Statistics 5101, Fall 2000, Geyer

Homework Solutions #10

Problem L5-3

Let $Y_i = X_i^k$, then $Y_1, Y_2, \ldots$ is a sequence of independent identically distributed random variables (functions of independent random variables are independent by Theorem 13 of Chapter 3 in Lindgren) with expectation

$$\mu_Y = E(Y_i) = E(X^k).$$

Then the LLN says

$$\overline{Y}_n \xrightarrow{P} \mu_Y$$

but this is just other notation for

$$\frac{1}{n} \sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k).$$

Problem L5-6

Write Y for the weight of the 100 booklets. Then

$$\begin{aligned} E(Y) &= 100 \\ \operatorname{var}(Y) &= 100 \times .02^2 = .04 \end{aligned}$$

so

$$P(Y > 100.5) = 1 - P(Y < 100.5) = 1 - \Phi\left(\frac{100.5 - 100}{\sqrt{100} \times .02}\right) = 1 - \Phi(2.5) = .0062$$

Problem L5-9

Let $Y \sim \mathcal{U}(-0.5, 0.5)$ be one error, then from the appendix on brand name distributions

$$\begin{aligned} E(Y) &= 0 \\ \operatorname{var}(Y) &= \frac{1}{12} \end{aligned}$$

If W is the sum of n i. i. d. such errors then

$$\begin{aligned} E(W) &= 0 \\ \operatorname{var}(W) &= \frac{n}{12} \end{aligned}$$

Thus

$$\begin{aligned}
P\left(|W| < \sqrt{n}/2\right) &= P\left(-\sqrt{n}/2 < W < \sqrt{n}/2\right) \\
&= \Phi\left(\frac{\sqrt{n}/2 - 0}{\sqrt{n/12}}\right) - \Phi\left(\frac{-\sqrt{n}/2 - 0}{\sqrt{n/12}}\right) \\
&= 1 - 2\Phi(-\sqrt{3}) \\
&= 0.9167355
\end{aligned}$$

Problem L6-13

By direct count, the probability of a sum of 5 or less rolling a pair of dice is 5/18. Thus, if Y is the number of such rolls in 72 tries, $Y \sim \text{Bin}(72, 5/18)$, and

$$\begin{aligned}
E(Y) &= 72 \times \frac{5}{18} = 20 \\
\text{var}(Y) &= 72 \times \frac{5}{18} \times \frac{13}{18} = 14.4444 \\
\text{sd}(Y) &= \sqrt{14.4444} = 3.8006
\end{aligned}$$

So, using a continuity correction,

$$P(Y \geq 28) = 1 - \Phi\left(\frac{27 + 0.5 - 20}{3.8006}\right) = .0242$$

Problem L6-86

From a picture of the triangular density, the two inside intervals have three times the probability of the outside intervals. Thus the probabilities of the intervals are $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, and $\frac{1}{8}$.

Let X_1, X_2, X_3, and X_4 be the counts in the cells (1, 2, 2, 1), then this is a multinomial random vector and the probability of these counts is

$$\begin{aligned}
\binom{n}{x_1, x_2, x_3, x_4} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4} &= \frac{6!}{1!\,2!\,2!\,1!}\left(\frac{1}{8}\right)^1\left(\frac{3}{8}\right)^2\left(\frac{3}{8}\right)^2\left(\frac{1}{8}\right)^1 \\
&= 180 \cdot \frac{3^4}{8^6} \\
&= 0.0556183
\end{aligned}$$

Problem L12-12

Since it is a linear transformation of a multivariate normal random vector, (X, Y) is also multivariate normal with mean vector zero because

$$\begin{aligned}
E(X) &= E(U) + 2E(V) = 0 \\
E(Y) &= 3E(U) - E(V) = 0
\end{aligned}$$

and variance matrix $\mathbf{M}$ with components

$$\begin{aligned}
m_{11} &= \operatorname{var}(X) \\
&= \operatorname{var}(U + 2V) \\
&= \operatorname{var}(U) + 4\operatorname{var}(V) \\
&= 5 \\
m_{22} &= \operatorname{var}(Y) \\
&= \operatorname{var}(3U - V) \\
&= 9\operatorname{var}(U) + \operatorname{var}(V) \\
&= 10 \\
m_{12} &= \operatorname{cov}(X, Y) \\
&= \operatorname{cov}(U + 2V, 3U - V) \\
&= 3\operatorname{var}(U) - 2\operatorname{var}(V) \\
&= 1 \\
m_{21} &= m_{12}
\end{aligned}$$

Problem N5-7

From the variance formula for the multinomial in the appendix on brand name distributions

$$\begin{aligned}
\operatorname{var}(X_i - X_j) &= \operatorname{var}(X_i) + \operatorname{var}(X_j) - 2\operatorname{cov}(X_i, X_j) \\
&= np_i(1 - p_i) + np_j(1 - p_j) + 2np_ip_j \\
&= n[p_i + p_j - (p_i - p_j)^2]
\end{aligned}$$

Problem N5-10

The problem is to specialize the formula

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\det(\mathbf{M})^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

for the density of the multivariate normal to the two-dimensional case, when the mean vector is

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$$

and the variance matrix is

$$\mathbf{M} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

Using the hints

$$\det(\mathbf{M}) = \sigma_X^2\sigma_Y^2(1 - \rho^2)$$

and

$$\begin{aligned}\mathbf{M}^{-1} &= \frac{1}{\sigma_X^2 \sigma_Y^2 (1 - \rho^2)} \begin{pmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix} \\ &= \frac{1}{(1-\rho^2)} \begin{pmatrix} \frac{1}{\sigma_X^2} & -\frac{\rho}{\sigma_X\sigma_Y} \\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{pmatrix}\end{aligned}$$

The constant part of the density is now done

$$\frac{1}{(2\pi)^{n/2} \det(\mathbf{M})^{1/2}} = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

because $n = 2$. So the only thing left is to match up the quadratic form in the exponent.

In general a quadratic form is written out explicitly in terms of components as

$$\begin{aligned}\mathbf{z}'\mathbf{A}\mathbf{z} &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} z_i z_j \\ &= \sum_{i=1}^n a_{ii} z_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_{ij} z_i z_j\end{aligned}$$

In this case the quadratic form in the exponent is

$$\begin{aligned}&(\mathbf{x} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &\qquad = \frac{1}{(1-\rho^2)} \left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right)\end{aligned}$$

which is the quadratic form in the formula to be proved. So we're done.

Problem N5-11

In this case the elements of the partitioned variance matrix are all scalars

$$\begin{aligned}\mathbf{M}_{11} &= \sigma_X^2 \\ \mathbf{M}_{12} &= \rho\sigma_X\sigma_Y \\ \mathbf{M}_{22} &= \sigma_Y^2 \\ \mathbf{M}_{22}^{-1} &= \frac{1}{\sigma_Y^2}\end{aligned}$$

Hence

$$
\begin{aligned}
E(X \mid Y) &= \boldsymbol{\mu}_1 + \mathbf{M}_{12}\mathbf{M}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \\
&= \mu_X + \rho\sigma_X\sigma_Y \cdot \frac{1}{\sigma_Y^2}(Y - \mu_Y) \\
&= \mu_X + \rho\frac{\sigma_X}{\sigma_Y^2}(Y - \mu_Y) \\
\operatorname{var}(X \mid Y) &= \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21} \\
&= \sigma_X^2 - \rho\sigma_X\sigma_Y \cdot \frac{1}{\sigma_Y^2}\rho\sigma_X\sigma_Y \\
&= \sigma_X^2(1 - \rho^2)
\end{aligned}
$$

Problem N5-12

We are to calculate $P\{q(\mathbf{X}) < d\}$ for given d, where

$$q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

and

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{M})$$

Now Problem 12-32 in Lindgren referred to in the hint says almost the same what we want

$$q_2(\mathbf{Y}) = \mathbf{Y}'\mathbf{M}^{-1}\mathbf{Y} \sim \text{chi}^2(p)$$

where

$$\mathbf{Y} \sim \mathcal{N}(0, \mathbf{M})$$

The only differences are (1) we have no means subtracted off in q_2 and (2) $\mathbf{Y}$ has mean zero. However,

$$q(\mathbf{X}) = q_2(\mathbf{X} - \boldsymbol{\mu})$$

and

$$\mathbf{X} - \boldsymbol{\mu} \sim \mathcal{N}(0, \mathbf{M})$$

so we can apply the 12-32 to this problem obtaining

$$q(\mathbf{X}) \sim \text{chi}^2(p)$$

Thus

$$P\{q(\mathbf{X}) < d\} = F(d),$$

where F is the the c. d. f. of the $\text{chi}^2(p)$ distribution.

Problem N5-13

(a) Write

$$\mathbf{Z} = \begin{pmatrix} U - V \\ V - W \end{pmatrix}$$

Then

$$\mathbf{Z} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} U \\ V \\ W \end{pmatrix}$$

thus is a linear transformation of multivariate normal, hence multivariate normal with

$$E(\mathbf{Z}) = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$\operatorname{var}(\mathbf{Z}) = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

(b) From the formula for the variance,

$$\operatorname{var}(Z_1) = \operatorname{var}(Z_2) = 2$$

and

$$\operatorname{cor}(Z_1, Z_2) = -\frac{1}{2}$$

Thus the conditional distribution of Z_1 given Z_2 is normal with mean

$$E(Z_1 \mid Z_2) = -\frac{1}{2} \cdot Z_2$$

and variance

$$\operatorname{var}(Z_1 \mid Z_2) = 2 \left[1 - \left(-\frac{1}{2} \right)^2 \right] = \frac{3}{2}$$