STK1100 våren 2018

Kontinuerlige stokastiske variabler
Forventning og varians
Momentgenererende funksjoner

Svarer til avsnittene 4.1 og 4.2 i læreboka

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

En diskret stokastisk variabel variabel kan anta endelig mange eller tellbart uendelig mange mulige verdier.

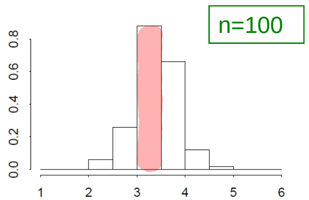
En kontinuerlig stokastisk variabel kan (i prinsippet) anta alle verdier i et intervall på tallinja (eventuelt hele tallinja). Eksempel: X = «vekt til nyfødt jente».

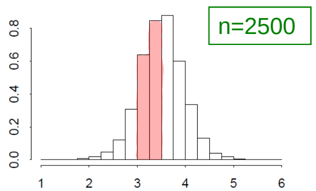
Sannsynlighetsfordelingen for en diskret stokastisk variabel er gitt ved punktsannsynligheten $p(x)=P(X=x)$ (jf. kapittel 3 i læreboka).

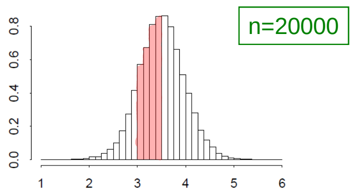
Vi kan ikke gi sannsynlighetsfordelingen for en kontinuerlig stokastisk variabel på samme måte.

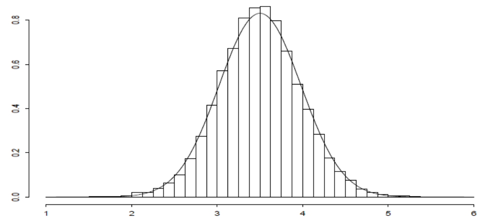
Vil se hvordan vi kan angi sannsynlighets-fordelingen til X = «vekt til nyfødt jente».

Tegner histogrammer for fødselsvektene til tilfeldige utvalg av 100, 2500 og 20000 «fullbårne» jenter (svangerskapslengde minst 37 uker) . Histogrammene er normert slik at arealet av en søyle er lik den relative frekvensen for det intervallet søylen dekker.









Når n er stor, kan histogrammet tilnærmes med funksjonen

$$f(x) = \frac{1}{0.48\sqrt{2\pi}} e^{-\frac{1}{2\cdot 0.48^2}(x-3.5)^2}$$

$f(x)$ kalles sannsynlighetstettheten til X og svarer til histogrammet for «uendelig mange» fødselsvekter.

Vi har da at $P(a \le X \le b) = \int_a^b f(x)dx$

Generelt har vi følgende:

La X være en kontinuerlig stokastisk variabel med sannsynlighetstetthet $f(x)$. Da er

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Ofte sier vi tetthet i stedet for sannsynlighetstetthet

Merk at vi for enhver sannsynlighetstetthet har at

- $f(x) \geq 0$ for alle x
- $\int_{-\infty}^{\infty} f(x)dx = 1$

La X være en kontinuerlig stokastisk variabel med sannsynlighetstetthet $f(x)$. Da er

$$P(X = c) = \lim_{\varepsilon \to 0} P(c - \varepsilon \leq X \leq c + \varepsilon) = \lim_{\varepsilon \to 0} \int_{c-\varepsilon}^{c+\varepsilon} f(x)dx = 0$$

Når X er kontinuerlig fordelt, har vi at

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b)$$

Slik er det ikke hvis X er diskret fordelt

Eksempel: Uniform fordeling
Detaljer på forelesningen (jf. læreboka side 161)

Eksempel: Tid mellom biler
Detaljer på forelesningen (jf. eksempel 4.5 i læreboka)

For en kontinuerlig stokastisk variabel X er den kumulative fordelingsfunksjonen gitt ved

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

Omvendt har vi at $F'(x) = f(x)$

Merk at: $P(X > a) = 1 - F(a)$

$$P(a \leq X \leq b) = F(b) - F(a)$$

Eksempel: Uniform fordeling (forts)
Detaljer på forelesningen (jf. eksempel 4.6)

Eksempel: Tid mellom biler (forts)
Detaljer på forelesningen

La p være et tall mellom 0 og 1

100p-persentilen i fordelingen til X kaller vi $\eta(p)$

Den er gitt ved:

$$p = F[\eta(p)] = \int_{-\infty}^{\eta(p)} f(y)dy$$

Persentilene kalles også kvantiler eller fraktiler

Medianen $\tilde{\mu}$ er 50-persentilen, så $0.50 = F(\tilde{\mu})$

Eksempel: Tid mellom biler (forts)
Detaljer på forelesningen

La X være en kontinuerlig stokastisk variabel med sannsynlighetstetthet $f(x)$. Da er forventningsverdien (forventningen) til X gitt ved

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

Forventningsverdien eksisterer så sant

$$\int_{-\infty}^{\infty} |x| f(x)dx < \infty$$

Eksempel: Tid mellom biler (forts)
Detaljer på forelesningen

Regneregler for forventning

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

$$E(aX + b) = a \cdot E(X) + b$$

(Bevis av siste resultat på forelesningen)

Eksempel: "Broken stick" model
Detaljer på forelesningen (jf. eksempel 4.11)

La X være en kontinuerlig stokastisk variabel med sannsynlighetstetthet $f(x)$ og forventnings-verdi μ. Da er variansen til X gitt ved

$$\sigma_X^2 = V(X) = E\left\{(X-\mu)^2\right\} = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$$

Standardavviket til en stokastisk variabel X er gitt ved $\sigma_X = SD(X) = \sqrt{V(X)}$

Regneregler for varians (bevis på forelesningen)

$$V(X) = E(X^2) - [E(X)]^2$$

$$V(aX + b) = a^2 \cdot V(X)$$

Tilnærmelser for forventning og varians

La X være en stokastisk variabel med forventning μ og varians σ^2

Vi antar at σ^2 er «liten», slik at X med stor sannsynlighet er «nær» μ

Ved en Taylor utvikling har vi da at

$$h(X) \approx h(\mu) + h'(\mu)(X - \mu)$$

Derfor er

$$E[h(X)] \approx h(\mu)$$

$$V[h(X)] \approx [h'(\mu)]\sigma^2$$

Momentgenererende funksjon (mgf)

La X være en kontinuerlig stokastisk variabel med sannsynlighetstetthetsynlighet $f(x)$

Den momentgenererende funksjon for X er gitt ved

$$M_X(t) = E\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)dx$$

Den momentgenererende funksjonen eksisterer hvis det fins et tall $t_0 > 0$ slik at

$$\int_{-\infty}^{\infty} e^{tx} \cdot f(x)dx < \infty \quad \text{for alle} \quad t \in (-t_0, t_0)$$

Eksempel 4.15 i læreboka
Detaljer på forelesningen

Momentgenererende funksjoner for kontinuerlige stokastiske variabler har tilsvarende egenskaper som for diskrete stokastiske variabler.

Spesielt: Hvis $M_X(t) = M_Y(t)$ så har X og Y samme fordeling.

Andre viktige egenskaper (bevis på forelesningen)

1) $M_X(0) = 1$
2) $M_X^{(r)}(0) = E(X^r)$
3) Hvis $R_X(t) = \ln\{M_X(t)\}$, så er
   $R_X'(0) = E(X)$ og $R_X''(0) = V(X)$

Eksempel 4.17 i læreboka
Detaljer på forelesningen