Kapittel 4: Forventning og varians

I kapitlet:

- Forventning.
- Varians og kovarians.
- Regneregler som **forenkler** regning av forventning og varians i mange tilfeller.
- Tilnærma regning ved Taylor-utvikling.
- Tsjebysjeffs ulikhet.

Hovedmål:

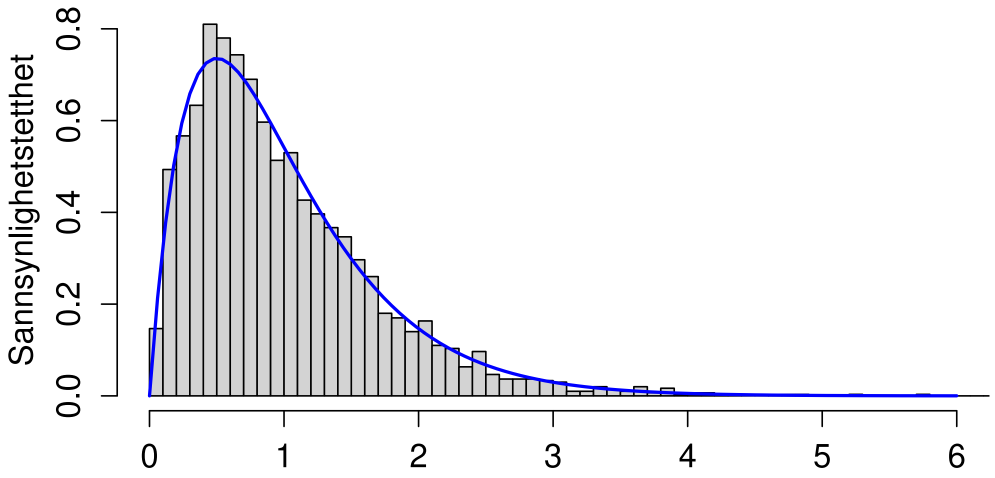
Lage sentralmål for fordelinger.

Lage spredningsmål for fordelinger.

Lage mål for sammenheng mellom to fordelinger.

Teoretisk modell eller observerte data

Teoretisk modell		**Datasett**
Stokastisk variabel	X	Observasjoner $x_1, x_2, \ldots, x_n$.
Sannsynlighetsfordeling	$f(x)$	Histogram fra relative frekvenser.
Forventningsverdi	$E(X)$	Gjennomsnitt $\overline{x}$. Sentralmål
Varians	$Var(X)$	Empirisk varians s^2. Spredningsmål
Korrelasjon	$Cor(X, Y)$	Empirisk korrelasjon (har ikke sett på denne). Sammenheng to variabler



4.1 Forventningsverdi av stokastisk variabel

Forventningsverdi (D 4.1):

Forventningsverdien (forventninga) til en **diskret** stokastisk variabel er definert ved

$$\mu = E(X) = \sum_x xP(X = x) = \sum_x xf(x)$$

„Expectation"

For **kontinuerlig** variabel:

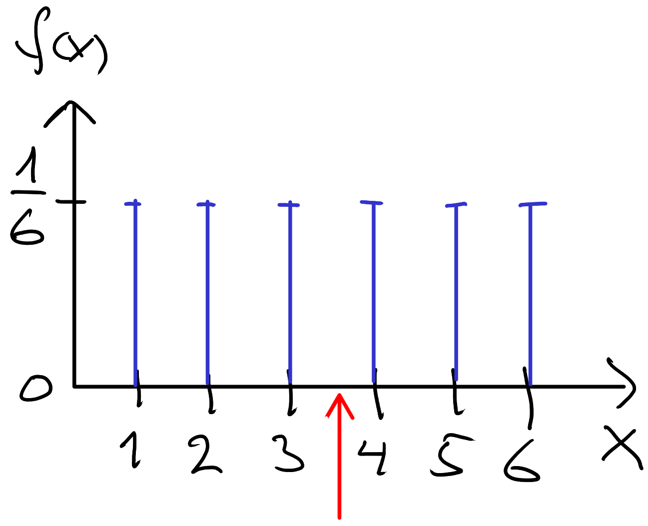
$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$

- Tyngdepunkt i fordelinga til X.
- Teoretisk gjennomsnitt for fordelinga. Tilsvarer gjennomsnittsverdi av uendelig mange forsøk ("gjennomsnitt i det lange løp").

$$E(X) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \lim_{n\to\infty} \overline{X}$$

Eksempel: Terningkast. La X være resultatet av et terningkast med en vanlig terning.

x	1	2	3	4	5	6
f(x) = P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6

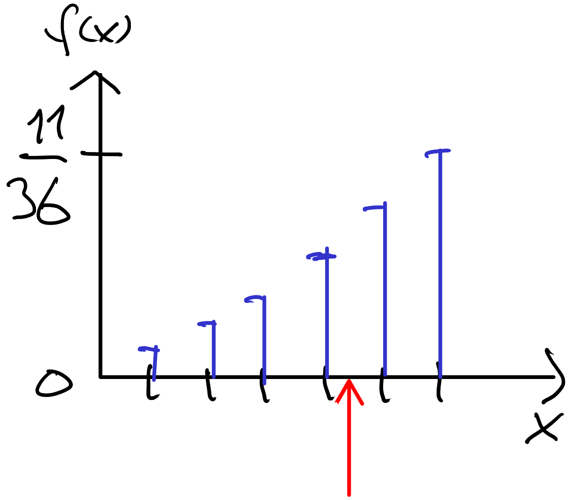


$$E(X) = \sum_x x \cdot f(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6}$$

$$= 21/6 = 3{,}5$$

Eksempel: Terningkast. La X være resultatet av et terningkast med en "vekta" terning.

x	1	2	3	4	5	6
f(x) = P(X=x)	1/36	3/36	5/36	7/36	9/36	11/36



$$E(X) = \sum_x x \cdot f(x) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + \cdots + 6 \cdot \frac{11}{36}$$

$$= \frac{131}{36} = 4{,}47$$

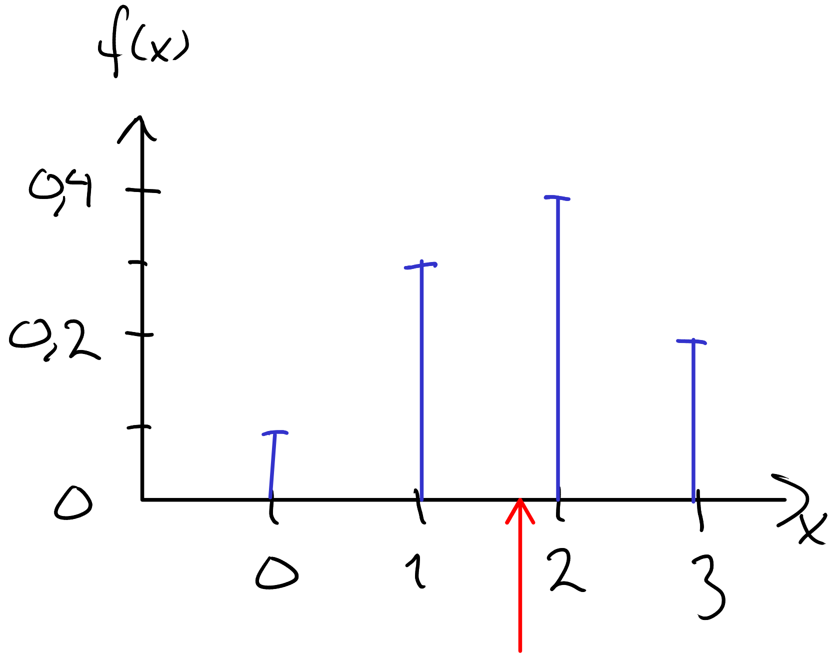
Eksempel: Valpekull. La Y være antall levende valper i kull fra en hunderase.

y	0	1	2	3
f(y) = P(Y=y)	0.1	0.3	0.4	0.2

$$E(Y) = \sum_y y \cdot f(y)$$

$$= 0 \cdot 0{,}1 + 1 \cdot 0{,}3 + 2 \cdot 0{,}4 + 3 \cdot 0{,}3$$

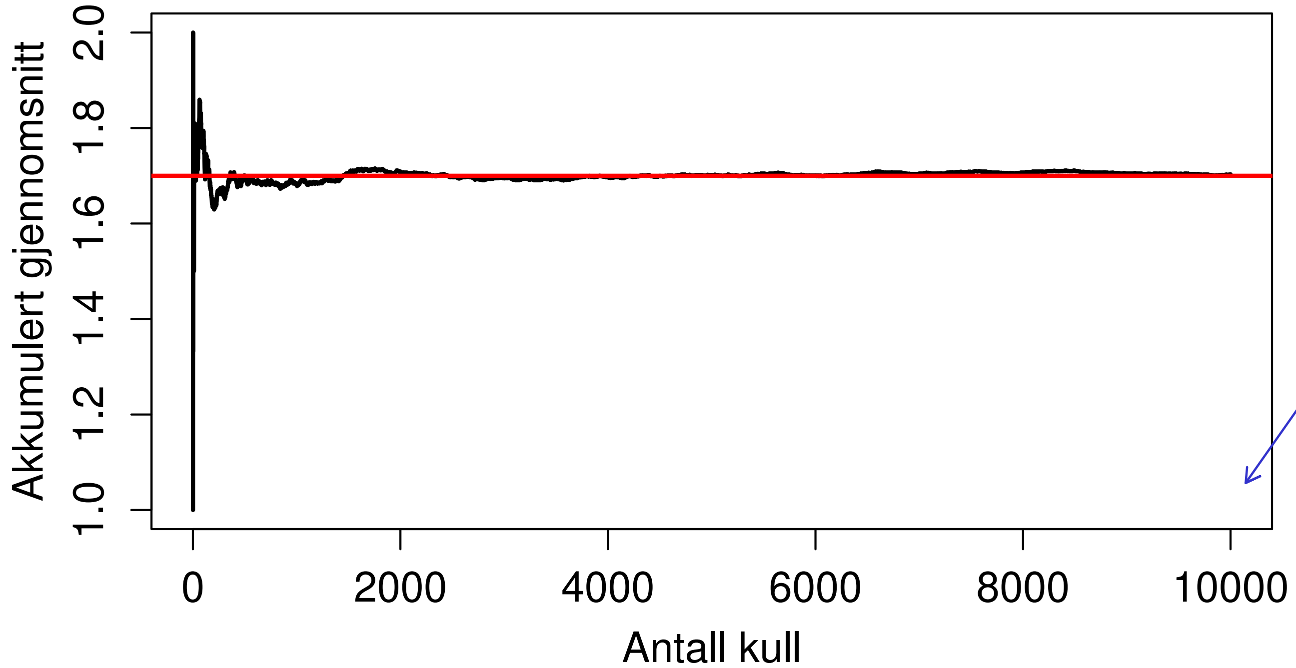
$$= \underline{\underline{1{,}7}}$$



- Har påstått at gjennomsnittet av mange forsøk vil konvergere mot forventningsverdien.
- Vil derfor gjøre et et eksperiment der vi trekker $n = 10000$ “valpekull” fra denne fordelinga, hva skjer med gjennomsnittet?

R-kode:

```
valper = c(0,1,1,1,2,2,2,2,3,3)     #`Urne' med 10 mulige resultat.
n = 10000
kull = sample(valper,n,replace=T) # Trekker 10000 ganger (med tilbakelegging).
akku.snitt = vector(length=n)  # Lager vektor til akkumulerte gjennomsnitt.
for (i in 1:n) akku.snitt[i] = mean(kull[1:i]) # Oppdaterer gjennomsnittet.
plot(akku.snitt,type="l",ylab="Akkumulert gjennomsnitt",xlab="Antall kull",lwd=2)
abline(1.7,0,col="red",lwd=2)
```



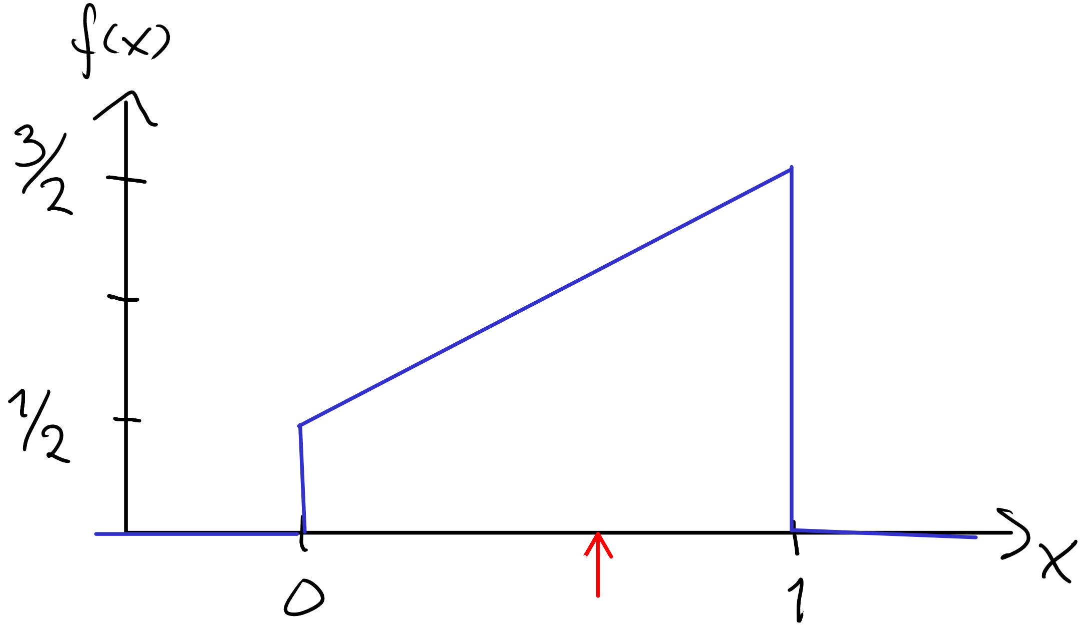
Gjennomsnitt av mange forsøk konvergerer mot forventningsverdien.

Eksempel: Metallprøve (kontinuerlig).

La X være hardhet av tilfeldig valgt metallprøve.

$$f(x) = \begin{cases} 1/2 + x, & 0 \leq x \leq 1. \\ 0, & \text{ellers.} \end{cases}$$

Figur:



$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_0^1 x \cdot \left(\frac{1}{2} + x\right) dx \quad \left[\frac{x^2}{4} + \frac{x^3}{3}\right]_0^1 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12} = 0{,}583$$

Tolkning: Om vi målen hardhet av ∞ mange metallprøven vil gjsnittet bli 0,583.

Forventningsverdi, funksjoner av stokastiske variabler

Eksempel: Valpekull forts.

La Y være antall levende valper i kull fra en hunderase.
Utgifter per kull: 4000 kroner. Salgspris for en valp: 5000 kroner.
Hva blir forventninga til fortjenesten V for et kull.

y	f(y)	v	f(v)
0	0.1	-4000	0,1
1	0.3	1000	0,3
2	0.4	6000	0,4
3	0.2	11000	0,2

$$V = \underbrace{5000 \cdot Y - 4000}_{\text{funksjon av } Y}$$

Finn først fortjenesten v og punktsannsynlighetene $f(v)$. (Se tabell)

$$E(V) = \sum_v v \cdot f(v) = (-4000)\cdot 0{,}1 + 1000 \cdot 0{,}3 + 6000 \cdot 0{,}4 + 11000 \cdot 0{,}2$$
$$= -400 + 300 + 2400 + 2200 = \underline{\underline{4500}}$$

(Gj.snittlig fortjeneste over ∞ mange kull.)

Generaliserer til veldig **viktig teorem**:

Forventningsverdi til funksjon av stokastisk variabel (T 4.1):

La X være en stokastisk variabel med sannsynlighetsfordeling $f(x)$.
Da blir forventningsverdien til en **funksjon $g(X)$**:

$$\mu_{g(X)} = E[g(X)] = \begin{cases} \sum_x g(x)\, f(x), & \text{X diskret.} \\ \int_{-\infty}^{\infty} g(x)\, f(x)\, dx, & X \text{ kontinuerlig.} \end{cases}$$

Bruker samme punkt s.s. som for X.

(OBS: $g(x)$ er her **ikke** en punktsannsynlighet/tetthetsfunksjon, men en funksjon av x.)

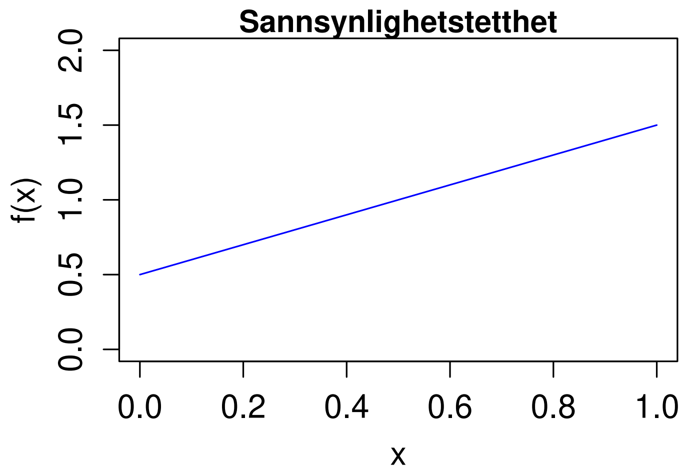
Kommentar: Om vi kjenner fordelinga til X, $f(x)$, kan vi også finne forventa verdi av enhver funksjon av X.

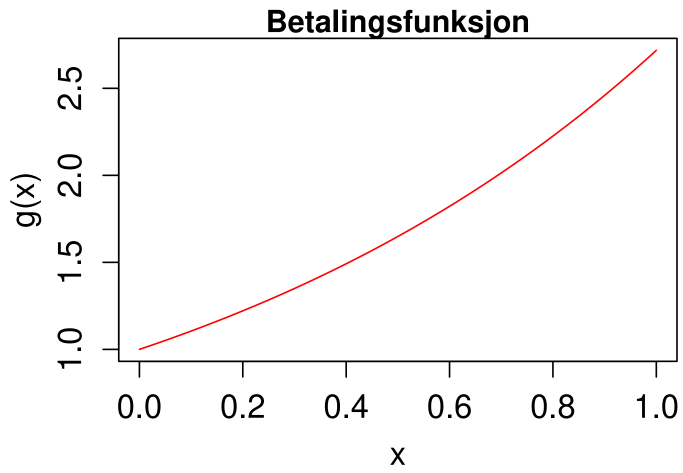
$E[2X]$	Gjennomsnitt i det lange løp for $2X$.
$E[X^2]$	Gjennomsnitt i det lange løp for X^2.
$E[e^X]$	Gjennomsnitt i det lange løp for e^X.

Eksempel: Metallprøve forts.

Anta at metallkjøpere vil ha stor hardhet, og at en produsent derfor får betalt etter følgende funksjon:

$$g(x) = e^x$$





Hva blir forventa betaling?

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,dx = \int_0^1 e^x\left(\frac{1}{2} + x\right)dx = \frac{1}{2}\int_0^1 e^x\,dx + \int_0^1 x\,e^x\,dx$$

delvis integr. $\longrightarrow = \left[x\,e^x - \frac{1}{2}e^x\right]_0^1 = \frac{1}{2}(e+1) = 1{,}86$

Eksempel: St. Petersburg-paradokset.

På et kasino tilbyr de deg et spill hvor du kaster en mynt til du får "kron" for første gang, og gevinsten du får er gitt ved

$$g(x) = 2^x$$

$2^1 = 2$
$2^2 = 4$
$2^3 = 8$
etc.

der x er antall kast til første "kron".

Hvor mye ville du ha **vært villig til å betale** for å spille spillet?
(Hint: Finn først sannsynlighetsfordelinga for antall kast, deretter forventa gevinst.)

$$\begin{aligned} f(1) &= P(\text{kron på første kast}) = P(K) = 0.5 \\ f(2) &= P(\text{kron første gang på andre kast}) = P(MK) = 0.5^2 \\ &\vdots \end{aligned}$$

Generelt: $f(x) = \left(\frac{1}{2}\right)^x, \quad x = 1, 2, \ldots$

Punkts.s. for antall kast til „Kron".

$$E[g(X)] = E\left(2^X\right) = \sum_x 2^x \cdot f(x) = \sum_{x=1}^{\infty} 2^x \cdot \left(\frac{1}{2}\right)^x$$

$$= \sum_{x=1}^{\infty} 1 = \underline{\underline{\infty}}$$

<u>Konklusjon</u>: Uansett hvor mye du betaler ser det ut til å lønne seg?

<u>Kommentar</u>: Forventa utbetaling er ∞, men er det relevant for spilleren? Det som skjer i det veldig lange løp kan være irrelevant for oss som lever i endelig tid.

Utvidelse: Funksjon av fleire variabler

Forventningsverdi til funksjon av stokastisk variabler (D 4.2):

La X og Y være være stokastiske variabler med simultan sannsynlighetsfordeling $f(x,y)$. Da blir forventningsverdien til en **funksjon $g(X,Y)$**:

$$\mu_{g(X,Y)} = E[g(X,Y)] = \begin{cases} \sum\limits_{x}\sum\limits_{y} g(x,y)\, f(x,y), & X,Y \text{ diskrete.} \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dxdy, & X,Y \text{ kontinuerlige.} \end{cases}$$

Kan utvides til fleire enn to variabler.

Eksempel: Trykkeri.

X = antall kundehenvendelser per dag. Y = antall bestillinger per dag.

		x			
		0	1	2	$P(Y=y)$
y	0	0.2	0.2	0.2	0.6
	1	0.0	0.1	0.2	0.3
	2	0.0	0.0	0.1	0.1
$P(X=x)$		0.2	0.3	0.5	1.0

- Hva er forventa differanse mellom henvendelser og bestillinger, altså $X - Y$?

$$\begin{aligned} g(x,y) &= x - y \\ E[g(X,Y)] &= E(X-Y) = \sum_{x=0}^{2}\sum_{y=0}^{2}(x-y)f(x,y) \\ &= (0-0)\cdot 0{,}2 + (0-1)\cdot 0{,}0 + \cdots + (2-2)\cdot 0{,}1 \\ &= \underline{\underline{0{,}8}} \end{aligned}$$

- Alternativ framgangsmåte? Hint: Innfør ny variabel

$$Z = X - Y$$

Finn fordelinga til Z og regn ut E(Z).

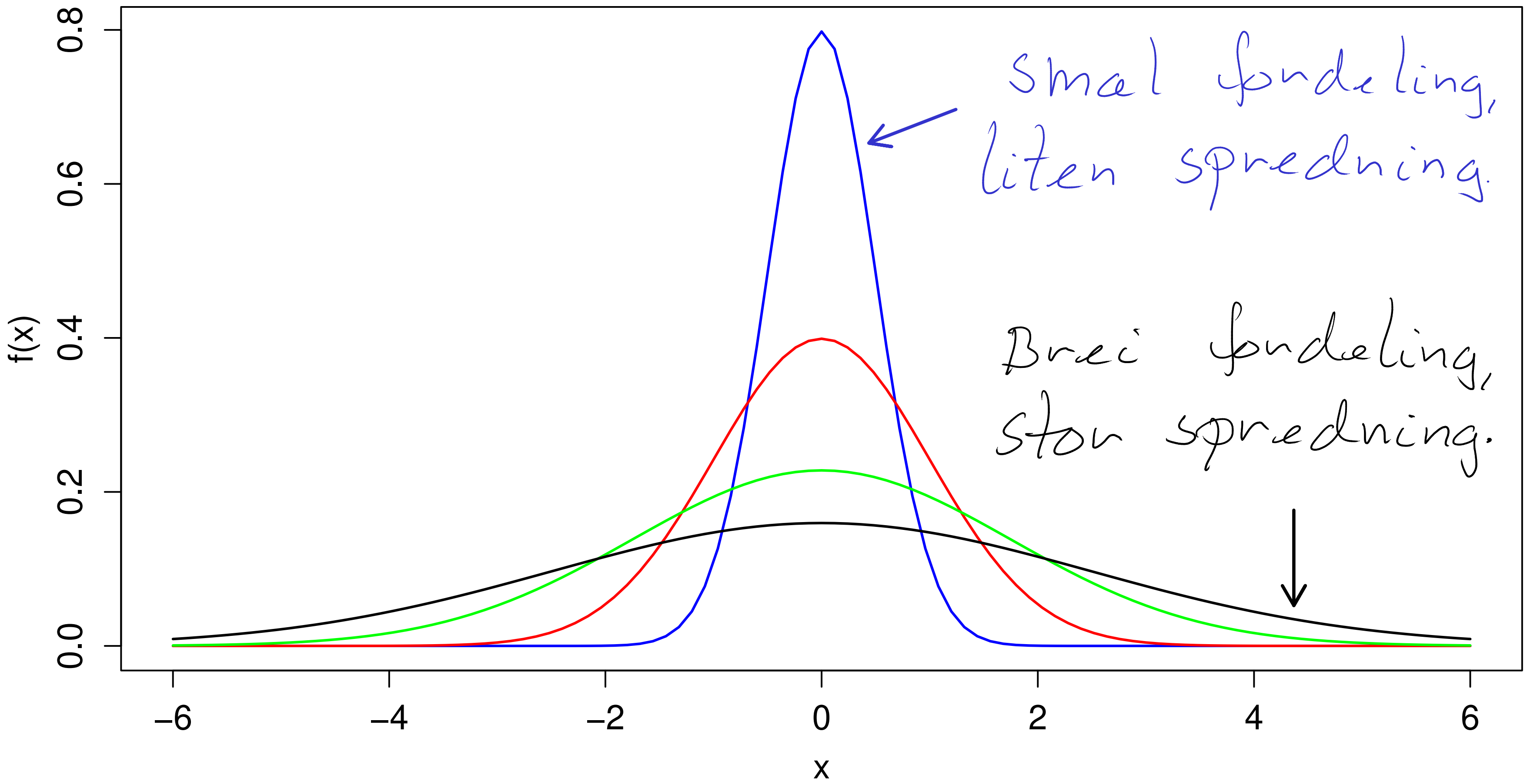
$z = x-y$	-2	-1	0	1	2
$f(z)$	0	0	$0{,}4$	$0{,}4$	$0{,}2$

$$E(Z) = \sum_z z \cdot f(z) = 1 \cdot 0{,}4 + 2 \cdot 0{,}2 = \underline{\underline{0{,}8}}$$

4.2 Varians og kovarians

Varians gir mål for spredning (i fordelinga) av en stokastisk variabel.

Eksempler på fordelinger med samme forventningsverdi (0), men ulik **spredning**:



Definisjon varians

$E(X) = \mu$

Varians og standardavvik (D 4.3):

For en stokastisk variabel X med sannsynlighetsfordeling $f(x)$ og forventningsverdi μ er **variansen** definert som forventningsverdien av funksjonen $g(X) = (X - \mu)^2$:

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 f(x), & X \text{ diskret.} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx, & X \text{ kontinuerlig.} \end{cases}$$

Standardavvik:

$$\sigma = SD(X) = \sqrt{\sigma^2}$$

Variansen er altså forventa ("gjennomsnittlig") **kvadratavvik** fra forventningsverdien μ.

Empirisk varians konvergerer mot variansen i fordelinga (blir lik ved uendelig mange forsøk).

$$Var(X) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \lim_{n\to\infty} S^2$$

Variansregning: Alternativ formel

Denne formelen er ofte enklere å bruke:

Alternativ variansformel (T 4.2):

Variansen til en stokastisk variabel X er:

$$\sigma^2 = Var(X) = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$$

Bevis: (Kontinuerlig tilfelle.)

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{\infty} (x^2 - 2x\mu + \mu^2) f(x)\,dx$$

$$= \underbrace{\int_{-\infty}^{\infty} x^2 f(x)\,dx}_{E(X^2)} - 2\mu \underbrace{\int_{-\infty}^{\infty} x f(x)\,dx}_{E(X)} + \mu^2 \underbrace{\int_{-\infty}^{\infty} f(x)\,dx}_{1}$$

$$= E(X^2) - 2\mu\mu + \mu^2 = E(X^2) - \mu^2$$

Eksempel: Valpekull forts.

Finn variansen og standardavviket til antall valper, Y, per kull.

$$Var(Y) = E(Y^2) - (E(Y))^2 = \sum_y y^2 \cdot f(y) - 1{,}7^2$$

$$= (0^2 \cdot 0{,}1 + 1^2 \cdot 0{,}3 + 2^2 \cdot 0{,}4 + 3^2 \cdot 0{,}2) - 1{,}7^2$$

$$= 3{,}7 - 1{,}7^2 = \underline{\underline{0{,}81}} \qquad SD(Y) = \sqrt{0{,}81} = \underline{\underline{0{,}9}}$$

Eksempel: Metallprøve forts.

Finn variansen i hardhet, X, av en metallprøve.

$$Var(X) = E(X^2) - (E(X))^2 = \int_0^1 x^2 \cdot \left(\frac{1}{2} + x\right) dx - \left(\frac{7}{12}\right)^2$$

$$= \left[\frac{1}{6}x^3 + \frac{1}{4}x^4\right]_0^1 - \left(\frac{7}{12}\right)^2 = \frac{1}{4} + \frac{1}{6} - \left(\frac{7}{12}\right)^2$$

$$= \underline{\underline{\frac{11}{144} = 0{,}076}} \qquad SD(X) = \sqrt{0{,}076} = \underline{\underline{0{,}276}}$$

Varians, funksjoner av stokastiske variabler

Varians til en funksjon av en stokastisk variabel (T 4.3):

La X være være en stokastisk variabel med sannsynlighetsfordeling $f(x)$. Da blir variansen til en **funksjon $\boldsymbol{g(X)}$**:

$E[g(X)]$

$$\begin{aligned}\sigma^2_{g(X)} &= Var[g(X)] = E\left\{[g(X) - \mu_{g(X)}]^2\right\} \\ &= \begin{cases} \sum_x [g(x) - \mu_{g(X)}]^2 f(x), & X \text{ diskret.} \\ \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)\, dx, & X \text{ kontinuerlig.} \end{cases}\end{aligned}$$

Her er $\mu_{g(X)} = E[g(X)]$.

Dette følger direkte fra teorem 4.1.

Kovarians, sammenheng mellom stokastiske variabler

Kovarians (D 4.4):

La X og Y være være stokastiske variabler med simultan sannsynlighetsfordeling $f(x,y)$ og forventningsverdier μ_X og μ_Y. Da er **kovariansen** mellom X og Y:

$$\begin{aligned}\sigma_{XY} &= Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \\ &= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)\, f(x,y), & X \text{ diskret.} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f(x,y)\, dxdy, & X \text{ kontinuerlig.} \end{cases}\end{aligned}$$

Kovariansen er er mål på samvariasjonen mellom X og Y (korrelasjon kommer seinere).

Spørsmål: Hva er sammenhengen mellom kovarians og varians?

Varians blir spesialtilfelle av kovarians.

$Cov(X,X) = Var(X)$

Illustrasjoner, kovarians

Rett linja

Kovariansverdien er et mål på den **lineære** sammenhengen mellom X og Y.

La oss se på tre tilfeller:

$Cov(X,Y) > 0$

Verdien av X og Y "følger" hverandre.

Liten/stor X tenderer til å komme sammen med liten/stor Y,

$Cov(X,Y) < 0$

Verdien av X og Y "går motsatt".

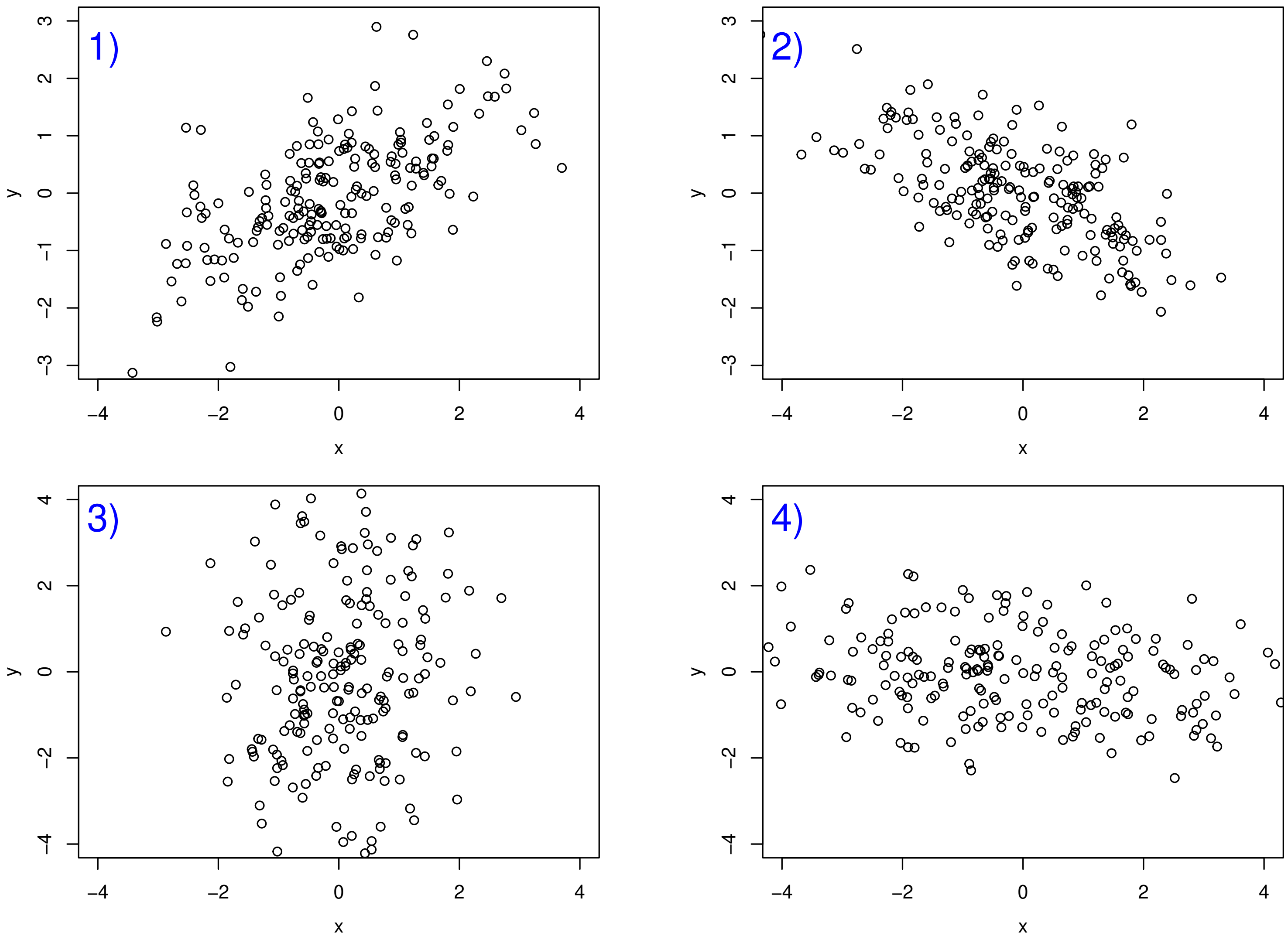
Liten/stor X tenderer til å komme sammen med stor/liten Y.

$Cov(X,Y) = 0$

Ingen lineær sammenheng mellom X og Y.

Illustrasjoner, kovarians

Eksempler på data fra kontinuerlige fordelinger med ulik kovarians:



Sen ut til å komma fra fordeling med:

1) $\sigma_{xy} > 0$

2) $\sigma_{xy} < 0$

3) $\sigma_{xy} = 0$

4) $\sigma_{xy} = 0$

Eksempel: Diskrete fordelinger med positiv, negativ og ingen kovrians.

1)

		x			
		1	2	3	$h(y)$
	1	4/20	2/20	0/20	6/20
y	2	2/20	4/20	2/20	8/20
	3	0/20	2/20	4/20	6/20
$g(x)$		6/20	8/20	6/20	1

$E(X) = \mu_X = 2 \qquad E(Y) = \mu_Y = 2$

$$Cov(X,Y) = \sum_{x=1}^{3} \sum_{y=1}^{3} (x-\mu_X)(y-\mu_Y) f(x,y)$$

$$= (1-2)\cdot(1-2)\cdot\frac{4}{20} + \cdots + (3-2)(3-2)\cdot\frac{4}{20}$$

$$= \underline{\underline{8/20}}$$

2)

		x			
		1	2	3	$h(y)$
	1	0/20	2/20	4/20	6/20
y	2	2/20	4/20	2/20	8/20
	3	4/20	2/20	0/20	6/20
$g(x)$		6/20	~~7~~/20 8	6/20	1

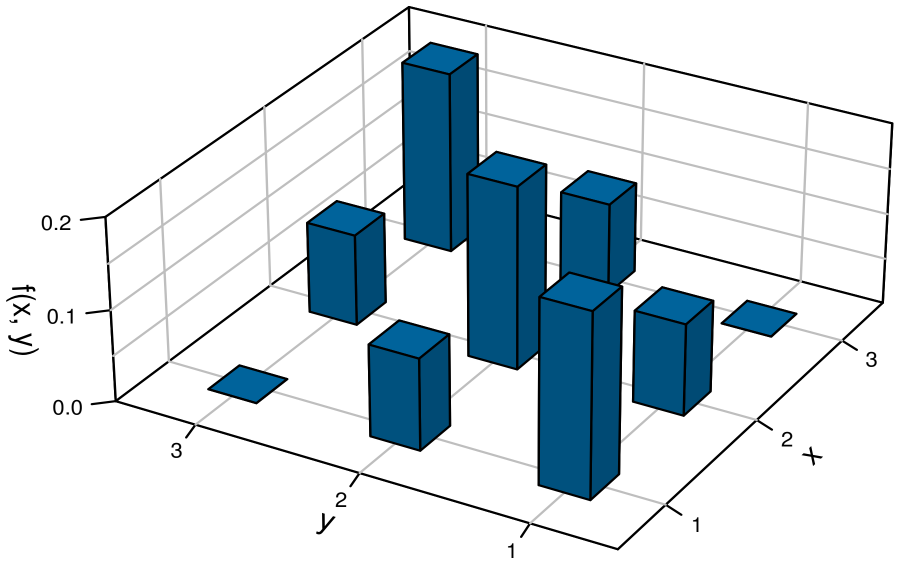
$$Cov(X,Y) = \underline{\underline{-8/20}}$$

3)

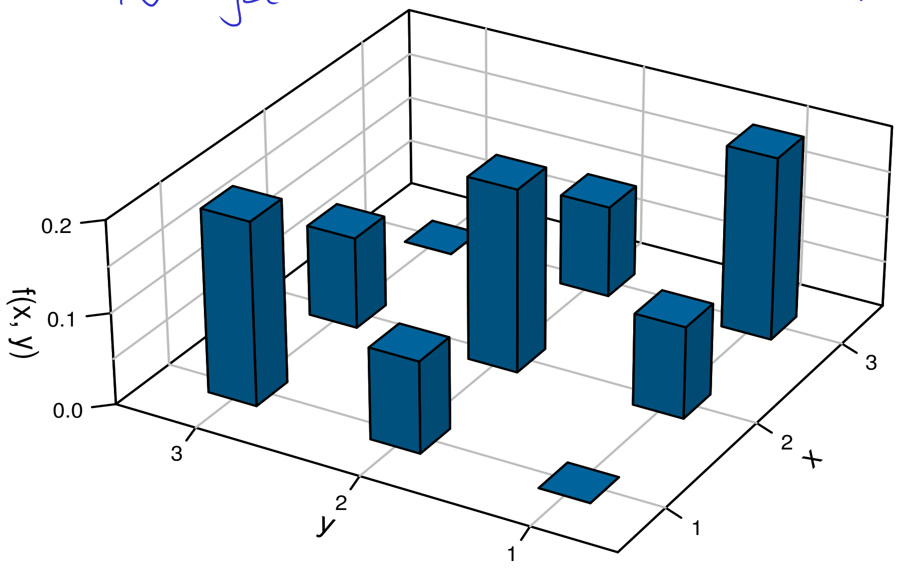
		x			
		1	2	3	$h(y)$
	1	2/20	1/20	2/20	5/20
y	2	4/20	2/20	4/20	10/~~19~~ 20
	3	2/20	1/20	2/20	5/20
$g(x)$		8/20	4/20	8/20	1

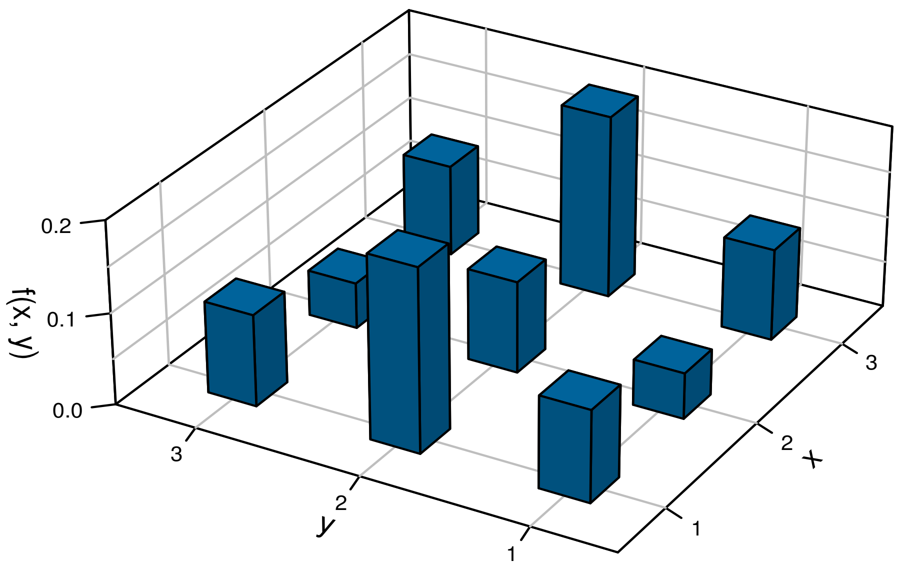
$$Cov(X,Y) = \underline{\underline{0}}$$

Positiv kovarians:



Negativ kovarians:





← kovarians 0.

(Merk at her er X og Y uavhengige, betinga fordelingen lik marginal.)

Kovariansregning: Alternativ formel

Som for varians eksisterer det en alternativ formel.

Alternativ kovariansformel (T 4.4):

Kovariansen mellom to stokastiske variabler X og Y er:

$$\sigma_{XY} = Cov(X,Y) = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y$$

Denne er vanligvis noe enklere i bruk.

Bevis: (Kontinuerlig tilfelle.)

$$\sigma_{XY} = Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)\,dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dxdy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\mu_Y f(x,y)\,dx\,dy$$

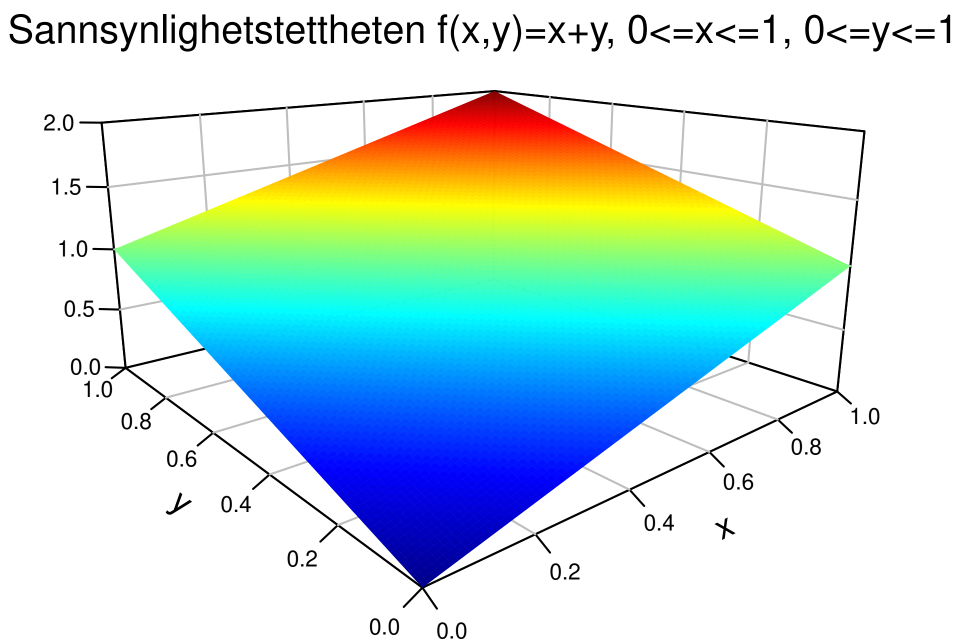
$$- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mu_X y\, f(x,y)\,dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mu_X\mu_Y f(x,y)\,dxdy$$

$$= E(XY) - \mu_Y \int_{-\infty}^{\infty} x \overbrace{\int_{-\infty}^{\infty} f(x,y)\,dy}^{g(x)}\,dx$$

$$- \mu_X \int_{-\infty}^{\infty} y \underbrace{\int_{-\infty}^{\infty} f(x,y)\,dx}_{h(y)}\,dy + \mu_X\mu_Y \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy}_{1}$$

$$= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y = \underline{\underline{E(XY) - \mu_X\mu_Y}}$$

Eksempel: Hardhet, X, og slitasjeindeks, Y, for metallprøve



$E(X) = 7/12$

$E(Y) = 7/12$

$Var(X) = 11/144$

$Var(Y) = 11/144$

Hva blir kovariansen mellom hardhet og slitasjeindeks?

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy = \int_0^1\int_0^1 xy(x+y)\,dx\,dy = \int_0^1\int_0^1 (x^2y + xy^2)\,dx\,dy$$

$$= \int_0^1 \left(\frac{1}{3}y + \frac{1}{2}y^2\right) dy = \underline{1/3}$$

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y = \frac{1}{3} - \frac{7}{12}\cdot\frac{7}{12} = -\underline{\underline{\frac{1}{144}}}$$

Korrelasjon, skalert kovarians

Vanligvis oppgis ikke kovariansen, men **korrelasjonen**.

Korrelasjon (D 4.5):

Dersom X og Y er stokastiske variabler med kovarians σ_{XY} og standardavvik σ_X og σ_Y, er **korrelasjonen** mellom X og Y (bruker ofte bokstaven ρ (rho) for denne):

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

- ρ_{XY} er et tall mellom -1 og 1.
- Fortegnet angir **positiv** eller **negativ** lineær sammenheng.
- Absoluttverdien angir styrken i den lineære sammenhengen.

Jo nærmere ± 1, jo sterkere.

Eksempel: Hardhet, X og slitasjeindeks, Y for metallprøve.

Hva blir korrelasjonen mellom variablene?

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \frac{-\frac{1}{144}}{\sqrt{\frac{11}{144} \cdot \frac{11}{144}}}$$

$$= -\frac{1}{11}$$

Svak negativ korrelasjon.

Uavhengige stokastiske variabler

Uavhengighet (T 4.8):

Dersom X og Y er to **uavhengige** stokastiske variabler vil

$$E(XY) = E(X) \cdot E(Y)$$

(Obs: Det motsatte trenger ikke være tilfelle.)

Bevis: (Kontinuerlig tilfelle.)

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f(x,y)\,dxdy \overset{\text{uavh}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, g(x)h(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} y\,h(y) \int_{-\infty}^{\infty} x\,g(x)\,dx\,dy = \int_{-\infty}^{\infty} y\,h(y)\cdot \mu_X\,dy$$

$$= \mu_X \int_{-\infty}^{\infty} y\,h(y)\,dy = \underline{\underline{\mu_X \mu_Y}}$$

Uavhengige stokastiske variabler forts.

Viktig: Hva blir **kovariansen** og **korrelasjonen** av uavhengige variabler? (K 4.5)

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y = \mu_X \mu_Y - \mu_X \mu_Y = \underline{\underline{0}}$$

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \underline{\underline{0}}$$

Så uavhengighet $\Rightarrow$ $Cov(X,Y) = 0$

- Men $Cov(X,Y) = 0$ medfører ikke nødvendigvis uavhengighet.

4.3 Linærkombinasjoner av stokastiske variabler

Dette er egentlig ulike spesialtilfeller av teorem 4.1, der vi kan kraftig forenkle utrekningene.

T 4.5, K 4.1, K 4.2:

Dersom a og b er konstanter og X en stokastisk variabel har vi

$$E(a\,X + b) = a\,E(X) + b$$

Spesialtilfeller:

g(X) alltid lik b (konst.)

$$\begin{aligned} a = 0: \quad & E(b) &= \quad & b \\ b = 0: \quad & E(a\,X) &= \quad & a\,E(X) \end{aligned}$$

(Hva betyr dette?)

("Forventninga av lineærkombinasjonen er lineærkombinasjonen av forventninga.")

Bevis:

Kontinuerlig:

$$E(aX+b) = \int_{-\infty}^{\infty}(ax+b)f(x)\,dx = a\underbrace{\int_{-\infty}^{\infty} x f(x)\,dx}_{E(X)} + b\underbrace{\int_{-\infty}^{\infty} f(x)\,dx}_{1}$$

$$= a\mu + b$$

Diskret:

$$E(aX+b) = \sum_x (ax+b)f(x) = a\underbrace{\sum_x x f(x)}_{E(X)} + b\underbrace{\sum_x f(x)}_{1}$$

$$= a\mu + b$$

Sum eller differanse av to funksjoner

T 4.6, T 4.7, K 4.3, K 4.4:

Dersom g og h er funksjoner av en eller to stokastiske variabler:

$$
\begin{aligned}
E[g(X) \pm h(X)] &= E[g(X)] \pm E[h(X)] \\
E[g(X,Y) \pm h(X,Y)] &= E[g(X,Y)] \pm E[h(X,Y)]
\end{aligned}
$$

Spesialtilfeller:

$$
\begin{aligned}
E[g(X) \pm h(Y)] &= E[g(X)] \pm E[h(Y)] \\
E(X \pm Y) &= E(X) \pm E(Y)
\end{aligned}
$$

Altså: Forventning til en sum er summen av forventningene, tilsvarende for differanse.

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

Bevis: (Kontinuerlig tilfelle.)

$$E[g(X,Y) \pm h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(g(x,y) \pm h(x,y)\right) f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,dx\,dy \pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\,dx\,dy$$

$$= E[g(X,Y)] \pm E[h(X,Y)]$$

Eksempel: Meklerbetaling.

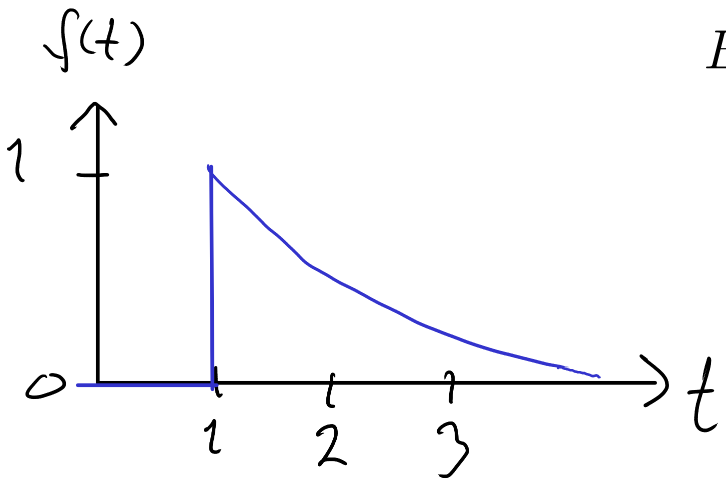
La salgsprisen av ei leilighet, T (i millioner) ha følgende sannsynlighetstetthet:

$$f(t) = e^{-(t-1)}, \quad t > 1$$

- Hva blir forventa pris?
- Hva blir forventa betaling til mekleren som skal ha 50000 + 10% av delen av beløpet som overstiger 1 million? Altså betalingsfunksjon:

$$B = 0.05 + (T - 1) \cdot 0.10$$

1)



$$E(T) = \int_{-\infty}^{\infty} t \cdot f(t)\,dt = \int_1^{\infty} t \cdot e^{-(t-1)}\,dt$$

$$= e \int_1^{\infty} t e^{-t}\,dt = e\left[-te^{-t} - e^{-t}\right]_1^{\infty}$$

$$= e\, 2 e^{-1} = \underline{\underline{2}}$$

2) $B = 0{,}05 + 0{,}1 \cdot (T-1)$

$$E(B) = E(0{,}05 + 0{,}1 \cdot (T-1))$$

$$= 0{,}05 + 0{,}1 \cdot E(T-1)$$

$$= 0{,}05 + 0{,}1 \cdot (E(T) - 1)$$

$$= -0{,}05 + 0{,}1 \cdot E(T)$$

$$= -0{,}05 + 0{,}1 \cdot 2$$

$$= \underline{\underline{0{,}15}} \qquad 150\,000 \text{ kr.}$$

Eksempel: Metallprøve.

$$f(x,y) = \begin{cases} x+y & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{ellers} \end{cases}$$

Anta at produsenten får betalt etter både hardhet (X) og slitasjeindeks (Y) etter denne funksjonen:

$$g(x,y) = 3\,e^{x} - 2y$$

Kjent fra tidligere:

$$E(X) = E(Y) = \frac{7}{12} \qquad E\left(e^{X}\right) = \frac{1}{2}(1+e)$$

Hva blir forventa betaling?

$$E(3e^X - 2Y) = E(3e^X) - E(2Y) = 3E(e^X) - 2E(Y)$$

$$= 3 \cdot \frac{1}{2} \cdot (1+e) - 2 \cdot \frac{7}{12} = \frac{1}{3} + \frac{3}{2}e = 4{,}41$$

Varians av lineærkombinasjoner

(Egentlig ulike spesialtilfeller av teorem 4.1/4.3, der vi kan kraftig forenkle utrekningene.)

T 4.9, K 4.7, K 4.8:

For stokastisk variabel X og konstanter a og b har vi

$$Var(a\,X + b) = a^2 Var(X) = a^2 \sigma_X^2$$

Spesialtilfeller:

$$\begin{aligned} Var(X + b) &= Var(X) \\ Var(a\,X) &= a^2 Var(X) \\ Var(b) &= 0 \end{aligned}$$

Varians av „konstant".

Bevis: ~~(Kontinuerlig tilfelle.)~~

Husk: $Var[g(X)] = E\left\{[g(X) - \mu_{g(X)}]^2\right\}$.

$$
\begin{aligned}
Var(aX+b) &= \cancel{\int_{-\infty}^{\infty} [(ax+b) - E(aX+b)]^2 f(x)dx} \\
&= E\left\{\left[aX+b-E(aX+b)\right]^2\right\} \\
&= E\left\{\left[aX+b-aE(X)-b\right]^2\right\} \\
&= E\left[(aX-a\mu)^2\right] \\
&= E\left[a^2(X-\mu)^2\right] \\
&= a^2 E\left[(X-\mu)^2\right] \\
&= \underline{\underline{a^2 Var(X)}}
\end{aligned}
$$

Eksempel: Meklerbetaling forts.

Hva blir variansen til salgsprisen?

- Hva blir variansen i betalinga til mekleren som skal ha 10% av total pris?
- Hva blir variansen i betalinga til mekleren som skal ha 50000 + 10% av delen av beløpet som overstiger 1 million?
- Hva blir variansen i betalinga til mekleren som skal ha 100000 uansett pris?

$$\mathrm{Var}(T) = E(T^2) - \mu^2 = \int_{-\infty}^{\infty} t^2 \cdot f(t)\,dt - 2^2$$

$$= \int_1^{\infty} t^2 e^{-(t-1)}\,dt - 4 \underset{v=t-1}{\overset{\uparrow}{=}} \int_0^{\infty} (v+1)^2 e^{-v}\,dv - 4$$

$$= \int_0^{\infty} v^2 e^{-v}\,dv + 2\int_0^{\infty} v e^{-v}\,dv + \int_0^{\infty} e^{-v}\,dv - 4$$

Gamma-funksjon, se tabellhefte.

$$= \Gamma(3) + 2\Gamma(2) + \Gamma(1) - 4 = 2! + 2\cdot 1! + 0! - 4 = \underline{\underline{1}}$$

1) $B_1 = 0{,}1T$

$E(B_1) = E(0{,}1T) = 0{,}1 \cdot E(T) = 0{,}1 \cdot 2 = \underline{\underline{0{,}2}}$

$Var(B_1) = Var(0{,}1T) = 0{,}1^2\, Var(T) = \underline{\underline{0{,}01}}$

$SD(T) = \underline{\underline{0{,}1}}$

2) $B_2 = 0{,}05 + 0{,}1 \cdot (T-1)$

$E(B_2) = \underline{\underline{0{,}15}}$ (s. 38)

$Var(B_2) = Var(0{,}05 + 0{,}1 \cdot (T-1)) = Var(0{,}1 \cdot (T-1))$

$= 0{,}1^2 \cdot Var(T-1) = 0{,}1^2 \cdot Var(T) = \underline{\underline{0{,}01}}$

$SD(T) = \underline{\underline{0{,}1}}$

3) $B_3 = 0{,}1$

$E(B_3) = E(0{,}1) = \underline{\underline{0{,}1}}$

$Var(B_3) = Var(0{,}1) = \underline{\underline{0}}$

Varians av lineærkombinasjoner forts.

T 4.9, K 4.9, K 4.10:

For stokastiske variabler X og Y og konstanter a, b og c har vi

$$\begin{aligned} Var(a\,X + b\,Y + c) &= a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y) \\ &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY} \end{aligned}$$

Spesialtilfeller: Anta X og Y er **uavhengige**, hva blir da

$$\begin{aligned} Var(a\,X + b\,Y) &= a^2 Var(X) + b^2 Var(Y) \\ Var(a\,X - b\,Y) &= a^2 Var(X) + (-b)^2 Var(Y) = a^2 Var(X) + b^2 Var(Y) \\ Var(X + Y) &= Var(X) + Var(Y) \\ Var(X - Y) &= Var(X) + (-1)^2 Var(Y) = Var(X) + Var(Y) \end{aligned}$$

Merk spesielt de to siste!

Bevis:

$$Var(aX + bY + c) = E\left\{\left[aX + bY + c - E(aX + bY + c)\right]^2\right\}$$

$= $ Se øvingsoppgave.

Eksempel: Trykkeri forts.

		x			
		0	1	2	$P(Y=y)$
	0	0.2	0.2	0.2	0.6
y	1	0.0	0.1	0.2	0.3
	2	0.0	0.0	0.1	0.1
$P(X=x)$		0.2	0.3	0.5	1.0

Hva blir forventninga og variansen til differansen mellom henvendelser X og bestillinger Y?

$$
\begin{aligned}
E(X) &= 0\cdot 0.2 + 1\cdot 0.3 + 2\cdot 0.5 = 1.3 \\
E(Y) &= 0\cdot 0.6 + 1\cdot 0.3 + 2\cdot 0.1 = 0.5 \\
Var(X) &= E(X^2) - \mu_X^2 = (0^2\cdot 0.2 + 1^2\cdot 0.3 + 2^2\cdot 0.5) - 1.3^2 = 0.61 \\
Var(Y) &= E(Y^2) - \mu_Y^2 = (0^2\cdot 0.6 + 1^2\cdot 0.3 + 2^2\cdot 0.1) - 0.5^2 = 0.45 \\
E(XY) &= 0\cdot 0\cdot 0.2 + \cdots + 2\cdot 2\cdot 0.1 = 0.9 \\
Cov(X,Y) &= E(XY) - \mu_X\mu_Y = 0.9 - 1.3\cdot 0.5 = 0.25
\end{aligned}
$$

$$E(X-Y) = E(X) - E(Y) = 1{,}3 - 0{,}5 = \underline{\underline{0{,}8}}$$

$$\begin{aligned}
Var(X-Y) &= Var(X) + (-1)^2 Var(Y) + 2 \cdot 1 \cdot (-1) \, Cov(X,Y) \\
&= Var(X) + Var(Y) - 2 \cdot Cov(X,Y) \\
&= 0{,}61 + 0{,}45 - 2 \cdot 0{,}25 \\
&= \underline{\underline{0{,}56}}
\end{aligned}$$

Eksempel: Eksamensspøsmål NTNU.

Vi har to variabler X og Y der vi antar at:

$$E(X) = 10, \quad Var(X) = 4, \quad E(Y) = 8, \quad Var(Y) = 9, \quad Cov(X,Y) = 5$$

Finn tallverdier for uttrykkene:

$$E(2X - Y) \qquad Var(2X - Y) \qquad E[(X-3)(Y-5)]$$

$$E(2X - Y) = 2E(X) - E(Y) = 2 \cdot 10 - 8 = \underline{\underline{12}}$$

$$Var(2X-Y) = 2^2 Var(X) + (-1)^2 Var(Y) + 2 \cdot 2 \cdot (-1) Cov(X,Y)$$

$$= 4 \cdot 4 + 9 - 4 \cdot 5 = \underline{\underline{5}}$$

$$E[(X-3)(Y-5)] = E(XY - 5X - 3Y + 15) = E(XY) - 5E(X) - 3E(Y) + 15$$

$$= 85 - 5 \cdot 10 - 3 \cdot 8 + 15 = \underline{\underline{26}}$$

$$Cov(X,Y) = E(XY) - E(X) \cdot E(Y) \Rightarrow E(XY) = 5 + 10 \cdot 8 = 85$$

Forventning og varians til lineærkombinasjon av n variabler

Generaliseringer:

For stokastiske variabler $X_1, \ldots, X_n$ og konstanter $a_1, \ldots, a_n$ har vi at:

$$E\left(\sum_{i=1}^{n} a_i\, X_i\right) = \sum_{i=1}^{n} a_i\, E(X_i)$$

Om variablene er **uavhengige** (K 4.11):

Så kovariansen er 0.

$$Var\left(\sum_{i=1}^{n} a_i\, X_i\right) = \sum_{i=1}^{n} a_i^2\, Var(X_i)$$

Se også formler fra "Tabeller og formler ...":

$$\begin{aligned} E\left(\sum_{i=1}^n a_i X_i + b\right) &= \sum_{i=1}^n a_i E(X_i) + b \\ Var\left(\sum_{i=1}^n a_i X_i + b\right) &= \sum_{i=1}^n a_i^2 Var(X_i) + 2\sum_{i=\cancel{n}\,1}^n \sum_{j=1}^{i-1} \cancel{Cov(X_i, X_j)}\; a_i a_j Cov(X_i, X_j) \end{aligned}$$

Eksempel: Valpekull forts.

Y = antall levende valper.

$$E(Y) = 1.7 \qquad Var(Y) = 0.81$$

Oppdretteren har faste utgifter på 4000 per valpekull, og tjener 5000 per levende valp.
Finn forventning og varians for fortjeneste (V) for ett valpekull.
Finn forventning og varians for total fortjeneste (T) av 10 antatt **uavhengige** valpekull.

$$V = 5000\,Y - 4000 \qquad T = \sum_{i=1}^{10} V_i$$

$$\begin{aligned}
E(V) &= 5000 \cdot E(Y) - 4000 = 5000 \cdot 1.7 - 400 = 4500 \\
Var(V) &= 5000^2 \cdot Var(Y) = 5000^2 \cdot 0.81 = 20250000 \qquad SD(V) = 4500 \\
E(T) &= E(V_1 + V_2 + \cdots + V_{10}) = E(V_1) + \cdots + E(V_{10}) = \underline{\underline{45000}} \\
Var(T) &= Var(V_1 + \cdots + V_{10}) \overset{\text{uavh.}}{=} Var(V_1) + \cdots + Var(V_{10}) \\
&= \underline{\underline{202500000}} \qquad SD(T) = 14230,2
\end{aligned}$$

Et nyttig triks: Indikatorvariabler

- Kan ofte skrive en stokastisk variabel som summen av 0/1-variabler.
- Enkeltvis kan disse være enkle å finne forventning/varians av.
- Kan til slutt bruke setningene om forventning/varians av lineærkombinasjoner.

- La A være en hendelse som har sannsynlighet p for å inntreffe i et stokastisk forsøk.
- La I_A være en stokastisk variabel som er slik at: $I_A = \begin{cases} 1, & \text{hendelsen A inntraff.} \\ 0, & \text{hendelsen A inntraff ikke.} \end{cases}$
- Da blir fordelinga (punktsannsynligheten) til I_A:

i	0	1
$P(I_A = i)$	$1-p$	p

- Kan da vise at:

$$E(I_A) = p \qquad \text{og} \qquad Var(I_A) = p(1-p)$$

$E(I_A) = 0 \cdot (1-p) + 1 \cdot p = \underline{\underline{p}},$

$E(I_A^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$

$Var(I_A) = E(I_A^2) - (E(I_A))^2$

$= p - p^2 = \underline{\underline{p(1-p)}}$

Eksempel: 6-ere på terningkast

- La N være antall 6-ere på 3 terningkast. Finn forventinga av N.
  Hint: Kan skrive N som
  $$N = I_1 + I_2 + I_3$$
  der I_i er indikatorvariabel for hendelsen "6-er på kast i".

Fordeling og forventning av en indikatorvariabel:

i	0	1
$P(I = i)$	$5/6$	$1/6$

$$E(I) = P(I = 1) = 1/6$$

$$E(N) = E(\underbrace{I_1 + I_2 + I_3}_{\#\text{ 6-ere totalt}}) = E(I_1) + E(I_2) + E(I_3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$$

Kan òg finne varians på samme måte.

Kort sammenfatting: Forventning og varians til funksjoner av X

- Lineærkombinasjoner: Enkle regler gjør at vi kan forenkle problemet. Noen eksempler:

$$\begin{aligned} E\left(2X + 3e^Y - 3\right) &= E(2X) + E\left(3e^Y\right) - 3 = 2E(X) + 3E\left(e^Y\right) - 3 \\ Var\left(3Y^2 + 2\right) &= Var\left(3Y^2\right) = 3^2 Var\left(Y^2\right) \\ Var(2X - 3Y) &= 2^2 Var(X) + (-3)^2 Var(Y) + 2 \cdot 2 \cdot (-3) Cov(X, Y) \end{aligned}$$

- For å komme videre må du allerede kjenne $E(X)$, $E(e^Y)$, $Var(X)$, $Var(Y^2)$, etc. Eller du må ta i bruk "grunnteorema" (4.1, 4.3) for forventning og varians:

$$\begin{aligned} E(X) &= \int x f(x)\, dx & Var(X) &= E\left[(X - \mu_X)^2\right] \\ E[g(X)] &= \int g(x) f(x)\, dx & Var[g(X)] &= E\left[(g(X) - \mu_{g(X)})^2\right] \end{aligned}$$

- Ofte komplisert å finne forventning/varians av $\underbrace{\text{ikke-lineær funksjon } g(X)}_{\text{som } X^2,\ e^X,\ \log(X),\ 1/X}$, finnes alternativ:
  - Erstatt $g(X)$ med ei Taylor-rekkeutvikling.
  - Også nyttig om du kjenner forventning/varians til X men ikke fullstendig sannsynlighetsfordeling $f(x)$.

Rekkeutvikling, forventning

Andre ordens Taylorrekke-approksimasjon for funksjon $g(x)$ om $\mu_X = E(X)$:

$$g(x) \approx g(\mu_X) + \left.\frac{\partial g(x)}{\partial x}\right|_{x=\mu_X} \cdot (x-\mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \cdot \frac{(x-\mu_X)^2}{2}$$

Forventning av dette:

$$\begin{aligned} E[g(X)] &\approx E[g(\mu_X)] + \left.\frac{\partial g(x)}{\partial x}\right|_{x=\mu_X} \cdot E(X-\mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \cdot \frac{E[(X-\mu_X)^2]}{2} \\ &= g(\mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \cdot \frac{\sigma_X^2}{2} \end{aligned}$$

$E(X-\mu_X) = E(X) - \mu_X$

$= \mu_X - \mu_X = 0$

Lar seg generalisere til funksjoner av fleire (uavhengige) variabler:

$$E[g(X,Y)] \approx g(\mu_X,\mu_Y) + \left.\frac{\partial^2 g(x,y)}{\partial x^2}\right|_{\substack{x=\mu_X \\ y=\mu_Y}} \cdot \frac{\sigma_X^2}{2} + \left.\frac{\partial^2 g(x,y)}{\partial y^2}\right|_{\substack{x=\mu_X \\ y=\mu_Y}} \cdot \frac{\sigma_Y^2}{2}$$

Rekkeutvikling, varians

Første ordens Taylorrekke-approksimasjon for funksjon $g(x)$ om $\mu_X = E(X)$:

$$g(x) \approx g(\mu_X) + \left.\frac{\partial g(x)}{\partial x}\right|_{x=\mu_X} \cdot (x - \mu_X)$$

Varians av dette:

$$\begin{aligned} Var[g(X)] &\approx \overbrace{Var[g(\mu_X)]}^{0} + \left[\frac{\partial g(x)}{\partial x}\right]^2_{x=\mu_X} \cdot \underbrace{Var(X - \mu_X)} \\ &= \left[\frac{\partial g(x)}{\partial x}\right]^2_{x=\mu_X} \cdot \sigma_X^2 \end{aligned}$$

$Var(X - \mu_X) = Var(X)$

konst.

Lar seg tilsvarende generalisere:

$$Var[g(X,Y)] \approx \left[\frac{\partial g(x,y)}{\partial x}\right]^2_{\substack{x=\mu_X \\ y=\mu_Y}} \cdot \sigma_X^2 + \left[\frac{\partial g(x,y)}{\partial y}\right]^2_{\substack{x=\mu_X \\ y=\mu_Y}} \cdot \sigma_Y^2$$

Eksempel: Metallprøve. La X være hardhet av tilfeldig valgt metallprøve.

$$f(x) = \begin{cases} 1/2 + x, & 0 \leq x \leq 1. \\ 0, & \text{ellers.} \end{cases}$$

$$E(X) = \frac{7}{12} \qquad Var(X) = \frac{11}{144}$$

Vi antar produsenten får betalt etter følgende funksjon: $g(x) = e^x$

Finn forventning og varians for $g(X)$, eksakt og tilnærma.

Eksakt:

$$\begin{aligned} E\left(e^X\right) &= \int_{-\infty}^{\infty} g(x)f(x)dx = \int_0^1 e^x \left(x + \frac{1}{2}\right) dx = \frac{e+1}{2} = 1.859 \quad \text{(fra tidligere)} \\ Var\left(e^X\right) &= E[(g(X) - \mu_{g(X)})^2] = E[g(X)^2] - \{E[g(X)]\}^2 \\ &= \int_0^1 (e^x)^2 \left(x + \frac{1}{2}\right) dx - \left(\frac{e+1}{2}\right)^2 = \frac{e^2}{4} - \frac{1}{2} - \frac{1}{4} = 0.238 \end{aligned}$$

Tilnærma:

$$g'(x) = e^x$$

$$g''(x) = e^x$$

$$E\left(e^X\right) \approx e^{\mu_X} + \left.\frac{\partial^2(e^X)}{\partial X^2}\right|_{X=\mu_X} \cdot \frac{\sigma_X^2}{2} = e^{\mu_X} + e^{\mu_X} \cdot \frac{\sigma_X^2}{2}$$

$$= e^{7/12}\left(1 + \frac{1}{2} \cdot \frac{11}{144}\right) = \frac{299}{288} \cdot e^{7/12} = \underline{\underline{1{,}860}}$$

$$Var\left(e^X\right) \approx \left(\left.\frac{\partial(e^X)}{\partial X}\right|_{X=\mu_X}\right)^2 \cdot \sigma_X^2 = \left(e^{\mu_X}\right)^2 \cdot \sigma_X^2$$

$$= \left(e^{7/12}\right)^2 \cdot \frac{11}{144} = \underline{\underline{0{,}245}}$$

4.4 Tsjebysjeffs teorem

Anta at vi vil regne ut en sannsynlighet, for eksempel

$$P(1 < X < 5)$$

Kjenner ikke fordelinga til X, bare forventning og varians. (Kjenner et sentralmål for fordelinga og har et mål på spredninga, men kjenner ellers ikke formen på fordelinga.)
Kan da ikke regne eksakt sannsynlighet, men kan finne grense (skranke) for sannsynligheten:

Tsjebysjeff (T 4.10):

Sannsynligheten for at verdien på en variabel X tar verdi innafor k standardavvik fra forventningsverdien er minst $1 - \frac{1}{k^2}$:

$$P\left(\mu_X - k \cdot \sigma_X < X < \mu_X + k \cdot \sigma_X\right) \geq 1 - \frac{1}{k^2}$$

Tar ikke med bevis.

Eksempel: Bussventetid.

Du går til holdeplassen for å ta bussen. Men du kjenner ikke den eksakte ruta, bare at det kommer en buss akkurat hvert 10. minutt. Fordelinga for ventetida:

$$f(x) = \begin{cases} \frac{1}{10}, & 0 \leq x \leq 10 \\ 0, & \text{ellers} \end{cases} \qquad \mu = \frac{10}{2} = 5 \qquad \sigma = \frac{5}{\sqrt{3}} \approx 2.89$$

Hva blir sannsynligheten for at du skal vente $\mu \pm k\sigma$?

negativ s.s., altså ≥ 0 her

	Sannsynlighet	Eksakt	k	Skranke
$k\sigma = 1$:	$P(4 < X < 6)$	$\frac{2}{10} = 0.20$	$k = \frac{1}{\sigma} = 0.346$	$\geq 1 - 1/k^2 = 1 - 1/0{,}346^2 = -7{,}3$
$k\sigma = 4$:	$P(1 < X < 9)$	$\frac{8}{10} = 0.80$	$k = \frac{4}{\sigma} = 1{,}386$	$\geq 1 - 1/k^2 = 1 - 1/1{,}386^2 = 0{,}429$

Ganske dårlige resultater? Dette skyldes at teoremet gjelder uansett formen på fordelinga.

Vil gi noe bedre resultat for enkelte andre fordelinger.