Chalmers tekniska högskola
Matematiska vetenskaper
MVE090 Matematisk statistik Z3
Projekt 1 VT 2021
Ursprungligen författat av kursens förre lärare Tommy Norberg
Typografiskt uppdaterat av kursens nuvarande lärare Patrik Albin

Simulering av ett enkelt kösystem

Kunder ankommer slumpmässigt till ett betjäningssystem enligt en Poissonprocess med intensitet λ st/minut. Detta innebär att tiderna $T_1, T_2, \ldots$ mellan kundankomster är oberoende och exponentialfördelade med väntevärde $1/\lambda$ minuter. Tiderna $\xi_1, \xi_2, \ldots$ som kund nr $1, 2, \ldots$ tillbringar i betjäning är oberoende och har en fördelning med täthet f (specificeras på sida 2). Denna täthet beror av två parametrar $\alpha > 0$ och $\beta > 0$. Dessutom gäller att betjäningstiderna $\xi_1, \xi_2, \ldots$ är oberoende av ankomstprocessen. (Tänk igenom vad detta innebär praktiskt.)

Tanken med projektet är att du ska simulera detta kösystem med hjälp av MATLAB-programmet som finns att ladda ner ifrån kurshemsidan

`www.math.chalmers.se/Stat/Grundutb/CTH/mve090/2021/Laborations/Matlabkod_Projekt_1.m`

En viss del av programmet ska du själv definiera och koda. I början av det ska du dessutom själv skriva in ett antal parametervärden. Dessa erhålles från kursens projektlärare Anton Johansson via individuella Canvas meddelande i god tid innan arbetet med projektet är tänkt börja efter första maj helgen. Det finns inte två likadana parameteruppsättningar och varje deltagare i kursen måste använda sina egna – det är förbjudet använda samma lapp parametervärden som någon annan.

Det är viktigt att du läser igenom MATLAB-programmet noga och försöker att förstå det. Gör gärna detta tillsammans med någon eller några kompisar. Parametrarna λ, β och α heter i programmet `lbd`, `bta` och `alf`.

Du ska totalt göra två simuleringar. Utdata från varje simulering är en plott (som du ska studera noggrant och förstå—jobba även här gärna i grupp) och en datamängd. Räkna ut L, W och λ_e i båda fallen.[1] Plotten visar antalet kunder i systemet som funtion av tiden under en kort tidsperiod. Datamängden är antalet kunder vid ekvidistanta tidpunkter valda så långt ifrån varandra att de kan anses vara oberoende observationer av ett typiskt kundantal. Estimeringen som ska göras i den andra datamängden kräver betydligt fler observationer än den i den första. I den andra har därför data samplats något tätare i tiden. Du måste själv sätta variablerna `alf`, `bta` och `ant` till `alf2`, `bta2` och `ant2` innan den andra simuleringen. Och innan du gör den första måste du i programmets början på för detta avsedd plats, lägga in dina värden på parametrarna `prm`, `seed`, `lbd`, `alf1`, `bta1`, `alf2`, `bta2` samt `ant1` och `ant2` (du hittar värdena på ditt parameterblad). Programmet går inte att exekvera förrän du gjort ovanstående och dessutom lagt in ytterligare lite kod som du

[1] Se bifogat blad om köer

själv ska ta fram (se uppgift (a) på sida 2). Inga andra förändringar av programkoden får göras.

Till sist: Tänk på att alltid spara data som ligger till grund för beräkningar! Både nu och i din framtida yrkesverksamhet. Du kan ju bli tvingad att verifiera dina resultat (här skattningar och konfidensintervall) i ett senare skede. Klarar du ej detta, ligger du kanske illa till. Det är också viktigt att du följer instruktionerna på sida 2 för projektet noga eftersom inlämnade projekt till stor del kommer att rättas automatiskt.

Uppgifter att utföra och besvara

Antag att

$$f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} \quad \text{för} \quad x > 0$$

för lämpligt valda $\alpha, \beta > 0$. Härled en formel för simulering av betjäningstiden ξ.

(a) Koda formeln i MATLAB. Koden ska omvandla slumptalet u till en simulerad observation x av ξ. Värdet av parametrarna α och β finns redan i variablerna `alf` resp `bta`, så dem ska du inte ändra på utan endast använda. Likaså finns värdet av u redan i variabeln `u`, så inte heller den ska du ändra—endast använda. Koden skall inledas med raden

```
% Kod 1a start
```

fortsättas med att variabeln `x` tilldelas värdet av x och avslutas av raden

```
% Kod 1a slut
```

Lägg in din kod på därför avsedd plats i MATLAB-programmet. Exekvera det med dina egna värden på inparametrarna `prm`, `seed`, `lbd`, `bta1`, `bta2`, `alf1`, `alf2`, `ant1` och `ant2`. Programmets resultat finns i ett endimensionellt fält kallat `data`. Denna data ska du använda till att

(b) punkt- och intervallskatta förväntat antal kunder i systemet under stationära förhållanden. Du kan anta att data är hyfsat normalfördelade (så att c.g.s. kan tillämpas). Önskad konfidensgrad är ca 99%. Svaret skall beräknas med en korrekt decimal på formen `estimat` $\pm$ `error`.

Du får gärna använda standardkommandon som `mean`, `std`, `tinv`, etc i MATLAB. Ändra nu rad 22 `alf=alf1; bta=bta1; ant=ant1;` till `alf=alf2; bta=bta2; ant=ant2;` och exekvera programmet en andra gång. Nu är datamängden betydligt större och du ska använda den till att

(c) punkt- och intervallskatta sannolikheten att antalet kunder i systemet är minst 2. Önskad konfidensgrad är ca 95%. Svaret skall beräknas med tre korrekta decimaler på formen `estimat` $\pm$ `error`.

(d) Är normalapproximation tillåten (varför)?

(e) Beräkna förväntat antal kunder under stationära förhållanden i de två fallen och förklara i ord vad som är orsaken till att kösystemet beter sig så olika.

Den sista frågan kräver att du läser och förstår bifogade kommentarer om kösystem. Obs att frågor på innehållet kan komma på tentamen. Svaren på (a), (b) och (c) ska ej motiveras. Men kom ihåg att spara datamängderna ifall du i ett senare skede måste motivera svaren i (b) eller (c).

Gör inga ändringar i programvaran annat än de du har fått instruktion

Angående projekt 1

Svaren på de fem uppgifterna skall sändas i ett e-mejl till Anton Johansson på e-mejl-adressen `johaant@chalmers.se` med rubriken ("subject") `Z3-Projekt 1 VT 2021-prm`, där du byter ut `prm` mot ditt värde på variabeln. I e-brevet ska finnas personliga uppgifter samt svar på uppgifterna enl följande mall:

```
Namn:
Personnummer:
e-post:

% Kod 1a start

% Kod 1a slut

estb=
errb=

estc=
errc=

svard:

svare:
```

Använd Matlab-syntax i Kod 1a. Värdet av `prm` hittar du på din privata lapp med parametervärden. Glöm inte att ange namn, personnummer och e-postadress. Personnummret ska bestå av 10 siffror utan mellanrum. Mellan raderna `% Kod 1a start` och `% Kod 1a slut` lägger du din kod från deluppgift (a). Variablerna `estb` resp `estc` skall tilldelas skattningen i fråga (b) resp (c). Variablerna `errb` resp `errc` skall tilldelas skattningen av felet i fråga (b) resp (c). Obs konventionen `estx` $\pm$ `errx` som redan nämnts ovan. Utgå gärna ifrån den mall som finns tillgänglig för projektinlämning från kurshemsidan

`www.math.chalmers.se/Stat/Grundutb/CTH/mve090/2021/Laborations/Mall_Projekt_1.txt`

Observera att inlämnade projekt som ej följer ovanstående mall riskerar att hamna i papperskorgen utan åtgärd och utan besked om detta till inlämnaren.

Observera även att projektet senast kan lämnas in läsperiodens sista dag söndagen den 6 juni 2021. För elever som får "retur" på sina projekt (dvs de godkännes ej utan ändringar) kan korrektioner lämnas in till och med (allra senast) söndagen den 13 juni 2021.

Appendix. Några ytterligare kommentarer om köer

Vi har simulerat ett kösystem med oändligt många betjäningsstationer och oändlig kapacitet. Man använder ofta beteckningen G/G/c/K för kösystem, där servicetiderna samt tiderna mellan ankomster är oberoende och likafördelade. Det första G:et betyder att ankomstprocessen är generell, ej Poisson, i vilket fall man skriver M. Det andra G:et betyder att betjäningsfördelningen är generell. Man använder M om betjäningstiderna är exponentialfördelade. Bokstaven c betecknar antalet betjäningsstationer och K systemets kapacitet, d.v.s det maximala antalet kunder i kö och under betjäning. Den sistnämnda brukar uteslutas då kapaciteten är oändlig. Du har simulerat två tämligen olika M/G/∞-köer.

Teorin för M/M/c/K-köer är enkel och man kan t.ex visa i fallet $c < \infty$, $K = \infty$, att kösystemet är stationärt precis då betjäningsgraden $\rho = (\lambda\mu)/c < 1$ (här är μ väntevärdet av en typisk betjäningstid) samt att det inte exploderar om $\rho \leq 1$. Man säger att kön exploderar om totala antalet kunder som är under betjäning växer obegränsat.Att en kö är stationär betyder bl.a att tiden tills dess att den är tom har ändligt väntevärde. Att det är så i det först simulerade fallet tror man kanske inte då man ser plotten.

Köer med begränsad kapacitet ($K < \infty$) är alltid stationära eftersom kunder som anländer då kön är full avviker och aldrig kommer tillbaka. För sådana köer är istället sannolikheten $P(N = K)$ att kösystemet är fullt av intresse. Här betecknar N det totala antalet kunder i systemet vid en godtycklig tidpunkt, så att om, t.ex, $P(N = K) = 0.35$ så är systemet fullt 35% av tiden och effektiv ankomstintensitet, λ_e, är lika med 0.65λ, ty det är bara då $N < K$ som kunder tas emot.

Mycket forskningstid gick åt under 1900-talet till att generalisera naturliga resultat, som är enkla att visa för M/M/c/K-köer, till G/G/c/K-fallet. En viktig insats publicerades av J. D. C. Little 1961. Little visade att

$$L/W = L_q/W_q = \lambda_e$$

gäller för i princip alla stationära G/G/c/K-köer. Här är L_q förväntat antal kunder i kö och L är förväntat totalt antal kunder i kösystemet (alltså de i kö plus de under betjäning). Vidare är W_q förväntad tid i kö och $W - W_q$ förväntad betjäningstid. Dessutom är λ_e effektiv ankomstintensitet (som är $= P(N < K)\lambda < \lambda$ om $K < \infty$).

För M/G/∞-kön som vi simulerat gäller att $L/W = \lambda$. För köer med $c = \infty$ är ju $L_q = 0$ och $W_q = 0$. Alla anländande kunder blir ju direkt betjänade och $\lambda_e = \lambda$ eftersom inga kunder avvisas.