Please begin every answer to a numbered part of this question on a new sheet of paper.

Suppose X_1, X_2, X_3 are iid Bernoulli(p) random variables for $p \in [0,1]$ and write $\mathbf{X} = (X_1, X_2, X_3)$. Define the statistics

$$T(\mathbf{X}) = X_1 + X_2 + X_3 \quad \text{and} \quad S(\mathbf{X}) = \begin{cases} 0 & \text{if } T(\mathbf{X}) = 0 \text{ or } T(\mathbf{X}) = 1 \\ T(\mathbf{X}) & \text{otherwise} \end{cases}.$$

1. Completely specify the conditional distributions of $T \mid S = s$ for $s = 0, 2,$ and 3. (Give possible values and corresponding conditional probabilities for each of the three distributions.)

2. Completely specify the conditional distributions of $\mathbf{X} \mid S = s$ for $s = 0, 2,$ and 3. (Note that $\mathbf{X}$ takes values in $\{0,1\}^3$.)

3. Based on your answer to **2.**, argue carefully that the statistic $S(\mathbf{X})$ is *not* sufficient for the parameter p.

4. Evaluate the Fisher Information in S about the parameter p. (Show that this is a ratio of polynomials in p, but you need NOT simplify. In fact, this is less than the Fisher Information in $\mathbf{X}$ about p.)

5. Suppose that only $S(\mathbf{X})$ (and not $\mathbf{X}$ or $T(\mathbf{X})$) is available for use in inference about p. Find the maximum likelihood estimator of p based on $S(\mathbf{X})$.

6. If *a priori* $p \sim \mathrm{U}(0,1)$ and $S = 0$, write out a ratio of definite integrals that could be evaluated to obtain the posterior (conditional) probability that $p < 0.5$. (You need NOT actually evaluate this ratio, but produce an explicit form that gives the correct number.)

Now suppose that $S_1, S_2, S_3, \ldots, S_n$ are iid with marginal pmf f specified below.

s	0	2	3
$f(s)$	$(1-p)^3 + 3(1-p)^2 p$	$3(1-p)p^2$	p^3

(Note that $(1-p)^3 + 3(1-p)^2 p$ simplifies slightly to $(1-p)^2(1+2p)$.)

7. Define the sample mean $\bar{S}_n = \frac{1}{n}\sum_{i=1}^{n} S_i$. For $p = .05$ what is an approximate distribution for $(\bar{S}_{100})^2$? Argue carefully that your approximation is correct.

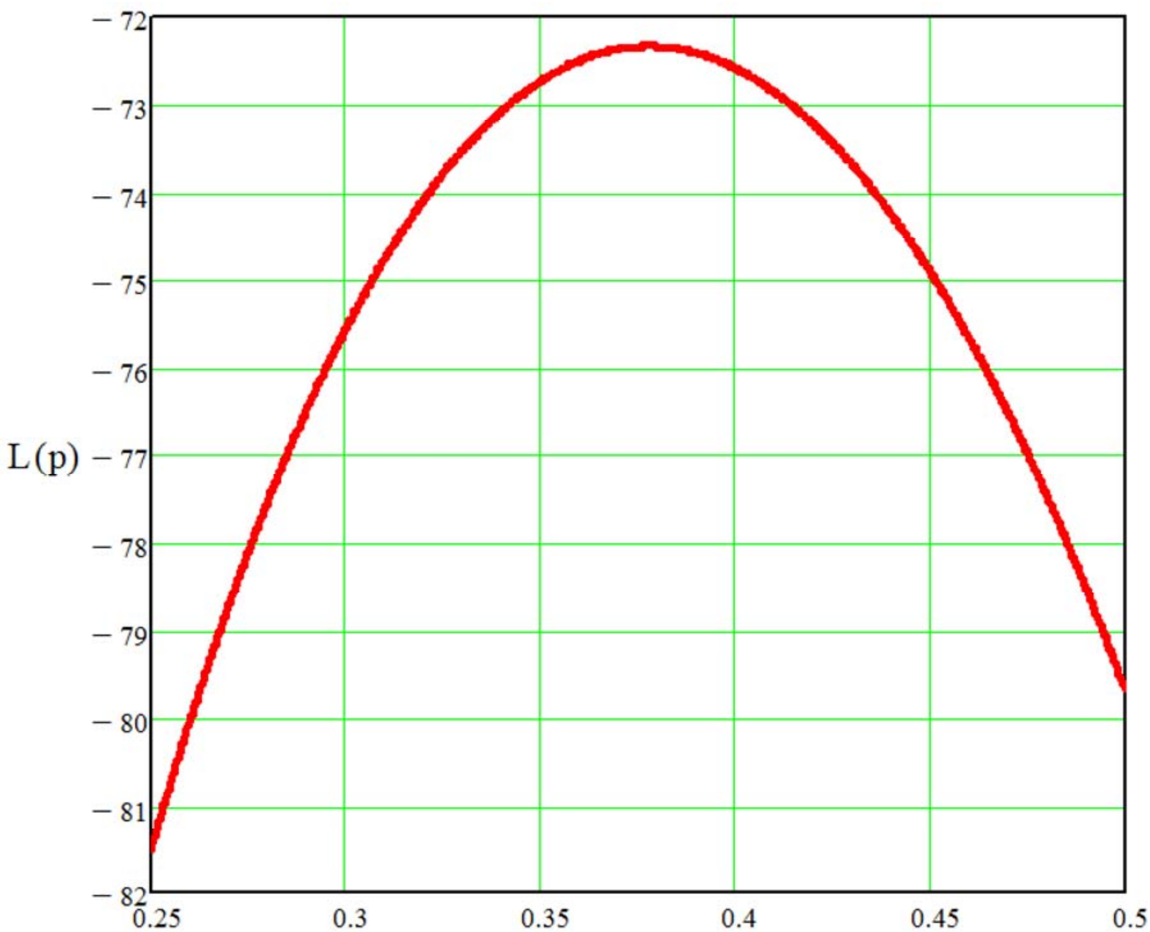
8. Let

n_0 = the number of S_i taking the value 0
n_2 = the number of S_i taking the value 2
n_3 = the number of S_i taking the value 3

Give a formula for the log-likelihood function based on the S_i, $L(p)$.

A sample of size $n = 100$ produces $n_0 = 64, n_2 = 29,$ and $n_3 = 7$ and a log-likelihood that is plotted below. Further, some numerical analysis can be done to show that

$$L(0.378) \approx -72.331, L'(0.378) \approx 0 \text{ and } L''(0.378) \approx -1011$$



9. Give an approximate p-value for testing $\mathrm{H}_0{:}p = 0.5$ vs $\mathrm{H}_\mathrm{a}{:}p \neq 0.5$ based on the information above.

10. Give "Wald" and "LRT" two-sided approximate 90% confidence limits for p based on the information above.