Laboration 2: Absorberande Markovkedjor

Benjamin Kjellson (tidigare version skriven av Anders Björkström)

2016-05-26

Viktigt: Innan ni läser vidare

Gör först följande:

1. I den övre menyn i RStudio, tryck på `Tools --> Global Options...`. I rutan som dyker upp, där det står "Default text encoding", tryck "Change" och välj "UTF-8".
2. Gå in på kurshemsidan och ladda ner mallen för denna laboration (döp den till något vettigt när ni sparar den på er dator). Öppna den i RStudio. Skriv er rapport i denna fil.

För dig som använder en av skolans datorer

Om du har sparat din `.Rmd`-fil t.ex. i mappen `Documents` eller mappar i den, och du får ett konstigt felmeddelande när du trycker på `Knit HTML`, så finns det ett par sätt att lösa detta problem. På kurshemsidan, under samma dag som denna labb, så finns det en länk VIKTIGT! Du som använder skolans laptop. Klicka på den för att läsa hur du löser problemet.

Krav för laboration 2

Tänk på följande **krav** på er rapport:

- Rapporten måste kunna läsas av någon som inte har läst labbinstruktionerna. Så ni måste skriva vad det är ni ska göra innan ni gör det, och berätta vad syftet är.
- Rapporten måste vara skriven i **R Markdown** eller knitr.
- All kod som används måste synas i labbrapporten, men ska inte beskrivas i detalj i rapporten.
- Alla **tabeller och diagram** måste förses med **numrering och beskrivande text**, och refereras till i rapportens vanliga text på rätt sätt. Diagram måste ha lämpliga rubriker på axlar och tabeller lämpliga rubriker på kolumner.

Introduktion

I förra laborationen undersökte vi vad som händer när $n \to \infty$ i en reguljär Markovkedja. Även när Markovkedjor inte är reguljära så är det oftast långtidsuppträdandet som är det intressanta, från tillämpad synpunkt sett. Betrakta till exempel en person som kommer till ett kasino med ett visst startbelopp, X_0 kronor, på fickan. Låt X_n beteckna hens kapital efter n spelomgångar. Talföljden $X_1, X_2, X_3 \dots$ rör sig slumpartat uppåt och nedåt, men om spelaren är tillräckligt envis kan det bara sluta på endera av två sätt: Antingen blir hen pank eller så blir kasinot ruinerat. Det är intressant att räkna ut sannolikheterna för de båda möjligheterna, för både spelaren och kasinot.

Kim är stamkund på det populära kasinot Rodmans Roulette. Hen har en krona med sig när hen kommer, och satsar en krona i varje spelomgång. Vid vinst så ökar Kims kapital med en krona, och vid förlust så minskar det med en krona. Chansen att vinna är 50% (kasinot drivs utan vinstsyfte). Kim går hem till sin residens antingen när hen är pank eller när hen är ägare till 5 kronor. Vi ska genom praktiska experiment undersöka sannolikheten att hen lyckas vinna fem kronor.

Uppgift 1

(a)

Följande funktion tar en vinstsannolikhet `p` och ett kapital `kapital` som input, och ger som output det nya kapitalet efter en spelomgång, enligt definitionen som gavs in introduktionen ovan.

```
en_spelomgang <- function(p, kapital)
  if (runif(1) < p) {    # om vi vunnit
    return(kapital + 1)
  } else {               # annars har vi förlorat
    return(kapital - 1)
  }
}
```

Kommandot `runif(1)` genererar en slumpvariabel U från en likformig fördelning på intervallet $[0, 1]$. Som du bör känna till vid det här laget så gäller det att $\mathbb{P}(U < p) = p$ för $p \in [0, 1]$.

Din uppgift är nu att skriva en funktion `kim_spelar`, som simulerar Kims spelande enligt definitionen som gavs i introduktionen. Funktionen ska ta som input en vinstsannolikhet `p` och ett startkapital `kapital`. Som output ska funktionen returnera en 1:a (dvs `return(1)`) om Kims kapital når 5 kronor (Kim vinner), och en 0:a (dvs `return(0)`) om Kims kapital når 0 kronor (Kim förlorar). Funktionen `kim_spelar` **ska använda sig av** funktionen `en_spelomgang` ovan. Här är ett skelett:

```
kim_spelar <- function(p, kapital) {
  # fyll i innehållet
}
```

Tips:

- Kanske du vill definiera en tillfällig variabel som håller koll på det nuvarande kapitalet, och uppdatera denna variabel efter varje spelomgång.
- Kommandot `repeat { ... }` fungerar som en oändlig loop, men kan brytas genom att använda `break` eller `return(...)` någonstans innanför `{` och `}`. Här ersätts `...` med kod.

(b)

Skriv en funktion `sim_kim`, som gör `n` simuleringar av Kims spelande och räknar hur många gånger hen går med vinst. Funktionen ska använda sig av `kim_spelar` ovan, och ska som argument ta en vinstsannolikhet `p`, ett startkapital `kapital`, och antalet simuleringar `n` (default = 1000). Output ska vara antalet simuleringar som slutade med vinst. Här är ett skelett:

```
sim_kim <- function(p, kapital, n = 1000) {
  # fyll i innehållet
}
```

Gör nu tusen simuleringar med hjälp av funktionen du just skrev. Hur många gånger lyckades Kim nå sitt mål, respektive blev pank? Antag att Kim startar med 1 krona i kapital vid varje simulering, och att vinstsannolikheten vid varje spelomgång är 0.5. Använd funktionen `set.seed` med ditt 6-siffriga personnummer som argument innan du anropar `sim_kim`.

(c)

Om Kim har två kronor att starta med så är det rimligt att tro att chansen att lyckas nå fram till fem kronor är bättre. Undersök genom simulering hur det går för en spelare som har en, två, tre eller fyra kronor i startkapital. Gör tusen experiment för vart och ett av dessa tre fall också, och anteckna hur ofta spelet slutar lyckligt. Redovisa:

- Text som beskriver vad du gör.
- Koden du skrivit.
- I en tabell: Hur många av de tusen simulerade spelen som slutade lyckligt för Kim, när startkapitalet var 1, 2, 3 och 4 kronor.
- Vilka slutsatser du drar.

Anmärkning:

- Poängen med att skriva funktionerna ovan är att uppgift (c), samt senare uppgifter i denna labbrapport, nu blir väldigt enkela att lösa. I tidigare versioner av denna labb så har studenterna ofta löst uppgifterna genom att kopiera och ändra i koden för varje nytt startkapital och vinstsannolikhet. Detta leder till mer jobb, en grötigare rapport, och ökad risk för att göra fel. Lärdom: generalisera ditt problem, dela upp det i mindre delproblem, och skriv små funktioner för att lösa vart och ett av dessa delproblem. Använd dig sedan av de mindre funktionerna för att lösa det större problemet.

Uppgift 2

Ett annat sätt att undersöka Kims chans att lyckas är så här: Om Kims kapital efter n spelomgångar är X_n kronor (där X_0 betecknar startkapitalet) så kommer följden X_n att bilda en Markovkedja. Den kommer förr eller senare att hamna i något av de båda absorberande tillstånden 0 eller 5. Låt oss beteckna dessa båda händelser med $X_\infty = 0$ respektive $X_\infty = 5$. Sannolikheten $\mathbb{P}(X_\infty = 5)$ kan vi approximativt räkna ut genom att utnyttja en ekvation som säger att om $\mathbf{P}$ är övergångsmatrisen för en absorberande Markovkedja så gäller att

$$\mathbf{P}^n \to \begin{pmatrix} \mathbf{0} & \mathbf{SR} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

då $n \to \infty$. Här ska matrisen $\mathbf{P}$ vara skriven på "standardform", dvs de fyra transienta tillstånden (1–4) ska komma först, och de båda rekurrenta sist. Sätt tillstånd 0 sist.

(a)

Ett av elementen i matrisen **SR** beskriver sannolikheten att absorption sker i tillstånd 5 när spelaren startar med en krona. Vilket element är det?

(b)

Ta en bit papper och skriv ned övergångsmatrisen $\mathbf{P}$ för en generell vinstsannolikhet p (behöver ej redovisas). När du är säker på att du fått till rätt svar, skriv en funktion `kims_matris` i R, som tar som input en vinstsannolikhet `p` och ger som output övergångsmatrisen för denna vinstsannolikhet. Du får följande till hjälp—allt du behöver göra är att byta ut några av 0:orna mot rätt sannolikhet.

```
kims_matris <- function(p) {
  matrix(c( 0  , 0  , 0  , 0, 0, 0  ,
            0  , 0  , 0  , 0, 0, 0  ,
            0  , 0  , 0  , 0, 0, 0  ,
            0  , 0  , 0  , 0, 0, 0  ,
            0  , 0  , 0  , 0, 0, 0  ,
            0  , 0  , 0  , 0, 0, 0   ),
         nrow = 6, byrow = TRUE)
}
```

(c)

Skriv en funktion `hitta_SR` som tar som input en övergångsmatris `P` skriven på standardform och ger som output motsvarande matris **SR**. Matrisen **SR** kan räknas ut på teoretisk väg eller genom att avgöra för vilket n som $\mathbf{P}^n \approx \mathbf{P}^{n+1}$ med fyra decimalers noggrannhet

Om du väljer att skriva funktionen med hjälp av de teoretiska resultaten så måste du förklara (i ord och symboler) vad dessa säger först, och var du hittat dem. Om du väljer det andra sättet så kan du få använding av funktionerna `mpow` och `matrices_equal` från förra laborationen. Redovisa i detta fall för vilket n som $\mathbf{P}^n \approx \mathbf{P}^{n+1}$ i fallet då vinstsannolikheten är $p = 0.5$.

Hur många av de tusen spelen i uppgift 1 "borde" ha slutat lyckligt för Kim, när startkapitalet var 1, 2, 3 respektive 4 kronor? Stämmer talen i **SR** överens med de tal du kom fram till i uppgift 1 (c)?

Kommentar

Chansen att lyckas nå till fem kronor tycks öka med ökande startkapital enligt en ganska enkel formel. Genom att titta på resultaten hittills kan du förmodligen gissa hur formeln ser ut.

Uppgift 3

De flesta kasinon drivs inte ideellt, utan spelarna förlorar pengar i det långa loppet. Vi ska nu undersöka hur chansen att nå femkronorsnivån ändras när sannolikheten att förlora ett spel är större än chansen att vinna. Ett sådant spel kan beskrivas med en Markovkedja med en likadan övergångsmatris som i uppgift två, med skillnaden att p inte är 50%. Undersök chansen att nå från en krona till fem kronor då $p = 30\%, 40\%, 50\%, 60\%,$ och 70%. Vi bryr oss inte om att räkna på något annat startkapital än 1 krona. Använd dels simulering, som i uppgift 1, dels samma metod som i uppgift 2. Här får ni användning av de funktioner ni skrivit tidigare. Tänk på att det n som krävs för konvergens ändras då p ändras.

Redovisa sedan en tabell med tre kolumner och en rad för varje värde på p:

Vinstsannolikhet p	$\mathbb{P}(X_\infty = 5)$ enligt simulering	$\mathbb{P}(X_\infty = 5)$ avläst i **SR**
30%		
40%		

och så vidare (tabellen fortsätter nedåt till 70%). Här är lite kod ni kan använda (där det står `...` behöver ni fylla i något):

```
kims_df <- data.frame(Vinstsannolikhet = 3:7 / 10,
                      sim_P = ...
```

```r
                          SR_P = ...)
names(kims_df)[-1] <- c("$\\mathbb{P}(X_\\infty = 5)$ enligt simulering",
                          "$\\mathbb{P}(X_\\infty = 5)$ avläst i $\\mathbf{SR}$")
knitr::kable(kims_df, digits = 2, caption = "Tabell X: ...")
```

Uppgift 4

På Rodmans Roulette måste man inte satsa exakt en krona vid varje spel. Det är tillåtet att satsa ett valfritt belopp. Om man vinner får man tillbaka dubbla insatsen. Robin brukar besöka Rodmans Roulette ibland. Hen har alltid 1 krona i startkapital, men följer en djärvare strategi än Kim: inför varje spelomgång satsar hen alla pengar hen har. Dock är Robin, precis som Kim, nöjd om hen kan nå upp till 5 kronor. Om hen i ett visst ögonblick äger 3 eller 4 kronor satsar hen alltså bara 2 respektive 1 krona. Vi ska undersöka om Robin har bättre eller sämre chans att nå fram till femkronorsmålet än vad Kim har.

(a)

Robins kapital utvecklas med tiden på ett sätt som beskrivs av en Markovkedja. Beskriv kedjans övergångsmatris om vinstsannolikheten är p, genom att skriva en R-funktion `robins_matris` som ger denna övergångsmatris som output, givet input `p`.

(b)

Räkna ut matrisen **SR** (se uppgift 2c) för Robins övergångsmatris och avläs Robins vinstchans, då $p = 30\%, 40\%, 50\%, 60\%$, och 70%. Skriv *inte* ut dessa matriser; visa istället resultaten i en tabell tillsammans med motsvarande vinstchanser för Kim. Ha inte med för många decimaler i tabellen och kom ihåg att startkapitalet är 1 krona.

För vilka värden på p bör man spela djärvt respektive försiktigt?

Redovisning

Laborationen skall utföras i grupper om högst två personer. Redovisning sker genom skriftliga svar på uppgifterna 1–4, som ska vara inlämnade senast det datum som anges på schemat. De som har skrivit sitt namn på redogörelsen ska vara beredda på att besvara muntliga följdfrågor som examinatorn kan vilja ställa. Var noga med att följa kraven som ställdes i början av denna instruktion!

Efterarbete (behöver inte redovisas)

Det går att härleda en formel för sannolikheten att den "försiktige" Kim ska lyckas nå fram till fem kronor, när hens startkapital är i kronor och vinstsannolikheten är p. Härledningen ingår i kursen (fast vi kommer nog inte att hinna gå igenom den på föreläsningen). Läs avsnitt 4.5.1 i Ross (sid. 220–221 i upplaga 11, sid. 230–232 i upplaga 10, sid. 217–219 i upplaga 9). Där kan du också få bekräftat om din gissning i slutet av uppgift 2 var riktig. Du kan också kontrollera om dina övriga resultat stämmer med Ross formel (4.14) för P_i.

Det går också att härleda en formel för sannolikheten att Robins djärva strategi leder till framgång, när hens startkapital är 1 krona och vinstsannolikheten är p. Se övning 2 på de blad som delades ut i anslutning till avsnitt 4.6. Jämför svaret i den uppgiften med uttrycket för Kims vinstchans, för samma p och samma startkapital. För vilka värden på p är Kims respektive Robins strategi bäst? Kontrollera att detta stämmer med svaret på uppgift 4 (b).