Sammenheng mellom diskrete og kontinuerlige variable

Det er egentlig matematisk unødvendig 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. Tilsvarende formler

$$Var(aX + b) = a^2 Var(X)$$

og

$$Var(X + Y) = Var(X) + Var(Y)$$

gjelder for variansen. Vi beviser den første for det kontinuerlige tilfellet. La $\mu = E(X)$. Da blir $a\mu + b = E(aX + b)$, og

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

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

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

Vi beviser denne for den kontinuerlige tilfellet. Husk at $\int_\Omega f(x)\,dx = 1$.

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

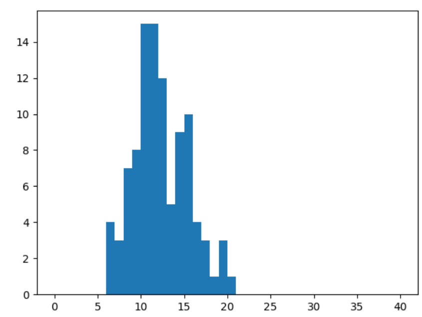
Sentralgrenseteoremet

Pythonkoden

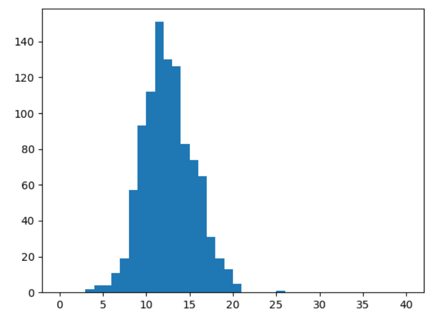
```
import numpy as np
import matplotlib.pyplot as plt

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

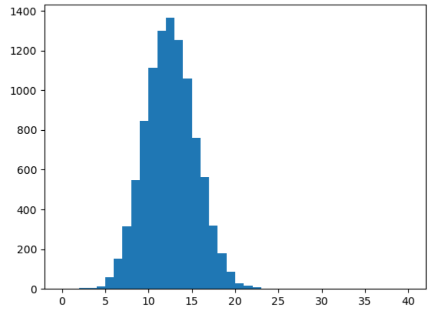
kjører hundre binomiske forsøk med $n = 40$ og $p = 0.3$. Her er et histogram av alle utfallene.



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.

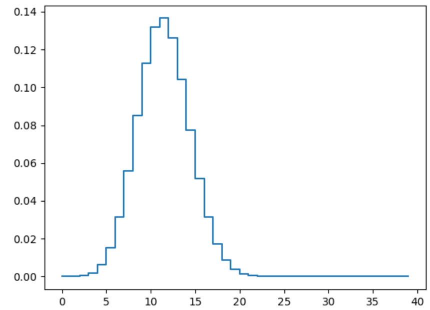
Teorem. *Dersom X_1, X_2,...,X_n er identisk fordelte variable med forventning μ og standardavvik σ, er*

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

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

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 $p \to 0$ og $n \to \infty$ mens np er konstant, vil binomisk fordeling $\to$ poissonfordelingen.

Poisson og eksponensialfordelingen