

MVE051/MSG810 2016 Föreläsning 4

Petter Mostad

Chalmers

November 13, 2016

Multivariata (simultana) fördelningar

- ▶ De (diskreta och kontinuerliga) stokastiska variabler vi pratat om så långt är *univariata*: Det är bara ett enkelt värde som är osäkert.
- ▶ Mera nyttiga är *multivariata* modeller: Här är det flera reella värden som är osäkra, och som beror på varann.
- ▶ Teorin i detta kapitel rör sig om *bivariata* modeller, där vi har två värden som är osäkra, men teorin kan lätt generaliseras.

Diskreta simultana fördelningar

- ▶ En diskret bivariat fördelning kan specificeras genom en frekvensfunktion som för varje par av reella tal (x, y) ger ett reellt tal $f_{XY}(x, y)$ som uppfyller
 - ▶ $f_{XY}(x, y) \geq 0$.
 - ▶ $\sum_{\text{alla } x} \sum_{\text{alla } y} f_{XY}(x, y) = 1$.

Vi har då att $\Pr[X = x \text{ och } Y = y] = f_{XY}(x, y)$.

- ▶ Exempel: En tabell som anger $f_{XY}(x, y)$ för varje kombination av ett ändligt antal x och ett ändligt antal y , (medan $f_{XY}(x, y) = 0$ för alla andra par).
- ▶ Vi definierar *marginalfördelningarna* f_X och f_Y som

$$f_X(x) = \sum_{\text{alla } y} f_{XY}(x, y) = \Pr[X = x]$$

$$f_Y(y) = \sum_{\text{alla } x} f_{XY}(x, y) = \Pr[Y = y]$$

Märk att detta är frekvensfunktioner.

Kontinuerliga simultana fördelningar

- ▶ En kontinuerlig bivariat fördelning kan specificeras genom en täthetsfunktion som för varje par av reella tal (x, y) ger ett reellt tal $f_{XY}(x, y)$ som uppfyller
 - ▶ $f_{XY}(x, y) \geq 0$.
 - ▶ $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$.

Vi har då att $\Pr[X \in [a, b] \text{ och } Y \in [c, d]] = \int_a^b \int_c^d f_{XY}(x, y) dx dy$.

- ▶ Exempel: $f_{XY}(x, y) = 2xy + 1/2$ för $0 \leq x \leq 1$ och $0 \leq y \leq 1$.
- ▶ Vi definierar *marginalfördelningarna* f_X och f_Y som

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dx \end{aligned}$$

Märk att detta är täthetsfunktioner.

Oberoende stokastiska variabler

- ▶ Om A och B är *händelser* så är A och B oberoende om $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$.
- ▶ Två diskreta stokastiska variabler X och Y är oberoende om, för *alla* x och y ,

$$\Pr[X = x \text{ och } Y = y] = \Pr[X = x] \cdot \Pr[Y = y]$$

- ▶ Detta är det samma som att, för alla x och y ,

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

- ▶ Även för kontinuerliga stokastiska variabler gäller definitionen X och Y är oberoende om

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

- ▶ Om (X, Y) är en bivariat stokastisk variabel så definierar vi väntevärdet $\mathbb{E}[H(X, Y)]$ med

$$\mathbb{E}[H(X, Y)] = \sum_{\text{alla } x} \sum_{\text{alla } y} H(x, y) f_{XY}(x, y)$$

om (X, Y) är diskret och, om (X, Y) är kontinuerlig,

$$\mathbb{E}[H(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) f_{XY}(x, y) dx dy.$$

- ▶ Vi får räkneregler som $\mathbb{E}[cX] = c \mathbb{E}[X]$ och $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- ▶ OM X och Y ÄR OBEROENDE, så får vi

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y],$$

annars gäller inte detta alltid.

- ▶ Om (X, Y) är en bivariat stokastisk variabel så definierar vi *kovariansen* med

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- ▶ Med räknereglerna får vi lätt att

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- ▶ Vi ser att om X och Y är oberoende så är kovariansen noll; motsatsen gäller inte.
- ▶ Kovariansen mäter "samvariationen" mellan X och Y , och kan i princip vara vilket reellt tal som helst.
- ▶ Några räkneregler: $\text{Cov}[cX, Y] = c \text{Cov}[X, Y]$ och $\text{Cov}[X, cY] = c \text{Cov}[X, Y]$. Också: $\text{Cov}[X, Y] = \text{Cov}[Y, X]$.

- ▶ Korrelationen mellan X och Y defineras som

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}} = \frac{\text{Cov}[X, Y]}{\text{Sd}[X] \text{Sd}[Y]}$$

- ▶ Korrelationen mäter också "samvariationen", men är alltid mellan -1 och 1.
- ▶ Vi har $|\rho_{XY}| = 1$ om och enbart om $Y = \beta_0 + \beta_1 X$ för reela värden β_0 och β_1 och $\beta_1 \neq 0$.
- ▶ Kom i håg att korrelationen (enbart) mäter linjära beroenden. Positiv korrelation betyder att linjen lutar uppåt, negativ korrelation att den lutar neråt.

Betingade fördelningar

- ▶ För händelser A och B har vi $\Pr[A | B] = \frac{\Pr[A \cap B]}{\Pr[B]}$ (om $\Pr B > 0$).
- ▶ För diskreta variabler X och Y definierar vi den betingade fördelningen $f_{X|Y}$ med (om $f_Y(y) > 0$)

$$f_{X|Y}(x) = \Pr[X = x | Y = y] = \frac{\Pr[X = x \text{ och } Y = y]}{\Pr[Y = y]} = \frac{f_{XY}(x, y)}{f_Y(y)}$$

- ▶ Märk att detta är en frekvensfunktion, och att $f_{X|Y}(x)f_Y(y) = f_{XY}(x, y)$.
- ▶ Även för kontinuerliga variabler X och Y definierar vi

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

Variabelbyten för multivariata fördelningar

Antag (X, Y) har en bivariat kontinuerlig fördelning med täthetsfunktion $f_{XY}(x, y)$, och att (U, V) har en bivariat kontinuerlig fördelning med täthetsfunktion $f_{UV}(u, v)$. Antag att

$$(X, Y) = (h_1(U, V), h_2(U, V))$$

definerar en en-till-en transformation från området där f_{UV} är positiv till området där f_{XY} är positiv, och att båda h_1 och h_2 har kontinuerliga partiella derivater. Då har vi att

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J|$$

där J är *Jacobi-matrisen* definerad med

$$J = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

och där $|J| \neq 0$ alltså är determinanten av denna matrisen.