Study Questions #4 for STA 3000

These are for study only, not to hand in for credit

Q1: Consider a decision problem with data space $\mathcal{X} = \{1, 2\}$, parameter space $\Omega = \{\theta_0, \theta_1\}$, and action space $\aleph = \{a_0, a_1\}$. Suppose that the loss function is

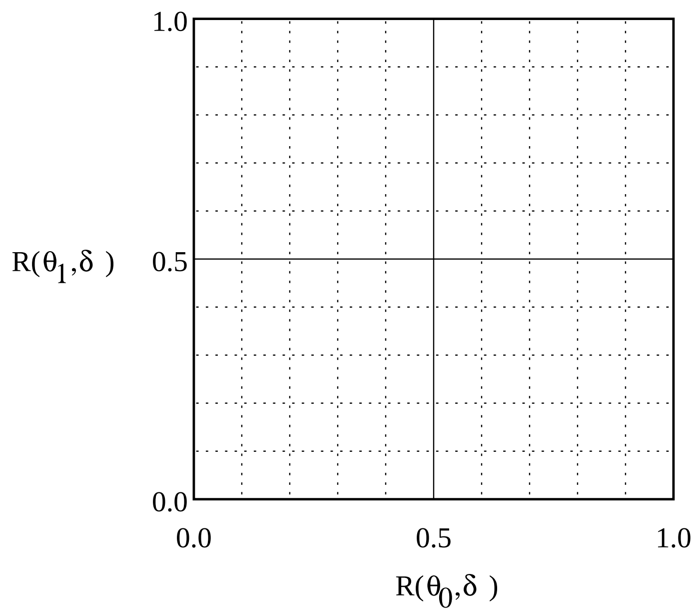
$$L(\theta, a) = \begin{cases} 0.2 & \text{if } \theta = \theta_0 \text{ and } a = a_0 \\ 0.8 & \text{if } \theta = \theta_0 \text{ and } a = a_1 \\ 0.7 & \text{if } \theta = \theta_1 \text{ and } a = a_0 \\ 0.4 & \text{if } \theta = \theta_1 \text{ and } a = a_1 \end{cases}$$

and that the data distributions are given by

$$P_{\theta_0}(x) = \begin{cases} 0.5 & \text{if } x = 1 \\ 0.5 & \text{if } x = 2 \end{cases}$$

$$P_{\theta_1}(x) = \begin{cases} 0.0 & \text{if } x = 1 \\ 1.0 & \text{if } x = 2 \end{cases}$$

1. Draw the risk set for this decision problem in the diagram below:



2. Find a minimax rule for this decision problem. (Write an explicit formula for the decision rule.) Is this the unique minimax rule? Is it admissible? Explain your answers (geometric explanations are OK).

3. Find a Bayes rule for this decision problem when the prior gives equal probability to θ_0 and θ_1. (Write an explicit formula for the decision rule.) Is this Bayes rule unique? Is it admissible? Explain.

4. Are all Bayes rules for this decision problem (for all possible priors) admissible? Explain.

Q2: Suppose a positive integer X is observed from a distribution of the following form:

$$P_\theta(X = x) \;=\; \begin{cases} 3/4 & \text{if } x = \theta \\ 1/4 & \text{if } x = \theta + 1 \\ 0 & \text{otherwise} \end{cases}$$

where the parameter θ is a positive integer (ie, $\Omega = \{1, 2, 3, \ldots\}$). We wish to find a decision rule that estimates θ with $0-1$ loss — ie,

$$L(\theta, \delta(x)) \;=\; \begin{cases} 0 & \text{if } \delta(x) = \theta \\ 1 & \text{if } \delta(x) \neq \theta \end{cases}$$

(1) Find the risk functions for each of the estimators $\delta_1(x) = x$, $\delta_2(x) = x + 1$, and $\delta_3(x) = x - 1$.

(2) Find a Bayes rule for each of the following prior distributions:

(a) $P(\Theta = \theta) \;=\; 2^{-\theta}$
(b) $P(\Theta = \theta) \;=\; 3 \cdot 4^{-\theta}$
(c) $P(\Theta = \theta) \;=\; 2 \cdot 3^{-\theta}$

Determine whether or not each of these rules is admissible.

(3) Which of the rules $\delta_1(x) = x$, $\delta_2(x) = x + 1$, and $\delta_3(x) = x - 1$ are admissible?

(4) Find a minimax rule for this decision problem.

Q3: Consider the null hypothesis that n observed points, $(x_1, y_1), \ldots, (x_n, y_n)$, are independently drawn from a distribution that is uniform over some convex region of the (x, y) plane. Devise a pure significance test of this hypothesis — ie, a way of computing a p-value from the data for which the p-value will be uniformly distributed over $(0, 1)$ if the null hypothesis is true. Design your test so that it will be sensitive to departures from the null hypothesis in which the data is uniformly distributed over some non-convex region, or non-uniformly distributed over some convex region. You needn't worry about departures from the assumption of independence.

If you can, think of more than one pure significance test, and compare their properties, such as whether or not they are invariant to translating, rotating, and re-scaling the data (including scaling x and y differently). Also, discuss what one would have to do to address this problem using Bayesian decision theory (with the action being to declare that the null hypothesis is either true or false, and with the loss being 0 if this declaration is correct, and 1 if it is incorrect). How easy or hard do you think using Bayesian decision theory would be?

Don't worry about computational issues for this problem. It is OK if your significance test doesn't reduce to some simple formula, and it is OK if the only way you can think of computing the p-value by computer is inefficient. (There must be *some* way of computing the p-value, however, even if it is inefficient.) I don't expect you to be able to derive quantitative properties of the tests analytically.

Hint: Find a sufficient statistic for the model assumed by the null hypothesis, and then use the method described by Cox and Hinkley in Section 3.3, of basing a test statistic on the conditional distribution of the full data given the sufficient statistic, which does not depend on the unknown parameter (in this case, the convex region).