

Funksjoner av stokastiske variable.

Dekkes av pensumssidene i kap. 7

I ulike sammenhenger, blant annet for å finne fordelingen til estimatorer, er vi interesserte i fordelingen til funksjoner av stokastiske variable, f.eks. fordelingen til $Y = u(X)$ når vi kjenner fordelingen til X , eller fordelingen til $W = g(X_1, \dots, X_n)$ når vi kjenner fordelingen til $X_1, \dots, X_n$.

Transformasjoner (7.2)

Vi kjenner fordelingen til X . Ønsker å finne fordelingen til $Y = u(X)$. (Boka omhandler også situasjonen når vi har to funksjoner av to variable - vi skal hovedsaklig se på situasjonen med en variabel.)

Eksempel: Vi kjenner fordelingen til hastigheten til molekylerne i en gass, X , hva er fordelingen til energien $Y = \frac{1}{2}mX^2$? □

Diskrete variable: Les selv i boka.

Kontinuerlige variable:

La X være en kontinuerlig stok. var. med sanns. tetthet $f(x)$. Anta videre at $Y = u(X)$ definerer en *en-til-en* transformasjon fra X til Y . Vi ønsker å finne fordelingen til Y . Siden $u(X)$ er en-til-en har vi også en entydig transformasjon fra Y til X , $X = w(Y) = u^{-1}(Y)$. Sannsynlighetstettheten til Y er da gitt ved **transformasjonsformelen**:

$$g(y) = f(w(y))|w'(y)|$$

(Se side 34 i formelsamlingen/teorem 7.3 i boka)

Bevis:

1. $u(X)$ voksende:

$$F_Y(y) = P(Y \leq y) = P(X \leq w(y)) = F_X(w(y))$$

$$g(y) = F'_Y(y) = \frac{d}{dy} F_X(w(y)) = f(w(y))w'(y)$$

2. $u(X)$ avtakende:

$$\begin{aligned}F_Y(y) &= P(Y \leq y) = P(X \geq w(y)) \\&= 1 - P(X \leq w(y)) = 1 - F_X(w(y)) \\g(y) &= \frac{d}{dy}(1 - F_X(w(y))) = -f(w(y))w'(y) \\&= f(w(y))|w'(y)| \text{ siden } w'(y) \text{ negativ.}\end{aligned}$$

Alternativt bevis: Se boka. $\square$

Eksempel: Fordelingen til hastigheten, X , til molekylene i en gass er gitt ved:

$$f(x) = 4x^3 e^{-x^4}, \quad x > 0$$

Hva blir fordelingen til kinetisk energi til molekylene, dvs fordelingen til $Y = \frac{1}{2}mX^2$?

Siden $X > 0$ er transformasjonen en-til-en.

$X = \sqrt{2Y/m} = w(Y)$ og vi får:

$$\begin{aligned}g(y) &= 4\left(\frac{2y}{m}\right)^{(3/2)} e^{-(2y/m)} \frac{1}{2} \left(\frac{2y}{m}\right)^{-1/2} \frac{2}{m} \\&= \frac{4}{m} \frac{2y}{m} e^{-(2y/m)} \\&= \frac{8y}{m^2} e^{-4y^2/m^2}, \quad y > 0 \quad \square\end{aligned}$$

Eksempel: $X \sim N(\mu, \sigma^2)$. Hva er fordelingen til $Z = \frac{X-\mu}{\sigma}$?

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

$X = \sigma Z + \mu = w(Z)$ er en-til-en og vi får

$$\begin{aligned} g(z) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(\sigma z + \mu - \mu)^2}{\sigma^2}} \sigma \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}, \quad -\infty < z < \infty \end{aligned}$$

Dvs $Z \sim N(0, 1)$. □

Merk: Det er ikke alltid slik som i eksemplet over at en lineærkombinasjon har samme fordeling (bare med andre parametre) som den opprinnelige variabelen.

Hva gjør vi dersom transformasjonen ikke er en-til-en?

Eksempel: $X \sim N(\mu, \sigma^2)$. Hva er fordelingen til $Y = (\frac{X-\mu}{\sigma})^2 = Z^2$? Ikke en-til-en for alle mulige Z , men kan deles opp i to en-til-en funksjoner:

- for $z < 0$: $z = -\sqrt{y} = w_1(y)$
- for $z \geq 0$: $z = \sqrt{y} = w_2(y)$

Kan vises (se boka) at da er:

$$g(y) = f(w_1(y))|w'_1(y)| + f(w_2(y))|w'_2(y)|$$

I vårt eksempel gir dette:

$$\begin{aligned} g(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-\sqrt{y})^2} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \frac{1}{2\sqrt{y}} \\ &= \frac{1}{2^{1/2}\sqrt{\pi}} e^{-\frac{y}{2}} \frac{1}{\sqrt{y}} \\ &= \frac{1}{2^{1/2}\Gamma(1/2)} y^{-1/2} e^{-y/2}, y \geq 0 \end{aligned}$$

Dette er en χ_1^2 -fordeling! □

Summer av stok. var.

Eksempel: Du kjøper to lyspærer til lampen din. La X_1 og X_2 være funksjonstiden til de to pærene. Når den første har feilet setter du i den neste slik at den samlede funksjonstiden blir $Y = X_1 + X_2$. Hva blir fordelingen til Y ? Anta at X_1 og X_2 er uavh. og eksponensialfordelte med parameter β .

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X_1 + X_2 \leq y) \\ &= \int \int_{x_1 + x_2 \leq y} \frac{1}{\beta} e^{-x_1/\beta} \frac{1}{\beta} e^{-x_2/\beta} dx_2 dx_1 \\ &= \int_0^y \int_0^{y-x_1} \frac{1}{\beta} e^{-x_1/\beta} \frac{1}{\beta} e^{-x_2/\beta} dx_2 dx_1 \\ &= \int_0^y \frac{1}{\beta} e^{-x_1/\beta} \int_0^{y-x_1} \frac{1}{\beta} e^{-x_2/\beta} dx_2 dx_1 \\ &= \int_0^y \left(\frac{1}{\beta} e^{-x_1/\beta} - \frac{1}{\beta} e^{-y/\beta} \right) dx_1 \\ &= 1 - e^{-y/\beta} - \frac{y}{\beta} e^{-y/\beta} \end{aligned}$$

$$\begin{aligned}
f_Y(y) &= F'_Y(y) \\
&= \frac{1}{\beta}e^{-y/\beta} - \frac{1}{\beta}e^{-y/\beta} + \frac{y}{\beta^2}e^{-y/\beta} \\
&= \frac{1}{\beta^2}ye^{-y/\beta}
\end{aligned}$$

Dette er en gamma-fordeling med parametre $\alpha = 2$ og β . Videre kan det på samme måte vises at dersom $X_1, \dots, X_n$ er uavh. eksponensialfordelte med parameter β så er $Y = \sum_{i=1}^n X_i$ gamma-fordelt med parametre $\alpha = n$ og β . □

La oss se på denne typen problemer mer generelt. La X_1 og X_2 være to uavhengige stok. var. med sanns. tettheter $f_{X_1}(x_1)$ og $f_{X_2}(x_2)$. Vi ønsker å finne fordelingen til $Y = X_1 + X_2$.

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(X_1 + X_2 \leq y) \\
&= \int \int_{x_1 + x_2 \leq y} f_{X_1}(x_1) f_{X_2}(x_2) dx_2 dx_1 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{y-x_1} f_{X_1}(x_1) f_{X_2}(x_2) dx_2 dx_1
\end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) \int_{-\infty}^{y-x_1} f_{X_2}(x_2) dx_2 dx_1 \\
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) F_{X_2}(y-x_1) dx_1 \\
f_Y(y) &= F'_Y(y) \\
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y-x_1) dx_1
\end{aligned}$$

Uttrykket i siste linje kalles konvolusjonsformelen. Tilsvarende kan vises for diskrete stok. var. (integral erstattes da av summer) slik at vi totalt får:

Konvolusjonsformelen: Fordelingen til $Y = X_1 + X_2$ der X_1 og X_2 er uavh. stok. var. finnes ved:

$$f_Y(y) = \begin{cases} \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(y-x) dx & , \text{kont.} \\ \sum_x f_{X_1}(x) f_{X_2}(y-x) & , \text{diskr.} \end{cases}$$

Konvolusjonsformelen kan generaliseres til summer av flere variable.

Eksempel: Mellom to byer går det to veier. Antall trafikkulykker per år på den ene veien, X_1 , er Poisson-fordelt med forventning μ_1 , mens antall ulykker på den andre veien, X_2 , er Poisson-fordelt med forventning μ_2 . Hva er fordelingen til totalt antall ulykker på de to veiene, $Y = X_1 + X_2$?

$$\begin{aligned}
 f_Y(y) &= \sum_x f_{X_1}(x) f_{X_2}(y-x) \\
 &= \sum_{x=0}^y \frac{\mu_1^x}{x!} e^{-\mu_1} \frac{\mu_2^{y-x}}{(y-x)!} e^{-\mu_2} \\
 &= \sum_{x=0}^y \frac{y!}{x!(y-x)!} \mu_1^x \mu_2^{y-x} e^{-(\mu_1+\mu_2)} \frac{1}{y!} \\
 &= \frac{1}{y!} e^{-(\mu_1+\mu_2)} \sum_{x=0}^y \binom{y}{x} \mu_1^x \mu_2^{y-x} \\
 &= \frac{1}{y!} e^{-(\mu_1+\mu_2)} (\mu_1 + \mu_2)^y \\
 &= \frac{(\mu_1 + \mu_2)^y}{y!} e^{-(\mu_1+\mu_2)}
 \end{aligned}$$

Y er altså Poisson-fordelt med forventningsverdi $\mu_1 + \mu_2$.

Noen fordelinger (f.eks. normal, binomisk, Poisson og χ^2) har den egenskapen, som vist for Poisson, at summer av uavhengige variable er i samme fordelingsklasse (dvs summer har samme fordeling bare med andre parameterverdier). Se side 34/35 i formelsamlingen for en oversikt.

Merk at det bare er noen fordelinger som har denne egenskapen - det er ikke en generell egenskap ved fordelinger. F.eks. så vi at summen av to identisk eksponensialfordelte variable ble gammafordelt.