QUESTION 2

Wen Fan(wf2255), Banruo Xie(bx2168), Hanjun Li(hl3339)

2.1

According to the definition of variance:

$Var(\bar{X}) = E((\bar{X})^2) - E(\bar{X})^2$

$Var(\bar{X}) = Var(\frac{1}{n}(X_1 + X_2 + ... + X_n)) = \frac{1}{n^2}Var((X_1 + X_2 + ... + X_n))$

$= \frac{1}{n^2}(Var(X_1) + Var(X_2) + ... + Var(X_n)) = \frac{Var(X)}{n}$

$E(\bar{X}) = E(\frac{1}{n}(X_1 + X_2 + ... + X_n)) = \frac{1}{n}(E(X_1) + E(X_2) + ... + E(X_n)) = E(X)$

Since $X \sim Poisson(\lambda)$**,** $E(X) = \lambda$**,** $Var(X) = \lambda$**,**

$Var(\bar{X}^2) = Var(\bar{X}) + (E(\bar{X}))^2 = \frac{\lambda}{n} + \lambda^2$

2.2

Since $E(X^2) = Var(X) + (E(X))^2 = \lambda + \lambda^2$**,**

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

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

$= \frac{1}{n-1}(n\lambda + n\lambda^2) - (\lambda + n\lambda^2)) = \lambda$**, Hence Proved.**

2.3

$$E(Y_i) = E((X_i - \lambda))^2 - E(X_i)$$

$$= E((X_i)^2) - 2\lambda E(X_i) + \lambda^2 - E(X_i) = \lambda + \lambda^2 - 2\lambda^2 + \lambda^2 - \lambda = 0$$

$$Var(Y_i) = E((Y_i)^2) - (E(Y_i))^2 = E((Y_i)^2)$$

$$= E(((X_i)^2 - (2\lambda + 1)X_i + \lambda^2))^2$$

$$= E((X_i)^4 - 4\lambda X_i{}^3 + 6\lambda^2 X_i{}^2 - 4\lambda^3 X_i + \lambda^4 - 2X_i{}^3 + 4\lambda X_i{}^2 - 2\lambda^2 X_i + X_i{}^2)$$

By using moment generating function of Poisson distribution:

$$E(X) = \lambda, E(X^2) = \lambda^2 + \lambda, E(X^3) = \lambda^3 + 3\lambda + \lambda$$

$$E(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$$

Plugging back to the equation, we can have: $Var(Y_i) = 2\lambda^2$

2.4

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

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

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

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

2.5

By LLN and Slutsky Theorem $\bar{X} - \lambda \xrightarrow[n->\infty]{P} 0$**, therefore,** $n(\bar{X} - \lambda)^2 \xrightarrow[n->\infty]{P} 0$ **and** $\frac{\bar{X}}{\sqrt{n}} \xrightarrow[n->\infty]{P} 0$

As a result, $s^2 - \bar{X} = \frac{1}{n-1}(\sum_{i=1}^{n} Y_i - n(\bar{X} - \lambda)^2 + \bar{X})$ **will converge to** $\frac{1}{n-1}\sum_{i=1}^{n} Y_i = \frac{n}{n-1}\bar{Y}$

as n goes to ∞**.**

And by CLT, $\sqrt{n}\bar{Y} \xrightarrow[n->\infty]{D} N(0, 2\lambda^2)$**,** $\frac{n}{n-1} \xrightarrow{n->\infty} 1$**, we would have** $\frac{n}{n-1}\sqrt{n}\bar{Y} \xrightarrow[n->\infty]{D} N(0, 2\lambda^2)$**.**

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

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

2.6

If $E(\bar{X}) = E(s^2)$**,then we calculate** $P(|\frac{\sqrt{n}}{\sqrt{2}}\frac{s^2-\bar{X}-E(s^2-\bar{X})}{\bar{X}}| \leq Z_{1-\frac{a}{2}}) = a$**.**

We compare $\frac{\sqrt{n}}{\sqrt{2}}\frac{s^2-\bar{X}}{\bar{X}}$ **with** $Z_{1-\frac{a}{2}}$**, to determine whether reject** H_0

If the former value is less than the z-score, we fail toreject, otherwise, we reject

2.7

In [40]:
```
import numpy as np
from scipy.stats import norm
```

In [41]:
```
def model_1(n):
    a = np.random.poisson(5,500*50)
    mean = np.mean(a)
    variance = np.var(a,ddof=1)
    return [mean, variance]
```

In [42]:
```
def model_2(n):
    b = []
    for i in range(n):
        gamma = np.random.gamma(2.5,2,1)
        b = b + list(np.random.poisson(gamma,50))
    mean = np.mean(np.array(b))
    variance = np.var(np.array(b),ddof=1)
    return [mean,variance]
```

In [43]:
```
[mean1,var1]=model_1(500)
[mean2,var2]=model_2(500)
```

In [44]:
```
print("For model 1, it is ", np.sqrt(250)*(var1-mean1)/mean1)
print("For model 2, it is ", np.sqrt(250)*(var2-mean2)/mean2)
norm.ppf(0.975)
```

```
For model 1, it is  0.028281667787889257
For model 2, it is  30.288903569535357
```

Out[44]:
```
1.959963984540054
```

From the result above, we fail to reject H_0 for model 1 and we reject H_0 for model 2 at 5% level

2.8

In [85]:
```
def model_3(fre):
    data = []
#    print(len(fre))
    for i in range(len(fre)):
        for j in range(fre[i]):
#            print(fre[i])
            data.append(i)
    mean = np.mean(np.array(data))
    var = np.var(np.array(data),ddof=1)
    return [mean,var,np.array(data)]
```

In [86]:
```
fre = [1,4,15,31,39,55,54,49,47,31,16,9,8,4,3]
[mean3,var3,data] = model_3(fre)
```

In [87]:
```
np.sqrt(len(data)/2)*(var3-mean3)/mean3
```

Out[87]:
```
0.9786745788901424
```

the result above is smaller than 1.96, Hence we fail to reject H_0 at 5% level

Therefore, there is no overdispersions.