Laboration 3

Sannolikhetsteori I

*Monir Bounadi**

2018-12-26

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. Resultaten illustrerar en sats som kallas för Stora talens lag. Börja med att skapa ett dataset med 300 observationer från en exponentialfördelning med väntevärde 1:

```
# Först ger vi ett "frö" (seed) till slumptalsgeneratorn genom funktionen set.seed
# Detta gör att jag kan få exakt samma slumptal, och således resultat, som dig
# när jag kör din kod
set.seed(19880210) # fyll i ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

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

Ta en titt på slumptalen:

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

Följande kod ger en vektor medel med 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 (skriv t.ex. `?cumsum` i Console i RStudio, och testa denna funktion på exempelvis vektorn `c(1, 1, 1)`).

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

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

Studera plotten. Vad händer med medelvärdet? Ser det ut att konvergera mot något? Gör nya slumptal och en ny plot med ett annat frövärde, på följande vis:

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

```r
set.seed(19880210) # fyll i ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

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

plot(medel2,
  type = "l", # plotta en linje, inte ringar (default)
  main = "Kumulativt löpande medelvärde",
  xlab = "Antal observationer",
  ylab = "Medelvärde")
abline(a = 1, b = 0, col = "gray60") # Horisontell linje vid y = 1
```

Ser det likadant ut nu? Ni kan också generera flera slumptal—skriv `rexp(n, rate = 1)` där ni väljer n större än 300—och se vad som händer. Nedan visas hur 1000 slumptal genereras

```r
set.seed(19880210) # byt ut 19880210 mot ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

slumptal3 <- rexp(1000, rate = 1)
medel3 <- cumsum(slumptal3) / 1:length(slumptal3)
# Plotta sedan vektorn medel3
```

Generera nu slumptal från en exponentialfördelning med ett annat väntevärde m (välj ett värde på m) ochkopiera och ändra i koden ovan för att se hur medelvärdet av talen utvecklar sig. Dvs, generera $n = 1000$ slumptal. Jämför med väntevärdet m i den fördelning talen kommer ifrån. Pröva med några olika värden på m. Kom ihåg att till funktionen rexp i R så ger du 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 väntevärdet i fördelningen!

Viktigt

- Kom ihåg att använda funktionen set.seed med ditt födelsedatum innan du använder slumptalsgeneratorn (någon av funktionerna `runif`, `rexp`, `rnorm`). Detta gäller alla uppgifter ovan.
- Kom ihåg att när du ska skapa en plot med kod lik den ovan, så måste du där det står `abline(a = 1, b = 0, col = "gray60")` byta ut `a = 1` mot `a = m`, där `m` är väntevärdet för den fördelning du genererar slumptal för.

Redovisning av uppgift 1

Redovisningen skall inkludera följande:

- En inledande beskrivning av vad Stora talens lag säger, i egna ord.
- Svar på följande frågor:
  - Hur utvecklas medelvärdet för de exponentialfördelade slumptalen?
  - Vad tycks det konvergera mot?
  - Gäller samma sak för normal- och likformigt fördelade slumptal?
- 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.
- En slutsats om hur diagrammen du visat (simuleringarna du 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 heter Centrala gränsvärdessatsen.

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

```
set.seed(19880210) # 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 din rapport, men gör det på din egen dator för att få en uppfattning av hur en sådan mängd likformiga tal är fördelade.

Fördelningen borde ha 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 $\text{Re}(0, 1)$:

```
set.seed(19880210) # 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)
```

Du 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 din rapport, men gör det på din egen dator för att övertyga dig om innehållet i t.ex. rad 3 av matrisen.

Skapa en vektor summa med 1000 summor genom att summera elementen i `likf_matris` kolumnvis

```
summa <- colSums(likf_matris)
# alternativt sätt att göra detsamma: summa <- apply(likf_matris, 2, sum)
# Skriv ?apply i Console för att läsa mer om funktionen apply

# 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 du se om du 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)
```

```
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 ditt födelsedatum innan du 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 du skapar. Vilken fördelning liknar det?

Redovisning av uppgift 2

Redovisningen skall inkludera följande:

- En inledande beskrivning av vad Centrala gränsvärdessatsen säger, i ord.
- Svar på följande frågor: – Vilken fördelning verkar de standardiserade summorna av likformiga slumptal få? – Gäller samma sak för exponentialfördelade slumptal? – Vad händer när ni ändrar antalet slumptal n som summeras?
- Endast två histogram för de standardiserade summorna av likformiga slumptal, med olika antal slumptal (n) i summorna.
- Endast två histogram för de standardiserade summorna för Exp(1)-fördelade slumptal, med olika antal slumptal i summorna. Ange hur många slumptal summorna består av (t.ex. 3, 30 eller 300).
- En slutsats om hur diagrammen du visat (simuleringarna du har utfört) hör ihop med Centrala gränsvärdessatsen.

Tips

I uppgift 3 ovan kan du skapa data för de standardiserade summorna, och rita upp histogrammet, i ett och samma stycke kod enligt:

```
set.seed(19880210) # fyll i ditt egna födelsedatum. Om ni jobbar i par, välj den enas.

n <- 30 # antalet rader = antalet tal i summan
M <- 1000 # antalet kolonner

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

mu <- 1/2 # Väntevärdet
sigma2 <- 1/12 # Variansen
stand <- (summa - n * mu) / sqrt(n * sigma2)

hist(stand,
  main = "Standardiserade observationer",
  ylab = "Antal",
  xlab = "Värde")
```

Då är det lätt att byta ut värdet på n (men låt $M = 1000$, alltid!) samt att ändra fördelning (byt ut `runif` mot `rexp`, samt ange rätt parameter till den funktionen och ändra `mu` och `sigma2` mot rätt väntevärde och varians).