STK1100 våren 2023

Transformasjoner og simulering av kontinuerlige stokastiske variabler

Svarer til avsnitt 4.7 og 4.8 i læreboka

Matematisk institutt
Universitetet i Oslo

Problemstilling

La X være en kontinuerlig stokastisk variabel med sannsynlighetstetthet $f_X(x)$ og kumulativ fordeling $F_X(x)$

La $Y = g(X)$, der g er en kjent transformasjon

Hva er den kumulative fordelingen $F_Y(y)$ og sannsynlighetstettheten $f_Y(y)$ til Y?

Eksempel: eksponentialfordelingen

X er eksponentialfordelt med parameterer $\lambda > 0$

Sannsynlighetstetthet:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

Kumulativ fordeling:

$$F_X(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 - e^{-\lambda x} & \text{for } x > 0 \end{cases}$$

Sett $Y = cX$ for $c > 0$

Kumulativ fordeling for $Y = cX$ (for $y > 0$):

$$F_Y(y) = P(Y \leq y) \quad = P(cX \leq y)$$

$$= P(X \leq y / c) \quad = F_X(y / c)$$

$$= 1 - e^{-\lambda(y/c)} \quad = 1 - e^{-(\lambda/c)y}$$

Sannsynlighetstetthet for Y (for $y > 0$):

$$f_Y(y) = F_Y'(y) \quad = (\lambda / c)e^{-(\lambda/c)y}$$

Y er eksponentialfordelt med parameterer λ / c

Eksempel: normalfordelingen

X har sannsynlighetstetthet:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{for} \quad -\infty < x < \infty$$

$Y = aX + b \quad (a \neq 0).$ Hva er fordelingen til Y?

Anta først at $a > 0$. Da har Y kumulativ fordeling:

$$F_Y(y) = P(Y \le y) = P(aX + b \le y)$$

$$= P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

Sannsynlighetstettheten for Y:

$$f_Y(y) = F_Y'(y) = F_X'\left(\frac{y-b}{a}\right)\frac{1}{a} = f_X\left(\frac{y-b}{a}\right)\frac{1}{a}$$

Anta så at $a < 0$. Da er

$$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P\left(X \ge \frac{y-b}{a} \right)$$

$$= 1 - P\left(X \le \frac{y-b}{a} \right) = 1 - F_X\left(\frac{y-b}{a} \right)$$

Det gir

$$f_Y(y) = F_Y'(y) = -F_X'\left(\frac{y-b}{a} \right)\frac{1}{a} = f_X\left(\frac{y-b}{a} \right)\frac{1}{|a|}$$

Dermed for $a \neq 0$

$$f_Y(y) = f_X\left(\frac{y-b}{a} \right)\frac{1}{|a|} = \frac{1}{\sigma|a|\sqrt{2\pi}} e^{-\frac{1}{2a^2\sigma^2}(y-a\mu-b)^2}$$

Altså $Y \sim N(a\mu + b, a^2\sigma^2)$

Stokastisk simulering

Datamaskiner kan generere stokastiske variabler som er uniformt fordelt over (0,1)

Hvordan kan det brukes til å generere stokastiske variabler med en gitt fordeling?

Anta at X er uniformt fordelt på (0,1). Da er

$$f_X(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{ellers} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Hvordan skal vi velge g for at $Y = g(X)$ skal ha kumulativ fordeling $F(y)$, der $F(y)$ er en gitt strengt voksende kumulativ funksjon?

Vi skal velge g som den inverse funksjonen til F, dvs $Y = F^{-1}(X)$

For da har vi at:

$$
\begin{aligned}
P(Y \le y) &= P(F^{-1}(X) \le y) \\
&= P(X \le F(y)) \\
&= F(y)
\end{aligned}
$$

så Y har kumulativ fordeling $F(y)$

Eksempel: eksponentialfordelingen

Vi vil at Y skal ha kumulativ fordeling

$$F(y) = \begin{cases} 0 & \text{for } y \leq 0 \\ 1 - e^{-y} & \text{for } y > 0 \end{cases}$$

Den inverse funksjonen er (for x>0)

$$F^{-1}(x) = -\ln(1 - x)$$

Så hvis X er uniformt fordelt på (0,1), så er $Y = -\ln(1 - X)$ eksponentialfordelt med parameter 1

Transformasjon til uniform fordeling

Anta at X har kumulativ fordeling $F_X(x)$

Hva er da fordelingen til $Y = F_X(X)$?

Vi har at (for $0 \le y \le 1$) :

$$F_Y(y) = P(Y \le y) = P(F_X(X) \le y)$$

$$= P(X \le F_X^{-1}(y)) = F_X(F_X^{-1}(y))$$

$$= y$$

Så Y er uniformt fordelt på (0,1)

Et generelt resultat

Anta at X har sannsynlighetstetthet $f_X(x)$ og la $Y = g(X)$, der $g(x)$ er en strengt monoton deriverbar funksjon med invers funksjon $h(y)$ med derivert $h'(y)$

Da har Y sannsynlighetstetthet

$$f_Y(y) = f_X(h(y)) \, | \, h'(y) \, |$$

Læreboka bruker dette resultatet for å finne sannsynlighetstettheten til Y i flere eksempler

Men det er ofte enklere å gå via den kumulative fordelingen i hvert enkelt eksempel (slik vi har gjort ovenfor)

Eksempel: ikke-monoton transformasjon

Z er standardnormalfordelt:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{for} \quad -\infty < z < \infty$$

Sett $Y = Z^2$.

Da har Y kumulativ fordeling (y>0):

$$F_Y(y) = P(Y \le y)$$

$$= P(Z^2 \le y)$$

$$= P\left(-\sqrt{y} \le Z \le \sqrt{y}\right)$$

$$= F_Z\left(\sqrt{y}\right) - F_Z\left(-\sqrt{y}\right)$$

Sannsynlighetstettheten for Y (for y>0):

$$f_Y(y) = F'_Y(y) = \frac{d}{dy}\left\{F_Z\left(\sqrt{y}\right) - F_Z\left(-\sqrt{y}\right)\right\}$$

$$= \frac{1}{2\sqrt{y}} F'_Z\left(\sqrt{y}\right) + \frac{1}{2\sqrt{y}} F'_Z\left(-\sqrt{y}\right)$$

$$= \frac{1}{2\sqrt{y}} f_Z\left(\sqrt{y}\right) + \frac{1}{2\sqrt{y}} f_Z\left(-\sqrt{y}\right)$$

$$= \frac{1}{\sqrt{2\pi y}} e^{-y/2} = \frac{1}{2^{1/2}\Gamma(1/2)} y^{1/2-1} e^{-y/2}$$

Y er gamma-fordelt med $\alpha = 1/2$ og $\beta = 2$ (som også kalles kji-kvadratfordelingen med 1 frihetsgrad)

Kji-kvadratfordelingen er et spesialtilfelle av gammafordelingen, der vi har $\alpha = \nu / 2$ og $\beta = 2$

En stokastisk variabel X er kjikvadratfordelt med ν frihetsgrader hvis den har sannsynlighetstetthet

$$f(x;\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

Vi har at

$$E(X) = \alpha\beta = \nu \qquad V(X) = \alpha\beta^2 = 2\nu$$

Kji-kvadratfordelingen brukes til å beskrive funksjoner av stokastiske variabler

F.eks. hvis $Z \sim N(0,1)$ så er Z^2 kji-kvadratfordelt med ν=1