Föreläsning 6 sammanfattning

Litteratur

Från kursboken: 5.3-5.6

Väntevärde för en funktion av flerdimensionell s.v.

(X, Y) är en tvådimensionell s.v. och $g(X, Y)$ är en reellvärd funktion. Då gäller

$$E(g(X,Y)) = \begin{cases} \sum_j \sum_k g(j,k) p_{X,Y}(j,k) & \text{diskreta s.v.} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) & \text{kontinuerliga s.v.} \end{cases}$$

Intressanta fall är t.ex.

$$g(X,Y) = X + Y,\ g(X,Y) = X \text{ samt } g(X,Y) = XY$$

Linjäritet av väntevärde

Då X och Y är s.v. med väntevärden $E(X)$ och $E(Y)$ samt $a,\ ,b$ och c är konstanter så gäller att

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

Detta samband är helt generellt dvs gäller vid *godtyckliga beroende mellan X och Y.*

X och Y oberoende

Om X och Y är oberoende s.v. så gäller att

$$E(XY) = E(X)E(Y).$$

Detta utvidgas till fler än två oberoende s.v.

Beroendemått

Man kan kvantifiera beroendet mellan två s.v. X och Y med ett beroendemått. Beroendemåttet beskriver graden av *linjärt* beroende/samband.

Om en tvådimensionell s.v. (X, Y) har ändliga väntevärden μ_X och μ_Y samt standardavvikelser $D(X) = \sigma_X$ och $D(Y) = \sigma_Y$ så uttrycks *kovariansen* mellan X och Y som

$$C(X, Y) = E((X - \mu_X)(Y - \mu_Y)).$$

Kovariansen beskriver hur förändringar för den ena s.v. är linjärt associerade till den andra s.v.

Räkneregler för kovarianser

X, Y och Z är s.v., a, b, $, c$ och d är konstanter. Då gäller

$$\begin{cases} C(X, X) = V(X), \quad C(X, Y) = C(Y, X), \\ C(aX + b, cY + d) = acC(X; Y), \\ C(X + Y, Z) = C(X, Z) + C(Y, Z). \end{cases}$$

Korrelationskoefficienten $\rho(X, Y)$

Korrelationskoefficienten antar värden i intervallet $[-1, 0]$ och beräknar styrkan av relationen för hur variablerna rör sig. Den definieras enligt

$$\rho(X, Y) = \frac{C(X, Y)}{D(X)D(Y)} = \frac{C(X, Y)}{\sigma_X \sigma_Y}$$

Korrelationskoefficienten är ytterligare dimensionslös, m.a.o. en punktskattning.

Beräkning av kovarians

Kovariansen kan beräknas enligt följande

$$C(X, Y) = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y.$$

Ifall X och Y är oberoend så är de naturligtvis även okorrelerade.
Bevis:

$$E(XY) = E(X)E(Y) \Rightarrow C(X, Y) = 0$$

Men det är inte säkert att okorrelerade s.v. är oberoende!

Linjärkombination

Kovariansen är viktig för att förstå *variation* i summor av s.v. Sätt

$$a_1X_1 + a_2X_2 + ... + a_nX_n + b$$

för n s.v. $X_1, ..., X_n$ och konstanterna $a_1, a_2, ..., a_n, b$.

Väntevärdet fås enligt

$$E(\sum_{i=1}^{n} a_iX_i + b) = \sum_{i=1}^{n} a_iE(X_i) + b,$$

Och variansen

$$V(\sum_{i=1}^{n} a_iX_i + b) = \sum_{i=1}^{n} a_i^2V(X_i) + 2\sum_{1\leq j<k\leq n} a_ja_kC(X_j, X_k).$$

För oberoende s.v. så gäller

$$E(\sum_{i=1}^{n} a_iX_i + b) = \sum_{i=1}^{n} a_iE(X_i) + b,$$

$$V(\sum_{i=1}^{n} a_iX_i + b) = \sum_{i=1}^{n} a_i^2V(X_i).$$

Om alla s.v. är oberoende var och en med väntevärdet μ och standardavvikelsen σ och

$$\bar{X} = \frac{\sum_{i=1}^{n}(X_1 + ... + X_n)}{n}$$

så gäller att

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \sigma^2/n, \quad D(\bar{X} = \sigma/\sqrt{n}).$$

Stora talens lag

En av de viktigaste resultaten inom sannolikhetsteorin. Den utsäger att det *aritmetiska medelvärdet av flera oberoende s.v. med samma väntevärde μ ligger nära μ,* bara antalet är tillräckligt stort.

Definition

$X_1, X_2, ...$ är en följd oberoende och likafördelade s.v. med samma väntevärde $E(X_i) = \mu$ och standardavvikelse $D(X_i) = \sigma < \infty$. Låt

$$\bar{X} = \frac{1}{n}(X_1 + ... + X_n)$$

vara medelvärdet av de n första variablerna. Då gäller för godtyckligt $\epsilon > 0$ att

$$P(|\bar{X} - \mu| > \epsilon) \rightarrow 0 \text{ då } n \rightarrow \infty.$$

I ord säger uttrycket att en s.v. $\bar{X}_n$ som uttrycker medelvärdet av de n stycken s.v. har en fördelning som koncentrerar sig runt $\mu = E(X_i)$.

En tolkning till detta resultat är att medelvärde är en bra uppskattning av väntevärdet.

Stora talens lag är en av grundstenarna inom empirisk vetenskap. Rent praktiskt så kan man säga att om man genomför oberoende observationer av någon s.v. som exemeplvis livslängden hos personer i ett försäkringskollektiv så kommer medelvärdena av observationerna att ligga nära det *sanna* väntevärdet. Man kan alltså uppskatta det sanna väntevärdet genom att skatta väntevärdet med medelvärdet.