# Kapitel 2

## Estimation og testning i en binomialfordeling

Vi betragter følgende model: $X$ er binomialfordelt med (kendt) antalsparameter $n$ og ukendt sandsynlighedsparameter $p \in ]0,1[$. Vi vil diskutere hvordan man, efter observation af en værdi $x$ af $X$, estimerer $p$ og tester en hypotese af formen $p = p_0$.

Denne model er ikke så anvendelig i sig selv, hvilket tydeligt ses af det lidt kunstige eksempel vedrørende kast med en mønt, som vi bruger til illustration. Men eksemplet er velegnet til at forklare tankegang og begreber i forbindelse med estimation og testning. Desuden indgår modellen som byggesten i mere komplicerede binomialfordelingsmodeller.

Eksempel 2.1. En mønt kastes 100 gange. 61 af kastene giver krone, de resterende 39 plat. Hvad er et fornuftigt gæt på sandsynligheden $p$ for at mønten viser krone? Er det rimeligt at antage, at den er 1/2?

Møntkasteksemplet er så nemt at forstå, fordi det relaterer til noget vi alle kender. Desværre er det meget dårligt i en anden henseende. En ganske almindelig mønt, som ikke har været udsat for overlast af nogen art, vil naturligvis vise krone med en sandsynlighed der er meget tæt på 1/2. De meget små støbefejl eller skævheder, der kan være tale om, kan næppe afsløres ved 100 kast. Det vender vi tilbage til. Indtil videre anlægger vi en naivt måbende holdning, hvor vi lader som om vi ikke har den fjerneste anelse om hvad sandsynligheden for krone er, og at det kunne være vældigt interessant at få opklaret, om det kan tænkes at den simpelthen er 1/2.

Likelihoodfunktionen bliver

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x}$$

og logaritmen til den (idet vi ignorerer binomialkoefficienten, da likelihoodfunktioner jo kun har interesse på nær en multiplikativ konstant)

$$l(p) = x \log p + (n-x)\log(1-p).$$

### 2.1. Estimation.

Idet vi ser bort fra de ekstreme tilfælde $x = 0$ og $x = n$, ses at differentialkvotienten

$$l'(p) = \frac{x}{p} - \frac{n-x}{1-p}$$

er aftagende med værdier nær $+\infty$ når $p$ er i nærheden af 0 og værdier nær $-\infty$ for $p$ tæt ved 1. Heraf følger umiddelbart at likelihoodfunktionen har et entydigt maksimum, bestemt ved likelihoodligningen

$$\frac{x}{p} - \frac{n-x}{1-p} = 0$$

som har løsningen

$$\hat{p} = \frac{x}{n}.$$

Altså: ML–estimatet for parameteren i en binomialfordeling er den observerede relative hyppighed.

Hvis man vil angive en usikkerhed på dette estimat kan man benytte at binomalfordelingen, for store værdier af $n$ og værdier af $p$ som ikke er alt for tæt på 0 eller 1, er godt approksimeret ved normalfordelingen med samme middelværdi og varians. Da en normeret normalfordelt variabel med sandsynlighed 0.95 ligger mellem grænserne $\pm 1.96$, angiver man ofte estimatet $\hat{p}$ med tilføjelsen "$\pm 1.96\sigma$", hvor $\sigma$ betegner standardafvigelsen for $\hat{p}$, udregnet (i mangel af bedre) under antagelse af at $p = \hat{p}$. Altså

$$\sigma = \sqrt{\frac{1}{n^2} n\hat{p}(1-\hat{p})} = \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}.$$

Eksempel 2.1 fortsat. I møntkasteksemplet ($n = 100$, $x = 61$) bliver $\hat{p} = 0.61$. Den estimerede standardafvigelse for $\hat{p}$ er

$$\sigma = \sqrt{\frac{1}{100} 0.61 \times 0.39} = 0.04877$$

og vi vil derfor (idet $1.96 \times 0.04877 = 0.096 \approx 0.10$) angive estimatet med usikkerhed på formen "$p = 0.61 \pm 0.10$".

Bemærk, at intervallet $[0.61 - 0.10, 0.61 + 0.10] = [0.51, 0.71]$ ikke indeholder 0.5. Hermed har vi indirekte sagt, at $p$ nok ikke er 0.5. Det er netop udsagn af den slags, vi skal beskæftige os nærmere med nu.

## 2.2. Testning.

Antag, at vi for et givet tal $p_0 \in ]0,1[$ ønsker at teste hypotesen $p = p_0$. Kvotienttestørrelsen er

$$-2\log q = 2\left(l(\hat{p}) - l(p_0)\right)$$

$$= 2\left(x\log\frac{x}{n} + (n-x)\log\frac{n-x}{n} - x\log p_0 - (n-x)\log(1-p_0)\right).$$

Hypotesen forkastes for ekstremt store værdier af denne størrelse, som ifølge asymptotiske resultater er approksimativt $\chi^2$–fordelt med 1 frihedsgrad. Som tommelfingerregel plejer man at sige, at denne approksimation kan bruges når både $np_0$ og $n(1-p_0)$ er $\geq 5$.

Eksempel 2.1 fortsat. I møntkasteksemplet fås, for $p_0 = \frac{1}{2}$,

$$-2\log q = 2\left(61\log 0.61 + 39\log 0.39 - 61\log 0.5 - 39\log 0.5\right) = 4.880.$$

Opslag i tabellen (Ssr. side 151) viser at 95% fraktilen i $\chi^2$–fordelingen med 1 frihedsgrad er 3.841. Hypotesen må altså forkastes, hvis testet foretages på "standard–niveauet" $\alpha = 0.05$.

Som tidligere nævnt ville man i praksis næppe mistænke en ganske almindelig mønt for at være skæv af den grund. Det sker trods alt ca. hver 28. gang, at man får mindst 61 plat eller 61 krone. Men antag for tankeeksperimentets skyld, at vi havde fået 85 krone og 15 plat. Kvotientteststørrelsen ville så have været

$$-2\log q = 2\left(85\log 0.85 + 15\log 0.15 - 85\log 0.5 - 15\log 0.5\right) = 54.09$$

som ligger ekstremt langt ude i halen af $\chi^2$ fordelingen med 1 frihedsgrad. Faktisk gælder der (som man ikke kan se af tabellen) at sandsynligheden for at en stokastisk variabel med denne fordeling er $\geq 54.09$ er af størrelsesorden $10^{-13}$. Og det er faktisk også rigtigt, som man kan overbevise sig om ved eksakte udregninger (jvf. Ssr. opgave 4.5.3, hvor den tilsvarende sandsynlighed for $x = 80$ er vurderet), at om man så kastede med symmetriske mønter i hele sit liv ville man næsten med sikkerhed aldrig opleve at få over 85 kroner i 100 kast. Så hvis det virkelig skete ville det være en sensation, som helt naturligt ville føre til at man undersøgte den pågældende mønt (eller møntkaster) lidt nærmere.

---

Opgave 2.2.1. A (ikke slynglen fra opgave 1.7.1, men en mere tungsindig type som også hedder A) har besluttet at bruge 10 år af sit liv på at forsøge at komme i Guiness Rekordbog, ved som den første at slå mindst 80 gange krone i 100 kast med en symmetrisk mønt. Nat og dag sidder han, overvåget af notarer og retsskrivere, og kaster 100 kast i minuttet. Gør rede for, at det næppe vil lykkes. (Vink: Benyt evt. opgave 4.5.3 i Ssr.)

### 2.3. Pearson's $\chi^2$–test.

Betragt funktionen

$$f(\hat{p}) = -2\log q =$$

$$2n\left(\hat{p}\log\hat{p} + (1-\hat{p})\log(1-\hat{p}) - \hat{p}\log p_0 - (1-\hat{p})\log(1-p_0)\right)$$

der udtrykker kvotientteststørrelsen som funktion af $\hat{p} = \frac{x}{n}$. Med henblik på at opskrive Taylorudviklingen til og med anden orden af denne funktion omkring $p_0$ differentierer vi den to gange:

$$
\begin{aligned}
f'(\hat{p}) &= 2n\left(1 + \log\hat{p} - (1 + \log(1-\hat{p})) - \log p_0 + \log(1-p_0)\right) \\
&= 2n\left(\log\hat{p} - \log(1-\hat{p}) - \log p_0 + \log(1-p_0)\right).
\end{aligned}
$$

Åbenbart er $f'(p_0) = 0$. Differentiation endnu en gang giver

$$
f''(\hat{p}) = 2n\left(\frac{1}{\hat{p}} + \frac{1}{1-\hat{p}}\right).
$$

Ifølge Taylors formel har vi nu, for $\hat{p}$ i nærheden af $p_0$,

$$
\begin{aligned}
f(\hat{p}) &\approx f(p_0) + f'(p_0)(\hat{p}-p_0) + \frac{1}{2}f''(p_0)(\hat{p}-p_0)^2 \\
&= 0 + 0 + \frac{1}{2} \times 2n\left(\frac{1}{p_0} + \frac{1}{1-p_0}\right)(\hat{p}-p_0)^2 = \frac{(x-np_0)^2}{np_0(1-p_0)}.
\end{aligned}
$$

Denne størrelse, *Pearson's* $\chi^2$*–størrelse*, er således en approksimation til $-2\log q$, og benyttes ofte i stedet for denne, blandt andet fordi den har en simpel fortolkning som kvadratet på den observerede værdi af den standardiserede eller normerede binomialfordelte variabel

$$
\frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{X - \mathrm{E}X}{\sqrt{\mathrm{var}X}}.
$$

Af den centrale grænseværdisætning (se også bemærkninger side 28–29 i Ssr.) følger jo, at denne variabel for store værdier af $n$ er approksimativt normeret normalfordelt, og dermed at kvadratet på den er approksimativt $\chi^2$–fordelt med 1 frihedsgrad (jvf. definitionen af $\chi^2$–fordelingen Ssr. side 125–126). Da vi endvidere har indset, at Pearson's $\chi^2$–størrelse approksimerer $-2\log q$, har vi hermed antydet et bevis for at $-2\log Q$ er approksimativt $\chi^2$–fordelt under hypotesen $p = p_0$ — noget som vi ellers generelt må afstå fra, da det er for svært.

Da kvadratet på en normeret normalfordelt stokastisk variabel er $\chi^2$–fordelt med 1 frihedsgrad, kan man også udføre Pearson's $\chi^2$–test på følgende måde: Udregn

$$
u = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}
$$

og find ved opslag i en normalfordelingstabel (f.eks. Ssr. side 150) testets P–værdi som

$$
P = P\left(U^2 \geq u^2\right) = P\left(|U| \geq |u|\right) = 2P\left(U \geq |u|\right),
$$

hvor $U$ betegner en normeret normalfordelt stokastisk variabel.

EKSEMPEL 2.1 fortsat. For $n = 100$, $x = 61$ og $p_0 = 0.5$ fås

$$u = \frac{61 - 50}{\sqrt{100 \times 0.5 \times 0.5}} = 2.20$$

så Pearson's $\chi^2$–størrelse er $2.20^2 = 4.84$, altså ret tæt på $-2\log q = 4.880$. Ved opslag i $\chi^2$–tabellen fås da også samme konklusion som for kvotienttestet. Hvis man vil have en mere præcis P–værdi kan man gå ind i normalfordelingstabellen og finde, at $2P(U \geq 2.20) \approx 0.028$.

---

OPGAVE 2.3.1. Betragt følgende eksempel, hvor tallene fra eksempel 2.1 er "blæst op" med en faktor 10: En mønt kastes 1000 gange, 610 af kastene giver "krone". Hvad bliver konklusionen nu? Udregn både $-2\log q$ og Pearson's teststørrelse, og gør rede for hvordan disse størrelser ændrer sig når man på denne måde ganger $n$ og $x$ med den samme faktor.

## 2.4. Pearson's $\chi^2$–størrelse med kontinuitetskorrektion.

Som tegningerne side 97 og 98 i Ssr. antyder er det sådan, at halesandsynligheden $P(X \geq x)$ for $X$ binomialfordelt $(n, p_0)$, der ifølge den centrale grænseværdisætning kan approksimeres med integralet af den tilsvarende normalfordelings tæthed fra $x$ til $+\infty$, er endnu bedre approksimeret ved integralet af denne tæthed fra $x - \frac{1}{2}$ til $+\infty$. Det betyder, at størrelsen

$$\left( \frac{|X - np_0| - \frac{1}{2}}{\sqrt{np_0(1 - p_0)}} \right)^2,$$

kaldet Pearson's $\chi^2$–størrelse *med kontinuitetskorrektion,* har en fordeling, hvor halesandsynlighederne er bedre approksimeret ved $\chi^2$–fordelingens halesandsynligheder end tilfældet er for den ukorrigerede Pearson–størrelse. Derfor benytter man ofte denne størrelse, især når $\chi^2$–approksimationen er tvivlsom, fordi $np_0$ eller $n(1 - p_0)$ er lille.

EKSEMPEL 2.1 fortsat. For $X$ binomialfordelt med antalsparameter 100 og sandsynlighedsparameter 0.5 fås ved eksakt beregning

$$P(X \geq 61) = 0.0176.$$

Den sædvanlige approksimation ved halesandsynligheden i en normalfordeling (uden kontinuitetskorrektion) giver åbenbart (idet $U$ som sædvanlig betegner en normeret normalfordelt variabel)

$$P(X \geq 61) \approx P\left(\sqrt{100 \times 0.5 \times 0.5}U + 0.5 \times 100 \geq 61\right)$$

$$= P\left(U \geq \frac{61-50}{5}\right) = 0.0139$$

medens approksimationen med kontinuitetskorrektion giver

$$P(X \geq 61) \approx P\left(\sqrt{100 \times 0.5 \times 0.5}U + 0.5 \times 100 \geq 61 - \frac{1}{2}\right)$$

$$= P\left(U \geq \frac{61-50-\frac{1}{2}}{5}\right) = 0.0179.$$

Selv i dette tilfælde, hvor både $np_0$ og $n(1-p_0)$ er væsentligt større end 5 (de er nemlig begge 50), får man altså en væsentligt bedre approksimation til den rigtige sandsynlighed ved at benytte kontinuitetskorrektionen. Pearson's teststørrelse med kontinuitetskorrektion bliver i dette tilfælde

$$\frac{\left(|61-50|-\frac{1}{2}\right)^2}{100 \times 0.5 \times 0.5} = 4.41.$$

### 2.5. Det eksakte binomialtest.

Når vi her har lagt så stor vægt på approksimation af teststørrelsernes fordeling, er det fordi man ikke i almindelighed kan beregne P–værdien

$$P = P\left(-2\log Q \geq -2\log q\right)$$

eksakt. Men det kan man faktisk i den meget lille model, som dette kapitel handler om. Det er overkommeligt (f.eks. ved hjælp af et lille program i Pascal eller C, eller ved anvendelse af forfatterens lille tabelprogram WinT) at beregne P–værdien direkte som en sum af punktsandsynligheder i binomialfordelingen.

For $p_0 = 0.5$ er det eksakte test særligt nemt at udføre, fordi P–værdien kan udregnes som en dobbelt halesandsynlighed i binomialfordelingen. Dette følger af, at $-2\log q$ som funktion af $x$ er symmetrisk omkring $x = \frac{n}{2}$, aftagende for $x \leq \frac{n}{2}$ og voksende for $x \geq \frac{n}{2}$, så P–værdien

$$P = P(-2\log Q \geq -2\log q)$$

kan omskrives til

$$P = P(|X - \frac{n}{2}| \geq |x - \frac{n}{2}|) = \begin{cases} 2\sum_{k=x}^{n}\binom{n}{k}0.5^n & \text{for } x \geq \frac{n}{2}, \\ 2\sum_{k=0}^{x}\binom{n}{k}0.5^n & \text{for } x \leq \frac{n}{2}. \end{cases}$$

I ord: P–værdien er den dobbelte halesandsynlighed, regnet fra $x$ og udefter (væk fra $\frac{n}{2}$).

For $p_0 \neq 0.5$ gælder approksimativt et lignende resultat. Også i dette tilfælde er $-2\log q$ som funktion af $x$ aftagende for $x \leq np_0$ og voksende for $x \geq np_0$, men kun approksimativt symmetrisk omkring $np_0$.

Den approksimative symmetri følger af at Pearson's $\chi^2$–størrelse, som jo approksimerer $-2\log q$, er symmetrisk om $np_0$. Da binomialfordelingen med parametre $(n, p_0)$ desuden (på grund af sin lighed med en normalfordeling) er approksimativt symmetrisk omkring $np_0$, følger det, at P–værdien bliver en sum af to halesandsynligheder, der er omtrent lige store, hvoraf igen ses at

$$P \approx \begin{cases} 2\sum_{k=x}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} & \text{for } x \geq np_0, \\ 2\sum_{k=0}^{x} \binom{n}{k} p_0^k (1-p_0)^{n-k} & \text{for } x \leq np_0. \end{cases}$$

I ord: P–værdien er approksimativt lig med den dobbelte halesandsynlighed regnet fra $x$ og udefter (væk fra $np_0$).

---

OPGAVE 2.5.1. I perioden 1977 blev der i Københavns kommune født 2560 drenge og 2401 piger. Vi betragter den model, der fortolker antal drengefødsler som udfaldet af en binomialfordelt stokastisk variabel med antalsparameter 4961 og ukendt sandsynlighedsparameter $p$. Angiv maksimaliseringsestimatoren for $p$ (med usikkerhed), og foretag testet for hypotesen $p = \frac{1}{2}$.

## 2.6. Ensidet test.

Når man tester en hypotese må man i almindelighed angive både hvad hypotesen går ud på og hvad alternativet er. Indtil nu har vi underforstået, at alternativet til hypotesen $p = p_0$ er $p \in ]0,1[, \;\; p \neq p_0$. Det kan i visse tilfælde være relevant at teste $p = p_0$ mod alternativet $p > p_0$. Man taler i så fald om et *ensidet* test, i modsætning til det *tosidede* test som vi hidtil har beskæftiget os med. Det svarer til, at vi som udgangspunkt tager modellen med parameterområde indskrænket til $p \in [p_0, 1[$. Udtrykket for likelihoodfunktionen bliver naturligvis det samme som før, blot er funktionen nu kun defineret på det mindre parameterområde $[p_0, 1[$. For $x < np_0$ vil punktet $\frac{x}{n}$, hvor funktionen før antog sit maksimum, ligge uden for definitionsområdet. Men da funktionen i dette tilfælde er aftagende på hele sit definitionsområde, ser man let at maksimaliseringsestimatoren for $p$ bliver

$$\hat{p} = \begin{cases} \frac{x}{n} & \text{for } x \geq np_0, \\ p_0 & \text{for } x < np_0. \end{cases}$$

Fordelingen af kvotienttestørrelsen

$$-2\log Q = \begin{cases} 2\left(X \log \frac{X}{np_0} + (n-X)\log \frac{n-X}{n(1-p_0)}\right) & \text{for } X \geq np_0, \\ 0 \text{ for } X < np_0. \end{cases}$$

kan i dette tilfælde *ikke* approksimeres med en $\chi^2$–fordeling. Men den kan åbenbart approksimeres med en fordeling af "blandet type" (jvf.

Ssr. side 82, eksempel 5.3.4), som har halvdelen af sin masse placeret i et atom i 0 (svarende til $X \leq np_0$) og den anden halvdel (svarende til $X > np_0$) fordelt som i en $\chi^2$–fordeling med 1 frihedsgrad. Det betyder, at testets P–værdi bliver 1 for $x \leq np_0$, medens den for $x > np_0$ bliver halvt så stor som P–værdien i det sædvanlige test mod det tosidede alternativ.

Eksempel 2.1 fortsat. Som tidligere bemærket fås for $n = 100$, $x = 61$ og $p_0 = 0.5$ en eksakt P–værdi ved det tosidede test på $2 \times 0.0176 = 0.0352$. Det betyder, at en symmetrisk mønt med sandsynlighed 0.0352 ville give et resultat der er mindst lige så ekstremt. Ved "mindst lige så ekstremt" forstår vi her, at vi enten får mindst 61 krone eller mindst 61 plat, dvs. enten $X \geq 61$ eller $X \leq 39$.

Men antag nu, at vi har foretaget forsøget, fordi vi har en mistanke om at denne mønt har for stor tilbøjlighed til at give krone. Det vi ønsker at teste er i så fald hypotesen $p = 0.5$ imod alternativet $p > 0.5$. I så fald er en observation $\leq 39$ formelt ikke i modstrid med hypotesen $p = 0.5$, tværtimod. Derimod må observationen $x = 61$ regnes for "dobbelt så ekstrem" som den var under det tosidede alternativ, fordi den ikke alene ligger langt fra den forventede værdi 50, men også netop ligger i den retning, som alternativet har forudsagt. Den korrekte P–værdi i dette tilfælde er således 0.0176, ikke 0.0352.