11th problem set

This will not be turned in, but some of its topics will be covered on the final.

1. A null hypothesis $H_0 : \theta = \theta_0$ will be rejected in favor of the alternative hypothesis $H_1 : \theta \neq \theta_0$ if and only if a certain statistic R is too big. The cumulative probability distribution function of R is continuous regardless of the value of θ. If $\theta_1 \neq \theta_0$ then $\mathbf{Pr}(R \text{ is too big} \mid \theta = \theta_1) > \mathbf{Pr}(R \text{ is too big} \mid \theta = \theta_0)$; in other words, this test is more likely to reject a false null hypothesis than a true one.

   Given the observation that $R = r$, the "p-value" is $\mathbf{Pr}(R \geq r \mid \theta = \theta_0)$.

   (a) Show that the p-value is a statistic.

   (b) If you are testing at the 4% level, for what values of the p-value would you reject H_0?

   (c) Find the probability distribution of the p-value given that $\theta = \theta_0$.

   (d) Show that the p-value has a higher probability of being small if $\theta =$ anything besides θ_0 than if $\theta = \theta_0$ regardless of how small is considered "small."

2. In #3(c)(ii) on the 10th problem set, the null hypothesis says $b = c$. The null hypothesis does <u>not</u> assert independence of rows and columns, unlike the null hypothesis in #3(c)(i), which <u>does</u> assert independence. Do the test, and show that the result is significant (i.e., the null hypothesis would be rejected) at the 10% level.

3. Suppose $X_1, X_2, X_3 \sim$ i. i. d. and $f_{X_1}(x) = \alpha x^{\alpha-1}$ for $0 \leq x \leq 1$ and $f_{X_1}(x) = 0$ if either $x > 1$ or $x < 0$. It is observed that $X_1 = 0.8$, $X_2 = 0.8$, and $X_3 = 0.9$.

   (a) Estimate α by the method of moments.

   (b) Estimate α by the method of maximum likelihood.

4. Suppose $X \sim \text{Poisson}(\lambda)$. It is observed that $3 \leq X \leq 5$.

   (a) Find the likelihood function.

   (b) Take 4 to be a first approximation to the maximum-likelihood estimate of λ. Use Newton's method to find the second approximation.

CONTINUED⟶

5. Suppose you want to estimate α and β in the logistic regression model

$$\text{logit}\,\mathbf{Pr}(Y_i = 1) = \alpha(x_i - \beta)$$

given these data:

x	Y
200	0
220	1
222	0

(a) Explain why 221 would be a reasonable first approximation to the maximum-likelihood estimate of β.

(b) Would the maximum-likelihood estimate of α be positive or negative? Explain.

6. In class (Session #34) we developed a method of estimating the "accident-rate" of insurance customers. It was asserted that that method is not reliable unless the sample-size is large. Suppose the numbers of accidents suffered by five customers in one unit of time are as follows:

$$\begin{aligned} X_1 &= 3 \\ X_2 &= 4 \\ X_3 &= 2 \\ X_4 &= 1 \\ X_5 &= 2 \end{aligned}$$

Estimate each customer's "accident-rate" and comment on the assertion made in class.

7. Use the Kolmogorov-Smirnov test to test the null hypothesis that the following are independent observations from an exponential distribution with expectation 1.

1.13
0.19
1.48
0.28
0.14
1.70
0.44
0.62