Homework 4

1. ***Using conditional probability for estimation***

Suppose $x : \Omega \to U$ and $w : \Omega \to V$ are independent random variables with pmfs p^x and p^y. Define the random variable y by

$$y = x + w$$

We would like to measure y and hence determine x, without measuring w. We have

$$V = \{1, 2, 3, 4, 5\} \qquad \text{and} \qquad U = \{1, 2, 3, 4, 5, 6\}$$

and

$$p^w = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.2 \\ 0.2 \end{bmatrix} \qquad \text{and} \qquad p^x = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.15 \\ 0.15 \\ 0.25 \\ 0.25 \end{bmatrix}$$

(a) Show that

$$\text{Prob}(x = a \text{ and } y = b) = p^x(a) p^w(b - a)$$

where $p^w(b - a)$ is understood to be zero for $b - a < 1$ or $b - a > 5$.

(b) Find the induced pmf of y.

(c) Now suppose we measure $y_{\text{meas}} = 7$. Find the conditional pmf q where

$$q(a) = \text{Prob}(x = a \mid y = y_{\text{meas}})$$

Plot q against U.

(d) Now simulate the random variables w and x, and perform $m = 10000$ trials. Keep data for x and w, and for each trial, compute $y = w + x$.

Compute the observed frequencies of y, and plot these. Compare with the pmf for y.

(e) Now look at the same data set, and write a loop that discards those data items for which $y \neq y_{\text{meas}}$. On the remaining data, compute the observed frequencies of x. Plot the observed frequencies of x, and compare with the conditional pmf of x given y_{meas}.

(f) Now you have to pick one estimate of x based on your measurement. Which is the estimate that minimizes your probability of error?

2. ***Radar***

A radar has conditional probability mass functions

$$p^{\text{nothing}} = 10^{-2} \begin{bmatrix} 4 & 7 & 11 & 12 & 17 & 16 & 13 & 9 & 6 & 4 & 1 & 0 & 0 \end{bmatrix}$$
$$p^{\text{something}} = 10^{-2} \begin{bmatrix} 0 & 0 & 2 & 6 & 8 & 10 & 15 & 15 & 15 & 13 & 9 & 5 & 2 \end{bmatrix}$$

Here X_1 is the event that nothing is visible, and X_2 is the event that something is visible, and Y_i is the event that the radar returns i pings (at least one ping is always returned.) Then

$$\text{Prob}(Y_i \mid X_1) = p_i^{\text{nothing}} \qquad \text{Prob}(Y_i \mid X_2) = p_i^{\text{something}}$$

The prior probabilities are

$$\text{Prob}(X_1) = 0.6 \qquad \text{Prob}(X_2) = 0.4;$$

(a) Find the classifier with the minimum probability of error, and the corresponding probability of a correct classification.

(b) Suppose we significantly prefer false positives to false negatives, and so decided to minimize

$$J_1 + \mu J_2$$

where

$$J_1 = \text{Prob(false positive)} \qquad J_2 = \text{Prob(false negative)}$$

First, let's arbitrarily choose $\mu = 3$.

Find the optimal decision rule to minimize the weighted sum cost. For this rule, find the probability of a false positive error, a false negative error, and the probability of a correct result.

(c) To do this more systematically, we need to plot the *operating characteristic*. As usual when plotting trade-off curves, it's a good idea to choose a wide range of values for μ. Pick a suitable set of values for μ, and for each compute the classifier K which minimizes $J_1 + \mu J_2$, as well as the optimal values of J_1 and J_2. Plot J_1 against J_2. As you can see, it's not a very good radar.

(d) Now we would like to find the classifier which minimizes the probability of false positives, subject to the constraint that the probability of false negatives is less than 0.1. First find the smallest μ you can use that achieves this level of probability of false negatives, then find the corresponding optimal classifier. Again, for this rule, find the probability of a false positive error, a false negative error, and the probability of a correct result.

(e) In fact there are only a finite number of possible classifiers? How many are there? (Don't only count the threshold rules, count all of them.)

(f) Write a routine that for each possible decision rule, computes the corresponding probability of false positives and false negatives. Plot another copy of the operating characteristic, and add to it all of these new points.

(g) The plot you obtained in part (f) has a symmetry. State precisely what this symmetry is. Can you explain how it arises?

(h) What is the *worst* possible classifier you could use; i.e., the one that achieves the maximum possible probability of error?

(i) Now among *all* possible decision rules, find the one that achieves the smallest possible probability of false positives, subject to the constraint that the probability of false negatives is less than 0.1. What is the decision rule, the probability of a false positive error, a false negative error, and the probability of a correct result. Can you explain why this decision rule was not found by the approach in part (d).

(j) Let's also explore what would happened if the radar were better, i.e., the conditional pmfs were further apart. Suppose

$$\text{Prob}(Y_i \mid X_2) = p^{\text{something}}_{i-k}$$

i.e, the conditional probabilities given something is visible are shifted to the right by k positions. For each $k \in \{0, 1, 2, \ldots, 5\}$ plot the operating characteristic, (all on one plot).