Bayesian Statistics

Exercises 5

1. Mr. Rubin has determined that his utility function for a change in fortune on the interval $r \in [-100, 500]$ is

$$U(r) = 0.62 \log(0.004r + 1).$$

   a) He is offered a choice between \$100 and the chance to participate in a gamble wherein he wins \$0 with probability 2/3 and \$500 with probability 1/3. Which should he choose?

   b) Suppose he is offered instead a chance to *pay* \$100 to participate in the gamble. Should he do it?

2. Suppose your final exam is a multiple choice exam, and you have to assign, for each answer A, B, C and D a probability that it is correct. That is the "action" is a probability vector (q_A, q_B, q_C, q_D) of non-negative numbers summing to 1. The unknown quantity θ that affects your response is which answer is actually correct. That is $\Theta = \{A, B, C, D\}$.

   Consider the following three loss functions:

   i) (Brier score) $L(\theta; q) = \sum_{i \in \Theta}(q_i - I(\theta = i))^2$

   ii) ("log loss") $L(\theta; q) = -\log(q_\theta)$.

   iii) $L(\theta; q) = 1 - q_\theta$.

   Suppose that after reading the question and answers you have decided that answers A to D have, for you, (posterior) probabilities (p_A, p_B, p_C, p_D). Then under loss functions i) and ii) your posterior expected loss would be minimized by reporting $q = p$, that is, they are "proper scoring rules". However, under loss function iii) your posterior expected loss would be minimized by a q other than p.

   Exercise: show that i), and ii) are proper scoring rules, but iii) is not. Find the optimal q (as a function of p) for iii)

3. Consider two forecasters, who are predicting whether it will rain tomorrow. Each gives their probability that it will rain tomorrow, and these predictions are later compared against the actual observation of whether or not it rains. In a particular 10-day period the probabilities given by each forecaster are:

```
1: 0.59 0.79 0.54 0.36 0.77 0.62 0.37 0.18 0.39 0.14
2: 0.32 0.57 0.60 0.53 0.85 0.21 0.46 0.07 0.45 0.29
```

   and the actual observations were

```
1 1 0 0 1 1 1 0 1 0
```

   where 1 indicates rain, and 0 indicates no rain. Based on these data, who do you assess to be the better forecaster, and why? Provide analyses for both the Brier score loss function and the log-loss function.

4. (continuation of 3. above) Now suppose that you were forecaster 1 in the above example, and you were invited to participate in a forecasting competition, where they are going to score your forecasts using the Loss $L_1(i, p) = 1 - p_i$, where p_i is the probability that you assign to outcome i ($i = 0, 1$). a) For the 10-day period considered in 3 above, what would you have *reported* to the competition as your probabilities for rain on each day, in order to minimize your expected loss L_1? b) For these reported probabilities what would your average loss L_1 have been for this 10-day period? Compare this with the loss you would have achieved if you had instead reported your actual probabilities from 3. above.

5. Suppose x is Poisson with mean θ and that it is desired to estimate θ under the loss $L(\theta, d) = (\theta - d)^2/\theta$.

   What is the (frequentist) risk function $R(\theta, \delta)$ for the estimator $\delta_0(x) = x$?

   Now suppose that we have a Gamma(α, β) prior on θ, with $\alpha > 1$. Find the Bayes estimator of θ (that is, the estimator that minimizes the integrated risk).

6. Let X be Poisson with mean $\theta \in \Theta$, where $\Theta = (0, \infty)$ and the action space $\mathcal{A} = [\ 0\ , \infty)$. The loss function $L(\theta, a) = (\theta - a)^2$. Consider decision rules of the form $\delta_c(x) = cx$. Assume $\pi(\theta) = e^{-\theta}$ is the prior density.

   a) Find the risk function $R(\theta, \delta_c)$.
   b) Show that δ_c is inadmissible if $c > 1$.
   c) Find the integrated risk $r(\pi, \delta_c)$.
   d) Find the value of c that minimizes $r(\pi, \delta_c)$.