Tõenäosusteooria ja statistika II

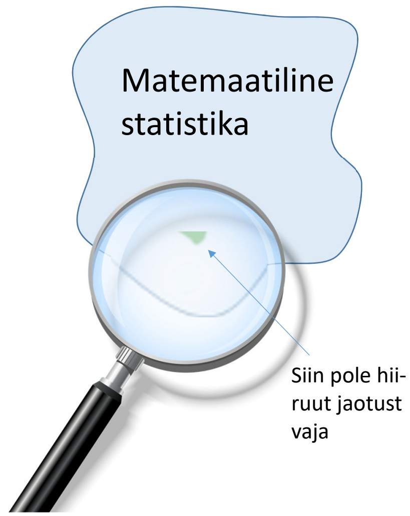
Loeng 7

Hii-ruut jaotus

Märt Möls

mart.mols@ut.ee

Hii-ruut jaotus



Teststatistiku jaotuseks hii-ruut jaotus:

* Pearsoni hii-ruut test
* Tõepärasuhte test
* log-rank test
* Cochran-Mantel-Haenzel test
* G-test
* Kruskal-Wallis test
* McNemar test
* Mudeli sobivuse testid
* usaldusintervall dispersioonile
* .....

Kasutatakse tuletuskäigus:

* T-test
* Usaldusintervall keskväärtusele
* F-test (ANOVA/dispersioonanalüüs)
* Testid regressioonmudeli parameetritele
* Dispersioonide võrdlemine
* .....

Hii-ruut jaotuse definitsioon

Juhuslik suurus *X* on hii-ruut jaotusega vabadusastmete arvuga $f > 0$ (vahel: $f \in \mathbb{N}$), $X \sim \chi^2(f)$, kui tema tihedusfunktsioon avaldub kujul

$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}},\ x \geq 0$$

kus $\Gamma(.)$ on gammafunktsioon, $\Gamma(z) = \int_0^\infty y^{z-1} e^{-y} dy.$

Kas definitsioon on korrektne?

$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}},\ x \geq 0$$

$$\Gamma(z) = \int_0^\infty y^{z-1} e^{-y} dy.$$

$$f(x) \geq 0 \quad \forall x \geq 0$$

$$\int_0^\infty f_X(x) dx = 1$$

$$\int_0^\infty x^{\frac{f}{2}-1} e^{-\frac{x}{2}} dx = \int_0^\infty 2^{\frac{f}{2}-1} (x/2)^{\frac{f}{2}-1} e^{-\frac{x}{2}} dx$$

$$= \int_0^\infty 2^{\frac{f}{2}-1} (x/2)^{\frac{f}{2}-1} e^{-\frac{x}{2}} 2\, d(x/2)$$

$$= 2^{\frac{f}{2}} \int_0^\infty (x/2)^{\frac{f}{2}-1} e^{-\frac{x}{2}} d(x/2) \ = 2^{f/2}\Gamma\left(\frac{f}{2}\right)$$

Hii-ruut jaotuse definitsioon

Juhuslik suurus *X* on hii-ruut jaotusega vabadusastmete arvuga $f > 0$ (vahel: $f \in \mathbb{N}$), $X \sim \chi^2(f)$, kui tema tihedusfunktsioon avaldub kujul

$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}},\ x \geq 0$$

kus $\Gamma(.)$ on gammafunktsioon, $\Gamma(z) = \int_0^\infty y^{z-1} e^{-y} dy.$

$$\Gamma(1) = \ldots$$

Hii-ruut jaotuse definitsioon

Juh
arv
tihe
kus

$$\Gamma(z+1) = \int_0^\infty y^z e^{-y} dy = \int_0^\infty y^z \left(-e^{-y}\right)' dy$$
$$= -\frac{y^z}{e^y}\bigg|_0^\infty - \int_0^\infty -e^{-y} (y^z)' dy$$
$$= 0 + z \int_0^\infty y^{z-1} e^{-y} dy = z \cdot \Gamma(z)$$

$$\Gamma(1) = 1$$
$$\Gamma(z+1) = z\Gamma(z), z > 0$$

Hii-ruut jaotuse definitsioon

Juhuslik suurus *X* on hii-ruut jaotusega vabadusastmete arvuga $f > 0$ (vahel: $f \in \mathbb{N}$), $X \sim \chi^2(f)$, kui tema tihedusfunktsioon avaldub kujul

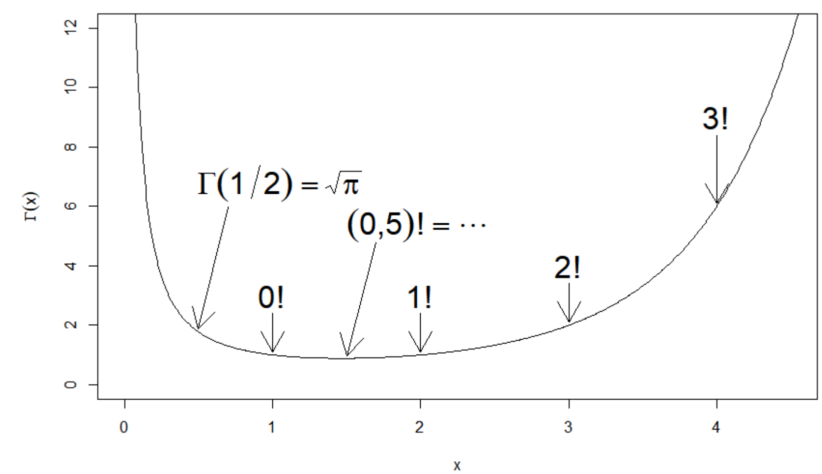
$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}},\ x \geq 0$$

kus $\Gamma(.)$ on gammafunktsioon, $\Gamma(z) = \int_0^\infty y^{z-1} e^{-y} dy.$

$$\Gamma(1) = 1$$
$$\Gamma(z+1) = z\Gamma(z), z > 0 \qquad \Gamma(2) = 1 \qquad \Gamma(3) = 2$$
$$\Gamma(4) = 6 \qquad \Gamma(5) = 24$$
$$\Gamma(6) = 120$$

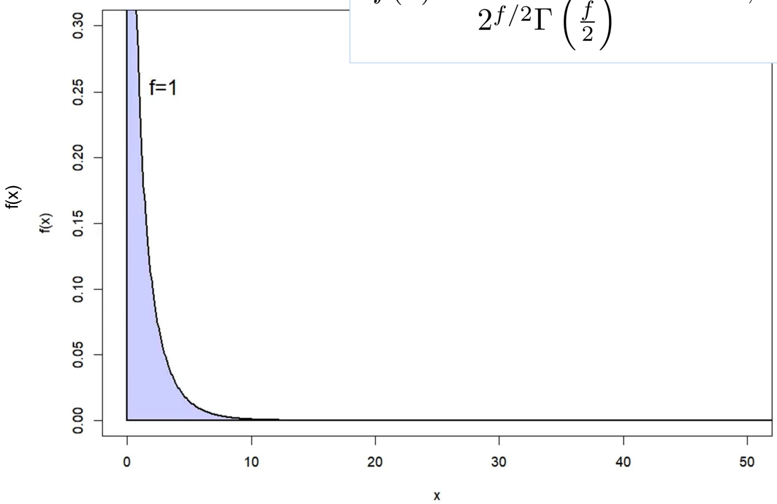
Gammafunktsioon

$$\Gamma(z) = \int_0^\infty y^{z-1} e^{-y} dy.$$



Hii-ruut jaotuste näiteid

$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}}, \ x \geq 0$$



Milline näeb välja kahe vabadusastmega (*f*=2) hii-ruut jaotuse tihedusfunktsioon?

$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}}, \ x \geq 0$$

f=2

$$f(x) = \frac{1}{2} e^{-\frac{x}{2}},$$

eksponentjaotus

$\chi^2(2)$ = Exp(1/2)

Hii-ruut jaotus momente genereeriv funktsioon

$M_X(t) = \mathrm{E}(e^{tX}), t \in \mathcal{R}$

$$\mathrm{E}(e^{tX}) = \int_{x=0}^{x=\infty} e^{tx} \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}} dx$$

$$= \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} \int_{x=0}^{x=\infty} x^{\frac{f}{2}-1} e^{-\frac{(1-2t)x}{2}} dx$$

$$y = \frac{(1-2t)x}{2} \quad x = \frac{2y}{(1-2t)}$$

$$= \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} \int_{y=0}^{y=\infty} \left(\frac{2y}{1-2t}\right)^{\frac{f}{2}-1} e^{-y} dx$$

kui t<1/2

Hii-ruut jaotus momente genereeriv funktsioon

$M_X(t) = \mathrm{E}(e^{tX}), t \in \mathcal{R}$

$$\mathrm{E}(e^{tX}) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} \int_{y=0}^{y=\infty} \left(\frac{2y}{1-2t}\right)^{\frac{f}{2}-1} e^{-y} \frac{2}{1-2t} dy$$

$$= \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} \left(\frac{2}{1-2t}\right)^{\frac{f}{2}} \int_0^{\infty} y^{\frac{f}{2}-1} e^{-y} dy$$

$$\Gamma(z) = \int_0^{\infty} y^{z-1} e^{-y} dy.$$

$$= \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} \left(\frac{2}{1-2t}\right)^{\frac{f}{2}} \Gamma(f/2)$$

Hii-ruut jaotus
momente genereeriv funktsioon

$M_X(t) = \mathrm{E}(e^{tX}), t \in \mathcal{R}$

$$\mathrm{E}(e^{tX}) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} \int_{y=0}^{y=\infty} \left(\frac{2y}{1-2t}\right)^{\frac{f}{2}-1} e^{-y} \frac{2}{1-2t} dy$$

$$= \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)} \left(\frac{2}{1-2t}\right)^{\frac{f}{2}} \int_0^\infty y^{\frac{f}{2}-1} e^{-y} dy$$

$$\Gamma(z) = \int_0^\infty y^{z-1} e^{-y} dy.$$

$$= \frac{1}{\cancel{2^{f/2}}\cancel{\Gamma\left(\frac{f}{2}\right)}} \left(\frac{\cancel{2}}{1-2t}\right)^{\frac{f}{2}} \cancel{\Gamma(f/2)} = \left(\frac{1}{1-2t}\right)^{\frac{f}{2}}$$

Hii-ruut jaotus
momente genereeriv funktsioon

$M_X(t) = \mathrm{E}(e^{tX}), t \in \mathcal{R}$

$m_k = M_X^{(k)}(0)$

$\mathrm{E}(X) = M_X'(0)$

$$M_X(t)' = f\,(1-2t)^{-\frac{f}{2}-1}$$

$\mathrm{E}(X) = f$

$$M_X(t)'' = f(-f/2-1)\,(1-2t)^{-\frac{f}{2}-2}\,(-2)$$
$$= f(f+2)\,(1-2t)^{-\frac{f}{2}-2}$$

$\mathrm{E}(X^2) = f(f+2)$

$$\mathrm{D}(X) = \mathrm{E}X^2 - (\mathrm{E}X)^2$$
$$= f(f+2) - f^2 = 2f$$

$$M_X(t) = (1-2t)^{-\frac{f}{2}}$$

Teoreem 1

Kui $X_1 \sim \chi^2(f_1)$ ja $X_2 \sim \chi^2(f_2)$ on sõltumatud juhuslikud suurused siis

$$X_1 + X_2 \sim \chi^2(f_1+f_2)$$

Tõestus

$$M_{X_1}(t) = (1-2t)^{-\frac{f_1}{2}} \qquad M_{X_2}(t) = (1-2t)^{-\frac{f_2}{2}}$$

$M_{X+Y}(t) = M_X(t)\, M_Y(t)$

$$M_{X_1+X_2}(t) = (1-2t)^{-\frac{f_1}{2}} \cdot (1-2t)^{-\frac{f_2}{2}}$$
$$= (1-2t)^{-\frac{f_1+f_2}{2}} \quad \sim \chi^2(f_1+f_2)$$

M. O. T. T.

Seos jaotusega N(0;1)

Teoreem 2

Kui X_1, X_2, …, X_n on sõltumatud standardse normaaljaotusega N(0;1) juhuslikud suurused, siis

$$\sum_{i=1}^{n} {X_i}^2 \sim \chi^2(n)$$

Tõestus

Alustame juhust $n = 1$. Kui $X_1 \sim N(0;1)$, kas siis

$$X_1{}^2 \sim \chi^2(1)?$$

Kui $Y = X^2$ siis on Y tihedusfunktsioon leitav valemiga (vaata 2. loeng):

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

Standardne normaaljaotus on aga 0-punkti suhtes sümmeetriline ja seega

$$f_X(\sqrt{y}) = f_X(-\sqrt{y}) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\left(\sqrt{y}\right)^2}{2}\right).$$

Asetades N(0;1) tiheduse $f_Y(y)$ arvutusvalemisse saame:

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}}\exp\left(-\frac{y}{2}\right).$$

Tõestus

Kas tulemuseks saadud tihedus

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}}\exp\left(-\frac{y}{2}\right).$$

on hii-ruut jaotuse tihedus vabadusastmete arvuga *f*=1?

Hii-ruut jaotuse (vabadusastmete arvuga *f*) tihedusfunktsioon on kujul:

$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)}x^{\frac{f}{2}-1}e^{-\frac{x}{2}},\ x \geq 0$$

$$f(x) = \frac{1}{2^{1/2}\Gamma\left(\frac{1}{2}\right)}x^{\frac{1}{2}-1}e^{-\frac{x}{2}},\ x \geq 0$$

Tõestus

Kas tulemuseks saadud tihedus

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}}\exp\left(-\frac{y}{2}\right).$$

on hii-ruut jaotuse tihedus vabadusastmete arvuga *f*=1?

Hii-ruut jaotuse (vabadusastmete arvuga *f*) tihedusfunktsioon on kujul:

$$f(x) = \frac{1}{2^{f/2}\Gamma\left(\frac{f}{2}\right)}x^{\frac{f}{2}-1}e^{-\frac{x}{2}},\ x \geq 0$$

$$f(y) = \frac{1}{\sqrt{2}\sqrt{\pi}}\frac{1}{\sqrt{y}}e^{-\frac{y}{2}},\ y \geq 0$$

Tegemist on tõesti hii-ruut jaotusega!

Tõestus

Kui $X_1 \sim N(0;1)$, siis

$$X_1{}^2 \sim \chi^2(1).$$

Kui aga $X_1{}^2 \sim \chi^2(1)$ ja $X_2{}^2 \sim \chi^2(1)$ on sõltumatud, siis (teoreem 1):

$$X_1{}^2 + X_2{}^2 \sim \chi^2(2).$$

Kui aga $X_1{}^2 + \ldots + X_{n-1}{}^2 \sim \chi^2(n-1)$ ja $X_n{}^2 \sim \chi^2(1)$ on sõltumatud, siis (teoreem 1):

$$X_1{}^2 + \ldots + X_n{}^2 \sim \chi^2(n).$$

q.e.d.

Teoreem 3

Kui X_1, X_2, …, X_n on sõltumatud standardse normaaljaotusega N(0;1) juhuslikud suurused, siis

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 \sim \chi^2(n-1)$$

Tõestus

Kasutame tulemust, et normaaljatusega juhuslike suuruste vektori $\mathbf{X}$~N($\boldsymbol{\mu}$; $\boldsymbol{\Sigma}$) korral

$$\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}; \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$$

Antud juhul $\mathbf{X}=(X_1, X_2, \dots, X_n)^T$ jaotuseks on

$$\mathbf{X} \sim N(\mathbf{0}; \mathbf{I})$$

Ortogonaalse maatriksi **O** korral ($\mathbf{OO}^T=\mathbf{I}$, $\mathbf{O}^T\mathbf{O}=\mathbf{I}$):

$$\mathbf{OX} \sim N(\mathbf{0}; \mathbf{I})$$

Valime maatriksi O selliselt, et tema esimene rida koosneks suurustest $1/\sqrt{n}$.

Siis vektori **Y**=**OX** esimene element on $\sqrt{n}\bar{X}$, vektori **Y** elementide ruutude summa on aga sama mis vektori **X** elementide ruutude summa:

$$\mathbf{Y}^T\mathbf{Y} = (\mathbf{OX})^T\mathbf{OX} = \mathbf{X}^T\mathbf{O}^T\mathbf{OX} = \mathbf{X}^T\,\mathbf{I}\,\mathbf{X} = \mathbf{X}^T\mathbf{X}$$

$$n\bar{X}^2 + Y_2^2 + \dots + Y_n^2 = \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=2}^{n} Y_i^2 = \underbrace{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}_{\sum_{i=1}^{n}(X_i^2 - \bar{X}^2) = \sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$$

$$\mathbf{Y} = \mathbf{OX} \sim N(\mathbf{0}; \mathbf{I})$$

n-1 sõltumatu N(0,1) juhusliku suuruse ruudu summa

Järeldus (Valimikeskmise ja dispersiooni sõltumatus)

Kui X_i, $i = 1, 2, \dots, n$ on sõltumatud normaaljaotusega $N(0,1)$ juhuslikud suurused, siis $\sum_{i=1}^n (X_i - \bar{X})^2$ ja $\bar{X}$ on sõltumatud juhuslikud suurused.

Tõestus. Teoreemist järeldus, et Y_1 on sõltumatu suurusest Y_i, $i = 2, 3, \dots, n$ ning seega $\sqrt{n}\bar{X}$ on sõltumatu suurusest

$$\sum_{i=2}^n Y_i^2 = \sum_{i=1}^n (X_i - \bar{X})^2.$$

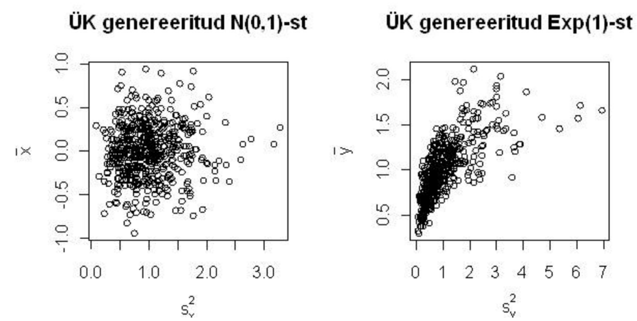
m.o.t.t.

NB. Järeldus ütleb ka, et valimikeskmine $\bar{X}$ ja valimidispersioon s^2 on sõltumatud normaaljaotusega ÜK korral. Teiste ÜK jaotuste korral see omadus ei pruugi kehtida.

Illustreerimi simuleerimisnäite põhjal, et s^2 ja $\bar{x}$ on tõepoolest sõltumatud jaotuse $N(0,1)$ korral. Kontrnäitena kasutame jaotust $Exp(1)$.

Algoritm:

1. Genereerida üks valim mahuga $n = 10$ jaotusest $N(0,1)$ ja teine valim jaotusest $Exp(1)$.
2. Arvutada valimite põhjal keskmised $\bar{x}$, $\bar{y}$ ja dispersioonid s_x^2, s_y^2.
3. Kanda punkt $(\bar{x}, s_x^2)$ ühele graafikule ja punkt $(\bar{y}, s_y^2)$ teisele.
4. Korrata sammud 1-3 $k = 500$ korda.



Järeldus (Hii-ruut jaotuse seos jaotusega $N(\mu, \sigma)$)

Sõltumatute $X_i \sim N(\mu, \sigma)$ korral kehtib:

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n),$$

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1).$$

Tõestus. Kuna $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (\frac{X_i - \mu}{\sigma})^2$ ja $\frac{X_i - \mu}{\sigma} \sim N(0,1)$, siis omaduse 3 põhjal on järelduse 1. seos tõestatud.
Edasi

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu - (\bar{X} - \mu))^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2,$$

kus $Z_i = (X_i - \mu)/\sigma \sim N(0,1)$ ja $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$. Rakendades om. 3 saamegi järelduse 2. väidet samutigi tõestatud.