Laboration 3

Sannolikhetsteori I

Emil Erikson och Leo Levenius*

2024-09-27

Förberedelser

Före en laboration, gör följande:

1. **Läs igenom filen "Laborationskrav" på kurshemsidan.**
2. För instuktioner på hur man skapar korrekta tabeller och diagram samt generella tips, läs igenom följande filer på kurshemsidan:
    - "Skapa tabeller från vektorer + text och numrering av tabeller och diagram"
    - "Introduktion till Rstudio, R Markdown, och R"
    - "Exempel R och R Markodown"
    - "Matematiska symboler i R Markdown/LaTeX"
3. Ladda ned laborationsmallen.
4. Under "Tools/Global Options/Code/Saving/Default text encoding", tryck "Change" och välj "UTF-8".

*Tidigare versioner av Monir Bounadi, Benjamin Kjelsson, Maria Deijfen, Andreas Nordvall Lagerås, Tom Britton, Jens Malmros, och OE.

Inledning

Syftet med laborationen är att använda simuleringar för att illustrera två av de viktigaste gränsvärdessatserna i sannolikhetsteorin: Stora talens lag och Centrala gränsvärdessatsen.

Uppgift 1: Medelvärde av summor av stokastiska variabler

I denna uppgift ska vi generera slumptal och se hur deras medelvärde utvecklas då antalet observatiner ökar. Resultaten illustrerar en sats som kallas för *Stora talens lag*. Börja med att skapa ett dataset med $n = 300$ observationer från en exponentialfördelning med väntevärde $m = 1$:

```
# Först ger vi ett "frö" (seed) till slumptalsgeneratorn genom funktionen set.seed
# Detta gör att vi får exakt samma slumptal varje gång vi kör koden.
set.seed(19690420) # fyll i ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

m <- 1 # väntevärde
slumptal <- rexp(300, rate = 1/m) # Notera att argumentet rate = 1 / väntevärde.
# Hade vi velat generera slumptal från en exponentialfördelning med
# väntevärde 2 så hade vi alltså skrivit rate = 1 / 2 istället.
```

Vi tar en titt på slumptalen.

```
hist(slumptal,
  main = "Histogram för exponentialfördelade slumptal",
  ylab = "Antal",
  xlab = "Värde")
```

Följande kod ger en vektor, `medel`, med $n = 300$ element där element i anger stickprovsmedelvärdet av de i första slumptalen och producerar sedan en plot över denna vektor. Gå igenom koden och se till att ni förstår vad som händer (testa t.ex. med olika vektorer som `c(1, 1, 1)` eller kolla vad funktionerna gör genom att skriv t.ex. `?cumsum` i konsollen i `R`).

```
medel <- cumsum(slumptal) / 1:length(slumptal)

plot(medel,
  type = "l", # plotta en linje
  main = "Kumulativt löpande medelvärde",
  ylab = "Medelvärde",
  xlab = "Antal observationer")
abline(a = m, b = 0, col = "gray60") # Horisontell linje vid y = 1 = väntevärdet
```

Studera plotten. Vad händer med medelvärdet? Ser det ut att konvergera mot något? Testa att simulera nya slumptal och göra en ny plot på följande vis.

```
set.seed(19690420) # För att få nya slumptal måste du byta seed från förra simuleringen.

slumptal2 <- rexp(300, rate = 1/m)
medel2 <- cumsum(slumptal2) / 1:length(slumptal2)

plot(slumptal2)
# Plotta nu själv vektorn medel2
```

Ser det likadant ut nu? Vad händer när ni ökar antalet slumpade tal? Nedan visas hur 1000 slumptal genereras.

```
set.seed(19690420) # fyll i ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

slumptal3 <- rexp(1000, rate = 1/m)
medel3 <- cumsum(slumptal3) / 1:length(slumptal3)

# Plotta nu själv vektorn medel3
```

Generera nu $n = 1000$ slumptal från en exponentialfördelning med ett annat väntevärde m (välj själv ett annat värde på m) och kopiera samt ändra i koden ovan för att se hur medelvärdet av talen utvecklar sig. Pröva för några olika värden på m. Kom ihåg att till funktionen `rexp` i `R` så ger ni inte väntevärdet m, utan $1/m$, till argumentet `rate`.

Se också vad som händer om vi låter slumptalen komma ifrån en $\mathcal{N}(0,1)$-fördelning eller en $\text{Re}(0,1)$-fördelning (`slumptal <- rnorm(n)` respektive `slumptal <- runif(n)`). Använd samma n som ovan. Se hur medelvärdet utvecklas och jämför med fördelningens väntevärde!

Viktigt

- Kom ihåg att använda funktionen `set.seed` med födelsedatum innan ni använder slumptalsgeneratorn (någon av funktionerna `runif`, `rexp` eller `rnorm`). Detta gäller alla uppgifter ovan.

Redovisning av uppgift 1

- Skriv en inledande beskrivning av vad Stora talens lag säger, med egna ord.
- Svara på följande frågor:
  1. Hur utvecklas medelvärdet för de exponentialfördelade slumptalen?
  2. Vad tycks det konvergera mot?
  3. Gäller samma sak för normal- och likformigt fördelade slumptal?
- Skapa medelvärdesplottar för $\text{Exp}(1/m)$-fördelade slumptal för två olika värden på m (välj själv), samt för *antingen* $\mathcal{N}(0,1)$-fördelning eller $\text{Re}(0,1)$-fördelning.
  - Diagrammen **måste** ha numrering samt en beskrivande text under diagrammet. *Tips*: Använd `fig.cap`.
- Skriv en slutsats om hur diagrammen ni visat (simuleringarna ni har utfört) hör ihop med Stora talens lag.

Uppgift 2: Fördelning för summa av stokastiska variabler

I den här uppgiften ska vi undersöka vad summan av ett antal stokastiska variabler har för fördelning. Resultatet kommer att illustrera en sats som kallas för *Centrala gränsvärdessatsen*.

Skapa först ett dataset med $n = 300$ observationer från en likformig fördelning:

```
set.seed(19690420) # fyll i ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

likf <- runif(300)
```

Ta en titt på slumptalen:

```
hist(likf,
  main = "Likformigt fördelade slumptal",
  ylab = "Antal",
  xlab = "Värde")
```

Gör inte detta i er rapport, men gör det på er egen dator för att få en uppfattning av hur en sådan mängd likformiga tal är fördelade.

Fördelningen har väntevärde $\mu = 1/2$ och varians $\sigma^2 = 1/12$. Skapa nu en matris med 30 rader där varje rad innehåller 1000 observationer från en likformig fördelning ($\mathrm{Re}(0, 1)$).

```
set.seed(19690420) # fyll i ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

n <- 30 # antalet rader
M <- 1000 # antalet kolumner

likf_matris <- matrix(runif(n * M), nrow = n, ncol = M)
```

Ni kan göra ett histogram av observationerna på exempelvis rad 3 genom att skriva

```
hist(likf_matris[3, ],
  main = "Likformigt fördelade slumptal",
  ylab = "Antal",
  xlab = "Värde")
```

Gör inte detta i er rapport, men gör det på er egen dator för att övertyga dig om innehållet i t.ex. rad 3 av matrisen.

Genom att summera elementen i matrisen `likf_matris` kolumnvis genom funktionen `colSums` kan vi skaen vektor `summa` med 1000 element.

```
summa <- colSums(likf_matris)
# Notera att det finns en liknande funktion för att summera över rader: rowSums
# Ännu fler liknande funktioner som fungerar på matriser och data.frames
# kan ni se om ni skriver ?colSums i Console.
```

Varje element i `summa` är alltså en summa av $n = 30$ stycken likformigt fördelade slumptal. Gör ett histogram för `summa`. Vad ser det ut att vara för fördelning?

Prova också att standardisera observationerna genom att dra bort (det teoretiska) väntevärdet och dividera med (den teoretiska) standardavvikelsen:

```
mu <- 1/2 # Väntevärdet
sigma2 <- 1/12 # Variansen
stand <- (summa - n * mu) / sqrt(n * sigma2) # standardisera observationerna
hist(stand,
  main = "Standardiserade observationer",
  ylab = "Antal",
  xlab = "Värde")
```

Gör nu om samma sak som ovan, fast med $Exp(1)$-fördelade slumptal istället! Glöm inte att ändra väntevärdet μ och variansen σ^2 i standardiseringen ovan till väntevärdet och variansen i exponentialfördelningen.

Glöm heller inte att använda funktionen `set.seed` med födelsedatum innan ni använder funktionen `rexp`. Hur ser histogrammet för `summa` ut nu? Prova att summera fler eller färre slumptal än 30, t.ex. $n = 3$ och $n = 300$. Behåll samma antal kolonner $M = 1000$ i de nya matriserna ni skapar. Vilken fördelning liknar det?

Redovisning av uppgift 2

- Skriv en inledande beskrivning av vad Centrala gränsvärdessatsen säger, med egna ord.
- Svara på följande frågor:
  1. Vilken fördelning verkar de standardiserade summorna av likformiga slumptal få?
  2. Gäller samma sak för exponentialfördelade slumptal?
  3. Vad händer när ni ändrar antalet slumptal n som summeras?
- Skapa **två** histogram för de standardiserade summorna av likformiga slumptal, med olika antal slumptal n i summorna. Ange hur många slumptal summorna består av (t.ex. 3, 30 eller 300).
- Skapa **två** histogram för de standardiserade summorna av exponentiella slumptal, med olika antal slumptal n i summorna. Ange hur många slumptal summorna består av (t.ex. 3, 30 eller 300).
  - Alla diagram **måste** ha numrering samt en beskrivande text under diagrammet. *Tips*: Använd `fig.cap`.
- Skriv en slutsats om hur diagrammen ni visat (simuleringarna ni har utfört) hör ihop med Centrala gränsvärdessatsen.