

Diskrete og kontinuerlige variable

Det er egentlig ikke nødvendig med noe skille mellom diskrete og kontinuerlige stokastiske variable, men man gjør det allikevel av pedagogiske grunner. Hvis du triller en terning, må den lande på en av de seks sidene, og alle barn i barnehagen forstår at sannsynligheten er $1/6$ for at den lander på en bestemt side. Hvis man derimot skulle starte med den kontinuerlige stokastiske variabelen X , definert $i - 1/2 < X < i + 1/2 = i$ øyne, sannsynlighetstetthetsfunksjonen $f(x) = 1/6$, og definert

$$\begin{aligned} P(i \text{ øyne}) &= P(i - 1/2 < X < i + 1/2) \\ &= \int_{i-1/2}^{i+1/2} f(x) dx = \frac{1}{6}, \end{aligned}$$

ville mange falt av lasset ganske fort. Men får fint å gjøre det, og således er diskrete variable kun spesialtilfeller av de kontinuerlige.

Regneregler for forventning og varians

Vi kommer til å trenge noen viktige regneregler for forventning og varians. Dersom X er en stokastisk variabel, og a og b reelle tall, gjelder.

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

Dersom X og Y er identisk fordelte stokastiske variable, gjelder

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

Bevisene for disse ligningene er trivielle, både for diskrete og kontinuerlige variable. For variansen gjelder at

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

og

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

dersom X og Y er uavhengige. Vi beviser den første. La $\mu = E(X)$. Da blir $a\mu + b = E(aX + b)$, og

$$\begin{aligned} \text{Var}(aX + b) &= \int_{\Omega} (ax + b - a\mu - b)^2 f(x) dx \\ &= a^2 \int_{\Omega} (x - \mu)^2 f(x) dx = a^2 \text{Var}(X). \end{aligned}$$

Til slutt en formel som kan være grei å bruke i blant

$$\text{Var}(X) = (E(X - \mu)^2) = E(X^2) - \mu^2.$$

Vi beviser denne også. Husk at $\int_{\Omega} f(x) dx = 1$.

$$\begin{aligned} \text{Var}(X) &= \int_{\Omega} (x - \mu)^2 f(x) dx \\ &= \int_{\Omega} (x^2 - 2x\mu + \mu^2) f(x) dx \\ &= \int_{\Omega} x^2 f(x) dx - 2\mu \int_{\Omega} x f(x) dx + \mu^2 \int_{\Omega} f(x) dx \\ &= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2. \end{aligned}$$

Eksempel. Jeg har en gammel strekk i foten. Modellen for hvor lenge jeg holder ut med en arbeidsoppgave kan også brukes til å beskrive antall kilometer jeg holder ut på løpetur før foten blir vond å løpe på. Torkil er mye sprekere enn meg, så han løper dobbelt så fort, men han har den samme skaden, så han holder ikke på så lenge han heller. Så om vi skal på løpetur gir vi meg et forsprang på en kilometer. En stokastisk variabel som beskriver hvor langt han løper dersom jeg løper X km, er

$$Y = 2X - 1.$$

Hans forventede løpslengde blir

$$E(Y) = E(2X - 1) = 2E(X) - 1 = 2 \cdot \frac{4}{3} - 1 = \frac{7}{3} \triangle$$

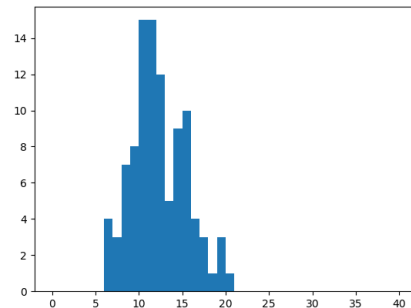
Sentralgrenseteoremet

Pythonkoden

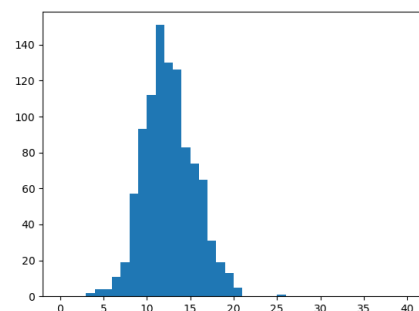
```
import numpy as np
import matplotlib.pyplot as plt

p=.3
n=40
x=np.random.binomial(n, .3, 100)
plt.hist(x, bins=np.arange(n+1))
plt.show()
```

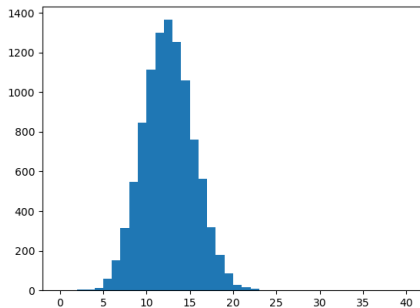
kjører hundre binomiske forsøk med $n = 40$ og $p = 0.3$. Her er et histogram av alle utfallene.



Kjører vi det binomiske eksperimentet tusen ganger, får vi histogrammet



For moro skyld kjører vi det binomiske eksperimentet ti tusen ganger, og får



Merk hvordan utfallene fordeler seg på en måte som ligner mistenkelig på normalfordelingskurven. Sentralgrenseteoremet forteller oss at dette alltid skjer.

Teorem. Dersom X_1, X_2, \dots, X_n er identisk fordelte variable med forventning μ og standardavvik σ , er

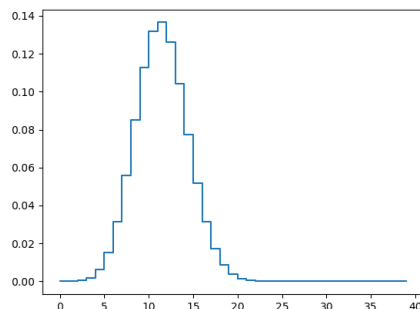
$$X = \frac{1}{n} \sum_{i=1}^n X_i$$

tilnærmet normalfordelt med forventning μ og standardavvik σ/\sqrt{n} .

Beviset skal vi ikke ta, det tar minst en A4-side å skrive opp. Merk at X_i ikke trenger være normalfordelt, jf. eksemplet over.

Binomisk og normal

Følgende plot viser sannsynlighetene for et binomisk forsøk med $n = 40$ og $p = 0.3$. Merk at $\mu = 40 \cdot 0.3 = 12$ og $\sigma = 40 \cdot 0.3 \cdot 0.7 = 8.4$. Dersom n er stor,



er binomisk fordeling tilnærmet normalfordelt med $\mu = np$ og $\sigma = np(1-p)$. Dette er nyttig dersom du ønsker å finne kumulative binomiske sannsynligheter, og ikke kan programmere.

Binomisk og poisson

Hvis man lar $p \rightarrow 0$ og $n \rightarrow \infty$ mens np er konstant, vil binomisk fordeling \rightarrow poissonfordelingen.

Poisson og eksponensialfordelingen

Ventetiden mellom hendelser i en poissonprosess er eksponensialfordelt. La X være en poissonfordelt variabel med forventning $\lambda = \alpha t$, og T en eksponensialfordelt variabel med parameter α . Parameteren α beskriver antall forekomster per tidsenhet i poissonprosessen, og dette er da enheten på tidsaksen i eksponensialfordelingen. En lett måte å huske dette på, er å huske at

$$P(X = 0) = e^{-\alpha t} = P(T > t).$$

Eksempel. Ballspark IL scorer vanligvis to mål per kamp. Sannsynligheten for at de scorer ett mål på en kamp er:

$$P(X = 1) = 2e^{-2} = 0.27.$$

Sannsynligheten for at de ikke scorer mål i første omgang er:

$$P\left(T > \frac{1}{2}\right) = e^{-1} = 0.37. \quad \triangle$$