Solution to Assignment #10

1. **(Rao-Blackwell Theorem)** Let \( X_1, \ldots, X_n \) be an independent sample from a Bernoulli(\( \theta \)) distribution, so that \( P(X_i = 1) = \theta \) and \( P(X_i = 0) = 1 - \theta \).

   (a) Find a minimal sufficient statistic \( U \) for \( \theta \).
   
   Solution: For two samples \( s_1 = (x_1, \ldots, x_n) \) and \( s_2 = (y_1, \ldots, y_n) \), the likelihood ratio simplifies to
   \[
   \left( \frac{\theta}{1 - \theta} \right)^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}
   \]
   which does not depend on \( \theta \) if and only if \( \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \), so either \( \bar{x} \) or \( \sum_{i=1}^{n} x_i \) could be a minimal sufficient statistic. Here we use \( U = \sum_{i=1}^{n} x_i \).

   (b) Find the maximum likelihood estimator \( \hat{\theta} \) for \( \theta \).
   
   Solution: The derivative of the log-likelihood set to zero is the equation
   \[
   \frac{\partial \log L(\theta | s)}{\partial \theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n - \sum_{i=1}^{n} x_i}{1 - \theta} = 0
   \]
   which has solution \( \hat{\theta} = \bar{x} \). (The second derivative is \( -\sum_{i=1}^{n} x_i / \theta^2 - (n - \sum_{i=1}^{n} x_i) / (1 - \theta)^2 \) < 0 when \( \sum_{i=1}^{n} x_i \) is not 0 or \( n \). In these extremes, the likelihood is monotonic with a maximum at 0 or 1, respectively, so \( \bar{x} \) is a maximum.)

   (c) Suppose a statistician wanted to use \( T(s) = (X_1 + X_2)/2 \) for an estimator. Using the sufficient statistic \( U \) you found in part (a), find the Rao-Blackwell estimator \( T_U = E(T | U) \).
   
   Solution: First, by the linearity of expectations and exchangeability of the random variables,
   \[
   E(T_U | U) = (1/2)(E(X_1 | U) + E(X_2 | U)) = E(X_1 | U).
   \]
   Now conditional on \( U \), the sample contains exactly \( U \) ones and \( n - U \) zeros, so the conditional expected value of \( X_1 \) given this information is \( U/n = \bar{X} \) regardless the value of \( \theta \). Thus, \( T_U = \bar{X} \), the maximum likelihood estimator in this case. Note that \( T \) is the mean of the first two observations and \( T_U \) is the mean of all \( n \) observations.

   (d) Compute the exact MSE for \( T \) and for \( T_U \) and verify (assuming \( n > 2 \)) the claim of the Rao-Blackwell theorem in this example.
   
   Solution: Both \( T \) and \( T_U \) are unbiased, so the MSEs are the respective variances. Also, recall from the Bernoulli distribution that \( \text{Var}(X_i) = \theta(1 - \theta) \). \( \text{MSE}(T) = \theta(1 - \theta)/2 \) and \( \text{MSE}(T_U) = \theta(1 - \theta)/n \). The ratio \( \text{MSE}(T)/\text{MSE}(T_U) = n/2, \) so whenever \( n > 2, \text{MSE}(T) > \text{MSE}(T_U) \), which is consistent with the Rao-Blackwell theorem.

2. **(Cramer-Rao Lower Bound)** Let \( X_1, \ldots, X_n \) be an independent sample from a Geometric(\( 1/\theta \)) distribution with probability function \( p(x) = (1/\theta)(1 - 1/\theta)^x \) for \( x = 0, 1, 2, \ldots \) where \( \theta > 1 \).

   (a) Find a minimal sufficient statistic for \( \theta \).
   
   Solution: For two samples \( s_1 = (x_1, \ldots, x_n) \) and \( s_2 = (y_1, \ldots, y_n) \), the likelihood ratio simplifies to
   \[
   (1 - (1/\theta))^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}
   \]
   which does not depend on \( \theta \) if and only if \( \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \), so either \( \bar{x} \) or \( \sum_{i=1}^{n} x_i \) could be a minimal sufficient statistic.
(b) Find the maximum likelihood estimator $\hat{\theta}$.

Solution: The derivative of the log-likelihood set to zero is the equation

$$\frac{\partial \log L(\theta | s)}{\partial \theta} = -n + \sum_{i=1}^{n} \frac{x_i}{\theta} + \sum_{i=1}^{n} x_i \frac{1}{\theta - 1} = 0$$

which has solution $\hat{\theta} = 1 + \bar{x}$. The second derivative is

$$\frac{n + \sum_{i=1}^{n} x_i}{\theta^2} - \frac{\sum_{i=1}^{n} x_i}{(\theta - 1)^2}$$

which simplifies to

$$\frac{n(1 + \bar{x})}{(1 + \bar{x})^2} - \frac{n \bar{x}}{\bar{x}^2} = -\frac{n}{\bar{x}(1 + \bar{x})} < 0$$

at $\theta = 1 + \bar{x}$ (when this is positive). When $\bar{x} = 0$, the maximum is achieved at $\theta = 0$. So, $\hat{\theta} = \bar{x} + 1$ is the MLE.

(c) Find the Fisher information $I(\theta)$ for one observation and $nI(\theta)$ for a sample.

Solution: Compute $I(\theta) = -E(\partial^2 \log f_\theta(X)/\partial \theta^2)$. Note that $\log f_\theta(X) = -(X + 1) \log \theta + X \log(\theta - 1)$. Recall that for a Geometric($\gamma$) distribution, the mean is $(1 - \gamma)/\gamma$, so in this reparameterization, $E(X) = (1 - 1/\theta)/(1/\theta) = \theta - 1$.

$$I(\theta) = -E(\partial^2 \log f_\theta(X)/\partial \theta^2) = -\frac{E(X) + 1}{\theta^2} + \frac{E(X)}{(\theta - 1)^2} = -\frac{1}{\theta} + \frac{1}{\theta - 1} = \frac{1}{\theta(\theta - 1)}$$

The Fisher Information for a sample is $n/(I(\theta)(\theta - 1))$.

(d) Is the MLE an unbiased estimator?

Solution: Yes.

$$E(\hat{\theta}) = E(1 + \bar{X}) = 1 + E(X) = 1 + \theta - 1 = \theta$$

(e) What does the Cramer-Rao Lower Bound imply about variance of the MLE?

By the Cramer-Rao lower bound, $\text{Var}(\hat{\theta}) \geq (nI(\theta))^{-1} = \theta(\theta - 1)/n$. Since $\text{Var}(1 + \bar{X}) = \text{Var}(\bar{X}) = \text{Var}(X)/n$, and the variance of a Geometric($1/\theta$) distribution is $(1 - 1/\theta)/(1/\theta)^2 = \theta^2 - \theta = \theta(\theta - 1)$. Therefore $\text{Var}(\hat{\theta}) = \theta(\theta - 1)/n = (nI(\theta))^{-1}$, so among all unbiased estimators, the MLE achieves the lowest possible variance.

3. (Likelihood Ratio Tests)

We observe 250 random variables, each of which takes on a value from 0, 1, 2, 3, 4. The table of observations is

<table>
<thead>
<tr>
<th>$x$</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>103</td>
</tr>
<tr>
<td>1</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
so 103 of the 250 random variables were observed to be zero, and so on. Consider these two models: (1) $X_i \sim \text{Binomial}(4, \theta)$; (2) $X_i \sim \text{Multinomial}(5)$.

Find the MLE for each model and conduct a likelihood ratio test for the binomial model versus the multinomial model. State hypotheses, calculate the value of the test statistic, compare the value of this statistic to a reference distribution, and compute a p-value.

Solution: Let $p_k = P(X_i = k)$. The null hypothesis of the binomial distribution is formally this:

$$H_0: p_k = \binom{4}{k} \theta^k (1-\theta)^{4-k} \text{ for some } \theta \in (0, 1), \ k = 0, 1, \ldots, 4$$

versus the alternative hypothesis

$$H_A: p_k \geq 0, \sum_{k=0}^{4} p_k = 1$$

The likelihood ratio statistic is $-2 \log \Lambda$ where $\Lambda$ is the ratio of the likelihood maximized under $H_0$ over the likelihood maximized over the entire parameter space.

Under the null hypothesis, the likelihood is

$$f_\theta(s) = \prod_{k=0}^{4} \left( \binom{4}{k} \theta^k (1-\theta)^{4-k} \right)^{x_k}$$

and the log-likelihood is

$$\log f_\theta(s) = \sum_{k=0}^{4} x_k \left( \log \binom{4}{k} + k \log \theta + (4-k) \log(1-\theta) \right)$$

which has derivative

$$\frac{\partial \log f_\theta(s)}{\partial \theta} = \sum_{k=0}^{4} x_k \left( \frac{k}{\theta} - \frac{4-k}{1-\theta} \right) = 0$$

which has solution $\sum_{k=0}^{4} k x_k / (4n)$. We do the calculations in R.

```r
> k = 0:4
> x = c(103, 73, 45, 25, 4)
> n = sum(x)
> theta.hat = sum(k * x)/(4 * n)
> print(theta.hat)
[1] 0.254
```

The estimated binomial probabilities are

```r
> p0 = dbinom(0:4, 4, theta.hat)
> print(p0)
[1] 0.309710058 0.421803511 0.215425118 0.048898999 0.004162314
```

The maximum log-likelihood under the null hypothesis is
> logl0 = sum(x * log(p0))
> print(logl0)


Under the alternative hypothesis, the maximum probability estimates are the proportions in each category.

> p1 = x/n
> print(p1)

[1] 0.412 0.292 0.180 0.100 0.016

> logl1 = sum(x * log(p1))
> print(logl1)

[1] -332.4677

The test statistic is then calculated to be

> r = -2 * (logl0 - logl1)
> print(r)

[1] 35.46667

and the p-value, is

> 1 - pchisq(r, 4 - 1)

[1] 9.70791e-08

as the full model has 4 free parameters and the null model has but one.

Since this p-value is so small, there is overwhelming evidence against the null hypothesis that the binomial model is appropriate for this data.