1. (*20pts, 2pts for each*)

   (a) True.

   (b) False. Two independent random variables have zero correlation coefficient. The converse statement is not true in general.

   (c) False. The mean and variance of Cauchy distribution do not exist. Therefore, we cannot apply the version of WLLN taught in class for this case. **Actually,** $\overline{X}_n$ **has the same distribution as** X_1 (check LN, Ch1-6, p.89, item 6).

   (d) False. A point estimate is a value while a point *estimator* is a random variable.

   (e) True.

   (f) False. When θ is fixed, the summation of probabilities (or integration of density) over all possible data is one. The likelihood function represents the probability of observing *same* data under different values of θ. Therefore, the summation (or integration) of likelihood function over θ need not be one.

   (g) False. It should be "no *unbiased* estimator".

   (h) True.

   (i) False. T contains all information *about* θ. Some information that is not related to θ could be lost during the transformation T.

   (j) False. Because the function $n(\overline{X}_n - \mu)^2$ contains an unknown parameter μ, it is not a statistic.

2. (*18pts, 3pts for each*)

   (a) $X \sim$ Hypergeometric(r, n, m), where $r = 6$, $n = 6$, $m = 53 - 6 = 47$.

   (b) $X \sim$ Poisson(λ), where λ is unknown.

   (c) $X \sim$ Normal(μ, σ^2), where $\mu = 0$ and σ^2 is unknown.

   (d) $X \sim$ Uniform(1, 2, ..., N), where N is unknown.

   (e) Let θ be the probability of getting a head, which is unknown, then $X_1 \sim$ Binomial(3, θ), $X_2 \sim$ Geometric(θ), and X_1, X_2 are independent.

   (f) $(X_1, X_2, X_3, X_4) \sim$ Multinomial(n, p_1, p_2, p_3, p_4), where p_1, p_2, p_3, p_4 are unknown.

3. (a) (*6pts*) The cdf of Y_n is

$$
\begin{aligned}
F_{Y_n}(y) &= P(Y_n \le y) = P(X_1 \le y, \ldots, X_n \le y) \\
&= \prod_{i=1}^{n} P(X_i \le y) = \left(\frac{y}{\theta}\right)^n
\end{aligned}
$$

for $0 \le y \le \theta$. So,

$$\begin{aligned} P(|Y_n - \theta| < \epsilon) &= P(\theta - \epsilon < Y_n < \theta + \epsilon) = P(\theta - \epsilon < Y_n < \theta) = 1 - P(Y_n \le \theta - \epsilon) \\ &= 1 - F_{Y_n}(\theta - \epsilon) = 1 - \left(\frac{\theta - \epsilon}{\theta}\right)^n \\ &= 1 - \left(1 - \frac{\epsilon}{\theta}\right)^n \longrightarrow 1, \quad \text{as } n \to \infty. \end{aligned}$$

(Note. Compare the result with the consistent property of MLE.)

(b) (*6pts*) The cdf of Z_n is

$$\begin{aligned} F_{Z_n}(z) &= P(Z_n \le z) = P(n(\theta - Y_n) \le z) = P\left(\theta - \frac{z}{n} \le Y_n\right) = 1 - P\left(Y_n < \theta - \frac{\epsilon}{n}\right) \\ &= 1 - F_{Y_n}\left(\theta - \frac{z}{n}\right) = 1 - \left(\frac{\theta - z/n}{\theta}\right)^n \\ &= 1 - \left(1 + \frac{(-z/\theta)}{n}\right)^n \end{aligned}$$

for $0 < z < n\theta$. Because

$$F_{Z_n}(z) = 1 - \left(1 + \frac{(-z/\theta)}{n}\right)^n \longrightarrow 1 - e^{-z/\theta}, \quad \text{as } n \to \infty,$$

for any $z \in (0, \infty)$, and $1 - e^{-z/\theta}$ is the cdf of the Exponential distribution $E(1/\theta)$, it is proved that Z_n converge in distribution to Z.
(Note. Compare the result with the asymptotic normality property of MLE. Can you see how different they are?)

4. (a) (*6pts*) Because the joint pdf is:

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} \left[(\theta+1)x_i^\theta\right] = (\theta+1)^n \left(\prod_{i=1}^{n} x_i\right)^\theta,$$

the log-likelihhod function is:

$$l(\theta) = \log f(x_1, \ldots, x_n|\theta) = n\log(\theta+1) + \theta\sum_{i=1}^{n}\log(x_i).$$

By setting

$$l'(\theta) = \frac{n}{\theta+1} + \sum_{i=1}^{n}\log(x_i) = 0,$$

we can get the solution is

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n}\log(x_i)} - 1.$$

Because

$$l''(\theta) = -n(\theta+1)^{-2} < 0, \quad \text{for any } \theta,$$

$\hat{\theta}$ is the MLE.

(e) (*6pts*) The Fisher information contained in $X_1, \ldots, X_n$ is

$$E(-l''(\theta)) = E\left[\frac{n}{(\theta+1)^2}\right] = \frac{n}{(\theta+1)^2}.$$

Notice that the Fisher information contained in *a single observation* X_i is

$$I(\theta) = \frac{1}{(\theta+1)^2}.$$

Therefore, the asymptotic variance of the MLE is

$$\frac{1}{E(-l''(\theta))} = \frac{1}{nI(\theta)} = \frac{(\theta+1)^2}{n},$$

and the asymptotic distribution of the MLE is Normal distribution with mean θ and variance $\frac{(\theta+1)^2}{n}$.

5. (a) (*4pts*) Because $\mu_1 = \mathrm{E}(Y_1) = \theta + \frac{1}{2}$, the moment estimator is $\hat{\theta}_1 = \overline{Y}_n - \frac{1}{2}$. $\hat{\theta}_1$ is unbiased because

$$\mathrm{E}(\hat{\theta}_1) = \mathrm{E}(\overline{Y}_n - \frac{1}{2}) = \mathrm{E}(\overline{Y}_n) - \frac{1}{2} = \left(\theta + \frac{1}{2}\right) - \frac{1}{2} = \theta.$$

(b) (*2pts*)

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}(\overline{Y}_n - \frac{1}{2}) = \mathrm{Var}(\overline{Y}_n) = \frac{1}{n}\mathrm{Var}(Y_1) = \frac{1}{n}\mathrm{Var}(Y_1 - \theta) = \frac{1}{12n},$$

because $Y_1 - \theta \sim U(0,1)$. Hence, the standard error of $\hat{\theta}_1$ is $\frac{1}{\sqrt{12n}}$.

(c) (*6pts*) From the hint (i)(ii) and Thm 2.8 in LN, Ch1-6, p.36, the pdf of $T_{(n)}$ is

$$f(t) = nt^{n-1}, \quad 0 < t < 1. \tag{I}$$

Therefore,

$$\mathrm{E}(T_{(n)}) = \int_0^1 t \cdot nt^{n-1} dt = \frac{n}{n+1}. \tag{II}$$

Hence, $\mathrm{E}(\hat{\theta}_2) = \mathrm{E}(Y_{(n)} - \theta) + \theta - \frac{n}{n+1} = \mathrm{E}(T_{(n)}) + \theta - \frac{n}{n+1} = \theta$.

(d) (*4pts*) From equation (I),

$$\mathrm{E}(T_{(n)}^2) = \int_0^1 t^2 \cdot nt^{n-1} dt = \frac{n}{n+2} \tag{III}$$

From hint (ii) and equations (II) and (III),

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}(Y_{(n)}) = \mathrm{Var}(T_{(n)}) = \mathrm{E}(T_{(n)}^2) - (\mathrm{E}(T_{(n)}))^2 = \frac{n}{(n+1)^2(n+2)}.$$

Therefore, the standard error of $\hat{\theta}_2$ is $\sqrt{\frac{n}{(n+1)^2(n+2)}}$.

(e) (*2pts*) The relative efficiency is:

$$\text{eff}_n(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} = \frac{12n^2}{(n+1)^2(n+2)}.$$

The asymptotic relative efficiency is $\lim_{n\to\infty} \text{eff}_n(\hat{\theta}_1, \hat{\theta}_2) = 0$.

(f) (*4pts*) Because $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased,

$$MSE(\hat{\theta}_i) = \text{Var}(\hat{\theta}_i),$$

for $i = 1, 2$. We know that $\text{eff}_n(\hat{\theta}_1, \hat{\theta}_2) < 1$ for $n > 7$, which implies $\text{Var}(\hat{\theta}_1) > \text{Var}(\hat{\theta}_2)$ when the sample size is greater than 7. Therefore, $\hat{\theta}_2$ has a smaller MSE and is better.

6. (a) (*5pts*) The joint pdf of $X_1, \ldots, X_r$ is:

$$\begin{aligned} f(x_1, \ldots, x_r|p) &= \prod_{i=1}^{r} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \\ &= \left[\prod_{i=1}^{r} \binom{n}{x_i}\right] \left[p^{\sum_{i=1}^{r} x_i}\right] \left[(1-p)^{nr-\sum_{i=1}^{r} x_i}\right] \\ &= \left[\left(\frac{p}{1-p}\right)^{\sum_{i=1}^{r} x_i}\right] \left[(1-p)^{nr}\right] \left[\prod_{i=1}^{r} \binom{n}{x_i}\right] \\ &= \left\{\exp\left[\left(\sum_{i=1}^{r} x_i\right) \cdot \log\left(\frac{p}{1-p}\right)\right]\right\} \left[(1-p)^{nr}\right] \left[\prod_{i=1}^{r} \binom{n}{x_i}\right]. \end{aligned}$$

This is an *exponential family*, which implies $T = \sum_{i=1}^{r} X_i$ is sufficient and complete.

(b) (*3pts*)

$$E(U) = 1 \cdot P(U=1) = P(X_1 = 0) = \binom{n}{0} p^0(1-p)^n = (1-p)^n = \theta.$$

(c) (*5pts*) For $t \le n(r-1)$,

$$\begin{aligned} P(U=1|T=t) &= \frac{P(X_1 = 0, T = \sum_{i=1}^{r} X_i = t)}{P(T=t)} = \frac{P(X_1 = 0, \sum_{i=2}^{r} X_i = t)}{P(T=t)} \\ &= \frac{P(X_1 = 0)P(\sum_{i=2}^{r} X_i = t)}{P(T=t)} \\ &= \frac{\left[\binom{n}{0} p^0(1-p)^n\right]\left[\binom{n(r-1)}{t} p^t(1-p)^{n(r-1)-t}\right]}{\binom{nr}{t} p^t(1-p)^{nr-t}} \\ &= \frac{\binom{n}{0}\binom{n(r-1)}{t}}{\binom{nr}{t}} = \frac{\binom{n(r-1)}{t}}{\binom{nr}{t}} \end{aligned}$$

(note that from the Hint, $\sum_{i=2}^{r} X_i \sim B(n(r-1), p)$ and $T = \sum_{i=1}^{r} X_i \sim B(nr, p)$). When $t > n(r-1)$, it is impossible that X_1 can be zero. Therefore $P(U = 1|T = t) = 0$.

(d) (*3pts*) Because T is sufficient and complete for θ (note that $\theta = (1-p)^n$ is a one-to-one transformation) and U is an unbiased estimator of θ, $E(U|T)$ is a UMVUE. The UMVUE is:

$$E(U|T) = 1 \cdot P(U = 1|T) = \begin{cases} \dfrac{\binom{n(r-1)}{T}}{\binom{nr}{T}}, & \text{if } T \leq n(r-1) \\ 0, & \text{if } T > n(r-1) \end{cases}.$$