Tommy Norberg
Matematisk statistik
Chalmers & GU
1 oktober 2004

Introduktion till stokastisk simulering

Vi lär oss först en standardmetod som alltid går att tillämpa, men som kan vara beräkningsmässigt ineffektiv i vissa situationer. Sedan ska vi titta på hur man med probabilistiska metoder kan hantera flera viktiga specialfall.

Slumptal

Slumptal $u_1, u_2, \ldots$ genereras rekursivt i datorer av speciella programkommandon. I Matlab tilldelar kommandot `rn=rand(1,5)` fältet `rn` en vektor bestående av 5 st slumptal. Observera att sekvensen $u_1, u_2, \ldots$ är helt deterministisk. Den är dock konstruerad i syfte att verka totalt slumpmässig och så att varje u_i antar varje värde i enhetsintervallet $(0,1)$ med lika stor sannolikhet. Detta kan göras mer eller mindre bra och man bör vara uppmärksam när man börjar arbeta i ett nytt programspråk, vars slumptalsgenerator man inte har tidigare erfarenhet av.

Det innebär praktiskt att om man har låtit sitt program generera 5 st slumptal $u_1, \ldots, u_5$, så kan vi tänka oss dem som 5 st oberoende observationer av en stokastisk variabel U, som är likformigt fördelad på $(0,1)$. Jag påminner om att U är likformigt fördelad på $(0,1)$, vilket ofta skrives $U \sim \mathrm{U}(0,1)$, om U:s täthetsfunktion är

$$f(u) = \begin{cases} 1 & \text{om} \quad 0 < u < 1 \\ 0 & \text{f.ö} \end{cases}$$

Notera också att U:s fördelningsfunktion är

$$F(u) = P(U \leq u) = \begin{cases} 0 & \text{om} \quad u \leq 0 \\ u & \text{om} \quad 0 < u < 1 \\ 1 & \text{om} \quad u \geq 1 \end{cases}$$

Att u_1, u_2 är oberoende observationer av $U \sim \mathrm{U}(0,1)$, betyder att den gemensamma tätheten för u_1, u_2 faktoriseras enligt

$$f_{U_1,U_2}(u_1,u_2) = f_{U_1}(u_1) f_{U_2}(u_2) = 1$$

I situationer som denna är det ofta enklare (och bättre) att bara skriva $f(u_1,u_2)$ istället för $f_{U_1,U_2}(u_1,u_2)$, $f(u_1)$ istället för $f_{U_1}(u_1)$ och $f(u_2)$ istället för $f_{U_2}(u_2)$. Att u_1, u_2 är oberoende observationer innebär då helt enkelt att

$$f(u_1,u_2) = f(u_1) f(u_2)$$

Ett matematiskt ekvivalent sätt att uttrycka oberoendet är att den betingade fördelningen för den andra, U_2, givet att den första, U_1, erhöll värdet u_1, ej beror av u_1. M.a.o

$$f_{U_2|U_1=u_1}(u_2) = f_{U_2}(u_2) = 1 \quad \text{för} \quad 0 < u_2 < 1$$

(I ord: den betingade tätheten för U_2, givet att $U_1 = u_1$, beror ej av u_1.) Beteckningen $f_{U_2|U_1=u_1}(u_2)$ är klumpig. Ofta är det tydligare att istället skriva $f(u_2|u_1)$ eller $f(u_2|U_1 = u_1)$, trots att åtminstone den först nämnda av många matematiker anses slarvig. Att $f(u_2|u_1)$ ej beror av u_1 innebär då helt enkelt att

$$f(u_2|u_1) = f(u_2)$$

där vi som ovan skrivit $f(u_2)$ istället för det klumpigare $f_{U_2}(u_2)$.

Man kan mycket väl tänka sig att två $\mathrm{U}(0,1)$-fördelade variabler ej är oberoende. Ett trivialt exempel fås genom att låta U_1 vara $\mathrm{U}(0,1)$-fördelad och sätta $U_2 = U_1$. Nedan följer ett liknande, men aningen mer komplicerat exempel på två beroende $\mathrm{U}(0,1)$-fördelade variabler.

Exempel 1 Antag att U_1, U_2 har tätheten $f(u_1, u_2) = 2$ då $0 < u_1, u_2 < 0.5$ eller $0.5 \leq u_1, u_2 < 1$. M.a.o., U_1, U_2 är likformigt fördelad på unionen av de två rektanglarna $(0, 0.5) \times (0, 0.5)$ och $[0.5, 1) \times [0.5, 1)$. Då gäller att både U_1 och U_2 är $\mathrm{U}(0,1)$. (Rita figur och tänk!) Marginalfördelningarna är alltså $\mathrm{U}(0,1)$. Men observera att U_1 och U_2 ej är oberoende. Har man observerat värdet av U_1 så vet man vilken av händelserna $0 < U_2 < 0.5$ och $0.5 \leq U_2 < 1$ som har inträffat. $\square$

Övning 1 Visa matematiskt att marginalfördelningarna för den bivariata fördelningen i exemplet ovan är just $\mathrm{U}(0,1)$.

Övning 2 Låt paret U_1, U_2 vara likformigt fördelat på

$$S = \{(u_1, u_2) : 0 < u_1,\ 0 < u_2,\ u_1 + u_2 < 1\}$$

Är marginalfördelningarna $\mathrm{U}(0,1)$? Är U_1, U_2 oberoende?

Simulering av diskreta stokastiska variabler

Antag att X kan anta de tre värdena $1, 2, 3$ med sannolikheterna $p_1 = 0.25, p_2 = 0.40, p_3 = 0.35$. Vi börjar med att konstruera en stokastisk variabel med samma fördelning som X. Tag $U \sim \mathrm{U}(0,1)$ och definiera

$$\widetilde{X} = \begin{cases} 1 & \text{då} \quad U \leq 0.25 \\ 2 & \text{då} \quad 0.25 < U \leq 0.65 \\ 3 & \text{då} \quad U > 0.65 \end{cases}$$

Att X och $\widetilde{X}$ har samma fördelning kan skrivas $\widetilde{X} \stackrel{d}{=} X$.

Om u nu är ett slumptal, så är u, enligt ovan, en simulerad observation av U. Men då är ju x, definierad av

$$x = \begin{cases} 1 & \text{då} \quad u \leq 0.25 \\ 2 & \text{då} \quad 0.25 < u \leq 0.65 \\ 3 & \text{då} \quad u > 0.65 \end{cases}$$

en simulerad observation av $\widetilde{X}$ och också av X eftersom $\widetilde{X} \stackrel{d}{=} X$.

Om nu X allmänt har utfallsrummet $S = \{x_1, x_2, \ldots\}$ och $P(X = x_i) = p_i$, så kan vi simulera en observation x av X, genom att generera ett slumptal u och låta

$$x = \begin{cases} x_1 & \text{då} \quad u \leq p_1 \\ x_2 & \text{då} \quad p_1 < u \leq p_1 + p_2 \\ x_3 & \text{då} \quad p_1 + p_2 < u \leq p_1 + p_2 + p_3 \\ \vdots & \end{cases}$$

Något bevis av detta behövs knappast. Det räcker att referera till argumentationen i specialfallet ovan.

I några övningar i detta häfte ska du ge som svar en algoritm för någon slags stokastisk simulering. Åtminstone i en del fall är det praktiskt att ge svaret i form av ett Matlab-program.

Övning 3 (a) Ge en algoritm för simulering av ett kast med en symmetrisk sex-sidig tärning. (b) Antag att tärningen är osymmetrisk på ett sådant sätt att sannolikheten för 6:a är förhöjd med 10% och sannolikheten för 1:a är minskad med motsvarande belopp och övriga sidor har oförändrade sannolikheter. Ge en algoritm för simulering av kast med en sådan tärning.

Övning 4 Ge en algoritm för simulering av en $\mathrm{Bin}(4, 0.2)$-fördelad stokastisk variabel.

Övning 5 Ge en algoritm för simulering av en $\mathrm{Poi}(2)$-fördelad stokastisk variabel.

Vissa stokastiska variabler, som de $\mathrm{Bin}(n, p)$-fördelade, har en probabilistisk tolkning som man kan använda i en simulering. För att exemplifiera tekniken ska vi ge en alternativ lösning till övning 4 ovan. Låt $X \sim \mathrm{Bin}(4, 0.2)$. Då kan vi tänka oss X som antalet lyckade försök utav $n = 4$ oberoende, där sannolikheten för att ett visst försök ska lyckas är $p = 0.2$. Generera 4 slumptal $u_1, \ldots, u_4$ och låt x vara antalet i sådana att $u_i \leq p = 0.2$. Då är x en simulerad observation av $\mathrm{Bin}(4, 0.2)$-fördelningen.

Matlab-tips

I "Statistics toolbox" finns kommandon för beräkning av de vanligaste av de diskreta mass- och fördelningsfunktionerna. Kommandot `help stats` ger en förteckning över alla tillgängliga kommandon. Jag nöjer mig med att nämna här `binopdf` och `binocdf` samt `poisspdf` och `poisscdf`, med vars hjälp man kan beräkna binomial- och Poissonsannolikheter. Använd `help`-funktionen för att ta reda på hur de ska användas och vad de gör.

I Matlab finns även kommandon för generering av t.ex binomialfördelade och Poissonfördelade slumptal. Det är inte meningen att man ska använda denna typ av kommandon i lösningarna till övningarna i detta häfte. Övningarna syftar ju till att man ska bygga upp en viss förståelse för hur sådana kommandon fungerar.

Simulering av kontinuerliga stokastiska variabler

Antag nu att den stokastiska variabeln X är kontinuerligt fördelad med täthet $f(x)$ och fördelningsfunktion $F(x)$. Vi börjar med att konstatera att om $X \stackrel{d}{=} h(U)$, där $U \sim \mathrm{U}(0,1)$, så är $x = h(u)$ en simulerad observation av X om u är ett slumptal.

Sambandet mellan $f(x)$ och $F(x)$ är

$$f(x) = F'(x) \quad \Leftrightarrow \quad F(x) = \int_{-\infty}^{x} f(y)\,dy$$

Dessutom gäller

$$F(x) = P(X \leq x)$$

Notera att $F(x)$ är strikt växande i X:s utfallsrum. Då existerar en entydigt bestämd invers $F^{-1}(p)$ (som ibland kallas *kvantilfunktionen*), definierad för $0 < p < 1$ av att

$$p = F(x) \quad \Leftrightarrow \quad x = F^{-1}(p)$$

och det gäller, eftersom $F(x)$ är strikt växande och kontinuerlig, att

$$F^{-1}(p) \leq x \quad \Leftrightarrow \quad p \leq F(x)$$

Det finns ett resultat, som ibland kallas *simulerarnas huvudsats*, som säger att om $U \sim \mathrm{U}(0,1)$, så följer att den stokastiska variabeln $\widetilde{X} = F^{-1}(U)$ har samma fördelning som X. Beviset utnyttjar att utsagorna (påståendena) $F^{-1}(p) \leq x$ och $p \leq F(x)$ är ekvivalenta och är mycket enkelt:

$$P(\widetilde{X} \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x) = P(X \leq x)$$

Två variabler som har samma fördelningsfunktion har naturligtvis också samma täthet. Den senare är ju derivatan av den förra.

Resultatet säger att om u är ett slumptal och $x = F^{-1}(u)$, så är x en simulerad observation av den stokastiska variabeln X. Observera att x fås genom att man löser ekvationen $u = F(x)$ för ett givet slumptal u. Ibland är detta mycket enkelt och kan göras en gång för alla. Ibland får man ta till numeriska lösningsmetoder.

Exempel 2 Antag att $T \sim \mathrm{Exp}(\lambda)$. Då är T:s täthet och fördelningsfunktion

$$f(t) = \lambda e^{-\lambda t} \quad \text{för} \quad t > 0$$

$$F(t) = \int_0^t \lambda e^{-\lambda s}\,ds = 1 - e^{-\lambda t} \quad \text{för} \quad t > 0$$

Observera nu att

$$u = F(t) \quad \Leftrightarrow \quad t = -\frac{1}{\lambda}\ln(1-u)$$

Således är $t = -(1/\lambda)\ln(1-u)$ en simulerad observation av T om u är ett slumptal. Men notera att om u är ett slumptal, så är $1-u$ det likaså. Alltså kan vi i formeln ovan ersätta $1-u$ med u, och erhåller den ur beräkningssynpunkt något mer ekonomiska regeln

$$t = -\frac{1}{\lambda}\ln u$$

för simulering av $\mathrm{Exp}(\lambda)$. □

Ibland behöver man för att visa eller kanske snarare motbevisa vissa påståenden simulera från en fördelning som varken har väntevärde eller varians. En sådan fördelning är Cauchyfördelningen.

Exempel 3 Den s.k Cauchyfördelningen definieras av att tätheten är

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{för} \quad -\infty < x < \infty$$

Man kan visa genom att integrera fram fördelningsfunktionen $F(x)$ och invertera denna att $\tan\phi$ är Cauchyfördelad om ϕ är likformigt fördelad på intervallet $(-\pi/2,\ \pi/2)$. Således gäller, om u är ett slumptal, att

$$x = \tan\left[\pi(u - 0.5)\right]$$

är en simulering av en observation från Cauchyfördelningen. □

Ibland går det emellertid inte att explicit lösa x ur $u = F(x)$. Då får man ta till numeriska ekvationslösningsverktyg. Ett exempel på detta är när X är normalfördelad, ty då är

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(y-\mu)^2/2\sigma^2}\, dy$$

Inversen till denna funktion går inte att explicit beräkna, men det är naturligtvis inget problem att numeriskt lösa x ur

$$u = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(y-\mu)^2/2\sigma^2}\, dy$$

för slumptal u. Matlab har ett kommando som heter `norminv` som gör just detta. Att det finns andra enklare sätt att simulera normalfördelade observationer återkommer vi till.

De flesta slumptalsgeneratorer behöver initieras. Man har då möjlighet att antingen ge ett förutbestämt s.k frö ("seed" på engelska). Fördelen med detta är att man ska kunna upprepa en simulering och få exakt samma resultat. En andra möjlighet är att ge ett totalt slumpmässing frö. Det bör man göra vid produktionskörning, men det kan vara en god idé att spara fröet, så att körningen kan göras om. I Matlab kan man ge ett slumpmässigt frö så här: `seed=sum(100*clock); rand('state',seed);`.

Övning 6 Den s.k Paretofördelningen används ofta av försäkringsbolag. Den parametriseras något olika i olika sammanhang. I något har jag valt att parametrisera den så att tätheten blir

$$f(x) = \frac{\alpha - 1}{x_T}\left(\frac{x}{x_T}\right)^{-\alpha} \quad \text{för} \quad x \geq x_T$$

där $x_T > 0$ och $\alpha > -1$. Härled en algoritm för simulering av observationer av stokastiska variabler med denna täthet.

Övning 7 Den s.k generaliserade Paretofördelning har fördelningsfunktionen

$$F(x) = 1 - \left(1 + \gamma\frac{x}{\sigma}\right)^{-1/\gamma} \quad \text{för} \quad x > 0 \text{ sådana att } 1 + \gamma\frac{x}{\sigma} > 0$$

För parametrarna σ (som ger skalan) och γ (som ger fördelningens form) gäller

$$\sigma > 0, \quad -\infty < \gamma < \infty$$

Observera att då $\gamma = 0$ är $F(x) = 1 - \lim_{\gamma \to 0}(1 + \gamma x/\sigma)^{-1/\gamma} = 1 - e^{-x/\sigma}$ för $x > 0$. Det är bara då $\gamma < 0$ som kravet $1 + \gamma x/\sigma > 0$ innebär en begränsning. Då gäller att en generaliserad Paretovariabel har utfallsrummet $0 < x < -\sigma/\gamma$. Annars är utfallsrummet $x > 0$. Härled en algoritm för simulering av observationer från den generaliserade Paretofördelningen.

Simulering av oberoende observationer

Antag att vi ska simulera n oberoende observationer av en stokastisk variabel X med täthet (eller massfunktion) $f(x)$ och fördelningsfunktion

$$F(x) = \int_{-\infty}^{x} f(z)\, dz$$

Då genererar vi helt enkelt n slumptal u_i och löser x_i ur

$$u_i = F(x_i)$$

Då är $x_1, \ldots, x_n$ simulerade oberoende observationer av X.

Övning 8 Skriv en algoritm (eller ett Matlab-program) för simulering av n oberoende observationer av en Exp(λ)-variabel.

Övning 9 Skriv en algoritm (eller ett Matlab-program) för simulering av oberoende observationer $x_1, x_2, \ldots$ av en Exp(λ)-variabel, som avbryts då $\sum_{i=1}^{n} x_i > 1$ första gången.

Simulering av beroende observationer

Låt $f(x, y)$ vara en bivariat täthet för de två kontinuerliga stokastiska variablerna X och Y. Definiera

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

där

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

Då är $f(y|x)$ är den betingade tätheten för y givet att $X = x$ ($f(x)$ är X:s täthet). Motsvarande fördelningsfunktioner ska vi beteckna $F(y|x)$. Låt dessutom $F(x)$ beteckna X:s fördelningsfunktion.

Börja simuleringen av det bivariata paret X, Y med att simulera en observation x av X. Enligt vad som sagts ovan görs detta genom att vi genererar ett slumptal u och löser x ur ekvationen $u = F(x)$. Nu kan vi tänka oss att vi vet att $X = x$ och Y:s betingade täthet är då $f(y|x)$. Simulerar vi en observation y från denna täthet, så följer att x, y är en simulerad observation av X, Y. Detta sker genom att vi genererar ytterligare ett slumptal v och sedan löser y ur ekvationen $v = F(y|x)$.

Alltså: vid simulering av två stokastiska variabler X, Y, generera två slumptal u, v och lös först x ur $u = F(x)$ och sedan y ur $v = F(y|x)$.

Antag nu att $f(x, y, z)$ är täthet för trippeln X, Y, Z. Vi bildar först X, Y:s gemensamma täthet

$$f(x, y) = \int_{-\infty}^{\infty} f(x, y, z)\, dz$$

Sedan bildar vi den betingade fördelningen för Y, givat att $X = x, Y = y$, enligt

$$f(z|x, y) = \frac{f(x, y, z)}{f(x, y)}$$

Låt $F(z|x, y)$ vara motsvarande fördelningsfunktion. Ska vi simulera en observation av trippeln X, Y, Z börjar vi med att generera två slumptal u, v och räkna fram x, y enligt ovan. Sedan genererar vi ytterligare ett slumptal w och löser z ur ekvationen $w = F(z|x, y)$. Då är trippeln x, y, z en simulerad observation av X, Y, Z.

Alltså: vid simulering av tre stokastiska variabler X, Y, Z, generera tre slumptal u, v, w och lös först x ur $u = F(x)$, sedan y ur $v = F(y|x)$ och till sist z ur $w = F(z|x, y)$.

Medelst iterering av ovanstående kan man simulera beroende observationer från fler än 3 stokastiska variabler.

Probabilistiska metoder

Ovan har vi lärt oss generell metod att simulera oberoende såväl som beroende observationer av slumpmässiga fenomen. Nu ska vi titta på några speciella situationer.

Simulering av $X \sim \mathbf{Bin}(n, p)$

Vi kan tänka oss att X är antalet lyckade utav n oberoende försök, där varje försök lyckas med sannolikheten p. Generera n slumptal $u_1, \ldots, u_n$. Antalet i sådana att $u_i \leq p$ är en simulerad observation av $\mathrm{Bin}(n, p)$-fördelningen.

Simulering av $X \sim \mathbf{Geo}(p)$

Generera succesivt slumptal u_i. Sluta då $u_i \leq p$ inträffar för första gången. Antalet genererade slumptal u_i är en simulerad observation av $\mathrm{Geo}(p)$-fördelningen.

Simulering av $X \sim \mathbf{Poi}(\lambda)$

Beräkna successivt $x_i = -\ln u_i$, där $u_1, u_2, \ldots$ är slumptal. Sluta då $\sum_{i=1}^{n} x_i > \lambda$ för första gången. J.f.r övning 9. Om n är antalet genererade slumptal, så är $n - 1$ en simulerad observation av $\mathrm{Poi}(\lambda)$.

Simulering av $X \sim \mathbf{N}(\mu, \sigma)$

Använd Box-Mullers metod, som säger att z_1, z_2, givna av

$$z_1 = \sqrt{-2 \ln u_1} \cos 2\pi u_2$$
$$z_2 = \sqrt{-2 \ln u_1} \sin 2\pi u_2$$

kan betraktas som två oberoende observationer av $\mathrm{N}(0,1)$ om u_1, u_2 är slumptal. Då är $x_1 = \mu + \sigma z_1$ och $x_2 = \mu + \sigma z_2$ en simulering av två oberoende observationer av $\mathrm{N}(\mu, \sigma)$.

Box-Muller visade att två stokastiska variabler X, Y är oberoende $\mathrm{N}(0,1)$-fördelade om, och endast om, motsvarande vektors längd

$$R = \sqrt{X^2 + Y^2}$$

och vinkel mot X-axeln

$$\phi = \arctan \frac{X}{Y}$$

är oberoende och sådana att ϕ är $\mathrm{U}[0, 2\pi)$-fördelad medan R:s täthet är

$$f(r) = re^{-r^2/2} \quad \text{för} \quad r > 0$$

Genom att simulera en observation $r = \sqrt{-2 \ln u_1}$ av R och en oberoende observation $2\pi u_2$ av ϕ och gå från polära koordinater (r, ϕ) till kartesiska (x, y) får vi då enl Box-Mullers resultat att x, y är två oberoende observationer från $\mathrm{N}(0,1)$-fördelningen.

Simulering av $X \sim$ LN(μ, σ)

Använd att $X \sim \mathrm{LN}(\mu, \sigma)$ om, och endast om, $Y = \ln X \sim \mathrm{N}(\mu, \sigma)$. Så simulera en observation y av Y, t.ex m.h.a Box-Mullers metod (se ovan), och låt $x = e^y$. Då är x en simulerad observation av $X \sim \mathrm{LN}(\mu, \sigma)$.

Simulering av $X_1, X_2 \sim$ N$(\mu_1, \mu_2; \sigma_1, \sigma_2, \rho)$

Använd Box-Mullers metod (se ovan) för att simulera två oberoende observationer z_1, z_2 av $\mathrm{N}(0,1)$, d.v.s, låt

$$\begin{aligned} z_1 &= \sqrt{-2 \ln u_1} \cos 2\pi u_2 \\ z_2 &= \sqrt{-2 \ln u_1} \sin 2\pi u_2 \end{aligned}$$

där u_1, u_2 är slumptal. Bilda sedan

$$\begin{aligned} x_1 &= \mu_1 + \sigma_1 z_1 \\ x_2 &= \mu_2 + \sigma_2 \left(\rho\, z_1 + \sqrt{1 - \rho^2}\, z_2 \right) \end{aligned}$$

Då är x_1, x_2 en simulerad observation av $\mathrm{N}(\mu_1, \mu_2; \sigma_1, \sigma_2, \rho)$. Se övning 10 nedan. Observera att $-1 \leq \rho \leq 1$ och att det är bara då $\rho = 0$ som observationerna är oberoende. Observera också att observationerna är linjärt beroende ifall $\rho = \pm 1$.

Övning 10 Visa att om Z_1, Z_2 är $\mathrm{N}(0,1)$ och oberoende, så är paret X_1, X_2, givet av

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 Z_1 \\ X_2 &= \mu_2 + \sigma_2 \rho\, Z_1 + \sigma_2 \sqrt{1 - \rho^2}\, Z_2 \end{aligned}$$

$\mathrm{N}(\mu_1, \mu_2; \sigma_1, \sigma_2, \rho)$-fördelat. (Ledning: Linjärkombinationer av oberoende normalfördelade variabler är normalfördelade.)

Övning 11 Visa att om $X_1, X_2 \sim \mathrm{N}(\mu_1, \mu_2; \sigma_1, \sigma_2, \rho)$, så är Z_1, Z_2, givna av

$$Z_1 = \frac{X_1 - \mu_1}{\sigma_1}$$
$$Z_2 = \frac{1}{\sqrt{1-\rho^2}} \left(\frac{X_2 - \mu_2}{\sigma_2} - \rho \, \frac{X_1 - \mu_1}{\sigma_1} \right)$$

standardiserat normalfördelade och oberoende. (Ledning: Linjärkombinationer av X_1, X_2 är normalfördelade.)

Litteraturhänvisningar

Jay Devore & Nicholas Farnum: Applied Statistics for Engineers and Scientists. Duxbury press 1999. Begreppet slumptalsgenerator finns nämnt på sidan 159. I övrigt finns inget om stokastiska simulering i denna lärobok.

J Susan Milton & Jesse C Arnold: Introduction to Probability and Statistics; Principles and applications for engineering and the computing sciences. Fourth edition. McGraw-Hill 2003. Läs om simulering av diskreta stokastiska variabler i avsnitt 3.9 på sidorna 78-90. Simulering av kontinuerliga variabler finns beskrivet i avsnitt 4.9 på sidan 134.

G. E. P. Box & Mervin E. Muller: A note on the generation of random normal deviates. Annals of Mathematical Statistics, vol **29** (1958), pp 610-611.

Svar eller lösningar till vissa övningar

1) Av symmetrin följer att $U_1 \stackrel{d}{=} U_2$. Den förras täthet är $f(u_1) = \int_{-\infty}^{\infty} f(u_1, u_2)\, du_2$. Om $0 < u_1 < 0.5$ får vi att $f(u_1) = \int_0^{0.5} 2\, du_2 = 1$ och om $0.5 \leq u_1 < 1$ får vi istället $f(u_1) = \int_{0.5}^{1} 2\, du_2 = 1$. Vi ser att $f(u_1) = 1$ för $0 < u_1 < 1$, vilket visar att $U_1 \sim \mathrm{U}(0,1)$.

2) Marginal fördelningarna är ej $\mathrm{U}(0,1)$, U_1, U_2 är ej oberoende.

3) (a) $x = [6u - 1]$, där $[y]$ betyder heltalsdelen av y. (b) T.ex $x = 1 \Leftrightarrow u \leq 9/60$, $x = 6 \Leftrightarrow 9/60 < u \leq 20/60 = 2/6$, $x = 3 \Leftrightarrow 2/6 < u \leq 3/6$, $x = 4 \Leftrightarrow 3/6 < u \leq 4/6$, $x = 5 \Leftrightarrow 4/6 < u \leq 5/6$, $x = 6 \Leftrightarrow u > 5/6$.

4) $\mathrm{Bin}(4, 0.2)$ har massfunktionen $p(k) = 4!/(k!(4-k)!)\, 0.2^k\, 0.8^{4-k}$ för $k = 0, 1, 2, 3, 4$. Låt $x = 0$ då $u \leq p(0) = 0.4096$, $x = 1$ då $0.4096 < u \leq p(0) + p(1) = 0.8192$, $x = 2$ då $0.8192 < u \leq p(0) + p(1) + p(2) = 0.9728$, $x = 3$ då $0.9728 < u \leq p(0) + p(1) + p(2) + p(3) = 0.9984$ och $x = 4$ då $u > 0.9984$.

5) $\mathrm{Poi}(2)$ har massfunktionen $p(k) = e^{-2} 2^k / k!$ för $k = 0, 1, \ldots$. Om vi nöjer oss med 4 decimalers noggranhet blir algoritmen: Låt $x = 0$ då $u \leq p(0) = 1353$, $x = 1$ då $0.1353 < u \leq p(0) + p(1) = 0.4060$, $x = 2$ då $0.4060 < u \leq \sum_{i=0}^{2} p(i) = 0.6767$, $x = 3$ då $0.6767 < u \leq \sum_{i=0}^{3} p(i) = 0.8571$, $x = 4$ då $0.8571 < u \leq \sum_{i=0}^{4} p(i) = 0.9473$, $x = 5$ då $0.9473 < u \leq \sum_{i=0}^{5} p(i) = 0.9834$, $x = 6$ då $0.9834 < u \leq \sum_{i=0}^{6} p(i) = 0.9955$, $x = 7$ då $0.9955 < u \leq \sum_{i=0}^{7} p(i) = 0.9989$, $x = 8$ då $0.9989 < u \leq \sum_{i=0}^{8} p(i) = 0.9998$, $x = 9$ då $0.9998 < u \leq= 1.0000$. Denna algoritm har en öppenbar nackdel.

6) Paretofördelningens fördelningsfunktion är $F(x) = \int_{x_T}^{x} \frac{\alpha-1}{x_T} \left(\frac{x}{x_T}\right)^{-\alpha} dx = 1 - (x/x_T)^{-\alpha}$ för $x > x_T$. Låt u beteckna ett slumptal. Löser vi ut x ur $u = F(x)$, så får vi att $x = x_T (1-u)^{-1/(\alpha-1)}$ och precis som för exponentialfördelningen kan vi byta $1 - u$ mot u och erhåller $x = x_T u^{-1/(\alpha-1)}$.

7) $x = \sigma \left(u^{-\gamma} - 1\right)/\gamma$

10) Enligt ledningen är paret X_1, X_2 N-fördelat. Notera nu att $E[X_1] = E[\mu_1 + \sigma_1 Z_1] = \mu_1 + \sigma_1 E[Z_1] = \mu_1$ och $E[X_2] = E[\mu_2 + \sigma_2 \rho\, Z_1 + \sigma_2 \sqrt{1-\rho^2}\, Z_2] = \mu_2 + \sigma_2 \rho E[Z_1] + \sigma_2 \sqrt{1-\rho^2} E[Z_2] = \mu_2$. Variablerna har alltså rätt väntevärden. Att de har rätt standardavvikelser ser vi ur $\mathrm{Var}[X_1] = \mathrm{Var}[\mu_1 + \sigma_1 Z_1] = \mathrm{Var}[\sigma_1 Z_1] = \sigma_1^2 \mathrm{Var}[Z_1] = \sigma_1^2$ och $\mathrm{Var}[X_2] = \mathrm{Var}[\mu_2 + \sigma_2 \rho\, Z_1 + \sigma_2 \sqrt{1-\rho^2}\, Z_2] = \mathrm{Var}[\sigma_2 \rho\, Z_1 + \sigma_2 \sqrt{1-\rho^2}\, Z_2] = \mathrm{Var}[\sigma_2 \rho\, Z_1] + \mathrm{Var}[\sigma_2 \sqrt{1-\rho^2}\, Z_2] = \sigma_2^2 \rho^2 \mathrm{Var}[Z_1] + \sigma_2^2 (1-\rho^2) \mathrm{Var}[Z_2] = \sigma_2^2 \rho^2 + \sigma_2^2 (1-\rho^2) = \sigma_2^2$. Och att deras korrelation är ρ ser vi ur $\mathrm{Kov}[X_1, X_2] = E[(X_1 - \mu_1)(X_2 - \mu_2)] = E[\sigma_1 Z_1 (\sigma_2 \rho\, Z_1 + \sigma_2 \sqrt{1-\rho^2}\, Z_2)] = \sigma_1 \sigma_2 \rho E[Z_1^2] + \sigma_1 \sigma_2 \sqrt{1-\rho^2} E[Z_1 Z_2] = \sigma_1 \sigma_2 \rho$ ty $E[Z_1 Z_2] = E[Z_1] E[Z_2] = 0$.

11) Enligt ledningen är paret Z_1, Z_2 N-fördelat. Att visa att $E[Z_1] = E[Z_2] = 0$, $\mathrm{Var}[Z_1] = \mathrm{Var}[Z_2] = 1$ och $\mathrm{Kov}[Z_1, Z_2] = 0$ (vilket medför $\rho = 0$) överlåtes åt läsaren.