

TMS136

Föreläsning 5

Två eller flera stokastiska variabler

- I många situationer är det av intresse att betrakta fler än en s.v. åt gången
- Speciellt gör man det ”i statistik” där man nästan alltid jobbar med stickprov
- Med ett stickprov menar vi här inte observerade data utan en uppsättning oberoende (definieras lite längre fram) och likafördelade s.v.

Till att börja med

- I vissa situationer är man intresserade av hur två s.v. beror av varandra
- Till exempel kan man vara intresserad av att prediktera ett aktiepris utifrån ett index

Två diskreta

- För två diskreta s.v. X och Y kan vi definiera en *gemensam massfunktion (joint probability mass function)* $f_{XY}(x, y)$ som är sådan att

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

2. $\sum_{\text{alla } x} \sum_{\text{alla } y} f_{XY}(x, y) = 1$

3. $f_{XY}(x, y) = P(X = x, Y = y)$

Två kontinuerliga

- För två kontinuerliga s.v. X och Y kan vi definiera en *gemensam täthetsfunktion (joint probability density function)* $f_{XY}(x, y)$ som är sådan att

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

2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$

3. $P((X, Y) \in B) = \iint_B f_{XY}(x, y) dx dy$

Exempel

- Låt X vara tiden (i ms) tills din dator får kontakt med en viss server och Y vara tiden (i ms) (från dess du begärde kontakt med servern) tills servern ”släpper in” dig.
- Det är klart att $X < Y$
- Baserat på historiska data har vi kommit fram till att

$$f_{XY}(x, y) = 6 \times 10^{-6} \exp(-0.001x - 0.002y) \text{ för } x < y$$

Exempel fortsättning

- Observera att

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy \\ &= 6 \times 10^{-6} \int_0^{\infty} \int_x^{\infty} \exp(-0.001x - 0.002y) dy dx \\ &= 3 \times 10^{-3} \int_0^{\infty} \exp(-0.003x) dx = 1 \end{aligned}$$

Exempel fortsättning

- Vi har även att (kolla att integralerna "stämmer")

$$P(X < 1000, Y < 2000)$$

$$= 6 \times 10^{-6} \int_0^{1000} \int_x^{2000} \exp(-0.001x - 0.002y) dy dx$$

$$= 3 \times 10^{-3} \int_0^{1000} \exp(-0.003x) - \exp(-4 - 0.001x) dx$$
$$\approx 0.915$$

”Marginaler”

- Om man för en gemensam massfunktion summerar över den ena s.v.’n:s möjliga värden får man (*marginal*)massfunktionen för den andra s.v.’n

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

och

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

”Marginaler”

- Om man för en gemensam täthetsfunktion integrerar över den ena s.v.'n:s värdemängd får man *(marginal)tätheten* för den andra s.v.'n

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

och

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

Exempel fortsättning

- Om vi "serverexemplet" enbart är intresserade av $P(Y > 2000)$ kan vi få denna ur (kom ihåg att $Y > X$)

$$P(Y > 2000)$$

$$\begin{aligned} &= 6 \times 10^{-6} \int_0^{2000} \int_{2000}^{\infty} \exp(-0.001x - 0.002y) dy dx \\ &+ 6 \times 10^{-6} \int_{2000}^{\infty} \int_x^{\infty} \exp(-0.001x - 0.002y) dy dx \approx 0.05 \end{aligned}$$

Exempel fortsättning

- Vi kan också beräkna $P(Y > 2000)$ ur marginaltätheten för Y
- Marginaltätheten fås ur (kom ihåg att $X < Y$)

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dx \\ &= 6 \times 10^{-6} \int_0^y \exp(-0.001x - 0.002y) dx \\ &= 6 \times 10^{-3} \exp(-0.002y)(1 - \exp(-0.001y)) \end{aligned}$$

Exempel fortsättning

- Vi har då att

$$\begin{aligned} & P(Y > 2000) \\ &= 0.006 \int_{2000}^{\infty} \exp(-0.002y) (1 - \exp(-0.001y)) dy \\ &\approx 0.05 \end{aligned}$$

Betingade fördelningar

- För diskreta s.v. X och Y kan vi definiera *betingade massfunktioner* enligt

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

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

Betingade fördelningar

- Det gäller att

1. $f_{X|y}(x) \geq 0$

2. $\sum_{\text{alla } x} f_{X|y}(x) = 1$

- "Samma" gäller så klart även för $f_{Y|x}(y)$

Betingade fördelningar

- För kontinuerliga s.v. X och Y kan vi definiera *betingade täthetsfunktioner* enligt

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

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

Betingade fördelningar

- Det gäller att

1. $f_{X|y}(x) \geq 0$

2. $\int_{-\infty}^{\infty} f_{X|y}(x) dx = 1$

3. $P(Y \in B | X = x) = \int_B f_{X|y}(x) dx$

- "Samma" gäller så klart även för $f_{Y|x}(y)$

Exempel forts

- I "serverexemplet" har vi (kolla detta)

$$f_X(x) = 0.003\exp(-0.003x)$$

- Det följer (för $x > 0$ och $x < y$) att

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} = 0.002\exp(0.002x - 0.002y)$$

Exempel forts

- Till exempel kan vi beräkna

$$P(Y > 2000 | X = 1500)$$

$$= \int_{2000}^{\infty} f_{Y|1500}(y) dy = 0.002e^3 \int_{2000}^{\infty} e^{-0.002y} dy$$

$$\approx 0.368$$

Betingat väntevärde/varians

- Man kan definiera betingat väntevärde enligt

$$\mu_{Y|x} = E(Y|x) = \begin{cases} \int_{-\infty}^{\infty} y f_{Y|x}(y) dy \\ \sum_{\text{alla } y} y f_{Y|x}(y) \end{cases}$$

Betingat väntevärde/variants

- Och betingad variants enligt

$$\sigma^2_{Y|x} = V(Y|x) = \begin{cases} \int_{-\infty}^{\infty} (y - \mu_{Y|x})^2 f_{Y|x}(y) dy \\ \sum_{\text{alla } y} (y - \mu_{Y|x})^2 f_{Y|x}(y) \end{cases}$$

Exempel forts.

- I "serverexemplet" kan vi till exempel beräkna (minns att $Y > X$ och använd partiell integration)

$$\begin{aligned} E(Y|X = 1500) &= 0.002e^3 \int_{1500}^{\infty} ye^{-0.002y} dy \\ &= 2000 \end{aligned}$$

Oberoende s.v.

- Vi säger att två s.v. X och Y om något (och därmed alla) av följande ekvivalenta samband gäller

$$1. f_{XY}(x, y) = f_X(x)f_Y(y)$$

$$2. f_{X|Y}(x) = f_X(x)$$

$$3. f_{Y|X}(y) = f_Y(y)$$

$$4. P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Exempel

- Antag att X och Y är längden respektive diametern för en viss typ av skruv
- Antag vidare att X och Y är oberoende, X är normalfördelad med väntevärde 10.5 mm och varians 0.0025 (mm)^2 och Y är normalfördelad med väntevärde 3.2 mm och varians 0.0036 (mm)^2
- Beräkna $P(10.4 < X < 10.6, 3.15 < Y < 3.25)$

Exempel forts

- Vi har då att

$$\begin{aligned} & P(10.4 < X < 10.6, 3.15 < Y < 3.25) \\ &= P(10.4 < X < 10.6)P(3.15 < Y < 3.25) \\ &= P\left(\frac{10.4 - 10.5}{0.05} < Z < \frac{10.6 - 10.5}{0.05}\right) \\ &\times P\left(\frac{3.15 - 3.2}{0.06} < Z < \frac{3.25 - 3.2}{0.06}\right) \\ &\approx P(-2 < Z < 2)P(-0.83 < Z < 0.83) \\ &= (2\Phi(2) - 1)(2\Phi(0.83) - 1) \approx 0.57 \end{aligned}$$

Fler än två s.v.

- Man kan generalisera allt som gäller för två s.v. till att gälla flera s.v.
- Till exempel kan man definiera multivariata tätheter och massfunktioner analogt med det vi sett ovan
- Marginalerna (för en eller flera s.v.) fås genom att integrera/summera över den/de s.v. "man vill ha bort"

Fler än två s.v.

- Om mass-/täthetsfunktionen för $X_1, \dots, X_p$ är $f_{X_1 \dots X_p}(x_1, \dots, x_p)$ gäller att $X_1, \dots, X_p$ är oberoende om och endast om

$$f_{X_1 \dots X_p}(x_1, \dots, x_p) = f_{X_1}(x_1) \cdots f_{X_p}(x_p)$$

för alla $x_1, \dots, x_p$