1. $X_1, \ldots, X_n$ is a random variable having the bernoulli distribution with the parameter θ.

   (a) **(2 mark)** Let $S = \sum_{i=1}^n X_i$, find the distribution of S.

   (b) **(2 marks)** Find the maximum likelihood estimate of $\theta(1-\theta)$.

   (c) **(2 marks)** Is it unbiased estimator? No mark will be given if the answer is "Yes" or "No".

   (d) **(2 marks)** Find the Cramer-Rao Lower Bound for the variance of an unbiased estimator for $\theta(1-\theta)$.

   (e) **(2 mark)** Does the variance of any unbiased estimator for $\theta(1-\theta)$ achieve this bound? Why? Explain in details.

   (f) **(3 marks)** Find the limiting distribution of the maximum likelihood estimate for $\theta(1-\theta)$ by Central Limit Theorem and Delta method. What phenomenon do you observe?

Solutions:

(a) Since $X_i \overset{\text{i.i.d.}}{\sim}$ Bernoulli(θ), we have

$$\begin{aligned} M_{X_i}(t) &= (1-\theta) + \theta e^t \\ M_S(t) &= M_{\sum_{i=1}^n X_i}(t) = \Pi_{i=1}^n M_{X_i}(t) = ((1-\theta) + \theta e^t)^n \end{aligned}$$

Therefore, $S = \sum_{i=1}^n X_i \sim$ Bin(n,θ).

(b)

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\ l(\theta) &= log f_{\mathbf{X}}(\mathbf{x}) = (\sum_{i=1}^n x_i) log\theta + (n - \sum_{i=1}^n x_i) log(1-\theta) \\ l'(\theta) &= \frac{\sum_{i=1}^n x_i}{\theta} + \frac{n - \sum_{i=1}^n x_i}{\theta - 1} \end{aligned}$$

Taking $l'(\theta) = 0$, we have $\theta = \bar{x}$ and $l''(\theta) < 0$ at $\theta = \bar{x}$.Therefore, we get the MLE for θ, which is $\hat{\theta} = \bar{X}$. Then by the invariance property of MLE, we get the MLE of $\theta(1-\theta)$ is $\hat{\theta}(1-\hat{\theta}) = \bar{X}(1-\bar{X})$.

(c) By part (a), we have

$$\begin{aligned} E(S) &= n\theta \\ Var(S) &= n\theta(1-\theta) \\ E(S^2) &= n\theta(1-\theta+n\theta) \end{aligned}$$

Since $\hat{\theta} = \bar{X} = \frac{1}{n}S$, we have

$$\begin{aligned} E(\hat{\theta}) &= \frac{1}{n}E(S) = \theta \\ E(\hat{\theta}^2) &= \frac{1}{n^2}E(S^2) = \frac{\theta(1-\theta+n\theta)}{n} \\ E(\hat{\theta}(1-\hat{\theta})) &= \theta - \frac{\theta(1-\theta+n\theta)}{n} = \frac{n-1}{n}\theta(1-\theta) \neq \theta(1-\theta) \end{aligned}$$

Therefore, it is not unbiased.

(d) By part(c), take T = $\frac{n}{n-1}\bar{X}(1-\bar{X})$ and E(T) = $\theta(1-\theta)$, so T is an unbiased estimator of $g(\theta) = \theta(1-\theta)$.

$$
\begin{aligned}
logf_{X_i}(x_i;\theta) &= x_i log\theta + (1-x_i)log(1-\theta) \\
\frac{\partial^2}{\partial\theta^2}logf_{X_i}(x_i;\theta) &= -\frac{x_i}{\theta^2} - \frac{1-x_i}{(1-\theta)^2} \\
E(\frac{\partial^2}{\partial\theta^2}logf_{X_i}(x_i;\theta)) &= -\frac{1}{\theta} - \frac{1}{1-\theta} = -\frac{1}{\theta(1-\theta)} \\
g'(\theta) = 1-2\theta & \\
CR\ Lower\ Bound &= -\frac{g'(\theta)^2}{nE(\frac{\partial^2}{\partial\theta^2}logf_{X_i}(x_i;\theta))} \\
&= \frac{(1-2\theta)^2}{\frac{n}{\theta(1-\theta)}}
\end{aligned}
$$

(e)

$$
\frac{\partial}{\partial\theta}logf_{\mathbf{X}}(\mathbf{x};\theta) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n-\sum_{i=1}^n x_i}{1-\theta} = \frac{\sum_{i=1}^n x_i - n\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}(\frac{\sum_{i=1}^n x_i}{n} - \theta)
$$

There is no function $A(n,\theta)$ s.t. $\frac{\partial}{\partial\theta}logf_{\mathbf{X}}(\mathbf{x};\theta) = A(n,\theta)[h(\mathbf{x})-g(\theta)]$, where $g(\theta) = \theta(1-\theta)$ and $h(\mathbf{x})$ is an unbiased estimator of $g(\theta)$. Therefore, no unbiased estimator of $g(\theta)$ achieve the CR lower bound.

(f) By C.L.T., we have $\hat{\theta} \to$ N(θ, $\frac{\theta(1-\theta)}{n}$). Then by Delta Method, we have g($\hat{\theta}$) $\to$ N($g(\theta)$, $\frac{g'(\theta)^2\theta(1-\theta)}{n}$) = N($\theta(1-\theta)$, $\frac{(1-2p)^2p(1-\theta)}{n}$). So the limiting variance of MLE can achieve the CR lower bound.

2. $X_1, X_2, ..., X_n$ are observations of a random sample of size n from the geometric distribution with probability distribution $f(x,\theta) = \theta(1-\theta)^x$

(a) **(2 marks)** Find the distribution of $\sum_{i=1}^{n} X_i$.

(b) **(2 mark)** Find the estimator from the method of moment.

(c) **(2 mark)** Find the estimator from the method of maximum likelihood.

(d) **(2 marks)** Find the maximum likelihood estimator for $E(X)$.

(e) **(2 marks)** Is the maximum likelihood estimator for $E(X)$ unbiased? If yes, find its variance. If no, find its mean squared error.

(f) **(2 marks)** Is the variance of any unbiased estimator for $E(X)$ equal to the Cramer-Rao Lower Bound? No need to find the Cramer-Rao Lower Bound.

Solutions:

(a) Since $X_i \overset{\text{i.i.d.}}{\sim}$ Geometric(θ), we have

$$
\begin{aligned}
M_{X_i}(t) &= \frac{\theta}{1-(1-\theta)e^t} \\
M_S(t) &= M_{\sum_{i=1}^n X_i}(t) = \Pi_{i=1}^n M_{X_i}(t) = (\frac{\theta}{1-(1-\theta)e^t})^n
\end{aligned}
$$

Therefore, $S = \sum_{i=1}^n X_i \sim$ NB(n,θ).

(b) For MME:

$$
\begin{aligned}
M_1' &= \widetilde{E(X)} \\
\frac{1}{n}\sum_{i=1}^n x_i &= \frac{\widetilde{1-\theta}}{\theta} \\
\tilde{\theta} &= \frac{1}{1+\bar{X}}
\end{aligned}
$$

(c)

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \theta^n(1-\theta)^{\sum_{i=1}^n x_i} \\
l(\theta) &= log f_{\mathbf{X}}(\mathbf{x}) = nlog\theta + (\sum_{i=1}^n x_i)log(1-\theta) \\
l'(\theta) &= \frac{n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta-1}
\end{aligned}
$$

Taking $l'(\theta) = 0$, we have $\theta = \frac{1}{1+\bar{x}}$ and $l''(\theta) < 0$ at $\theta = \frac{1}{1+\bar{x}}$.Therefore, we get the MLE for θ, which is $\hat{\theta} = \frac{1}{1+\bar{X}}$.

(d) EX $= \frac{1-\theta}{\theta}$. Therefore, by the invariant property of MLE, we have the MLE of EX is $\frac{1-\hat{\theta}}{\hat{\theta}}$ $= \bar{X}$

(e)

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

Therefore, the MLE of EX is unbiased.

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

(f)

$$
\begin{aligned}
\frac{\partial}{\partial\theta} log f_{\mathbf{X}}(\mathbf{x};\theta) &= \frac{n}{\theta} - \frac{\sum_{i=1}^n x_i}{1-\theta} = \frac{n(1-\theta) - \sum_{i=1}^n x_i\theta}{\theta(1-\theta)} \\
&= \frac{-n}{1-\theta}(\bar{x} - \frac{1-\theta}{\theta})
\end{aligned}
$$

Therefore, the variance of the unbiased estimator $\bar{X}$ is equal to the CR lower bound.

3. Let $X_1, \ldots, X_n$ are independently uniformly distributed on $(\theta, 2\theta)$.