Sanstat 14/9 - 2011

Kovarians

Vi har sagt.

Väntevärdet är <u>linjärt</u>.

$$E\left(\sum_i a_i X_i\right) = \sum a_i E(X_i)$$

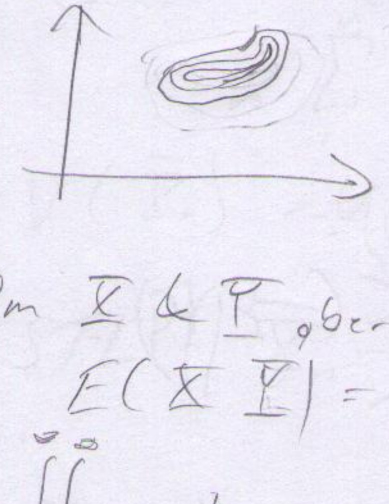
<u>Def</u>: $C(X,Y) = E[(X - E(X))(Y - E(Y))] =$

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

Mäter <u>linjärt</u> beroende

<u>Korelation</u>: $\rho(X,Y) = \dfrac{C(X,Y)}{D(X)\,D(Y)}$

Det gäller $|\rho(\ )| \le 1$



<u>OBS!</u> $X \in N(0,1)$ & $Y = X^2$

har $C(X,Y) = 0$

Om $X$ & $Y$ oberoende så

$E(XY) = \iint_{-\infty}^{\infty} f_{XY}(xy)\,dxdy$

$= \iint_{-\infty}^{\infty} xy\, f_X(x)\, f_Y(y)\,dxdy = \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy$

$= E(x)E(y)$

Alltså $C(X,Y) = 0$

Slutsats $(X,Y)$ oberoende $\Rightarrow$ $(X,Y)$ oberoende
ingen korelation

Omvändning <u>falsk</u>

Räkneregler för kovarians

(1) $V(X) = C(X, X)$ $\quad E(X)$ etc

(2) $C(X+Y, Z) = E[(X+Y-(\mu_X+\mu_Y))(Z-\mu_Z)]$

$= E[(X-\mu_X)(Z-\mu_Z) + (Y-\mu_Y)(Z-\mu_Z)] =$

$= C(X,Z) + C(Y,Z)$

(3) $C(X,Y) = C(Y,X)$

Alltså Kovariansen är bilinjär.

$$C\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_i \sum_j a_i b_j C(X_i, Y_j)$$

Tänk: Multiplikation av parantesuttryck

Ex: $(x+2y)(6x-4y) = 6x\,x - 4xy + 26xy - 24yy$

$C(X+2Y, 6X-4Y) = 6C(X,X) - 4C(Y,X) + 26C(X,Y)$
$- 24C(X,Y)$

Ex: $V(X+Y) = C(X+Y, X+Y) =$

$= C(X,X) + C(X,Y) + C(Y,X) + C(Y,Y) =$

$= V(X) + V(Y) + 2C(X,Y)$

Notera: Om $X$ & $Y$ är oberoende eller okorrelerade

$V(X+Y) = V(X) + V(Y)$

Allmänt: Varians för summa av obur/ekorr

= summa av varianserna

Notera också: $V(aX) = C(aX, aX) =$

$= aa \cdot C(X, X) = a^2 V(X)$

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

Ex: $X, Y$ oberoende eller okorrelerade $\Rightarrow$

$V(X - Y) = V(X + (-Y)) = V(X) + V(-Y) =$

$= V(X) + (-1)^2 V(Y) = V(X) + V(Y)$

Stora talens lag

$X_1, X_2, \ldots$ oberoende stokastiska variabler med samma fördelning, v.v. $\mu$, varians $\sigma^2$

Låt $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$

Då gäller $E(\bar{X}_n) = E\left(\frac{1}{n}\sum_1^n X_k\right)$

$= \frac{1}{n}\sum_1^n E(X_n) = \mu$

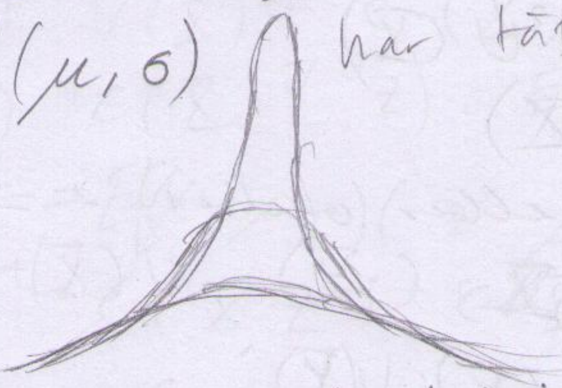
$V(\bar{X}_n) = V\left(\frac{1}{n}\sum_1^n X_k\right) = \frac{1}{n^2}\sum_1^n V(X_k) = \sigma^2/n$

STL $P(|\bar{X}_n - \mu| > \varepsilon) \to 0$ då $n \to \infty$

för alla $\varepsilon > 0$

# Normalfördelning (Gausfördelning)

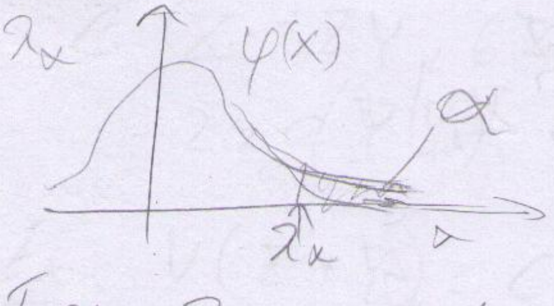
$N(\mu, \sigma)$



har tätheten: $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$

För Standard N-fördelning $(\mu = 0, \sigma = 1)$
kallas täthet $\varphi$ och fördelfun $\Phi$

Om $X \in N(\mu, \sigma)$
$P(a \leq X \leq b)$ ges av normalcdf för
Kvantilen betecknas $\lambda_\alpha$




T.ex $\lambda_{0,05} = 1,96$

Om $X$ är normal fördelad så är $aX$
$aX + b$ också Normalfördelad
(med v.v $aE(x)+b$, var $a^2 V(X)$)

Normalfördelningar är en lägesskal-familj
inte alltid sant

Visa genom att sätta på fördelningsfunktionen

$$F_{aX+b}(x) = P(aX + b \le x) = P\left(X \le \frac{x-b}{a}\right)$$

$$F_X\left(\frac{x-b}{a}\right), \text{ så } f_{aX+b}(x) = \frac{1}{a} f_X\left(\frac{x-b}{a}\right)$$

Speciellt: om $X \in N(\mu, \sigma)$ så är

$$(X - \mu)/\sigma \in N(0,1)$$

Alltså:

$$P(X \le c) = P\Big(\underbrace{\frac{X-\mu}{\sigma}}_{\in N(0,1)} \le \frac{c-\mu}{\sigma}\Big) = \Phi\left(\frac{c-\mu}{\sigma}\right)$$

Sats: En linjärkombination av ober. N-fördelade s.v. är N-fördelad

Centrala gränsvärdessatsen

$X_1, X_2, \ldots$ oberoende lika-fördelade s.v. med v.v. $\mu$, varians $\sigma^2$

Då gäller $\sum_{k=1}^{n} X_k \in N(n\mu, \sqrt{n}\sigma)$

Pas för alla x

$$P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}\right) = P\left(\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \le x\right) \to \underline{\underline{\Phi(x)}}$$