11. Binomial- og normalfordeling

Normalapproksimationen

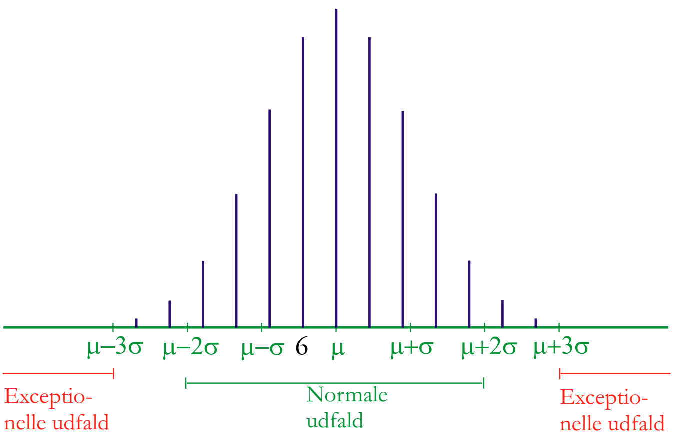
Som det vil være fremgået af det foregående, er der mange fælles træk mellem binomialfordelingen og normalfordelingen. Bl. a. er det tydeligt, at der er mange fælles træk i den måde, man beregner sandsynligheder på. Den mest markante forskel er, at binomialfordelingen er en *diskret* fordeling, mens normalfordelingen er en *kontinuert fordeling*.

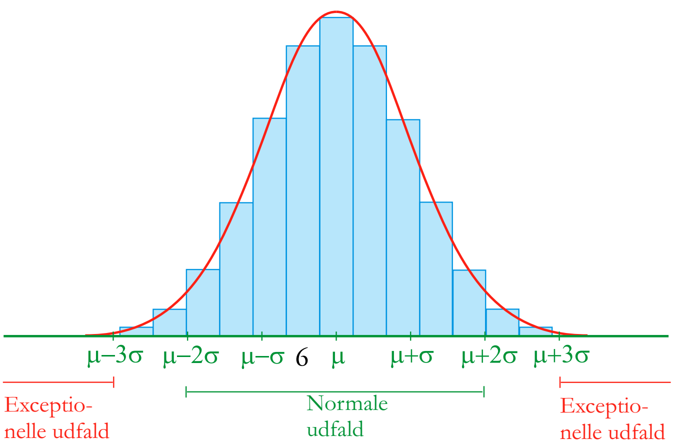
På tegningerne nedenfor ser vi øverst et stolpediagram for en binomialfordeling. Nedenunder har vi bredt stolperne ud, så de danner et histogram. Alle rektanglerne har bredden 1, så deres arealer er lig med deres højder. Den glatte kurve er tæthedskurven for en normalfordeling med binomialfordelingens middelværdi $\mu = np$ og spredning $\sigma = \sqrt{np(1-p)}$. Det er tydeligt, at de to fordelinger ligner hinanden meget. Ser vi fx på $P(X \leq 6)$ i binomialfordelingen, er denne sandsynlighed lig med sum-

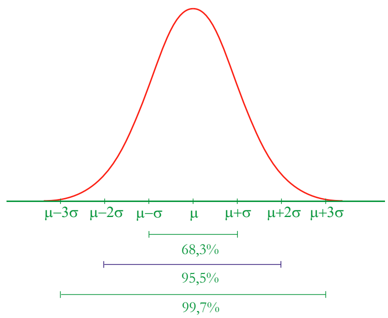
Hvornår binomialfordeling

1. Der foreligger et eksperiment med netop to mulige udfald: Succes eller ikke succes.
2. Sandsynligheden p for succes er konstant.
3. De enkelte udførelser af eksperimentet er uafhængige af hinanden.

Hvis disse betingelser er opfyldt, er antallet af succeser ved n udførelser af eksperimentet binomialfordelt $b(n,p)$.



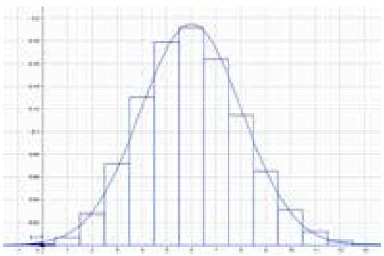




Søjlediagram for en binomialfordeling overlejret med tæthedskurven for en normalfordeling med binomialfordelingen middelværdi og spredning men af arealerne af de seks rektangler længst til venstre. Det er meget tæt på at være det samme som arealet under normalfordelingskurven til venstre for 6. Dette areal er jo netop værdien $F(6)$ fra normalfordelingens fordelingsfunktion. Normalfordelingens fordelingsfunktion giver en god tilnærmelse til den kumulerede binomialsandsynlighed.

Mange statistiske modeller bygger på teorien for normalfordelte observationer. Selv om mange observationer kommer fra binomialfordelinger, er det altså muligt at udnytte normalfordelingsmodellerne. Man kan benytte binomialfordelingens middelværdi og spredning, og derefter anvende normalfordelingens fordelingsfunktion.

Normalfordelingen er en symmetrisk fordeling, og det er binomialfordelingen ikke altid. Hvis p er meget lille eller meget stor, bliver binomialfordelingen skæv, og er derfor vanskelig at approksimere (tilnærme) med en normalfordeling. En meget brugt tommelfingerregel er, at når $np > 5$ *og* $n(1-p) > 5$, kan man bruge approksimationen. Der gælder:



Sådan ser det ud, når p = 6 og n = 20. Man kan godt se, at binomialfordelingen er lidt skæv, men approksimationen er ganske god

Sætning 11.1: Normalapproksimationen

For en binomialfordeling med antalsparameter n og sandsynligheds-parameter p gælder med god nøjagtighed, hvis $np > 5$ og $n(1-p) > 5$:

Antal succeser kan regnes normalfordelt $N(\mu = np, \sigma = \sqrt{np(1-p)})$.

Altså en normalfordeling med binomialfordelingens middelværdi og spredning.

Eksempel 11.2: Falsk terning

En skummel person har fremstillet en falsk terning, som har en sandsynlighed på 0,175 for at slå en sekser (en normal terning har sandsynligheden $1/6 \approx 0{,}167$). Den falske terning kastes 6000 gange. Antal seksere er binomialfordelt $b(n = 6000, p = 0{,}175)$.

Denne fordeling har middelværdien $\mu = np = 1050$ og spredningen:

$$\sigma = \sqrt{np(1-p)} = \sqrt{6000 \cdot 0{,}175 \cdot 0{,}825} = 29{,}43$$

1. Sandsynligheden for, at der fremkommer 1000 eller flere seksere, beregnes ved hjælp af normalapproksimationen til $1 - P(X \leq 999) = 0{,}9584$.

Med den eksakte binomialfordeling bliver resultatet 0,9576.