Foundations of Statistics I: Decision Theory
Problem Set 5

Problem 1 *10 points Consider a population of patients, some of which have a genetic marker $g = 1$ and some of which don't. An assay gives you a measurement z that depends on the marker. The assay is has cdf $F_g(z)$. If we set a cutoff point z_0 on the assay and declare "positive" all individuals that for whom $z > z_0$, this declaration will lead to true positives but also some false positives. A Receiver Operating Characteristics curve is a graph of the true positive fraction (sensitivity) on the vertical scale versus the false positive fraction (one minus specificity) on the horizontal scale as we vary z_0. Let p_1 and p_2 be the horizontal and vertical coordinates of the ROC curve, respectively. Formally the ROC curve R is given by the equation $p_2 = R(p_1) = 1 - F_1[F_0^{-1}(1 - p_1)]$. Here $1 - p_1$ is specificity and p_2 is sensitivity.*

Show that the area under the ROC curve is equal to the probability that a randomly selected person with the genetic marker has an assay that is greater than a randomly selected person without the marker.

Problem 2 *10 points A forecaster announces probabilities π for a sequence of binary events (1 = rain, 0 not). The frequency of forecast π is $\nu(\pi)$. Among the events that are assigned probability π the frequency of actual successes is $\bar{x}(\pi)$*

Show that

$$BS = \sum_{\pi \in \Pi} \nu(\pi) \left[\bar{x}(\pi)(\pi - 1)^2 + (1 - \bar{x}(\pi))\pi^2 \right].$$

can be rewritten as

$$BS = \sum_{\pi \in \Pi} \nu(\pi) \left[\pi - \bar{x}(\pi) \right]^2 + \sum_{\pi \in \Pi} \nu(\pi) \left[\bar{x}(\pi)(1 - \bar{x}(\pi)) \right].$$

Suppose the sequence of events is independent with probability of success .4 Evaluate the terms in this expression for the following four forecasters:

Charles: *always says .4*
Mary: *randomly chooses between .3 and .5*
Qing: *either says .2 or .3; when he says .2 it never rains, when he sais .3 it always rains.*
Ana: *follows this table*

	rain	*no rain*
$\pi = .3$	*.15*	*.35*
$\pi = .5$	*.25*	*.25*

comment on the calibration and refinement of these forecasters.