Besvarelse af opgave 16.7

I tabellen nedenfor ses fordelingen af antal "kroner" i 400 kast med 12 mønter.

Antal krone	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
Antal kast	3	3	8	28	56	86	82	70	44	15	5	0	0	400

a) Opstil en model for materialet og estimer sandsynligheden for krone i et enkelt møntkast.

Vi betragter nu et forsøg med 400 uafhængige observationer i et endeligt udfaldsrum, der består i antallet af mulige antal "kroner" ved et enkelt kast. Dvs. udfaldsrummet består i et antal "kroner" tilhørende intervallet {0, 1, …, 12}. Der er altså tretten mulige hændelser ved hvert kast.
Vi betragter derefter en stokastisk variabel Z_x, der beskriver antal gange, hvor udfaldet "x kroner ved kast af tolv mønter" forekommer. Denne stokastiske variabel har et udfaldsrum bestående af alle vektorer $(z_0, \ldots, z_{12}\}$ bestående af hele tal, hvor summen af vektorens elementer er lig med 400. Med andre ord er Z_x multinomialfordelt med antalsparameter n = 400 og sandsynlighedsparametre $p_0, \ldots, p_{12}$.

Dvs. $Z_x \sim \text{Mult}\,(n = 400, p_0, p_1, \ldots, p_{12})$

Vi mangler imidlertid at bestemme sandsynlighedsparametrene $p_0, \ldots, p_{12}$. Der må imidlertid gælde at disse kan findes som punktsandsynligheder i en binomialfordeling, da sandsynlighederne p_x netop beskriver sandsynligheden for at få x kroner ved kast af tolv mønter. Hvis Y er en stokastisk variabel, der beskriver antal "kroner" i et kast med 12 mønter, så gælder der

$$p_i = \text{P}(Y = i \mid \theta = \hat{\theta}) \qquad Y \sim \text{Bin}\,(m = 12,\ \theta = \hat{\theta})$$

$\hat{\theta}$ er naturligvis maximum likelihood estimatet for sandynlighedsparameteren i binomialfordelingen, denne er givet ved (jf. s. 545)

$$\hat{\theta} = \frac{1}{n \cdot m} \cdot \sum_{x=0}^{m} z_x \cdot x = \frac{1}{400 \cdot 12} \cdot \left(3 \cdot 0 + 3 \cdot 1 + 8 \cdot 2 + 28 \cdot 3 + 56 \cdot 4 + \ldots + 0 \cdot 12\right) = \frac{1}{4800} \cdot 2276 = 0{,}4742$$

b) Kontroller modellen

Vi skal nu kontrollere modellen, dette gøres ved det sædvanlige q-test i multinomialfordelingen, der fx findes på s. 544:

$$Q = \sum_{x=0}^{12} \frac{\left(Z_x - n \cdot \hat{p}_x\right)}{n \cdot \hat{p}_x}$$

Bemærk, en vigtig forudsætning for anvendelse af denne teststørrelse er, at de forventede værdier $n \cdot \hat{p}_x$ ikke er for små. En tommelfingerregel er at disse størrelser ikke må være mindre end 3. Se nedenstående tabel.

Antal krone, x	*Antal kast, z_x*	*Estimeret sandsynlighed for x, $\hat{p}_x$*	*Forventet antal kast $n \cdot \hat{p}_x$*
0	3	0,00045	0,18
1	3	0,00484	1,93
2	8	0,02398	9,59
3	28	0,07209	28,83
4	56	0,14626	58,50
5	86	0,21102	84,41
6	82	0,22200	88,80
7	70	0,17159	68,63
8	44	0,09670	38,68
9	15	0,03876	15,50
10	5	0,01048	4,19
11	0	0,00172	0,69
12	0	0,00013	0,05
Total	400	1	400

Det ses af ovenstående tabel, at udfaldene $x = 0$, $x = 1$, $x = 11$ og $x = 12$ *ikke* overholder, at $n \cdot \hat{p}_x > 3$. Derfor slås udfaldene 0, 1 og 2 sammen og udfaldene 10, 11 og 12 slås ligeledes sammen. Se nedenstående tabel:

Antal krone, x	*Antal kast, z_x*	*Estimeret sandsynlighed for x, $\hat{p}_x$* [1]	*Forventet antal kast $n \cdot \hat{p}_x$*
0-2	14	0,02926	11,71
3	28	0,07209	28,83
4	56	0,14626	58,50
5	86	0,21102	84,41
6	82	0,22200	88,80
7	70	0,17159	68,63
8	44	0,09670	38,68
9	15	0,03876	15,50
10-12	5	0,01233	4,93
Total	400	1	400

[1] Bemærk at for de to grupper af udfald, der er sandsynligheden for eksempelvis 0-2 udregnet i binomialfordelingen som summen af punktsandsynlighederne for 0, 1 og 2.

Nu, hvor alle $n \cdot \hat{p}_x > 3$, kan vi udregne q-teststørrelsen:

$$q = \left(\frac{(14 - 11{,}71)^2}{11{,}71} + \frac{(28 - 28{,}83)^2}{28{,}83} + \ldots + \frac{(5 - 4{,}93)^2}{4{,}93} \right) = 1{,}907$$

Denne skal sammenholdes med 95% fraktilen i en χ^2 fordeling. Frihedsgradsantallet fås som antallet af udfaldsmuligheder minus antal estimerede parametre minus én. Man skal være opmærksom på at når man grupperer over udfald, så mister man frihedsgrader. Nu med de grupperede udfald, hvor der er 9 grupper, så gælder det, at

$$Q \sim \chi^2(9 - 1 - 1 = 7)$$

95% fraktilen i en χ^2 fordeling med 7 frihedsgrader slås op enten på lommeregner eller i Erlang S til 14,067. Dermed må hypotesen klart accepteres – men hvilken hypotese er det i virkeligheden vi tester?

Q-testet tester, hvorvidt $\hat{p}_x$'erne kan siges at antage nogle givne værdier. I dette tilfælde om, hvorvidt sandsynlighedsparametrene i den opstillede multinomialfordeling, $\hat{p}_x$, kan findes ud fra den opstillede binomialfordeling med det dertil hørende ML-estimat for binomialfordelingens parameter. Altså må binomialfordelingen som grundlag for denne multinomialfordeling klart accepteres.

Signifikanssandsynligheden ρ udregnes til:

$\rho = \mathrm{P}(Q > q) = \mathrm{P}(Q > 1{,}907) = 0{,}9648$

Igen fortæller denne os, at hypotesen om p_x som fundet ud fra en binomialfordeling må accepteres.

c) Kan det antages, at mønterne med lige stor sandsynlighed viser krone og plat?

Dette svarer til et ganske normalt test i en binomialfordeling om fordelingens sandsynlighedsparameter. Denne er vha. ML tidligere fundet til 0,4742. Kan det tænkes, at den sande værdi er 0,5?

H_0: $\theta = 0{,}5$ vs. H_1: $\theta \neq 0{,}5$

Testet er et approksimativt test i en standardiseret normalfordeling og den approksimerede signifikanssandsynlighed er givet ved[2]:

[2] idet formlerne på s. 318 for tests i diskrete fordelinger er kombineret med formlerne for middelværdi og varians i binomialfordelingen på s. 95.

$$\rho = 2 \cdot \min\left\{1 - \Phi\left(\frac{x - \frac{1}{2} - n \cdot \theta_0}{\sqrt{n \cdot \theta_0 \cdot (1 - \theta_0)}}\right) ;\ \Phi\left(\frac{x + \frac{1}{2} - n \cdot \theta_0}{\sqrt{n \cdot \theta_0 \cdot (1 - \theta_0)}}\right)\right\}$$

Denne signifikanssandsynlighed udregnes dermed vha. opslag i den standardiserede normalfordeling til:

$$\rho = 2 \cdot \min\left\{1 - \Phi\left(\frac{2276 - \frac{1}{2} - 4800 \cdot 0{,}5}{\sqrt{4800 \cdot 0{,}5 \cdot (1 - 0{,}5)}}\right) ;\ \Phi\left(\frac{2276 + \frac{1}{2} - 4800 \cdot 0{,}5}{\sqrt{4800 \cdot 0{,}5 \cdot (1 - 0{,}5)}}\right)\right\} =$$

$$2 \cdot \min\{1 - \Phi(-3{,}59) ; \Phi(-3{,}57)\} = 2 \cdot \min\{0{,}99984 ; 0{,}00018\} = 0{,}00036 < \alpha = 0{,}05$$

Derfor må hypotesen om en sandsynlighedsparameter i binomialfordelingen på 0,5 altså klart afvises. Mønten er tydeligvis ikke fair.