2005
STAT354
ADDENDA

The purpose of this document is to record material produced in response to students requests. This will be done by Chapters, and will record answers to queries, worked examples and topics to supplement the Study Guide.
Chapters 1 and 2

Query

The first request is to answer a question about how to tackle Q1 in Assignment 1.

Answer

Look at the Unit Notes pages 8 to 9.

Query

Moment Generating Functions (MGFs):

MGFs are a most powerful tool in statistics as they allow you to solve problems about the sums, differences and linear combinations of random variables quickly without the need for transformations of variables. The procedure can be used for univariate and multivariate problems.

Revision

The MGF of a random variable $Y$ is defined as

$$E(e^{yt}).$$

The MGF generates the (gasp!) moments of the distribution, viz,

$$E(e^{yt}) = \int e^{yt} f(y)dy = \int (1 + yt + (yt)^2/2 + (yt)^3/6 + \ldots) f(y)dy$$

$$= 1 + \mu t + \mu' t^2/2 + \ldots$$

where the moments are about the origin. Thus

$$\mu'_1 = \mu = \int yf(y)dy, \mu'_2 = \int y^2f(y)dy$$

e tc.

The key reason for the utility of the MGF is the uniqueness property. That is, each distribution has a unique set of moments. Or, the moments define the distribution uniquely and vice versa.

Two examples are given to demonstrate this property.
Bernoulli distribution

The pdf is

$$P(Y|\pi) = \pi^y(1 - \pi)^{1-y}, \ y = 0, 1$$

with mgf

$$M(t) = Ee^{Yt} = \sum_{y=0,1} e^{yt}\pi^y(1 - \pi)^{1-y} = e^{1t}\pi + e^{0t}(1 - \pi) = e^t\pi + (1 - \pi)$$

$$= (1 + t + t^2/2 + \ldots)\pi + 1 - \pi = \pi + \pi t + \pi t^2/2 + \ldots + 1 - \pi$$

and so

$$M(t) = 1 + \pi t + \pi t^2/2 + \ldots$$

and so all the moments for the Bernoulli are the same ($\pi$).

Thus if you are told that an unknown distribution has all its moments equal, you can identify it as a Bernoulli distribution.

Normal distribution

$$Y \sim N(\mu, \sigma^2)$$

What is the MGF?

$$M(t) = E(e^{Yt}) = \int e^{yt} \frac{1}{\sigma\sqrt{2\pi}} e^{-(y - \mu)^2/2\sigma^2} dy$$

$$= \int \frac{1}{\sigma\sqrt{2\pi}} e^{yt} - (y - \mu)^2/2\sigma^2 dy$$

Consider the term in the exponent, ie, on expansion

$$yt - (y - \mu)^2/2\sigma^2 = yt - \frac{y^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} + \frac{2\mu y}{2\sigma^2}$$

$$= \frac{2yt\sigma^2 - y^2 - \mu^2 + 2\mu y}{2\sigma^2}$$

$$= \frac{2y(t\sigma^2 + \mu) - y^2 - \mu^2}{2\sigma^2}$$

Now completing the square gives

$$= \frac{2y(t\sigma^2 + \mu) - y^2 - (\mu + t\sigma^2)^2 + (t\sigma^2)^2 + 2\mu t\sigma^2}{2\sigma^2}$$
\[
\frac{-(y - \mu - t\sigma^2)^2 + t^2\sigma^4 + 2\mu t\sigma^2}{2\sigma^2} = \frac{-(y - \mu - t\sigma^2)^2}{2\sigma^2} + t^2\sigma^2/2 + \mu t
\]

And so the mgf becomes

\[
M(t) = \int e^{t^2\sigma^2/2 + \mu t} \frac{1}{\sigma\sqrt{2\pi}} e^{-(y - \mu - t\sigma^2)^2/2\sigma^2} dy
\]

\[
= e^{t^2\sigma^2/2 + \mu t} \int \frac{1}{\sigma\sqrt{2\pi}} e^{-(y - \mu - t\sigma^2)^2/2\sigma^2} dy
\]

\[
= e^{t^2\sigma^2/2 + \mu t}
\]

Thus

\[
\ln M(t) = \mu t + \frac{t^2\sigma^2}{2}
\]

The function \(\ln M(t)\) is called the cumulant generating function, and the moments generated from such a function are called cumulants. Thus the Normal distribution has only two cumulants. This feature is peculiar to the Normal and in fact characterises the distribution. Thus any unknown distribution with only two cumulants can be identified as a Normal distribution, with the two cumulants being the mean and the variance.
Query

Deriving and manipulating the joint distribution function.

Example

A point $X$ is chosen from a uniform distribution on $(0, 1)$. After $X = x$ has been observed, a point $Y$ is then chosen from a uniform distribution on the interval $(x, 1)$.

1. Find the conditional distribution of $Y$ given $X = x$.

2. Hence give the joint distribution of $X$ and $Y$. Verify that it is indeed a pdf.

3. Find the marginal distribution of $Y$. Verify that it is a pdf.

Solution

1. By definition

   \[ f_2(y|x) = \frac{1}{1-x}, \quad x < y < 1 \]

   This is a pdf, since

   \[ \int_x^1 \frac{1}{1-x} dy = \frac{1}{1-x} \int_x^1 dy = 1 \]

2. Since

   \[ f_2(y|x) = \frac{f(x, y)}{f_1(x)} \]

   by definition then

   \[ f(x, y) = f_2(y|x) \cdot f_1(x) = \frac{1}{1-x} \cdot 1 = \frac{1}{1-x}, \quad 0 < x < y < 1 \]

   Is this a pdf? To check verify that \( \int \int f(x, y) dx dy = 1 \). The range of definition is a triangle contained in a unit square, with vertices $(0,0)$, $(0,1)$ and $(1,1)$. Thus we need to show that

   \[ \int_{y=0}^{y=1} \int_{x=0}^{x=y} f(x, y) dx dy = 1 \]
Now
\[
\int_{y=0}^{y=1} \int_{x=0}^{x=y} \frac{1}{1-x} \, dx \, dy = \int_0^1 [- \log(1-x)]_0^y \, dy = \int_0^1 - \log(1-y) \, dy = 1
\]
by parts,
\[
\int_0^1 - \log(1-y) \, dy = [-y \log(1-y)]_0^1 - \int_0^1 \frac{y}{1-y} \, dy
\]
\[
= [-y \log(1-y)]_0^1 + [y + \log(1-y)]_0^1
\]
\[
= [(1-y) \log(1-y)]_0^1 + 1 = 1
\]
as required.

3. The marginal of \(Y\) is given by
\[
f_2(y) = \int f(x,y) \, dx = \int_{x=0}^{x=y} \frac{1}{1-x} \, dx = [- \log(1-x)]_0^y = - \log(1-y), \quad 0 < y < 1.
\]

As this is the final integrand in 2., then \(f_2(y)\) is indeed a pdf.
Query

A joint distribution that is a combination of discrete and continuous random variables.

A quality control plan calls for randomly selecting 3 items from the daily output (assumed large) and observing the number of defectives \( Y \). The proportion of defectives \( \pi \) varies and has the uniform distribution on \((0, 1)\). Find the expected number of defectives to be observed. (Hint: common sense tells us \( E(Y) = ? \)) Describe the joint distribution.

Solution

Now \( Y|\pi \sim B(3, \pi) \) and so

\[
P(Y = y|\pi) = \binom{3}{y} \pi^y (1 - \pi)^{3-y}, \quad y = 0, 1, 2, 3
\]

and

\[
f(\pi) = 1, \quad 0 < \pi < 1
\]

The joint distribution is found via

\[
P(Y|\pi) = \frac{f(y, \pi)}{f(\pi)}
\]

to give

\[
f(y, \pi) = \binom{3}{y} \pi^y (1 - \pi)^{3-y}, \quad y = 0, 1, \ldots, 3 : \quad 0 < \pi < 1
\]

which is a series of 'vanes'.

There are two methods of calculating the expected number of defectives \( E(Y) \):

1. Now

\[
E(Y|\pi) = 3\pi
\]

and

\[
E(Y) = E_X(E_Y(Y|X)) = E_\pi(E_Y(Y|\pi)) = E_\pi(3\pi) = 3 \cdot \frac{1}{2} = 1.5
\]
2. Also

\[ E(Y) = \int y f_2(y) dy \]

and

\[ f_2(y) = P(Y = y) = \int_0^1 f(y, \pi) d\pi = \int_0^1 \left( \frac{3}{y} \right) \pi^y (1 - \pi)^{3-y} d\pi \]

which gives

\[ P(Y = y) = 1/4, \ y = 0, 1, 2, 3. \]

For example

\[ P(Y = 3) = \int_0^1 \pi^3 d\pi = \left[ \frac{\pi^4}{4} \right]_0^1 = 1/4 \]

Thus

\[ E(Y) = (1 + 2 + 3) \cdot 1/4 = 6/4 = 1.5 \]

as before.
Contours of $2x^2 + xy + 2y^2$

![Contour Plot](image)

Figure 1: Contour plot of $2x^2 + xy + 2y^2$

**Eigenvalues and eigenvectors**

These are best described via a simple example, using coordinate geometry.

Consider the ellipse

$$2x^2 + xy + 2y^2 = c$$

where $c$ is a constant. The level curves (plots of the ellipse for varying $c$) are shown in the contour plot in Figure 1.

The R code below produced the plot.

```r
x <- seq(-10, 10, len=100)
y <- x
f <- outer(x,y,function(x,y) 2*x^2 + x * y + 2* y^2 )
oldpar <- par(no.readonly=T)
```
par(pty="s")
contour(x,y,f,xlab="x",ylab="y",main="Contours of \(2x^2 + xy + 2y^2\)"

The calculation of eigenvalues and eigenvectors can be interpreted as finding the simplified form of this ellipse centered at the origin following a simple linear coordinate transformation.

To calculate the eigenvalues we solve the equation

\[ |A - \lambda I| = 0 \]

where \(A\) is the matrix for the quadratic form describing the equation of the ellipse, viz,

\[
A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}
\]

where the ellipse is written as

\[
x'Ax = c
\]

using

\[
x' = (x, y)
\]

The equation

\[ |A - \lambda I| = 0 \]

becomes

\[(2 - \lambda) = \pm 1\]

giving \(\lambda = 1, 3\).

To calculate the eigenvectors we solve the equation

\[
xA = \lambda x
\]

Thus for \(\lambda = 1\) we get

\[
\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}
\]

which gives \(x + y = 0\) and so the eigenvector is \(\begin{pmatrix} 1 \\ -1 \end{pmatrix}\)

For \(\lambda = 3\) we find that the eigenvector is \(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\)
Thus if we choose a transformation of coordinates

\[
\begin{pmatrix}
X \\
Y
\end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix}
x \\
y
\end{pmatrix}
\]

then this should produce an ellipse of the form

\[X^2 + 3Y^2 = c\]

as shown in the contour plot. (The ratio of the major to the minor semi axes is in proportion to the inverse square roots of the eigenvalues \ldots)

This derived equation can be verified by substituting the back transformed variables

\[x = \frac{X + Y}{2}\]

and

\[y = \frac{Y - X}{2}\]

into the original equation of the ellipse.

Note that to get the scaling to be correct, each of the eigenvectors should be normalised.

Thus to get \( P \) such that

\[P'AP = \text{diag}(\lambda_1, \ldots, \lambda_p)\]

we use

\[P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\]

You can verify that

\[P'AP = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\]

as required.

Note that the direction given by the coordinate transformation is consistent with the orientation of the ellipse given in the plot.

**Q8 Assignment 1**

Full Solution

Eigenvalues : 

11
\[ A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & -1 \\ 1 & -1 & -2 \end{pmatrix} \]

\[ |A - \lambda I| = \begin{vmatrix} 1 - \lambda & 2 & 1 \\ 2 & 1 - \lambda & -1 \\ 1 & -1 & -2 - \lambda \end{vmatrix} = 0 \]

Expanding by cofactors gives:

\[ a_{11} \]

\[ (1 - \lambda) \begin{vmatrix} 1 - \lambda & -1 \\ -1 & -2 - \lambda \end{vmatrix} = -\lambda^3 + 4\lambda - 3 \]

\[ -a_{12} \]

\[ -2 \begin{vmatrix} 2 & 1 \\ 1 & -2 - \lambda \end{vmatrix} = 6 + 4\lambda \]

\[ a_{31} \text{ (could use } a_{13}) \]

\[ +1 \begin{vmatrix} 2 & 1 - \lambda \\ 1 & -1 \end{vmatrix} = -3 + \lambda\lambda \]

to give

\[ |A - \lambda I| = -\lambda^3 + 9\lambda = 0 \]

and so

\[ \lambda = 0, 3, -3 \]

Eigen vectors:

The equation \[ Ax = \lambda x \]
gives:

\[ \lambda = 0 \]

\[ x_1 + 2x_2 + x_3 = 0 \]

\[ 2x_1 + x_2 - x_3 = 0 \]

\[ x_1 - x_2 - 2x_3 = 0 \]
Equation 1 plus equation 2 gives $x_1 = -x_2$ which for the 3rd equation gives $x_3 = x_1$, and so the vector is

\[
\begin{pmatrix}
1 \\ -1 \\ 1
\end{pmatrix}
\]

$\lambda = 3$

\[
x_1 + 2x_2 + x_3 = 3x_1 \\
2x_1 + x_2 - x_3 = 3x_2 \\
x_1 - x_2 - 2x_3 = 3x_3
\]

Equation 2 minus 2 times equation 3 give $x_3 = 0$ which for the first equation gives $x_1 = x_2$, and so the vector is

\[
\begin{pmatrix}
1 \\ 1 \\ 0
\end{pmatrix}
\]

$\lambda = -3$

\[
x_1 + 2x_2 + x_3 = -3x_1 \\
2x_1 + x_2 - x_3 = -3x_2 \\
x_1 - x_2 - 2x_3 = -3x_3
\]

Equation 1 plus equation 2 give $x_1 = -x_2$ which for the third equation gives $x_3 = -2x_1$, and so the vector is

\[
\begin{pmatrix}
1 \\ -1 \\ -2
\end{pmatrix}
\]

Normalising each vector gives

\[
P = \begin{pmatrix}
1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\
-1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\
1/\sqrt{3} & 0 & -2/\sqrt{6}
\end{pmatrix}
\]
Chapter 3

Jacobians

In terms of change of variable, the Jacobian is simply a scaling factor, resulting from the change of scale when moving from one variable ($X$) to another ($Y$), say.

Assuming that we have a 1:1 transformation $Y = \phi(X)$, then we can say that

$$P(x < X < x + \delta x) = P(y < Y < y + \delta y)$$

and since the probabilities can be approximated by the height of the density times the interval, we have

$$f(x)\delta x \approx f(y)\delta y$$

which in the limit becomes

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

To be strictly correct, using $Y = \phi(X)$ this should be

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(X = \psi(Y)) \left| \frac{d\psi(x)}{dy} \right|$$

where $X = \psi(Y)$ is the inverse function of $Y = \phi(X)$. Thus the RHS is a function of $y$, as per the Notes p17.

Example

What is the distribution of $Y = X/\sigma$ if $X \sim N(0, \sigma^2)$?

Now

$$Y = \phi(X) = X/\sigma$$

and so

$$X = \psi(Y) = \sigma Y$$

giving the Jacobian as

$$|J| = \frac{dx}{dy} = \frac{d\psi(y)}{dy} = \sigma.$$ 

Since

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x - 0)^2/2\sigma^2}$$
then

\[ f(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\sigma y - 0)^2}{2\sigma^2} |J|} \]

\[ f(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-y^2/2\sigma} = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \]

which is the familiar unit normal distribution.

In two dimensions, the Jacobian becomes a matrix, but as we only need its determinant, it again boils down to a scaling factor, as the determinant is a scalar.

See pages 24/25 of the Notes for a full description.
Exercise page 22

Solutions

1. The PIT is $Y$ such that $Y = F(X)$.
   
   - If $X \sim U(a, b)$ then $f(x) = 1/(b - a)$, $a < x < b$ and then
     
     $$F(x) = P(X \leq x) = \int_{t=a}^{t=x} \frac{1}{b - a} dt = \left[ \frac{x - a}{b - a} \right]$$
     
     Thus
     
     $$Y = \frac{x - a}{b - a} = \phi(X)$$
     
     and
     
     $$X = (b - a)Y + a = \psi(Y)$$
     
   - $F(x) = \int_{t=0}^{t=x} 2tdt = x^2$
     
     and so
     
     $$Y = X^2 = \phi(X)$$
     
     with
     
     $$X = \sqrt{Y} = \psi(Y)$$
     
     since $x > 0$.

2. - The Uniform distribution :

   $$f(y) = f(\psi(y)) \cdot |J| = \frac{1}{b - a} \cdot \left| \frac{d\psi(y)}{dy} \right| = \frac{1}{b - a} \cdot (b - a) = 1$$

   as required.

   - The Triangular distribution :

   $$f(y) = f(\psi(y)) \cdot |J| = 2\sqrt{y} \cdot \left| \frac{d\psi(y)}{dy} \right| = 2\sqrt{y} \cdot \frac{1}{2\sqrt{y}} = 1$$

   as expected.
3. If you wish to generate random numbers with a pdf $f(x)$, then generate $U(0, 1) = Y$ random numbers and use the PIT $X = F^{-1}(Y)$ to produce the desired distribution. For example, say you wish to generate exponentially distributed random numbers, viz,

$$f(x) = e^{-x}, \; x > 0.$$ 

The inverse transformation is the logarithmic, ie, $X = \log(Y)$, and so $\log U(0, 1)$ will generate the required distribution. (Try it!)
Query

Could you illustrate the method of completing the square using the exponential value half-way down page 26 of the Unit Notes?

Solution

\[
\frac{U^2 + V^2 - 2UV + V^2}{2} = \frac{U^2}{2} + V^2 - UV = \frac{U^2}{2} + (V^2 - 2V(U/2)) = \frac{U^2}{2} + (V^2 - 2V(U/2) + U^2/4 - U^2/4) = \frac{U^2}{4} + (V - U/2)^2
\]

as per page 26 with a minus sign.
Exercise page 32

Solution

This will be given later.
Chapter 4

Exercise page 43

Solution

\[ \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \]

\[ \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \]

\[ \mathbf{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \]

\[ |\Sigma| = \left| \sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2 \right| = \sigma_1^2 \sigma_2^2 \left| 1 - \rho^2 \right| \]

\[ \Sigma^{-1} = \begin{pmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix} / \sigma_1 \sigma_2 (1 - \rho^2) \]

\[ = \left( \begin{array}{cc} -\frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{array} \right) / (1 - \rho^2) \]

The exponent is:

\[ = -\frac{1}{2(1 - \rho^2)} (x_1 - \mu_1, x_2 - \mu_2) \left[ \begin{array}{cc} 1/\sigma_1^2 & -\rho/\sigma_1 \sigma_2 \\ -\rho/\sigma_1 \sigma_2 & 1/\sigma_2^2 \end{array} \right] \left( \begin{array}{c} x_1 - \mu_1 \\ x_2 - \mu_2 \end{array} \right) \]

\[ = -\frac{1}{2(1 - \rho^2)} (x_1 - \mu_1, x_2 - \mu_2) \left[ \begin{array}{c} (x_1 - \mu_1)/\sigma_1^2 - \rho(x_2 - \mu_2)/\sigma_1 \sigma_2 \\ -\rho(x_1 - \mu_1)/\sigma_1 \sigma_2 + (x_2 - \mu_2)/\sigma_2^2 \end{array} \right] \]

This gives the form of the exponent on page 41 of the Notes, with

\[ k = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \]
Exercise page 47

Solution

The MGF is:

$$M_X(t) = e^{t'\mu + t'\Sigma t/2}$$

giving the CGF as:

$$K_X(t) \overset{\text{def}}{=} \log M_X(t) = t'\mu + t'\Sigma t/2$$

The exponent in the MGF ( = CGF) becomes for the Bivariate Normal:

$$(t_1, t_2) \left( \begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right) + (t_1, t_2) \left( \begin{array}{cc} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array} \right) \left( \begin{array}{c} t_1 \\ t_2 \end{array} \right)$$

$$= \mu_1 t_1 + \mu_2 t_2 + (t_1^2 \sigma_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)/2 = \mathcal{E}$$

MGF:

$$\frac{dM(t)}{dt_1} = e^{\mathcal{E}} \left( \mu_1 + \sigma_1^2 t_1 + \rho\sigma_1\sigma_2 t_2 \right) = e^{\mathcal{E}} (\mu_1 + \ldots)$$

So

$$\left. \frac{dM(t)}{dt_1} \right|_{t=0} = \mu_1$$

and so

$$\left. \frac{dM(t)}{dt_2} \right|_{t=0} = \mu_2$$

$$\left. \frac{d^2M(t)}{dt_1^2} \right|_{t=0} = e^{\mathcal{E}} (\mu_1 + \ldots)(\mu_1 + \ldots) + e^{\mathcal{E}} (\sigma_1^2)$$

$$\left. \frac{d^2M(t)}{dt_1^2} \right|_{t=0} = \mu_1 + \sigma_1^2 = E(X_1^2)$$

and

$$\left. \frac{d^2M(t)}{dt_2^2} \right|_{t=0} = \mu_2 + \sigma_2^2$$

These give $V(X_1) = E(X_1^2) - (EX_1)^2 = \sigma_1^2$ and $V(X_2) = \sigma_2^2$

Now the fifth moment is

$$\left. \frac{d^2M(t)}{dt_2 dt_1} \right|_{t=0} = e^{\mathcal{E}} (\mu_1 + \ldots)(\mu_2 + \sigma_2^2 t_2 + \rho\sigma_1\sigma_2 t_1) + e^{\mathcal{E}} (\rho\sigma_1\sigma_2)$$

$$\left. \frac{d^2M(t)}{dt_2 dt_1} \right|_{t=0} = \mu_1 \mu_2 + \rho\sigma_1\sigma_2 = E(X_1 X_2)$$

21
So $\text{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \rho \sigma_1 \sigma_2$

The CGF yields directly:

$$\frac{dK(t)}{dt_1} = \mu_1 + \sigma_1^2 t_1 + \rho \sigma_1 \sigma_2 t_2$$

$$\frac{dK(t)}{dt_1}_{t=0} = \mu_1$$

$$\frac{d^2K(t)}{dt_1^2}_{t=0} = \sigma_1^2$$

with

$$\frac{d^2K(t)}{dt_2dt_1}_{t=0} = \rho \sigma_1 \sigma_2$$

and the results for $\mu_2$ and $\sigma_2^2$ are similar to $\mu_1$ and $\sigma_1^2$. So the CGF is easier to deal with than the MGF. Thus there are only five cumulants, which define the Bivariate Normal uniquely.


**Chapters 5 and 6**

**Exercises**

1. For a random sample of size \( n \) from a \( U(0, 1) \) distribution, show that the distribution of the \( r \) th order statistic is a Beta distribution with parameters \( r \) and \( n - r + 1 \).

2. For the situation described in 1., give the distributions of the smallest and largest order statistics.

3. Interpret graphically the distributions obtained in 2., using asymptotic arguments.

4. For example 3 given on page 81 of the Notes, calculate the expected value of the estimator \( kY_n \), and examine the asymptotic case.

5. Show that the square of a non-central \( t \) is a non-central \( F \).

The results from 4. are now demonstrated via a practical example.

**Example**


During WWII, a statistical procedure was developed for estimating German war production. It was based on *serial numbers* (order statistics).

Every piece of German equipment was stamped with a serial number that indicated the order of manufacture. So if the total number of tanks type T produced by a given date was \( N \), each tank of that type would bear the numbers 1 to \( N \). As the war progressed, some of these number became known to the Allies, either by direct capture of a tank or from records seized when a command post was overrun.

The problem was to estimate \( N \) using only the sample of \( n \) "captured" serial numbers,

\[
1 \leq Y_1 < Y_2 < \ldots < Y_n \leq N.
\]

As a model, it was assumed that the \( Y_i \)'s were one of the \( \binom{N}{n} \) possible ordered sets of \( n \) integers from 1 to \( N \) and that each set was equally probable.
That is,
\[ P(Y_1 = y_1 < Y_2 < \ldots < Y_n = y_n) = \left( \frac{N}{n} \right)^{-1} \]

The procedure originally proposed for estimating \( N \) was to add to the largest order statistic the average ”gap” in the \( Y_i \)’s. The estimator was \( W_1 \) where
\[ W_1 = Y_n + \frac{1}{n-1} \sum_{i>j} (Y_i - Y_j - 1) \]
which is just
\[ W_1 = Y_n + \frac{Y_n - Y_1}{n-1} - 1 \]
due to cancellations. Thus if 5 tanks gave the numbers 14, 28, 92, 146 and 298, the estimate of the total production would be 386 since
\[ w_1 = 298 + \frac{298 - 14}{4} - 1 = 368 \]
in contrast to the MLE (maximum likelihood estimator) of 298!

The estimator \( W_1 \) is unbiased, with variance
\[ V(W_1) = \frac{n(N-n)(N+1)}{(n-1)(n+1)(n+2)} \]

Another unbiased estimator (similar to \( kY_n \) from 4.) is
\[ W_2 = \left( \frac{n+1}{n} \right) Y_n - 1 \]
with variance
\[ V(W_2) = \frac{(N+1)(N-n)}{n(n+2)} \]
This estimator is better than \( W_1 \) since
\[ \frac{V(W_2)}{V(W_1)} = 1 - \frac{1}{n^2} \]

When the war was over and the official records of the Speer Ministry were impounded, it was found that estimates derived from procedures similar to the one just described were far more accurate than those based on other sources of information.

For example, the serial number estimate for German tank production in 1942 was 3400, which was very close to the actual figure. The ”official” Allied
estimate, based on intelligence and espionage was 18,000!

Note that an equivalent problem is the estimation of the total number of bus routes by a visitor to a unfamiliar city, where the bewildered traveller has only the numbers from the handful of buses that have not stopped for him!

Solutions

1. Now

\[ f_{Y_r}(y_r) = \frac{n!}{(n - r)!(r - 1)!} \left[ F(y_r) \right]^{r-1} \left[ 1 - F(y_r) \right]^{n-r} f(y_r) \]

and since \( f_{y_r} = 1 \) and \( F_{y_r} = y_r \) then

\[ f_{Y_r}(y_r) = \frac{n!}{(n - r)!(r - 1)!} \left[ y_r \right]^{r-1} \left[ 1 - y_r \right]^{n-r} \]

which is

\[ f_{Y_r}(y_r) = \frac{\Gamma(n + 1)}{\Gamma(n - r + 1)\Gamma(r)} \left[ y_r \right]^{r-1} \left[ 1 - y_r \right]^{n-r}, \quad 0 < y_r < 1, \]

the density of a Beta distribution with parameters \( r \) and \( n - r + 1 \).

2. The distribution of the smallest observation is

\[ f_{Y_1}(y_1) = n[1 - y_1]^{n-1}, \quad 0 < y_1 < 1 \]

and the largest gives

\[ f_{Y_n}(y_n) = ny_n^{n-1}, \quad 0 < y_n < 1 \]

3. The distribution of the smallest observation is skewed to the right with the mode at zero. The largest distribution is skewed to the left with the mode at 1. The distributions are shown for \( n = 10 \).
Smallest

\[ 10 \times (1 - y)^9 \]

Largest

\[ 10 \times y^9 \]
As the sample size becomes large, the distribution of the smallest collapses onto zero, while the distribution of the largest collapses onto 1, as expected.

4. From page 85 of the Notes,
\[
E(kY_n) = kE(Y_n) = \frac{(n + 2)}{(n + 1)} \frac{\theta n}{(n + 1)} = \frac{\theta (n + 2)n}{(n + 1)^2} = \theta \frac{(1 + 2/n)}{(1 + 1/n)^2}
\]
and so as \( n \to \infty \) then
\[
E(kY_n) \to \theta
\]
and so the estimator is unbiased, asymptotically.

5. What we are asked to show is that
\[
t^2_n(\lambda) = F_{1,n}(\lambda)
\]
\((\lambda = 0)\) This is certainly true, from elementary statistics, via \( \mu = 0 \).
\((\lambda \neq 0)\)
\[
T'(\lambda) = \frac{X}{\sqrt{W/n}}
\]
and so
\[
T'^2(\lambda) = \frac{X^2}{W/n} = \frac{W_1(\lambda)/n_1}{W_2(0)/n_2} = \frac{\chi^2(\lambda)/1}{\chi^2(0)/n}
\]
which is distributed as \( F_{1,n}(\lambda) \).