# Normalfordelingsapproksimationen

En stokastisk variabel $X$ er binomialfordelt med sandsynlighedsparameter $p$ og antalsparameter $n$ hvis

$$P(X=k)=\left(\genfrac{}{}{0pt}{}{n}{k}\right)·p^{k}·\left(1-p\right)^{n-k}$$

Vi har desuden en kort notation: $X\~B\left(n,p\right)$.

Man kan vise at middelværdien er $n⋅p$ og spredningen er $\sqrt{n⋅p⋅\left(1-p\right)}$.

Hvis $n·p\geq 5$ og $n·\left(1-p\right)\geq 5$, så er normalfordelingen med samme middelværdi og spredning en god approksimation til binomialfordelingen.

### Opgave 1

Undersøg hvor god en approksimation normalfordelingen er til binomialfordelingen.
F.eks. ved at beregne $P\left(21\leq X\leq 39\right)$ hvis $X$ er binomialfordelt med sandsynlighedsparameter $p=0,3$ og antalsparameter $n=100$ og sammenligne med hvad normalfordelingen giver.

*Tip: det var noget med et sumtegn når man skulle bestemme sandsynligheden for flere udfald med binomialfordelingen.*

### Opgave 2

Nedenfor ses en binomialtest med sandsynlighedsparameter $p=0,3$, antalsparameter $n=100$ og et signifikansniveau på $5 \%$.

1. Hvordan er acceptområdet i testen bestemt?
2. Approksimér acceptområdet vha. de normale udfald i normalfordelingen.
3. Find på et eksempel hvor testen kunne bruges.



Lad $X$ være binomialfordeling med sandsynlighedsparameter $p$ og antalsparameter $n$. Da binomialfordelingen kan approksimeres med normalfordelingen, har vi fra normalfordelingen at

$$P\left(X-2σ\leq μ\leq X+2σ\right)≈95 \%$$

Vi er interesseret i at lave et konfidensinterval for sandsynlighedsparameteren $p$. Hvis vi indsætter middelværdi og spredning, får vi:

$$P\left(X-2⋅\sqrt{n⋅p⋅\left(1-p\right)}\leq n⋅p\leq X+2⋅\sqrt{n⋅p⋅(1-p)}\right)≈95 \%$$

$$P\left(\frac{X}{n}-2⋅\frac{\sqrt{n⋅p⋅\left(1-p\right)}}{n}\leq p\leq \frac{X}{n}+2⋅\frac{\sqrt{n⋅p⋅(1-p)}}{n}\right)≈95 \%$$

Hvis vi har en stikprøve $x$ af vores binomialfordelte $X$ indfører vi $\hat{p}=\frac{x}{n}$ og kan estimere et $95 \%$ konfidensinterval for $p$ således:

$$\left[\hat{p}-2⋅\frac{\sqrt{n⋅\hat{p}⋅\left(1-\hat{p}\right)}}{n},\hat{p}+2⋅\frac{\sqrt{n⋅\hat{p}⋅\left(1-\hat{p}\right)}}{n} \right]$$

Vi har tidligere set at $\hat{p}=\frac{x}{n}$ er den sandsynlighedsparameter hvor sandsynligheden for at få stikprøven $x$ er størst, så det giver god mening at den er midten af vores estimat. Intervallet kan omskrives til det nedenstående (som vi har brugt tidligere):

$$\left[\hat{p}-2⋅\sqrt{\frac{\hat{p}⋅\left(1-\hat{p}\right)}{n}},\hat{p}+2⋅\sqrt{\frac{\hat{p}⋅\left(1-\hat{p}\right)}{n}} \right]$$

### Eksempel

Megafon laver en opinionsundersøgelse i forhold til næste folketingsvalg hvor de spørger 1000 personer og 480 svarer at de ville stemme på rød blok. Hvis vi antager at de 1000 personer er udvalgt tilfældigt blandt vælgerne, dvs. de udgør en repræsentativ stikprøve, så har vi antallet som svarer at de stemmer på rød blok er binomialfordelt med antalsparameter $n=1000$ og en ukendt sandsynlighedsparameter $p$ som angiver andelen af vælgerne som svarer at de stemmer på rød blok. Vi har desuden vores stikprøve
$x=480$. Dermed har vi estimatet

$$\hat{p}=\frac{480}{1000}=0,48$$

Som fortæller at vi ud fra stikprøven estimerer at $48 \%$ af vælgerne svarer at de stemmer på rød blok. Hvis vi gerne vil vide hvor sikre vi er på dette estimat, så kan vi lave et $95 \%$ konfidensinterval for $p$:

$$\left[0,48-2⋅\sqrt{\frac{0,48⋅\left(1-0,48\right)}{1000}},0,48+2⋅\sqrt{\frac{0,48⋅\left(1-0,48\right)}{1000}} \right]=[0,448;0,512]$$

Dermed har vi estimeret et $95 \%$ konfidensinterval for $p$ til mellem $44,8 \%$ og $51,2 \%$.

### Opgave 3

1. Afgør vha. konfidensintervallet i eksemplet ovenfor om nulhypotesen nedenfor kan forkastes på et $5 \%$ signifikansniveau.

$50 \%$ *af vælgerne svarer at de vil stemme på rød blok.*

1. Hvordan ville man vha. et konfidensinterval afgøre om en nulhypotese kunne forkastes på et $1 \%$ signifikansniveau?