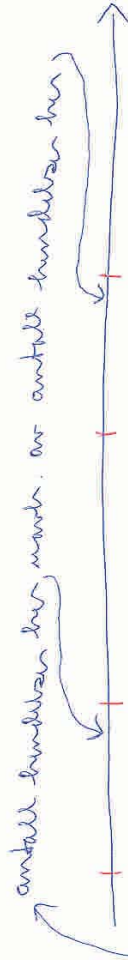


Poisson-fordeling (5.6)

Typisk for Poisson-forsøk er at vi ser på antall forekomster av en hendelse innenfor et tidsintervall eller et spesifisert område.

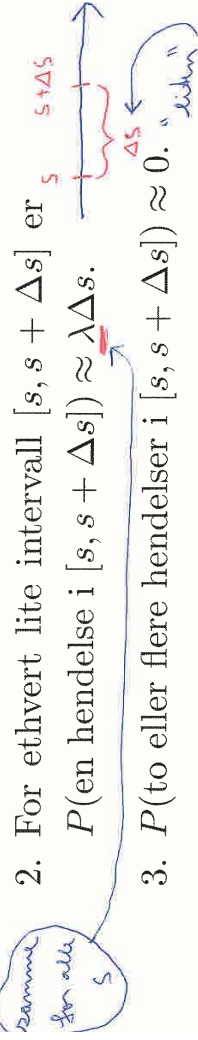
Eksempler på ting som typisk er Poisson-fordelt:

- Antall telefoner til et kontor i løpet av en time.
- Antall visninger av en webside per time.
- Antall trafikkulykker på en veistrekning i løpet av ett år.
- Antall utrykninger fra en brannstasjon per uke.
- Antall trykkfeil per side i en bok.
- Antall trær per mål i en skog. $\square$



Forekomsten av hendelser følger en *Poisson-prosess* dersom:

1. Antallene hendelser i ikke-overlappende intervaller er uavhengige.



2. For ethvert lite intervall $[s, s + \Delta s]$ er $P(\text{en hendelse i } [s, s + \Delta s]) \approx \lambda \Delta s$.

3. $P(\text{to eller flere hendelser i } [s, s + \Delta s]) \approx 0$. "liten"

Dersom dette er oppfylt er $X =$ "antall hendelser i $[0, t]$ " Poisson-fordelt med parameter $\mu = \lambda t$.

Det kan da vises at:

$$f(x) = p(x; \lambda t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}, \quad x = 0, 1, 2, \dots$$

og at $E(X) = \lambda t$ og $\text{Var}(X) = \lambda t$.

Parameteren λ representerer forventet antall hendelser per enhet og kalles *intensiteten* til prosessen.

Vi skal lære mer om Poisson-prosesser senere i kurset.

NB! Vi kan også ha Poisson-fordeling i andre situasjoner (men generelle Poisson-prosesser etc.)

Poisson-fordeling som tilnærming til binomisk fordeling.

$$n > 50$$

$$p < 0.05$$

Når n er stor og p nær 0 (eller 1) kan binomisk fordeling tilnærmes med Poisson-fordeling med $\mu = np$. (Nyttig fordi det gir enklere regning).

Eksempel: Antall vinnerrekker i lotto. Sanns. → 7 rulle

for at en rekke gir 7 rette er $p = 1 / \binom{34}{7} =$

$1/5379616$. Antall innleverte rekker en uke er ca $n = 23500000$. La $X =$ "antall rekker som har 7 rette". Vi har da at $X \sim B(n, p) \approx \text{Poisson}(\mu)$, der $\mu = np = 23500000/5379616 = 4.37$. Dvs:

$$f(0) = P(X=0) \approx \frac{4.37^0}{0!} e^{-4.37} = 0.013$$

$$f(1) = P(X=1) \approx \frac{4.37^1}{1!} e^{-4.37} = 0.055$$

x	0	1	2	3	4
$f(x)$	0.013	0.055	0.121	0.176	0.192

x	5	6	7	8	9
$f(x)$	0.168	0.122	0.076	0.042	0.020

Eksempel

Ventilen i flyktetaster for bruk.

Vet at 10% ikke taster flyktetaster

Uavh. testresultat.

La $Y =$ antall av 12 ventiler som ikke taster taster

$X =$ antall ventiler som vi taster for man har funnet 10 som taster taster

Fordeling?

Y : i) 12 uavh. enkeltforsøk

ii) taster / ikke taster

(i) $p = 0.10$ for ikke taster i alle forsøk

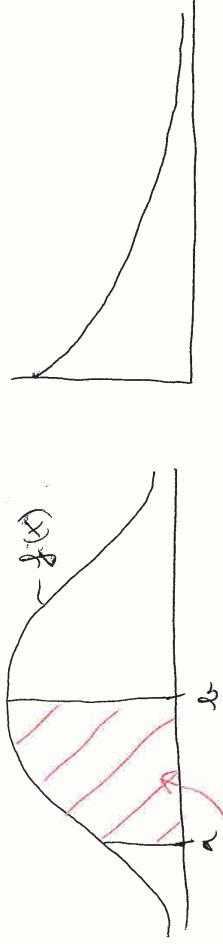
$$\Rightarrow Y \sim B(12, 0.10)$$

X : i) Uavh. enkeltforsøk, imidlertid 10 "taster"

ii) taster / ikke taster

(i) $p = 0.90$ for "taster" i hvert forsøk

$$\Rightarrow X \sim \text{mønstret binomisk}, k_2 = 10, p = 0.90$$



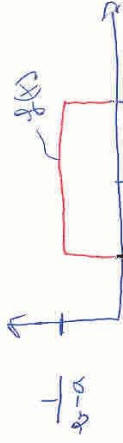
Kontinuerlige sannsynlighetsfordelinger.

Dekkes av kap. 6 og deler av kap. 8.5 i boka.

Husk: $f(x)$ er sannsynlighetstettheten til en kontinuerlig X dersom:

1. $f(x) \geq 0$ for alle $x \in \mathcal{R}$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. $P(a < X < b) = \int_a^b f(x) dx$

Vi skal nå se på noen av de mest brukte kontinuerlige sannsynlighetsfordelingene. (For hver slik fordeling er $f(x)$ gitt ved en bestemt parametrisert funksjon).



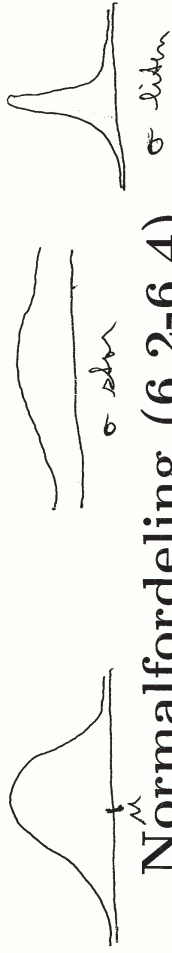
Kontinuerlig uniformfordeling (6.1)

Dersom X kun kan anta verdier i et intervall $[a, b]$, og det er like sannsynlig at X faller i et hvilket som helst område av samme utstrekning innen dette intervallet er X uniformfordelt (f.eks. hvor på en wire et brudd oppstår). Da er

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

og det kan vises at $E(X) = \frac{(a+b)}{2}$ og $\text{Var}(X) = \frac{(b-a)^2}{12}$. $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \dots = \frac{a+b}{2}$

Kontinuerlig uniformfordeling brukes mye i forbindelse med simulering. De fleste randomgeneratorene genererer uniformfordelte tall mellom 0 og 1. Ved hjelp av transformasjoner o.l. kan man ut fra dette få generert tall fra andre sannsynlighetsfordelinger.



Normalfordeling (6.2-6.4)

Normalfordelingen er den viktigste og mest brukte av alle sannsynlighetsfordelinger. En lang rekke fenomener i natur, teknologi, økonomi, etc kan modelleres ved en normalfordeling. Normalfordeling er også svært sentral i statistisk teori pga at mange fordelinger under visse betingelser kan tilnærmes til normalfordeling.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right), \quad -\infty < x < \infty,$$

$E(X) = \mu$ og $\text{Var}(X) = \sigma^2$. Vi skriver at $X \sim N(\mu, \sigma^2)$.

Siden $f(x)$ er komplisert å integrere, regner vi ut normalfordelings sannsynligheter ved å gjøre transformasjonen

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} = \frac{X - \mu}{\sigma}$$

der Z er standard normalfordelt (sjekk selv at $E(Z) = 0$ og $\text{Var}(Z) = 1$) og så bruke tabell

over standard normalfordelingen. F.eks.

$$P(X < a) = P\left(\frac{X - \mu}{\sigma} < \frac{a - \mu}{\sigma}\right) = P\left(Z < \frac{a - \mu}{\sigma}\right)$$

der $P(Z < \frac{a-\mu}{\sigma})$ finnes fra tabellen. (Går ikke nærmere inn på dette da det antas kjent fra før.)

Sentralgrenseteoremet (8.5)

Summer/gjennomsnitt av mange uavh. stok. var. (fra en hvilken som helst fordeling!) vil være tilnærmet normalfordelt!

La $X_1, \dots, X_n$ være uavh. variable fra en fordeling med forventning μ og varians σ^2 . Vi har da (sjekk selv!) at $E(\bar{X}) = \mu$ og $\text{Var}(\bar{X}) = \sigma^2/n$.

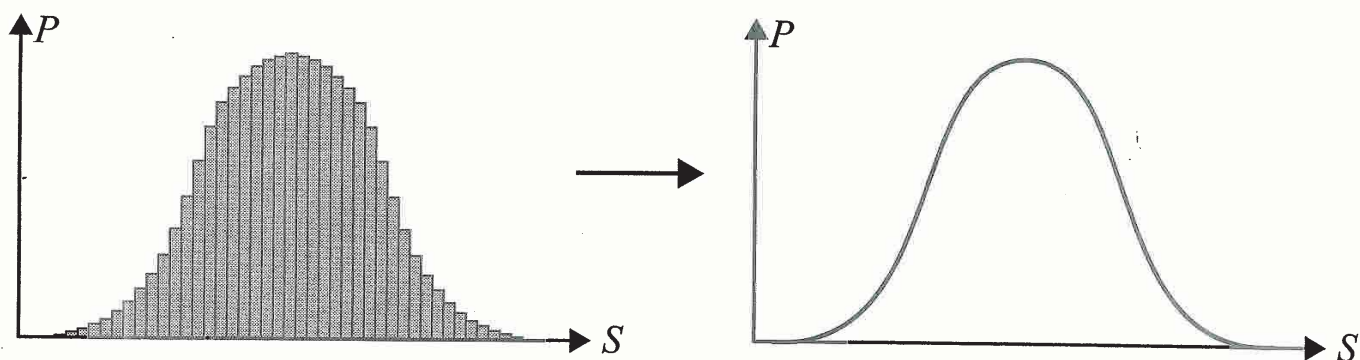
Sentralgrenseteoremet (SGT) sier nå at

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \approx N(0, 1)$$

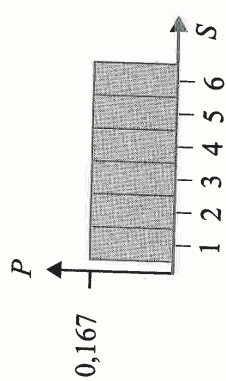
Generelt OK tilnærming når $n \geq 30$.

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

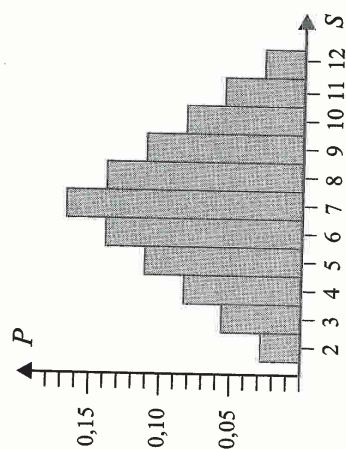
$$\bar{X} \approx \frac{1}{n} \sum_{i=1}^n X_i$$



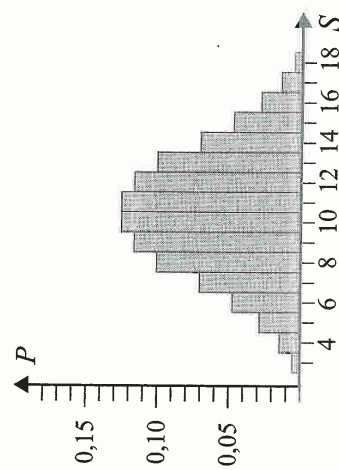
Figur 2.30 Sannsynlighetsfordeling for kast med åtte terninger og normalfordeling



Figur 2.27 Sannsynlighetsfordeling for kast med en terning



Figur 2.28 Sannsynlighetsfordeling for kast med to terninger



Figur 2.29 Sannsynlighetsfordeling for kast med tre terninger

Vi ser at dette også gir at

$$Z = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^n X_i - E(\sum_{i=1}^n X_i)}{\sqrt{\text{Var}(\sum_{i=1}^n X_i)}} \approx N(0, 1)$$

Dvs, generelt

$$Z = \frac{V - E(V)}{\sqrt{\text{Var}(V)}} \approx N(0, 1)$$

der $V = \bar{X}$ eller $V = \sum_{i=1}^n X_i$.

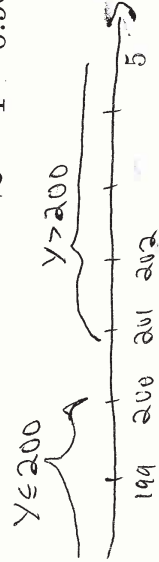
Eksempel: Du kaster en terning 50 ganger.

Hva er sannsynligheten for at summen av alle terningkastene blir større enn 200?

La X_i være resultatet av det i te terningkast og $Y = \sum_{i=1}^{50} X_i$. Vi vet at (se diskret uniform fordeling) $E(X_i) = 3.5$ og $\text{Var}(X_i) = 2.92$.

$$\begin{aligned} P(Y > 200) &= 1 - P(Y \leq 200) \\ &= 1 - P\left(\frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}} \leq \frac{200 - E(Y)}{\sqrt{\text{Var}(Y)}}\right) \\ &= 1 - P\left(Z \leq \frac{200 - 50 \cdot 3.5}{\sqrt{50 \cdot 2.92}}\right) \\ &= 1 - P(Z \leq 2.07) \end{aligned}$$

$$\begin{aligned} SGT &\approx 1 - 0.981 = \underline{\underline{0.019}} \quad \square \end{aligned}$$



Binomisk fordeling tilnærmet med normalfordeling (6.5)

X = "antall suksesser i n forsøk" er en sum (la "suksess" = 1 og "ikke suksess" = 0) og kan dermed i følge SGT tilnærmes med normalfordeling. Når $X \sim B(n, p)$ husker vi at $E(X) = np$ og $\text{Var}(X) = np(1-p)$, og vi får dermed at

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \approx N(0, 1)$$

$\nwarrow E(X)$
 $\nwarrow \text{Var}(X)$

OK tilnærming når $np > 5$ og $n(1-p) > 5$. Se figurene 6.22-6.25 i boka. $E_{nt.} \text{ } np(1-p) > 5$

$$\begin{aligned} P(a \leq X \leq b) &\approx P\left(\frac{a - 0.5 - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right) \\ &= P\left(Z \leq \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right) - P\left(Z \leq \frac{a - 0.5 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Her er -0.5 i nedre grense og +0.5 i øvre grense korreksjoner vi gjør for å få en bedre tilnærming (se fig. 6.23).