NTNU

Kunnskap for ei betre verd

Varians og kovarians

Torstein Fjeldstad

Institutt for matematiske fag, NTNU

31.01.2019

1. Repetisjon
2. Varians og kovarians (kap 4.2)

Repetisjon I

Forventningsverdi til stokastisk variabel

La X vere ein stokastisk variabel med fordeling $f(x)$. Forventningen til X er då

$$E(X) = \mu \quad = \sum_{x} x f(x) \qquad X \text{ diskret}$$

$$E(X) = \mu \quad = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \qquad X \text{ kontinuerleg}$$

Repetisjon II

Forventningsverdi til funksjon av stokastisk variabel

La X vere ein SV med fordeling $f(x)$. Forventningsverdien til den stokastiske variabelen $g(X)$ er

$$E(g(X)) = \mu_{g(X)} = \sum_x g(x)f(x) \qquad X \text{ diskret}$$

$$E(g(X)) = \mu_{g(X)} = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x \qquad X \text{ kontinuerleg}$$

Repetisjon III

Forventningsverdi til sum/differanse av funksjonar av stokastisk variabel

Forventningsverdien til summen eller differansen av to funksjonar av den stokastiske variabelen X med fordeling $f(x)$ er:

$$E(g_1(X) \pm g_2(X)) = E(g_1(X)) \pm E(g_2(X)).$$

Merk: $E(aX + b) = aE(X) + b$.

Repetisjon IV

Forventningsverdi til funksjon av stokastiske variable

La X og Y vere stokastiske variable med simultanfordeling $f(x, y)$. La $g(X, Y)$ vere ein vilkårleg funksjon av X og Y. Då er:

$$E(g(X, Y)) = \sum_x \sum_y g(x, y) f(x, y) \qquad X, Y \text{ diskret}$$

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) \mathrm{d}x \mathrm{d}y \qquad X, Y \text{ kontinuerleg}$$

Repetisjon V

Forventningsverdi til sum/differanse av funksjonar av stokastiske variabele

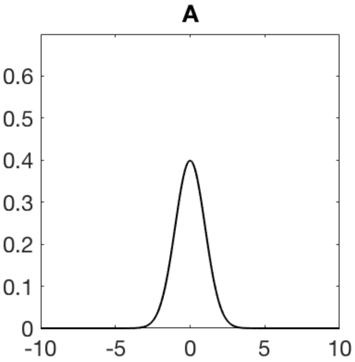
Forventningsverdien til summen eller differansen av to funksjonar av dei stokastiske variabelane X og Y med simultanfordeling $f(x, y)$ er:

$$E(g_1(X, Y) \pm g_2(X, Y)) = E(g_1(X, Y)) \pm E(g_2(X, Y)).$$

Merk: $E(X + Y) = E(X) + E(Y)$.

Varians

Eksempel varians



Eksempel (prosjektstyring)

X: tid brukt på datainnsamling
Y: tid brukt på dataanalyse

$y \backslash x$	1	2	3	$h(y)$
1	0.03	0.05	0.02	0.10
2	0.03	0.14	0.03	0.20
3	0.03	0.17	0.10	0.30
4	0.01	0.24	0.15	0.40
$g(x)$	0.1	0.6	0.3	1

Varians, kovarians og korrelasjon I

Varians

La X vere ein stokastisk variabel med fordeling $f(x)$ og forventning $\mu = E(X)$. Variansen til X er

$$\sigma^2 = Var(X) = E((X-\mu)^2) = \sum_x (x-\mu)^2 f(x) \qquad X \text{ diskret}$$

$$\sigma^2 = Var(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x \qquad X \text{ kontinuerleg}$$

Den positive kvadratrota av variansen til X, $\sigma = \sqrt{Var(X)} = SD(x)$, kallast standardavviket til X.

Varians, kovarians og korrelasjon II

Rekneregel for varians

For ein stokastisk variabel X og to konstantar a og b har me

$$\sigma^2_{aX+b} = Var(aX+b) = a^2 Var(X) = a^2 \sigma^2_X .$$

Varians, kovarians og korrelasjon III

Kovarians

La X og Y vere to stokastiske variable med simultanfordeling $f(x, y)$ og forventning $\mu_X = E(X)$ og $\mu_Y = E(Y)$. Kovariansen (samvariasjonen) til X og Y er:

$$\begin{aligned}
\sigma_{XY} &= Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) \\
&= \sum_x \sum_y (x - \mu_y)(y - \mu_Y) f(x, y) \qquad X, Y \text{ diskret}
\end{aligned}$$

$$\begin{aligned}
\sigma_{XY} &= Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) \\
&= \int_{-\infty}^{\infty} \int_{\infty}^{\infty} \sum_y (x - \mu_y)(y - \mu_Y) f(x, y) \mathrm{d}x \mathrm{d}y \qquad X, Y \text{ kontinuerleg}
\end{aligned}$$

Varians, kovarians og korrelasjon IV

Rekneregel varians sum av to variable

For to stokastiske variable X og Y og to konstantar a og b har me:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(y) + 2abCov(X, Y)$$

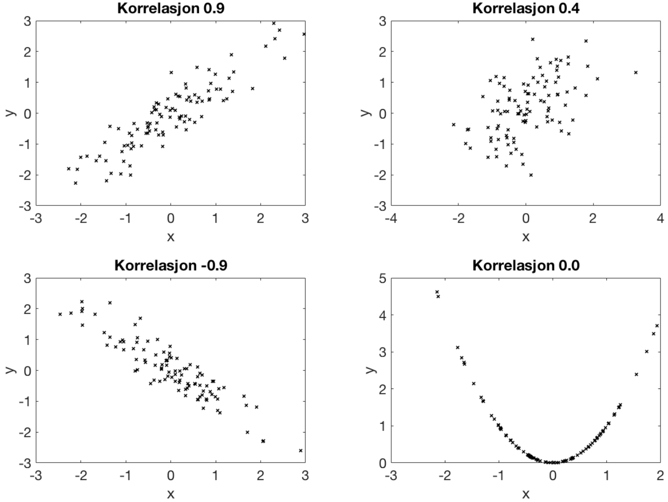
Korrelasjon

Korrelasjon

La X og Y vere to stokastiske variable med kovarians σ_{XY} og variansar σ_X^2 og σ_Y^2. Korrelasjonskoeffisienten til X og Y er då

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Eksempel korrelasjon



Neste veke

— Diskrete fordelingar