200B HW#1 solution

3.7 Multivariate Distributions

3. (a) We have

$$\int_0^{+\infty}\int_0^{+\infty}\int_0^{+\infty} f(x_1,x_2,x_3)dx_1dx_2dx_3 = \frac{c}{6}$$

Since the value of this integral must be equal to 1, it follows that $c = 6$.

(b) For $x_1 > 0$ and $x_3 > 0$,

$$f_{13}(x_1,x_3) = \int_0^{+\infty} f(x_1,x_2,x_3)dx_2 = 3e^{-(x_1+3x_3)}.$$

(c) The marginal joint p.d.f of X_2 and X_3 is

$$f_{23}(x_2,x_3) = \int_0^{+\infty} f(x_1,x_2,x_3)dx_1 = 6e^{-(2x_2+3x_3)}.$$

The conditional p.d.f. of X_1 given that $X_2 = x_1$ and $X_3 = x_3$ is

$$g_1(x_1|x_2,x_3) = \frac{f(x_1,x_2,x_3)}{f_{23}(x_2,x_3)} = e^{-x_1}$$

Therefore,

$$\begin{aligned} P(X_1 < 1|X_2 = 2, X_3 = 1) &= \int_0^1 g_3(x_3|x_2 = 2, x_2 = 1)\,dx_1 \\ &= \int_0^1 e^{-x_1}dx_1 = 1 - e^{-1}. \end{aligned}$$

8. For any given value x of X, the random variables $Y_1,\ldots,Y_n$ are i.i.d., each with the p.d.f. $g(y|x)$. Therefore, the conditional joint p.d.f. of $Y_1,\ldots,Y_n$ given that $X = x$ is

$$h(y_1,\ldots,y_n|x) = g(y_1|x)\cdots g(y_n|x) = \begin{cases} \frac{1}{x^n} & \text{for } 0 < y_i < x (i = 1,\ldots,n), \\ 0 & \text{otherwise.} \end{cases}$$

This joint p.d.f. is positive if and only if each $y_i > 0$ and x is greater than every y_i. In other words, x must be greater than $m = \max\{y_1, \ldots, y_n\}$.

(a) For $y_i > 0 (i = 1, \ldots, n)$, the marginal joint p.d.f. of $Y_1, \ldots, Y_n$ is

$$g_0(y_1, \ldots, y_n) = \int_{-\infty}^{\infty} f(x)h(y_1, \ldots, y_n|x)dx = \int_{m}^{\infty} \frac{1}{n!}\exp(-x)dx = \frac{1}{n!}\exp(-m).$$

(b) For $y_i > 0 (i = 1, \ldots, n)$, the conditional joint p.d.f. of X given that $Y_i = y_i (i = 1, \ldots, n)$ is

$$g_1(x|y_1, \ldots, y_n) = \frac{f(x)h(y_1, \ldots, y_n|x)}{g_0(y_1, \ldots, y_n)} = \begin{cases} \exp(-(x-m)) & \text{for } x > m, \\ 0 & \text{otherwise.} \end{cases}$$

3.8 Functions of Random Variable

7.(a) $g_1(y_1) = f(\sqrt{y_1})|\frac{dx}{dy_1}| = \frac{1}{2\sqrt{y_1}}$, $0 < y_1 < 1$.

(b) $g_2(y_2) = f(-y_2^{-\frac{1}{3}})|\frac{dx}{dy_2}| = \frac{1}{3y_2^{\frac{2}{3}}}$, $-1 < y_2 < 0$.

(c) $g_3(y_3) = f(y_3^2)|\frac{dx}{dy_3}| = 2y_3$, $0 < y_3 < 1$.

3.9 Functions of Two or More Random Variables

6. By Eq. (3.9.2) (with a change in notation),

$$g(z) = \int_{-\infty}^{\infty} f(z-t, t)dt \quad \text{for} -\infty < z < \infty$$

However, the integrand is positive only for $0 \le z - t \le t \le 1$. Therefore, for $0 \le z \le 1$, it is positive only for $z/2 \le t \le z$ and we have

$$g(z) = \int_{z/2}^{z} 2z dt = z^2.$$

For $1 < z < 2$, the integrand is positive only for $z/2 \le t \le 1$ and we have

$$g(z) = \int_{z/2}^{1} 2z dt = z(2-z).$$

14.

$$\begin{aligned} G(Y \le y) &= \text{Pr(at least n-1 less equal to y)} \\ &= \Pr((\cup_{i=1}^{n}\{x_j \le y, , j \ne i, , x_i > y\}) \cup \{x_j \le y\}) \\ &= \sum_{i=1}^{n} \Pr(\{x_j \le y, , j \ne i, , x_i > y\}) + \Pr(\{x_j \le y\})) \\ &= nF(y)^{n-1}(1 - F(y)) + F(y)^n \\ &= nF(y)^{n-1} - (n-1)F(y)^n \end{aligned}$$

$$g(y) = \frac{dG(y)}{dy} = n(n-1)(F(y)^{n-2}f(y) - F(y)^{n-1}f(y))$$

3.11 Supplementary Exercises

16. For $0 < x < 1$, the marginal p.d.f of X is

$$f_1(x) = \int_x^1 2(x+y)dy = 1 + 2x - 3x^2.$$

Therefore, $P(X < 1/2) = \int_0^{1/2} f_1(x)dx = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} = \frac{5}{8}$.

For $0 < x < y < 1$, the conditional p.d.f of Y given $X = x$ is

$$g_2(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{2(x+y)}{1+2x-3x^2}.$$

4.4 Moments

10. The m.g.f. of Z is

$$\begin{aligned}
\psi_1(t) &= E(\exp(tZ)) = E[\exp(t(2X - 3Y + 4))] \\
&= \exp(4t)E(\exp(2tX)\exp(-3ty)) \\
&= \exp(4t)E(\exp(2tX))E(\exp(-3tY)) \quad \text{Since } X \text{ and } Y \text{ are independent} \\
&= \exp(4t)\psi(2t)\psi(-3t) \\
&= \exp(4t)\exp(4t^2 + 6t)\exp(9t^2 - 9t) \\
&= \exp(13t^2 + t)
\end{aligned}$$

4.6 Covariance and Correlation

8.

$$\begin{aligned}
\mathrm{Cov}(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) &= E[(\sum_{i=1}^{n} a_i X_i - E(\sum_{i=1}^{n} a_i X_i))(\sum_{j=1}^{m} b_j Y_j - E(\sum_{j=1}^{m} b_j Y_j))] \\
&= E[(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i E(X_i))(\sum_{j=1}^{m} b_j Y_j - \sum_{j=1}^{m} b_j E(Y_j))] \\
&= E[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j (X_i - E(X_i)(Y_j - E(Y_j)))] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j E[(X_i - E(X_i)(Y_j - E(Y_j)))] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \mathrm{Cov}(X_i, Y_j)
\end{aligned}$$

7.1 Statistical Inference

3. The random variables of interest are the observable $Z_1, Z_2, \ldots$, the times at which successive particles hit the target, and β, the hypothetically observable (parameter) rate of the Poisson process. The hit times occur according to a Poisson process with rate β conditional on β. Other random variables of interest are the observable inter-arrival times $Y_1 = Z_1$ and $Y_k = Z_k - Z_{k-1}$ for $k \geq 2$.

6. The random variables of interest are the observable number X of Mexican-American grand jurors and the hypothetically observable (parameter) P. The conditional distribution of X given $P = p$ is the binomial distribution with parameters 220 and p. Also, P has the beta distribution with parameters α and β, which have not yet been specified.

7.5 Maximum Likelihood Estimators

2. For $1 \leq i \leq n$, let the random variable $X_i = 1$ if the purchases of a certain brand of breakfast cereal are made by women and $X_i = 0$ if they are made by men. Then $X_1, \ldots, X_n$ form a random sample from the Bernoulli distribution with parameter p. Base on the observed values $x_1, \ldots, x_n$, the likelihood function is

$$f_n(x|p) = \prod_{i=1}^{n} p_i^x (1-p)^{(1-x_i)} = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$$

The log-likelihood function is

$$L(p) = \log(f_n(x|p)) = \left(\sum_{i=1}^{n} x_i\right)\log(p) + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-p)$$

Let

$$\frac{dL(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = \frac{\sum_{i=1}^{n} x_i - np}{p(1-p)} = 0$$

we have $\hat{p} = \sum_{i=1}^{n} x_i/n = \bar{x}_n$. And it can be verified that the second derivative of $L(p)$ at $\hat{p}$ is negative, so the M.L.E is $\bar{x}_n = 58/70 = 29/35$.

3. It can be seen that $\frac{dL(p)}{dp} > 0$ for $p < \bar{x}_n = 58/70$, which implies $L(p)$ is increasing for $1/2 \le p \le 2/3$. The log-likelihood and hence the likelihood function achieves the maximum at $p = 2/3$. Namely, the M.L.E $\hat{p} = 2/3$.

9.

$$L(\theta) = \log(f_n(x|\theta)) = \log(\theta^n (X_1X_2 \dots X_n)^{\theta-1}) = n\log(\theta) + (\theta-1)\log(X_1X_2\dots X_n)$$

Take the derivative we have

$$\frac{dL(\theta)}{d\theta} = \frac{n}{\theta} + \log(X_1X_2\dots X_n).$$

Then to check concaveness we take the second derivative

$$\frac{dL(\theta)^2}{d^2\theta} = -\frac{n}{\theta^2} < 0.$$

So the M.L.E can be solved by $\frac{dL(\theta)}{d\theta} = 0$ which yields $\theta = \frac{n}{-\log(X_1X_2\dots X_n)}$.

11. The p.d.f. of each observation can be written as follows:

$$f(x|\theta_1,\theta_2) = \begin{cases} \frac{1}{\theta_2-\theta_1} & \text{for } \theta_1 \le x \le \theta_2, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the likelihood function is

$$f_n(x|\theta_1,\theta_2) = \frac{1}{(\theta_2-\theta_1)^n}$$

for $\theta_1 \le \min\{x_1,\dots,x_n\} \le \max\{x_1,\dots,x_n\} \le \theta_2$, and $f_n(x|\theta_1,\theta_2) = 0$ otherwise. Hence, $f_n(x|\theta_1,\theta_2)$ will be a maximum when $\theta_2 - \theta_1$ is made as small as possible. Since the smallest possible value of θ_2 is $\max\{x_1,\ldots,x_n\}$ and the largest possible value of θ_1 is $\min\{x_1,\ldots,x_n\}$, the M.L.E.'s of (θ_1,θ_2) are $(\min\{x_1,\ldots,x_n\}$,$\max\{x_1,\ldots,x_n\})$.

12. The likelihood function is $f_n(x|\theta_1,\ldots,\theta_k)=\theta_1^{n_1}\cdots\theta_k^{n_k}$.

If we let $L(\theta_1,\ldots,\theta_k)=\log f_n(x|\theta_1,\ldots,\theta_k)$ and let $\theta_k=1-\sum_{i=1}^{k-1}\theta_i$, then

$$\frac{\partial L(\theta_1,\ldots,\theta_k)}{\partial\theta_i}=\frac{n_i}{\theta_i}-\frac{n_k}{\theta_k}\quad\text{for } i=1,\ldots,k-1.$$

If each of these derivatives is set equal to 0, we obtain the relations

$$\frac{\theta_1}{n_1}=\frac{\theta_2}{n_2}=\cdots=\frac{\theta_k}{n_k}.$$

The Hessian matrix H is

$$\begin{bmatrix} \frac{\partial L(\theta_1,\ldots,\theta_k)^2}{\partial^2\theta_1} & \frac{\partial L(\theta_1,\ldots,\theta_k)^2}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial L(\theta_1,\ldots,\theta_k)^2}{\partial\theta_1\partial\theta_{k-1}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial L(\theta_1,\ldots,\theta_k)^2}{\partial\theta_1\partial\theta_{k-1}} & \frac{\partial L(\theta_1,\ldots,\theta_k)^2}{\partial\theta_2\partial\theta_{k-1}} & \cdots & \frac{\partial\partial L(\theta_1,\ldots,\theta_k)^2}{\partial^2\theta_{k-1}} \end{bmatrix}=\begin{bmatrix} -\frac{n_1}{\theta_1^2}-\frac{n_k}{\theta_k^2} & -\frac{n_k}{\theta_k^2} & \cdots & -\frac{n_k}{\theta_k^2} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{n_k}{\theta_k^2} & -\frac{n_k}{\theta_k^2} & \cdots & -\frac{n_{k-1}}{\theta_{k-1}^2}-\frac{n_k}{\theta_k^2} \end{bmatrix}$$

Write H into H_1+H_2 where

$$H_1=\begin{bmatrix} -\frac{n_1}{\theta_1^2} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -\frac{n_{k-1}}{\theta_{k-1}^2} \end{bmatrix},$$

$$H_2=\begin{bmatrix} -\frac{n_k}{\theta_k^2} & -\frac{n_k}{\theta_k^2} & \cdots & -\frac{n_k}{\theta_k^2} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{n_k}{\theta_k^2} & -\frac{n_k}{\theta_k^2} & \cdots & -\frac{n_k}{\theta_k^2} \end{bmatrix}=-\frac{n_k}{\theta_k^2}\mathbf{1}\mathbf{1}^T,$$

where $\mathbf{1}=(1,\ldots,1)^T$. Then for arbitary non-zero vector $x=(x_1,\ldots,x_{k-1})^T$

$$x^TH_1x=-\sum_{j=1}^{k-1}\frac{n_j}{\theta_j^2}x_j^2<0,$$

$$x^TH_2x=-\frac{n_k}{\theta_k^2}x^T\mathbf{1}\mathbf{1}^Tx=-\frac{n_k}{\theta_k^2}(x^T\mathbf{1})^2\leq 0.$$

So it is a concave function. If we let $\theta_i = \alpha n_i$ for $i = 1, \ldots, k$, then

$$1 = \sum_{i=1}^{k} \theta_i = \alpha \sum_{i=1}^{k} n_i = \alpha n$$

Hence $\alpha = 1/n$. It follows that $\hat{\theta}_i = n_i/n$ for $i = 1, \ldots, k$.

13. It follows from Eq. (5.10.2) (with x_1 and x_2 now replaced by x and y) that the likelihood function is

$$f_n(x, y|\mu_1, \mu_2) \propto \exp\left\{-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left[\left(\frac{x_i - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_i - \mu_1}{\sigma_1}\right)\left(\frac{y_i - \mu_2}{\sigma_2}\right) + \left(\frac{y_i - \mu_2}{\sigma_2}\right)^2\right]\right\}$$

If we let $L(\mu_1, \mu_2) = \log f(x, y|\mu_1, \mu_2)$, then

$$\frac{\partial L(\mu_1, \mu_2)}{\partial \mu_1} = \frac{1}{1-\rho^2}\left[\frac{1}{\sigma_1^2}\left(\sum_{i=1}^{n} x_i - n\mu_1\right) - \frac{\rho}{\sigma_1\sigma_2}\left(\sum_{i=1}^{n} y_i - n\mu_2\right)\right],$$

$$\frac{\partial L(\mu_1, \mu_2)}{\partial \mu_2} = \frac{1}{1-\rho^2}\left[\frac{1}{\sigma_2^2}\left(\sum_{i=1}^{n} y_i - n\mu_2\right) - \frac{\rho}{\sigma_1\sigma_2}\left(\sum_{i=1}^{n} x_i - n\mu_1\right)\right].$$

The Hessian matrix

$$H = \begin{bmatrix} \frac{\partial L(\mu_1,\mu_2)^2}{\partial^2 \mu_1} & \frac{\partial L(\mu_1,\mu_2)^2}{\partial \mu_1 \partial \mu_2} \\ \frac{\partial L(\mu_1,\mu_2)^2}{\partial \mu_1 \partial \mu_2} & \frac{\partial L(\mu_1,\mu_2)^2}{\partial^2 \mu_2} \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} -\frac{n}{\sigma_1^2} & \frac{n\rho}{\sigma_1\sigma_2} \\ \frac{n\rho}{\sigma_1\sigma_2} & -\frac{n}{\sigma_1^2} \end{bmatrix}.$$

For any non-zero vector $x = (x_1\, x_2)^T$,

$$x^T H x = -\frac{n}{1-\rho^2}\left(\left(\frac{x_1}{\sigma_1} - \frac{\rho x_2}{\sigma_2}\right)^2 + (1-\rho^2)\frac{x_2^2}{\sigma^2}\right) < 0.$$

So $L(\mu_1, \mu_2)$ is concave. When these derivatives are set equal to 0, the unique solution is $\mu_1 = \bar{x}_n$ and $\mu_2 = \bar{y}_n$. Hence, these values are the M.L.E.'s.

7.10 Supplementary Exercises

5.

$$\begin{aligned} L(\mu) = \log(f_n(x|\mu)) &= \log\left(\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(X_1 - b_1\mu)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(X_2 - b_2\mu)^2}{2\sigma_2^2}}\right) \\ &= -\log(2\pi\sigma_1\sigma_2) - \frac{(X_1 - b_1\mu)^2}{2\sigma_1^2} - \frac{(X_2 - b_2\mu)^2}{2\sigma_2^2}, \end{aligned}$$

$$\frac{dL(\mu)}{d\mu} = \frac{b_1(X_1 - b_1\mu)}{\sigma_1^2} + \frac{b_2(X_2 - b_2\mu)}{\sigma_2^2},$$

$$\frac{dL(\mu)^2}{d^2\mu} = -\frac{b_1^2}{\sigma_1^2} - \frac{b_2^2}{\sigma_2^2} < 0,$$

So the M.L.E can be solved by $\frac{dL(\mu)}{d\mu} = 0$ which yields $\mu = \frac{b_1\sigma_2^2 X_1 + b_2\sigma_1^2 X_2}{b_1^2\sigma_2^2 + b_2^2\sigma_1^2}$.

Additional Problem

1.Show that for a random sample $X_1, ..., X_n$ it holds that

i) $\min_a \sum_{i=1}^n (X_i - a)^2 = \sum_{i=1}^n (X_i - \bar{X})^2$.

$$\begin{aligned}
\sum_{i=1}^n (X_i - a)^2 &= \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - a)^2 \\
&= \sum_{i=1}^n [(X_i - \bar{X})^2 + (\bar{X} - a)^2 + 2(X_i - \bar{X})(\bar{X} - a)] \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - a)^2 + 2(\bar{X} - a)\sum_{i=1}^n (X_i - \bar{X}) \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - a)^2 \\
&\geq \sum_{i=1}^n (X_i - \bar{X})^2
\end{aligned}$$

ii) $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$.

$$\begin{aligned}
\sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\
&= \sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2 \\
&= \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\
&= \sum_{i=1}^n X_i^2 - n\bar{X}^2
\end{aligned}$$

iii) $\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{2n}\sum_{i,j}(X_i - X_j)^2$.

$$\begin{aligned}
\sum_{i,j}(X_i - X_j)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} - (X_j - \bar{X}))^2 \\
&= \sum_{i,j}((X_i - \bar{X})^2 - 2(X_i - \bar{X})(X_j - \bar{X}) + (X_j - \bar{X})^2) \\
&= \sum_{i,j}((X_i - \bar{X})^2 + (X_j - \bar{X})^2) - 2\sum_{i,j}(X_i - \bar{X})(X_j - \bar{X}) \\
&= 2n\sum_{i=1}^{n}(X_i - \bar{X})^2 - 2(\sum_{i=1}^{n}(X_i - \bar{X}))(\sum_{j=1}^{n}(X_j - \bar{X})) \\
&= 2n\sum_{i=1}^{n}(X_i - \bar{X})^2
\end{aligned}$$

iv) If the population from which the random sample is drawn has mean μ and variance σ^2, then furthermore $E(\bar{X}) = \mu, \mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}$.

$$E(\bar{X}) = E(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

$$\mathrm{var}(\bar{X}) = \mathrm{var}(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(X_i) = \frac{\sigma^2}{n}$$

v) The moment generating function (m.g.f) of $\bar{X}$ is $\phi(t) = [\phi_0(t/n)]^n$, where $\phi_0(t) = E(e^{tX})$ is the m.g.f of X.

$$\phi(t) = E(e^{t\bar{X}}) = E(e^{\sum_{i=1}^{n}\frac{tX_i}{n}}) = E(\prod_{i=1}^{n} e^{\frac{tX_i}{n}}) = \prod_{i=1}^{n} E(e^{\frac{tX_i}{n}}) = [\phi_0(t/n)]^n$$