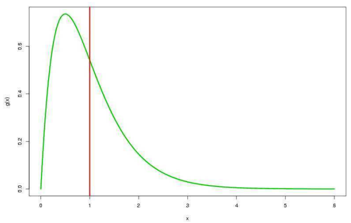
Kapittel 4: Matematisk forventning

TMA4240 Statistikk (F2 og E7)

Univariate tilfeller foreleses onsdag 1.september, 2004

Multivariate tilfeller foreleses mandag 6.september, 2004



4.1 Forventing til en stokastisk variabel

DEF 4.1: La X være en stokastisk variabel med sannsynlighetsfordeling $f(x)$.
Forventningsverdien (mean, expected value) til X er

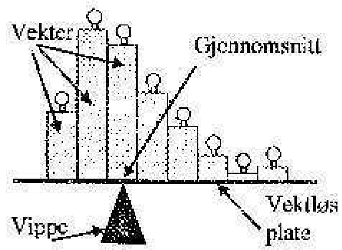
$$\mu = E(X) = \sum_{\forall x} x f(x)$$

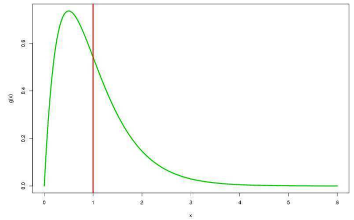
hvis X er diskret, og

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

hvis X er kontinuerlig.

Tyngdepunkt





Togforsinkelsen (forts.)

- Deler av oppgave 1, eksamen desember 2003.
  - I denne oppgaven kan du bruke uten å vise det at

$$\int_0^\infty x^r e^{-ax} dx = \frac{r!}{a^{r+1}} \text{ når } a > 0 \text{ og } r \text{ er et heltall} \geq 0$$

  - Vi betrakter ankomst- og oppholdstider for et bestemt lokaltog på en jernbanestasjon. Toget skal etter rutetabellen ankomme hver hverdag klokka 8:00, men kommer alltid etter dette tidspunktet.
  - La X (minutter) betegne togets forsinkelse på en tilfeldig valgt hverdag. Vi antar at X er en stokastisk variabel med sannsynlighetstetthet

$$f_X(x) = \begin{cases} kxe^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

  der $k > 0$ er en konstant.

i) Har vist at $k = 4$.

ii) Hva er forventingsverdien til X?

Prosjektstyring

- X = tid for å samle inn data
- Y = tid for å analysere data

x y

	1	**2**	**3**	$f_Y(y)$
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
$f_X(x)$	0.10	0.60	0.30	1.00

Prosjektstyring (forts.)

- Ser på tid brukt til datainnsamling (X)
- Kunden har betalt 1200 kr for datainnsamlingen, og prosjektarbeideren som skal utføre datainnsamlingen får 500 kr timen.
- Hva er forventet inntekt for datainnsamlingen?

		x			
		1	**2**	**3**	$f_Y(y)$
	1	0.03	0.05	0.02	0.10
	2	0.03	0.14	0.03	0.20
y	**3**	0.03	0.17	0.10	0.30
	4	0.01	0.24	0.15	0.40
	$f_X(x)$	0.10	0.60	0.30	1.00

Forventing til funksjon av en stokastisk variabel

TEO 4.1: La X være en stokastisk variabel med sannsynlighetsfordeling $f(x)$.
Forventningsverdien til den stokastiske variablen $g(X)$ er

$$\mu_{g(X)} = E[g(X)] = \sum_{\forall x} g(x) f(x)$$

hvis X er diskret, og

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

hvis X er kontinuerlig.

$E(aX + b)$

TEO 4.5: Hvis a og b er konstanter, så er

$$E(aX + b) = aE(X) + b$$

COR 1: Setter vi $a = 0$ ser vi at $E(b) = b$

COR 2: Setter vi $b = 0$ ser vi at $E(aX) = aE(X)$

E(sum eller differanse)

TEO 4.6: Forventningsverdien til summen eller differansen av to eller flere funksjoner av den stokastiske variable X, er summen eller differansen til forventningsverdiene til funksjonene. Det vil si,

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

siden

$$\begin{aligned} g(X) &= g_1(X) \pm g_2(X) \\ E(g(X)) &= E(g_1(X) \pm g_2(X)) \\ &= \int_{-\infty}^{\infty} [g_1(X) \pm g_2(X)] \cdot f(x) dx \\ &= E[g_1(X)] \pm E[g_2(X)]. \end{aligned}$$

4.2 Varians (og kovarians)

DEF 4.3: La X være en stokastisk variabel med sannsynlighetsfordeling $f(x)$ og forventning $\mu = E(X)$. *Variansen* til X er

$$\sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

hvis X er diskret, og

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

hvis X er kontinuerlig.
Den positive kvadratroten av variansen, $\sigma = SD(X)$, kalles standard avviket til X.

TEO 4.2: Variansen til en stokastisk variabel X er

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

Togforsinkelsen (forts.)

$$\int_0^{\infty} x^r e^{-ax} dx = \frac{r!}{a^{r+1}} \text{ når } a > 0 \text{ og } r \text{ er et heltall} \geq 0$$

$$f_X(x) = \begin{cases} 4xe^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

- E(X)=1
- Hva er variansen til X?

Varians til funksjon av en stokastisk variabel

TEO 4.3: La X være en stokastisk variabel med sannsynlighetsfordeling $f(x)$. Variansen til den stokastiske variablen $g(X)$ er

$$\sigma^2_{g(X)} = E[(g(X) - \mu_{g(X)})^2] = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$

hvis X er diskret, og

$$\sigma^2_{g(X)} = E[(g(X) - \mu_{g(X)})^2] = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x) dx$$

hvis X er kontinuerlig.

Varians til lineærkombinasjon av stokastisk variabel

TEO 4.9: Hvis a og b er konstanter, så er

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

COR 1: Setter vi $a = 1$ ser vi at

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

COR 2: Setter vi $b = 0$ ser vi at

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

Prosjektstyring (forts.)

- Ser på aktivitet A.
- Kunden har betalt 1200 kr for aktivitet A, og prosjektarbeideren som skal utføre aktivitet A får 500 kr timen.
- Inntekt for aktivitet A: $g(X) = 1200 - 500 \cdot X$
- Forventning: $E(g(X)) = 100$.
- Hva er $\text{Var}(g(X))$?

	1	**2**	**3**
$f_X(x)$	0.10	0.60	0.30

4.4. Chebyshevs teorem

TEO 4.11: Chebyshevs teorem Sannsynligheten for at en stokastisk variabel X vil anta en verdi innen k standardavvik fra forventningsverdien er minst $1 - 1/k^2$. Det vil si,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

- k=1: $P(\mu - \sigma < X < \mu + \sigma) \geq 1 - \frac{1}{1^2} = 0$
- k=2: $P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{2^2} = 0.75$
- k=3: $P(\mu - 3\sigma < X < \mu + 3\sigma) \geq 1 - \frac{1}{3^2} = 0.89$

Oppsummering

Diskret stokastisk variabel	**Kontinuerlig stokastisk variabel**
$\mu = E(X) = \sum_x x f(x)$	$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$
$\mu_{g(X)} = E(g(X)) = \sum_x g(x) f(x)$	$\mu_{g(X)} = E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$
$g(X) = aX + b$: $\mu = E(g(X)) = aE(X) + b$	$g(X) = aX + b$: $\mu = E(g(X)) = aE(X) + b$
$g(X) = g_1(X) \pm g_2(X)$: $\mu = E(g(X)) = E(g_1(X) \pm g_2(X)) = E(g_1(X)) \pm E(g_2(X))$	$g(X) = g_1(X) \pm g_2(X)$: $\mu = E(g(X)) = E(g_1(X) \pm g_2(X)) = E(g_1(X)) \pm E(g_2(X))$
$\sigma^2 = \text{Var}(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x) = E(X^2) - \mu^2$	$\sigma^2 = \text{Var}(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) = E(X^2) - \mu^2$
$\sigma^2_{g(X)} = \text{Var}(g(X)) = E[(g(X) - \mu_{g(X)})^2] = \sum_x (g(x) - \mu_{g(X)})^2 f(x) = E(g(X)^2) - \mu^2_{g(X)}$	$\sigma^2_{g(X)} = \text{Var}(g(X)) = E[(g(X) - \mu_{g(X)})^2] = \int_{-\infty}^{\infty} (g(x) - \mu_{g(X)})^2 f(x) = E(g(X)^2) - \mu^2_{g(X)}$
$\sigma^2_{aX+b} = \text{Var}(aX+b) = a^2\text{Var}(X) = a^2\sigma^2_X$	$\sigma^2_{aX+b} = \text{Var}(aX+b) = a^2\text{Var}(X) = a^2\sigma^2_X$