Tõenäosusteooria ja statistika II

Loeng 5

Juhuslike suuruste transformatsioonid

ja nende jaotus

Märt Möls

mart.mols@ut.ee

Juhuslike suuruste summa. Konvolutsioon.

Olgu *X* ja *Y* diskreetsed juhuslikud suurused. Milline on nende summa *Z* = *X*+*Y* jaotus?

Summa üle kõigi Y_i võimalike väärtuste

$$P(Z = z) = \sum_{i=1}^{k} P(Y = y_i) P(X = z - y_i | Y = y_i)$$

Lemma

Olgu X ja Y pidevad sõltumatud juhuslikud suurused tihedusfunksioonidega f_X ja f_Y. Sel juhul juhusliku suuruse $Z = X + Y$ tihedusfunktsioon avaldub kujul

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)\, dy, \quad z \in \mathbb{R}.$$

$$F_Z(z) = P(X + Y \le z)$$

$$= \iint\limits_{X+Y \le z} f_{X,Y}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{z-y} f_X(x)\, dx\, dy = \int_{-\infty}^{\infty} f_Y(y) F_X(z-y)\, dy$$

$$f_Z(z) = \frac{\partial F_Z(z)}{\partial z}$$

$$F_Z(z) = \int_{-\infty}^{\infty} f_Y(y) F_X(z-y)\, dy$$

$$f_Z(z) = \frac{\partial}{\partial z} \int_{-\infty}^{\infty} f_Y(y) F_X(z-y)\, dy$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) \frac{\partial F_X(z-y)}{\partial z}\, dy$$

Kui tuletisega võib integraali märgi alla liikuda – näiteks kui tuletis eksisteerib ja on pidev või kui …

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)\, dy$$

Näide

$$X \sim U(0;1) \qquad Z = X + Y$$
$$Y \sim U(0;1)$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)\, dy$$
$$= \int_{-\infty}^{\infty} I(0 \le y \le 1) I(0 \le z - y \le 1)\, dy$$
$$= \int_{z-1}^{z} I(0 \le y \le 1)\, dy$$

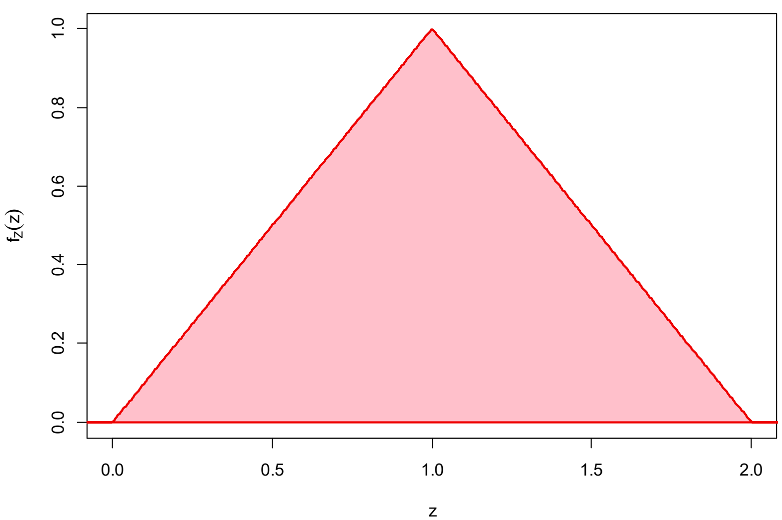
Näide

$$X \sim U(0;1) \qquad Z = X + Y$$
$$Y \sim U(0;1)$$

$$f_Z(z) = \int_{z-1}^{z} I(0 \le y \le 1)\, dy$$
$$= 0, \text{ kui } z < 0$$
$$= 0, \text{ kui } z > 2$$

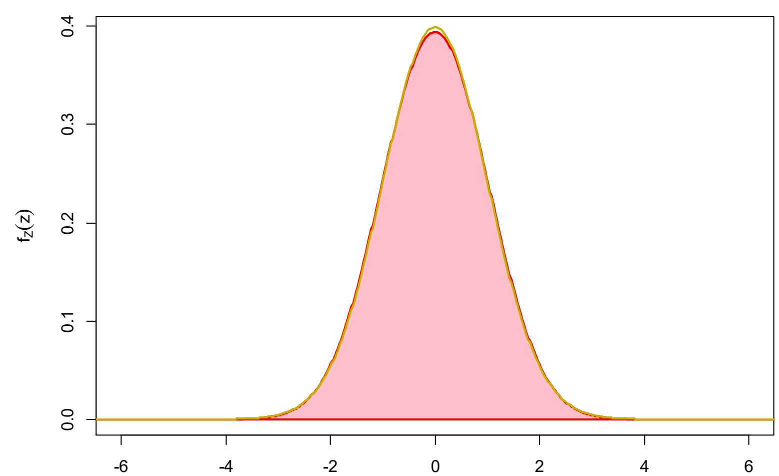
Kui $0 \le z \le 1$
$$= \int_{0}^{z} 1\, dy = z$$

Kui $1 < z \le 2$
$$= \int_{z-1}^{1} 1\, dy = 2 - z$$



$Z = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12} - 6$

Võrdlus standardse normaaljaotusega



Näide 2 – Tegelikud mõjud ja mõjude hinnangud

$X \sim F_X$

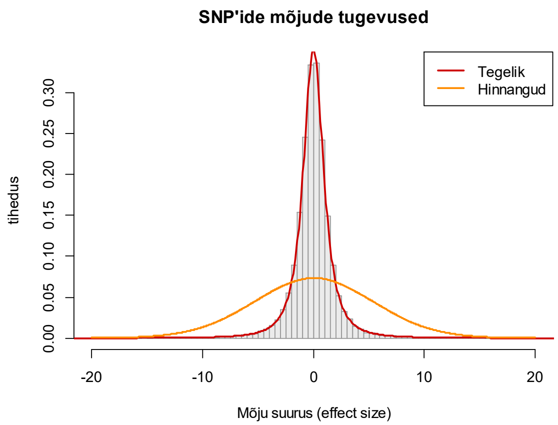
mutatsioonide mõjud

$\varepsilon \sim N(0; \sigma_\varepsilon^2)$

hindamisvead

$Y = X + \varepsilon$

uuringutes nähtud hinnangud



Lemma

Olgu X ja Y sõltumatud pidevad juhuslikud suurused tihedusfunktsioonidega f_X ja f_Y. Siis juhusliku suuruse $Z = \frac{X}{Y}$ tihedusfunktsioon avaldub kujul

$$f_Z(z) = \int_{-\infty}^{\infty} |y|\, f_X(zy) f_Y(y)\, dy.$$

$$\begin{aligned}
F_Z(z) &= P(X/Y \le z) \\
&= P(X/Y \le z \cap Y < 0) + P(X/Y \le z \cap Y > 0) \\
&= P(X \ge zY \cap Y < 0) + P(X \le zY \cap Y > 0) \\
&= \int_{-\infty}^{0} \int_{zy}^{\infty} f_X(x) f_y(y)\, dx\, dy + \int_{0}^{\infty} \int_{-\infty}^{zy} f_X(x) f_y(y)\, dx\, dy
\end{aligned}$$

$$\begin{aligned}
F_Z(z) &= \int_{-\infty}^{0} \int_{zy}^{\infty} f_X(x) f_y(y)\, dx\, dy + \int_{0}^{\infty} \int_{-\infty}^{zy} f_X(x) f_y(y)\, dx\, dy \\
&= \int_{-\infty}^{0} f_y(y) \int_{zy}^{\infty} f_X(x)\, dx\, dy + \int_{0}^{\infty} f_y(y) \int_{-\infty}^{zy} f_X(x)\, dx\, dy \\
&= \int_{-\infty}^{0} f_y(y)(1 - F_X(zy))\, dy + \int_{0}^{\infty} f_y(y) F_X(zy)\, dy
\end{aligned}$$

$$f_Z(z) = \int_{-\infty}^{0} f_y(y) \left[-f_X(zy) y\right]\, dy + \int_{0}^{\infty} f_y(y) f_X(zy) y\, dy$$

$$f_Z(z) = \int_{-\infty}^{0} f_y(y) f_X(zy) |y|\, dy + \int_{0}^{\infty} f_y(y) f_X(zy) |y|\, dy$$

Juhusliku suuruse transformatsiooni jaotus

$Y = g(X)$

Olgu $g(X)$ pidev ja pööratav (mingis vahemikus):

$\exists g^{-1}$, nii et $X = g^{-1}(Y)$

$\Rightarrow$ $g^{-1}(Y)$ ka pidev ja pööratav

pidev ja pööratav $\Rightarrow$ rangelt monotoonne (kas kasvav või kahanev)

kasvav: $y_1 < y_2 \;\Rightarrow\; g^{-1}(y_1) < g^{-1}(y_2)$

kahanev: $y_1 < y_2 \;\Rightarrow\; g^{-1}(y_1) > g^{-1}(y_2)$

$Y = g(X),\ X = g^{-1}(Y)$

Variant A:

$$
\begin{aligned}
&g^{-1} \text{ kasvav: } y_1 < y_2 \ \Rightarrow\ g^{-1}(y_1) < g^{-1}(y_2)\\
&F_Y(y) = P(Y \le y)\\
&\qquad = P(g(X) \le y)\\
&\qquad = P(g^{-1}(g(X)) \le g^{-1}(y))\\
&\qquad = P(X \le g^{-1}(y))\\
&\qquad = F_X(g^{-1}(y))\\
&f_Y(y) = \frac{\partial}{\partial y} F_X(g^{-1}(y))\\
&\qquad = f_X(g^{-1}(y)) \frac{\partial}{\partial y} g^{-1}(y)
\end{aligned}
$$

$Y = g(X),\ X = g^{-1}(Y)$

Variant B:

$$
\begin{aligned}
&g^{-1} \text{ kahanev: } y_1 < y_2 \ \Rightarrow\ g^{-1}(y_1) > g^{-1}(y_2)\\
&F_Y(y) = P(Y \le y)\\
&\qquad = P(g(X) \le y)\\
&\qquad = P(g^{-1}(g(X)) \ge g^{-1}(y))\\
&\qquad = 1 - P(X < g^{-1}(y))\\
&\qquad = 1 - F_X(g^{-1}(y))\\
&f_Y(y) = \frac{\partial}{\partial y} \left[1 - F_X(g^{-1}(y))\right]\\
&\qquad = -f_X(g^{-1}(y)) \frac{\partial}{\partial y} g^{-1}(y)
\end{aligned}
$$

pideva juhusliku suuruse korral P(X≤x) = P(X<x)

$Y = g(X),\ X = g^{-1}(Y)$

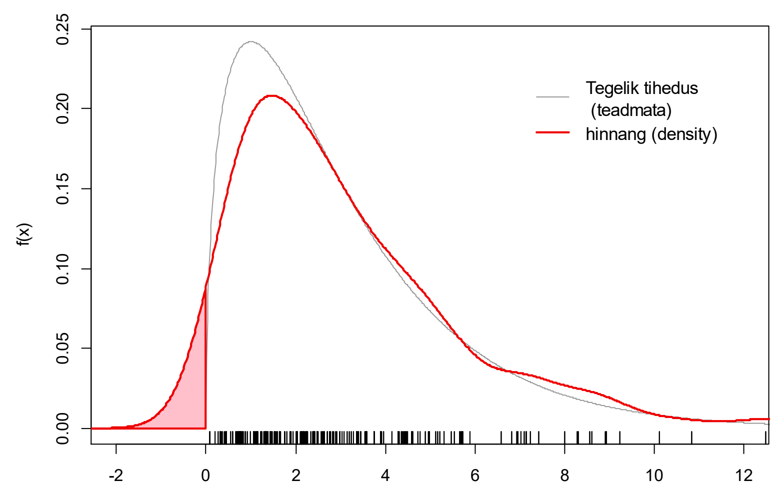
$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(y)}{\partial y} \right|$$

$$Y_1 = g_1(X_1, X_2), Y_2 = g_2(X_1, X_2)$$
$$X_1 = g_1^{-1}(Y_1, Y_2), X_2 = g_2^{-1}(Y_1, Y_2)$$

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2))|J|$$

$$J = \begin{vmatrix} \frac{\partial g_1^{-1}(y_1,y_2)}{\partial y_1} & \frac{\partial g_1^{-1}(y_1,y_2)}{\partial y_2} \\ \frac{\partial g_2^{-1}(y_1,y_2)}{\partial y_1} & \frac{\partial g_2^{-1}(y_1,y_2)}{\partial y_2} \end{vmatrix}$$

Kasutusnäide – tiheduse hindamine



Kasutusnäide – tiheduse hindamine

algne juhuslik suurus

$$Y \quad (0\ldots\infty) \xrightarrow[\displaystyle X=\ln(Y)]{\displaystyle g^{-1}(X)} X \quad (-\infty\ldots\infty)$$

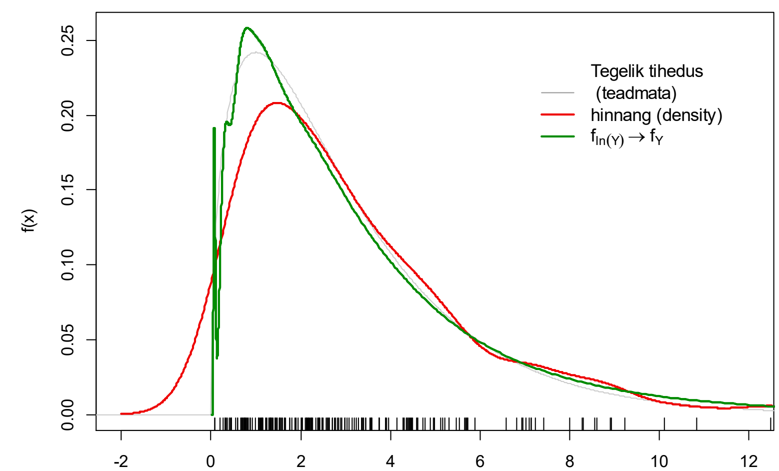
$$Y \quad (0\ldots\infty) \xleftarrow[\displaystyle g(X)]{\displaystyle Y=\exp(X)} X \quad (-\infty\ldots\infty)$$

hindame tiheduse $f_X(x)$

arvutame $f_Y(y)$

$$f_Y(y) = f_X(g^{-1}(y))\frac{\partial g^{-1}(y)}{\partial y}$$
$$= f_X(\ln(y))\cdot\frac{1}{y}$$

Kasutusnäide – tiheduse hindamine



Näide R'i programmina

```
# Andmed (antud näites genereeritud)
  set.seed(10); n=200;
  y=rchisq(n,df=3)

# Algne hinnang
  tihedus=density(y)

# Joonistame algse tihedusfunktsiooni hinnangu
# (hinnang ei arvesta, et juhusliku suuruse väärtused peavad olema mittenegatiivsed)
  plot(tihedus$x, tihedus$y, type="l")

# Hindame juhusliku suuruse X=log(Y) tiheduse
  tihedus_X=density(log(y))

# Arvutame log(Y) tihedusest Y tiheduse ja joonistame selle
  lines(exp(tihedus_X$x), tihedus_X$y/exp(tihedus_X$x), col=2)
```

$Y = g(X) = \exp(X)$ $\quad f_X(g^{-1}(y))$ $\quad \frac{1}{y}$

Transformatsioon $Y=X^2$

Üldjuhul pole teisendus $g(x)=x^2$ pööratav, sestap ei saa me kasutada f_Y leidmiseks äsja tõestatud abitulemust. Proovime siis leida lahendust veidi teisiti.

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(X^2 \le y) \\ &= P\left(-\sqrt{y} \le X \le \sqrt{y}\right) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \frac{\partial F_Y(y)}{\partial y} \\ &= \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right) \end{aligned}$$