Introduction to probability and statistic: Assignment 2

Darix, SAMANI SIEWE

October 26, 2024

Exercise 1

Let Y be a r.v. that has a Poisson distribution with a parameter $\lambda > 0$ and X a r.v. such that X given Y = y a has a binomial distribution with parameters (y, p), $(0 < p < 1)$.

1. Show that the distribution of X is a Poisson distribution with parameter λp.

   $Y \sim \mathbb{P}(\lambda)$ and $X/Y \sim \mathbb{B}(y,p)$

   $P(Y=y) = \dfrac{\lambda^y}{y!}e^{-\lambda}$ and $P(X=x/Y=y) = \dbinom{y}{x}p^x(1-p)^{y-x}$

$$\begin{aligned}
P(X=x) &= \sum_{y=x}^{\infty} P(X=x, Y=y) \\
&= \sum_{y=x}^{\infty} P(Y=y)P(X=x/Y=y) \\
&= \sum_{y=x}^{\infty} \frac{\lambda^y}{y!}e^{-\lambda}\binom{y}{x}p^x(1-p)^{y-x} \\
&= \sum_{y=x}^{\infty} \frac{\lambda^y}{y!}e^{-\lambda}\frac{y!}{x!(y-x)!}p^x(1-p)^{y-x} \\
&= \sum_{y=x}^{\infty} \frac{\lambda^{y-x}\lambda^x}{x!(y-x)!}e^{-\lambda}p^x(1-p)^{y-x} \\
&= \sum_{y=x}^{\infty} \frac{(\lambda p)^x}{x!(y-x)!}e^{-\lambda}[\lambda(1-p)]^{y-x} \\
&= \frac{(\lambda p)^x}{x!}e^{-\lambda}\sum_{y=x}^{\infty}\frac{[\lambda(1-p)]^{y-x}}{(y-x)!} \qquad\qquad z = y-x \\
&= \frac{(\lambda p)^x}{x!}e^{-\lambda}\sum_{z=0}^{\infty}\frac{[\lambda(1-p)]^{z}}{(z)!} \\
&= \frac{(\lambda p)^x}{x!}e^{-\lambda}e^{\lambda(1-p)} \\
&= \frac{(\lambda p)^x}{x!}e^{-\lambda}e^{\lambda}e^{-\lambda p} \\
&= \frac{(\lambda p)^x}{x!}e^{-(\lambda p)}
\end{aligned}$$

   Based on the expression of $P(X=x) = \frac{(\lambda p)^x}{x!}e^{-(\lambda p)}$ we can say that X follows the poison distribution with parameter (λp)

2. Show that the distribution of $Y - X$ is a Poisson distribution with parameter $\lambda(1-p)$.

   we know that the momment generative function of poison distribution is $M_X(t) = e^{\lambda(e^t-1)}$

   In order to show that let's compute the moment generative function of $Y - X$

$$
\begin{aligned}
P(Y-X=z) &= \sum_{y=z}^{\infty} P(Y-X=z/Y=y)P(Y=y) \\
&= \sum_{y=z}^{\infty} P(X=y-z/Y=y)P(Y=y) \\
&= \sum_{y=z}^{\infty} \binom{y}{y-z} p^{y-z}(1-p)^{y-y+z}\frac{\lambda^y}{y!}e^-\lambda \\
&= \sum_{y=z}^{\infty} \frac{y!}{(y-z)!z!} p^{y-z}(1-p)^{z}\frac{\lambda^y}{y!}e^-\lambda \\
&= \frac{(1-p)^z e^-\lambda}{z!}\sum_{y=z}^{\infty} \frac{p^{y-z}\lambda^y}{(y-z)!} \\
&= \frac{(1-p)^z e^-\lambda}{z!}\sum_{y=z}^{\infty} \frac{p^{y-z}\lambda^{y-z}\lambda^z}{(y-z)!} \\
&= \frac{(1-p)^z e^-\lambda}{z!}\lambda^z\sum_{y=z}^{\infty} \frac{(p\lambda)^{y-z}}{(y-z)!} \qquad\qquad k=y-z \\
&= \frac{(1-p)^z e^-\lambda}{z!}\lambda^z\sum_{k=0}^{\infty} \frac{(p\lambda)^{k}}{(k)!} \\
&= \frac{(\lambda(1-p))^z e^-\lambda}{z!}e^{p\lambda} \\
&= \frac{(\lambda(1-p))^z e^{-\lambda+p\lambda}}{z!} \\
&= \frac{(\lambda(1-p))^z e^{-(1-p)\lambda}}{z!}
\end{aligned}
$$

based on the expression of $P(Y-X=z) = \dfrac{(p(1-p))^z e^{-(1-p)\lambda}}{z!}$ we can say that $Y-X$ follows the poison distribution with parameter $\lambda(1-p)$

3. Are X and $Y-X$ independent ?

   In Order to answer to this question let's Calculate $P(X=x/Y-Z=z)$ and based on this probability we will able de answer to this question:

$$
\begin{aligned}
P(X=x/Y-X=z) &= \frac{P(X=x, Y-X=Z)}{P(Y-X=z)} \\
&= \frac{P(X=x, Y=x+z)}{P(Y-X=z)} \\
&= \frac{P(Y=z+x)P(X=x/Y=z+x)}{P(Y-X=z)} \\
&= \frac{\frac{\lambda^{z+x}}{(z+x)!}\binom{z+x}{x}p^x(1-p)^z}{\frac{[\lambda(1-p)]^z}{z!}e^{-\lambda(1-p)}} \\
&= \frac{\frac{\lambda^{z+x}}{(z+x)!}\frac{(z+x)!}{x!z!}p^x(1-p)^z}{\frac{[\lambda(1-p)]^z}{z!}e^{-\lambda(1-p)}} \\
&= \frac{\lambda^x}{x!}p^x e^{-\lambda P} \\
&= \frac{(\lambda p)^x}{x!}e^{-\lambda p} \\
&= P(X=x)
\end{aligned}
$$

so $P(X = x/Y - X = z) = P(X = x)$ that mean X and $X - Y$ are independent variable aleatoric

Exercise 2

Let $X_1, ..., X_n$ be a sample of n independent and identically distributed (i.i.d.) random variables with a standard normal distribution.

1. Show that X_i^2, i = 1,...,n has a gamma distribution.

$$\begin{aligned} P(X_i^2 \leq \sqrt{x}) &= P(\sqrt{X_i} < X_i \leq \sqrt{x}) \\ &= P(X_i \leq \sqrt{x}) - P(X_i \leq -\sqrt{x}) \\ &= 2\Phi(\sqrt{x}) - 1 \end{aligned}$$

$$\begin{aligned} f_{X_i^2}(x) = F'_{X_u^2}(x) &= [P(X_i^2 \leq \sqrt{x})]' = [2\Phi(\sqrt{x}) - 1]' \\ &= 2\Phi'(\sqrt{x})(\sqrt{x})' \\ &= 2\phi(\sqrt{x})\frac{1}{2\sqrt{x}} \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}\frac{1}{\sqrt{x}} \\ &= \frac{\frac{1}{2}^{\frac{1}{2}}}{\sqrt{\pi}}x^{-\frac{1}{2}}e^{-\frac{1}{2}x} \\ &= \frac{\frac{1}{2}^{\frac{1}{2}}}{\sqrt{\pi}}x^{\frac{1}{2}-1}e^{-\frac{1}{2}x} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} \qquad \text{where} \qquad \alpha = \beta = \frac{1}{2} \qquad \text{and} \qquad \Gamma(\alpha) = \sqrt{\pi} \end{aligned}$$

let's prove that $\Gamma(\alpha) = \Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$\begin{aligned} \Gamma(\frac{1}{2}) &= \int_0^{+\infty} x^{-\frac{1}{2}e^{-x}dx} \\ &= \int_0^{+\infty} \frac{1}{\sqrt{x}e^{-x}dx} \text{let's donote } y = \sqrt{x}\frac{dy}{dx} = \frac{1}{2\sqrt{x}} \implies dx = 2\sqrt{x}dy = 2ydy \\ &= \int_0^{+\infty} \frac{1}{y}e^{-y^2}2ydy \\ &= 2\int_0^{+\infty} e^{-y^2}dy \\ &= \int_{-\infty}^{+\infty} e^{-y^2}dy \qquad \text{we recognize the Gaussian integral} \\ &= \sqrt{\pi} \end{aligned}$$

we have prove that the probability density function of $f_{X_i^2}(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ where $\alpha = \beta = \dfrac{1}{2}$ and $\Gamma(\alpha) = \sqrt{\pi}$
so, we can conclute that X_i^2 follows the Gamma distribution with parameter $\alpha = \beta = \frac{1}{2}$

2. Let consider $K_n = \sum_{i=0}^{n} X_i^2$

Show that K_n follows a gamma distribution with parameter $(\frac{n}{2}, \frac{1}{2})$ which is also called a chi-square distribution with n degrees of freedom, denoted $\chi^2(n)$

we know that $X_i^2, \ldots, X_n^2$ are n identically independent random variable because $X_i, \ldots, X_n$ are n identically independent random variable and we also known that the moment generative function of random variable X that follows the gamma distribution is $M_X(t) = (\dfrac{\beta}{\beta - t})^{\alpha}$

In order to prove that let's computer the moment generative function of K_n

$$\begin{aligned}
M_{K_n}(t) &= M_{\sum_{i=0}^{n} X_i^2}(t) \\
&= \prod_{i=0}^{n} M_{X_i^2}(t) \\
&= \prod_{i=0}^{n} (\frac{\frac{1}{2}}{\frac{1}{2} - t})^{\frac{1}{2}} \qquad \text{because } X_i^2 \sim \mathbb{G}(\frac{1}{2}, \frac{1}{2}) \\
&= (\frac{\frac{1}{2}}{\frac{1}{2} - 1})^{\frac{n}{2}}
\end{aligned}$$

$M_{K_n}(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - 1}\right)^{\frac{n}{2}}$ that mean K_n follows the gamma distribution with parameter $(\frac{n}{2}, \frac{1}{2})$

Exercise 3

Let f_1 and f_2, be two probability density functions such that :
$\int x f_1(x) = \mu_1$ and $\int x f_2(x) = \mu_2$
assume that $mu_1 > \mu_2$
Let X be a random variable with probability density function :
$f_X(x) = \theta f_1(x) + (1 - \theta) f_2(x). \quad \theta \in [0, 1]$

1. Check that f is a probability density function.

   In order to answer to this question we need to prove that f is positive and $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

   - Let's prove that $f_X(x) \geq 0$:
     we know that f_1 and f_2 are the density probability that mean that $f_1(x) \geq 0$ and $f_2(x) \geq 0$ and $\theta \in [0, 1]$ and $(\theta) \in [0, 1]$
     so, f is a summation of positive value that mean $\forall \theta \in [0, 1] f(x) \geq 0$
   - now, let's prove that $\int_{-\infty}^{+\infty} f_X(x)dx = 1$:

$$\begin{aligned}
\int_{-\infty}^{+\infty} f_X(x)dx = 1 &= \int_{-\infty}^{+\infty} \theta f_1(x) + (1 - \theta) f_2(x) \\
&= \theta \int_{-\infty}^{+\infty} f_1(x)dx + (1 - \theta) \int_{-\infty}^{+\infty} f_2(x)dx \\
&= \theta * 1 + (1 - \theta) * 1 \\
&= 1
\end{aligned}$$

   so f is a probability density function

2. Express the expectation of X for any $\theta \in [0, 1]$.

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x f_X(x)dx \\
&= \int_{-\infty}^{+\infty} x[\theta f_1(x) + (1-\theta) f_2(x)]dx \\
&= \int_{-\infty}^{+\infty} x\theta f_1(x)dx + \int_{-\infty}^{+\infty} x(1-\theta) f_2(x)]dx \\
&= \theta \int_{-\infty}^{+\infty} x f_1(x)dx + (1-\theta) \int_{-\infty}^{+\infty} x f_2(x)]dx \\
&= \theta\mu_1 + (1-\theta)\mu_2
\end{aligned}
$$

so $E(X) = \theta\mu_1 + (1-\theta)\mu_2$

3. Let $(X_1, \ldots, X_n)$ be n i.i.d r.v.'s with the p.d.f. $f_X(x)$.

   (a) Use the method of moments to propose an estimator of θ.

   - Theatrical value:
     $E(X) = \theta\mu_1 + (1-\theta)\mu_2$
   - empirically value: $\bar{X_n} = \sum_{i=0}^{n} X_i$

$$
\begin{aligned}
E(X) &= \bar{X_n} \\
\theta\mu_1 + (1-\theta)\mu_2 &= \bar{X_n} \\
\theta(\mu + 1 - \mu_2) &= \bar{X_n} - \mu_2 \\
\implies \theta &= \frac{\bar{X_n} - \mu_2}{\mu_1 - \mu_2}
\end{aligned}
$$

   so $MME(\theta) = \dfrac{\bar{X_n} - \mu_2}{\mu_1 - \mu_2}$

   (b) Is this estimator unbiased ?

   To answer to this question let's compute $E(\hat{\theta})$:

$$
\begin{aligned}
E(\hat{\theta}) &= E(\frac{\bar{X_n} - \mu_2}{\mu_1 - \mu_2}) \\
&= \frac{1}{\mu_1 - \mu_2} E(\bar{X_n} - \mu_2) \\
&= \frac{1}{\mu_1 - \mu_2} \left[E(\bar{X_n}) - \mu_2\right] \\
&= \frac{1}{\mu_1 - \mu_2} \left[E(\frac{1}{n}\sum_{i=0}^{n} X_i) - \mu_2\right] \\
&= \frac{1}{\mu_1 - \mu_2} \left[\frac{1}{n}\sum_{i=0}^{n} E(X_i) - \mu_2\right] \\
&= \frac{1}{\mu_1 - \mu_2} \left[\frac{1}{n} n E(X_i) - \mu_2\right] \\
&= \frac{1}{\mu_1 - \mu_2} \left[\theta\mu_1 + (1-\theta)\mu_2 - \mu_2\right] \\
&= \frac{1}{\mu_1 - \mu_2} \left[\theta\mu_1 + \mu_2 - \theta\mu_2 - \mu_2\right] \\
&= \frac{1}{\mu_1 - \mu_2} \left[\theta\mu_1 - \theta\mu_2\right] \\
&= \frac{1}{\mu_1 - \mu_2} \left[\mu_1 - \mu_2\right]\theta \\
&= \theta
\end{aligned}
$$

$E(\hat{\theta}) = \theta$ that mean estimator $\hat{\theta}$ is unbiased

Exercise 4

Let $(X_1, \ldots, X_n)$ be n independent and identically distributed random variables following a log-normal distribution with parameter (μ, σ^2).

$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}$

1. Find the $\hat{\mu}$ and $\hat{\sigma}^2$, the maximum likelihood estimator of μ and σ^2

   - Likelihood:

$$
\begin{aligned}
L(\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} X_i} e^{-\frac{(\log(X_i)-\mu)^2}{2\sigma^2}} \\
&= (\frac{1}{\sqrt{2\pi}})^n (\frac{1}{\sigma})^n \left(\prod_{i=1}^{n} \frac{1}{X_i}\right) e^{-\frac{1}{2\sigma^2}\sum_{i=0}^{n}(\log(X_i-\mu))^2}
\end{aligned}
$$

   - log-likelihood:

$$
\log(L(\mu, \sigma^2)) = n\log(\frac{1}{\sqrt{2\pi}}) - \frac{n}{2}\log(\sigma^2) + \log(\prod_{i=0}^{n} \frac{1}{X_i}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log(X_i) - \mu)^2
$$

   - likelihood-Equation:

$$
\begin{cases} \frac{d}{d\mu}\log(L(\mu, \sigma^2)) = 0 \\ \frac{d}{d\sigma^2}\log(L(\mu, \sigma^2)) = 0 \end{cases} \implies \begin{cases} \frac{1}{\sigma^2}\sum_{i=0}^{n}(\log(X_i) - \mu) = 0 \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=0}^{n}(\log(X_i - \mu))^2 = 0 \end{cases} \implies \begin{cases} \sum_{i=0}^{n}(\log(X_i) - \mu) = 0 \\ -n\sigma^2 + \sum_{i=0}^{n}(\log(X_i - \mu))^2 = 0 \end{cases}
$$

$$
\begin{cases} \sum_{i=0}^{n}\log(X_i) - \sum_{i=0}^{n}\mu = 0 \\ \sum_{i=0}^{n}(\log(X_i - \mu))^2 = n\sigma^2 \end{cases} \implies \begin{cases} \mu = \frac{1}{n}\sum_{i=0}^{n}\log(X_i) \\ \sigma^2 = \frac{1}{n}\sum_{i=0}^{n}(\log(X_i - \mu))^2 \end{cases}
$$

so we have found the value :

$\begin{cases}\hat{\mu} = \frac{1}{n}\sum_{i=0}^{n}\log(X_i) \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=0}^{n}(\log(X_i) - \hat{\mu}))^2\end{cases}$

2. Show that $\hat{\mu}$ is unbiased.

$$\begin{aligned}
E(\hat{\mu}) &= E(\frac{1}{n}\sum_{i=0}^{n}\log(X_i)) \\
&= \frac{1}{n}E(\sum_{i=0}^{n}\log(X_i)) \\
&= \frac{1}{n}\sum_{i=0}^{n}E(\log(X_i)) \\
&= \frac{1}{n}\sum_{i=0}^{n}E(Y_i) \qquad \text{where } Y_i \sim \mathbb{N}(\mu, \sigma^2) \\
&= \frac{1}{n}\sum_{i=0}^{n}\mu \\
&= \frac{1}{n}n\mu \\
&= \mu
\end{aligned}$$

$E(\hat{\mu}) = \mu$ that mean that $\hat{\mu}$ is unbiased

3. Show that $\hat{\sigma}^2$ is biased but asymptotically unbiased.

$$
\begin{aligned}
E(\hat{\sigma}^2) &= E(\frac{1}{n}\sum_{i=0}^{n}(\log(X_i)-\hat{\mu}))^2) \\
&= \frac{1}{n}E\left(\sum_{i=0}^{n}(\log(X_i)-\hat{\mu})^2\right) \\
&= \frac{1}{n}E\left(\sum_{i=0}^{n}\left((\log(X_i))^2 - 2\hat{\mu}\log(X_i)+\hat{\mu}^2\right)\right) \\
&= \frac{1}{n}E\left(\sum_{i=0}^{n}(\log(X_i))^2 - 2\hat{\mu}\sum_{i=0}^{n}\log(X_i)+\sum_{i=0}^{n}\hat{\mu}^2)\right) \\
&= \frac{1}{n}E\left(\left(\sum_{i=0}^{n}(\log(X_i))^2 - 2\hat{\mu}n\hat{\mu}+n\hat{\mu}^2\right)\right) \\
&= \frac{1}{n}E\left(\sum_{i=0}^{n}(\log(X_i))^2 - n\hat{\mu}^2\right) \\
&= \frac{1}{n}\left(E\left(\sum_{i=0}^{n}(\log(X_i))^2\right) - nE(\hat{\mu}^2)\right) \\
&= \frac{1}{n}\left(\sum_{i=0}^{n}E\left(\log(X_i)\right)^2 - nE\left(\frac{1}{n^2}\left(\sum_{i=0}^{n}\log(X_i)\right)^2\right)\right) \\
&= \frac{1}{n}\left(\sum_{i=0}^{n}E\left(Y_i^2\right) - \frac{1}{n}E\left(\sum_{i=0}^{n}Y_i\right)^2\right) \qquad \text{where } Y_i \sim \mathbb{N}(\mu,\sigma^2) \\
&= \frac{1}{n}\left(\sum_{i=0}^{n}(\mu^2+\sigma^2) - \frac{1}{n}E\left(\sum_{i=0}^{n}Y_i^2+\sum_{i\neq j}^{n}X_iX_j\right)\right) \\
&= \frac{1}{n}\left(n(\mu^2+\sigma^2) - \frac{1}{n}\left(\sum_{i=0}^{n}E(Y_i^2)+\sum_{i\neq j}^{n}E(X_iX_j)\right)\right) \\
&= \frac{1}{n}\left(n(\mu^2+\sigma^2) - \frac{1}{n}\left(n(\mu^2+\sigma^2)+\sum_{i\neq j}^{n}E(X_i)E(X_j)\right)\right) \\
&= \frac{1}{n}\left(n(\mu^2+\sigma^2) - \frac{1}{n}\left(n(\mu^2+\sigma^2)+n(n-1)\mu^2\right)\right) \\
&= \frac{1}{n}\left(n(\mu^2+\sigma^2) - (\mu^2+\sigma^2) - (n-1)\mu^2\right) \\
&= \frac{1}{n}\left(n\mu^2+n\sigma^2-\mu^2-\sigma^2-n\mu^2+\mu^2)\right) \\
&= \frac{1}{n}(n\sigma^2-\sigma^2) \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$ that mean $\hat{\sigma}^2$ is biased
$\lim_{x\to\infty} E(\hat{\sigma}^2) = \lim_{x\to\infty}\frac{n-1}{n}\sigma^2 = \sigma^2$ that mean $\hat{\sigma}^2$ is asymptotically unbiased.

4. Find the distribution of $\hat{\mu}$ and the distribution of $\hat{\sigma^2}$

$$
\begin{aligned}
M_{\hat{\mu}} &= E(e^{\frac{t}{n}\sum_{i=0}^{n}\log(X_i)}) \\
&= E(\prod_{i=0}^{n} e^{\frac{t}{n}\log(X_i)}) \\
&= \prod_{i=0}^{n} E(e^{\frac{t}{n}\log(X_i)}) \\
&= (M_{\log(X_i)}(\frac{t}{n}))^n \\
&= (M_{Y_i}(\frac{t}{n}))^n \qquad \text{where } Y_i \sim \mathbb{N}(\mu, \sigma^2) \\
&= \left(e^{\mu(\frac{t}{n})+\frac{1}{2}\sigma^2(\frac{t}{n})^2)}\right)^n \\
&= e^{\mu t+\frac{1}{2}\sigma^2(\frac{t^2}{n}))} \\
&= e^{\mu t+\frac{1}{2}\frac{\sigma^2}{n}t^2))}
\end{aligned}
$$

so $M_{\hat{\mu}} = e^{\mu t+\frac{1}{2}\frac{\sigma^2}{n}t^2))}$ it's a moment generative function of normal distribution with parameter $(\mu, \frac{\sigma^2}{n})$
so $\hat{\mu} \sim \mathbb{N}(\mu, \frac{\sigma^2}{n})$

$$
\begin{aligned}
M_{\hat{\sigma}^2} &= E(e^{\frac{t}{n}\sum_{i=0}^{n}(\log(X_i)-\hat{\mu}))^2}) \\
&= E(e^{\frac{t}{n}\sum_{i=0}^{n}((\log(X_i))^2-2\hat{\mu}\log(X_i))+\hat{\mu}^2)}) \\
&= E(e^{\frac{t}{n}((\sum_{i=0}^{n}(\log(X_i))^2-2\hat{\mu}\sum_{i=0}^{n}\log(X_i))+\sum_{i=0}^{n}\hat{\mu}^2)}) \\
&= E(e^{\frac{t}{n}((\sum_{i=0}^{n}(\log(X_i))^2-\hat{\mu}n\hat{\mu}+n\hat{\mu}^2)}) \\
&= E(e^{\frac{t}{n}((\sum_{i=0}^{n}(\log(X_i))^2-n\hat{\mu}^2)})
\end{aligned}
$$