Foundations of Statistics I: Decision Theory
Problem Set 1.

Problem 1 *[10 points] courtesy of K. Laskey at George mason U.*

A. Assume that the probability that a woman of age 40 in a given population has breast cancer (the prevalence of breast cancer) is 0.5%. The probability that the disease is detected by a mammogram (the sensitivity of the mammogram) is 80%. The probability that a woman without breast cancer will have a negative mammogram (the specificity of the mammogram) is 92%. When asked for the probability of breast cancer in a woman of age 40 from the general population who has a positive mammogram, Dr. Smith estimated a probability of about 80%. Would you agree with Dr. Smith? Why or why not?

B. Consider a pregnancy test with the same sensitivity and specificity as the mammogram. However, assume that the probability of being pregnant in a woman who takes the test is 20%. What is the posterior probability that a woman testing positive for pregnancy is pregnant? Explain the difference between this and the previous problem.

C. In the pregnancy problem, what are the parameter, the parameter space, the sample space, the data, the likelihood, and the prior?

Problem 2 *[20 points]*

Consider a sample $x = (x_1, x_2, \ldots, x_n)$ *where* $x_i|\lambda \sim Pois(\lambda)$*, and assume that prior information about* λ *allows us to specify the prior* $\pi(\lambda) = e^{-\lambda}$.

1. *Derive an expression for the posterior distribution, that is the distribution of* λ *given* x
2. *Derive an expression for the mean of this distribution.*

3. *Write a computer program to perform the following tasks:*
   - *Simulate a sample* $(x_1, x_2, \ldots, x_5)$ *by simulating a value of lambda from the prior and, conditional on that, simulating the vector* $(x_1, x_2, \ldots, x_5)$*;*
   - *graph the prior and the sequence of posterior distributions* $p(\lambda|x_1)$, $p(\lambda|x_1, x_2)$, ... $p(\lambda|x_1, ..., x_5)$ *on the same plot.*
4. *Consider now the alternative prior*

$$\pi^*(\lambda) = pe^{-\lambda} + (1-p)ae^{-a\lambda}$$

   where $a = .01$ *and* $p = .9$*, and derive the form of the posterior distribution using this prior. Show that it is a mixture of two gamma distributions and derive the mixture weights.*
5. *Assume the sample* $x = (1.1, 1.3, .45, .6, 15.3)$ *was observed and graph, on the same plot, the posterior distributions obtained using both priors* π *and* π^*.

		States of nature		
		θ_1	θ_2	θ_3
	a_1	£100	£110	£120
Actions	a_2	£90	£100	£120
	a_3	£120	£110	£100

Table 1 A simple decision problem with three possible actions, four possible outcomes, expressed as cash flows in pounds, and three possible states of the world.

Please write your own code and turn in both the code and the graphs. R would be perfect for this. You can use existing random number generators for known distributions.

Problem 3 *[10 points] Consider the actions described in Table 1 in which the consequences are monetary payoffs. Convert this problem into one of choosing between lotteries, as defined in the von Neumann – Morgenstern theory. The decision maker holds the following indifferences with reference lotteries:*

£100 $\sim$ £120 w.p 1/2; £90 w.p 1/2;
£110 $\sim$ £120 w.p 4/5; £90 w.p 1/5;

Assume that $\pi(\theta_1) = \pi(\theta_3) = 1/4$ *and* $\pi(\theta_2) = 1/2$. *Which is the optimal action according to the expected utility principle?*

Problem 4 (10 pts) *Suppose that there are* $k \geq 2$ *horses in a race and that a gambler believes that* π_i *is the probability that horse* i *will win* ($\sum_{i=1}^{k} \pi_i = 1$). *Suppose that the gambler has decided to wager an amount* a *to be divided among this* k *horses. If he or she wagers* a_i *on horse* i *and that horse wins, the utility of the gambler is* $\log(c_i a_i)$, *where* $c_1, \cdots, c_k$ *are known positive numbers. Find values* $a_1, \cdots, a_k$ *to maximize expected utility.*