SIF 5060 Statistikk

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen nov. 2000

Oppgave 1

a)

Antall mulige tallkombinasjoner blir $10^4 = 10000$. Sannsynligheten for å gjette riktig blir da $p = \frac{1}{10000}$.
X er geomtrisk fordelt med parameter $p = \frac{1}{10000}$ fordi det er bare to mulige utfall, gjette riktig eller feil, med konstant sannsynlighet, hvert gjett er uavhengig av de forrige gjettene, og X er antall gjett til og med først suksess (riktig gjett).

$$P(X = x) = p(1-p)^{x-1} \qquad \text{(geomtrisk fordeling)}$$
$$\Downarrow$$
$$P(X = 300) = \frac{1}{10000}\left(1 - \frac{1}{10000}\right)^{299} = \frac{1}{10303,53} = \underline{\underline{9,7 \cdot 10^{-5}}}$$

b)

Antall måter å plassere to 7-tall blandt 4 posisjoner er

$$\binom{4}{2} = \frac{4 \cdot 3}{2 \cdot 1} = 6.$$

Antall mulige siffer på to gjenværende posisjoner er

$$9^2 = 81.$$

Dette gir

$$m = 6 \cdot 81 = \underline{\underline{486}}$$

Dermed blir X geometrisk fordelt med $p = 1/486$,

$$E[X] = \frac{1}{p} = 486.$$

Forventet tid til første vinner blir dermed

$$486 \cdot 1/2 \text{ min } = 243 \text{ min } = 4 \text{ timer } 3 \text{ minutter }.$$

c)

Utfallsrommet til $Y = \{1, 2, 3, \ldots, m\}$. Definerer hendelsen

$$F_i : \text{Innringer nummer } i \text{ gjetter riktig kode}$$

der $i = 1, 2, \ldots, m$. Setter opp sannsynligheten

$$\begin{aligned}
P(Y = y) &= P(F_1^c \cap F_2^c \cap \ldots \cap F_{y-1}^c \cap F_y) \\
&= P(F_y | F_1^c \cap F_2^c \cap \ldots \cap F_{y-1}^c) P(F_1^c \cap F_2^c \cap \ldots \cap F_{y-1}^c) \\
&= \ldots = P(F_y | F_1^c \cap F_2^c \cap \ldots \cap F_{y-1}^c) \cdot P(F_{y-1}^c | F_1^c \cap F_2^c \cap \ldots \cap F_{y-2}^c) \\
&\qquad \ldots \cdot P(F_2^c | F_1^c) \cdot P(F_1^c) \\
&= \frac{1}{m - (y-1)} \cdot \frac{m - (y-1)}{m - (y-2)} \cdot \ldots \cdot \frac{m-2}{m-1} \cdot \frac{m-1}{m} \\
&= \frac{1}{m}.
\end{aligned}$$

Da blir

$$\underline{\underline{f(y) = P(Y = y) = \frac{1}{m}}} \text{ for } y = 1, 2, \ldots, m$$

Forventet tid til noen vinner premien:

$$\begin{aligned}
E(Y) &= \sum_{y=1}^{m} y \cdot f(y) = \sum_{y=1}^{m} y \cdot \frac{1}{m} = \frac{1}{m} \sum_{y=1}^{m} y \\
&= \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2}
\end{aligned}$$

Forventet tid blir da

$$E(Y) \cdot 1/2 \text{ min } = \underline{\underline{\frac{m+1}{4} \text{ min } = 121 \text{ min } 45 \text{ sec}}}$$

Oppgave 2

a)
For at X skal være tilnærmet binomisk fordelt må en ha at $N >> n$ slik at det ikke betyr noe at samme person ikke vil bli spurt mer enn en gang.

$$P(X = 9) = \binom{n}{9} \theta^9 (1 - \theta)^{n-9} = \binom{20}{9} 0.5^9 (1 - 0.5)^{20-9} = \underline{\underline{0.1602}}$$

$$P(X > 9) = 1 - P(X \leq 9) \overset{\text{tabell}}{=} 1 - 0.412 = \underline{\underline{0.588}}$$

$$\begin{aligned} P(X>9|X\le 12) &= \frac{P(X>9\cap X\le 12)}{P(X\le 12)} = \frac{P(10\le X\le 12)}{P(X\le 12)} \\ &= \frac{P(X\le 12)-P(X\le 9)}{P(X\le 12)} \overset{\text{tabell}}{=} \frac{0.868-0.412}{0.868} = \underline{\underline{0.525}} \end{aligned}$$

b)
$X \sim b(x; n, \theta)$

$$L(\theta) = P(X=x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$l(\theta) = ln\binom{n}{x} + xln\theta + (n-x)ln(1-\theta)$$

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

$$l'(\theta) = 0$$
$$\Downarrow$$
$$\frac{x}{\theta} = \frac{n-x}{1-\theta}$$
$$x(1-\theta) = \theta(n-x)$$
$$x - x\theta = \theta n - \theta x$$
$$\theta = \frac{x}{n}$$

dvs. SME er $\underline{\underline{\hat{\theta} = \frac{X}{n}}}$

$$\mathrm{E}(\hat{\theta}) = \mathrm{E}\left(\frac{X}{n}\right) = \frac{1}{n}\mathrm{E}(X) = \frac{1}{n}n\theta = \underline{\underline{\theta}}$$

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X) = \frac{1}{n^2}n\theta(1-\theta) = \underline{\underline{\frac{\theta(1-\theta)}{n}}}$$

c)
$H_0 : \theta = \frac{1}{2}$ mot $H_1 : \theta \neq \frac{1}{2}$
Benytter testobservator

$$Z = \frac{X - n\frac{1}{2}}{\sqrt{n\frac{1}{2}(1-\frac{1}{2})}} \approx N(z; 0, 1)$$

når H_0 er riktig.

Dvs. forkaster H_0 dersom

$$Z < -z_{\frac{\alpha}{2}} \quad \text{eller} \quad Z > z_{\frac{\alpha}{2}}$$
$$\Updownarrow$$
$$\frac{X-\frac{n}{2}}{\frac{\sqrt{n}}{2}} < -z_{\frac{\alpha}{2}} \quad \text{eller} \quad \frac{X-\frac{n}{2}}{\frac{\sqrt{n}}{2}} > z_{\frac{\alpha}{2}}$$

Innsatt tall:
Observert verdi av testobservator:

$$Z = \frac{2562 - \frac{5000}{2}}{\frac{\sqrt{5000}}{2}} = 1.75$$

$\alpha = 0.10$ gir $z_{\frac{\alpha}{2}} = 1.645$
Dvs. forkaster H_0 og erklærer G som vinner av valget.

Oppgave 3

a)

$$\begin{aligned}
E(\hat{r}) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n}E(\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} r \\
&= \frac{1}{n} \cdot nr \\
&= r.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\hat{r}) &= \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \frac{r^2}{\alpha} \\
&= \frac{1}{n^2} \cdot n \cdot \frac{r^2}{\alpha} \\
&= \frac{r^2}{n\alpha}.
\end{aligned}$$

Sentralgrenseteoremet sier at dersom $X_1, X_2, \ldots, X_n$ er uavhengige, og $E(X_i) = \mu$ og $\text{Var}(X_i) = \sigma^2$, så vil fordelinga til

$$\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma}$$

konvergere mot en standard normalfordeling når n går mot uendelig.

I vår situasjon er $\bar{X} = \hat{r}$, $\mu = r$, $\sigma^2 = \frac{r^2}{\alpha}$, slik at fordelingen til

$$\begin{aligned} Z &= \sqrt{n} \cdot \frac{\hat{r} - r}{\sqrt{\frac{r^2}{\alpha}}} \\ &= \sqrt{n\alpha} \cdot \frac{\hat{r} - r}{r} \end{aligned}$$

vil konvergere mot en standard normalfordeling. Dermed blir Z tilnærmet normalfordelt når n er stor.

b)

$$\begin{aligned} P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) &= 1 - \alpha \\ P(-z_{\frac{\alpha}{2}} \leq \sqrt{n\alpha} \cdot \frac{\hat{r} - r}{r} \leq z_{\frac{\alpha}{2}}) &= 1 - \alpha \\ P(\frac{\hat{r}}{1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}} \leq r \leq \frac{\hat{r}}{1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}}) &= 1 - \alpha \end{aligned}$$

fordi

$$\begin{aligned} -z_{\frac{\alpha}{2}} &\leq \sqrt{n\alpha} \cdot \frac{\hat{r} - r}{r} \\ \frac{-r z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}} &\leq \hat{r} - r \\ r(1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}) &\leq \hat{r} \\ r &\leq \frac{\hat{r}}{1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}} \end{aligned}$$

og

$$\begin{aligned} \sqrt{n\alpha} \cdot \frac{\hat{r} - r}{r} &\leq z_{\frac{\alpha}{2}} \\ \hat{r} - r &\leq \frac{r z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}} \\ \hat{r} &\leq r(1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}) \\ \frac{\hat{r}}{1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}} &\leq r \end{aligned}$$

.

Et $(1-\alpha)\cdot 100\%$ konfidensitervall blir dermed

$$[\frac{\hat{r}}{1+\frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}},\frac{\hat{r}}{1-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n\alpha}}}]$$

.

Alternativt må man også godta at en utleder et symetrisk intervall basert på at $\sqrt{n\alpha}\cdot(\hat{r}-r)/\hat{r}\approx n(\cdot;0,1)$ dersom dette begrunnes skikkelig.
Innsatt tallene $\hat{r}=16.99$, $\alpha=5$, $n=20$, $a=0.05 \Rightarrow z_{\frac{\alpha}{2}}=1.960$, får vi konfidensitervallet $[14.21, 21.13]$.

c) Vi vet at

$$\begin{aligned} z_A &= \sqrt{n\alpha}\cdot\frac{\hat{r}_A-r_A}{r_A} \\ &\approx n(z_A;0,1) \\ z_B &= \sqrt{n\alpha}\cdot\frac{\hat{r}_B-r_B}{r_B} \\ &\approx n(z_B;0,1). \end{aligned}$$

Under H_0 vil man ha at

$$\begin{aligned} z_A-z_B &= \sqrt{n\alpha}\cdot\frac{\hat{r}_A-\hat{r}_B}{r} \\ &\approx n(\cdot;0,\sqrt{1+1}), \end{aligned}$$

der r er felles (ukjent) verdi for r_A og r_B. Siden n er stor erstatter vi r med dens estimater; $\hat{r}=1/2\cdot(\hat{r}_A+\hat{r}_B)$. Testobservatoren blir dermed

$$\begin{aligned} U &= \frac{1}{\sqrt{2}}\sqrt{n\alpha}\cdot\frac{\hat{r}_A-\hat{r}_B}{\frac{1}{2}(\hat{r}_A+\hat{r}_B)} \\ &= \sqrt{2n\alpha}\cdot\frac{\hat{r}_A-\hat{r}_B}{\hat{r}_A+\hat{r}_B} \end{aligned}$$

som er tilnærmet standard-normalfordelt når n er stor. Forkaster dermed H_0 dersom $U>z_{\frac{\alpha}{2}}$ eller $U<-z_{\frac{\alpha}{2}}$.

Oppgave 4

$$y=x^2$$

$$\Downarrow$$

$$x=\sqrt{y} \qquad \text{eller} \qquad x=-\sqrt{y}$$

$$\frac{dx}{dy}=\frac{1}{2\sqrt{y}} \qquad\qquad \frac{dx}{dy}=-\frac{1}{2\sqrt{y}}$$

X er standard normalfordelt,

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$$

slik at

$$\begin{aligned}
f_Y(y) &= f_x(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_x(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\
&= 2 \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(\sqrt{y})^2} \cdot \frac{1}{2\sqrt{y}} \\
&= \underline{\underline{\frac{1}{2^{1/2}\Gamma(1/2)}\, y^{1/2-1}\, e^{-\frac{y}{2}}}}
\end{aligned}$$

Dette er fordelingsfunksjonen til kji-kvadrat fordelingen med en frihetsgrad, og Y er dermed χ^2_1-fordelt.