Week 2. Lab Code

* doc : https://docs.scipy.org/doc/scipy/reference/stats.html

Hw1.

Tumor counts : trying to estimate the rate of tumorigenesis in two strains of mice A,B.

```
#count data : A =10, B=13
#A : well studied, Poisson-distributed with a mean of 12
### strong prior
#B : unknown, but related to type A mice.
### weak prior
```

a) find the posterior distributions, means, variancees and 95% quantile based confidence intervals for theta A and B.

```
import scipy
import scipy.stats as st
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import nbinom
from scipy.stats import gamma

# Likelihood
def likelihood(theta, n, sy):
    return (theta**sy)*np.exp(-n*theta)

# prior parameters for thetaA ~ gamma(a1,b1)
# A : poisson distributed and mean is 12, E(thetaA) = 12 = a/b =12
# B : unknown, but related to type A mice.
a1 = 120
b1 = 10
prior1 = st.gamma(a1, scale= 1/b1)

# prior parameters for thetaB ~ gamma(a2,b2)
# A : poisson distributed and mean is 12, E(thetaA) = 12 = a/b =12
a2 = 12
b2 = 1
prior2 = st.gamma(a2,scale= 1/b2)


# data in group A and posterior distribution
# n=number of mice, sy=sum of their numbers of Yi
nA = 10
```

```
syA = 117 #ya = (12,9,12,14,13,13,15,8,15,6)
postA = st.gamma(a1+syA, scale=1/(b1+nA))

# data in group B and posterior distribution
nB = 13
syB = sum([11,11,10,9,9,8,7,10,6,8,8,9,7]) #113
postB = st.gamma(a2+syB, scale=1/(b2+nB))
```

Group A Bayesian Analysis

```
# posterior mean (Group A)
(a1+syA)/(b1+nA)
```

```
11.85
```

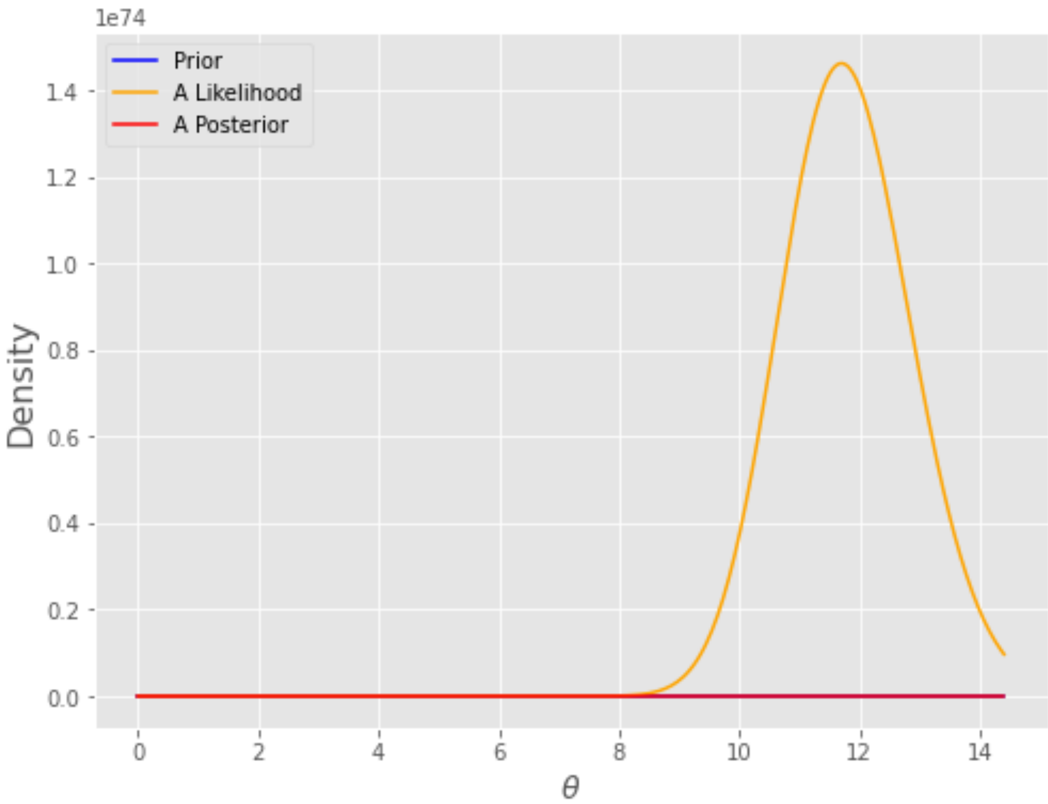
```
# posterior mode (Group A)
(a1+syA-1)/(b1+nA)
```

```
11.8
```

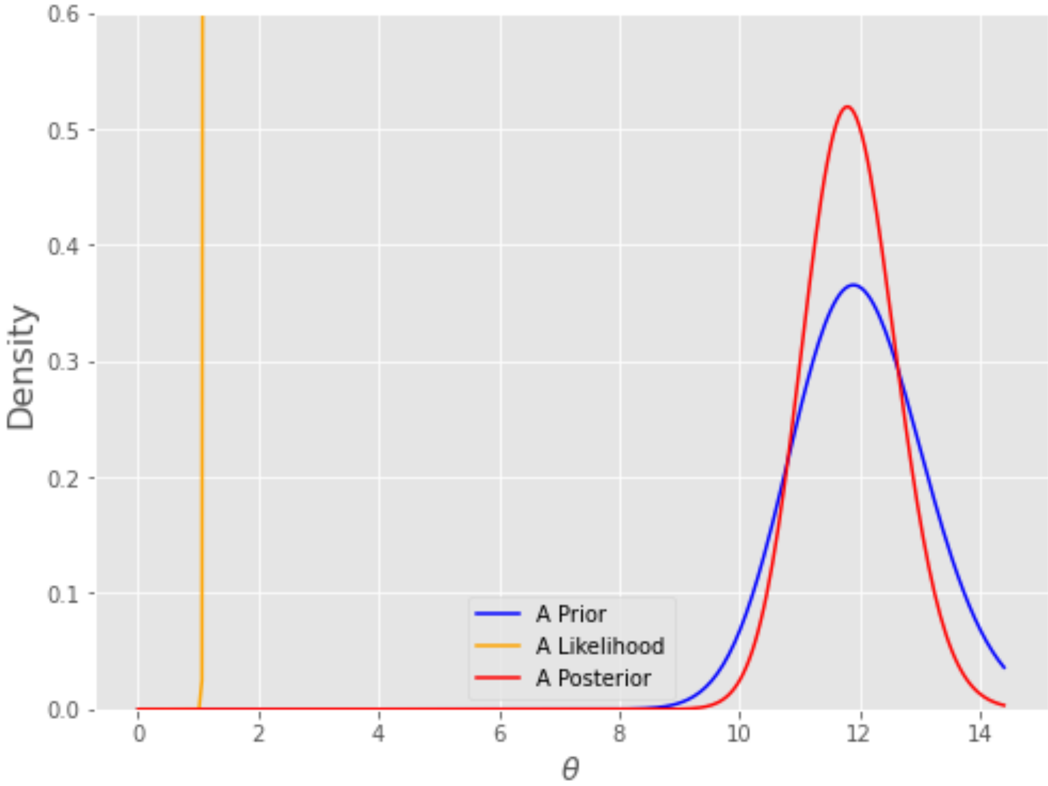
```
# posterior 95% CI (Group A)
# upper and lower bounds
lbA=gamma.ppf(0.025, a1+syA, scale=1/(b1+nA))
ubA=gamma.ppf(0.975, a1+syA, scale=1/(b1+nA))
(lbA, ubA)
```

```
(10.389238190941795, 13.405448325642006)
```

```
thetas = np.linspace(0.001, ubA+1, 300)
plt.figure(figsize=(8, 6))
plt.style.use('ggplot')
plt.plot(thetas, prior1.pdf(thetas), label='Prior', c='blue')

# 시각화를 위해 likelihood 함수에 임의의 상수를 곱함.
plt.plot(thetas, likelihood(thetas, nA, syA), label='A Likelihood', c='orange')
plt.plot(thetas, postA.pdf(thetas), label='A Posterior', c='red')
plt.xlabel(r'$\theta$', fontsize=14)
plt.ylabel('Density', fontsize=16)
plt.legend();
```




```
thetas = np.linspace(0.001, ubA+1, 300)
plt.figure(figsize=(8, 6))
plt.style.use('ggplot')
plt.plot(thetas, prior1.pdf(thetas), label='A Prior', c='blue')
plt.plot(thetas, likelihood(thetas, nA, syA), label='A Likelihood', c='orange')
plt.plot(thetas, postA.pdf(thetas), label='A Posterior', c='red')
plt.xlabel(r'$\theta$', fontsize=14)
plt.ylabel('Density', fontsize=16)
plt.ylim([0, 0.6])
plt.legend();
```



Group B Bayesian Analysis

```
# posterior mean (Group B)
(a2+syB)/(b2+nB)
```

```
8.928571428571429
```

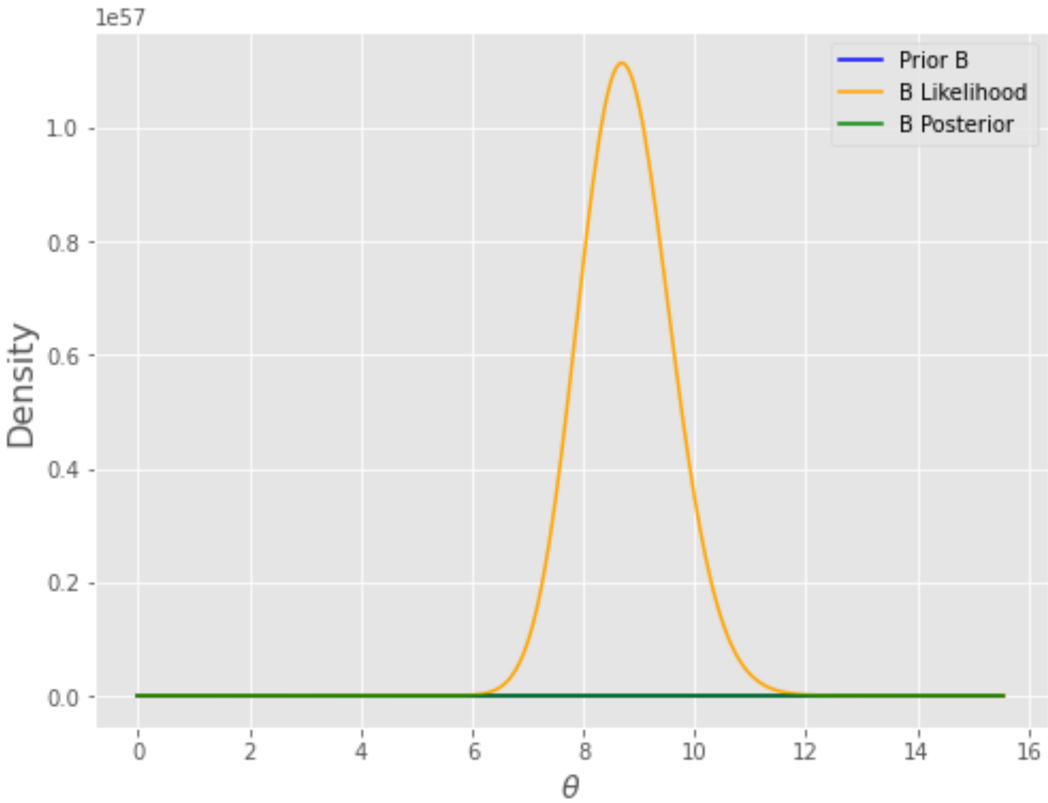
```
# posterior mode (Group B)
(a2+syB-1)/(b2+nB)
```

```
8.857142857142858
```

```
# posterior 95% CI (Group B)
# upper and lower bounds
lbB=gamma.ppf(0.025, a2+syB, scale=1/(b2+nB))
ubB=gamma.ppf(0.975, a2+syB, scale=1/(b2+nB))
(lbB,ubB)
```

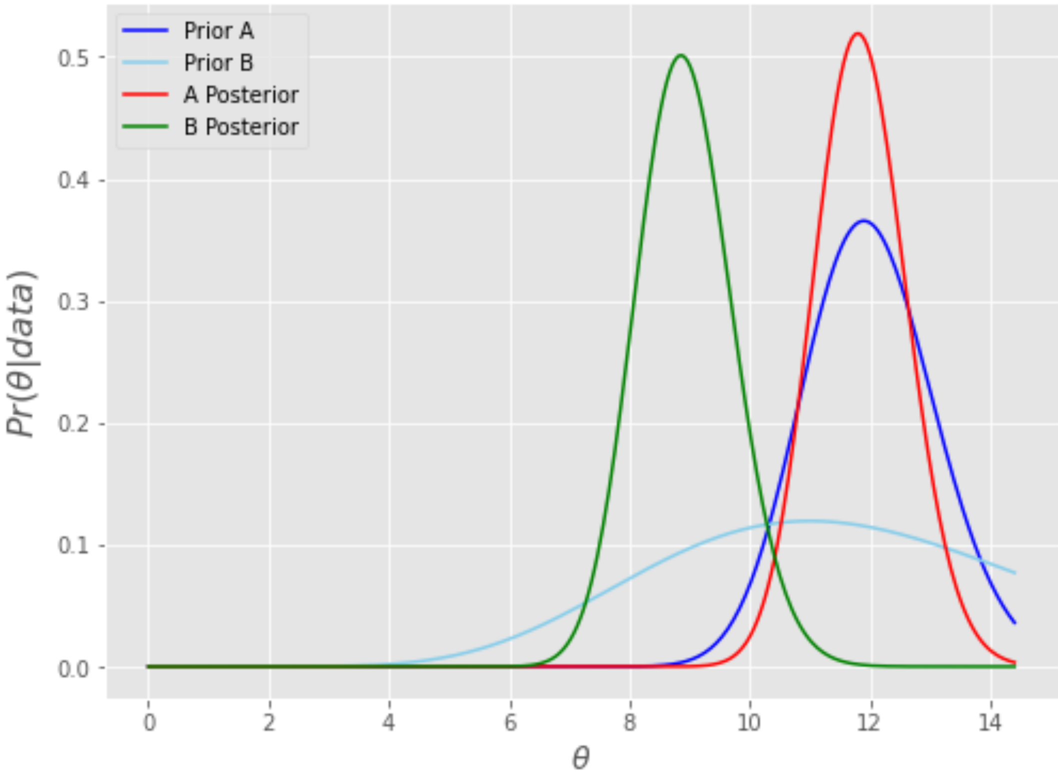
```
(7.432064219464302, 10.560308149242363)
```

```python
thetas = np.linspace(0.001, ubB+5, 300)
plt.figure(figsize=(8, 6))
plt.style.use('ggplot')
plt.plot(thetas, prior2.pdf(thetas), label='Prior B', c='blue')

# 시각화를 위해 likelihood 함수에 임의의 상수를 곱함.
plt.plot(thetas, likelihood(thetas, nB, syB), label='B Likelihood', c='orange')
plt.plot(thetas, postB.pdf(thetas), label='B Posterior', color='green')

plt.xlabel(r'$\theta$', fontsize=14)
plt.ylabel('Density', fontsize=16)
plt.legend();
```



Comparing Two Posterior Distributions

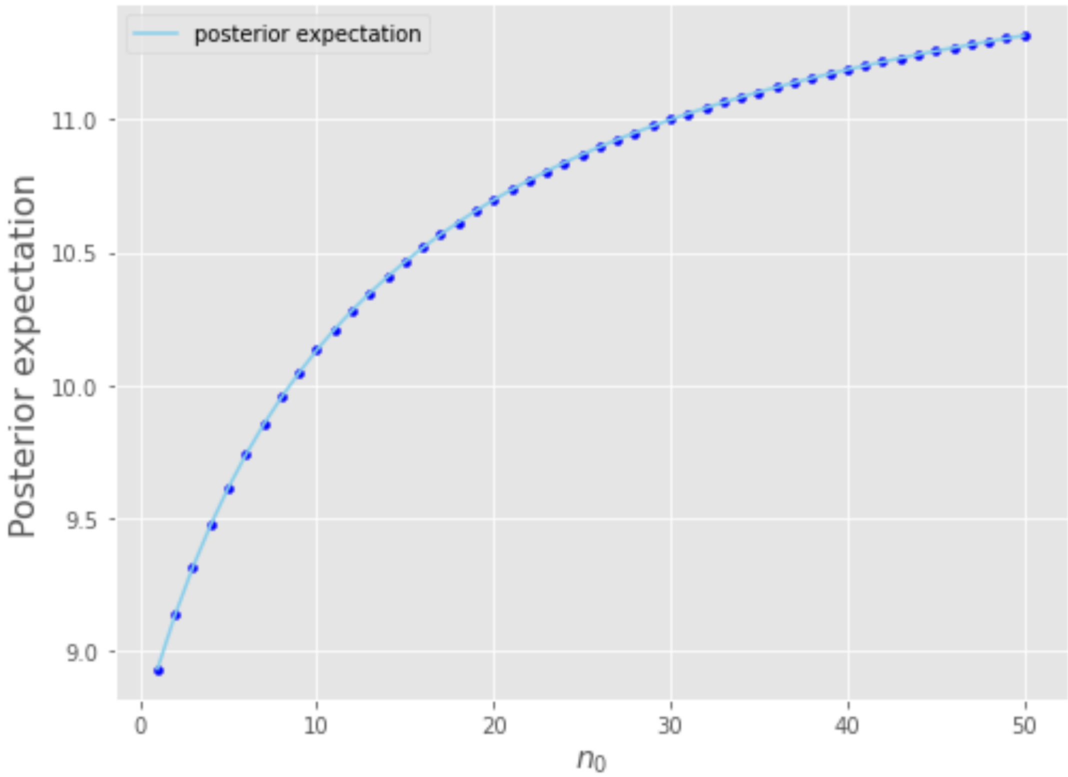
```
thetas = np.linspace(0.001, ubA+1, 300)
plt.figure(figsize=(8, 6))
plt.style.use('ggplot')
plt.plot(thetas, prior1.pdf(thetas), label='Prior A', c='blue')
plt.plot(thetas, prior2.pdf(thetas), label='Prior B', c='skyblue')
plt.plot(thetas, postA.pdf(thetas), label='A Posterior', c='red')
plt.plot(thetas, postB.pdf(thetas), label='B Posterior', c='green')
plt.xlabel(r'$\theta$', fontsize=14)
plt.ylabel(r'$Pr(\theta|data)$', fontsize=16)
plt.legend();
```



b) Compute and plot the posterioor expectation of theta B under theta B~ gamma(12*n0,n0)

```
nzeros = np.linspace(1, 50, 50) # n0={1,2,3,,,,,50}

plt.figure(figsize=(8, 6))
plt.style.use('ggplot')

#describe what sort of prior beliefs about theta B would be necessary in order
of the
#posterior expectation of thetaB

#gamma expectation = a/b
plt.plot(nzeros, (12*nzeros+syB)/(nzeros+nB), label='posterior expectation',
c='skyblue')
plt.scatter(nzeros, (12*nzeros+syB)/(nzeros+nB), c='blue', s=15)
plt.xlabel(r'$n_{0}$', fontsize=14)
```

```
plt.ylabel("Posterior expectation", fontsize=16)
plt.legend();
```



c. Discuss whether or not it makes sense to have

```
p(thetaA,thetaB) = p(thetaA)x p(thetaB)
```

theta A 와 thetaB가 독립인지 물어보는 것인데, A가 연구가 잘 되어 더 정확하고 큰 prior 을 주었고 B가 이와 관련되었다는 것을 알 수 있다. 하지만 이를 통해서 모든것을 정확히 알 수는 없기 때문에, 독립이라고 보아도 무방하다.

2. Data 가 binomial distribution 일 때, likelihood 를 exponential families 형태로 변환해보기.

① Likelihood

$$p(y_1, y_2 \cdots y_n|\theta) = \prod_{i=1}^{n} p(y_i|\theta)$$

$$= \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum y_i}(1-\theta)^{\sum(1-y_i)}$$

$$= \theta^{y}(1-\theta)^{n-y} = \left(\frac{\theta}{1-\theta}\right)^{y}(1-\theta)^{n}$$

$$= e^{\phi y}(1+e^{\phi})^{-n} \quad \left(\phi = \log\frac{\theta}{1-\theta}\right)$$

② Beta distribution → Conjugacy?

exponential families conjugacy

· Likelihood) $f(y_1 \cdots y_n|\theta) = \prod_{i=1}^{n} h(y_i)\cdot c(\phi)\cdot e^{\phi k(y_i)} \propto c(\phi)\cdot e^{\phi \sum k(y_i)}$

• binomial likelihood) $e^{\phi y}(1+e^{\phi})^{-n}$ ↔ $\phi = \log\frac{\theta}{1-\theta}$

• prior) $p(\phi|y) \propto k(n_0,t_0)\,c(\phi)e^{n_0t_0\phi}$

$\propto c(\phi)^{n_0} e^{n_0t_0\phi}$

beta prior) $p(\theta) \sim Beta(n_0t_0,\ n_0(1-t_0))$

$p(\theta) \propto \theta^{n_0t_0-1}(1-\theta)^{n_0(1-t_0)-1}$ ← $\phi=\log\frac{\theta}{1-\theta}$

$\theta^{n_0t_0-1}(1-\theta)^{-n_0t_0+n_0-1}$ $\quad e^{\phi} = \frac{\theta}{1-\theta}$

$\theta^{n_0t_0-1}\left(\frac{1}{(1-\theta)}\right)^{n_0t_0}(1-\theta)^{n_0}$ $\quad 1+e^{\phi} = 1+\frac{\theta}{1-\theta}\ \frac{1}{1-\theta}$

$\propto e^{\phi(n_0t_0-1)}(1+e^{\phi})^{2-n_0}$

$\propto e^{n_0t_0\phi}\cdot c(\phi)^{n_0}$

• posterior) $p(\phi|y) \propto p(\phi)f(y|\phi)\cdots$

$= c(\phi)^{n_0+n}\cdot \exp\{\phi(n_0t_0+ n\cdot\frac{\sum k(y_i)}{n}\}$

Beta posterior) $p(\phi|y) \propto p(\phi)\cdot f(y|\phi)$

$\propto e^{\phi(n_0t_0-1)}(1+e^{\phi})^{2-n_0}\cdot e^{\phi y}\cdot(1+e^{\phi})^{-n}$

$\propto e^{\phi(n_0t_0-1+y)}(1+e^{\phi})^{2-n_0-n}$

$\phi = \log\frac{\theta}{1-\theta}$ $\quad \propto \left(\frac{\theta}{1-\theta}\right)^{n_0t_0-1+y}\left(\frac{1}{1-\theta}\right)^{2-n_0-n}$

$e^{\phi} = \frac{\theta}{1-\theta}$ $\quad \propto \theta^{n_0t_0-1+y}(1-\theta)^{-\cancel{n_0t_0}+1-y+n+\cancel{n_0}-2}$

$\propto \theta^{n_0t_0-1+y}(1-\theta)^{n_0(1-t_0)+n-y-1}$

$\longrightarrow Beta(n_0t_0+y,\ n_0(1-t_0)+n-y)$

3. 증명 문제들

① prove the following

$Poi(\lambda) = \lim_{r\to\infty} NB\left(r, \overset{/\!/P}{\frac{\lambda}{r+\lambda}}\right)$ $\quad \frac{r+\lambda}{r+\lambda}$

$p(x) = \binom{r-1+x}{x}(1-p)^r p^x = \binom{r-1+x}{x}\left(1-\frac{\lambda}{r+\lambda}\right)^r\left(\frac{\lambda}{r+\lambda}\right)^x$

$= {}_{r-1+x}C_x\left(\frac{r}{r+\lambda}\right)^r\left(\frac{\lambda}{r+\lambda}\right)^x$

$= \frac{(r-1+x)!}{x!(r-1)!}\left(\frac{r}{r+\lambda}\right)^r\left(\frac{\lambda}{r+\lambda}\right)^x$

$\lim_{r\to\infty} p(x) = \lim_{r\to\infty}\frac{(r-1+x)!}{x!(r-1)!}\times\left(\frac{r}{r+\lambda}\right)^r\times\left(\frac{\lambda}{r+\lambda}\right)^x$

$$= \lim_{r\to\infty} \frac{(r-1+x)!}{x!(r-1)!} \times \frac{\left(\cancel{\frac{r+\lambda}{r}}\right)}{\cancel{\left(\frac{1}{1+\frac{\lambda}{r}}\right)^r}} \times \frac{\lambda^x}{(r+\lambda)^x}$$

$\left(e^x = \lim_{n\to\infty} \left(1+\frac{x}{n}\right)^n \right)$

ex) $\lim_{x\to\infty} \left(1+\frac{1}{x}\right)^{-2x} = \frac{1}{e^2}$

$$= \lim_{r\to\infty} \frac{(r-1+x)!}{x!(r-1)!} \times \cancel{\frac{1}{\left(1+\frac{\lambda}{r}\right)^r}} \left(1+\frac{\lambda}{r}\right)^{-r} \times \frac{\lambda^x}{(r+\lambda)^x}$$

$$= \lim_{r\to\infty} \frac{(r-1+x)!}{x!(r-1)!} \times e^{-\lambda} \cdot \frac{\lambda^x}{(r+\lambda)^x} = \lim_{r\to\infty} \cancel{\frac{(r-1+x)!}{(r-1)!(r+\lambda)^x}}\, e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

$$\lim_{r\to\infty} \frac{\overbrace{(r-1+x)(r-2+x)\cdots(r+x-(r-1))}^{r^x\text{승}, \ x\text{개}}}{(r+\lambda)^x} = 1.$$

$$= \lim_{r\to\infty} p(x) = \frac{e^{-\lambda}\cdot\lambda^x}{x!} = poi(\lambda)$$

---

② compare the variance of each distribution

Show that negative binomial is always overdispersed

↳ 과대산포. (변동성 ↑) → variance 더 큰것.

consider $\theta \sim gamma(\alpha, \beta)$, mean: $\frac{\alpha}{\beta}$, variance: $\frac{\alpha}{\beta^2}$ → $\alpha = \beta = \frac{1}{\sigma^2}$

(σ^2 만들게)

→ $Pr(Y=y) = \frac{\Gamma(\alpha+y)}{y!\,\Gamma(\alpha)} \cdot \frac{\beta^\alpha \mu^y}{(\mu+\beta)^{\alpha+y}}$ → $E(Y)=\mu$, if $\alpha=\beta=\frac{1}{\sigma^2}$

$Var(Y) = \mu(1+\sigma^2\mu)$, $\sigma^2>0$ so it's always bigger than μ (variance of poisson variance)

...음%

$$r\cdot\frac{(1-p)}{p^2} = \frac{r\left(\frac{r}{\lambda+r}\right)}{\left(\frac{\lambda}{\lambda+r}\right)^2} = \frac{\frac{r^2}{\cancel{(\lambda+r)}}}{\frac{\lambda^2}{\cancel{(\lambda+r)^2}}} \quad \frac{\frac{r^2}{\lambda^2}}{(\lambda+r)} = \frac{r^2(\lambda+r)}{\lambda^2} > \lambda$$

$(r>0)$

$$\left(\frac{r}{\lambda}\right)^2(\lambda+r) > \lambda$$

---

③ likewise, prove the following:

$Y \sim Binom(n,p)$ where $p(y) = \binom{n}{x} p^y p^{n-y}$

let mean $np=\lambda$ → $p=\frac{\lambda}{n}$

$poi(\lambda) = \lim_{n\to\infty} Binom(n, \frac{\lambda}{n})$

→ $Y\sim B(n,p) \Rightarrow p(y) = {}_nC_y\, p^y(1-p)^{n-y}$

$$= {}_nC_y \left(\frac{\lambda}{n}\right)^y \left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$\Rightarrow \frac{n!}{(n-y)!\,y!} \left(\frac{\lambda}{n}\right)^y \left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-y)!\,y!} \left(\frac{\lambda}{n}\right)^y \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{e^{-\lambda}} \left(1-\cancel{\frac{\lambda}{n}}\right)^{-y}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-y)!\,y!} 1^{-y} \cdot \frac{\lambda^y}{n^y} \cdot \cancel{\left(1-\frac{\lambda}{n}\right)^n} e^{-\lambda}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-y)!\,y!} \times \frac{1}{n^y} \cdot \frac{\lambda^y}{1} \cdot e^{-\lambda}$$

$$= \underbrace{\lim_{n\to\infty} \frac{n!}{(n-y)!\,n^y}}_{1} \times \frac{\lambda^y \cdot e^{-\lambda}}{y!}$$

$$\lim_{n\to\infty} \frac{\overbrace{n\times(n-1)\times\cdots(n-y+1)}^{y\text{개이므로 } n^y}\cancel{(n-y)!}}{\cancel{(n-y)!}\,n^y}$$

$$= 1$$

$$poi = \frac{e^{-\lambda}\lambda^y}{y!}$$