CAS CS 365
Lab 5

1. Multinormal Distribution (n, π)
    a. Example: Rolling a fair dice n = 60 times
        i. Probabilities, π = (⅙, …, ⅙)
        ii. Observed data, X = (10, 9, 11, 10, 8, 12)
    b. Observed data, x = (x_1,..., x_m)
    c. Constraint: x_1 + … + x_m = n
    d. Probabilities, π = (theta$_1$, …, theta$_m$)
    e. Constraint: theta$_1$ + … + theta$_m$ = 1
    f. Multinomial PMF:

$$PMF = \frac{n!}{\Pi_{i=1}^{n} x_i!} \Pi_{i=1}^{n} (p(x_i))^{x_i}$$

2. Consider following problem
    a. Suppose X = (200, 34, 38, 98) is a sample from Mult(n=370, π)

$$\pi_\theta \;=\; \left(\frac{1}{2} + \frac{1}{4}\theta,\; \frac{1}{4}(1-\theta),\; \frac{1}{4}(1-\theta),\; \frac{1}{4}\theta\right).$$

    We want to solve for theta
        i. Find the MLE without using EM
            1. Likelihood, L(theta;x) is then given by

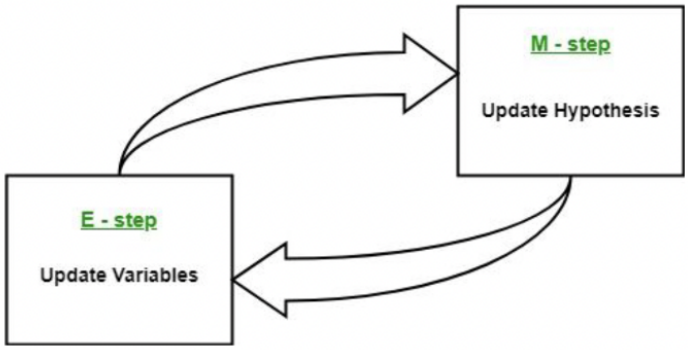
$$L(\theta; \mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!}\left(\frac{1}{2}+\frac{1}{4}\theta\right)^{x_1}\left(\frac{1}{4}(1-\theta)\right)^{x_2}\left(\frac{1}{4}(1-\theta)\right)^{x_3}\left(\frac{1}{4}\theta\right)^{x_4}$$

            2. So that the log-likelihood l(theta;x) is

$$l(\theta; \mathbf{x}) \;=\; C \;+\; x_1 \ln\left(\frac{1}{2}+\frac{1}{4}\theta\right) \;+\; (x_2+x_3)\ln(1-\theta) \;+\; x_4 \ln(\theta)$$

            3. Final step is to solve $\frac{dl}{d\theta} = 0$
3. Slight change to the problem
    a. Suppose Y = (y1, y2, y3 = 34, y4 = 38, y5 = 98) is a sample from Mult(n=370, π) but we only observe X (original problem)
    b. (y1 + y2 = 200)

c. $$\pi_\theta^* = \left(\frac{1}{2},\ \frac{1}{4}\theta,\ \frac{1}{4}(1-\theta),\ \frac{1}{4}(1-\theta),\ \frac{1}{4}\theta\right).$$

4. Classical EM
    a. EM algorithm is used for obtaining MLEs of parameters when some of the data is missing or unobserved
    b. 

5. E Step
    a. Given the statistical model which generates a set X of observed data, a set of unobserved latent data Y, and a vector of unknown parameters theta, along with the log likelihood function l(theta; X,Y)
    b. E-step of the EM algorithm computes expected value of l(theta; X,Y) given the observed data, X, and the current parameter estimate, $\text{theta}_{\text{old}}$, say. In particular, we define

$$\begin{aligned} Q(\theta;\theta_{old}) &:= \mathsf{E}\left[l(\theta;\mathcal{X},\mathcal{Y}) \mid \mathcal{X},\theta_{old}\right] \\ &= \int l(\theta;\mathcal{X},y)\, p(y \mid \mathcal{X},\theta_{old})\, dy \end{aligned}$$

where p(y| X, $\text{theta}_{\text{old}}$) is the conditional density of Y given the observed data, X, and assuming theta = $\text{theta}_{\text{old}}$.
6. Y-Problem Likelihood
    a. $$\mathcal{L}(\theta;\mathcal{X},\mathcal{Y}) = \frac{n!}{y_1!y_2!y_3!y_4!y_5!}\left(\frac{1}{2}\right)^{y_1}\left(\frac{1}{4}\theta\right)^{y_2}\left(\frac{1}{4}(1-\theta)\right)^{y_3}\left(\frac{1}{4}(1-\theta)\right)^{y_4}\left(\frac{1}{4}\theta\right)^{y_5}$$
    b. $$l(\theta;\mathcal{X},\mathcal{Y}) = C + y_2\ln(\theta) + (y_3+y_4)\ln(1-\theta) + y_5\ln(\theta)$$

7. E Step continued
    a. $$l(\theta;\mathcal{X},\mathcal{Y}) \;=\; C \;+\; y_2\ln(\theta) \;+\; (y_3+y_4)\ln(1-\theta) \;+\; y_5\ln(\theta)$$
    b. Recalling that $Q(\theta;\theta_{old}) := \mathsf{E}\left[l(\theta;\mathcal{X},\mathcal{Y}) \mid \mathcal{X},\theta_{old}\right]$, we have
    $$Q(\theta;\theta_{old}) \;:=\; C \;+\; \mathsf{E}\left[y_2\ln(\theta) \mid \mathcal{X},\theta_{old}\right] \;+\; (y_3+y_4)\ln(1-\theta) \;+\; y_5\ln(\theta)$$
8. Expected value of y2 term
    a. Given n=370 and Y = (y1, y2, y3 = 34, y4 = 38, y5 = 98), y1 + y2 = 200
    b. We can think of choosing between y2 and y1 as a binomial distribution. (Y = y2 in the formula)
        i. $$f(\mathcal{Y} \mid \mathcal{X},\theta) \;=\; \mathsf{Bin}\left(y_1+y_2,\; \frac{\theta/4}{1/2+\theta/4}\right).$$
9. E Step Complete
    a. Recalling that $Q(\theta;\theta_{old}) := \mathsf{E}\left[l(\theta;\mathcal{X},\mathcal{Y}) \mid \mathcal{X},\theta_{old}\right]$, we have
    $$\begin{aligned} Q(\theta;\theta_{old}) \;&:=\; C \;+\; \mathsf{E}\left[y_2\ln(\theta) \mid \mathcal{X},\theta_{old}\right] \;+\; (y_3+y_4)\ln(1-\theta) \;+\; y_5\ln(\theta) \\ &=\; C \;+\; (y_1+y_2)p_{old}\ln(\theta) \;+\; (y_3+y_4)\ln(1-\theta) \;+\; y_5\ln(\theta) \end{aligned}$$
    where $p_{old} := \dfrac{\theta_{old}/4}{1/2+\theta_{old}/4}.$
10. M Step
    a. M Step consists of maximizing over theta the expectation computed. That is, we set $\theta_{new} \;:=\; \max_{\theta}\; Q(\theta;\theta_{old}).$
    b. We then set $\theta_{old} = \theta_{new}$
    c. $$\begin{aligned} Q(\theta;\theta_{old}) \;&:=\; C \;+\; \mathsf{E}\left[y_2\ln(\theta) \mid \mathcal{X},\theta_{old}\right] \;+\; (y_3+y_4)\ln(1-\theta) \;+\; y_5\ln(\theta) \\ &=\; C \;+\; (y_1+y_2)p_{old}\ln(\theta) \;+\; (y_3+y_4)\ln(1-\theta) \;+\; y_5\ln(\theta) \end{aligned}$$
    d. We now maximize Q(theta; θ_{old}) to find θ_{new}. Taking the derivative we obtain
    $p_{old} := \dfrac{\theta_{old}/4}{1/2+\theta_{old}/4}.$
    $$\frac{dQ}{d\theta} \;=\; \frac{(y_1+y_2)}{\theta}p_{old} \;-\; \frac{(y_3+y_4)}{1-\theta} \;+\; \frac{y_5}{\theta}$$
    which is zero when we take theta = θ_{new} where
    $$\theta_{new} \;:=\; \frac{y_5+p_{old}(y_1+y_2)}{y_3+y_4+y_5+p_{old}(y_1+y_2)}$$

11. Writing Code to find Theta
    a. Start with initial guess of theta_old
    b. From E Step: $$p_{old} := \frac{\theta_{old}/4}{1/2 + \theta_{old}/4}.$$
    c. From M step: solve for theta_new

$$\theta_{new} \;\; := \;\; \frac{y_5 + p_{old}(y_1 + y_2)}{y_3 + y_4 + y_5 + p_{old}(y_1 + y_2)}$$

    d. Repeat b-c for N iterations
12. Notice the following
    a. From M-Step (update rule)

$$\theta_{new} \;\; := \;\; \frac{y_5 + p_{old}(y_1 + y_2)}{y_3 + y_4 + y_5 + p_{old}(y_1 + y_2)}$$

    b. Our update rule is based on our data and p_old from the E step

$$p_{old} := \frac{\theta_{old}/4}{1/2 + \theta_{old}/4}.$$