fra andre diskrete fordelinger
 - hvordan simulere fra andre kontinuerlige fordelinger
 - hvordan sjekke om pseudo-tilfeldige tall «ser tilstrekkelig tilfeldige ut»
 - hvor stor bør n være for å få en god approksimasjon