Australian National University

Venue: ______________________________

Student Number: |___|___|___|___|___|___|___|___|

Research School of Finance, Actuarial Studies and Statistics Examinatation

Semester 1 - Mid-semester Exam 2017

STAT3013/STAT4027/STAT8027 Statistical Inference

Writing Time: 180 minutes
Reading Time: 15 minutes minutes

Exam Conditions:
Central Examination
Students must return the examination paper at the end of the examination
This examination paper is not available to the ANU Library archives

Materials Permitted In The Exam Venue:
(No electronic aids are permitted e.g. laptops, phones)
Two sheets of A4 paper with notes on both sides
Calculator (any-programmable or not)
Unannotated paper-based dictionary (no approval required)

Materials to Be Supplied To Students:
Scribble Paper

Marks

Question 1	Question 2	Question 3	Question 4	Total

INSTRUCTIONS:

1.) This exam paper comprises a total of 22 pages. Please ensure your paper has the correct number of pages.

2.) The exam includes a total of 4 questions.

3.) After each question there are four blank pages to write your solutions. You may use both sides of each page to write your solutions.

4.) Each question appears on the following pages [marks are indicated]:

- Question 1 is on page 3 [**25 marks**].
- Question 2 is on page 8 [**25 marks**].
- Question 3 is on page 13 [**25 marks**].
- Question 4 is on page 18 [**25 marks**].

5.) Include all workings for each question, as marks will not be awarded for answers that do not include workings.

6.) Draw a $\boxed{\text{box}}$ around each final answer.

7.) Ensure you include your student number on this exam book.

8.) A table of probability distributions is provided with the exam.

Total Marks = 100

This exam is a redeemable exam. It will be worth either 20% or 0% of your final grade based on your final exam mark.

Question 1 [**25 marks**]: Let $X \sim \text{normal}(\mu, \text{variance} = \sigma^2)$ and $Y \sim \text{normal}(\gamma, \text{variance} = \sigma^2)$. Suppose X and Y are independent.

a. [**10 marks**] Fully derive the **moment generating function** for X. Do not just state the end result from the table.

b. [**5 marks**] Let $U = X + Y$. Fully derive the distribution of U. Make sure to specify its mean and variance.

c. [**5 marks**] Let $V = X - Y$. Derive the distribution of V. Make sure to specify its mean and variance.

d. [**5 marks**] Show that U and V are independent. Make sure to clearly outline any assumptions you make.

(a.) Let's derive the moment generating function. As the result is on the table, all steps must clearly be shown.

$$
\begin{aligned}
M_X(t) = E\left[exp(xt)\right] &= \int_{-\infty}^{\infty} exp(xt) f_x(x) dx \\
&= \int_{-\infty}^{\infty} exp(xt)(2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right) dx \\
&= \int_{-\infty}^{\infty} exp(xt)(2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(x^2-2x\mu+\mu^2\right)\right) dx \\
&= \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} exp\left(xt-\frac{1}{2\sigma^2}\left(x^2-2x\mu+\mu^2\right)\right) dx \\
&= \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(x^2-2x\mu-2\sigma^2 xt+\mu^2\right)\right) dx \\
&= \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(x^2-2x(\mu+\sigma^2 t)+\mu^2\right)\right) dx
\end{aligned}
$$

Now let $a = \mu + \sigma^2 t$ and $b = \mu^2$. So we have:

$$
\begin{aligned}
x^2 - 2x(\mu+\sigma^2 t) + \mu^2 &= x^2 - 2xa + b \\
&= (x-a)^2 - a^2 + b
\end{aligned}
$$

$$
\begin{aligned}
E\left[exp(xt)\right] &= \int_{\infty}^{\infty} (2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left((x-a)^2-a^2+b\right)\right) dx \\
&= exp\left(-\frac{1}{2\sigma^2}\left(-a^2+b\right)\right) \underbrace{\int_{\infty}^{\infty} (2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(x-a\right)^2\right) dx}_{=1} \\
&= exp\left(-\frac{1}{2\sigma^2}\left(-(\mu+\sigma^2 t)^2+\mu^2\right)\right) \\
&= exp\left(-\frac{1}{2\sigma^2}\left(-\mu^2-2\mu\sigma^2 t-\sigma^4 t^2+\mu^2\right)\right) \\
&= exp\left(\mu^2/(2\sigma^2)+2\mu\sigma^2 t/(2\sigma^2)+\sigma^4 t^2/(2\sigma^2)-\mu^2/(2\sigma^2)\right) \\
&= exp\left(\mu^2/(2\sigma^2)+\mu t+\sigma^2 t^2/2-\mu^2/(2\sigma^2)\right) \\
&= exp\left(\mu t+\sigma^2 t^2/2\right)
\end{aligned}
$$

So we have the result.

(b.)-(d.) We can solve all 3 questions through one go. First we have:

$$U = X + Y; \quad V = X - Y \quad \Rightarrow \quad X = \frac{U+V}{2}; \quad Y = \frac{U-V}{2}$$

We will need the Jacobian to determine transformed joint distribution:

$$J = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

The determinant of the matrix $J = (1/2)(-1/2) - (1/2)(1/2) = -1/2$. This leads to the following joint distribution for U, V:

$$\begin{aligned}
f_{U,V}(u,v) &= f_{X,Y}\left(x = \frac{u+v}{2}, y = \frac{u-v}{2}\right)\Big| - \frac{1}{2}\Big| \\
&= (2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(\frac{u+v}{2} - \mu\right)^2\right)(2\pi\sigma^2)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(\frac{u-v}{2} - \gamma\right)^2\right)\Big| - \frac{1}{2}\Big| \\
&= \left((2\pi\left(2\sigma^2\right)\right)^{-1/2}\left(2\pi\left(2\sigma^2\right)\right)^{-1/2} exp\left(-\frac{1}{2\sigma^2}\left(\frac{u^2}{2} - u(\mu+\gamma) + \frac{v^2}{2} - v(\mu-\gamma) + \mu^2 + \gamma^2\right)\right) \\
&= \left((2\pi\left(2\sigma^2\right)\right)^{-1/2}\left(2\pi\left(2\sigma^2\right)\right)^{-1/2} exp\left(-\frac{1}{2\left(2\sigma^2\right)}\left(u^2 - 2u(\mu+\gamma) + v^2 - 2v(\mu-\gamma) + 2\mu^2 + 2\gamma^2\right)\right) \\
&= \left((2\pi\left(2\sigma^2\right)\right)^{-1/2}\left(2\pi\left(2\sigma^2\right)\right)^{-1/2} exp\left(-\frac{1}{2\left(2\sigma^2\right)}\left(u^2 - 2u(\mu+\gamma) + v^2 - 2v(\mu-\gamma) + (\mu+\gamma)^2 + (\mu-\gamma)^2\right)\right) \\
&= \underbrace{\left((2\pi\left(2\sigma^2\right)\right)^{-1/2} exp\left(-\frac{1}{2\left(2\sigma^2\right)}\left(u - (\mu+\gamma)\right)^2\right)}_{f_U(u)}\underbrace{\left((2\pi\left(2\sigma^2\right)\right)^{-1/2} exp\left(-\frac{1}{2\left(2\sigma^2\right)}\left(v - (\mu-\gamma)\right)^2\right)}_{f_V(v)}
\end{aligned}$$

There we see $U \sim \text{normal}(\mu+\gamma, 2\sigma^2)$; $V \sim \text{normal}(\mu-\gamma, 2\sigma^2)$. Also as $f_{U,V}(u,v) = f_U(u) \times f_V(v)$ we can state that U and V are independent.

Question 2 **[25 marks]**: Write pseudo-code to clearly outline **three** different methods **we have discussed in class** [label them algorithm (a), (b), and (c)] to obtain **at least** $S = 1,000$ samples from the distribution below. Assume **only** that you are able to draw independent random samples from a uniform$(0, 1)$ distribution. Be sure to work out all specific details. **Rank the three approaches based on efficiency (some approaches may be equally efficient). Discuss the reasoning for your ranking.**

$$f(x) = 4x^3; \quad 0 \leq x \leq 1.$$

- We will consider three different approaches to sample from this density. (a) the inverse CDF method, (b) the accept-reject method, and (c) the Metropolis-Hastings algorithm.

a. The Inverse CDF Method

The first thing we need to do is determine the CDF:

$$\begin{aligned} F_X(c) &= \int_0^c 4x^3 dx \\ &= c^4 \end{aligned}$$

So take the cdf and set it equal to U, which is a standard uniform random variable.

$$X^4 = U \quad \Rightarrow \quad X = U^{(1/4)}$$

Algorithm 1 Generate Samples from X - Inverse CDF Method

```
    let N = 1,000 be the number of samples we wish to generate
2:  for n in 1:N do
        sample U ~ uniform(0,1)
4:      calculate X = U^(1/4)
        store that value of X
6:  return the 1,000 values of X
```

b. The Accept-Reject Method

To make our life easy let $V \sim \text{uniform}(0,1)$. Now, let's figure out M:

$$M = sup_x \frac{f_X(x)}{f_V(x)} = max_x \frac{4x^4}{1} = 4 \quad \text{(i.e. this is maximized when } x = 1)$$

Algorithm 2 Generate Samples from X - Accept-Reject Method

```
    let c = 0
2:  while c < 1000 do
        sample V ~ uniform(0,1) and U ~ uniform(0,1)
4:      if u < (1/4)4v^3 then set x = v; store x; set c = c + 1
        if u ≥ (1/4)4v^3 then return to Step 3
6:  return the 1,000 values of X
```

We know that the $P(\text{Accept}) = 1/M = 1/4$, therefore we expect we will need roughly 4,000 runs of the algorithm.

c. The Metropolis-Hastings Algorithm

We will consider a symmetric proposal distribution. Let a proposed value x be $x^* \sim \text{uniform}(0,1)$. We will run the algorithm for $2,000$ scans, removing the first $1,000$ for burn-in.

Algorithm 3 Generate Samples from X - Metropolis-Hastings Algorithm

let $N = 1,000$ be the number of samples we wish to generate
let the starting value for $x_{(1)}$ be equal to c
for n in 2:(N+1000) **do**
    sample $x^* \sim$ uniform$(0, 1)$
    calculate the Metropolis-Hasting Ratio: $MR = \frac{f_X(x^*)}{f_X(x)} = \frac{4x^{*3}}{4x_{(n)}^3} = \frac{x^{*3}}{x_{(n)}^3}$
    calculate $\rho = min(MR, 1)$
    sample $U \sim$ uniform$(0, 1)$
    if $u \leq \rho$ **then** set the new value of $x_{(n+1)}$ equal to x^*; store $x_{(n+1)}$
    if $u > \rho$ **then** set the new value of $x_{(n+1)}$ equal to the previous value of $x_{(n)}$; store $x_{(n+1)}$
return 1,000 values of X after removing the first 1,000 stored values for burn-in

- The ranking of the methods is a > b > c. The reasoning is that (a) generates samples directly and those samples are independent. The samples for (b) are independent but they are not direct. So we will have to reject (1 - 1/M) candidates. Finally, for (c) the samples are not direct and they are not independent. The Markov chain will eventually converge to the target distribution, but this may take time.

Question 3 [**25 marks**]: Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x|\theta) = \left(\frac{x^3}{\theta^4 \; 3!}\right) \exp\left(-x/\theta\right);\; x > 0$. For the questions below, if a closed form analytical solution doesn't exist, clearly outline a computational solution via pseudo-code.

a. [**6 marks**] Derive the **Method of Moments** estimator for θ.

   i) Is it unbiased?
   ii) What is its variance?
   iii) If the estimator is biased can you determine an unbiased estimator based on it?

b. [**6 marks**] Based on the Method of Moments estimator ($\tilde{\theta}$) and the **Central Limit Theorem** what is an approximate distribution for $\tilde{\theta}$?

c. [**6 marks**] Derive the **Maximum Likelihood** estimator for θ.

   i) Is it unbiased?
   ii) What is its variance?
   iii) What is its mean squared error?
   iv) If the estimator is biased can you determine an unbiased estimator based on it?

d. [**7 marks**] Derive the **Maximum Likelihood** estimator for θ^2.

   i) Is it unbiased?
   ii) If the estimator is biased can you determine an unbiased estimator based on it?

(a.) The first thing to notice is that $X \sim \text{gamma}(a = 4, b = \theta)$. Note:

$$f_X(x) = \frac{1}{\theta^4 \; 3!} x^{4-1} exp(-x/\theta) = \frac{1}{\theta^4 \; \Gamma(4)} x^{4-1} exp(-x/\theta)$$

Now we just need to set the expected value of the distribution (the first moment) equal to the sample mean (sample first moment).

$$\begin{aligned} E[X] &= \bar{X} \\ 4\theta &= \bar{X} \\ \tilde{\theta} &= \bar{X}/4 \end{aligned}$$

Now let's get the expected value and variance of $\tilde{\theta}$.

$$\begin{aligned} E[\tilde{\theta}] &= E\left[\bar{X}/4\right] \\ &= \frac{1}{4} E\left[\bar{X}\right] \\ &= \frac{1}{4n} E\left[\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{4n} n E\left[X_i\right] \\ &= \frac{1}{4} E\left[X_i\right] \\ &= \frac{1}{4} 4\theta = \theta \end{aligned}$$

Therefore the Method of Moments estimator $\tilde{\theta}$ is an unbiased estimator of θ. Now let's get the variance of the estimator.

$$
\begin{aligned}
V[\tilde{\theta}] &= V\left[\bar{X}/4\right] \\
&= \frac{1}{4^2} V\left[\bar{X}\right] \\
&= \frac{1}{4^2 n^2} V\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{4^2 n^2} n V\left[X_i\right] \\
&= \frac{1}{4^2 n} V\left[X_i\right] \\
&= \frac{1}{4^2 n} 4\theta^2 \\
&= \frac{1}{4n} \theta^2
\end{aligned}
$$

b.) We notice that the MoM estimator is made up of a sum, so we can rely on the central limit theorem.

$$\tilde{\theta} \mathrel{\dot{\sim}} \text{normal}\left(\theta, \frac{\theta^2}{4n}\right)$$

(c.) Let's write out the likelihood:

$$
\begin{aligned}
L(\theta|\boldsymbol{x}) &= \prod_{i=1}^{n} \frac{1}{\theta^4\ \Gamma(4)} x_i^{4-1} exp(-x_i/\theta) \\
&= \left(\frac{1}{\theta^4\ \Gamma(4)}\right)^n \left[\prod_{i=1}^{n} x_i^{4-1}\right] exp(-\sum_{i=1}^{n} x_i/\theta)
\end{aligned}
$$

From here we can get the log-likelihood:

$$\ell(\theta|\boldsymbol{x}) = n\left(log(1) - 4log(\theta) - log(\Gamma(4)\right] + (4-1)\sum_{i=1}^{n} log(x_i) - \sum_{i=1}^{n} x_i/\theta$$

Now let's differentiate this, set it equal to zero, and solve for θ:

$$
\begin{aligned}
\frac{d\ \ell(\theta|\boldsymbol{x})}{d\theta} &= -\frac{4n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} \\
&\Rightarrow -\frac{4n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0 \\
& \frac{4n}{\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta^2} \\
& \hat{\theta} = \frac{\bar{X}}{4}
\end{aligned}
$$

Based on the previous results, we have $E[\hat{\theta}] = \theta$ and $V[\hat{\theta}] = \frac{\theta^2}{4n}$. So the MLE is an unbiased estimator of θ.

(d.) Based on the invariance property of MLEs we have:

$$\widehat{\theta^2} = \hat{\theta}^2 = \left(\frac{\bar{X}}{4}\right)^2 = \frac{1}{16}\left(\bar{X}\right)^2$$

Now let's find the expected value of the estimator:

$$\begin{aligned} E[\hat{\theta}^2] &= E\left[\frac{1}{16}\left(\bar{X}\right)^2\right] \\ &= \frac{1}{16}E\left[\left(\bar{X}\right)^2\right] \end{aligned}$$

Now consider:

$$\begin{aligned} E\left[\left(\bar{X}\right)^2\right] &= V(\bar{X}) + \left(E[\bar{X}]\right)^2 \\ &= \frac{4\theta^2}{n} + (4\theta)^2 \\ &= \left(\frac{4}{n} + 4^2\right)\theta^2 \end{aligned}$$

So we have:

$$E[\hat{\theta}^2] = \frac{1}{16}\left(\frac{4}{n} + 4^2\right)\theta^2 = \left(\frac{1}{4n} + 1\right)\theta^2$$

We can see the estimator is biased for θ^2. We can create a new estimator, based on this estimator which won't be biased:

$$\hat{\gamma} = \left(\frac{1}{4n} + 1\right)^{-1}\hat{\theta}^2$$

Question 4 [**25 marks**]: In families where one parent has a rare hereditary disease, the probability that that a particular child inherits the disease is p, where $0 < p < 1$. In a survey, only families of size k, **with at least one child with an inherited disease** were independently sampled. For the study, n such families were observed independently and there are r_i children with the disease in the i^{th} family $(i = 1, 2, \ldots, n)$. Determine the **Maximum Likelihood** estimator for p. if a closed form analytical solution doesn't exist, clearly outline a computational solution via pseudo-code.

- Let's first determine that in a family of size k, the probability there is **at least one** child with a hereditary disease. Note that this probability may be modelled as a binomial distibrution:

$$\begin{aligned} P(\text{at least child}) &= 1 - P(\text{no children}) \\ &= 1 - \binom{k}{0} p^0 (1-p)^{k-0} \\ &= 1 - (1-p)^k \end{aligned}$$

- Now, we have data on families of size k where at least on child has the disease. For each family the number of children which has a hereditary disease is r_i. This suggests that we have a conditional binomial distribution. For a single family we have:

$$P(X = r_i \big| r_i > 0) = \frac{\binom{k}{r_i} p^{r_i} (1-p)^{k-r_i}}{1-(1-p)^k}$$

- Based on n such families which were independently sampled, we have the the following likelihood for p:

$$\begin{aligned} L(p \big| r_1, \ldots, r_n) &= \prod_{i=1}^{n} \frac{\binom{k}{r_i} p^{r_i} (1-p)^{k-r_i}}{1-(1-p)^k} \\ &= \prod_{i=1}^{n} \binom{k}{r_i} p^{r_i} (1-p)^{k-r_i} \left[1-(1-p)^k\right]^{-1} \\ &= \left[1-(1-p)^k\right]^{-n} \left[\prod_{i=1}^{n} \binom{k}{r_i}\right] p^{\sum r_i} (1-p)^{nk - \sum r_i} \end{aligned}$$

- Let's get the log likelihood:

$$\ell(p \big| r_1, \ldots, r_n) = -nlog\left[1-(1-p)^k\right] + \sum \log \binom{k}{r_i} + \sum r_i log(p) + (nk - \sum r_i) log(1-p)$$

- Let's differentiate this with respect to p:

$$
\begin{aligned}
\frac{d\ell(p \big| r_1, \ldots, r_n)}{dp} &= -\frac{nk(1-p)^{k-1}}{1-(1-p)^k} + \frac{\sum r_i}{p} - \frac{nk - \sum r_i}{(1-p)} \\
&= \frac{-knp + \sum r_i \left[1-(1-p)^k\right]}{p(1-p)[1-(1-p)^k]} \\
&\Rightarrow \frac{-knp + \sum r_i \left[1-(1-p)^k\right]}{p(1-p)[1-(1-p)^k]} = 0 \\
&-knp + \sum r_i \left[1-(1-p)^k\right] = 0 \\
&\sum r_i \left[1-(1-p)^k\right] = knp
\end{aligned}
$$

- We see the we are unable to get a closed form solution for the MLE. Let's use the Newton-Raphson algorithm.

$$
U = -\frac{nk(1-p)^{k-1}}{1-(1-p)^k} + \frac{\sum r_i}{p} - \frac{nk - \sum r_i}{(1-p)}
$$

Now we determine H , the second derivative:

$$
H = \frac{d^2\ell(\cdot)}{dp^2} = \frac{kn\left[(1-p)^k + k - 1\right](1-p)^{k-2}}{\left[1-(1-p)^k\right]^2} - \frac{\sum r_i}{p^2} - \frac{nk - \sum r_i}{(1-p)^2}
$$

Algorithm 4 Newton-Raphson

```
let check = 10
let p_1 = 0.5
let c = 2
while check < 0.00001 do
    p_c = p_(c-1) - H^{-1}(p_(c-1)) U(p_(c-1))
    calculate check = |p_c - p_(c-1)|
    let c = c + 1
return the last value of p
```

End Of Examination