

**Stat 543 Assignment 1 (due Monday January 25, 2016)**  
**Likelihood Functions, Bayes Paradigm, and Simple Decision Theory**

1. In this problem we'll consider some issues concerning the defining and meaning of a "likelihood function." Underlying what follows, suppose that  $X_1, X_2, \dots, X_{10}$  are iid exponential with mean  $\theta$ . That is, we'll suppose that the marginal pdf is

$$f(x|\theta) = I[x \geq 0] \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$$

Here are two data sets (pseudo-randomly) generated from this model with  $\theta = 2$ .

Data Set #1 : 2.661, 3.764, 0.079, 1.286, 0.081, 0.233, 2.283, 2.600, 0.321, 0.421

Data Set #2 : 2.426, 1.852, 0.803, 11.524, 0.055, 1.266, 0.506, 0.448, 0.749, 1.930

- (a) Plot, on the same set of axes, the natural logarithms of the likelihood functions (as defined in class for jointly continuous observations) based on Data Sets #1 and #2. (It is often convenient to work on a log-likelihood scale, especially in models with independent observations.) Are these two functions exactly the same? Do they provide some reliable indication of how Vardeman generated the data?
- (b) (Type I left censoring) Suppose that what is observable is not  $X_i$ , but rather  $W_i = X_i I[X_i > 1]$ . That is, if an observation is less than 1, I observe a "0" (that I might as well take as a code for "observation was less than 1"). In this circumstance, a sensible definition of a likelihood function would seem to be

$$L_{\text{left}}(\theta) = \prod_{i=1}^{10} \left( I[w_i = 0] \left( 1 - \exp\left(-\frac{1}{\theta}\right) \right) + I[w_i \geq 1] \frac{1}{\theta} \exp\left(-\frac{w_i}{\theta}\right) \right)$$

Plot the logarithm of this likelihood function for Data Set #1. How does it compare to that for Data Set #1 in part (a)? Does it have a similar "vertical position" on the plot? Does it have a similar "shape"?

- (c) (Rounding/digitalization) Using the joint pdf as a likelihood function essentially means that regardless of how many digits are written down, a data value entered to make the likelihood is treated as an "exact" "real number." That is, "2.661" implicitly has an infinite number of trailing 0's behind it. Another (more realistic) point of view is to think of "2.661" as meaning "something between 2.6605 and 2.665." Under this interpretation, a more sensible likelihood function than the one used in part (a) is

$$L_{.0005}(\theta) = \prod_{i=1}^{10} (F(x_i + .0005|\theta) - F(x_i - .0005|\theta))$$

where  $F(x|\theta)$  is the exponential cdf. Plot the logarithm of  $L_{.0005}(\theta)$  for Data Set #1 and compare it to the loglikelihood plotted in part (a). Then do something similar, supposing that Data Set #1 had been recorded and reported to only 2 decimal places, and then to only 1 decimal place. How do these two loglikelihoods compare to  $\ln L_{.0005}(\theta)$  and to the one plotted in part (a)?

2. Suppose that  $X_1, X_2, \dots, X_5$  are iid  $N(\mu, \sigma^2)$ . Below are two data sets (pseudo-randomly) generated from this model with  $\mu = 0$  and  $\sigma = 1$ .

Data Set #1 : 0.041, -0.705, 0.088, -0.103, -1.9203

Data Set #2 : 0.195, 0.551, -0.821, 0.319, 0.457

Make contour plots of the loglikelihood functions for these two data sets (for  $-2 < \mu < 2$  and  $0 < \sigma < 2$ ) (using the definition of likelihood function given in class for jointly continuous observations). If you haven't before made a contour plot in R, see the syntax provided for the last part of Lab #8 at

<http://www.public.iastate.edu/~vardeman/stat401/stat401B.html>

3. There are two possible diagnostic tests for a certain medical condition. Individuals from some population (say infants born at hospital X) are tested for the condition. In 3 separate studies

- 10 of 100 tested with Test A alone had a "positive" test result
- 20 of 100 tested with Test B alone had a "positive" test result
- 50 tested with both Test A and Test B broke down as below

		Test B	
		+	-
Test A	+	4	2
	-	8	36

Think of running both tests on an individual as producing an outcome falling naturally into one of the 4 cells of the table below (with corresponding probabilities indicated).

		Test B	
		+	-
Test A	+	$p_{++}$	$p_{+-}$
	-	$p_{-+}$	$p_{--}$

Since  $p_{--} = 1 - (p_{++} + p_{+-} + p_{-+})$ , we might think of this as a statistical problem with a 3-dimensional parameter vector  $\mathbf{p} = (p_{++}, p_{+-}, p_{-+})$ . Write an appropriate loglikelihood for  $\mathbf{p}$  based on all data collected in the three studies.

4. The following are required

- B&D Problems 1.2.2, 1.2.3, 1.2.14
- B&D Problems 1.3.3, 1.3.9, 1.3.10, 1.3.11, 1.3.12, 1.3.18 (use  $n = 5$ )

5. B&D Problem 1.2.5 is optional, not required but recommended.

LATE ADJUSTMENT: Hold Problem 2, and B&D 1.3.11 and 1.3.12 for Assignment 2.

### Hint for Problem 1.2.14 of B&D

The general story in this problem is this:

For  $X_1, X_2, X_3, \dots, X_n, X_{n+1}$  iid  $f(x|\theta)$  and  $\theta \sim g(\theta)$ , the joint density of the observables and parameter is

$$f(\mathbf{x}, \theta) = g(\theta) \prod_{i=1}^{n+1} f(x_i|\theta)$$

and from this, one can (in theory) get the marginal of the vector of observables  $\mathbf{X}$  by integrating out  $\theta$ , i.e.

$$f_g(\mathbf{x}) = \int g(\theta) \prod_{i=1}^{n+1} f(x_i|\theta) d\theta$$

THEN, from this one can contemplate finding the conditional distribution of  $X_{n+1}|X_1, X_2, \dots, X_n$  that B&D call the predictive distribution (note that you don't get this "for free" in terms of assumptions ... the form of  $f_g$  depends on  $g$ ).

Now actually carrying out the above program is going to be very unpleasant except for "convenient" cases. In fact, the only multivariate distributions that are really "convenient" are the MVN distributions. Thankfully, (and not surprisingly) this problem uses them. That is, think of writing

$$X_i = \theta + \epsilon_i$$

for the  $\epsilon_i$  iid  $N(0, \sigma_0^2)$  independent of  $\theta \sim N(\theta_0, \tau_0^2)$ . Then

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ X_{n+1} \\ \theta \end{pmatrix} \sim \text{MNV}(\theta_0 \mathbf{1}, \Sigma)$$

for an appropriate  $\Sigma$  that you can figure out easily enough, since finding variances and covariances is straightforward. ( $\text{Var} X_i = \sigma_0^2 + \tau_0^2, \text{Cov}(X_i, X_j) = \tau_0^2, \text{Cov}(X_i, \theta) = \tau_0^2$ ). From this, the marginal of  $X_1, X_2, X_3, \dots, X_n, X_{n+1}$  is obvious (also MVN) and you can use facts about conditionals of MVN distributions to get the conditional distribution of  $X_{n+1}|X_1, X_2, \dots, X_n$ .

In doing the above, you will probably need to use a matrix fact about the inverse of a certain patterned matrix. The covariance matrix of  $\mathbf{X}$  (and of any subset of the entries of  $\mathbf{X}$ ) is of the form

$$a\mathbf{I} + b\mathbf{J}$$

for  $\mathbf{J}$  a matrix of 1's. The inverse of such a matrix has the form

$$\alpha\mathbf{I} + \beta\mathbf{J}$$

for  $\alpha = \frac{1}{a}$  and  $\beta = -\frac{b}{a(a+kb)}$ , where  $k$  is the dimension of the (square) matrix  $\mathbf{I}$  (or  $\mathbf{J}$ ).