Group: Zining Fan, Mutian Wang, Siyuan Wang
UNI: zf2234, mw3386, sw3418

## Graded Homework 1 - Exercise 2

1. Based on the moments generating function, $E[\bar{X}^2] = \phi''(0)$. Since $X_i \sim Poisson(\lambda)$, $\mu = \lambda, \sigma^2 = \lambda$. Using CLT, we could get that

$$\bar{X} \sim N(\lambda, \frac{\lambda}{n})$$

So, its moments generating function is

$$\phi(t) = e^{t\lambda + \frac{\lambda}{2n}t^2}$$

$$\phi''(t) = \frac{\lambda}{n}e^{t\lambda + \frac{\lambda}{2n}t^2} + (\lambda + \frac{\lambda t}{n})^2 e^{t\lambda + \frac{\lambda}{2n}t^2}$$

So,

$$E[\bar{X}^2] = \phi''(0) = \lambda^2 + \frac{\lambda}{n}$$

2. $E[s^2] = \frac{1}{n-1}\sum_{i=1}^{n} E[{X_1}^2] - \frac{n}{n-1}E[\bar{X}^2]$

Using the moments generating function, we could get that

$$E[{X_1}^2] = \Phi''(0) = \lambda^2 + \lambda$$

So,

$$E[s^2] = \frac{n}{n-1}(\lambda^2 + \lambda) - \frac{n}{n-1}(\lambda^2 + \frac{\lambda}{n})$$
$$= \frac{n}{n-1}\frac{n-1}{n}\lambda = \lambda$$

So, $E[s^2]$ is an unbiased estimator of $\lambda$.

3.

$$E[Y_i] = E[{X_i}^2 - 2\lambda X_i + \lambda^2] - E[X_i]$$
$$= E[{X_i}^2] - 2\lambda E[X_i] + \lambda^2 - \lambda$$
$$= \lambda^2 + \lambda - 2\lambda^2 + \lambda^2 - \lambda = 0$$

So, $E[Y_i] = 0$.

$$Var[Y_i] = Var[{X_i}^2 - 2\lambda X_i + \lambda^2 - X_i]$$
$$= Var[{X_i}^2 - 2\lambda X_i - X_i]$$
$$= Var[{X_i}^2] + (2\lambda + 1)^2 Var[X_i] - 2(2\lambda + 1)Cov({X_i}^2, X_i)$$

Among this,

$$Var[{X_i}^2] = E[{X_i}^4] - E[{X_i}^2]^2$$
$$Cov({X_i}^2, X_i) = E[{X_i}^3] - E[{X_i}^2]E[X_i]$$

Using the moments generating function, we could get that

$$E[{X_i}^2] = \phi''(0) = \lambda^2 + \lambda$$

$$E[{X_i}^3] = \phi'''(0) = \lambda^3 + 3\lambda^2 + \lambda$$

$$E[{X_i}^4] = \phi''''(0) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$$

Plug these in we could get that

$$Var[Y_i] = 2\lambda^2$$

4. $s^2 - \bar{X} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 - \bar{X} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \lambda + \lambda - \bar{X})^2 - \bar{X}$

$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \lambda)^2 + \frac{2}{n-1}\sum_{i=1}^{n}(\lambda X_i - \lambda^2 - \bar{X}X_i + \lambda\bar{X}) + \frac{1}{n-1}\sum_{i=1}^{n}(\lambda - \bar{X})^2 - \bar{X}$

As $Y_i = (X_i - \lambda)^2 - X_i$, $Y_i + X_i = (X_i - \lambda)^2$

$s^2 - \bar{X} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i + X_i) + \frac{2n\lambda\bar{X}}{n-1} - \frac{2n\lambda^2}{n-1} - \frac{2n\bar{X}^2}{n-1} + \frac{2\lambda n\bar{X}}{n-1} + \frac{n\lambda^2}{n-1} - \frac{2n\lambda\bar{X}}{n-1} + \frac{n\bar{X}^2}{n-1} - \frac{n\bar{X}-\bar{X}}{n-1}$

$= \frac{n\bar{Y}}{n-1} + \frac{n\bar{X}}{n-1} - \frac{n\lambda^2}{n-1} - \frac{n\bar{X}^2}{n-1} + \frac{2\lambda n\bar{X}}{n-1} - \frac{n\bar{X}-\bar{X}}{n-1}$

$= \frac{n\bar{Y}}{n-1} + \frac{\bar{X}}{n-1} - \frac{n\bar{X}^2}{n-1} + \frac{2\lambda n\bar{X}}{n-1} - \frac{n\lambda^2}{n-1}$

$= \frac{n}{n-1}\bar{Y} + \frac{1}{n-1}\bar{X} - \frac{n}{n-1}(\bar{X} - \lambda)^2$

5.

$$\sqrt{n}(s^2 - \bar{X}) = \sqrt{n}(\frac{n}{n-1}\bar{Y} + \frac{1}{n-1}\bar{X} - \frac{n}{n-1}(\bar{X} - \lambda)^2)$$

Since $X_i \sim Poisson(\lambda)$,

$$\bar{X} \xrightarrow[\infty]{D} N(\lambda, \frac{\lambda}{n})$$

From $\bar{X} \xrightarrow[\infty]{P} \lambda$ know that,

$$\frac{\sqrt{n}}{n-1}\bar{X} \xrightarrow[\infty]{P} 0$$

Also from $\bar{X} \xrightarrow[\infty]{P} \lambda$,we have

$$(\bar{X} - \lambda)^2 \xrightarrow[\infty]{P} 0$$

$$\frac{n^{\frac{3}{2}}}{n-1}(\bar{X} - \lambda)^2 \xrightarrow[\infty]{P} 0$$

$$\sqrt{n}(\frac{1}{n-1}\bar{X} - \frac{n}{n-1}(\bar{X} - \lambda)^2) \xrightarrow[\infty]{P} 0$$

As $E(Y_i) = 0 \; and \; Var(Y_i) = 2\lambda^2$, get that:

$$\bar{Y} \xrightarrow[\infty]{D} N(0, \frac{2\lambda^2}{n})$$

$$\sqrt{n}\bar{Y} \xrightarrow[\infty]{D} N(0, 2\lambda^2)$$

$\frac{n}{n-1} \xrightarrow[\infty]{P} 1$ and according to Slutsky's Theorem,

$$\frac{n}{n-1}\sqrt{n}\bar{Y} \xrightarrow[\infty]{D} N(0, 2\lambda^2)$$

$$\frac{n}{n-1}\sqrt{n}\bar{Y} + \sqrt{n}(\frac{1}{n-1}\bar{X} - \frac{n}{n-1}(\bar{X}-\lambda)^2) \xrightarrow[\infty]{D} N(0, 2\lambda^2)$$

The asymptotic distribution of $\sqrt{n}(s^2 - \bar{X})$ is $N(0, 2\lambda^2)$.
Similarly, as $\bar{X} \xrightarrow[\infty]{P} \lambda$, so:

$$\frac{1}{\sqrt{2\bar{X}}} \xrightarrow[\infty]{P} \frac{1}{\sqrt{2\lambda}}$$

.

According to Slustsky's Theorem,

$$\sqrt{\frac{n}{2}}\frac{s^2 - \bar{X}}{\bar{X}} \xrightarrow[\infty]{D} \frac{1}{\sqrt{2\lambda}} N(0, 2\lambda^2)$$

which is N(0,1).
So the asymptotic distribution of $\sqrt{\frac{n}{2}}\frac{s^2-\bar{X}}{\bar{X}}$ is $N(0,1)$.

6. The null hypotheses $H_0 : E[\bar{X}] = E[s^2]$ can be written as:

$$H_0 : E[s^2 - \bar{X}] = 0$$

We want to test whether overdispersion exists, so the alternative hypotheses is :

$$H_1 : E[s^2 - \bar{X}] > 0$$

From question 4, we know that , under the $H_0$ assumption and $X_i \sim Poisson(\lambda)$, we have that:

$$\sqrt{\frac{n}{2}}\frac{s^2 - \bar{X}}{\bar{X}} \sim N(0,1)$$

Let test statistic TS = $\sqrt{\frac{n}{2}}\frac{s^2-\bar{X}}{\bar{X}}$, the critical value should be $Z_{1-\alpha}$.
When TS $> Z_{1-\alpha}$, we reject the null hypothesis and conclude that there is overdispersion in the data. Otherwise we fail to reject $H_0$.

7. We set $\alpha = 0.05$.

   (a) The code to generate the 500 poission random samples with $\lambda = 5$ are shown below,

```
x=np.zeros(501)
error=0
for i in range(1,501):
    x=np.random.poisson(5,50)
    mean=np.mean(x)
    var=np.var(x)
    TS=np.sqrt(25)*(var-mean)/mean
    if TS>1.65:
        error=error+1
error=error/500
```

After running the code, we get the probability of rejecting $H_0$ when $H_0$ is true:

```
error
0.048
```

The value of error is the probability of rejecting $H_0$ when $H_0$ is true, we could find that it's close to 0.05, which equals to $\alpha$. So, the null hypothesis testing we proposed in the previous question is reasonable.

(b) The code to generate poission random variables with $\lambda$ having gamma distribution is shown below,

```
x=np.zeros(501)
error=0
for i in range(1,501):
    lamda=np.random.gamma(2.5)
    x=np.random.poisson(lamda,50)
    mean=np.mean(x)
    var=np.var(x)
    TS=np.sqrt(25)*(var-mean)/mean
    if TS>1.65:
        error=error+1
error=error/500
```

The probability is :

```
error
0.04
```

The value of error is the probability of rejecting $H_0$ when $H_0$ is true, we could find that it's close to 0.05, which equals to $\alpha$. So, the null hypothesis testing we proposed in the previous question is reasonable.

8. The code to compute the TS of the given data is shown below,

```
death=[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14]
freq=[1,4,15,31,39,55,54,49,47,31,16,9,8,4,3]
n=sum(freq)
mean=np.average(death, weights=freq)
var=np.sum((death-mean)*(death-mean)*freq)/(np.sum(freq)-1)
TS=np.sqrt(np.sum(freq)/2)*(var-mean)/mean
```

The observed value of TS is

```
TS
0.9786745788901388
```

Since $0.97 < 1.65$, we fail to reject $H_0$.