Ground rules: You must work alone on this quiz. Do not collaborate with anyone, either within or outside the class, to obtain answers or even hints. Questions of clarification should be directed to the instructor; in addition, the use of the Internet should be avoided. This quiz contains 5 questions, each of equal weight, making the quiz worth 50 points.

1. The distribution of starting salaries (in \$1000s) of USC undergraduates majoring in statistics is well modeled by a Pareto distribution. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of n recent graduates, modeled as arising from a Pareto pdf of the form

$$f_Y(y;\theta) = \begin{cases} \theta\nu^\theta y^{-(\theta+1)}, & y > \nu \\ 0, & \text{otherwise.} \end{cases}$$

where $\theta > 1$. The parameter ν represents the minimum starting salary; we will assume $\nu = 35$ (known). Find the form of the most powerful level α rejection region to test $H_0 : \theta = 2$ versus $H_a : \theta = 3$. Explain, in as much detail as possible, how to find the critical value for the test; you don't have to express the critical value in terms of a well known probability distribution. Just tell me how one would find it.

2. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from $f_Y(y;\theta)$, where

$$f_Y(y;\theta) = \begin{cases} \theta^2 y e^{-\theta y}, & y > 0,\ \theta > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Derive the UMP level α test for $H_0 : \theta = \theta_0$ versus $H_a : \theta < \theta_0$. The rejection region for this test should depend on a χ^2 quantile.
(b) Plot the power function of the test, $K(\theta)$, when $n = 15$, $\alpha = 0.01$ and $\theta_0 = 10.4$. Try to use R to plot the power function.

3. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a beta distribution with parameters $\alpha = 1$ and $\beta = \theta$.
(a) Derive the level α likelihood ratio test of $H_0 : \theta = 1$ versus $H_a : \theta \neq 1$. Explicitly state how all critical values are chosen. Don't confuse the Type I Error probability α here with the "α" parameter in the beta model.
(b) For **extra credit**, derive the power function $K(\theta)$ and plot it when $n = 10$ and the Type I Error probability is 0.10. If you can not do this, you can still take a guess of what you think the power function would look like.

4. Suppose that we have two **independent** samples:

$$\begin{aligned} &\text{Sample 1}: \quad Y_{11}, Y_{12}, ..., Y_{1n} \sim \text{iid Poisson}(\theta_1) \\ &\text{Sample 2}: \quad Y_{21}, Y_{22}, ..., Y_{2n} \sim \text{iid Poisson}(\theta_2), \end{aligned}$$

and that we are interested in comparing the population means θ_1 and θ_2.
(a) Show that the loglikelihood function of $\boldsymbol{\theta} \equiv (\theta_1, \theta_2)'$ is given by

$$\ln L(\boldsymbol{\theta}|\boldsymbol{y}_1, \boldsymbol{y}_2) \equiv \ln L(\theta_1, \theta_2|\boldsymbol{y}_1, \boldsymbol{y}_2) = \sum_{j=1}^{n} y_{1j} \ln\theta_1 + \sum_{j=1}^{n} y_{2j} \ln\theta_2 - n(\theta_1 + \theta_2) + c,$$

where c is a constant that depends neither on θ_1 nor θ_2, and $\boldsymbol{y}_i = (y_{i1}, y_{i2}, ..., y_{in})'$, for $i = 1, 2$.

(b) Derive the form of the rejection region for the level α LRT of $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 \neq \theta_2$. Here, the parameter space $\Omega = \{\boldsymbol{\theta} : \theta_1 > 0,\ \theta_2 > 0\}$ and the null space is $\Omega_0 = \{\boldsymbol{\theta} : \theta_1 > 0,\ \theta_2 > 0,\ \theta_1 = \theta_2\}$. Note that when H_0 is true, we can envision $Y_{11}, Y_{12}, ..., Y_{1n}, Y_{21}, Y_{22}, ..., Y_{2n}$ as an iid sample of size $2n$ from a Poisson distribution with mean θ, say, where $\theta = \theta_1 = \theta_2$; that is, finding the MLE over the restricted space requires only a univariate maximisation procedure (over Ω, it requires a two-variate procedure since there are two free parameters). State explicitly how all rejection region critical values are obtained. Do not use the two-sample test based on the large-sample distribution of $\overline{Y}_{1+} - \overline{Y}_{2+}$, as in Section 1.3 of the course notes. Do not use a test based on the t distribution either.

(c) Use the large-sample approximation to the likelihood ratio statistic λ (actually to $-2\ln\lambda$) to test $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 \neq \theta_2$ with the following data:

Left:	0	1	1	2	2	1	1	4	0	3
Right:	0	1	4	3	7	2	0	1	3	3

All theoretical regularity conditions are satisfied for the approximation to hold. What is your conclusion?

Note: These data represent the number of weeds growing in square-foot plots of ground in my left and right "wooded" areas in my back yard. That is, I have two wooded areas in my back yard; one on the left side of the yard, and one on the right side. Ten plots (i.e., their locations) were randomly selected from each side. The counts correspond to the number of weeds per plot. Assume that the Poisson model holds (it's likely reasonable). Do the left and right sides of my back yard have different weed levels? (I know the answer; the question is whether or not the data support it).

5. Oil-drilling technology is improving every day; however, finding productive wells among prospective sites is not an exact science. In region i, let p denote the probability of finding a productive well, and let Y_i denote the number of sites in the region that are drilled to find the **first** productive well in that region. A large excavation study is conducted in n regions. Conditional on p, where $0 < p < 1$, we will model the data $Y_1, Y_2, ..., Y_n$ as an iid sample from a geometric distribution with pmf

$$f_Y(y; p) = \begin{cases} p(1-p)^{y-1}, & y = 1, 2, ... \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, because of region-to-region variability, we assume that p is best regarded as a random variable with a beta(α, β) prior distribution, where α and β are both constants larger than zero.

(a) Geologists know that p is not large; in fact, they have suggested that p is close to 0.10. Pick a beta prior that seems reasonable in this situation (that is, give me appropriate values of α and β).

(b) Turn the Bayesian crank (i.e., go through the 5 steps in the notes) to find the posterior distribution of p, given the data $\boldsymbol{Y} = \boldsymbol{y}$.
(c) Suppose that $n = 10$ and that $\sum_{i=1}^{10} y_i = 134$. For your values of α and β in part (a), graph the posterior distribution $g(p|\boldsymbol{y})$ and find $E(p|\boldsymbol{Y} = \boldsymbol{y})$, the posterior mean of p.
(d) For **extra credit**, derive the Jeffreys noninformative prior for the geometric parameter p and reanalyse the data, assuming that $n = 10$ and that $\sum_{i=1}^{10} y_i = 134$. Graph the the posterior distribution $g(p|\boldsymbol{y})$ and find $E(p|\boldsymbol{Y} = \boldsymbol{y})$, the posterior mean of p, for this prior choice. Are your results from parts (c) and (d) largely different?