Homework 1

Michael Li

1. In lecture we found out that for the multinomial distribution the MLE for $\pi_i = \frac{n_i}{n}$

so $\pi_1 = \frac{54}{100} = \frac{27}{50}$, $\pi_2 = \frac{25}{100} = \frac{1}{4}$, $\pi_3 = \frac{9}{100}$

$\pi_4 = \frac{5}{100} = \frac{1}{20}$, $\pi_5 = \frac{2}{100} = \frac{1}{50}$, $\pi_6 = \frac{5}{100} = \frac{1}{20}$

2. heads $\Rightarrow$ 1, 2 w/ equal prob

tails $\Rightarrow$ 3, 4, 5, 6 w/ equal prob

heads prob θ, tails prob $(1-\theta)$

$$f_X(x) = \begin{cases} \frac{\theta}{2} & x = 1,2 \\ \frac{(1-\theta)}{4} & x = 3,4,5,6 \\ 0 & x \notin \{1,2,\dots,6\} \end{cases}$$

3. $L(\theta|x) = p(x|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$

$$= \prod_{i=1}^{n} \left(\frac{\theta}{2}\right)^{I(x_i=1,2)} \left(\frac{1-\theta}{4}\right)^{I(x_i=3,4,5,6)}$$

$$= \left(\frac{\theta}{2}\right)^{\sum_{i=1}^{n} I(x_i=1,2)} \left(\frac{1-\theta}{4}\right)^{\sum_{i=1}^{n} I(x_i=3,4,5,6)}$$

$$= \left(\frac{\theta}{2}\right)^{(n_1+n_2)} \left(\frac{1-\theta}{4}\right)^{(n_3+n_4+n_5+n_6)}$$

$$\ln L(\pi|x) = (n_1+n_2)\ln\left(\frac{\theta}{2}\right) + (n_3+n_4+n_5+n_6)\ln\left(\frac{1-\theta}{4}\right)$$

$$\frac{d}{d\theta}\ln L(\theta|x) = \frac{1}{2}\left(\frac{n_1+n_2}{(\theta/2)}\right) - \frac{1}{4}\left(\frac{n_3+n_4+n_5+n_6}{\left(\frac{1-\theta}{4}\right)}\right)$$

$$= \frac{n_1+n_2}{\theta} - \frac{n_3+n_4+n_5+n_6}{1-\theta} = 0$$

$$\frac{n_1+n_2}{\theta} = \frac{n_3+n_4+n_5+n_6}{1-\theta}$$

$$(1-\theta)(n_1+n_2) = \theta(n_3+n_4+n_5+n_6)$$

$$n_1+n_2 - \theta(n_1+n_2) = \theta(n_3+n_4+n_5+n_6)$$

$$n_1+n_2 = \theta(n_3+n_4+n_5+n_6+n_1+n_2)$$

$$\hat{\theta} = \frac{n_1+n_2}{n}$$ where $n = n_1+n_2+n_3+n_4+n_5+n_6$

4. $g_X(x) = \begin{cases} (1-\phi)^{x-1}\phi & x \in \mathbb{N} \\ 0 & x \notin \mathbb{N} \end{cases}$

$p(x|\phi) = (1-\phi)^{x-1}\phi$

$L(\phi|x) = p(X|\phi) = \prod_{i=1}^{n} p(x_i|\pi)$

$= \prod_{i=1}^{n} (1-\phi)^{x_i-1}\phi$

$= (1-\phi)^{\sum_{i=1}^{n} x_i - 1} \cdot \phi^n$

$= (1-\phi)^{S-n} \cdot \phi^n$ where $S = \sum_{i=1}^{n} x_i$

$\ln L(\phi|x) = (S-n)\ln(1-\phi) + n\ln(\phi)$

$\frac{d}{d\phi} \ln L(\phi|x) = -\frac{S-n}{1-\phi} + \frac{n}{\phi} = 0$

$\frac{n}{\phi} = \frac{S-n}{1-\phi}$

$n - n\phi = \phi(S-n)$

$n = \phi(S-n+n)$

$\hat{\phi} = \frac{n}{S}$

since $\hat{\phi} = \frac{n}{S}$, $\frac{1}{\hat{\phi}} = \frac{S}{n}$, which is the sample mean

5. (a) For $x \in \{1, \dots, 5\}$ geometric distribution

For $x = 6$ whatever is left

$$h_X(x|x{=}6) = 1 - \sum_{i=1}^{5} (1-\psi)^{i-1}\psi = 1 - \psi - \psi \sum_{i=1}^{4} (1-\psi)^i$$

case $x = 1$ is different

$$= 1 - \psi - \psi\left(\frac{(1-\psi) - (1-\psi)^5}{1-(1-\psi)}\right)$$

$$= 1 - \psi - \psi\left(\frac{(1-\psi) - (1-\psi)^5}{\psi}\right)$$

$$= 1 - \psi - (1-\psi) + (1-\psi)^5 = (1-\psi)^5$$

so $h_X(x) = \begin{cases} (1-\psi)^{x-1}\psi & x \in A \\ (1-\psi)^5 & x = 6 \\ 0 & \end{cases}$ where $A = \{1,2,3,4,5\}$

(b) $p(x_i|\psi) = [(1-\psi)^{x_i-1}\psi]^{I(x_i \in A)} [(1-\psi)^5]^{I(x_i=6)}$

$$L(\psi|x) = p(x|\psi) = \prod_{i=1}^{n} p(x_i|\psi)$$

$$= \prod_{i=1}^{n} (1-\psi)^{(x_i-1)(I(x_i \in A))} \psi^{I(x_i \in A)} (1-\psi)^{5 \cdot I(x_i=6)}$$

$$= (1-\psi)^{\sum_{i=1}^{n}(x_i-1)(I(x_i \in A))} \psi^{\sum_{i=1}^{n} I(x_i \in A)} (1-\psi)^{\sum_{i=1}^{n} 5 \cdot I(x_i=6)}$$

$$= (1-\psi)^{\sum_{i=1}^{n} x_i I(x_i \in A) - I(x_i \in A) + 5 \cdot I(x_i=6)} \psi^{(n-n_6)}$$

$$= (1-\psi)^{\left[\sum_{i=1}^{n} x_i I(x_i \in A)\right] - (n-n_6) + 5n_6} \psi^{(n-n_6)}$$

$$= (1-\psi)^{\left[\sum_{i=1}^{n} x_i - \underbrace{\sum_{i=1}^{n} x_i I(x_i=6)}_{\text{constant}}\right] + 6n_6 - n} \psi^{(n-n_6)}$$

$$= (1-\psi)^{\sum_{i=1}^{n} x_i - 6n_6 + 6n_6 - n} \psi^{(n-n_6)}$$

$$= (1-\psi)^{s-n} \psi^{(n-n_6)}$$ where $s = \sum_{i=1}^{n} x_i$

(c) $\ln L(\psi|x) = (s-n)\ln(1-\psi) + (n-n_6)\ln(\psi)$

$$\frac{d}{d\psi}\ln L(\psi|x) = -\left(\frac{s-n}{1-\psi}\right) + \frac{n-n_6}{\psi} = 0$$

$$\frac{n-n_6}{\psi} = \frac{s-n}{1-\psi}$$

$$(1-\psi)(n-n_6) = \psi(s-n)$$

$$n-n_6-\psi(n-n_6) = \psi(s-n)$$

$$n-n_6 = \psi(s-n)+\psi(n-n_6)$$

$$n-n_6 = \psi(s-n+n-n_6)$$

$$\frac{n-n_6}{s-n_6} = \hat{\psi} \quad \text{where } s = \sum_{i=1}^{n} x_i$$

6. $M_{twopart}$: $\hat{\theta} = \frac{n_1+n_2}{n} = \frac{54+25}{100} = \frac{79}{100}$

$$L(\hat{\theta}|x) = \left(\frac{\frac{79}{100}}{2}\right)^{79}\left(\frac{\frac{21}{100}}{4}\right)^{21} \quad \left(\frac{\theta}{2}\right)^{(n_1+n_2)}\left(\frac{1-\theta}{4}\right)^{(n_3+n_4+n_5+n_6)}$$

$$= 1.79687\times10^{-59} \qquad \ln(1.79687\times10^{-59}) \approx -135.266$$

M_{geom}: $\hat{\phi} = \frac{n}{s} = \frac{100}{191}$

$$L(\hat{\phi}|x) = (1-\phi)^{s-n}\cdot\phi^{n}$$

$$= \left(\frac{91}{191}\right)^{91}\cdot\left(\frac{100}{191}\right)^{100}$$

$$= 3.93907\times10^{-58} \qquad \ln(3.93907\times10^{-58}) = -132.179$$

M_{trunc}: $\hat{\psi} = \frac{n-n_6}{s-n_6} = \frac{100-5}{191-5} = \frac{95}{186}$

$$L(\hat{\psi}|x) = (1-\psi)^{s-n}\,\psi^{(n-n_6)}$$

$$= \left(\frac{91}{186}\right)^{91}\cdot\left(\frac{95}{186}\right)^{95}$$

$$= 1.06439\times10^{-56} \qquad \ln(1.06439\times10^{-56}) = -128.882$$

M_{five}: $L(\hat{\pi}|x) = \prod_{i=1}^{6}\pi_i^{n_i} = \left(\frac{27}{50}\right)^{54}\left(\frac{1}{4}\right)^{25}\left(\frac{9}{100}\right)^{9}\left(\frac{1}{20}\right)^{5}\left(\frac{1}{50}\right)^{2}\left(\frac{1}{20}\right)^{5}$

$$= 4.76108\times10^{-56} \qquad \ln(4.76108\times10^{-56}) = -127.384$$

By comparing the likelihoods, we can see the truncated model was the best out of all, the one params, but M_{five} was best overall.

7. $AIC_M = 2k_M - 2\ln L_M$

$\underset{\text{params}}{k_M}$ $\quad \underset{\text{likelihood}}{L_M}$

$M_{two-part}$: params $=1$, $L_M = 1.79687 \times 10^{-59}$

$AIC_{two-part} = 2 \cdot 1 - 2\ln(1.79687 \times 10^{-59})$

$= 272.533$

M_{geom}: params $=1$, $L_M = 3.93907 \times 10^{-58}$

$AIC_{geom} = 2 \cdot 1 - 2\ln(3.93907 \times 10^{-58})$

$= 266.358$

M_{trunc}: params $=1$, $L_M = 1.06439 \times 10^{-56}$

$AIC_{trunc} = 2 \cdot 1 - 2\ln(1.06439 \times 10^{-56})$

$= 259.765$

M_{five}: params $=5$, $L_M = 4.76108 \times 10^{-56}$

$AIC_{trunc} = 2 \cdot 5 - 2\ln(4.76108 \times 10^{-56})$

$= 264.769$

- The truncated geometric model has the best AIC score

8. The two models that are nested in M_{five} are $M_{two-part}$ and M_{trunc}. Since in M_{five} you basically assign the probabilities for each x_i, M_{five} can become $M_{two-part}$ by assigning $\pi_1, \pi_2 = \frac{\theta}{2}$ and $\pi_3, \pi_4, \pi_5, \pi_6 = \frac{1-\theta}{4}$ and M_{five} can become M_{trunc} by assigning $\begin{cases} \pi_i = (1-\psi)^{x-1}\psi & x \in \{1, \ldots, 5\} \\ \pi_i = (1-\psi)^5 & x = 6 \end{cases}$

LRTs

$M_{two-part}$: $L(\hat{\theta}|x) = 1.79687 \times 10^{-59}$ — from question 6

M_{five}: $L(\hat{\pi}|x) = 4.76108 \times 10^{-56}$

$$D = -2\ln\left(\frac{L(\hat{\theta}|x)}{L(\hat{\pi}|x)}\right) = -2\ln\left(\frac{1.79687 \times 10^{-59}}{4.76108 \times 10^{-56}}\right) = 15.7644$$

$P(\underbrace{\chi^2_4}_{5\text{ params} - 1\text{ param} = 4} > 15.7644) = 0.003 < 0.05$

so reject null hypothesis, so M_{five} is the better model

M_{trunc}: $L(\hat{\psi}|x) = 1.06434 \times 10^{-56}$

$D = -2\ln(1.06439 \times 10^{-56}) + 2\ln(4.76108 \times 10^{-56}) = 2.996$

$P(\chi^2_4 > 2.996) = 0.56 > 0.05$

so fail to reject null, M_{trunc} is the better model

9. M_{TA} has $p(x_i|\alpha) = \left(\frac{1}{2}\right)^{I(x_i=1)}\left(\frac{1}{4}\right)^{I(x_i=2)}\left(\frac{1}{8}\right)^{I(x_i=3)}\left(\frac{1}{16}\right)^{I(x_i=4)}\left(\frac{1}{32}\right)^{I(x_i=5,6)}$

$$L(\alpha|x) = p(x|\alpha) = \prod_{i=1}^{n} p(x_i|\alpha)$$

$$= \prod_{i=1}^{n}\left(\frac{1}{2}\right)^{I(x_i=1)}\left(\frac{1}{4}\right)^{I(x_i=2)}\left(\frac{1}{8}\right)^{I(x_i=3)}\left(\frac{1}{16}\right)^{I(x_i=4)}\left(\frac{1}{32}\right)^{I(x_i=5,6)}$$

$$= \left(\frac{1}{2}\right)^{\sum_{i=1}^{n} I(x_i=1)}\left(\frac{1}{4}\right)^{\sum_{i=1}^{n} I(x_i=2)}\left(\frac{1}{8}\right)^{\sum_{i=1}^{n} I(x_i=3)}\left(\frac{1}{16}\right)^{\sum_{i=1}^{n} I(x_i=4)}\left(\frac{1}{32}\right)^{\sum_{i=1}^{n} I(x_i=5,6)}$$

$$= \left(\frac{1}{2}\right)^{n_1}\left(\frac{1}{4}\right)^{n_2}\left(\frac{1}{8}\right)^{n_3}\left(\frac{1}{16}\right)^{n_4}\left(\frac{1}{32}\right)^{(n_5+n_6)}$$

$$= \left(\frac{1}{2}\right)^{54}\left(\frac{1}{4}\right)^{25}\left(\frac{1}{8}\right)^{9}\left(\frac{1}{16}\right)^{5}\left(\frac{1}{32}\right)^{7}$$

$$= 1.01958 \times 10^{-56}$$

$$\ln L(\alpha|x) = \ln(1.01958 \times 10^{-56}) = -128.925$$

Since all the probabilities are set, there are 0 parameters in M_{TA}

M_{TA} is nested in M_{five} and M_{trunc}. For M_{five} you can make $\pi_1 = \frac{1}{2}, \pi_2 = \frac{1}{4}, \pi_3 = \frac{1}{8}, \pi_4 = \frac{1}{16}, \pi_5 = \frac{1}{32}, \pi_6 = \frac{1}{32}$ and for M_{trunc} you can make $\hat{\varphi} = \frac{1}{2}$.

LRTs

M_{five}: $L(\hat{\pi}|x) = 4.76108 \times 10^{-56}$

$$D = -2\ln(1.01958 \times 10^{-56}) + 2\ln(4.76108 \times 10^{-56}) = 3.08217$$

$$P(\chi^2_5 > 3.08217) = 0.69 > 0.05$$

five params − 0 params

Fail to reject null

M_{TA} is the better model

M_{trunc}: $L(\hat{\varphi}|x) = 1.06439 \times 10^{-56}$

$$D = -2\ln(1.01958 \times 10^{-56}) + 2\ln(1.06439 \times 10^{-56}) = 0.086$$

$$P(\chi^2_1 > 0.086) = 0.77 > 0.05$$

1 param − 0 params

Fail to reject null

M_{TA} is the better model

100 samples

10. $M_{two\text{-}part}$: $\hat{\theta} = \frac{n_1+n_2}{n} = \frac{30+25}{100} = \frac{55}{100} = \frac{11}{20}$

$$L(\hat{\theta}|x) = \left(\frac{\frac{11}{20}}{2}\right)^{55}\left(\frac{\frac{9}{20}}{4}\right)^{45}$$

$$= 2.9185 \times 10^{-74}$$

$$AIC = 2\cdot 1 - 2\ln(2.9185 \times 10^{-74})$$

$$= 340.64$$

M_{five}: $L(\hat{\pi}|x) = \prod_{i=1}^{6} \pi_i^{n_i} = \left(\frac{3}{10}\right)^{30}\left(\frac{1}{4}\right)^{25}\left(\frac{3}{25}\right)^{12}\left(\frac{13}{100}\right)^{13}\left(\frac{1}{10}\right)^{10}\left(\frac{1}{10}\right)^{10}$

$$= 4.93829 \times 10^{-74}$$

$$AIC = 2\cdot 5 - 2\ln(4.93829 \times 10^{-74})$$

$$= 347.589$$

The AIC shows that $M_{two\text{-}part}$ is the beter model

1000 samples

$M_{two\text{-}part}$: $\hat{\theta} = \frac{n_1+n_2}{n} = \frac{300+250}{1000} = \frac{11}{20}$

$$L(\hat{\theta}|x) = \left(\frac{\frac{11}{20}}{2}\right)^{550}\left(\frac{\frac{9}{20}}{4}\right)^{450}$$

$$= 4.4835 \times 10^{-736}$$

$$AIC = 2\cdot 1 - 2\ln(4.4835 \times 10^{-736})$$

$$= 3388.4$$

M_{five}: $L(\hat{\pi}|x) = \prod_{i=1}^{6} \pi_i^{n_i} = \left(\frac{3}{10}\right)^{300}\left(\frac{1}{4}\right)^{250}\left(\frac{3}{25}\right)^{120}\left(\frac{13}{100}\right)^{130}\left(\frac{1}{10}\right)^{100}\left(\frac{1}{10}\right)^{100}$

$$= 8.6251 \times 10^{-734}$$

$$AIC = 2\cdot 5 - 2\ln(8.6251 \times 10^{-734})$$

$$= 3385.89$$

This time the AIC shows that M_{five} is a better model.

The AIC favors different models at different sample sizes because in the case of a small sample size, although M_{five} is a better fit and has a better log likelihood, the AIC penalizes the extra parameters M_{five} has and so $M_{two\text{-}part}$ has a better AIC. In a large sample size, the log likelihood matters more and overpowers the parameter penalty, so M_{five} has the beter AIC.