STA 114: Statistics

HW 8

Due Wed Nov 02 2011

1. To determine whether high protein diets offer higher gains in body weight in infant mammals, data were collected on 19 female rats, 12 of which were given a certain high protein diet, while the other 7 received a regular diet. For each rat, weight gain between 28th and 84th days after birth were recorded. Let $X_1, \cdots, X_n$ denote these measurements for the high-protein group, and $Y_1, \cdots, Y_m$ denote the same for the low-protein group. Consider modeling $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_1, \sigma_1^2)$, $Y_j \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_2, \sigma_2^2)$, X_i's and Y_j's are independent.

   (a) Assuming equal variances for the two groups, calculate 90%, 95%, 99% ML confidence intervals for $\eta = \mu_1 - \mu_2$, the expected additional weight gain due to the high protein diet, when observed data are as follows:

Diet	Observations
High	134 146 104 119 124 161 107 83 113 129 97 123
Low	70 118 101 85 107 132 94

   (b) Repeat part (a) but without assuming equal variances.

2. We want to compare Duke students' weekly expenditure on food between years 2010 and 2011. We have data $X = (X_1, \cdots, X_n)$ from 2010 on n students, and data $Y = (Y_1, \cdots, Y_m)$ from 2011 on another m students. Consider the model $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_1, \sigma^2)$, $Y_j \sim \mathsf{Normal}(\mu_2, \sigma^2)$, with $\xi(\mu_1, \mu_2, \sigma^2) = \mathsf{NN}\chi^{-2}(m_1, k_1, m_2, k_2, r, s)$. An expert's quantified beliefs about food expenditure are matched by $m_1 = m_2 = 125$, $k_1 = k_2 = 0.3$, $r = 1$ and $s = 53$ (i.e., among other things, the expert believes that the expenditure distribution has not changed between 2010 and 2011). The observed data are

Year	Expendtures
2010	139 210 75 131 184 140 140 127 100 129 145 145 120 237
2011	125 140 200 200 190 100 140 250 125 180 110 125 120 130 140 150 120 100 95 195 95 130

   (a) Give a central 95% posterior credible interval for $\eta = \mu_1 - \mu_2$.

   (b) Evaluate the posterior probability $\eta < 0$.

   (c) Based on these answers what are your conclusions on the changes in food expendture between 2010 and 2011?

3. For question 2, suppose our interest was instead on the difference $D^* = X^* - Y^*$ where $X^* = X_{n+1}$ and $Y^* = Y_{m+1}$, respectively, are expenditures reported by hypothetical students randomly sampled from the 2010 and 2011 populations. A natural choice of model is

$X_1, \cdots, X_n, X_{n+1} \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_1, \sigma^2)$, $Y_1, \cdots, Y_m, Y_{m+1} \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_2, \sigma^2)$. Suppose we assign (μ_1, μ_2, σ^2) a prior pmf $\xi(\mu_1, \mu_2, \sigma^2)$, such that $\xi(\mu_1, \mu_2, \sigma^2) = 1/18$ if $\mu_1 = 100, 125, 150$, $\mu_2 = 100, 125, 150$, $\sigma^2 = 25, 100$ and $\xi(\mu, \sigma^2) = 0$ otherwise [i.e., it is a discrete uniform prior over the set of 18 elements formed by the above choices for μ_1, μ_2 and σ^2].

(a) Calculate the posterior predictive probability of $D^* < 0$. [Hint: It might help to fill in the empty cells of the following table.]

(b) Calculate the same under the plug-in predictive distribution based on the MLE.

μ_1	μ_2	σ^2	$\xi(\mu_1, \mu_2, \sigma^2)$	$L_{x,y}(\mu_1, \mu_2, \sigma^2)$	$\xi(\mu_1, \mu_2, \sigma^2 \mid x, y)$	$P(D^* < 0 \mid \mu_1, \mu_2, \sigma^2)$
100	100	25	1/18			
100	100	100	1/18			
100	125	25	1/18			
100	125	100	1/18			
100	150	25	1/18			
100	150	100	1/18			
125	100	25	1/18			
125	100	100	1/18			
125	125	25	1/18			
125	125	100	1/18			
125	150	25	1/18			
125	150	100	1/18			
150	100	25	1/18			
150	100	100	1/18			
150	125	25	1/18			
150	125	100	1/18			
150	150	25	1/18			
150	150	100	1/18			

4. Again consider the situation of question 3, but now assume that (μ_1, μ_2, σ^2) is assigned the prior of question 2. Calculate the posterior predictive probability of $D^* < 0$.