Exercise2_cc4515_cy2550_ml4404

February 11, 2020

[26]:
```
import numpy as np
from scipy.misc import derivative
from scipy.stats import norm
from sympy import diff, symbols
import random
```

Exercise 2

Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404

(1)

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

$= \frac{1}{n}EX^2 + (1 - \frac{1}{n})(EX)^2$

$EX^2 = VarX + (EX)^2 = \lambda + \lambda^2$

$EX = \lambda$

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

(2)

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

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

(3)

$EY_i = EX_i^2 - 2\lambda EX_i + \lambda^2 - EX_i$

$= \lambda^2 + \lambda - 2\lambda^2 + \lambda^2 - \lambda = 0$

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

Using moment generating function, we can get

$EX = \lambda$

$EX^2 = \lambda^2 + \lambda$

$EX^3 = \lambda^3 + 3\lambda^2 + \lambda$

$EX^4 = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$

So $VarY_i = 2\lambda^2$

(4)

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

$Y_i = (X_i - \bar{X})^2 + (\bar{X} - \lambda)^2 + 2(X_i - \bar{X})(\bar{X} - \lambda) - X_i$

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

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

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

(5)

By CLT, $\sqrt{n}\bar{Y} \Rightarrow_D N(0, 2\lambda^2)$

$\frac{n}{n-1} \Rightarrow_P 1$

By LLN, $\bar{X} - \lambda \Rightarrow_P 0$

By CLT, $\sqrt{n}(\bar{X} - \lambda) \Rightarrow_D N(0, \lambda)$

Then Slutsky's Theorem tells that

$\sqrt{n}(\bar{X} - \lambda)(\bar{X} - \lambda) = \sqrt{n}(\bar{X} - \lambda)^2 \Rightarrow_D 0 * N(0, \lambda) = 0$

$\frac{\bar{X}}{\sqrt{n}} \Rightarrow_P 0$

So $(\sqrt{n}\bar{Y} - \sqrt{n}(\bar{X} - \lambda)^2 + \frac{\bar{X}}{\sqrt{n}}) \Rightarrow_D N(0, 2\lambda^2)$

Because $\frac{n}{n-1} \Rightarrow_P 1$

$\frac{n}{n-1}(\sqrt{n}\bar{Y} - \sqrt{n}(\bar{X} - \lambda)^2 + \frac{\bar{X}}{\sqrt{n}}) \Rightarrow_D N(0, 2\lambda^2)$

So $\sqrt{n}(s^2 - \bar{X}) \Rightarrow_D N(0, 2\lambda^2)$

$\bar{X} \Rightarrow_P \lambda$

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

(6)

For sample from Poisson distribution, when $E(\bar{X}) = E(s^2)$, $P(|\sqrt{\frac{n}{2}}\frac{(s^2-\bar{X}) - E(s^2-\bar{X})}{\bar{X}}| \leq Z_{1-\alpha/2}) = \alpha$

So if $|\sqrt{\frac{n}{2}}\frac{(s^2-\bar{X})}{\bar{X}}| \geq Z_{1-\alpha/2}$, we can reject H0.

(7)

[31]:
```
random.seed(1111)
model_1 = np.random.poisson(5,500*50)
m1,v1 = np.mean(model_1), np.var(model_1,ddof = 1)
```

[32]:
```
def model2_gen(n):
    a = []
    for i in range(n):
        a = a + list(np.random.poisson(np.random.gamma(2.5,2,1),50))
    return np.array(a)
```

[33]:
```
model_2 = model2_gen(500)
m2,v2 = np.mean(model_2), np.var(model_2,ddof = 1)
```

[34]:
```
print("Model A: ", np.sqrt(500/2)*(v1-m1)/m1)
print("Model B: ",np.sqrt(500/2)*(v2-m2)/m2)
```

```
Model A:  0.2692375458087828
Model B:  33.74859256105128
```

[35]:
```
norm.ppf(0.975)
```

[35]: 1.959963984540054

As we can see, model A $0.27 < 1.96$, model B $33.75 > 1.96$.

So for model A we can accept H0 at significant level 0.05.

For model B we should reject H0 at significant level 0.05.

(8)

[36]:
```
f = [1,4,15,31,39,55,54,49,47,31,16,9,8,4,3]
data = []
for i,fre in enumerate(f):
    for j in range(fre):
        data.append(i)
data = np.array(data)
```

[37]:
```
m3,v3 = np.mean(data), np.var(data,ddof = 1)
np.sqrt(len(data)/2)*(v3-m3)/m3
```

[37]: 0.9786745788901424

$0.98<1.96$, so we can accept H0 at significant level 0.05.

The test doesn't detect any overdispersions.

[60]:
```
#e = np.exp(1)
#t = symbols('x', real=True)
#l = symbols('l', real = True)
#diff(e**(l*(e**t-1)), t, 4).subs(t,0)
```