1 Binomialfordelingen

Definition 1 (Basisforsøg/Bernoulliforsøg): Et stokastisk eksperiment, hvortil der kun er to udfald, nemlig "succes" med sandsynligheden p og "fiasko" med sandsynligheden $1-p$, kaldes et **basisforsøg** (eller **basiseksperiment** eller **bernoulliforsøg**). ⧫

Definition 2 (Binomialforsøg): Et stokastisk eksperiment, hvor

1) vi betragter n basisforsøg (det vigtige er et **fast** antal)
2) basisforsøgene er uafhængige (dvs. udfaldene af hvert basisforsøg har ikke indflydelse på hinanden)
3) sandsynligheden for succes er den samme i hvert basisforsøg, nemlig p

kaldes et **binomialforsøg**. ⧫

Det vil sige, at når vi taler om et binomialforsøg, betragter vi:

n **uafhængige** basisforsøg med **samme sandsynlighed** p for succes i hvert basisforsøg.

Vi refererer til p som sandsynligheden for succes i basisforsøget.

Definition 3 (Binomialfordelingen): Lad en stokastisk variabel X betegne antallet af succeser i et binomialforsøg. Så siger vi, at X er binomialfordelt med **sandsynlighedsparameter** p og **antalsparameter** n.

Dette skrives som $X \sim b(n,p)$, hvor $b(n,p)$ kaldes ***binomialfordelingen*** og dækker over sandsynlighedsfordelingen for X, når X betegner antallet af succeser i et binomialforsøg. ⧫

Sætning 4: Lad X være binomialfordelt med sandsynlighedsparameter p og antalsparameter n, dvs. $X \sim b(n,p)$. Hvis r er antallet af succeser, så er

$$P\left(X=r\right) = K\left(n,r\right) \cdot p^r \cdot \left(1-p\right)^{n-r}, \quad r \in \{0,1,2,\ldots,n\},$$

hvor $K\left(n,r\right) = \frac{n!}{r!\cdot(n-r)!}$ kaldes også ***binomialkoefficienten***.

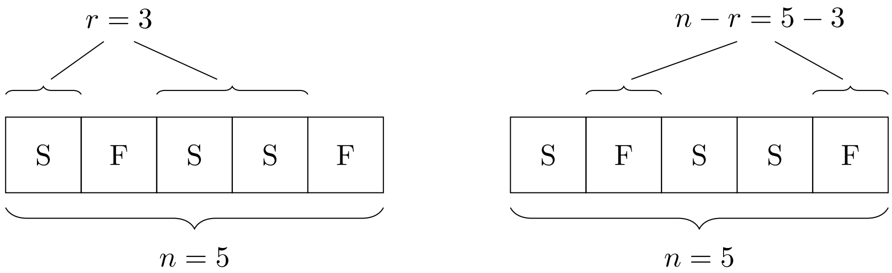
Hvad fortæller sætningen:

Sætningen fortæller hvad sandsynligheden er for at få præcis r succeser i et binomialforsøg.

Bevis (Forklaret i alle detaljer): Sætningen bevises gennem et eksempel, hvor vi vælger $n = 5$ og $r = 3$. Hvad skal vi så? Jo, vi skal finde et regneudtryk, som beregner sandsynligheden for at få 3 succeser i 5 udførte basisforsøg.

Vi starter med at kigge på et bestemt udfald af et binomialforsøg, hvortil vi beregner sandsynligheden for udfaldet. Vi refererer fremadrettet til et udfald som en konfiguration.

I figuren herunder ses konfigurationen, som vi indledende betragter, hvor "S" betegner en succes og "F" betegner en fiasko i et basisforsøg



Da hver af de 5 basisforsøg er uafhængige (det skal de jo være i et binomialforsøg!), beregnes sandsynligheden for netop ovenstående konfiguration ved

$$P(SFSSF) = P(S) \cdot P(F) \cdot P(S) \cdot P(S) \cdot P(F).$$

Idet p betegner sandsynligheden for succes, og $1 - p$ betegner sandsynligheden for fiasko, så er

$$\begin{aligned} P(SFSSF) = P(S) \cdot P(F) \cdot P(S) \cdot P(S) \cdot P(F) &= p \cdot (1-p) \cdot p \cdot p \cdot (1-p) \\ &= p \cdot p \cdot p \cdot (1-p) \cdot (1-p) \\ &= p^3 \cdot (1-p)^{5-3}. \end{aligned} \tag{1.1}$$

Betragter vi nu en anden konfiguration, hvor vi igen observerer 3 succeser, lad os sige

F	F	S	S	S

så kan vi (som før) beregne sandsynligheden for denne konfiguration ved

$$\begin{aligned} P(FFSSS) = P(F) \cdot P(F) \cdot P(S) \cdot P(S) \cdot P(S) &= (1-p) \cdot (1-p) \cdot p \cdot p \cdot p \\ &= p \cdot p \cdot p \cdot (1-p) \cdot (1-p) \\ &= p^3 \cdot (1-p)^{5-3}. \end{aligned} \tag{1.2}$$

Sammenligner vi sandsynligheden for de to udfald $SFSSF$ og $FFSSS$, altså (1.1) og (1.2), kan vi hurtigt overbevise os selv om, at sandsynligheden for enhver konfiguration med 3 succeser ud af 5 er givet ved

$$P(r = 3 \text{ succeser i en } \textbf{\textit{bestemt}} \text{ konfiguration}) = p^3 \cdot (1-p)^{5-3}.$$

Målet er dog at finde et regneudtryk for den samlede sandsynlighed for at observere 3 succeser ud af 5 og ikke blot sandsynligheden for bestemte konfigurationer.

Men den samlede sandsynlighed for at observere 3 succeser ud af 5 er jo lig med summen af alle sandsynlighederne for de enkelte konfigurationer, hvor der er 3 succeser.

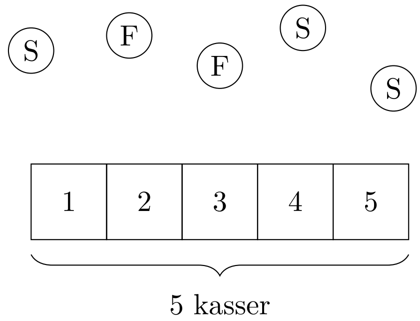
Dvs. at den samlede sandsynlighed for 3 succeser er givet ved

$$\begin{aligned}
&P(FFSSS) + P(FSFSS) + P(SFSSF) + \cdots + P(SSSFF) \\
&= \underbrace{p^3 \cdot (1-p)^{5-3} + p^3 \cdot (1-p)^{5-3} + p^3 \cdot (1-p)^{5-3} + \cdots + p^3 \cdot (1-p)^{5-3}}_{\text{Sum af et “antal led” af } \textbf{samme} \text{ størrelse}} \\
&= \{\text{antal led}\} \cdot p^3 \cdot (1-p)^{5-3} .
\end{aligned}$$

Men “antal led” må svare til antallet af mulige konfigurationer med 3 succeser ud 5.

Så spørgsmålet er nu, hvor mange mulige konfigurationer med 3 succeser ud af 5 der findes !! Jo, hertil skal vi anvende ***kombinatorik***.

Vi kan anskue dette problem som, at vi har fem bolde, tre bolde med “S” og to bolde med “F”, som skal fordeles med én bold i hver en kasse.



Da vi er interesseret i at observere en konfiguration med 3 succeser, vælger vi først de tre kasser, hvori vi kommer et “S”. Herefter kommer vi et “F” i de to sidste kasser.

Problemet er nu: På hvor mange forskellige måder kan vi vælge 3 kasser ud af 5? Dette beregner vi vha. binomialkoefficienten $K(5,3)$, da rækkefølgen på valget af de 3 kasser ikke har betydning[1].

Dvs. antallet af konfigurationer med 3 succeser ud af 5 er bestemt ved binomialkoefficienten $K(5,3)$, og derfor er den samlede sandsynlighed for $r = 3$ succeser i et binomialforsøg med $n = 5$ gentagelser bestemt ved

$$P(X = 3) = K(5,3) \cdot p^3 \cdot (1-p)^{5-3} , \tag{1.3}$$

hvor X er den stokastiske variabel, der tæller antallet af succeser i et binomialforsøg med 5 gentagelser.

Generelt: Uden at vælge konkrete tal for n og r kan en tilsvarende argumentation som ovenstående gennemføres. Vi kan derfor generalisere ovenstående ved at indsætte n i stedet for 5 og r i stedet for 3 i (1.3) for at konkludere, at den samlede sandsynlighed for r succeser i et binomialforsøg med n gentagelser er bestemt ved

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

hvor X er den stokastiske variabel, der tæller antallet af succeser i et binomialforsøg med n gentagelser. ■

[1] Husk, at $K(5,3)$ beregner, hvor mange forskellige måder der kan vælges 3 ud af 5 elementer, når rækkefølgen ikke har betydning.

1.1 Middelværdi, varians og spredning i en binomialfordeling - Forbeholdt A-niveau

Vi genopfrisker hér de generelle formler for middelværdi, varians og spredning for en **diskret** stokastisk variabel:

Generelle definitioner

Lad X være en **diskret** stokastisk variabel med udfaldene $\{x_1, x_2, x_3, \ldots, x_n\}$, så er

Middelværdi:

$$\mu = E\left[X\right] = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Varians:

$$\sigma^2 = Var\left[X\right] = E\left[(X-\mu)^2\right] = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot P(X = x_i)$$

Spredning:

$$\sigma = \sqrt{Var\left[X\right]}$$

Betragt nu en binomial stokastisk variabel X, dvs. $X \sim b(n,p)$. Husk, at en binomial stokastisk variabel er en diskret stokastisk variabel med udfaldene $\{0, 1, 2, \ldots, n\}$. Derfor er middelværdien, variansen og spredningen givet ved følgende:

Middelværdi-varians-spredning for en binomial stokastisk variabel

Lad $X \sim b(n,p)$, så er

Middelværdi:

$$\mu = E\left[X\right] = \sum_{r=0}^{n} r \cdot P(X = r)$$

Varians:

$$\sigma^2 = Var\left[X\right] = E\left[(X-\mu)^2\right] = \sum_{r=0}^{n} (r - \mu)^2 \cdot P(X = r)$$

Spredning:

$$\sigma = \sqrt{Var\left[X\right]}$$

For en binomial stokastisk variabel kan ovenstående formler for middelværdi og varians omskrives til to formler, som er nemmere at håndtere. Inden disse formler præsenteres, genopfrisker vi lige følgende for en binomial stokastisk variabel X:

$$P(X=r)=\frac{n!}{r!\cdot(n-r)!}\cdot p^r\cdot(1-p)^{n-r} \tag{1.4}$$

og summen af alle sandsynlighederne er lig med 1, dvs.

$$\sum_{r=0}^{n} P(X=r)=1. \tag{1.5}$$

Sætning 5: Hvis $X\sim b(n,p)$, så er

$$\mu=E[X]=n\cdot p \qquad \text{og} \qquad \sigma^2=Var[X]=n\cdot p\cdot(1-p)$$

Bevis (Sætning 5 middelværdi): Lad $X\sim b(n,p)$. Vi skal vise, at

$$E[X]=\sum_{r=0}^{n} r\cdot P(X=r)=n\cdot p,$$

hvor første lighed blot er definitionen på middelværdi, så det er den sidste lighed, vi skal vise. Betragt først $r\cdot P(X=r)$ for $r\geq 1$. Her er

$$\begin{aligned}
r\cdot P(X=r)&=r\cdot\frac{n!}{r!\cdot(n-r)!}\cdot p^r\cdot(1-p)^{n-r} && \text{pga. (1.4)}\\
&=\cancel{r}\cdot\frac{n\cdot(n-1)!}{\cancel{r}\cdot(r-1)!\cdot(n-r)!}\cdot p^r\cdot(1-p)^{n-r}\\
&=n\cdot\frac{(n-1)!}{(r-1)!\cdot(n-r)!}\cdot p\cdot p^{r-1}\cdot(1-p)^{n-r}\\
&=n\cdot p\cdot\frac{(n-1)!}{(r-1)!\cdot(n-r)!}\cdot p^{r-1}\cdot(1-p)^{n-r}. && (1.6)
\end{aligned}$$

Betragt nu middelværdien for X:

$$E[X]=\sum_{r=0}^{n} r\cdot P(X=r)=\sum_{r=1}^{n} r\cdot P(X=r), \tag{1.7}$$

da $0\cdot P(X=0)=0$. Vi skal nu regne videre på summen (1.7):

$$\begin{aligned}
E[X]=\sum_{r=1}^{n} r\cdot P(X=r)&=\sum_{r=1}^{n} n\cdot p\cdot\frac{(n-1)!}{(r-1)!\cdot(n-r)!}\cdot p^{r-1}\cdot(1-p)^{n-r} && \text{pga. (1.6)}\\
&=n\cdot p\cdot\sum_{r=1}^{n}\frac{(n-1)!}{(r-1)!\cdot(n-r)!}\cdot p^{r-1}\cdot(1-p)^{n-r}\\
&=n\cdot p\cdot\sum_{\tilde{r}=0}^{n-1}\frac{(n-1)!}{\tilde{r}!\cdot(n-1-\tilde{r})!}\cdot p^{\tilde{r}}\cdot(1-p)^{n-1-\tilde{r}} && \text{hvor } \tilde{r}=r-1\\
&=n\cdot p\cdot\sum_{\tilde{r}=0}^{\tilde{n}}\frac{\tilde{n}!}{\tilde{r}!\cdot(\tilde{n}-\tilde{r})!}\cdot p^{\tilde{r}}\cdot(1-p)^{\tilde{n}-\tilde{r}} && \text{hvor } \tilde{n}=n-1\\
&=n\cdot p,
\end{aligned}$$

hvor sidste lighed er opfyldt, idet $\frac{\tilde{n}!}{\tilde{r}!\cdot(\tilde{n}-\tilde{r})!} \cdot p^{\tilde{r}} \cdot (1-p)^{\tilde{n}-\tilde{r}}$ er sandsynlighedsfunktionen i binomialfordelingen $b(\tilde{n}, p)$, og vi ved fra (1.5), at summen af alle sandsynlighederne i en binomialfordeling (sandsynlighedsfordeling) er lig med 1. ■

For at bevise formlen for variansen skal vi anvende følgende sætning:

Sætning 6: Lad X være en stokastisk variabel, så gælder

$$Var\left[X\right] = E\left[X^2\right] - \left(E\left[X\right]\right)^2$$

Bemærk, at der ikke er antaget en bestemt fordeling af X. Det vil sige, at dette er et helt generelt resultat.

Bevis: Vi viser kun sætningen for en diskret stokastisk variabel. Lad X være en diskret stokastisk variabel med middelværdi $\mu = E\left[X\right]$, så er

$$\begin{aligned}
Var\left[X\right] &= E\left[(X-\mu)^2\right] \\
&= \sum_{i=1}^{n} (x_i - \mu)^2 \cdot P(X = x_i) \\
&= \sum_{i=1}^{n} (x_i^2 + \mu^2 - 2 \cdot x_i \cdot \mu) \cdot P(X = x_i) && \text{pga. 2. kvadratsætning} \\
&= \sum_{i=1}^{n} x_i^2 \cdot P(X = x_i) + \sum_{i=1}^{n} \mu^2 \cdot P(X = x_i) - \sum_{i=1}^{n} 2 \cdot x_i \cdot \mu \cdot P(X = x_i) \\
&= \sum_{i=1}^{n} x_i^2 \cdot P(X = x_i) + \mu^2 \cdot \sum_{i=1}^{n} P(X = x_i) - 2\mu \cdot \sum_{i=1}^{n} x_i \cdot P(X = x_i) \\
&= \sum_{i=1}^{n} x_i^2 \cdot P(X = x_i) + \mu^2 - 2\mu \cdot \sum_{i=1}^{n} x_i \cdot P(X = x_i) && \text{da } \sum_{i=1}^{n} P(X = x_i) = 1 \\
&= E\left[X^2\right] + \mu^2 - 2\mu \cdot E\left[X\right] && \text{definitionen på middelværdi} \\
&= E\left[X^2\right] + \mu^2 - 2\mu^2 && \text{da } \mu = E\left[X\right] \\
&= E\left[X^2\right] - \mu^2 \\
&= E\left[X^2\right] - \left(E\left[X\right]\right)^2 .
\end{aligned}$$

Hermed har vi vist, hvad vi skulle. ■

Bevis (Sætning 5 varians - Udfordring svær!): Lad $X \sim b(n,p)$, så er $E[X] = n \cdot p$ (dette har vi allerede vist). Kombineret med sætning 6 er

$$Var[X] = E\left[X^2\right] - (E[X])^2 = E\left[X^2\right] - (n \cdot p)^2 \tag{1.8}$$

Vi skal nu regne på $E[X^2]$:

$$\begin{aligned}
E\left[X^2\right] &= \sum_{r=0}^{n} r^2 \cdot P(X=r) && \text{definitionen på middelværdi} \\
&= \sum_{r=1}^{n} r^2 \cdot P(X=r) && \text{da } 0^2 \cdot P(X=0)=0 \\
&= \sum_{r=1}^{n} r \cdot n \cdot p \cdot \frac{(n-1)!}{(r-1)! \cdot (n-r)!} \cdot p^{r-1} \cdot (1-p)^{n-r} \\
&= n \cdot p \cdot \sum_{\tilde{r}=0}^{\tilde{n}} (1+\tilde{r}) \cdot \frac{\tilde{n}!}{\tilde{r}! \cdot (\tilde{n}-\tilde{r})!} \cdot p^{\tilde{r}} \cdot (1-p)^{\tilde{n}-\tilde{r}} && \text{hvor } \tilde{r} = r-1 \text{ og } \tilde{n} = n-1 \\
&= n \cdot p \cdot \left(1 + \sum_{\tilde{r}=0}^{\tilde{n}} \tilde{r} \cdot \frac{\tilde{n}!}{\tilde{r}! \cdot (\tilde{n}-\tilde{r})!} \cdot p^{\tilde{r}} \cdot (1-p)^{\tilde{n}-\tilde{r}}\right) \\
&= n \cdot p \cdot (1 + \tilde{n} \cdot p) \\
&= n \cdot p \cdot (1 + (n-1) \cdot p) && \text{da } \tilde{n} = n-1 \\
&= n \cdot p + n \cdot p^2 \cdot (n-1) \\
&= n \cdot p + n^2 \cdot p^2 - n \cdot p^2 \\
&= n^2 \cdot p^2 + n \cdot p - n \cdot p^2 \\
&= (n \cdot p)^2 + n \cdot p \cdot (1-p). && (1.9)
\end{aligned}$$

Kombinerer vi (1.8) og (1.9), er

$$Var[X] = E\left[X^2\right] - (n \cdot p)^2 = (n \cdot p)^2 + n \cdot p \cdot (1-p) - (n \cdot p)^2 = n \cdot p \cdot (1-p).$$

Hermed har vi vist, hvad vi skulle. ■

Opsummering - Formler til anvendelse

Lad $X \sim b(n,p)$, så er

Middelværdi:

$$\mu = E[X] = n \cdot p$$

Varians:

$$\sigma^2 = Var[X] = n \cdot p \cdot (1-p)$$

Spredning:

$$\sigma = \sqrt{Var[X]} = \sqrt{n \cdot p \cdot (1-p)}$$

Bemærk, at disse formler er langt nemmere at anvende, da vi kan regne middelværdi, varians og spredning direkte ud fra antalsparameteren og sandsynlighedsparameteren.