Probabilistic Modelling and Reasoning
Tutorial Sheet 5

School of Informatics, University of Edinburgh

Instructor: Dr Chris Williams

October 2007

1. [Bishop qu 2.7]. Consider the Bernoulli random variable X with mean θ, and suppose we have observed n_h occurrences of $X = 1$ and n_t occurrences of $X = 0$. The prior distribution for $\theta \sim \text{Beta}(\alpha_h, \alpha_t)$. Show that the posterior mean value of θ lies between the prior mean and the maximum likelihood estimate of θ. To do this, show that the posterior mean can be written as λ times the prior mean plus $(1 - \lambda)$ times the maximum likelihood estimate, for $0 \leq \lambda \leq 1$. HINT: to make this easier, verify that the required value of $\lambda = \alpha/(n + \alpha)$, where $\alpha = \alpha_h + \alpha_t$ and $n = n_h + n_t$. This illustrates the concept of the posterior distrubution being a compromise between the prior distribution and the maximum likelihood solution.

2. On Thursday the weather forecast for Saturday indicates a 60% chance of rain, and you are organizing an outdoor concert. The losses are as follows

$$\text{L(go ahead, fair)} = -1 \qquad \text{L(go ahead, rain)} = 2$$

$$\text{L(cancel, fair)} = 3 \qquad \text{L(cancel, rain)} = 0$$

Calculate the minimum risk strategy. Should you cancel the concert?

3. You have a machine that measures property x, the "orangeness" of liquids. You wish to discriminate between $\mathcal{C}_1$ = "IrnBru" and $\mathcal{C}_2$ = "Orangina". It is known that

$$p(x|\mathcal{C}_1) = \begin{cases} 10 & 1.0 \leq x \leq 1.1 \\ 0 & \text{otherwise} \end{cases}$$

$$p(x|\mathcal{C}_2) = \begin{cases} 200(x-1) & 1.0 \leq x \leq 1.1 \\ 0 & \text{otherwise} \end{cases}$$

The prior probabilities $P(\mathcal{C}_1) = 0.6$ and $P(\mathcal{C}_2) = 0.4$ are also known from experience. Calculate the optimal Bayes' classifier and $P(\text{error})$.

4. **(Bonus question)**. Suppose that instead of using the Bayes' decision rule to choose class k if $P(\mathcal{C}_k|\mathbf{x}) > P(\mathcal{C}_j|\mathbf{x})$ for all $j \neq k$, we use a randomized decision rule, choosing class j with probability $Q(\mathcal{C}_j|\mathbf{x})$. Calculate the error for this decision rule, and show that the error is minimized by using Bayes' decision rule.

5. **(Bonus question)**. Consider the beta distribution $p(\theta) = c(\alpha, \beta)\theta^{\alpha-1}(1-\theta)^{\beta-1}$, where $c(\alpha, \beta)$ is a normalizing constant. The mean of this distribution is $E[\theta] = \alpha/(\alpha + \beta)$. For

$\alpha, \beta > 1$ the distribution is unimodal (i.e. it has only one maximum). Find the value θ^* where this maximum is attained, and compare it to the mean. For what values of α and β to the mean and θ^* coincide?