STAT 32950

Robert Winter

HW #7

5. Differential entropy for cts RV X w/ PDF $f(x)$ is
$$H(X) = -\int_{\mathbb{R}} f(x)\log f(x)\,dx,$$
where $0\log 0 := 0$.

5a. Show a univariate normal RV w/ mean μ, variance σ^2 has entropy $\log(\sigma\sqrt{2\pi e})$.

- Let $X \sim N(\mu, \sigma^2)$. So,
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$
- Then
$$H(X) = -\int_{-\infty}^{\infty} f_X(x)\log f_X(x)\,dx$$
$$= -\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\right]dx$$
$$= -\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\left[\log\left(\left(\sqrt{2\pi\sigma^2}\right)^{-1}\right)-\frac{1}{2\sigma^2}(x-\mu)^2\right]dx$$
$$= -\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\left(-\log\left(\sqrt{2\pi\sigma^2}\right)\right)dx$$
$$-\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx$$
$$= \log(\sigma\sqrt{2\pi})\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx$$
$$+\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\frac{1}{2\sigma^2}(x-\mu)^2\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx$$
$$= \log(\sigma\sqrt{2\pi})\cdot 1 + \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\frac{1}{2\sigma^2}(x-\mu)^2\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx$$

$u = (x-\mu) \Rightarrow du = dx$

$dv = \frac{1}{2\sigma^2}(x-\mu)\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx$

$w = -\frac{1}{2\sigma^2}(x-\mu)^2$

$\Rightarrow dw = -\frac{1}{2\sigma^2}\cdot 2(x-\mu)dx \Rightarrow -\frac{1}{2}dw = \frac{1}{2\sigma^2}(x-\mu)dx$

$\Rightarrow dv = -\frac{1}{2}e^{w}\,dw$

$\rightarrow$

$$\left(\begin{aligned} \Rightarrow v &= -\tfrac{1}{2} e^{w} \\ &= -\tfrac{1}{2} \exp\{-\tfrac{1}{2\sigma^2}(x-\mu)^2\} \end{aligned} \right.$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{\sqrt{2\pi\sigma^2}}\left[(x-\mu)\left(-\frac{1}{2}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\right)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\left(-\frac{1}{2}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\right)dx \right]$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{\sqrt{2\pi\sigma^2}}\left[\lim_{x\to\infty}(x-\mu)\left(-\frac{1}{2}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\right) - \lim_{x\to-\infty}(x-\mu)\left(-\frac{1}{2}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\right)\right]$$

$$+ \frac{1}{2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{\sqrt{2\pi\sigma^2}}\left[\lim_{x\to\infty}\frac{-(x-\mu)}{2\exp\{\frac{1}{2\sigma^2}(x-\mu)^2\}} - \lim_{x\to-\infty}\frac{-(x-\mu)}{2\exp\{\frac{1}{2\sigma^2}(x-\mu)^2\}}\right] + \frac{1}{2}(1)$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{\sqrt{2\pi\sigma^2}}\left[0-0\right] + \frac{1}{2}$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{2}$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{2}\log(e)$$

$$= \log(\sigma\sqrt{2\pi}) + \log(e^{1/2})$$

$$= \log(\sigma\sqrt{2\pi}) + \log(\sqrt{e})$$

$$= \log(\sigma\sqrt{2\pi e}).$$

- That is,

$$H(X) = \log(\sigma\sqrt{2\pi e}),$$

as desired. $\square$

5b. Let X be a cts RV w/ mean 0, variance σ^2, PDF $f(x)$. Let $\varphi(x)$ be the PDF of a normal RV w/ mean 0, variance σ^2. Show

$$-\int_{\mathbb{R}} f(x) \log \varphi(x)\,dx = \log(\sigma\sqrt{2\pi e}).$$

- Note that $\mathbb{E}[X]=0$ & $Var(X)=\sigma^2$.
- Also,

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-0)^2\right\}$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

- So,

$$-\int_{\mathbb{R}} f(x) \log \varphi(x)\,dx = -\int_{-\infty}^{\infty} f(x) \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}\right]dx$$
$$= -\int_{-\infty}^{\infty} f(x)\left[-\log(\sqrt{2\pi\sigma^2}) - \frac{x^2}{2\sigma^2}\right]dx$$
$$= \int_{-\infty}^{\infty} f(x)\left[\log(\sqrt{2\pi\sigma^2}) + \frac{x^2}{2\sigma^2}\right]dx$$
$$= \int_{-\infty}^{\infty} f(x)\log(\sqrt{2\pi\sigma^2})\,dx + \int_{-\infty}^{\infty} f(x)\cdot\frac{1}{2\sigma^2}x^2\,dx$$
$$= \log(\sigma\sqrt{2\pi})\int_{-\infty}^{\infty} f(x)\,dx + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} x^2 f(x)\,dx$$
$$= \log(\sigma\sqrt{2\pi})\cdot 1 + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} x^2 f(x)\,dx$$

- Note that

$$\sigma^2 = Var(X)$$
$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
$$= \mathbb{E}[X^2] - 0^2$$
$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - 0$$
$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

- So, continuing,

$$-\int_{\mathbb{R}} f(x)\log\varphi(x)\,dx = \log(\sigma\sqrt{2\pi}) + \frac{1}{2\sigma^2}\cdot\sigma^2$$
$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{2}$$
$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{2}\log(e)$$
$$= \log(\sigma\sqrt{2\pi}) + \log(e^{1/2})$$
$$= \log(\sigma\sqrt{2\pi}) + \log(\sqrt{e})$$

$\longrightarrow$

$$\Rightarrow -\int_{\mathbb{R}} f(x) \log \varphi(x)\, dx = \log(\sigma\sqrt{2\pi e}),$$

as desired. $\square$

5c. Let X be any cts RV on $\mathbb{R}$ w/ mean 0, variance σ^2. Show $H(X)$ is smaller than the differential entropy of a normal RV w/ mean 0, variance σ^2; equality holds iff X is of normal dist. Hint: Jensen's Inequality.

- <u>Case I</u>: Suppose X is <u>not</u> of normal distribution.
  - Let $h(y) = \log(y)$. Then $h'(y) = \frac{1}{y}$, and $h''(y) = \frac{-1}{y^2} < 0$ $\forall y \in (0, \infty)$. Thus, h is strictly concave on its domain, $(0, \infty)$.
  - So, by Jensen's Inequality,
    $$h(E[Y]) > E[h(Y)]$$
    for some random variable Y.
  - Denote the PDF of X by $f(x)$. We also have $E[X] = 0$, $Var(X) = \sigma^2$.
  - Denote by $\varphi(x)$ the PDF of a random variable of distribution $N(0, \sigma^2)$.
  - We consider the random variable
    $$Y = \frac{\varphi(X)}{f(X)}.$$
  - By Jensen,
    $$h(E[Y]) > E[h(Y)]$$
    $$\Rightarrow \log\left(E\left[\frac{\varphi(X)}{f(X)}\right]\right) > E\left[\log\left(\frac{\varphi(X)}{f(X)}\right)\right]$$
    $$\Rightarrow \log\left(\int_{-\infty}^{\infty} \frac{\varphi(x)}{f(x)} f(x)\,dx\right) > \int_{-\infty}^{\infty} \log\left(\frac{\varphi(x)}{f(x)}\right) f(x)\,dx$$
    $$\Rightarrow \log\left(\int_{-\infty}^{\infty} \varphi(x)\,dx\right) > \int_{-\infty}^{\infty} [\log \varphi(x) - \log f(x)] f(x)\,dx$$
    $$\Rightarrow \log(1) > \int_{-\infty}^{\infty} [f(x) \log \varphi(x) - f(x) \log f(x)]\,dx$$
    $$\Rightarrow 0 > \int_{-\infty}^{\infty} f(x) \log \varphi(x)\,dx - \int_{-\infty}^{\infty} f(x) \log f(x)\,dx$$
    $$\Rightarrow -\int_{-\infty}^{\infty} f(x) \log \varphi(x)\,dx > -\int_{-\infty}^{\infty} f(x) \log f(x)\,dx$$
    $$\Rightarrow \log(\sigma\sqrt{2\pi e}) > -\int_{-\infty}^{\infty} f(x) \log f(x)\,dx \quad \text{by Part (b)}$$
    $$\Rightarrow -\int_{-\infty}^{\infty} \varphi(x) \log \varphi(x)\,dx > -\int_{-\infty}^{\infty} f(x) \log f(x)\,dx \quad \text{by Part (a)}$$

$\rightarrow$

$$\Rightarrow \quad H(N(0,\sigma^2)) > H(X)$$

by the definition of differential entropy, where $H(N(0,\sigma^2))$ denotes the differential entropy of a normal RV w/ mean 0 + variance σ^2.

- Thus, when X is <u>not</u> normally distributed, we have shown that $H(X)$ is strictly smaller than $H(N(0,\sigma^2))$.

- <u>Case II</u>: Suppose $X \sim N(0,\sigma^2)$.
  - Let $Z \sim N(0,\sigma^2)$.
  - Then, by Part (a),
    - $H(X) = \log(\sigma\sqrt{2\pi e})$, and
    - $H(Z) = \log(\sigma\sqrt{2\pi e})$.
  - So, $H(X) = H(Z)$.
  - Thus, when $X \sim N(0,\sigma^2)$, $H(X)$ equals $H(N(0,\sigma^2))$.

- Thus, we have proven that for any continuous RV on $\mathbb{R}$ w/ $E[X]=\mu$, $Var(X)=\sigma^2$,

$$H(X) \leq H(N(0,\sigma^2)),$$

with equality precisely when $X \sim N(0,\sigma^2)$. $\square$

5d. For $i=1,2$, let X_i be any cts RV on $\mathbb{R}$ w/ $E[X_i]=0$, $Var(X_i)=\sigma_i^2$. Let $Y=X_1+X_2$. Find $\max_{X_1,X_2} H(Y)$. Describe your choice of X_1, X_2 which yield maximum entropy of Y.

- $Var(Y) = Var(X_1+X_2)$

$$= Var(X_1)+Var(X_2)+2\,Cov(X_1,X_2)$$
$$= \sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2, \quad \text{where } \rho = Corr(X_1,X_2).$$

- $E[Y] = E[X_1+X_2]$

$$= E[X_1]+E[X_2]$$
$$= 0+0$$
$$= 0$$

- We assume $Y=X_1+X_2$ is continuous (which is not the case for certain choices of X_1 & X_2, such as $X_2=-X_1$), since otherwise, $H(Y)$ is not defined.
- We showed in Part (c) that, given mean 0 & a specific variance, differential entropy is maximized by a normal RV.
- So, since $E[Y]=0$ and $Var(Y)=\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2$, $H(Y)$ is maximized if $Y\sim N(0,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$.
- Let

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right),$$

So the marginal densities are $X_1\sim N(0,\sigma_1^2)$ & $X_2\sim N(0,\sigma_2^2)$.
- Since (X_1,X_2) is of bivariate normal distribution,

$Y=X_1+X_2\sim N(0+0,\ \sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)=N(0,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$.

We've found X_1 & X_2 such that the maximal $H(Y)$ will be attained!
- By Part (a), this maximal entropy is

$$H(Y)=\log\left(\sqrt{\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2}\,\sqrt{2\pi e}\right).$$

$\longrightarrow$

- Thus,
$$\max_{X_1, X_2} H(Y) = \log\left(\sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}\,\sqrt{2\pi e}\right)$$
where $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$

- If $Cov(X_1, X_2) = 0$, our answer simplifies:
$$\max_{X_1, X_2} H(Y) = \log\left(\sqrt{\sigma_1^2 + \sigma_2^2}\,\sqrt{2\pi e}\right)$$
where $X_1 \sim N(0, \sigma_1^2)$,
$X_2 \sim N(0, \sigma_2^2)$

6. Suppose $X = (X_1, \dots, X_p)$, $X_i \sim Ber(\nu_i)$, so

$$X_i = \begin{cases} 1 & w/p \ \nu_i \\ 0 & w/p \ 1-\nu_i \end{cases} \quad \text{for } i = 1, \dots, p; \ X_i \perp\!\!\!\perp X_j \ \forall i \neq j.$$

Then X is a p-variate Bernoulli random vector.

6a. Write out $\mathbb{E}[X]$ & indicate dimensions.

- For each i,
  - $\mathbb{E}[X_i] = \nu_i(1) + (1-\nu_i)(0) = \nu_i$.
- So,

$$\mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_p] \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_p \end{bmatrix} = \mathbb{E}[X]$$

- $\mathbb{E}[X]$ is a $(p \times 1)$ vector.

6b. Derive $Cov(X)$ & indicate dimensions.

$$\cdot\ Cov(X) = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_p) \\ Cov(X_1, X_2) & Var(X_2) & \cdots & Cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_1, X_p) & Cov(X_2, X_p) & \cdots & Var(X_p) \end{bmatrix}$$

$$\begin{aligned} \cdot\ Var(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \\ &= \nu_i(1)^2 + (1-\nu_i)(0)^2 - (\nu_i)^2 \\ &= \nu_i - \nu_i^2 \\ &= \nu_i(1-\nu_i) \end{aligned}$$

for each i.

- Since $X_i \perp\!\!\!\perp X_j \ \forall i \neq j$, we have $Cov(X_i, X_j) = 0 \ \forall i \neq j$.

$\rightarrow$

- Thus,

$$\mathrm{Cov}(X) = \begin{bmatrix} \nu_1(1-\nu_1) & 0 & \cdots & 0 \\ 0 & \nu_2(1-\nu_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \nu_p(1-\nu_p) \end{bmatrix}$$

- $\mathrm{Cov}(X)$ is a $(p \times p)$ matrix.

6c. Consider Y to be of mixture distribution: w/p π_c, Y is from a p-variate Bernoulli dist. w/ mean $\mu_c = (\mu_{c1}, \dots, \mu_{cp})$ + covariance Σ_c, with $\sum_{c=1}^{K} \pi_c = 1$. Denote $\pi = (\pi_1, \dots, \pi_K)$.

6c(i) Write expression for $\mu = \mathbb{E}[Y]$.

- We denote by Y_c the p-variate Bernoulli distribution with $\mathbb{E}[Y_c] = \mu_c$, $\mathrm{Cov}(Y_c) = \Sigma_c$.
- Then

$$\begin{aligned} \mu = \mathbb{E}[Y] &= \mathbb{E}\left[\textstyle\sum_{c=1}^{K} \pi_c Y_c\right] \\ &= \textstyle\sum_{c=1}^{K} \mathbb{E}[\pi_c Y_c] \\ &= \textstyle\sum_{c=1}^{K} \pi_c \mathbb{E}[Y_c] \end{aligned}$$

$$\Rightarrow \mu = \mathbb{E}[Y] = \sum_{c=1}^{K} \pi_c \mu_c = \pi_1 \mu_1 + \cdots + \pi_K \mu_K$$

- Element-wise, we may write

$$\mu = \mathbb{E}[Y] = \left(\sum_{c=1}^{K} \pi_c \mu_{c1}, \dots, \sum_{c=1}^{K} \pi_c \mu_{cp}\right)$$

6c(ii) Write an expression for the conditional probability $P(y|\mu_c)=\mathbb{P}(Y=y \mid y \text{ is from cluster } c \text{ w/ mean } \mu_c)$, where $y=(y_1,\ldots,y_p)$ is a realization of Y.

$$\cdot P(y|\mu_c)=\mathbb{P}(Y=y \mid y \text{ from cluster } c)$$
$$=\mathbb{P}\left(\begin{bmatrix}Y_1\\ \vdots \\ Y_p\end{bmatrix}=\begin{bmatrix}y_1\\ \vdots \\ y_p\end{bmatrix} \;\middle|\; y \text{ from cluster } c\right)$$
$$=\mathbb{P}(Y_1=y_1,\ldots,Y_p=y_p \mid y \text{ from cluster } c)$$
$$=\mathbb{P}(Y_1=y_1 \mid y \text{ from cluster } c)\cdots\mathbb{P}(Y_p=y_p \mid y \text{ from cluster } c)$$
since $Y_i \perp\!\!\!\perp Y_j \; \forall i\neq j$ for components of a p-variate Bernoulli distribution, which Y is, given that y is from cluster c

$$\Rightarrow P(y|\mu_c)=\mu_{c1}^{y_1}(1-\mu_{c1})^{1-y_1}\cdots\mu_{cp}^{y_p}(1-\mu_{cp})^{1-y_p}$$
$$=\prod_{i=1}^{p}\mu_{ci}^{y_i}(1-\mu_{ci})^{1-y_i}$$

6c(iii) Write an expression for $P(y|\pi,\mu_1,\ldots,\mu_K)=\mathbb{P}(Y=y \mid y \text{ is from a mixture model w/ parameters } \pi,\mu_1,\ldots,\mu_K)$

$$\cdot P(y|\pi,\mu_1,\ldots,\mu_K)=P(y|\mu_1)\mathbb{P}(y \text{ from cluster } 1)+\cdots+P(y|\mu_K)\mathbb{P}(y \text{ from cluster } K)$$
by the Law of Total Probability
$$=\pi_1\prod_{i=1}^{p}\mu_{1i}^{y_i}(1-\mu_{1i})^{1-y_i}+\cdots+\pi_K\prod_{i=1}^{p}\mu_{Ki}^{y_i}(1-\mu_{Ki})^{1-y_i}$$

$$\Rightarrow P(y|\pi,\mu_1,\ldots,\mu_K)=\sum_{c=1}^{K}\pi_c\prod_{i=1}^{p}\mu_{ci}^{y_i}(1-\mu_{ci})^{1-y_i}$$

6c(iv) Let C be class membership variable, $\mathbb{P}(C=c)=\pi_c$, $c=1,\dots,K$. Find $Cov(Y)$ using vector version of Law of Total Variance, $Var(X)=\mathbb{E}[Var(X|C)]+Var(\mathbb{E}[X|C])$.

- By the Law of Total Variance,
$$Cov(Y)=\mathbb{E}[Cov(Y|C)]+Cov(\mathbb{E}[Y|C]).$$
- By definition in the problem, $Cov(Y|C)=\Sigma_c$.
- So, $\mathbb{E}[Cov(Y|C)]=\mathbb{E}[\Sigma_c]$
$$=\sum_{c=1}^{K}(\pi_c\Sigma_c).$$
- Also by definition in the problem, $\mathbb{E}[Y|C]=\mu_c$.
- So, $Cov(\mathbb{E}[Y|C])=Cov(\mu_c)$
$$=\mathbb{E}[\mu_c\mu_c^T]-\mathbb{E}[\mu_c]\mathbb{E}[\mu_c^T]$$
$$=\sum_{c=1}^{K}\pi_c\mu_c\mu_c^T-\left(\sum_{c=1}^{K}\pi_c\mu_c\right)\left(\sum_{c=1}^{K}\pi_c\mu_c^T\right)$$
- Thus,
$$Cov(Y)=\sum_{c=1}^{K}\pi_c\Sigma_c+\sum_{c=1}^{K}\pi_c\mu_c\mu_c^T-\left(\sum_{c=1}^{K}\pi_c\mu_c\right)\left(\sum_{c=1}^{K}\pi_c\mu_c^T\right)$$
$$=\sum_{c=1}^{K}\pi_c(\Sigma_c+\mu_c\mu_c^T)-\left(\sum_{c=1}^{K}\pi_c\mu_c\right)\left(\sum_{c=1}^{K}\pi_c\mu_c^T\right)$$

6d. Sps $y^{(i)}$, $i=1,\dots,n$ are n independent observations of Y in Part (c) (which is of mixture dist). Each $y^{(i)}=(y_{i1},\dots,y_{ip})$ is from a p-variate Bernoulli dist. w/ mean μ_c & covariance Σ_c w/p π_c, w/ $\sum_{c=1}^{K}\pi_c=1$. Denote data as $Y=\{y^{(1)},\dots,y^{(n)}\}$.

6d(i) Write an expression (in terms of y_{ij}, π_c, μ_{ci}) for the likelihood fcn $L(\mu_1,\dots,\mu_K,\pi|Y)=P(Y|\mu_1,\dots,\mu_K,\pi)$.

- $L(\mu_1,\dots,\mu_K,\pi|Y)=P(Y|\mu_1,\dots,\mu_K,\pi)$
$$=\prod_{i=1}^{n}P(y^{(i)}|\mu_1,\dots,\mu_K,\pi)$$
since $y^{(1)},\dots,y^{(n)}$ are independent

$$\Rightarrow L(\mu_1,\dots,\mu_K,\pi|Y)=\prod_{i=1}^{n}\sum_{c=1}^{K}\pi_c\prod_{j=1}^{p}\mu_{cj}^{y_{ij}}(1-\mu_{cj})^{1-y_{ij}}$$
by Part (c)(iii)

6d(ii) Create a (non-degenerate) numerical dataset $[y_{ij}]_{n \times p}$ w/ $p=3$, $n=5$.

- We consider the arbitrary dataset

$$[y_{ij}]_{5\times3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

where each element is either 0 or 1 since each of the 5 observations are drawn from a trivariate Bernoulli distribution.

6d(iii) Write out the likelihood function in Part (i) in terms of π_c, μ_{ci} & numerical values of y_{ij}.

$$\begin{aligned}
\cdot\ L(\mu_1, \dots, \mu_K, \pi \mid Y) &= \prod_{i=1}^{5} \sum_{c=1}^{K} \pi_c \prod_{j=1}^{3} \mu_{cj}^{y_{ij}} (1-\mu_{cj})^{1-y_{ij}} \\
&= \Big[\sum_{c=1}^{K} \pi_c \prod_{j=1}^{3} \mu_{cj}^{y_{1j}} (1-\mu_{cj})^{1-y_{1j}}\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \prod_{j=1}^{3} \mu_{cj}^{y_{2j}} (1-\mu_{cj})^{1-y_{2j}}\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \prod_{j=1}^{3} \mu_{cj}^{y_{3j}} (1-\mu_{cj})^{1-y_{3j}}\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \prod_{j=1}^{3} \mu_{cj}^{y_{4j}} (1-\mu_{cj})^{1-y_{4j}}\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \prod_{j=1}^{3} \mu_{cj}^{y_{5j}} (1-\mu_{cj})^{1-y_{5j}}\Big]
\end{aligned}$$

$$\begin{aligned}
&= \Big[\sum_{c=1}^{K} \pi_c \big[\mu_{c1}^{y_{11}}(1-\mu_{c1})^{1-y_{11}} \cdot \mu_{c2}^{y_{12}}(1-\mu_{c2})^{1-y_{12}} \cdot \mu_{c3}^{y_{13}}(1-\mu_{c3})^{1-y_{13}}\big]\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \big[\mu_{c1}^{y_{21}}(1-\mu_{c1})^{1-y_{21}} \cdot \mu_{c2}^{y_{22}}(1-\mu_{c2})^{1-y_{22}} \cdot \mu_{c3}^{y_{23}}(1-\mu_{c3})^{1-y_{23}}\big]\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \big[\mu_{c1}^{y_{31}}(1-\mu_{c1})^{1-y_{31}} \cdot \mu_{c2}^{y_{32}}(1-\mu_{c2})^{1-y_{32}} \cdot \mu_{c3}^{y_{33}}(1-\mu_{c3})^{1-y_{33}}\big]\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \big[\mu_{c1}^{y_{41}}(1-\mu_{c1})^{1-y_{41}} \cdot \mu_{c2}^{y_{42}}(1-\mu_{c2})^{1-y_{42}} \cdot \mu_{c3}^{y_{43}}(1-\mu_{c3})^{1-y_{43}}\big]\Big] \\
&\cdot \Big[\sum_{c=1}^{K} \pi_c \big[\mu_{c1}^{y_{51}}(1-\mu_{c1})^{1-y_{51}} \cdot \mu_{c2}^{y_{52}}(1-\mu_{c2})^{1-y_{52}} \cdot \mu_{c3}^{y_{53}}(1-\mu_{c3})^{1-y_{53}}\big]\Big]
\end{aligned}$$

$\rightarrow$

$$= \left[\sum_{c=1}^{K} \pi_c \left[\mu_{c1}^{1}(1-\mu_{c1})^{0} \cdot \mu_{c2}^{1}(1-\mu_{c2})^{0} \cdot \mu_{c3}^{1}(1-\mu_{c3})^{0}\right]\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c \left[\mu_{c1}^{1}(1-\mu_{c1})^{0} \cdot \mu_{c2}^{1}(1-\mu_{c2})^{0} \cdot \mu_{c3}^{0}(1-\mu_{c3})^{1}\right]\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c \left[\mu_{c1}^{1}(1-\mu_{c1})^{0} \cdot \mu_{c2}^{0}(1-\mu_{c2})^{1} \cdot \mu_{c3}^{0}(1-\mu_{c3})^{1}\right]\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c \left[\mu_{c1}^{0}(1-\mu_{c1})^{1} \cdot \mu_{c2}^{1}(1-\mu_{c2})^{0} \cdot \mu_{c3}^{0}(1-\mu_{c3})^{1}\right]\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c \left[\mu_{c1}^{0}(1-\mu_{c1})^{1} \cdot \mu_{c2}^{1}(1-\mu_{c2})^{0} \cdot \mu_{c3}^{1}(1-\mu_{c3})^{0}\right]\right]$$

$$\Rightarrow L(\mu_1, \ldots, \mu_K, \pi \mid Y) = \left[\sum_{c=1}^{K} \pi_c \mu_{c1}\mu_{c2}\mu_{c3}\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c \mu_{c1}\mu_{c2}(1-\mu_{c3})\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c \mu_{c1}(1-\mu_{c2})(1-\mu_{c3})\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c (1-\mu_{c1})\mu_{c2}(1-\mu_{c3})\right]$$
$$\cdot \left[\sum_{c=1}^{K} \pi_c (1-\mu_{c1})\mu_{c2}\mu_{c3}\right]$$