

Sovellettu todennäköisyyslaskenta B

Antti Rasila

10. lokakuuta 2007

- 1 Kaksiulotteiset jakaumat, jatkoa
- 2 Reunajakaumien varianssit
- 3 Kovarianssi
 - Kovarianssiin liittyviä laskukaavoja
 - Korrelaatiokerroin
- 4 Ehdollinen jakauma
 - Ehdollinen odotusarvo
- 5 Moniulotteisia jakaumia

Esimerkki (Tuominen) 1/2

- Katkaistaan jana ympimähkään kahdesta kohdasta toisitaan riippumatta. Millä todennäköisyydellä janasta voidaan muodostaa kolmio?
- Oletusten nojalla

$$X \perp\!\!\!\perp Y, \quad X \sim \text{Tas}(0, 1), \quad Y \sim \text{Tas}(0, 1).$$

- Parin (X, Y) jakauma on siten tasainen yksikköneliössä

$$A(x, y) = \{(x, y) \mid 0 < x < 1, \quad 0 < y < 1\}.$$

- Koska tilanne on symmetrinen, voidaan tarkastella tapausta $X < Y$.
- Tällöin kolmion sivut ovat X , $Y - X$ ja $1 - Y$.

- Kolmio syntyy täsmälleen silloin, kun kahden palasen pituuksien summa on suurempi kuin kolmannen pituus. Siis

$$X < 1/2, Y > 1/2 \text{ ja } Y < X + 1/2.$$

- Vastaavasti tapaus $Y \geq X$ saadaan symmetrisesti. Merkitään suotuisten tapausten joukkoa B :llä.
- Siis

$$Pr((X, Y) \in B) = \frac{m(A \cap B)}{m(A)} = \frac{1}{4},$$

missä $m(\cdot)$ tarkoittaa ko. joukon pinta-alaa.

- Olkoon (X, Y) satunnaismuuttujapari, jonka yhteisjakauman pistetodennäköisyys- tai tiheysfunktioita merkitään $f_{XY}(x, y)$
- Diskreetissä tapauksessa:

$$\text{Var}(X) = E(X^2) - (E(X))^2,$$

missä $E(X^2) = \sum_x x^2 f_X(x)$

- Jatkuvassa tapauksessa:

$$\text{Var}(X) = E(X^2) - (E(X))^2,$$

missä $E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$

- Kovarianssi kuvaa kahden satunnaismuuttujan yhteisvaihtelua niiden yhteisjakauman painopisteen (μ_X, μ_Y) ympärillä.

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

- Laskukaavat diskreeteille ja jatkuville jakaumille:

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

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

- $\text{Cov}(X, X) = \text{Var}(X)$ ja $\text{Cov}(Y, Y) = \text{Var}(Y)$.

Esimerkki (Tuominen) 1/2

- Olkoon satunnaismuuttujaparilla (X, Y) tiheysfunktiona

$$f_{XY}(x, y) = \begin{cases} 8xy, & \text{kun } (x, y) \in A, \\ 0, & \text{muulloin,} \end{cases}$$

missä

$$A = \{(x, y) \in \mathbb{R}^2 \mid 0 < x < 1 \text{ ja } 0 < y < x\}.$$

- Tällöin

$$f_X(x) = \int_0^x 8xy \, dy = 4x^3, \text{ kun } x \in (0, 1),$$

$$f_Y(y) = \int_y^1 8xy \, dx = 4y(1 - y^2), \text{ kun } y \in (0, 1).$$

- Huomaa, että tässä tapauksessa X ja Y eivät ole riippumattomia, koska

$$f_{XY}(x, y) \neq f_X(x)f_Y(y).$$

Esimerkki (Tuominen) 2/2

- Reunatiheysfunktioiden avulla saadaan

$$E(X) = \int_0^1 x \cdot 4x^3 dx = \frac{4}{5},$$

$$E(Y) = \int_0^1 y \cdot 4y(1 - y^2) dy = \frac{8}{15}.$$

- Valitsemalla $g(X, Y) = XY$ voidaan laskea

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^x xy \cdot 8xy dx dy = \int_0^1 8x^2 \left(\int_0^x y^2 dy \right) dx \\ &= \int_0^1 8x^2 \cdot \frac{x^3}{3} dx = \frac{4}{9}. \end{aligned}$$

- Tästä saadaan X :n ja Y :n kovarianssi

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{4}{225}.$$

- Summan ja erotuksen varianssi yleisessä tapauksessa

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

- Olkoot (vakiot $a, b, c, d \in \mathbb{R}$)

$$W = a + bX$$

$$Z = c + dY$$

Tällöin

$$\text{Cov}(W, Z) = bd\text{Cov}(X, Y) = bd\sigma_{XY}$$

- Jos $X \perp Y$, niin $\text{Cov}(X, Y) = 0$, mutta käänteinen ei aina päde.

- Korrelaatiokerroin kuvaa kahden satunnaismuuttujan **lineaarisen** riippuvuuden voimakkuutta.

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

- Korrelaatiokertoimen ominaisuudet:
 - 1 $-1 \leq \text{Cor}(X, Y) \leq 1$
 - 2 Jos $X \perp\!\!\!\perp Y$, niin $\text{Cor}(X, Y) = 0$, mutta käänteinen ei aina päde.
 - 3 $\text{Cor}(X, Y) = \pm 1$ jos ja vain jos $Y = \alpha + \beta X$, missä vakiot $\alpha, \beta \in \mathbb{R}$ ja $\beta \neq 0$

- Korrelaatiokerroin voi olla 0 vaikka X :n ja Y :n välillä vallitsisi funktionaalinen riippuvuus.
- Jos esimerkiksi

$$X \sim N(0, 1), \quad Y = X^2,$$

niin $\text{Cov}(X, Y) = E(X^3) = 0$, koska funktio $f(x) = x^3$ on symmetrinen origon suhteen.

- Siten myös $\text{Corr}(X, Y) = 0$.

- Satunnaismuuttujan X ehdollinen jakauma satunnaismuuttujan Y suhteen eli ehdolla Y on

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

- Kahdella riippumattomalle satunnaismuuttujalla on seuraava ominaisuus

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

- Ehdollinen odotusarvo satunnaismuuttujalle X satunnaismuuttujan Y suhteen on diskreetissä tapauksessa

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

ja jatkuvassa tapauksessa

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

- Määritellään funktio $u: S_y \rightarrow \mathbb{R}$ kaavalla

$$u(y) = E(X|Y = y), \text{ kun } y \in S_y.$$

- Satunnaismuuttujaa u kutsutaan X :n odotusarvoksi ehdolla Y ja merkitään $E(X|Y)$.

Esimerkki (Niemi; Leon-Garcia) 1/2

- Oletetaan, että palvelupisteeseen saapuvien asiakkaiden määrä X aikavälillä $[0, t]$ on Poisson-jakautunut: $X \sim \text{Poisson}(\lambda_1 t)$.
- Oletetaan lisäksi, että asiakkaan palveluaika T on eksponenttijakautunut: $T \sim \text{Exp}(\lambda_2)$.
- Olkoon Y palvelupisteeseen saapuvien uusien asiakkaiden lukumäärä edellisen asiakkaan palvelun kestäessä. Oletetaan, että uusien asiakkaiden saapuminen ei riipu palveluajasta.
- Saadaan

$$\begin{aligned} \Pr(Y = k) &= \int_0^{\infty} \Pr(Y = k | T = t) \lambda_2 e^{-\lambda_2 t} dt \\ &= \int_0^{\infty} \frac{(\lambda_1 t)^k}{k!} e^{-\lambda_1 t} \lambda_2 e^{-\lambda_2 t} dt = \frac{\lambda_1^k}{\lambda_2 k!} \int_0^{\infty} t^k e^{-(\lambda_1 + \lambda_2)t} dt. \end{aligned}$$

- Sijoittamalla $r = (\lambda_1 + \lambda)t$ saadaan

$$\begin{aligned}\frac{\lambda_1^k \lambda_2}{k!} \int_0^\infty t^k e^{-(\lambda_1 + \lambda_2)t} dt &= \frac{\lambda_1^k \lambda_2}{k!(\lambda_1 + \lambda_2)^{k+1}} \int_0^\infty r^k e^{-r} dr \\ &= \frac{\lambda_1^k \lambda_2}{k!(\lambda_1 + \lambda_2)^{k+1}} \Gamma(k + 1).\end{aligned}$$

- Kun muistetaan, että $\Gamma(k + 1) = k!$, saadaan edelleen

$$\frac{\lambda_1^k \lambda_2}{k!(\lambda_1 + \lambda_2)^{k+1}} \Gamma(k + 1) = \frac{\lambda_1^k \lambda_2}{(\lambda_1 + \lambda_2)^{k+1}}.$$

- Satunnaismuuttuja Y on siis geometrisesti jakautunut.

Esimerkki (Tuominen) 1/2

- Olkoot $U \perp V$ ja $U, V \sim \text{Tas}(0, 1)$. Määritellään

$$X = \max\{U, V\}, \quad Y = \min\{U, V\}.$$

- Parille (X, Y) voidaan johtaa tiheysfunktio

$$f_{XY}(x, y) = \begin{cases} 2, & \text{kun } (x, y) \in A, \\ 0, & \text{muuten,} \end{cases}$$

missä

$$A = \{(x, y) \in \mathbb{R}^2 \mid 0 < x < 1 \text{ ja } 0 < y < x\}.$$

- Reunatiheysfunktioiksi saadaan tässä tapauksessa

$$f_X(x) = 2x, \quad x \in (0, 1),$$

$$f_Y(y) = 2(1 - y), \quad y \in (0, 1).$$

- Ehdolliset tiheysfunktiot ovat siten

$$f_{X|Y}(x|y) = \frac{1}{1-y}, \quad x \in (y, 1),$$

$$f_{Y|X}(y|x) = \frac{1}{x}, \quad y \in (0, x),$$

kun $x \in (0, 1)$ ja $y \in (0, 1)$.

- Kyseiset jakaumat ovat tasaisia.
- Ehdolliset odotusarvot ovat

$$E(X|Y = y) = \frac{y+1}{2}, \quad E(Y|X = x) = \frac{x}{2}.$$

- Multinomijakauma (esimerkkinä kolmiulotteinen):

$$(X_1, X_2, X_3) \sim \text{Multinom}(p_1, p_2, p_3, n)$$

$$f(x_1, x_2, x_3) = \binom{n}{x_1 \ x_2 \ x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3} = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

- Kaksiulotteinen normaalijakauma

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

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