

Note: Binomialfordeling med Nspire

KBJ, november 2019

Binomialfordelingens grundlag

Et sandsynlighedseksperiment består af gentagelser af et basiseksperiment E .

Ved basiseksperimentet er der fokus på én særlig hændelse, basishændelsen H .

Sandsynligheden for H når E udføres er hver gang p , kaldet *sandsynlighedsparameteren*.

Basiseksperimentet E udføres n gange, kaldet *antalsparameteren*.

Den stokastiske variabel X angiver antal gange H sker, når E udføres n gange.

Udfaldsrummet for X er da tallene $0, 1, 2, \dots, n$, idet H mindst sker 0 gange og højest n gange.

Sandsynlighedsfordelingen for X er da bestemt ved udtrykket:

$$P(X = r) = K(n, r) \cdot p^r \cdot (1 - p)^{n-r}$$

Hvor $K(n, r)$ (*binomialkoefficienten*) er antal kombinationer af r elementer valgt blandt n .

I Nspire kan $K(n, r)$ bestemmes med kommandoen: `nCr(n, r)`

At X er en stokastisk variabel som er binomialfordelt med sandsynlighedsparameter p og antalsparameter n skrives kort: $X \sim b(n, p)$

Eksempel

En stokastisk variabel X er binomialfordelt: $X \sim b\left(40, \frac{1}{3}\right)$.

Sandsynligheden for udfaldet $X = 10$ kan da bestemmes ved i Nspire at skrive:

$$\text{nCr}(40, 10) \cdot \left(\frac{1}{3}\right)^{10} \cdot \left(1 - \frac{1}{3}\right)^{40-10}$$

De fleste steder viser hvid baggrund det indtastede og grå baggrund resultatet ved tryk på "enter":

$$\text{nCr}(40, 10) \cdot \left(\frac{1}{3}\right)^{10} \cdot \left(1 - \frac{1}{3}\right)^{40-10} \quad 0.074864$$

På baggrund af beregningen ses at $P(X = 10) = 0,075$ svarende til 7,5%.

Generelt kan kommandoen `binompdf` også benyttes:

$$P(X = r) = \text{binompdf}(n, p, r)$$

<code>binompdf(40,1/3,10)</code>	
$\text{binomPdf}\left(40, \frac{1}{3}, 10\right)$	0.074864

Hændelser

En hændelse er en delmængde af udfaldsrummet. Ofte ser vi ved binomialfordelinger på delmængder som er sammenhængende intervaller. F.eks. $10 \leq X \leq 13$. Sandsynligheden for hændelsen kaldes da $P(10 \leq X \leq 13)$ og bestemmes som

$$P(10 \leq X \leq 13) = P(X = 10) + P(X = 11) + P(X = 12) + P(X = 13)$$

Med kommandoen `binomcdf` kan Nspire bestemme sandsynlighed for et interval af udfald:

$$P(a \leq X \leq b) = \text{binomcdf}(n, p, a, b)$$

<code>binomcdf(40,1/3,10,13)</code>	
$\text{binomCdf}\left(40, \frac{1}{3}, 10, 13\right)$	0.43315
$\text{binomPdf}\left(40, \frac{1}{3}, 10\right) + \text{binomPdf}\left(40, \frac{1}{3}, 11\right) + \text{binomPdf}\left(40, \frac{1}{3}, 12\right) + \text{binomPdf}\left(40, \frac{1}{3}, 13\right)$	0.43315

For hændelserne $X \leq r$ eller $X \geq r$ benyttes følgende specielle kommandoer:

$$P(X \leq r) = \text{binomcdf}(n, p, 0, r)$$

$$P(X \geq r) = \text{binomcdf}(n, p, r, n)$$

$\text{binomCdf}\left(40, \frac{1}{3}, 0, 15\right)$	0.76884
$\text{binomCdf}\left(40, \frac{1}{3}, 15, 40\right)$	0.342179

Til højre ses at $P(X \leq 15) = 0,769$, svarende til at sandsynligheden for ”højst 15” er 76,9%, samt $P(X \geq 15) = 0,342$, svarende til at sandsynligheden for ”mindst 15” er 34.2%.

Bemærk at kommandoen `binompdf(n, p)` producerer en liste med alle værdier for $P(X = r)$ samt at `binomcdf(n, p)` producerer en liste med alle værdier for $P(X \leq r)$.

Bestemmelse af q -fraktil

Det *mindste* tal a som gør at $P(X \leq a) > q$ kaldes for q -fraktilen og bestemmes med kommando:

$$a = \text{invbinom}(q, n, p)$$

F.eks. kan 25%-fraktilen for $X \sim b\left(40, \frac{1}{3}\right)$ bestemmes som vist herunder til venstre:

$\text{invbinom}(25\%, 40, 1/3)$	
$\text{invBinom}\left(25\%, 40, \frac{1}{3}\right)$	11

$\text{binomCdf}\left(40, \frac{1}{3}, 0, 11\right)$	0.273522
$\text{binomCdf}\left(40, \frac{1}{3}, 0, 10\right)$	0.171435

Det ses altså at 25%-fraktilen for X er 11. Det betyder, at hvis eksperimentet gentages et stort antal gange, så vil X antage værdier på 11 eller lavere ca. 25% af gangene. Som det ses af beregningerne til højre er $P(X \leq 11) = 27,4\% > 25\%$ og $P(X \leq 10) = 17,1\% < 25\%$.

Tallet $a = 11$ er således det mindste tal som gør at $P(X \leq a) > 25\%$.

Det *største* tal b som gør at $P(X \geq b) > q$ svarer til $(1 - q)$ -fraktilen og bestemmes ved:

$$b = \text{invbinom}(1-q, n, p)$$

$\text{invbinom}(95\%, 40, 1/3)$	
$\text{invBinom}\left(95\%, 40, \frac{1}{3}\right)$	18

$\text{binomCdf}\left(40, \frac{1}{3}, 18, 40\right)$	0.083202
$\text{binomCdf}\left(40, \frac{1}{3}, 19, 40\right)$	0.044086

Det vil sige at det største tal b så $P(X \geq b) > 5\%$ er $b = 18$.

Det ses således at $P(X \geq 18) = 8,3\% > 5\%$ og $P(X \geq 19) = 4,4\% < 5\%$.

Middelværdi og spredning

Middelværdien (forventet værdi) for X beregnes: $\mu = n \cdot p$

Spredningen for X beregnes: $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$

Hvis disse skal bruges senere kan de defineres som vist til højre.

$n := 40$	40.
$p := \frac{1}{3}$	0.333333
$m := n \cdot p$	13.3333
$s := \sqrt{n \cdot p \cdot (1 - p)}$	2.98142

For at tegne et søjlediagram over sandsynlighederne i en binomialfordeling skal oprettes to lister. Første liste skal indeholde alle hele tal fra 0 til n og kan i et beregninger-vindue defineres ved:

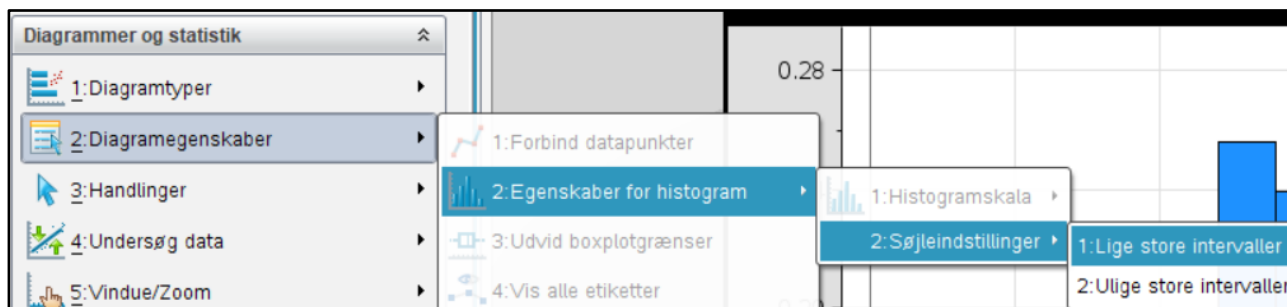
Bemærk at listens navn bliver "udfald". Anden liste skal indeholde sandsynlighederne:

Denne listes navn bliver "sandsynlighed".

```
sandsynlighed:=binompdf(40,1/3)
sandsynlighed:=binomPdf(40, $\frac{1}{3}$ )
{9.04377E-8,0.000002,0.000018,0.00011}
```

På y-aksen højreklikkes og vælges ”Tilføj Y-værdiliste”, hvorefter ”Sandsynlighed” vælges.

Vælg nu "2. Diagramegenskaber" → "2. Egenskaber for histogram" → "2. Søjleindstillinger" → "1. Lige store intervaller", som vist herunder:



Søjlestart: -0,25

Indstillinger for lige store søjlebredder

Bredde

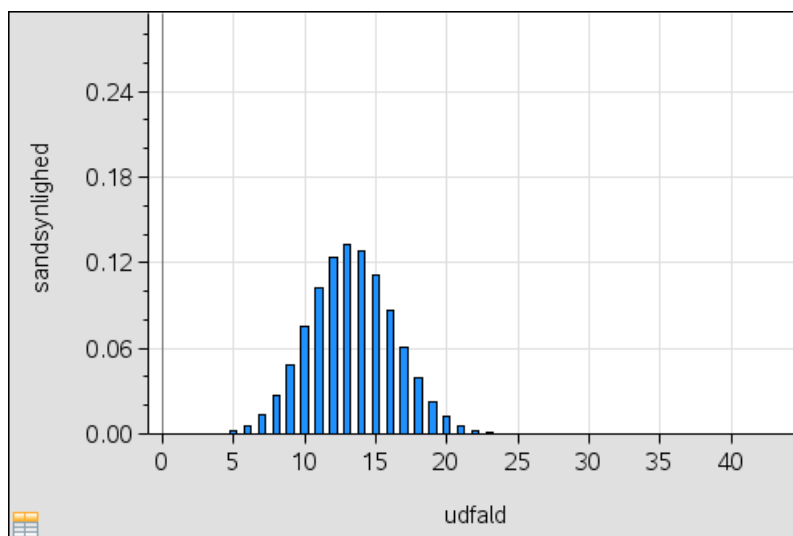
Søjlestart

Herefter tegnes søjlediagram over sandsynlighedernes fordeling mellem de mulige udfald fra 0 til 40, i det højden af søjlen angiver udfaldets sandsynlighed.

Det ses at den højeste søjle findes tæt på middelværdien

$$\mu = 40 \cdot \frac{1}{3} \approx 13,3$$

Det ses at langt fra μ er sandsynlighederne så små, at søjlerne ikke kan ses.



Ved at højreklikke i koordinatsystemet og vælge ”3. Zoom” → ”1. Indstillinger for vindue...” kan det udsnit af koordinatsystemet som ses justeres.



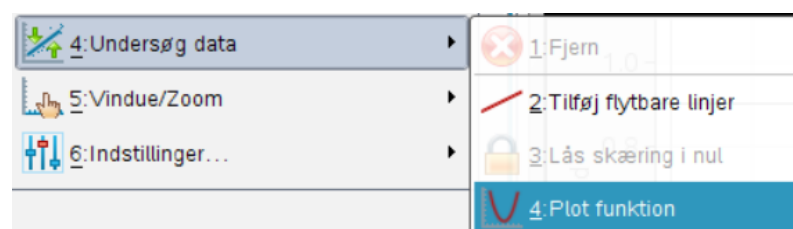
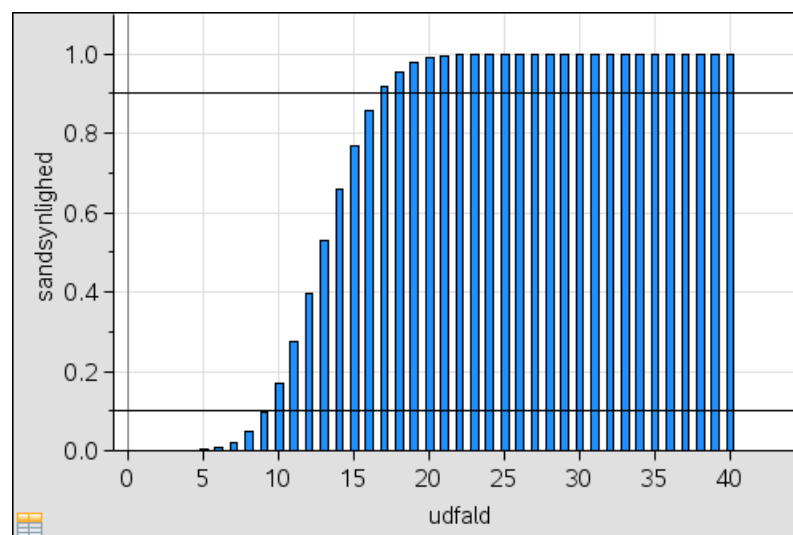
Ønskes i stedet et fordelingsplot over sandsynlighederne for at X mindst er tallene fra 0 til 40, skal listen sandsynligheder i stedet laves med kommandoen:

```
sandsynlighed:=binomcdf(n,p)
```

På figuren til venstre ses fordelingsplottet for $X \sim b\left(40, \frac{1}{3}\right)$.

Husk at YMaks under ”Indstillinger for vindue” (se ovenfor) bør sættes til 1,1.

Under ”4. Undersøg data” → ”4. Plot funktion” kan man tegne grafer for funktioner ved at indtaste forskriften. Indtegnes de vandrette linier $f_1(x) = 0,1$ og $f_2(x) = 0,9$ kan 10%- og 90%-fraktilerne aflæses til henholdsvis 10 og 16.



Binomialtest, signifikansniveau, nulhypotese og acceptmængde

Hvis der skal udføres to-sidet binomialtest på et signifikansniveau S , kan man med fordel bestemme den tilhørende acceptmængde. Det gøres ved at bestemme $\frac{S}{2}$ -fraktilen og $(1 - \frac{S}{2})$ -fraktilen:

$$a = \text{invbinom}(S/2, n, p)$$
$$b = \text{invbinom}(1-S/2, n, p)$$

Da vil acceptmængden for binomialtestet være $[a; b]$. Testet afgør om vi kan *forkaste* eller *ikke-forkaste* (også kaldet *acceptere*) en ved nulhypotesen H_0 antaget sandsynlighed p .

Hvis vores observerede værdi af X ligger i $[a; b]$ vil vi *ikke forkaste* nulhypotesen H_0 .

Hvis X ikke ligger i $[a; b]$ vil vi *forkaste* nulhypotesen H_0 og overveje en alternativ hypotese H_a .

Eksempel: To-sidet test

Vores nulhypotese $H_0: p = \frac{1}{3}$. Vi ønsker at lave et to-sidet test af nulhypotesen på et 5%-signifikansniveau. Vi bestemmer derfor 2,5%-fraktilen og 97,5%-fraktilen med Nspire:

$$\begin{array}{ll} \text{invBinom}\left(0.025, 40, \frac{1}{3}\right) & 8 \\ \text{invBinom}\left(0.975, 40, \frac{1}{3}\right) & 19 \end{array}$$

Det ses at acceptmængden i vores binomialtest er $[8; 19]$. Hvis vi ved eksperimentet observerer $X = 10$, ligger vi i acceptmængden og "accepterer", dvs. undlader at forkaste, H_0 . Hvis vi ved eksperimentet observerer $X = 5$, ligger vi uden for acceptmængden (i "den kritiske mængde") og *forkaster* H_0 . Dermed har vi bekræftet alternativhypotesen $H_a: p \neq \frac{1}{3}$.

Eksempel: En-sidet test

Hvis vores test er en-sidet til venstre, bestemmes S -fraktilen.

Acceptmængden på et 5%-signifikansniveau er altså $[9; 40]$.

Hvis vi forkaster H_0 bekræfter vi dermed $H_a: p < \frac{1}{3}$.

$$\text{invBinom}\left(0.05, 40, \frac{1}{3}\right) \quad 9$$

Hvis testet er en-sidet til højre, bestemmes $(1 - S)$ -fraktilen.

Acceptmængden på et 5%-signifikansniveau er altså $[0; 18]$.

Hvis vi forkaster H_0 bekræfter vi dermed $H_a: p > \frac{1}{3}$.

$$\text{invBinom}\left(0.95, 40, \frac{1}{3}\right) \quad 18$$

Normalfordelingsapproximation

En binomialfordelt stokastisk variabel $X \sim b(n, p)$ er en *diskret* sandsynlighedsfordeling. Det betyder at udfaldsrummet har et endeligt antal udfald. I en del eksperimenter kan udfaldene i princippet være alle tal på hele talaksen – også negative tal, brøker, mv. En sådan fordeling siges at være *kontinueret*. Det mest anvendte eksempel på en sådan fordeling er *normalfordelingen*.

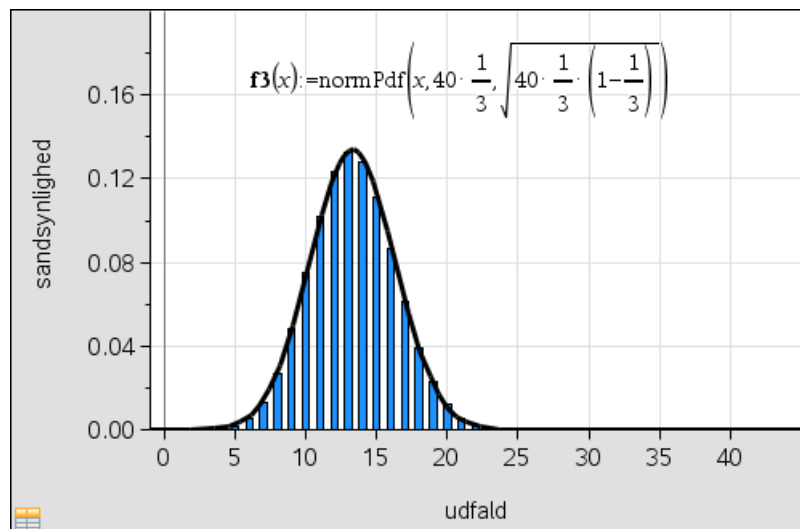
En normalfordeling er kendetegnet ved en *middelværdi* μ og en *spredning* σ . At en stokastisk variabel X er normalfordelt med middelværdi μ og spredning σ skrives: $X \sim N(\mu, \sigma)$.

Tæthedsfunktionen for $X \sim N(\mu, \sigma)$ har forskriften: $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$

I Nspire kan dette skrives: $f(x) = \text{normpdf}(x, \mu, \sigma)$

Det viser sig at en binomialfordelt stokastisk variabel X med middelværdi μ og spredning σ ofte kan beskrives tilnærmelsesvist som om den var normalfordelt $X \sim N(\mu, \sigma)$.

I et søjlediagram over $X \sim b(n, p)$ kan man med ”Plot funktion” (se side 5) indtegne grafen for normalfordelingsapproximationen $X \sim N(\mu, \sigma)$ sammen med søjlerne, som vist til højre.



En tommelfingerregel er at tilnærmelsen kan laves hvis $n \cdot p > 5$ og $n \cdot (1 - p) > 5$.

En normalfordelt stokastisk variabel $X \sim N(\mu, \sigma)$ opfylder følgende:

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95,45\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99,73\%$$

Denne egenskab nedarver $X \sim b(n, p)$ omtrent, hvis den er tilnærmelsesvist normalfordelt.

Vi kalder således intervallet $[\mu - 2\sigma; \mu + 2\sigma]$ for de *normale udfald*. Disse udgør således ca. 95% af alle udfald og svarer stort set til acceptmængden på et 5%-signifikansniveau.

Da kun 0,27% af udfaldene ligger udenfor intervallet $[\mu - 3\sigma; \mu + 3\sigma]$ kaldes sådanne udfald som ligger mere end tre spredninger fra middelværdien for *exceptionelle*.

OVERSIGT OVER NSPIRE-KOMMANDOER

I det følgende er X en binomialfordelt stokastisk variabel $X \sim b(n, p)$ hvor andet ikke fremgår

Beskrivelse	Kommando
Binomialkoefficient $K(n, r)$	<code>ncr(n, r)</code>
Sandsynligheden for netop r : $P(X = r)$	<code>binompdf(n, p, r)</code>
Sandsynligheden for højest r : $P(X \leq r)$	<code>binomcdf(n, p, 0, r)</code>
Sandsynligheden for mindst r : $P(X \geq r)$	<code>binomcdf(n, p, r, n)</code>
Sandsynlighed for mindst a og højest b : $P(a \leq X \leq b)$	<code>binomcdf(n, p, a, b)</code>
Mindste tal a så der gælder $P(X \leq a) > q$ (q -fraktil)	<code>invbinom(q, n, p)</code>
Største tal b så der gælder $P(X \geq b) > q$	<code>invbinom(1-q, n, p)</code>
Acceptmængde $[a; b]$ for to-sidet binomialtest på signifikansniveau S .	<code>a=invbinom(S/2, n, p)</code> <code>b=invbinom(1-S/2, n, p)</code>
Acceptmængde $[a; n]$ for en-sidet (venstresidet) binomialtest på signifikansniveau S .	<code>a=invbinom(S, n, p)</code>
Acceptmængde $[0; b]$ for en-sidet (højresidet) binomialtest på signifikansniveau S .	<code>b=invbinom(1-S, n, p)</code>
Listen "udfald" med tal fra 0 til n .	<code>udfald:=seq(x, x, 0, n)</code>
Listen "sandsynlighed" med alle $P(X = r)$	<code>sandsynlighed:=binompdf(n, p)</code>
Listen "sandsynlighed" med alle $P(X \leq r)$	<code>sandsynlighed:=binomcdf(n, p)</code>
Definere middelværdi $\mu = n \cdot p$ som m :	<code>m:=n*p</code>
Definere spredning $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$ som s	<code>s:=sqrt(n*p*(1-p))</code>
Tæthedsfunktion for normalfordelte $X \sim N(m, s)$:	<code>normpdf(x, m, s)</code>
For $X \sim N(m, s)$ bestemmes $P(X \leq r)$:	<code>normcdf(-infinity, r, m, s)</code>
For $X \sim N(m, s)$ bestemmes $P(X \geq r)$:	<code>normcdf(r, infinity, m, s)</code>
Normale udfald: $[\mu - 2\sigma; \mu + 2\sigma]$	<code>n*p-2*sqrt(n*p*(1-p))</code> <code>n*p+2*sqrt(n*p*(1-p))</code>
Grænse for exceptionelle udfald: $X < \mu - 3\sigma$ eller $X > \mu + 3\sigma$	<code>n*p-3*sqrt(n*p*(1-p))</code> <code>n*p+3*sqrt(n*p*(1-p))</code>