Mat-1.2620 Sovellettu todennäköisyyslaskenta B

6. harjoitukset / Ratkaisut

Aiheet: **Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat**
Moniulotteisia jakaumia

Avainsanat:

Diskreetti jakauma, Ehdollinen jakauma, Ehdollinen odotusarvo, Ehdollinen varianssi, Jatkuva jakauma, Kaksiulotteinen normaalijakauma, Karteesinen tulo, Kertymäfunktio, Korrelaatio, Korreloimattomuus, Korreloituneisuus, Kovarianssi, Kulmakerroin, Multinomijakauma, Odotusarvo, Pistetodennäköisyysfunktio, Regressiofunktio, Regressiosuora, Reunajakauma, Riippumattomuus, Riippuvuus, Satunnaismuuttuja, Suora, Tiheysfunktio, Varianssi, Yhteisjakauma, Yhteiskorrelaatiokerroin

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

Kaksiulotteiset satunnaismuuttujat

Olkoot X ja Y satunnaismuuttujia, joiden otosavaruudet ovat R ja S. Tällöin

$$X : S \to \mathbb{R}$$
$$Y : R \to \mathbb{R}$$

Olkoon $R\times S$ otosavaruuksien R ja S *karteesinen tulo*:

$$R \times S = \{(r,s) \mid r \in R, s \in S\}$$

Satunnaismuuttujien X ja Y *järjestetty pari* (X, Y) määrittelee **kaksiulotteisen satunnaismuuttujan**:

$$(X,Y) : S \times R \to \mathbb{R}^2$$

Diskreetti kaksiulotteinen jakauma

Olkoot X ja Y **diskreettejä** satunnaismuuttujia. Tällöin järjestetty pari

(X, Y)

määrittelee **diskreetin kaksiulotteisen satunnaismuuttujan**. Diskreetti kaksiulotteinen satunnaismuuttuja (X, Y) määrittelee **diskreetin kaksiulotteisen todennäköisyysjakauman**, jota kutsutaan satunnaismuuttujien X ja Y **yhteisjakaumaksi**.

Reaaliarvoinen funktio

$$f_{XY} : \mathbb{R}^2 \to \mathbb{R}$$

määrittelee *diskreettien satunnaismuuttujien* X ja Y **yhteisjakauman pistetodennäköisyysfunktion**, jos seuraavat ehdot pätevät:

(1) $f_{XY}(x,y) \geq 0$ kaikille x ja y

(2) $\sum_x \sum_y f_{XY}(x,y) = 1$

(3) $\Pr(X = x \text{ ja } Y = y) = f_{XY}(x,y)$

Jatkuva kaksiulotteinen jakauma

Olkoot X ja Y **jatkuvia** satunnaismuuttujia. Tällöin järjestetty pari

$$(X, Y)$$

määrittelee **jatkuvan kaksiulotteisen satunnaismuuttujan**. Jatkuva kaksiulotteinen satunnaismuuttuja (X, Y) määrittelee **jatkuvan kaksiulotteisen todennäköisyysjakauman**, jota kutsutaan satunnaismuuttujien X ja Y **yhteisjakaumaksi**.

Reaaliarvoinen funktio

$$f_{XY} : \mathbb{R}^2 \to \mathbb{R}$$

määrittelee **jatkuvien satunnaismuuttujien** X ja Y **yhteisjakauman tiheysfunktion**, jos seuraavat ehdot pätevät:

(1) $f_{XY}(x, y) \geq 0$ kaikille x ja y

(2) $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x, y)dydx = 1$$

(3) $$\Pr(a \leq X \leq b \text{ ja } c \leq Y \leq d) = \int_a^b\int_c^d f_{XY}(x, y)dydx$$

Kaksiulotteisen jakauman kertymäfunktio

Olkoon (X, Y) satunnaismuuttujien X ja Y muodostama *järjestetty pari*. Satunnaismuuttujien X ja Y *yhteisjakauman kertymäfunktio* F_{XY} määritellään kaavalla

$$F_{XY}(x, y) = \Pr(X \leq x \text{ ja } Y \leq y)$$

Diskreetin kaksiulotteisen jakauman kertymäfunktio

Olkoon $f_{XY}(x, y)$ *diskreetin kaksiulotteisen jakauman pistetodennäköisyysfunktio*. *Jakauman* **kertymäfunktio** saadaan kaavalla

$$F_{XY}(x, y) = \Pr(X \leq x \text{ ja } Y \leq y) = \sum_{x_i \leq x}\sum_{y_i \leq y} f_{XY}(x_i, y_i)$$

Jatkuvan kaksiulotteisen jakauman kertymäfunktio

Olkoon $f_{XY}(x, y)$ *jatkuvan kaksiulotteisen jakauman tiheysfunktio*. *Jakauman* **kertymäfunktio** saadaan kaavalla

$$F_{XY}(x, y) = \Pr(X \leq x \text{ ja } Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u, v)dvdu$$

Olkoon $F_{XY}(x, y)$ *jatkuvan kaksiulotteisen jakauman kertymäfunktio*. Jos *derivaatta*

$$\frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} = f_{XY}(x, y)$$

on olemassa ja on jatkuva, funktio $f_{XY}(x, y)$ on satunnaismuuttujien X ja Y yhteisjakauman *tiheysfunktio*.

Diskreetin kaksiulotteisen jakauman reunajakaumat

Olkoon $f_{XY}(x, y)$ *diskreetin kaksiulotteisen jakauman pistetodennäköisyysfunktio*.

Satunnaismuuttujan X **reunajakauman** *pistetodennäköisyysfunktio* on

$$f_X(x) = \Pr(X = x) = \sum_y f_{XY}(x, y)$$

Satunnaismuuttujan Y **reunajakauman** *pistetodennäköisyysfunktio* on

$$f_Y(y) = \Pr(Y = y) = \sum_x f_{XY}(x, y)$$

Satunnaismuuttujien X ja Y reunajakaumat yhtyvät satunnaismuuttujien X ja Y todennäköisyys-jakaumiin.

Jatkuvan kaksiulotteisen jakauman reunajakaumat

Olkoon $f_{XY}(x, y)$ *jatkuvan kaksiulotteisen jakauman tiheysfunktio*.

Satunnaismuuttujan X **reunajakauman** *tiheysfunktio* on

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy$$

Satunnaismuuttujan Y **reunajakauman** *tiheysfunktio* on

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dx$$

Satunnaismuuttujien X ja Y reunajakaumat yhtyvät satunnaismuuttujien X ja Y todennäköisyys-jakaumiin.

Satunnaismuuttujien riippumattomuus

Olkoon $f_{XY}(x, y)$ satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyys*- tai *tiheysfunktio*, $f_X(x)$ satunnaismuuttujan X *reunajakauman pistetodennäköisyys*- tai *tiheysfunktio* ja $f_Y(y)$ satunnais-muuttujan Y *reunajakauman pistetodennäköisyys*- tai *tiheysfunktio*.

Satunnaismuuttujat X ja Y ovat **riippumattomia**, jos ja vain, jos

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

Diskreetin kaksiulotteisen jakauman yleinen odotusarvo

Olkoon $f_{XY}(x, y)$ diskreettien satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* ja olkoon

$$g : \mathbb{R}^2 \to \mathbb{R}$$

jatkuva funktio. Tällöin satunnaismuuttujan $g(X, Y)$ **odotusarvo** on *vakio*

$$\mathrm{E}(g(X,Y)) = \sum_x \sum_y g(x,y) f_{XY}(x, y)$$

Jatkuvan kaksiulotteisen jakauman yleinen odotusarvo

Olkoon $f_{XY}(x, y)$ jatkuvien satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* ja olkoon

$$g : \mathbb{R}^2 \to \mathbb{R}$$

jatkuva funktio. Tällöin satunnaismuuttujan $g(X, Y)$ **odotusarvo** on *vakio*

$$\mathrm{E}(g(X,Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f_{XY}(x,y)dydx$$

Diskreetin kaksiulotteisen jakauman odotusarvot

Olkoon satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* $f_{XY}(x, y)$, satunnaismuuttujan X *reunajakauman pistetodennäköisyysfunktio* $f_X(x)$ ja satunnaismuuttujan Y *reunajakauman pistetodennäköisyysfunktio* $f_Y(y)$.

Satunnaismuuttujan X *odotusarvo* $\mathrm{E}(X) = \mu_X$ yhtyy satunnaismuuttujan X *reunajakauman odotusarvoon*:

$$\mathrm{E}(X) = \sum_x\sum_y xf_{XY}(x,y) = \sum_x x\sum_y f_{XY}(x,y) = \sum_x xf_X(x)$$

Satunnaismuuttujan Y *odotusarvo* $\mathrm{E}(Y) = \mu_Y$ yhtyy satunnaismuuttujan Y *reunajakauman odotusarvoon*:

$$\mathrm{E}(Y) = \sum_x\sum_y yf_{XY}(x,y) = \sum_y y\sum_x f_{XY}(x,y) = \sum_y yf_Y(y)$$

Jatkuvan kaksiulotteisen jakauman odotusarvot

Olkoon satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* $f_{XY}(x, y)$, satunnaismuuttujan X *reunajakauman tiheysfunktio* $f_X(x)$ ja satunnaismuuttujan Y *reunajakauman tiheysfunktio* $f_Y(y)$.

Satunnaismuuttujan X *odotusarvo* $\mathrm{E}(X) = \mu_X$ yhtyy satunnaismuuttujan X *reunajakauman odotusarvoon*:

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf_{XY}(x,y)dydx = \int_{-\infty}^{+\infty} x\int_{-\infty}^{+\infty} f_{XY}(x,y)dydx = \int_{-\infty}^{+\infty} xf_X(x)dx$$

Satunnaismuuttujan Y *odotusarvo* $\mathrm{E}(Y) = \mu_Y$ yhtyy satunnaismuuttujan Y *reunajakauman odotusarvoon*:

$$\mathrm{E}(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf_{XY}(x,y)dydx = \int_{-\infty}^{+\infty} y\int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = \int_{-\infty}^{+\infty} yf_Y(y)dy$$

Odotusarvon ominaisuudet

Satunnaismuuttujien X ja Y odotusarvojen muodostama järjestetty pari

$$(\mu_X, \mu_Y)$$

määrää satunnaismuuttujien X ja Y yhteisjakauman todennäköisyysmassan *painopisteen*.

Satunnaismuuttujien X ja Y **summan** $X + Y$ **odotusarvo**:

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$

Satunnaismuuttujien X ja Y **erotuksen** $X - Y$ **odotusarvo**:

$$\mathrm{E}(X-Y) = \mathrm{E}(X) - \mathrm{E}(Y)$$

Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, niin *tulon XY odotusarvo on odotusarvojen tulo*:

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y) = \mu_X \mu_Y$$

Huomautus:

Käänteinen ei päde: Siitä, että

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y) = \mu_X \mu_Y$$

ei seuraa, että satunnaismuuttujat X ja Y ovat riippumattomia.

Kaksiulotteisen jakauman varianssit ja standardipoikkeamat

Olkoot satunnaismuuttujien X ja Y *odotusarvot*

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{E}(Y) = \mu_Y$$

Satunnaismuuttujien X ja Y **varianssit** yhtyvät vastaavien *reunajakaumien variansseihin*:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \mathrm{E}[(X-\mu_X)^2]$$
$$\mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2 = \mathrm{E}[(Y-\mu_Y)^2]$$

Satunnaismuuttujien X ja Y varianssien kaavat voidaan kirjoittaa seuraaviin yhtäpitäviin muotoihin:

$$\mathrm{D}^2(X) = \mathrm{E}[(X-\mu_X)^2] = \mathrm{E}(X^2) - \mu_X^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$
$$\mathrm{D}^2(Y) = \mathrm{E}[(Y-\mu_Y)^2] = \mathrm{E}(Y^2) - \mu_Y^2 = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2$$

Satunnaismuuttujien X ja Y **standardipoikkeamat** yhtyvät vastaavien *reunajakaumien standardipoikkeamiin*:

$$\mathrm{D}(X) = \sigma_X = \sqrt{\mathrm{E}[(X-\mu_X)^2]}$$
$$\mathrm{D}(Y) = \sigma_Y = \sqrt{\mathrm{E}[(Y-\mu_Y)^2]}$$

Diskreetin kaksiulotteisen jakauman varianssit

Olkoon satunnaismuuttujan X *reunajakauman pistetodennäköisyysfunktio* $f_X(x)$ ja satunnaismuuttujan Y *reunajakauman pistetodennäköisyysfunktio* $f_Y(y)$.

Tällöin satunnaismuuttujien X ja Y *varianssit* ovat *vakioita*

$$\mathrm{D}^2(X) = \sum_x (x-\mu_X)^2 f_X(x)$$
$$\mathrm{D}^2(Y) = \sum_y (y-\mu_Y)^2 f_Y(y)$$

Jatkuvan kaksiulotteisen jakauman varianssit

Olkoon satunnaismuuttujan X *reunajakauman tiheysfunktio* $f_X(x)$ ja satunnaismuuttujan Y *reunajakauman tiheysfunktio* $f_Y(y)$.

$$\mathrm{D}^2(X) = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f_X(x)dx$$
$$\mathrm{D}^2(Y) = \int_{-\infty}^{+\infty} (y-\mu_Y)^2 f_Y(y)dy$$

Kovarianssi

Olkoot satunnaismuuttujien X ja Y *odotusarvot*

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{E}(Y) = \mu_Y$$

Satunnaismuuttujien X ja Y **kovarianssi** on *vakio*

$$\mathrm{Cov}(X,Y) = \sigma_{XY} = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)]$$

Erityisesti

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X) = \sigma_X^2$$
$$\mathrm{Cov}(Y,Y) = \mathrm{Var}(Y) = \sigma_Y^2$$

Satunnaismuuttujien X ja Y *kovarianssin* kaava voidaan kirjoittaa seuraaviin yhtäpitäviin muotoihin:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \mathrm{E}(XY) - \mu_X\mu_Y \\ &= \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)\end{aligned}$$

Jos satunnaismuuttujat X ja Y ovat *diskreettejä*, satunnaismuuttujien X ja Y *kovarianssi* on *vakio*

$$\mathrm{Cov}(X,Y) = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)$$

Jos satunnaismuuttujat X ja Y ovat *jatkuvia*, satunnaismuuttujien X ja Y *kovarianssi* on *vakio*

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)dxdy$$

Kovarianssin ominaisuudet

Jos

$$\mathrm{Cov}(X,Y) = \sigma_{XY} = 0$$

niin sanomme, että satunnaismuuttujat X ja Y ovat *korreloimattomia*.

Jos satunnaismuuttujat X ja Y ovat riippumattomia, niin ne ovat korreloimattomia, *mutta käänteinen ei päde*: Siitä, että satunnaismuuttujat X ja Y ovat korreloimattomia *ei seuraa*, että satunnaismuuttujat X ja Y ovat riippumattomia.

Olkoot satunnaismuuttujien X ja Y *varianssit*

$$\mathrm{Var}(X) = \sigma_X^2$$
$$\mathrm{Var}(Y) = \sigma_Y^2$$

ja kovarianssi

$$\mathrm{Cov}(X,Y) = \sigma_{XY}$$

Tällöin

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X,Y) = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY}$$

Jos

$$\mathrm{Cov}(X,Y) = \sigma_{XY} = 0$$

niin

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \sigma_X^2 + \sigma_Y^2$$

Korrelaatiokerroin

Olkoon satunnaismuuttujilla X ja Y on seuraavat *odotusarvot*, *varianssit* ja *kovarianssi*:

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2$$
$$\mathrm{E}(Y) = \mu_Y \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2$$
$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY}$$

Satunnaismuuttujien X ja Y **korrelaatiokerroin** on *vakio*

$$\begin{aligned}\mathrm{Cor}(X,Y) = \rho_{XY} &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \\ &= \frac{\mathrm{Cov}(X,Y)}{\mathrm{D}(X)\,\mathrm{D}(Y)} \\ &= \frac{\sigma_{XY}}{\sigma_X \sigma_Y}\end{aligned}$$

Korrelaatiokertoimen ominaisuudet

Huomaa, että

$$\mathrm{Cor}(X,Y) = \rho_{XY} = 0$$

täsmälleen silloin, kun

$$\mathrm{Cov}(X,Y) = \sigma_{XY} = 0$$

Jos siis

$$\mathrm{Cor}(X,Y) = \rho_{XY} = 0$$

niin satunnaismuuttujat X ja Y ovat *korreloimattomia*.

Jos satunnaismuuttujat X ja Y ovat riippumattomia, niin ne ovat korreloimattomia, *mutta käänteinen ei päde*: Siitä, että satunnaismuuttujat X ja Y ovat korreloimattomia *ei seuraa*, että satunnaismuuttujat X ja Y ovat riippumattomia.

Olkoon satunnaismuuttujien X ja Y *korrelaatiokerroin* Cor(X, Y). Tällöin

(i) $-1 \le \mathrm{Cor}(X,Y) \le +1$

(ii) Jos X ja Y ovat *riippumattomia*, niin $\mathrm{Cor}(X,Y) = 0$

(iii) $\mathrm{Cor}(X,Y) = \pm 1$, jos ja vain, jos

$$Y = \alpha + \beta X,$$

jossa α ja β ovat reaalisia *vakiota*, $\beta \neq 0$

Lineaarimuunnokset ja 2-ulotteisen jakauman tunnusluvut

Olkoot satunnaismuuttujilla X ja Y seuraavat odotusarvot, varianssit ja kovarianssi:

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{Var}(X) = \sigma_X^2$$
$$\mathrm{E}(Y) = \mu_Y \qquad \mathrm{Var}(Y) = \sigma_Y^2$$
$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY}$$

Olkoot

$$W = a + bX$$
$$Z = c + dY$$

jossa a, b, c, $d \in \mathbb{R}$ ovat reaalisia vakioita. Tällöin

$$\mathrm{E}(W) = a + b\,\mathrm{E}(X) = a + b\mu_X$$
$$\mathrm{E}(Z) = c + d\,\mathrm{E}(Y) = c + d\mu_Y$$
$$\mathrm{Var}(W) = b^2\,\mathrm{Var}(X) = b^2\sigma_X^2$$
$$\mathrm{Var}(Z) = d^2\,\mathrm{Var}(Y) = d^2\sigma_Y^2$$
$$\mathrm{Cov}(W,Z) = bd\,\mathrm{Cov}(X,Y) = bd\sigma_{XY}$$
$$\mathrm{Cor}(W,Z) = \mathrm{sgn}(bd)\,\mathrm{Cor}(X,Y) = \mathrm{sgn}(bd)\rho_{XY}$$

Ehdolliset jakaumat

Olkoon satunnaismuuttujien X ja Y yhteisjakauman *pistetodennäköisyys*- tai *tiheysfunktio* $f_{XY}(x, y)$ ja satunnaismuuttujien X ja Y *reunajakaumien* pistetodennäköisyys- tai tiheysfunktiot $f_X(x)$ ja $f_Y(x)$.

Satunnaismuuttujan X **ehdollinen jakauma** satunnaismuuttujan Y suhteen (ehdolla $Y = y$) on

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \text{ , jos } f_Y(y) > 0$$

Satunnaismuuttujan Y **ehdollinen jakauma** satunnaismuuttujan X suhteen (ehdolla $X = x$) on

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \text{ , jos } f_X(x) > 0$$

Ehdolliset jakaumat ja riippumattomuus

Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, satunnaismuuttujan X *ehdollinen jakauma* satunnaismuuttujan Y suhteen yhtyy satunnaismuuttujan X *reunajakaumaan*:

$$f_{X|Y}(x|y) = f_X(x) \text{ , jos } f_Y(y) > 0$$

Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, satunnaismuuttujan Y *ehdollinen jakauma* satunnaismuuttujan X suhteen yhtyy satunnaismuuttujan Y *reunajakaumaan*:

$$f_{Y|X}(y|x) = f_Y(y) \text{ , jos } f_X(x) > 0$$

Diskreetin kaksiulotteisen jakauman ehdolliset odotusarvot

Olkoot satunnaismuuttujat X ja Y *diskreettejä*.

Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on satunnaismuuttujan X ehdollisen jakauman odotusarvo:

$$\mathrm{E}(X \mid Y = y) = \sum_x x f_{X|Y}(x \mid y)$$

Satunnaismuuttujan Y **ehdollinen odotusarvo** satunnaismuuttujan X suhteen on satunnaismuuttujan Y ehdollisen jakauman odotusarvo:

$$\mathrm{E}(Y \mid X = x) = \sum_y y f_{Y|X}(y \mid x)$$

Jatkuvan kaksiulotteisen jakauman ehdolliset odotusarvot

Olkoot satunnaismuuttujat X ja Y *jatkuvia*.

Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on satunnaismuuttujan X ehdollisen jakauman odotusarvo:

$$\mathrm{E}(X \mid Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x \mid y)dx$$

Satunnaismuuttujan Y **ehdollinen odotusarvo** satunnaismuuttujan X suhteen on satunnais muuttujan Y ehdollisen jakauman odotusarvo:

$$\mathrm{E}(Y \mid X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y \mid x)dy$$

Ehdolliset odotusarvot ja riippumattomuus

Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, ehdolliset odotusarvot *yhtyvät* niiden *reuna-jakaumien odotusarvoihin*.

Jos siis X ja Y ovat *riippumattomia*, niin

$$\mathrm{E}(X \mid Y) = \mathrm{E}(X)$$

$$\mathrm{E}(Y \mid X) = \mathrm{E}(Y)$$

Iteroidun odotusarvon laki

Ehdolliset odotusarvot voidaan tulkita *satunnaismuuttujiksi* ehtomuuttujan suhteen.

Siten satunnaismuuttujan Y *ehdollisen odotusarvon odotusarvo* (satunnaismuuttujan X suhteen) on

$$\mathrm{E}_X\left[\mathrm{E}(Y \mid X)\right] = \mathrm{E}(Y)$$

Siten satunnaismuuttujan X *ehdollisen odotusarvon odotusarvo* (satunnaismuuttujan Y suhteen) on

$$\mathrm{E}_Y\left[\mathrm{E}(X \mid Y)\right] = \mathrm{E}(X)$$

Regressiofunktiot

Tarkastellaan satunnaismuuttujan X *ehdollista odotusarvoa*

$$\mathrm{E}(X \mid Y = y)$$

ehtomuuttujan Y arvojen y *funktiona*. Tätä funktiota kutsutaan satunnaismuuttujan X **regressio-funktioksi** satunnaismuuttujan Y suhteen.

Satunnaismuuttujan X regressiofunktio muuttujan Y suhteen määrittelee **regressiokäyrän**

$$x = g_y(y) = \mathrm{E}(X \mid Y = y)$$

Tarkastellaan satunnaismuuttujan Y *ehdollista odotusarvoa*

$$\mathrm{E}(Y \mid X = x)$$

ehtomuuttujan X arvojen x *funktiona*. Tätä funktiota kutsutaan satunnaismuuttujan Y **regressiofunktioksi** satunnaismuuttujan X suhteen.

Satunnaismuuttujan Y regressiofunktio muuttujan X suhteen määrittelee **regressiokäyrän**

$$y = g_x(x) = \mathrm{E}(Y \mid X = x)$$

Moniulotteisia jakaumia

Multinomijakauma

Multinomijakauma on *binomijakauman yleistys* useamman toisensa poissulkevan tapahtuman tilanteeseen.

Olkoon $A_1, A_2, \ldots , A_k$ otosavaruuden S *ositus*. Tällöin

$$A_i \cap A_j = \varnothing \text{ , } i \neq j$$

$$S = A_1 \cup A_2 \cup \cdots \cup A_k$$

Olkoot tapahtumien $A_1, A_2, \ldots , A_k$ *todennäköisyydet*:

$$\Pr(A_i) = p_i \text{ , } i = 1, 2, \ldots , k$$

$$p_1 + p_2 + \cdots + p_k = 1$$

Määritellään *satunnaismuuttujat* X_i , $i = 1, 2, \ldots , k$:

X_i = Tapahtuman A_i esiintymisten lukumäärä n-kertaisessa toistokokeessa

Tällöin

$$X_i \sim \mathrm{Bin}(n,\ p_i) \text{ , } i = 1, 2, \ldots, k$$

jossa

$$p_i = \Pr(A_i) \text{ , } i = 1, 2, \ldots , k$$

Lisäksi

$$X_1 + X_2 + \cdots + X_k = n$$

Multinomijakaumalla tarkoitetaan satunnaismuuttujien

$$X_1, X_2, \ldots , X_k$$

yhteisjakaumaa.

Huomautus:

Satunnaismuuttuja X_i *eivät ole riippumattomia*, koska niitä sitoo toisiinsa ehto

$$X_1 + X_2 + \cdots + X_k = n$$

jossa toistokokeiden lukumäärä n on *kiinteä* luku.

Satunnaismuuttujat $X_1, X_2, \ldots, X_k$ noudattavat $(k-1)$-ulotteista **multinomijakaumaa**, jos niiden *yhteisjakauman pistetodennäköisyysfunktio* on muotoa

$$\Pr(X_1 = n_1 \text{ ja } X_2 = n_2 \text{ ja } \ldots \text{ ja } X_k = n_k) = \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

jossa

$$p_1 + p_2 + \cdots + p_k = 1$$
$$n_1 + n_2 + \cdots + n_k = n$$

Merkintä:

$$(X_1, X_2, \ldots, X_k) \sim \text{Multinom}(p_1, p_2, \ldots, p_k; n)$$

Jos $k = 2$, niin multinomijakauma yhtyy *binomijakaumaan*:

$$\Pr_{Multinom}(X_1 = n_1 \text{ ja } X_2 = n - n_1) = \Pr_{Bin}(X_1 = n_1)$$

Multinomijakauman *yksiulotteiset reunajakaumat* ovat *binomijakaumia*.

Multinomitodennäköisyydet saadaan korottamalla *multinomi* $(p_1 + p_2 + \cdots + p_k)$ potenssiin n:

$$(p_1 + p_2 + \cdots + p_k)^n = \sum \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

jossa summa lasketaan yli kaikkien lukujen $n_1, n_2, \ldots, n_k$, joille pätee ehto

$$n_1 + n_2 + \cdots + n_k = n$$

2-ulotteinen normaalijakauma

Satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa **2-ulotteista normaalijakaumaa**, jos sen *tiheysfunktio* on

(1) $$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left\{-\frac{1}{2(1-\rho_{XY}^2)} Q(x,y)\right\}$$

jossa

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

ja

$$-\infty < \mu_X < +\infty \qquad \sigma_X > 0$$
$$-\infty < \mu_Y < +\infty \qquad \sigma_Y > 0$$
$$-1 \le \rho_{XY} \le +1$$

2-ulotteisen normaalijakauman tiheysfunktion määrittelevän yhtälön (1) hakasulkulauseke [·] määrää tiheysfunktion **tasa-arvokäyrät**. Kaikki tasa-arvokäyrät ovat **ellipsejä**, joiden yhtälöt voidaan ilmaista muodossa

$$Q(x,y)=\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2=c^2$$

missä c on vakio.

2-ulotteinen normaalijakauman tiheysfunktio (1) on *parametroitu* niin, että sen parametreina ovat satunnaismuuttujien X ja Y *odotusarvot*, *varianssit* ja *korrelaatio*.

Satunnaismuuttujien X ja Y **odotusarvot** ovat

$$\mu_X = \mathrm{E}(X)$$
$$\mu_Y = \mathrm{E}(Y)$$

Satunnaismuuttujien X ja Y **varianssit** ovat

$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}\left[(X-\mu_X)^2\right]$$
$$\sigma_Y^2 = \mathrm{Var}(Y) = \mathrm{E}\left[(Y-\mu_Y)^2\right]$$

Satunnaismuuttujien X ja Y **korrelaatio** on

$$\rho_{XY} = \mathrm{Cor}(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

jossa

$$\sigma_{XY} = \mathrm{Cov}(X,Y) = \mathrm{E}\left[(X-\mu_X)(Y-\mu_Y)\right]$$

on satunnaismuuttujien X ja Y *kovarianssi*.

2-ulotteista normaalijakaumaa noudattavan satunnaismuuttujien X ja Y parin (X, Y) **odotusarvo-vektori** on

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

ja **kovarianssimatriisi** on

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

2-ulotteisen normaalijakauman **reunajakaumat** ovat 1-*ulotteisia normaalijakaumia*:

$$Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$$
$$X \sim \mathrm{N}(\mu_X, \sigma_X^2)$$

2-*ulotteisen normaalijakauman tapauksessa* satunnaismuuttujien X ja Y *korreloimattomuudesta seuraa* niiden *riippumattomuus*. Muista, että *aina pätee* se, että *riippumattomuudesta seuraa korreloimattomuus*.

2-ulotteisen normaalijakauman **ehdolliset jakaumat** ovat 1-*ulotteisia normaalijakaumia*:

$$(Y \mid X) \sim \mathrm{N}(\mu_Y + \beta_{YX}(x-\mu_X)\,,\, \sigma_Y^2(1-\rho_{XY}^2))\,,\ \beta_{YX} = \rho_{XY}\frac{\sigma_Y}{\sigma_X}$$

$$(X \mid Y) \sim \mathrm{N}(\mu_X + \beta_{XY}(y-\mu_Y)\,,\, \sigma_X^2(1-\rho_{XY}^2))\,,\ \beta_{XY} = \rho_{XY}\frac{\sigma_X}{\sigma_Y}$$

Ehdollinen odotusarvo

$$\mathrm{E}(Y \mid X) = \mu_Y + \beta_{YX}(x-\mu_X)$$

on muuttujan Y **regressiofunktio** muuttujan X suhteen. Ehdollinen odotusarvo E($Y|X$) määrää *suoran*

$$y = \mu_Y + \beta_{YX}(x-\mu_X)$$

Suoran *kulmakerroin* on $\beta_{YX} = \rho_{XY}\dfrac{\sigma_Y}{\sigma_X}$ ja se kulkee pisteen (μ_X,μ_Y) kautta.

Ehdollinen odotusarvo

$$\mathrm{E}(X \mid Y) = \mu_X + \beta_{XY}(y-\mu_Y)$$

on muuttujan X **regressiofunktio** muuttujan Y suhteen. Ehdollinen odotusarvo E($X|Y$) määrää *suoran*

$$x = \mu_X + \beta_{XY}(y-\mu_Y)$$

Suoran *kulmakerroin* on $\beta_{XY} = \rho_{XY}\dfrac{\sigma_X}{\sigma_Y}$ ja se kulkee pisteen (μ_X,μ_Y) kautta.

Huomaa, että regressiosuorien kulmakertoimet β_{YX} ja β_{XY} toteuttavat yhtälön

$$\beta_{YX}\beta_{XY} = \rho_{XY}^2$$

2-ulotteisen normaalijakauman tapauksessa satunnaismuuttujien Y ja X **yhteiskorrelaatiokerroin** on

$$R_{Y|X} = \frac{\sigma_{YX}}{\sigma_Y\sigma_X} = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = R_{X|Y}$$

ja **ehdolliset varianssit** ovat

$$\sigma_{Y|X}^2 = \sigma_Y^2(1-\rho_{XY}^2)$$

$$\sigma_{X|Y}^2 = \sigma_X^2(1-\rho_{XY}^2)$$

2-ulotteisen normaalijakauman tiheysfunktiota muodon määräävien **tasa-arvoellipsien**

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 = c^2$$

pääakseleiden pituudet (oik. pituuksien suhteet) ja **suunnat** saadaan määräämällä *kovarianssimatriisin*

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

ominaisarvot ja **ominaisvektorit**.

Matriisin Σ **ominaisarvot** saadaan määräämällä determinanttiyhtälön

$$|\mathbf{\Sigma} - \lambda\mathbf{I}| = \begin{vmatrix} \sigma_X^2 - \lambda & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 - \lambda \end{vmatrix} = \lambda^2 - (\sigma_X^2 + \sigma_Y^2)\lambda + \sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 = 0$$

nollakohdat muuttujan λ suhteen. Yhtälöllä on (aina) kaksi reaalista ja ei-negatiivista nollakohtaa, jotka ovat siis matriisin Σ ominaisarvot.

Matriisin Σ ominaisarvoja λ_1 ja λ_2 vastaavat **ominaisvektorit**

$$\mathbf{q}_1 = (q_{11}, q_{21})$$

$$\mathbf{q}_2 = (q_{12}, q_{22})$$

saadaan yhtälöistä

$$\mathbf{\Sigma}\mathbf{q}_i = \lambda_i\mathbf{q}_i \text{ , } i = 1, 2$$

ottamalla huomioon ehdot

$$\mathbf{q}_i'\mathbf{q}_i = q_{1i}^2 + q_{2i}^2 = 1 \text{ , } i = 1, 2$$

Tasa-arvoellipsien

$$Q(x,y) = c^2$$

pääakseleiden *pituudet* suhtautuvat toisiinsa kuten *ominaisarvojen* λ_1 ja λ_2 *neliöjuuret* ja pääakseleiden *suunnat* yhtyvät vastaavien *ominaisvektoreiden suuntiin*.

Tehtävä 6.1.

Heitetään kahta virheetöntä noppaa (nopan virheettömyydellä tarkoitetaan sitä, että silmälukujen 1, 2, 3, 4, 5, 6 todennäköisyydet ovat yhtä suuria). Määritellään satunnaismuuttujat

X = tulos 1. nopan heitosta

Y = tulos 2. nopan heitosta

$U = \min(X,Y)$

$V = \max(X,Y)$

Määrää:

(a) Satunnaismuuttujan U jakauma.

(b) Satunnaismuuttujan V jakauma.

(c) Satunnaismuuttujien U ja V yhteisjakauma.

(d) $\mathrm{E}(U)$

(e) $\mathrm{E}(V)$

(f) Satunnaismuuttujan U ehdollinen jakauma ehdolla $V = 4$.

(g) Satunnaismuuttujan V ehdollinen jakauma ehdolla $U = 4$.

(h) $\mathrm{E}(U \mid V = 4)$

(i) $\mathrm{E}(V \mid U = 4)$

Tehtävä 6.1. – Mitä opimme?

Tehtävässä tarkastellaan *diskreettien* satunnaismuuttujien *yhteisjakaumaa*, *reunajakaumia* ja *ehdollisia jakaumia* sekä *tunnuslukuja*.

Tehtävä 6.1. – Ratkaisu:

Koska nopat oletettiin virheettömiksi, satunnaismuuttujien X ja Y *pistetodennäköisyysfunktiot*

$f_X(i) = \Pr(X = i)$, $i = 1, 2, 3, 4, 5, 6$

$f_Y(i) = \Pr(Y = i)$, $i = 1, 2, 3, 4, 5, 6$

voidaan esittää seuraavana taulukkona:

i	**1**	**2**	**3**	**4**	**5**	**6**
$f_X(i)$	1/6	1/6	1/6	1/6	1/6	1/6
$f_Y(i)$	1/6	1/6	1/6	1/6	1/6	1/6

(a) Muodostetaan heittotulosten *minimille*

$U = \min(X, Y)$

jossa

X = 1. nopan heiton tulos

Y = 2. nopan heiton tulos

seuraava aputaulukko:

$U = \min(X, Y)$		1. nopan heiton tulos X					
		1	2	3	4	5	6
2. nopan heiton tulos Y	1	1	1	1	1	1	1
	2	1	2	2	2	2	2
	3	1	2	3	3	3	3
	4	1	2	3	4	4	4
	5	1	2	3	4	5	5
	6	1	2	3	4	5	6

Satunnaismuuttujan $U = \min(X, Y)$ *pistetodennäköisyysfunktio*

$f_U(i) = \Pr(U = i)$, $i = 1, 2, 3, 4, 5, 6$

voidaan lukea tästä aputaulukosta. Pistetodennäköisyysfunktio voidaan esittää seuraavana taulukkona:

i	1	2	3	4	5	6
$f_U(i)$	11/36	9/36	7/36	5/36	3/36	1/36

Esimerkiksi silmäluku 5 voi tulla minimiksi täsmälleen kolmella eri tavalla:

1. nopan heiton tulos X	2. nopan heiton tulos Y	$U = \min(X, Y)$
5	5	5
5	6	5
6	5	5

(b) Muodostetaan heittotulosten *maksimille*

$$V = \max(X, Y)$$

jossa

X = 1. nopan heiton tulos

Y = 2. nopan heiton tulos

seuraava aputaulukko:

$V = \max(X, Y)$		1. nopan heiton tulos X					
		1	2	3	4	5	6
2. nopan heiton tulos Y	1	1	2	3	4	5	6
	2	2	2	3	4	5	6
	3	3	3	3	4	5	6
	4	4	4	4	4	5	6
	5	5	5	5	5	5	6
	6	6	6	6	6	6	6

Satunnaismuuttujan $V = \max(X, Y)$ *pistetodennäköisyysfunktio*

$$f_V(i) = \Pr(V = i)\ ,\ i = 1, 2, 3, 4, 5, 6$$

voidaan lukea tästä aputaulukosta. Pistetodennäköisyysfunktio voidaan esittää seuraavana taulukkona:

i	1	2	3	4	5	6
$f_V(i)$	1/36	3/36	5/36	7/36	9/36	11/36

Esimerkiksi silmäluku 2 voi tulla maksimiksi täsmälleen kolmella eri tavalla:

1. nopan heiton tulos X	2. nopan heiton tulos Y	$V = \max(X, Y)$
1	2	2
2	1	2
2	2	2

(c) Satunnaismuuttujien $U = \min(X, Y)$ ja $V = \max(X, Y)$ *yhteisjakauman pistetodennäköisyysfunktio*

$$f_{UV}(i, j) = \Pr(U = i \text{ ja } V = j)$$

voidaan esittää seuraavana taulukkona:

$f_{UV}(i,j)$		$U = \min(X, Y) = i$						
		1	2	3	4	5	6	Yht.
$V = \max(X, Y) = j$	1	1/36	0	0	0	0	0	1/36
	2	2/36	1/36	0	0	0	0	3/36
	3	2/36	2/36	1/36	0	0	0	5/36
	4	2/36	2/36	2/36	1/36	0	0	7/36
	5	2/36	2/36	2/36	2/36	1/36	0	9/36
	6	2/36	2/36	2/36	2/36	2/36	1/36	11/36
	Yht.	11/36	9/36	7/36	5/36	3/36	1/36	1

Esimerkiksi:

$$f_{UV}(5,3) = \Pr(U = 5 \text{ ja } V = 3) = 0$$

koska kahden luvun minimi ei voi olla maksimia suurempi.

Esimerkiksi:

$$f_{UV}(3,5) = \Pr(U = 3 \text{ ja } V = 5) = 2/36$$

koska tulos

$$\{U = 3 \text{ ja } V = 5\}$$

voi syntyä täsmälleen kahdella eri tavalla:

1. nopan heiton tulos X	2. nopan heiton tulos Y	$U = \min(X,Y)$	$V = \max(X,Y)$
3	5	3	5
5	3	3	5

(d) Satunnaismuuttujan U *odotusarvo* on

$$\begin{aligned}\mathrm{E}(U) &= \sum_{i=1}^{6} i f_U(i) = \sum_{i=1}^{6} i \Pr(U = i) \\ &= 1\times\frac{11}{36} + 2\times\frac{9}{36} + 3\times\frac{7}{36} + 4\times\frac{5}{36} + 5\times\frac{3}{36} + 6\times\frac{1}{36} = \frac{91}{36} = 2.528\end{aligned}$$

(e) Satunnaismuuttujan V *odotusarvo* on

$$\begin{aligned} \mathrm{E}(V) &= \sum_{j=1}^{6} j f_V(j) = \sum_{j=1}^{6} j \Pr(V=j) \\ &= 1\times\frac{1}{36} + 2\times\frac{3}{36} + 3\times\frac{5}{36} + 4\times\frac{7}{36} + 5\times\frac{9}{36} + 6\times\frac{11}{36} = \frac{161}{36} = 4.472 \end{aligned}$$

(f) Satunnaismuuttujan U *ehdolliset pistetodennäköisyysfunktiot*, ehdolla V saadaan kaavasta

$$f_{U|V}(i\,|\,j) = \frac{f_{UV}(i,j)}{f_V(j)}\,,\ i=1,2,\ldots,6\,,\ j=1,2,\ldots,6$$

Tuloksena saadaan ao. taulukko, jossa ehdolliset jakaumat ovat *riveinä*:

$f_{U\|V}(i\|j)$		$U = i \mid V = j$						
		1	2	3	4	5	6	Yht.
$V = j$	1	1	0	0	0	0	0	1
	2	2/3	1/3	0	0	0	0	1
	3	2/5	2/5	1/5	0	0	0	1
	4	**2/7**	**2/7**	**2/7**	**1/7**	**0**	**0**	**1**
	5	2/9	2/9	2/9	2/9	1/9	0	1
	6	2/11	2/11	2/11	2/11	2/11	1/11	1

Taulukko saadaan jakamalla (c)-kohdan taulukon solutodennäköisyydet

$f_{UV}(i,j) = \Pr(U = i \text{ ja } V = j)$

satunnaismuuttujan V reunajakauman todennäköisyyksillä

$f_V(j) = \Pr(V = j)$

jotka ovat (c)-kohdan taulukossa *rivisummina*.

Kysytty ehdollinen jakauma

$f_{U|V}(i|j = 4)$

on merkitty taulukkoon lihavoituna.

(g) Satunnaismuuttujan V *ehdolliset pistetodennäköisyysfunktiot*, ehdolla U saadaan kaavasta

$$f_{V|U}(j\,|\,i) = \frac{f_{UV}(i,j)}{f_U(i)}\ ,\ i = 1,2,\ldots,6\ ,\ j = 1,2,\ldots,6$$

Tuloksena saadaan ao. taulukko, jossa ehdolliset jakaumat ovat *sarakkeina*:

$f_{V\|U}(j\|i)$		$U = i$					
		1	2	3	**4**	5	6
$V = j \mid U = i$	1	1	0	0	**0**	0	0
	2	2/11	1/9	0	**0**	0	0
	3	2/11	2/9	1/7	**0**	0	0
	4	2/11	2/9	2/7	**1/5**	0	0
	5	2/11	2/9	2/7	**2/5**	1/3	0
	6	2/11	2/9	2/7	**2/5**	2/3	1
	Yht.	1	1	1	**1**	1	1

Taulukko saadaan jakamalla (c)-kohdan taulukon solutodennäköisyydet

$f_{UV}(i,j) = \Pr(U = i \text{ ja } V = j)$

satunnaismuuttujan U reunajakauman todennäköisyyksillä

$f_U(i) = \Pr(U = i)$

jotka ovat (c)-kohdan taulukossa *sarakesummina*.

Kysytty ehdollinen jakauma

$f_{U|V}(i|j = 4)$

on merkitty taulukkoon lihavoituna.

(h) *Ehdollinen odotusarvo*

$\mathrm{E}(U \mid V = 4)$

saadaan (f)-kohdan taulukosta:

$$\begin{aligned}\mathrm{E}(U \mid V = 4) &= \sum_{i=1}^{6} i f_{U|V}(i\,|\,j = 4)\\ &= 1\times\frac{2}{7} + 2\times\frac{2}{7} + 3\times\frac{2}{7} + 4\times\frac{1}{7} + 5\times 0 + 6\times 0 = \frac{16}{7} = 2.286\end{aligned}$$

(i) *Ehdollinen odotusarvo*

$$\mathrm{E}(V \mid U = 4)$$

saadaan (g)-kohdan taulukosta:

$$\mathrm{E}(V \mid U = 4) = \sum_{j=1}^{6} j f_{V|U}(j \mid i = 4)$$
$$= 1\times 0 + 2\times 0 + 3\times 0 + 4\times\frac{1}{5} + 5\times\frac{2}{5} + 6\times\frac{2}{5} = \frac{26}{5} = 5.2$$

Tehtävä 6.2.

Olkoon satunnaismuuttujien X ja Y yhteisjakauman pistetodennäköisyysfunktio

$$\Pr(X = -1, Y = 3) = \Pr(X = 0, Y = -2) = \Pr(X = 0, Y = 1) = \Pr(X = 2, Y = -2) = 1/4$$

Määrää:

(a) Satunnaismuuttujien X ja Y reunajakaumat.

(b) $\mathrm{Cov}(X, Y)$

(c) $\mathrm{Cor}(X, Y)$

(d) Satunnaismuuttujan Y ehdolliset jakaumat.

(e) $\mathrm{E}(Y \mid X)$

Tehtävä 6.2. – Mitä opimme?

Tehtävässä tarkastellaan *diskreettien* satunnaismuuttujien *yhteisjakaumaa*, *reunajakaumia* ja *ehdollisia jakaumia* sekä *tunnuslukuja*.

Tehtävä 6.2. – Ratkaisu:

(a) Satunnaismuuttujien X ja Y *yhteisjakauman*

$$f_{XY}(x, y) = \Pr(X = x \text{ ja } Y = y)$$

pistetodennäköisyysfunktio voidaan esittää seuraavana taulukkona:

$f_{XY}(x, y)$		x		
		–1	0	2
y	3	1/4	0	0
	1	0	1/4	0
	–2	0	1/4	1/4

Reunajakaumien

$$f_X(x) = \Pr(X = x) = \textstyle\sum_y f_{XY}(x, y)$$
$$f_Y(y) = \Pr(Y = y) = \textstyle\sum_x f_{XY}(x, y)$$

pistetodennäköisyysfunktiot saadaan tästä taulukosta rivi- ja sarakesummina:

$f_{XY}(x, y)$	x				$f_Y(y)$
		–1	0	2	
y	3	1/4	0	0	1/4
	1	0	1/4	0	1/4
	–2	0	1/4	1/4	1/2
$f_X(x)$		1/4	1/2	1/4	1

(b) Lasketaan satunnaismuuttujien X ja Y *kovarianssi* kaavalla

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

Lasketaan ensin satunnaismuuttujien X ja Y *odotusarvot*:

$$\mathrm{E}(X) = \sum_{i=1}^{3} x_i \Pr(X = x_i) = -1\times\frac{1}{4} + 0\times\frac{1}{2} + 2\times\frac{1}{4} = \frac{1}{4} = 0.25$$

$$\mathrm{E}(Y) = \sum_{i=1}^{3} y_i \Pr(Y = y_i) = -2\times\frac{1}{2} + 1\times\frac{1}{4} + 3\times\frac{1}{4} = 0$$

Edelleen

$$\mathrm{E}(XY) = \sum_{i=1}^{3}\sum_{j=1}^{3} x_i y_j \Pr(X = x_i, Y = y_j)$$

$$= (-1)\times 3\times\frac{1}{4} + 0\times(-2)\times\frac{1}{4} + 0\times 1\times\frac{1}{4} + 2\times(-2)\times\frac{1}{4} = -\frac{7}{4} = -1.75$$

Siten satunnaismuuttujien X ja Y *kovarianssi* on

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))] \\ &= \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) \\ &= -\frac{7}{4} - \frac{1}{4}\times 0 = -\frac{7}{4} = -1.75\end{aligned}$$

(c) Satunnaismuuttujien X ja Y *korrelaatio* on

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{D}(X)\,\mathrm{D}(Y)}$$

Satunnaismuuttujien X ja Y kovarianssi on laskettu (b)-kohdassa, mutta joudumme laskemaan vielä satunnaismuuttujien X ja Y *standardipoikkeamat*.

Lasketaan ensin satunnaismuuttujille X ja Y *2. origomomentit*:

$$\mathrm{E}(X^2) = \sum_{i=1}^{3} x_i^2 \Pr(X = x_i) = (-1)^2\times\frac{1}{4} + 0^2\times\frac{1}{2} + 2^2\times\frac{1}{4} = \frac{5}{4} = 1.25$$

$$\mathrm{E}(Y^2) = \sum_{i=1}^{3} y_i^2 \Pr(Y = y_i) = (-2)^2 \times \frac{1}{2} + 1^2 \times \frac{1}{4} + 3^2 \times \frac{1}{4} = \frac{18}{4} = \frac{9}{2} = 4.5$$

Satunnaismuuttujien X ja Y *varianssit* ovat:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \mathrm{E}[X - \mathrm{E}(X)]^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \frac{5}{4} - \left[\frac{1}{4}\right]^2 = \frac{19}{16} = 1.1875$$

$$\mathrm{Var}(Y) = \mathrm{D}^2(Y) = \mathrm{E}[Y - \mathrm{E}(Y)]^2 = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2 = \frac{9}{2} - 0^2 = \frac{9}{2} = 4.5$$

Siten

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{D}(X)\,\mathrm{D}(Y)} = \frac{-\frac{7}{4}}{\sqrt{\frac{19}{16}} \times \sqrt{\frac{9}{2}}} = -\frac{7\sqrt{2}}{\sqrt{171}} = -0.7570$$

(d) Muodostetaan satunnaismuuttujan *Y ehdollisten jakaumien pistetodennäköisyysfunktiot*, kun ehtomuuttujana on X:

$$f_{Y|X}(y) = \frac{f_{YX}(y,x)}{f_X(x)}$$

$x = -1$:

y	–2	1	3
$f_{Y\|X}(y)$	0	0	1

$x = 0$:

y	–2	1	3
$f_{Y\|X}(y)$	1/2	1/2	0

$x = 2$:

y	–2	1	3
$f_{Y\|X}(y)$	1	0	0

Ehdollisten jakaumien pistetodennäköisyydet saadaan jakamalla (a)-kohdassa esitetyn satunnaismuuttujien X ja Y yhteisjakauman pistetodennäköisyysfunktiota kuvaavan taulukon solutodennäköisyydet satunnaismuuttujan X reunajakauman todennäköisyyksillä.

(e) Ehdolliset odotusarvot

$$\mathrm{E}(Y \mid X = x) = \sum y f_{Y|X}(y)$$

saadaan kohdasta (d):

x	-1	0	2
$\mathrm{E}(Y\|X = x)$	3	$-1/2$	-2

Esimerkiksi:

$$\mathrm{E}(Y \mid X = 2) = \sum_{j=1}^{3} y_j \Pr(Y = y_j \mid X = 2) = -2\times 1 + 1\times 0 + 3\times 0 = -2$$

Tehtävä 6.3.

Olkoon satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio

$$f(x, y) = C(x + y)\ ,\ 0 \le x \le 1,\ 0 \le y \le 1 - x$$

jossa C on vakio.

Määrää:

(a) Vakio C.

(b) $\Pr(X \ge Y)$

(c) Satunnaismuuttujien X ja Y yhteisjakauman kertymäfunktio.

(d) Satunnaismuuttujien X ja Y reunajakaumat. Ovatko X ja Y riippumattomia?

(e) Tiheysfunktio satunnaismuuttujan X ehdolliselle jakaumalle ehdolla Y.

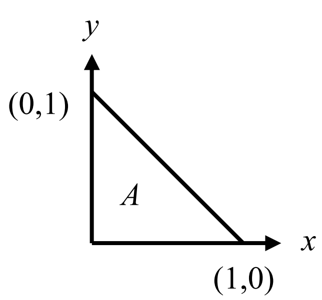
(f) Ehdollinen odotusarvo $\mathrm{E}(X \mid Y)$.

Tehtävä 6.3. – Mitä opimme?

Tehtävässä tarkastellaan *jatkuvien* satunnaismuuttujien *yhteisjakaumaa*, *reuna-jakaumia* ja *ehdollisia jakaumia* sekä *tunnuslukuja*.

Tehtävä 6.3. – Ratkaisu:

(a) Vakio C saadaan määrätyksi integroimalla satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio sen määrittelyalueen A yli:



Saamme siten seuraavan yhtälön vakion C määräämiseksi:

$$\begin{aligned}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dydx &= C\int_0^1\int_0^{1-x}(x+y)dydx \\ &= C\int_0^1\left[xy+\frac{1}{2}y^2\right]_0^{1-x}dx \\ &= C\int_0^1\left(\frac{1}{2}-\frac{1}{2}x^2\right)dx \qquad \text{z} \\ &= C\left[\frac{1}{2}x-\frac{1}{6}x^3\right]_0^1 \\ &= C\left(\frac{1}{2}-\frac{1}{6}\right)=\frac{1}{3}C=1\end{aligned}$$

Ratkaisuksi saadaan

$$C = 3$$

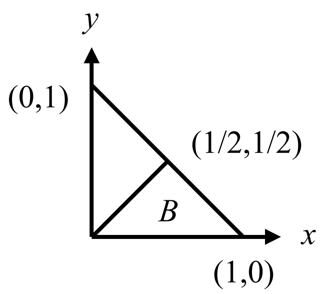
Siten satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* on muotoa

$$f(x, y) = 3(x + y)\ ,\ 0 \le x \le 1,\ 0 \le y \le 1 - x$$

(b) Todennäköisyys

$$\Pr(X \ge Y)$$

saadaan integroimalla yhteisjakauman tiheysfunktio alueen B yli:



Siten

$$\begin{aligned}\Pr\left(X \ge Y\right) &= 3\int_0^{1/2}\int_0^{x}(x+y)dydx + 3\int_{1/2}^{1}\int_0^{1-x}(x+y)dydx \\ &= 3\int_0^{1/2}\left[xy+\frac{1}{2}y^2\right]_0^{x}dx + 3\int_{1/2}^{1}\left[xy+\frac{1}{2}y^2\right]_0^{1-x}dx \\ &= 3\int_0^{1/2}\left(\frac{3}{2}x^2\right)dx + 3\int_{1/2}^{1}\left(\frac{1}{2}-\frac{1}{2}x^2\right)dx \\ &= 3\left[\frac{3}{6}x^3\right]_0^{1/2} + 3\left[\frac{1}{2}x-\frac{1}{6}x^3\right]_{1/2}^{1} = \frac{1}{2}\end{aligned}$$

(c) Satunnaismuuttujien X ja Y yhteisjakauman *kertymäfunktioksi* saadaan

$$\begin{aligned}
F_{XY}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)dvdu \\
&= 3\int_{0}^{x}\int_{0}^{y}(u+v)dvdu \\
&= 3\int_{0}^{x}\left[uv+\frac{1}{2}v^2\right]_0^y du \\
&= 3\int_{0}^{x}\left(uy+\frac{1}{2}y^2\right)du \\
&= 3\left[\frac{1}{2}u^2y+\frac{1}{2}uy^2\right]_0^x \\
&= \frac{3}{2}x^2y+\frac{3}{2}xy^2
\end{aligned}$$

Kertymäfunktio on määritelty alueella

$$A = \{(x,y) \mid 0 \le x \le 1,\ 0 \le y \le 1-x\}$$

(d) Satunnaismuuttujien X ja Y *reunajakaumat* saadaan integroimalla niiden yhteisjakauman tiheysfunktio vuorotellen kummankin muuttujan suhteen:

$$f_X(x) = 3\int_0^{1-x}(x+y)dy = 3\left[xy+\frac{1}{2}y^2\right]_0^{1-x} = \frac{3}{2}(1-x^2)$$

$$f_Y(y) = 3\int_0^{1-y}(x+y)dx = 3\left[\frac{1}{2}x^2+xy\right]_0^{1-y} = \frac{3}{2}(1-y^2)$$

Satunnaismuuttujat X ja Y *eivät ole riippumattomia*, koska

$$f_X(x)f_Y(y) = \frac{9}{4}(1-x^2)(1-y^2) = \frac{9}{4}(1-x^2-y^2+x^2y^2) \neq 3(x+y) = f_{XY}(x,y)$$

(e) Satunnaismuuttujan X *ehdollinen jakauma*, kun ehtomuuttujana on Y, on

$$f_{X|Y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{3(x+y)}{\frac{3}{2}(1-y^2)} = \frac{2(x+y)}{1-y^2}$$

Satunnaismuuttujan X ehdollinen jakauma ehdolla $Y = y$ *riippuu* y:stä, jolloin esimerkiksi myös sen odotusarvo ja varianssi riippuvat y:stä. Tämä on ymmärrettävää, koska satunnaismuuttujat X ja Y *eivät ole riippumattomia*.

(f) Satunnaismuuttujan X *ehdollinen odotusarvo*, kun ehtomuuttujana on Y, on

$$
\begin{aligned}
\mathrm{E}(X|Y) &= \int_{-\infty}^{+\infty} x f_{X|Y}(x)dx \\
&= \int_{0}^{1-y} x \frac{2(x+y)}{1-y^2} dx \\
&= \frac{2}{1-y^2} \int_{0}^{1-y} (x^2 + xy)dx \\
&= \frac{2}{1-y^2} \left[\frac{1}{3}x^3 + \frac{1}{2}x^2 y \right]_0^{1-y} \\
&= \frac{1}{3} \cdot \frac{(1-y)(2+y)}{1+y}
\end{aligned}
$$

Satunnaismuuttujan X ehdollinen odotusarvo ehdolla $Y = y$ eli satunnaismuuttujan X *regressiofunktio* satunnaismuuttujan Y suhteen *riippuu y*:stä. Tämä on ymmärrettävää, koska satunnaismuuttujat X ja Y *eivät ole riippumattomia*.

Tehtävä 6.4.

Olkoon satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio

$$f(x, y) = Ce^{x+y}\ ,\ 0 \le x \le 1,\ 0 \le y \le 1$$

jossa C on vakio.

Määrää:

(a) Vakio C.

(b) Satunnaismuuttujien X ja Y reunajakaumat.

(c) Ovatko X ja Y riippumattomia?

Tehtävä 6.4. – Mitä opimme?

Tehtävässä tarkastellaan *jatkuvien* satunnaismuuttujien *yhteisjakaumaa*, *reuna-jakaumia* ja *ehdollisia jakaumia* sekä *tunnuslukuja*.

Tehtävä 6.4. – Ratkaisu:

(a) Vakio C saadaan määrätyksi integroimalla satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio sen määrittelyalueen yli, koska aina pätee, että

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y)dydx \equiv 1$$

Saamme siten seuraavan yhtälön vakion C määräämiseksi:

$$\begin{aligned}
C\int_0^1\int_0^1 e^{x+y}dydx &= C\int_0^1 e^x\int_0^1 e^y dydx \\
&= C\int_0^1 \left[e^y\right]_0^1 dx \\
&= C(e-1)\int_0^1 e^x dx \\
&= C(e-1)\left[e^x\right]_0^1 \\
&= C(e-1)^2 = 1
\end{aligned}$$

Ratkaisuksi saadaan

$C = 1/(e-1)^2$

Siten satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* on

$$f(x,y) = \frac{1}{(e-1)^2}e^{x+y}\ ,\ 0 \le x \le 1\ ,\ 0 \le y \le 1$$

(b) Koska satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio faktoroituu muotoon

$$\begin{aligned}
f(x,y) &= \frac{1}{(e-1)^2}e^{x+y} \\
&= \frac{1}{e-1}e^x \times \frac{1}{e-1}e^y
\end{aligned}$$

niin satunnaismuuttujien X ja Y reunajakaumien tiheysfunktiot ovat

$$f_X(x) = \frac{1}{e-1}e^x$$

$$f_Y(y) = \frac{1}{e-1}e^y$$

(c) Satunnaismuuttujat X ja Y *ovat riippumattomia*, koska

$$f_X(x)f_Y(y) = f_{XY}(x,y)$$

Tehtävä 6.5.

Alla olevassa taulukossa on annettu satunnaismuuttujien X ja Y *reunajakaumat*:

$f_{XY}(x, y)$		x			$f_Y(y)$
		-1	0	2	
y	3				1/4
	1				1/4
	-2				1/2
$f_X(x)$		1/4	1/2	1/4	1

Täytä taulukon solut satunnaismuuttujien X ja Y yhteisjakauman pistetodennäköisyyksillä niin,
että satunnaismuuttujat X ja Y ovat *riippumattomia*.

Tehtävä 6.5. – Mitä opimme?

Tehtävässä tarkastellaan *riippumattomuuden* käsitettä.

Tehtävä 6.5. – Ratkaisu:

Diskreetit satunnaismuuttujat X ja Y ovat riippumattomia, jos ja vain jos niiden yhteisjakauman pistetodennäköisyysfunktio voidaan esittää sen reunajakauman pistetodennäköisyysfunktioiden tulona:

$$f_{XY}(x, y) = \Pr(X = x \text{ ja } Y = y) = \Pr(X = x)\Pr(Y = y) = f_X(x)f_Y(y)$$

Siten saamme *yhteisjakauman pistetodennäköisyyksiksi* seuraavan taulukon luvut:

$f_{XY}(x, y)$		x			$f_Y(y)$
		-1	0	2	
y	3	1/16	1/8	1/16	1/4
	1	1/16	1/8	1/16	1/4
	-2	1/8	1/4	1/8	1/2
$f_X(x)$		1/4	1/2	1/4	1

Tehtävä 6.6.

Oletetaan, että satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa kaksiulotteista normaalijakaumaa seuraavin parametrein:

$$\mathrm{E}(X) = +2 \qquad \mathrm{Var}(X) = 9$$
$$\mathrm{E}(Y) = -10 \qquad \mathrm{Var}(Y) = 4$$
$$\mathrm{Cov}(X, Y) = -5$$

Määrää:

(a) Satunnaismuuttujien X ja Y reunajakaumat.

(b) Cor(X, Y)

(c) Satunnaismuuttujan X ehdollinen jakauma, kun ehtomuuttujana on Y.

(d) Satunnaismuuttujan Y ehdollinen jakauma, kun ehtomuuttujana on X.

Tehtävä 6.6. – Mitä opimme?

Tehtävässä tarkastellaan 2-*ulotteisen normaalijakauman* ominaisuuksia.

Tehtävä 6.6. – Ratkaisu:

Oletuksen mukaan

$$\mathrm{E}(X) = \mu_X = +2 \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = 9$$
$$\mathrm{E}(Y) = \mu_Y = -10 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2 = 4$$
$$\mathrm{Cov}(X,Y) = \sigma_{XY} = -5$$

(a) Satunnaismuuttujien X ja Y *reunajakaumat* ovat normaalisia:

$X \sim \mathrm{N}(+2, 9)$

jossa

$$\mathrm{E}(X) = \mu_X = +2 \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = 9$$

ja

$Y \sim \mathrm{N}(-10, 4)$

jossa

$$\mathrm{E}(Y) = \mu_Y = -10 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2 = 4$$

(b) Määritelmän mukaan satunnaismuuttujien X ja Y *korrelaatiokerroin* on

$$\rho_{XY} = \mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{D}(X)\,\mathrm{D}(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-5}{3\times 2} = -\frac{5}{6} = -0.8333$$

(c) Satunnaismuuttujan X *ehdollinen jakauma*, kun ehtomuuttujana on Y, on normaalinen:

$X \mid Y \sim \mathrm{N}(\mathrm{E}(X \mid Y), \mathrm{Var}(X \mid Y))$

Satunnaismuuttujan X *ehdollisen jakauman* (ehdolla Y) odotusarvo eli satunnaismuuttujan X *ehdollinen odotusarvo* (ehdolla Y) on

$$\mathrm{E}(X \mid Y = y) = \mu_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = 2 - \frac{5}{6}\times\frac{3}{2}(y+10) = -\frac{5}{4}y - \frac{21}{4}$$

Satunnaismuuttujan X ehdollinen odotusarvo on ehtomuuttujan Y arvojen funktiona *lineaarinen*.

Satunnaismuuttujan X *ehdollisen jakauman* (ehdolla Y) varianssi eli satunnaismuuttujan X *ehdollinen varianssi* (ehdolla Y) on

$$\mathrm{Var}(X \mid Y = y) = (1-\rho_{XY}^2)\sigma_X^2 = \left(1-\left(\frac{5}{6}\right)^2\right)9 = \frac{11}{36}\times 9 = 2.75 < 9 = \sigma_X^2$$

Satunnaismuuttujan X ehdollinen varianssi (ehdolla Y) *ei riipu* ehtomuuttujan Y arvoista.

(d) Satunnaismuuttujan Y *ehdollinen jakauma*, kun ehtomuuttujana on X, on normaalinen:

$$Y \mid X \sim \mathrm{N}(\mathrm{E}(Y \mid X), \mathrm{Var}(Y \mid X))$$

Satunnaismuuttujan Y *ehdollisen jakauman* (ehdolla X) odotusarvo eli satunnaismuuttujan Y *ehdollinen odotusarvo* (ehdolla X) on

$$\mathrm{E}(Y \mid X = x) = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-\mu_X) = -10 - \frac{5}{6}\times\frac{4}{9}(x-2) = -\frac{10}{27}x - \frac{260}{27}$$

Satunnaismuuttujan Y ehdollinen odotusarvo on ehtomuuttujan X arvojen funktiona *lineaarinen*.

Satunnaismuuttujan Y *ehdollisen jakauman* (ehdolla X) varianssi eli satunnaismuuttujan Y *ehdollinen varianssi* (ehdolla X) on

$$\mathrm{Var}(Y \mid X = x) = (1-\rho_{XY}^2)\sigma_Y^2 = \left(1-\left(\frac{5}{6}\right)^2\right)4 = \frac{11}{36}\times 4 = \frac{11}{9} = 1.222 < 9 = \sigma_Y^2$$

Satunnaismuuttujan Y ehdollinen varianssi (ehdolla X) *ei riipu* ehtomuuttujan X arvoista.

Tehtävä 6.7.

Oletetaan, että satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa kaksiulotteista normaalijakaumaa seuraavin parametrein:

$$\mathrm{E}(X) = \mu_X = +2 \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = 4$$

$$\mathrm{E}(Y) = \mu_Y = -10 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2 = 9$$

$$\mathrm{Cor}(X,Y) = \rho_{XY} = -0.9$$

Määrää

$$\mathrm{Cov}(X, Y)$$

Tehtävä 6.7. – Mitä opimme?

Tehtävässä tarkastellaan 2-*ulotteisen normaalijakauman* ominaisuuksia.

Tehtävä 6.7. – Ratkaisu:

Määritelmän mukaan satunnaismuuttujien X ja Y *korrelaatiokerroin* on

$$\rho_{XY} = \mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{D}(X)\,\mathrm{D}(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

Siten

$$\sigma_{XY} = \mathrm{Cov}(X,Y) = \mathrm{Cor}(X,Y)\,\mathrm{D}(X)\,\mathrm{D}(Y) = \rho_{XY}\sigma_X\sigma_Y$$

Tässä

$$\mathrm{D}(X) = \sigma_X = 2$$
$$\mathrm{D}(Y) = \sigma_Y = 3$$
$$\mathrm{Cor}(X,Y) = \rho_{XY} = -0.9$$

joten

$$\mathrm{Cov}(X,Y) = \rho_{XY}\sigma_X\sigma_Y = -0.9\times 2\times 3 = -5.4$$

Tehtävä 6.8.

Oletetaan, että satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa kaksiulotteista normaalijakaumaa. Oletetaan lisäksi, että satunnaismuuttujan X regressiofunktio satunnaismuuttujan Y suhteen on

$$\mathrm{E}(X \mid Y = y) = -\frac{1}{5}y - \frac{4}{5}$$

satunnaismuuttujan Y regressiofunktio satunnaismuuttujan X suhteen on

$$\mathrm{E}(Y \mid X = x) = -\frac{5}{4}x - \frac{1}{4}$$

Määrää:

(a) Satunnaismuuttujien X ja Y odotusarvot.

(b) Satunnaismuuttujien X ja Y korrelaatio.

Tehtävä 6.8. – Mitä opimme?

Tehtävässä tarkastellaan 2-*ulotteisen normaalijakauman* ja sen *regressiofunktioiden* ominaisuuksia.

Tehtävä 6.8. – Ratkaisu:

Kaksiulotteisen normaalijakauman tapauksessa satunnaismuuttujan X *regressiofunktio* satunnaismuuttujan Y suhteen on muotoa

$$x = \mu_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

ja satunnaismuuttujan Y *regressiofunktio* satunnaismuuttujan X suhteen on muotoa

$$y = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

jossa

$$\mu_X = \mathrm{E}(X) \qquad \sigma_X^2 = \mathrm{D}^2(X) = \mathrm{Var}(X)$$
$$\mu_Y = \mathrm{E}(Y) \qquad \sigma_Y^2 = \mathrm{D}^2(Y) = \mathrm{Var}(Y)$$
$$\rho_{XY} = \mathrm{Cor}(X,Y)$$

Koska kaksiulotteisen normaalijakauman regressiofunktiot ovat *lineaarisia* ehtomuuttujien arvojen suhteen niitä kutsutaan *regressiosuoriksi*.

(a) Regressiosuorien yhtälöistä näkyy, että kumpikin suora kulkee satunnaismuuttujien X ja Y yhteisjakauman todennäköisyysmassan painopisteen

$$(\mu_X, \mu_Y)$$

kautta. Siten satunnaismuuttujien X ja Y odotusarvot saadaan määräämällä suorien leikkauspiste.

Siten leikkauspiste saadaan määräämällä lineaarisen yhtälöryhmän

$$\begin{cases} x = -\dfrac{1}{5}y - \dfrac{4}{5} \\ y = -\dfrac{5}{4}x - \dfrac{1}{4} \end{cases}$$

ratkaisu.

Sijoittamalla ensimmäinen yhtälö toiseen yhtälöön saadaan yhtälö

$$y = -\frac{5}{4}\left(-\frac{1}{5}y - \frac{4}{5}\right) - \frac{1}{4} = \frac{1}{4}y + \frac{3}{4}$$

josta saadaan ratkaisuksi

$$y = 1$$

Sijoittamalla tämä ensimmäiseen yhtälöön saadaan

$$x = -\frac{1}{5}\times 1 - \frac{4}{5} = -1$$

Siten satunnaismuuttujien X ja Y *odotusarvot* ovat

$$\mathrm{E}(X) = -1$$

$$\mathrm{E}(Y) = +1$$

(b) Kaksiulotteisen normaalijakauman regressiosuorien yleisistä lausekkeista näkyy, että suorien *kulmakertoimet*

$$\rho_{XY}\frac{\sigma_X}{\sigma_Y} \text{ ja } \rho_{XY}\frac{\sigma_Y}{\sigma_X}$$

toteuttavat yhtälön

$$\left(\rho_{XY}\frac{\sigma_X}{\sigma_Y}\right)\times\left(\rho_{XY}\frac{\sigma_Y}{\sigma_X}\right) = \rho_{XY}^2$$

Siten

$$\rho_{XY}^2 = \left(-\frac{1}{5}\right)\times\left(-\frac{5}{4}\right) = \frac{1}{4}$$

joten satunnaismuuttujien X ja Y *korrelaatiokerroin* on

$$\rho_{XY} = -\frac{1}{2}$$

koska regressiosuorien kulmakertoimet ovat *negatiivisia*.

Tehtävä 6.9.

Oletetaan, että satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa kaksiulotteista normaalijakaumaa seuraavin parametrein:

$\mathrm{E}(X) = 1 \qquad \mathrm{E}(Y) = -1$

$\mathrm{Var}(X) = \mathrm{D}^2(X) = 9 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = 4$

$\mathrm{Cor}(X, Y) = 0.5$

(a) Määrää muuttujien X ja Y kovarianssi.

(b) Määrää muuttujan X regressiofunktio muuttujan Y suhteen ja muuttujan Y regressiofunktio muuttujan X suhteen sekä vastaavat ehdolliset varianssit.

(c) Määrää kohdan (b) regressiofunktioita vastaavien suorien leikkauspiste ja vertaa sitä muuttujien X ja Y odotusarvojen vastaavaan pisteeseen. Mitä havaitset?

Tehtävä 6.9. – Mitä opimme?

Tehtävässä tarkastellaan 2-*ulotteisen normaalijakauman* ja sen *regressiofunktioiden* ominaisuuksia.

Tehtävä 6.9. – Ratkaisu:

Oletuksen mukaan

$$\mathrm{E}(X) = \mu_X = +1 \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = 9$$

$$\mathrm{E}(Y) = \mu_Y = -1 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2 = 4$$

$$\mathrm{Cor}(X,Y) = \rho_{XY} = 0.5$$

Lasketaan ensin muuttujien X ja Y *standardipoikkeamat*:

$\mathrm{D}(X) = 3$

$\mathrm{D}(Y) = 2$

(a) Satunnaismuuttujien X ja Y *kovarianssi*:

$$\mathrm{Cov}(X, Y) = \mathrm{Cor}(X, Y)\times\mathrm{D}(X)\times\mathrm{D}(Y) = 0.5\times3\times2 = 3$$

(b) Satunnaismuuttujan X *regressiofunktio* satunnaismuuttujan Y suhteen on suora, jonka kulmakerroin on

$$\mathrm{Cor}(X, Y)\times\mathrm{D}(X)/\mathrm{D}(Y) = 0.5\times3/2 = 0.75$$

Koska regressiosuora kulkee kaksiulotteisen normaalijakauman todennäköisyysmassan painopisteen

$$(\mathrm{E}(X), \mathrm{E}(Y)) = (+1,-1)$$

kautta, suoran yhtälö on muotoa

$$x - 1 = 0.75\times(y + 1)$$

Satunnaismuuttujan X *ehdollinen varianssi* satunnaismuuttujan Y suhteen on

$$(1 - \mathrm{Cor}(X, Y)^2)\times \mathrm{D}^2(X) = (1 - 0.5^2)\times 9 = 6.75 < \mathrm{D}^2(X)$$

Satunnaismuuttujan Y *regressiofunktio* satunnaismuuttujan X suhteen on suora, jonka kulmakerroin on

$$\mathrm{Cor}(X, Y)\times \mathrm{D}(Y)/\mathrm{D}(X) = 0.5\times 2/3 = 1/3$$

Koska regressiosuora kulkee kaksiulotteisen normaalijakauman todennäköisyysmassan painopisteen

$$(\mathrm{E}(X), \mathrm{E}(Y)) = (+1,-1)$$

kautta, suoran yhtälö on muotoa

$$y + 1 = (1/3)\times(x - 1)$$

Satunnaismuuttujan Y *ehdollinen varianssi* satunnaismuuttujan X suhteen on

$$(1 - \mathrm{Cor}(X, Y)^2)\times \mathrm{D}^2(Y) = (1 - 0.5^2)\times 4 = 3 < \mathrm{D}^2(Y)$$

(c) Koska kumpikin regressiosuorista kulkee kaksiulotteisen normaalijakauman todennäköisyysmassan painopisteen kautta, suorien leikkauspisteenä on piste

$$(\mathrm{E}(X), \mathrm{E}(Y)) = (+1,-1)$$

Tehtävä 6.10.

(a) Oletetaan, että satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa kaksiulotteista normaalijakaumaa parametrein

$$\mathrm{E}(X) = 0 \qquad \mathrm{E}(Y) = 1$$

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = 1 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = 4$$

$$\mathrm{Cov}(X, Y) = -1$$

Määrää muuttujien X ja Y korrelaatio ja muuttujan X regressiofunktio muuttujan Y suhteen ja muuttujan Y regressiofunktio muuttujan X suhteen sekä vastaavat ehdolliset varianssit.

(b) Oletetaan, että satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa kaksiulotteista normaalijakaumaa.

Olkoon satunnaismuuttujan X regressiosuora satunnaismuuttujan Y suhteen

$$y = -\frac{8}{3}x + \frac{14}{3}$$

ja satunnaismuuttujan Y regressiosuora satunnaismuuttujan X suhteen

$$y = -\frac{3}{2}x + \frac{7}{2}$$

Määrää muuttujien X ja Y odotusarvot.

Tehtävä 6.10. – Mitä opimme?

Tehtävässä tarkastellaan 2-*ulotteisen normaalijakauman* ja sen *regressiofunktioiden* ominaisuuksia.

Tehtävä 6.10. – Ratkaisu:

Oletuksen mukaan

$$\mathrm{E}(X) = \mu_X = 0 \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = 1$$
$$\mathrm{E}(Y) = \mu_Y = 1 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2 = 4$$
$$\mathrm{Cov}(X, Y) = \rho_{XY} = -1$$

Lasketaan ensin muuttujien X ja Y *standardipoikkeamat*:

$\mathrm{D}(X) = 1$

$\mathrm{D}(Y) = 2$

(a) Satunnaismuuttujien X ja Y korrelaatio:

$\mathrm{Cor}(X, Y) = \mathrm{Cov}(X, Y)/(\mathrm{D}(X)\times\mathrm{D}(Y)) = -1/(1\times2) = -0.5$

Satunnaismuuttujan X *regressiofunktio* satunnaismuuttujan Y suhteen on suora, jonka kulmakerroin on

$\mathrm{Cor}(X, Y)\times\mathrm{D}(X)/\mathrm{D}(Y) = -0.5\times1/2 = -1/4$

Koska regressiosuora kulkee kaksiulotteisen normaalijakauman todennäköisyysmassan painopisteen

$(\mathrm{E}(X), \mathrm{E}(Y)) = (0,1)$

kautta, suoran yhtälö on muotoa

$x = (-1/4)\times(y - 1)$

Vastaava *ehdollinen varianssi*:

$(1 - \mathrm{Cor}(X, Y)^2)\times\mathrm{D}^2(X) = (1 - (-0.5)^2)\times1 = 0.75 < \mathrm{D}^2(X)$

Satunnaismuuttujan Y *regressiofunktio* satunnaismuuttujan X suhteen on *suora*, jonka kulmakerroin on

$\mathrm{Cor}(X, Y)\times\mathrm{D}(Y)/\mathrm{D}(X) = -0.5\times2/1 = -1$

Koska regressiosuora kulkee kaksiulotteisen normaalijakauman todennäköisyysmassan painopisteen

$(\mathrm{E}(X), \mathrm{E}(Y)) = (0,1)$

kautta, suoran yhtälö on muotoa

$y - 1 = -x$

Vastaava *ehdollinen varianssi*:

$(1 - \mathrm{Cor}(X, Y)^2) \times \mathrm{D}^2(Y) = (1 - (-0.5)^2) \times 4 = 3 < \mathrm{D}^2(Y)$

(b) Suorat leikkaavat pisteessä (1,2), joten

$\mathrm{E}(X) = 1$

$\mathrm{E}(Y) = 2$

Tehtävä 6.11.

Oletetaan, että satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa kaksiulotteista normaalijakaumaa seuraavin parametrein:

$\mathrm{E}(X) = -10 \qquad \mathrm{E}(Y) = 5$

$\mathrm{Var}(X) = \mathrm{D}^2(X) = 16 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = 9$

$\mathrm{Cov}(X, Y) = 6$

(a) Määrää muuttujien X ja Y korrelaatio.

(b) Määrää muuttujan X regressiofunktio muuttujan Y suhteen ja muuttujan Y regressiofunktio muuttujan X suhteen sekä vastaavat ehdolliset varianssit.

(c) Kerro (b)-kohdassa määräämiesi regressiosuorien kulmakertoimet keskenään ja vertaa saatua tulosta (a)-kohdassa määräämäsi korrelaatiokertoimen neliöön. Mitä havaitset?

(d) Määrää kohdan (b) regressiofunktioita vastaavien suorien leikkauspiste ja vertaa sitä muuttujien X ja Y odotusarvojen vastaavaan pisteeseen. Mitä havaitset?

Tehtävä 6.11. – Mitä opimme?

Tehtävässä tarkastellaan 2-*ulotteisen normaalijakauman* ja sen *regressiofunktioiden* ominaisuuksia.

Tehtävä 6.11. – Ratkaisu:

Oletuksen mukaan

$$\mathrm{E}(X) = \mu_X = -10 \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = 16$$
$$\mathrm{E}(Y) = \mu_Y = 5 \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2 = 9$$
$$\mathrm{Cov}(X, Y) = \sigma_{XY} = 6$$

Lasketaan ensin muuttujien X ja Y *standardipoikkeamat*:

$\mathrm{D}(X) = 4$

$\mathrm{D}(Y) = 3$

(a) Muuttujien X ja Y *korrelaatio*:

$\mathrm{Cor}(X, Y) = \mathrm{Cov}(X, Y)/(\mathrm{D}(X)\times\mathrm{D}(Y)) = 6/(4\times3) = 1/2 = 0.5$

(b) Satunnaismuuttujan X *regressiofunktio* satunnaismuuttujan Y suhteen on *suora*, jonka kulmakerroin on

$\mathrm{Cor}(X, Y)\times\mathrm{D}(X)/\mathrm{D}(Y) = 0.5\times4/3 = 2/3 \approx 0.667$

Koska regressiosuora kulkee kaksiulotteisen normaalijakauman todennäköisyysmassan painopisteen

$(\mathrm{E}(X), \mathrm{E}(Y)) = (0,1)$

kautta, suoran yhtälö on muotoa

$x + 10 = (2/3)\times(y - 5)$

Satunnaismuuttujan X *ehdollinen varianssi* satunnaismuuttujan Y suhteen:

$(1 - \mathrm{Cor}(X, Y)^2)\times\mathrm{D}^2(X) = (1 - 0.5^2)\times16 = 12 < \mathrm{D}^2(X)$

Satunnaismuuttujan Y *regressiofunktio* satunnaismuuttujan X suhteen on *suora*, jonka kulmakerroin on

$\mathrm{Cor}(X, Y)\times\mathrm{D}(Y)/\mathrm{D}(X) = 0.5\times3/4 = 3/8 = 0.375$

Koska regressiosuora kulkee kaksiulotteisen normaalijakauman todennäköisyysmassan painopisteen

$(\mathrm{E}(X), \mathrm{E}(Y)) = (0,1)$

kautta, suoran yhtälö on muotoa

$y - 5 = (3/8)\times(x + 10)$

Satunnaismuuttujan Y *ehdollinen varianssi* muuttujan X suhteen:

$(1 - \mathrm{Cor}(X, Y)^2)\times\mathrm{D}^2(Y) = (1 - 0.5^2)\times9 = 27/4 = 6.75 < \mathrm{D}^2(Y)$

(c) Muuttujan X *regressiosuorassa* muuttujan Y suhteen kulmakerroin on

$$\mathrm{Cor}(X,Y)\times\frac{\mathrm{D}(X)}{\mathrm{D}(Y)}$$

Muuttujan Y *regressiosuorassa* muuttujan X suhteen kulmakerroin on

$$\mathrm{Cor}(X,Y)\times\frac{\mathrm{D}(Y)}{\mathrm{D}(X)}$$

Siten kulmakertoimien tuloksi saadaan

$$\mathrm{Cor}(X,Y)\times\frac{\mathrm{D}(X)}{\mathrm{D}(Y)}\times\mathrm{Cor}(X,Y)\times\frac{\mathrm{D}(Y)}{\mathrm{D}(X)}=[\mathrm{Cor}(X,Y)]^2$$

Tehtävässä

$$\mathrm{Cor}(X,Y)\times\frac{\mathrm{D}(X)}{\mathrm{D}(Y)}\times\mathrm{Cor}(X,Y)\times\frac{\mathrm{D}(Y)}{\mathrm{D}(X)}=\frac{2}{3}\times\frac{3}{8}=\frac{1}{4}=\left[\frac{1}{2}\right]^2=[\mathrm{Cor}(X,Y)]^2$$