131B HW#1 solution

3.7 Multivariate Distributions

1. (a) We have

$$\int_0^1 \int_0^1 \int_0^1 f(x_1, x_2, x_3) dx_1 dx_2 dx_3 = 3c$$

Since the value of this integral must be equal to 1, it follows that $c = 1/3$.

(b) For $0 \le x_1 \le 1$ and $0 \le x_3 \le 1$,

$$f_{13}(x_1, x_3) = \int_0^1 f(x_1, x_2, x_3) dx_2 = \frac{1}{3}(x_1 + 1 + 3x_3).$$

(c) The marginal joint p.d.f of X_1 and X_2 is

$$f_{12}(x_1, x_2) = \int_0^1 f(x_1, x_2, x_3) dx_3 = \frac{1}{3}(x_1 + 2x_2 + \frac{3}{2}).$$

The conditional p.d.f. of X_3 given that $X_1 = x_1$ and $X_2 = x_2$ is

$$g_3\left(x_3 | x_1, x_2\right) = \frac{f(x_1, x_2, x_3)}{f_{12}(x_1, x_2)} = \frac{x_1 + 2x_2 + 3x_3}{x_1 + 2x_2 + 3/2}$$

Therefore,

$$\begin{aligned} P\left(X_3 < \frac{1}{2} | X_1 = \frac{1}{4}, X_2 = \frac{3}{4}\right) &= \int_0^{1/2} g_3\left(x_3 | x_1 = \frac{1}{4}, x_2 = \frac{3}{4}\right) dx_3 \\ &= \int_0^{1/2} \left(\frac{7}{13} + \frac{12}{13} x_3\right) dx_3 = \frac{5}{13}. \end{aligned}$$

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

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

3.9 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

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

18. We need to transform (X, Y) to (Z, W), where $Z = X/Y$ and $W = Y$. The joint p.d.f. of (X, Y) is

$$f(x, y) = g_1(x|y) f_2(y) = \begin{cases} 3x^2 f_2(y)/y^3 & \text{if } 0 < x < y, \\ 0 & \text{otherwise.} \end{cases}$$

The inverse transformation is $x = zw$ and $y = w$. The jacobian is

$$J = \det \begin{pmatrix} w & z \\ 0 & 1 \end{pmatrix} = w.$$

The joint p.d.f. of (Z, W) is

$$g(z, w) = f(zw, w) w = 3z^2 w^2 f_2(w) w / w^3 = 3z^2 f_2(w), \quad \text{for } 0 < z < 1.$$

This is clearly factored in the appropriate way to show that Z and W are independent. Indeed, if we integrate $g(z, w)$ over w, we obtain the marginal p.d.f. of Z, namely $g_1(z) = 3z^2$, for $0 < z < 1$.

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

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

5.7 The Gamma Distributions

8. For any number $y > 0$,

$$\begin{aligned}
P(Y > y) &= P(X_1 > y, \ldots, X_k > y) = P(X_1 > y) \cdots P(X_k > y) \\
&= \exp(-\beta_1 y) \cdots \exp(-\beta_k y) = \exp(-(\beta_1 + \cdots + \beta_k)y),
\end{aligned}$$

which is the probability that an exponential random variable with parameter $\beta_1 + \cdots + \beta_k$ is greater than y. Hence, Y has the exponential distribution with parameter $\beta_1 + \cdots + \beta_k$.

6.5 Supplementary Exercises

9. Let $X_1, \ldots, X_{16}$ be the times required to serve each of the 16 customers, and each X_i has the exponential distribution with parameter $1/3$. According to Theorem 5.7.8, $E(X_1) = 3$

and $\mathrm{var}(X_1) = 9$. Let $Y = \sum_{k=1}^{16} X_k$ be the total time to serve the 16 customers. The central limit theorem approximation to the distribution of Y is the normal distribution with mean $E(Y) = \sum_{k=1}^{16} E(X_k) = 16 \times 3 = 48$ and variance $\mathrm{var}(Y) = \sum_{k=1}^{16} \mathrm{var}(X_k) = 16 \times 9 = 144$. The approximate probability that $Y > 60$ is

$$1 - \Phi\left(\frac{60 - 48}{\sqrt{144}}\right) = 1 - \Phi(1) = 0.1587.$$

7.5 Maximum Likelihood Estimators

5. Let $y = \sum_{i=1}^{n} x_i$. Then the likelihood function is

$$f_n(x|\theta) = \frac{\exp(-n\theta)\theta^y}{\prod_{i=1}^{n}(x_i!)}$$

(a) If $y > 0$ and we let $L(\theta) = \log f_n(x|\theta)$, then

$$\frac{\partial}{\partial\theta}L(\theta) = -n + \frac{y}{\theta}.$$

The maximum of $L(\theta)$ will be attained at the value of θ for which this derivative is equal to 0. In this way, we find that $\theta = y/n = \bar{x}_n$.

(b) If $y = 0$, then $f_n(x|\theta)$ is a decreasing function of θ. Since $\theta = 0$ is not a value in the parameter space, there is no M.L.E.

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 $\max\{x_1, \dots, x_n\}$ and the largest possible value of θ_1 is $\min\{x_1, \dots, x_n\}$, the M.L.E.'s of (θ_1, θ_2) are $(\min\{x_1, \dots, x_n\}, \max\{x_1, \dots, x_n\})$.

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

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

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

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

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

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

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

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

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.

7.6 Properties of Maximum Likelihood Estimators

6. The distribution of $Z = (X - \mu)/\sigma$ will be a standard normal distribution. Therefore,

$$0.95 = P(X < \theta) = P(Z < \frac{\theta - \mu}{\sigma}) = \Phi\left(\frac{\theta - \mu}{\sigma}\right).$$

Hence, from a table of the values of Φ it is found that $(\theta - \mu)/\sigma = 1.645$. Since $\theta = \mu + 1.645\sigma$, it follows that $\hat{\theta} = \hat{\mu} + 1.645\hat{\sigma}$. By example 7.5.6, we have

$$\hat{\mu} = \bar{X}_n \quad \text{and} \quad \hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right]^{1/2}.$$

14. For $1 \le i \le n$, let the random variable $X_i = 1$ if the monarch butterfly has a special type of marking on its wings and $X_i = 0$ otherwise. Then $X_1, \ldots, X_n$ form a random sample from the Bernoulli distribution with parameter p.

a. The number n of observations is random and has the pdf

$$P(n = k) = \binom{k-1}{4} p^5 (1-p)^{k-5} \quad \text{for } k = 5, 6, \ldots$$

The reason is $n = k$ follows from the fact that there are 4 monarch butterflies with special marking in the first $k-1$ captures and the last capture must be a butterfly with special marking. So the probability is $\binom{k-1}{4} p^4 (1-p)^{k-5} p = \binom{k-1}{4} p^5 (1-p)^{k-5}$.

Based on the observed values, the likelihood function is

$$p^5 (1-p)^{43-5} \binom{43-1}{4} p^5 (1-p)^{43-5} \propto p^{10} (1-p)^{76}$$

By taking the derivative of log-likelihood function and setting it to 0, we obtain the M.L.E. is 10/86=5/43.

b. The number of observations is random but does not depend on p. Based on the observed values, the likelihood function is

$$p^3 (1-p)^{58-3}$$

By taking the derivative of log-likelihood function and setting it to 0, we obtain the M.L.E. is 3/58.

Therefore, the M.L.E. of p is equal to the proportion of butterflies in the sample that have the special marking, regardless of the sampling plan.

Extra Problems

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 $1 \le i \le n$, 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 $\hat{p} = 2/3$.

6.5 Supplementary Exercises

9. Let $Y = \sum_{k=1}^{16} X_k$ be the total time to serve the 16 customers. The X_i's are independent and have the exponential distribution with parameter $1/3$. The m.g.f of Y is

$$\begin{aligned}
\psi(t) &= E(\exp(tY)) = E[\exp(t(\sum_{k=1}^{16} X_k))] \\
&= \prod_{k=1}^{16} E[\exp(tX_k)] \\
&= \prod_{k=1}^{16} \frac{1/3}{1/3 - t} \qquad \text{for } t < 1/3 \\
&= \left(\frac{1/3}{1/3 - t}\right)^{16} \qquad \text{for } t < 1/3
\end{aligned}$$

This corresponds to the m.g.f. of the gamma distribution with parameters 16 and 1/3. Thus,

$$P(Y > 60) = \int_{60}^{\infty} \frac{(1/3)^{16}}{\Gamma(16)} x^{15} e^{-x/3} dx = 0.1565.$$