

Väntevärden, korrelation och stora talens lag

Målet för denna föreläsning är stora talens lag som visar en viktig egenskap för summor av många oberoende s.v. Vi börjar med att undersöka väntevärde och variansen för en summa $X+Y$ av två s.v. För en reellvärd funktion $g(X,Y)$ av en tvådimensionell s.v. definieras väntevärdet som

$$E(g(X,Y)) = \sum_{j,k} g(j,k) p_{X,Y}(j,k), \quad \text{om } (X,Y) \text{ är en diskret s.v.}$$

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy, \quad \text{om } (X,Y) \text{ är en kont. s.v.}$$

så i specialfallet $g(X,Y) = X+Y$ gäller

$$\begin{aligned} E(X+Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f_{X,Y}(x,y) dx dy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y) dx \right) dy \\ &= E(X) + E(Y) \end{aligned}$$

om (X,Y) är en kontinuerlig s.v. och samma formel kan visat gälla även om (X,Y) är en diskret s.v. (se boken). Formeln generaliseras enkelt till $E(aX+bY+c) = aE(X)+bE(Y)+c$, så väntevärde är en linjär funktion. Från linjäriteten för väntevärde följer följande sats.

Sats: $V(aX+b) = a^2 V(X)$

$D(aX+b) = |a| D(X)$

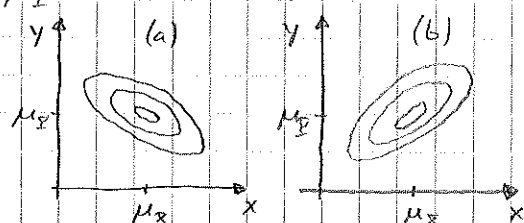
Bevis: Sätt $E(X) = \mu$. Då gäller $E(aX+b) = aE(X)+b = a\mu+b$ och vidare

$$V(aX+b) = E((aX+b - (a\mu+b))^2) = E((aX - a\mu)^2) = E(a^2(X-\mu)^2)$$

För att kunna utveckla variansen av $aX+bY+c$ $= a^2 E((X-\mu)^2) = a^2 V(X)$ på samma sätt som vi gjorde för väntevärdet, så behöver vi veta hur variationen hos X och Y hänger ihop.

Betrakta avvikelserna av X och Y från sina respektive väntevärden, dvs $X-\mu_X$ och $Y-\mu_Y$. Om $X-\mu_X$ och $Y-\mu_Y$ tenderar att ha samma tecken, så sägs de vara positivt korrelerade och om de tenderar att ha olika tecken så sägs de vara negativt korrelerade. Vi kvantifierar detta mha kovariansen.

Def: Kovariansen mellan de s.v. X och Y ges av $C(X,Y) = E((X-\mu_X)(Y-\mu_Y))$, där $\mu_X = E(X)$ och $\mu_Y = E(Y)$.



Nivåkurvor för $f_{X,Y}(x,y)$ vid (a) negativ respektive (b) positiv korrelation.

I boken visas att $C(X,Y) = E(XY) - E(X)E(Y)$ med ett bevis som är analogt med beviset av $V(X) = E(X^2) - (E(X))^2$. Ett ännu vanligare mått på korrelationen mellan två s.v. X och Y är följande:

Def: Korrelationskoefficienten mellan X och Y är $\rho(X,Y) = \frac{C(X,Y)}{D(X)D(Y)}$

Det går att visa att $-1 \leq \rho(X,Y) \leq 1$ samt att $\rho(X,Y) = 1$ om $Y = aX+b$ med $a > 0$ och att $\rho(X,Y) = -1$ om $Y = aX+b$ med $a < 0$ (se boken). Maximal korrelation uppträder alltså då X och Y är linjärt beroende.

Def: Om $C(X,Y) = 0$, så sägs de s.v. X och Y vara okorrelerade.

Notera att för okorrelerade s.v. så följer även att $\rho(X,Y) = 0$. Okorrelerade s.v. måste inte vara oberoende (se exempel 5.13 i boken), men vi ska visa att oberoende s.v. måste vara okorrelerade. För att visa detta behöver vi först en hjälpsats.

Sats: Om de s.v. X och Y är oberoende, så gäller $E(XY) = E(X)E(Y)$
(Satsen kan generaliseras till $E(X_1 \dots X_n) = E(X_1) \dots E(X_n)$ för obero. s.v. $X_1, \dots, X_n$).

Bevis: Om X och Y är kontinuerliga s.v. så är (för diskreta fallet, se boken)

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x,y) dx dy = \int_{-\infty}^{\infty} x f_X(x) \left(\int_{-\infty}^{\infty} y f_Y(y) dy \right) dx = E(X)E(Y)$$

$= f_X(x) f_Y(y)$

Sats: Om de s.v. X och Y är oberoende, så är de okorrelerade

Bevis: Enligt satsen ovan är kovariansen $C(X,Y) = E(XY) - E(X)E(Y) = 0$
 $= E(X)E(Y)$ ty X & Y är oberoende.

Med hjälp av kovariansen kan vi nu utveckla $V(aX+bY+c)$

Sats: $V(aX+bY+c) = a^2 V(X) + b^2 V(Y) + 2ab C(X,Y)$ (Tredje termen försvinner)
(om X och Y är oberoende.)

Bevis: Sätt $\mu_X = E(X)$ och $\mu_Y = E(Y)$. Då gäller:

$$V(aX+bY+c) = E((aX+bY+c) - (a\mu_X+b\mu_Y+c))^2 = E((a(X-\mu_X)+b(Y-\mu_Y)))^2$$

$$= E(a^2(X-\mu_X)^2 + b^2(Y-\mu_Y)^2 + 2ab(X-\mu_X)(Y-\mu_Y)) = a^2 V(X) + b^2 V(Y) + 2ab C(X,Y)$$

Satserna om väntevärde och varians för $aX+bY+c$ kan lätt generaliseras till linjärkombinationer av typen $a_1 X_1 + a_2 X_2 + \dots + a_n X_n + a_0$, där $X_1, \dots, X_n$ är s.v. och $a_0, a_1, \dots, a_n$ är konstanter. Vi gör detta utan bevis.

Sats: $E\left(\sum_{i=1}^n a_i X_i + a_0\right) = \sum_{i=1}^n a_i E(X_i) + a_0$, $V\left(\sum_{i=1}^n a_i X_i + a_0\right) = \sum_{i=1}^n a_i^2 V(X_i)$

I specialfallet $a_0=0$ och $a_i=1/n$, för $i=1, \dots, n$ fås:

Sats: Låt $X_1, \dots, X_n$ vara oberoende s.v. med väntevärde μ och standardavvikelse σ . Då uppfyller det aritmetiska medelvärdet $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ att $E(\bar{X}) = \mu$ och $V(\bar{X}) = \sigma^2/n$

$+ 2 \sum_{1 \leq i < j \leq n} a_i a_j C(X_i, X_j)$
(Sista termen försvinner)
(om X_i och X_j är oberoende.)

BILD 6 → Bilden visar täthetsfunktionen för medelvärdet av 1, 10, 100 och 1000 $N(1,1)$ -fördelade s.v. Som synes koncentreras fördelningen för medelvärdet alltmer kring väntevärdet ett (ty variansen går mot noll). Detta förklarar varför vi i statistikdelen av kursen kommer att använda det aritmetiska medelvärdet av en datamängd som skattning av väntevärdet för datans fördelning. Stora talens lag nedan, först formulerad av Bernoulli på 1690-talet, preciserar hur medelvärdet konvergerar mot väntevärdet.

Sats: För varje $\varepsilon > 0$, så gäller det att $IP(|\bar{X} - \mu| > \varepsilon) \rightarrow 0$ då $n \rightarrow \infty$.

Bevis: Vi bevisar stora talens lag med hjälp av Markovs olikhet. För kontinuerliga s.v. $Y \geq 0$, så gäller

$$E(Y) = \int_0^{\infty} y f_Y(y) dy = \int_0^a y f_Y(y) dy + \int_a^{\infty} y f_Y(y) dy \geq a \int_a^{\infty} f_Y(y) dy = a IP(Y \geq a)$$

≥ 0 $\geq a$

så $IP(Y \geq a) \leq E(Y)/a$, vilket kallas Markovs olikhet. Olikheten ger en övre gräns för sannolikheten för stora värden på Y . Vi sätter in $Y = (\bar{X} - \mu)^2 \geq 0$ och $a = \varepsilon^2$ i Markovs olikhet och får:

$$IP(|\bar{X} - \mu| > \varepsilon) = IP((\bar{X} - \mu)^2 \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} E((\bar{X} - \mu)^2) = \frac{1}{\varepsilon^2} \text{Var}(\bar{X}) = \frac{1}{\varepsilon^2} \frac{\sigma^2}{n}$$

$\uparrow$ Markovs olikhet $\uparrow$ $E(\bar{X})$

vilket för varje val av ε och σ går mot noll då $n \rightarrow \infty$.