EXERCISE ON BAYESIAN BASICS

(1) The following series z has been generated by a mechanism which made z independent across observations, with the probability of $z = 1$ equal to p_1 over the first part of the sample and p_2 over the second part of the sample. The probability of zero was $1 - p_i$ in each case. We don't know what p_1 and p_2 are, and we give them a joint uniform prior on the unit square. If n is the point in the sample (counting from left to right) after which the probability switches, plot the posterior pdf of n given this sample. You will probably want to use the computer to make the plot. Integrating out the parameters p_1 and p_2 does not have to be done numerically, since the integrals required have the form of the integral defining the Beta function — i.e. an integrand of the form $q^{\theta-1}(1-q)^{\phi-1}$. Matlab, Octave, R, etc. have functions that compute Beta.

$$z = [0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0]$$

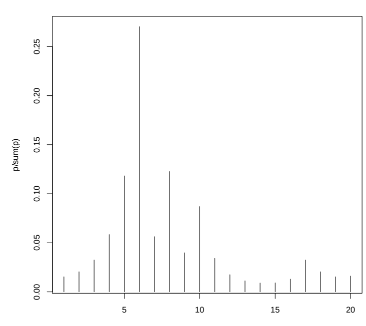
The likelihood as a function of n (the last observation in the first part of the sample), $n_j(n), j = 1, 2$ (the number of 1's in the p_j part of the sample), and $p_j, j = 1, 2$ is

$$p(z \mid n, p_1, p_2) = \prod_{t=1}^{n} p_1^{z_t}(1-p_1)^{1-z_t} \prod_{t=n+1}^{20-n} p_2^{z_t}(1-p_2)^{1-z_t}$$
$$= p_1^{n_1(n)}(1-p_1)^{n-n_1(n)} p_2^{n_2}(1-p_2)^{20-n-n_2(n)}$$

With a flat prior on p_i and a uniform prior on n, the posterior is proportional to the likelihood. Integrating the likelihood over p_1 and p_2 for a given n gives us

$$\text{Beta}(n_1+1, n-n_1+1)\,\text{Beta}(n_2+1, 20-n-n_2+1)\,.$$

The posterior is then the above function of n, normalized to integrate to one. Here is a plot of it:



(2) We have 10 i.i.d. draws of a random variable X that is a truncated version of an underlying variable X^*. X^* is distributed as $N(\mu, \sigma^2)$, and $X = \min(X^*, 2)$. In other words, any values of X^* above 2 are recorded as 2 in forming X.

Date: December 12, 2010.

(a) For each of samples (i) and (ii) below, plot the contours of the likelihood for μ and σ^2.

(i) 1.4, 1.0, 1.5, .76, 1.6, .72, .55, 1.2, .51, .68

(ii) .3, 2,2,2,2,2,2,2,2,2

Since we displayed the plots in class, I don't reproduce them here. For sample (i) the likelihood is symmetric about the sample mean as a function of μ. For sample (ii) it quite asymmetric, declining slowly as μ increases above 2. The likelihood function is

$$\prod_{t=1}^{T} \phi(x_t;\mu,\sigma^2)^{z_t}(1-\Phi(2;\mu,\sigma^2))^{1-z_t}\,,$$

where z_t is an indicator variable that is one for $x_t < 2$, zero otherwise, ϕ is the normal density function, and Φ is the normal cdf.

(b) Is the maximum likelihood estimate (MLE) for μ biased in sample (ii) and unbiased in sample (i)? Why or why not?

Bias is a frequentist property of an estimator. An estimator is biased or not before one sees the sample data, and seeing the sample data (and thus the likelihood function) cannot change whether the estimator is unbiased or not. A sample consisting entirely of 2's is always possible, no matter what the true values of μ and σ^2. But in such a sample, the likelihood is just $(1-\Phi(\mu,\sigma^2))^N$, where N is sample size. This approaches 1 as $\mu \to \infty$ for every σ^2, so the likelihood maximum is at $\mu = \infty$ with non-zero probability. The expectation of the MLE therefore does not exist, so it is certainly not unbiased.

(c) If you had quadratic loss for errors in estimating μ and a nearly flat prior, would it make sense to use the MLE in sample (i) but not in sample (ii)? Why or why not?

With quadratic loss (i.e. $E[(\mu-\hat{\mu})^2]$, the best choice for $\hat{\mu}$ would be the posterior mean of μ. (This is the principle of certainty equivalence for quadratic loss: The optimal decision can be found by replacing unknown quantities by their expectations and solving the resulting deterministic optimization.) In the first sample the likelihood is symmetric about $\hat{\mu}$, so the posterior mean with a nearly flat prior will be nearly the mean of the normalized likelihood, which is just the sample mean in this case. In the second sample, the likelihood is quite asymmetric, so the mean of the posterior will not coincide with the MLE.

Drawing contours in R:

```
x <- seq(1,5, by=.1)
y <- seq(-3,3, by=.1)
z <- outer(x,y, function(a,b) exp(-(a-3)^2 - (b^2)))
contour(x,y,z)
```

There are similar facilities in Octave and Matlab, though possibly in those languages the construction of the z matrix with `outer` would have to be replaced with loops over x and y.