

Simulering av Poissonprocesser

Olle Nerman, 2015-09-28

Gruppprojekt i MSG110, GU HT 2015 (max 5 personer/grupp)

Frågeställning: Hur åstadkommer man en realisering av en Poissonprocess på ett tidsintervall av given längd?

I detta laborationsprojekt skall vi jämföra två simuleringsmetoder för Poissonprocesser. För att vi skall få en bättre förståelse för simulering i största allmänhet så skall vi utgå från (pseudo-) slumpstal med oberoende sekvenser av de enklaste slumpstalen som är likformigt fördelade på intervallet $[0,1]$ och som ju typiskt finns tillgängliga ganska allmänt i diverse datorprogram och programmerbara miniräknare av alla de slag. Vi skall använda Matlab. Själva laborationsdelen är indelad i ett antal praktiska och konceptuella steg:

1. Hur åstadkommer man en följd av oberoende stokastiska variabler $X_1, X_2, X_3, \dots$ som är exponentialfördelade med givet väntevärde λ utgående från en sekvens av oberoende stokastiska variabler $U_1, U_2, U_3, \dots$ som alla är likformigt fördelade på intervallet $[0,1]$ teoretiskt? Börja läsa i kursens lärobok i kapitel 5. Hur gör man för att implementera praktiskt i Matlab? Denna andra fråga skall du besvara genom att implementera den teoretiska metoden ovan i en Matlab algoritm som genererar **300** oberoende exponentialfördelade X-variabler med väntevärde 2. Illustrera resultatet med ett histogram.
2. En Poissonprocess på $[0, \infty)$ med intensiteten λ karakteriseras av att tid till första händelsetidpunkten och alla tidsmellanrummen

mellan alla de succesiva händelsetidpunkterna är oberoende och exponentialfördelade λ . Låt $N(t)$ beteckna antalet pulser i en Poissonprocess med intensiteten $\lambda = 1/2$ för t i intervallet $[0, 100]$ och simulera en realisering n av $N(t)$ i intervallet med hjälp av X -variablerna ovan. Illustrera med ett funktionsdiagram med realiseringen av $n(t)$.

3. Totala antalet händelsetidpunkter i hela intervallet mellan 0 och 100 är Poissonfördelad med parameter $100 * 2 = 200$. Om man betingar med antalet händelsetidpunkter i hela intervallet $= k$, så är de succesiva pulserna i intervallet fördelade som de storleksordnade tidpunkterna i ett stickprov av storleken k från en likformig fördelning på intervallet $[0, 100]$. Leta upp en lämplig referens. Använd sedan detta resultat för att simulera en realisering av $N(t)$ på intervallet, genom att först simulera en observation från en Poissonfördelning (200) och sedan ett stickprov från en likformig fördelning på $[0, 100]$. Återigen skall ni utgå från slumpvalssekvenser från intervallet $[0, 1]$.
4. Man kan tänka sig många olika sätt att jämföra de simulerade processernas fördelningar, empiriskt med varandra och med teoretiska Poissonprocessfördelningar. Vi skall bara göra en enda sådan jämförelse genom att studera fördelningen för $Y =$ den 10:e händelsetidpunkten.

Vi börjar med att studera empiriska fördelningsfunktionerna för Y i processer som är simulerade med respektive metod. Simulera (t.ex med en lämplig loop) **2000** oberoende observationer av Y med respektive metod och rita in de båda empiriska fördelningsfunktionerna för respektive simuleringsmetod i ett och samma diagram.

Teorin säger att den tionde pulsen skall följa en viss gammafördelning (fördelningen är egentligen aningen störd av att vi inte har pulser efter tid $t=100$ i våra realiseringar).

Vad är parametrarna i denna gammafördelning?

Använd dessa parametrar och jämför i ett histogram empiriska fördelningen för stickprovet genererat av den andra metoden med den teoretiska sannolikhetsstätheten för gammafördelningen.

Studiedel:

Börja som sagt läsa om Monte Carlo simulering i läroboken och komplettera med läsning om metoderna i andra källor som ni själva letar upp. Observera att den så kallade inversa fördelningsfunktions-tekniken kräver lite modifiering när man skall simulera en diskret fördelning.

Skrivprojekt: Uppgiften är att i en ca **3-5** sidor lång text skriva om metoderna för att simulera en godtycklig exponentialfördelning respektive Poissonfördelning. För att få en bra struktur på hela texten skall ni använda en disposition med rubrikerna:

- **Allmän kort Introduktion:** Vad har man för nytta av (Monte Carlo) simulering. Vad är pseudoslumtpal . Hur genereras dessa i MatLab. Exempel på en tänkbar användning av simulering av en Poissonprocess för att studera någon lagom krånglig egenskap hos processen.
- **Sammanfattning:** Kort sammanfattning (ett "abstract") av rapportens syfte och innehåll.

- **Metodbeskrivningar** Presentation av de båda simuleringsmetoderna. I samband med detta skall ni presentera den teoretiska motiveringen för de båda simuleringsmetoderna med referenser till teori i boken eller från andra lämpliga källor. Redovisa också kort teorin för fördelningen för **10:e** händelsetidpunkten.
- **Implementering** (av alla metoderna i MatLab; redovisa med kod och lämplig beskrivning) Nästa steg är att presentera MatLab-implementeringen av de båda metoderna (Redovisa med kod och lämplig beskrivning). Redovisa också implementeringen av de två studierna av fördelningsjämförelser för **10:e** händelsetidpunkten.
- **Resultat:** Nästa steg är att presentera de succesiva resultaten av alla de använda metoderna i en resultatdel. Presentera alla resultaten av simuleringarna med diagrammen och bildtexter till dessa..
- **Diskussion:** Kort diskussion av implementeringsresultaten: Eventuella för och nackdelar med de två metoderna, andra valideringsmöjligheter än de implementerade o.s.v .
- **Referenser:** Välj ett standardiserat referenssystem.

Inskickning av grupparbetet skall ske senast 16 Oktober 12.00 till:

nerman.chalmers@analys.urkund.se

Urkund gör att ni måste formulera er själva och citera, om ni använder andras formuleringar. Annars blir det plagiat.

Lycka Till!

Olle

