

# Medical Tests: Part II

July 16, 2010

## **Abstract**

In this chapter we continue the discussion of diagnostic testing by considering instruments that are based on continuous outcomes. We also introduce additional diagnostic information in the form of covariates and additional tests. Our primary goal is to generalize the notion of PVP and PVN to a predictive probability of infection/disease conditional on all available diagnostic information. In addition, we generalize the concepts of sensitivity and specificity to the receiver operating characteristic (ROC) curve, which gives information about all possible sensitivity-specificity pairs over the range of all possible cutoffs for a continuous test for disease. The material presented here are mainly taken from work that is discussed in Choi, Johnson, and Thurmond (2006), Choi et al. (2006), and McInturff et. al (2004). This work and WinBUGS code are also catalogued at the website [www.epi.ucdavis/diagnostictests/](http://www.epi.ucdavis/diagnostictests/).

# 1 Introduction

Some tests, such as commercial pregnancy tests, are clearly dichotomous. But many tests provide numbers on a continuous scale. One approach to evaluating the performance of a continuous diagnostic outcome is to turn it into a dichotomous test. Pick a cutoff value  $c$  and decide that the test is positive if the continuous measurement is larger than  $c$ , or negative if smaller. In this instance, the methods of Chapter 14 apply. The cutoff value is typically chosen to give either good sensitivity or good specificity, or both, if possible. Increasing sensitivity generally decreases specificity, and vice versa. We may still be interested in making inferences about prevalence and other issues such as costs associated with misclassifying  $D$  and  $\bar{D}$  also come into play.

Dichotomizing continuous test scores is clearly an inefficient use of data since information is lost in the process i.e., the magnitudes of the individual test scores are ignored. From dichotomized data the predictive probability of disease is the same for *all*  $T^+$  (or  $T^-$ ) individuals regardless of their actual test scores. Subjects with test scores far exceeding the cutoff (or far below the cutoff) are treated as identical to subjects whose test values are barely above the cutoff (or barely below the cutoff). In fact, we expect the predictive probabilities of  $D$  and  $\bar{D}$  to become more pronounced as one gets farther from the cutoff.

We thus proceed to calculate the predictive probability of infection/disease given the precise diagnostic outcome information, namely without dichotomization. In addition, if multiple diagnostic outcomes are available, they are combined and ultimately a single predictive probability is obtained, thus

converting information to a simpler and hopefully more familiar, probability scale. If an individual has a 99% chance of being infected, it may well be concluded that they are indeed infected, and similarly, if there is only a 1% chance, they would likely be declared non infected. If their probability is 0.5, it would likely not be clear which is the case and more information should be collected for that individual.

The diagram in Figure 1 illustrates this idea. There are four individuals with test scores  $y_1 = 40$ ,  $y_2 = 64$ ,  $y_3 = 66$ , and  $y_4 = 110$ . The cutoff value is  $k = 65$  so that were the data to be dichotomized using  $k$ , individuals 1 and 2 would test negative while individuals 3 and 4 would test positive. Hence the dichotomized data for these four individuals would be  $(T_1, T_2, T_3, T_4) = (-, -, +, +)$ . Therefore, based on the dichotomized data, individuals 1 and 2 are treated as identical even though their test scores differ in magnitude by 24 units, and individuals 3 and 4 are treated as identical even though their test scores differ in magnitude by 44 units. However, individuals 2 and 3 are treated as different in spite of the fact that the magnitudes of their test scores differ by only 2 units.

The standard approach of dichotomizing the data simplifies the statistical modeling and data analysis and leads to easily interpretable parameters (such as sensitivity, specificity, predictive value positive, predictive value negative). This practice also facilitates the usual goal of testing, which is to classify individuals as diseased or not. These are key reasons for the longevity of the common practice of dichotomizing continuous test data.

In addition to diagnosis, we are also interested in ascertaining the quality of a continuous (or ordered categorical) test for infection or disease. One

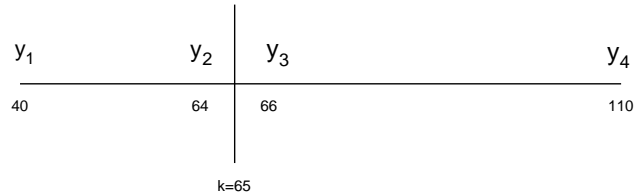


Figure 1: Test results in relation to a cutoff.

way to accomplish this is to estimate its corresponding receiver operating characteristic (ROC) curve. An ROC curve gives a graphical measure of the accuracy of a continuous medical test. Let  $y$  denote the continuous outcome. We assume that large values of  $y$  are associated with having the disease. Let  $\eta(c)$  and  $\theta(c)$  denote the test specificity and sensitivity that would result from using a dichotomous test with cutoff  $c$ . The ROC curve is a plot of the pairs  $((1 - \theta(c)), \eta(c))$  for all values of  $c$ . In other words, an ROC curve is a plot of the false positive probability (horizontal axis) versus the sensitivity (vertical axis) obtained for all  $c$ . For example, in the top panel of Figure 2, for every possible diagnostic score  $c$ , compute the probability that a non-diseased individual is above  $c$  and the probability that a diseased individual is above  $c$ . The ROC curve, given in the bottom panel, is a plot of these

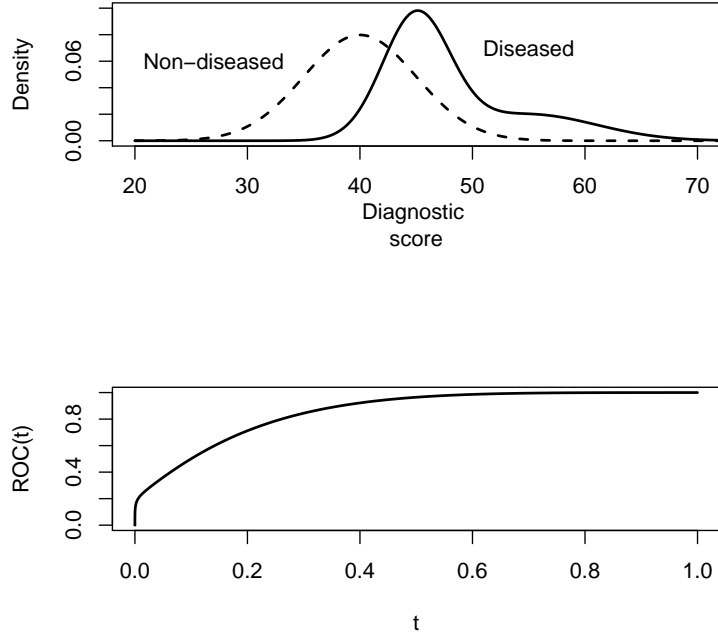


Figure 2: Illustration of distributions of continuous test scores (top panel) and their corresponding ROC curve (bottom panel)

pairs.

We require some notation. Denote the pdf for continuous test outcomes from the population  $\bar{D}$  as  $f_0$ , and denote the corresponding cdf as  $F_0$  and survival function as  $S_0 = 1 - F_0$ . Similarly, denote the pdf for outcomes from the population  $D$  as  $f_1$ , and denote the corresponding cdf as  $F_1$  and survival function as  $S_1 = 1 - F_1$ . The densities  $f_0$  and  $f_1$  are regarded as “truth.” For now, we define inferential objects of interest that are based on them. First, we define the predictive probability of disease/infection as  $\Pr(D|y)$  where  $y$

is a realization from either  $f_0$  or  $f_1$ . In general, we don't know which, but we do know that  $y$  was randomly sampled from a population with prevalence  $\pi$ . Thus, the marginal pdf for  $y$  is simply  $f(y) = (1 - \pi)f_0(y) + \pi f_1(y)$ , using the law of total probability. Then a simple application of Bayes theorem gives

$$\Pr(D|y, \pi) = \frac{\pi f_1(y)}{\pi f_1(y) + (1 - \pi)f_0(y)}.$$

Thus, if the two pdf's and the prevalence were all known, we could easily calculate this probability over a range of values for  $y$ .

*Exercise.* Give the formula for  $\Pr(D|y)$  when  $f_0$  corresponds to a  $N(0, 1)$  pdf and  $f_1$  corresponds to a  $N(1, 2)$  pdf. Furthermore, suppose you decided to conclude that an individual was  $D$  if this probability exceeded 0.95. Then calculate the log odds of  $D$ , simplify algebraically, and find the value  $y_c$  such that the rule  $\Pr(D|y) > 0.95 \Leftrightarrow \text{logit}(\Pr(D|y)) > \text{logit}(0.95) \Leftrightarrow y > y_c$ .

The ROC curve can be computed as

$$\text{ROC}(t) = S_1(S_0^{-1}(t))$$

for  $t \in [0, 1]$ . Pepe (2003, Chapter 4) gives an expanded discussion of ROC curves.

*Exercise.* Show this result. Hint: For a given cutoff  $c$ , let  $1 - \theta(c) = t$ . This implies that  $1 - \theta(c) = S_0(c) = t$ .

A parameter of interest for the ROC function is the area under the curve (AUC). The AUC is the probability that a randomly selected diseased individual has a test score greater than that for a randomly selected non-diseased

individual. In other words,  $\text{AUC} = \Pr(X > Y)$ , where  $X \sim f_1$  and, independently,  $Y \sim f_0$ . AUC can also be interpreted as the average sensitivity across all possible false positive probabilities.

*Exercise.* Show this result. Hint:

$$\Pr(X > Y) = \int_{-\infty}^{\infty} \int_y^{\infty} f_0(y)f_1(x)dx dy = \int S_1(y)f_0(y)dy.$$

then substitute  $y = S_0^{-1}(t)$  and use the previous representation for AUC. Obtain the AUC if  $X \sim N(1, 2)$  and  $Y \sim N(0, 1)$ . Hint: Use the fact that  $X - Y$  also has a known normal distribution.

If the support for diseased outcomes consists of a range of values that are all higher than the support values for non-diseased outcomes, the test can perfectly discriminate between the groups. In other words, there exists a cutoff value  $c$  that gives both perfect sensitivity and specificity. However, even in this case, one can do stupid things. If  $c$  is chosen too low, there will be positive probability of false positive outcomes, and perfect sensitivity. If  $c$  is too high, there will be positive probability of false negative outcomes and perfect specificity. In such cases, the ROC curve is not really defined because there are many sensitivities that go along with zero false negatives. Nonetheless, the ideal ROC curve would be the line  $\text{ROC}(t) = 1$  with corresponding  $\text{AUC} = 1$ . Tests that are completely worthless have measurements that are unrelated to the disease. In that case,  $f_0 = f_1$ , so for any cutoff, the probability of a true positive equals the probability of a false positive. The ROC curve is the line  $\text{ROC}(t) = t$  with corresponding  $\text{AUC} = 0.5$ .

pAUC is a partial area under the curve. For  $0 \leq t_0 < t_1 \leq 1$

$$\text{pAUC} = \int_{t_0}^{t_1} \text{ROC}(t) dt.$$

Often  $t_0 = 0$ , in which case a relatively small value of  $t_1$  emphasizes cutoff values that yield low false positive probability. These may represent the only sensitivity and specificity combinations that are of medical interest.

Another measure of accuracy for continuous tests is the likelihood ratio function,  $LR(y) = f_1(y)/f_0(y)$ . Our assumption that large values of  $y$  are indicative of disease is essentially the same as assuming that the likelihood ratio increases with  $y$ .

## 2 Diagnosis and ROC curves

We consider methods for disease diagnosis and ROC curve estimation based on a single test outcome. We both evaluate test accuracy and develop predictive models. For parametric extensions based on multiple continuous tests, see e.g., Choi et al. (2006). For nonparametric ROC analysis see, e.g., Hanson, Branscum, and Gardner (2008).

As before, we assume densities  $f_1$  and  $f_0$  for the diagnostic outcome data from diseased and non-diseased populations, respectively. However, we now assume that they are members of parametric families  $f_1(\cdot|\theta_1)$  and  $f_0(\cdot|\theta_0)$ . Appropriate choice of the parametric families is problem-specific but a common choice has been normal or log normal for each.

If a gold-standard (GS) is available, it may be the case that training data are available. This means that samples of diseased and non-diseased individuals are tested and diagnostic outcomes obtained. It is then possible



to estimate parameters associated with the two parametric families. However, it is generally our experience that even if training data are available, there are many instances where outcomes are obtained without the followup of a GS test. If the GS test were to always be applied, there would be no need whatsoever to perform additional (imperfect) diagnostic testing. In the absence of GS testing, test outcome data come from a mixture distribution where the prevalence,  $\pi$ , is the mixing proportion. As previously mentioned, we mainly consider the case without a GS test.

We assume independent prior densities for  $\theta_1$ ,  $\theta_0$ ,  $\pi$ , say  $p_1(\theta_1)$ ,  $p_0(\theta_0)$ ,  $p_*(\pi)$ . If we were lucky enough to have training data, they would be modeled (independently) as:

$$y_{01}, \dots, y_{0n_0} | \theta_0 \stackrel{iid}{\sim} f_0(\cdot | \theta_0), \quad \theta_0 \sim p_0(\theta_0),$$

$$y_{11}, \dots, y_{1n_1} | \theta_1 \stackrel{iid}{\sim} f_1(\cdot | \theta_1), \quad \theta_1 \sim p_1(\theta_1).$$

For data without a gold-standard, a random sample is modeled according to the mixture

$$y_1, \dots, y_n | \pi, \theta_1, \theta_0 \stackrel{iid}{\sim} \pi f_1(\cdot | \theta_1) + (1 - \pi) f_0(\cdot | \theta_0)$$

$$p(\theta_0, \theta_1, \pi) = p_0(\theta_0) p_1(\theta_1) p_*(\pi).$$

Constraints are placed on  $\theta_0$  and  $\theta_1$  and an informative prior placed on  $\pi$  to alleviate issues related to nonidentifiability.

Probably the most common parametric model has been the *binormal* model where  $f_0(\cdot | \theta_0) = N(\cdot | \mu_0, \sigma_0^2)$  and  $f_1(\cdot | \theta_1) = N(\cdot | \mu_1, \sigma_1^2)$ , either based on the original outcome data, or some transformation of it (like the log for

example). We note that the binormal model in the frequentist ROC setting refers to a semiparametric model in which the data are assumed to be normal after transformation by an unspecified monotone function.

Clinical diagnosis proceeds by calculating the predictive probability of disease conditional on an individual's (possibly transformed) diagnostic data  $y$ . This is given by

$$\begin{aligned} \Pr(D|y, \text{data}) &= \int \Pr(D|y, \theta_0, \theta_1, \pi) p(\theta_0, \theta_1, \pi | \text{data}) d\theta_0 d\theta_1 d\pi \\ &= \int \frac{f_1(y|\theta_1)\pi}{f_1(y|\theta_1)\pi + f_0(y|\theta_0)(1-\pi)} p(\theta_0, \theta_1, \pi | \text{data}) d\theta_0 d\theta_1 d\pi \end{aligned}$$

where  $\text{data}$  denotes the combined data for non-diseased and diseased individuals. A decision about whether or not an individual is diseased can be based on the magnitude of this predictive probability. However, it is also possible to monitor  $\Pr(D|y, \theta_0, \theta_1, \pi)$  in a Gibbs sampler and to obtain a probability interval for this object, which is the proportion of infected individuals in the population with serology value  $y$ . The posterior mean of this object thus serves two inferential purposes: (i) it's the predictive probability of disease for a particular individual with test outcome  $y$  and (ii) it's an estimate of the proportion of diseased individuals in the population having test outcome  $y$ .

Let  $\Phi(\cdot)$  denote the standard normal cdf. Here, the ROC curve is given by

$$\text{ROC}(t) = \Phi\left(\frac{\mu_1 - \mu_0}{\sigma_1} + \frac{\sigma_0}{\sigma_1} \Phi^{-1}(t)\right)$$

with corresponding AUC given by

$$\text{AUC} = \Phi \left( \frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \right).$$

Bayes estimates of these are their corresponding posterior means.

*Exercise.* Derive both of these results.

EXAMPLE 2.1. *Pancreatic cancer.*

Pepe (2003) illustrates methods for frequentist ROC data analysis using case-control cancer data from Wieand et al. (1989). These are individuals seen at the Mayo Clinic in Rochester, USA. Ninety cases have pancreatic cancer, and 51 controls were patients without cancer but who had chronic pancreatitis. Thus these data are training data and we perform a GS analysis based on them.

One goal was to determine which of two biomarkers, cancer antigen CA-125 and carbohydrate antigen CA-19-9, better discriminates cases from controls. The data were log transformed for analysis so that normal distributions could be used for modeling. The binormal model was used with diffuse independent  $N(0, 100)$  priors for means and  $\Gamma(0.01, 0.01)$  priors for precisions. The WinBUGS code below illustrates how to make inferences for the predictive probability of disease and ROC curve estimation.

Estimated ROC curves and 95% pointwise posterior bands from a binormal analysis are presented in Figure 3. Over a range of cutoff values with sufficiently high specificity (e.g.  $t \leq 0.2$ ), the ROC curve for CA-19-9 dominates the curve for cancer antigen 125, suggesting CA-19-9 has superior classification accuracy relative to CA-125. The posterior median and 95% interval for

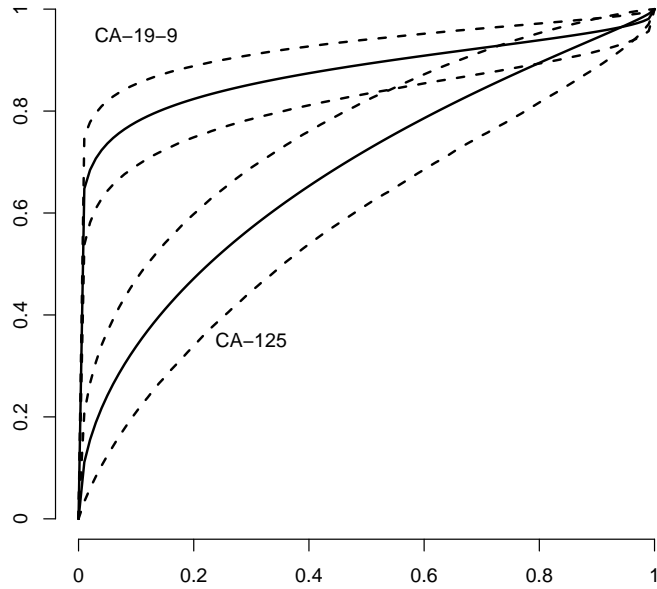


Figure 3: ROC curves (solid lines) with 95% pointwise probability bands for two pancreatic cancer biomarkers.

the AUC of CA-19-9 is 0.879 (0.818, 0.925) and is 0.680 (0.590, 0.760) for the CA-125 serum biomarker.

The WinBUGS code below generates output based on the two tests using the pancreatic data. The two tests are modeled in a way that output would be the same regardless of whether data for each test were analyzed separately. It is simply efficient to get output from a single code. In addition, if the test outcomes could be regarded as independent, conditional on disease status, then the same code could be modified to compare ROC values for the two

tests, for example by simply monitoring  $ROC_{199}(t) - ROC_{125}(t)$ , for any or all given values of  $t$ . If the test outcomes from the two tests were dependent, we would model outcomes as bivariate normal in each of the two populations as in Choi et al (2006). Of course this assumption would require scrutiny.

In the following code, the vector  $x$  used in the formulas for the ROC curves contains the values of  $\Phi^{-1}(t)$  across the grid  $t \in \{0, 0.01, 0.02, \dots, 1\}$ . For  $t = 0, 1$ , we set  $\Phi^{-1}(t)$  equal to  $-100$  and  $100$ , respectively. The following line of R code creates a text file stored on the C drive in the Temp directory (this location can be changed if desired) that contains the vector  $x$  in a format that can be copied and pasted into WinBUGS:

```
dput(c(-100,qnorm(seq(0.01,0.99,0.01)),100),file="C:\\Temp\\x.txt")
```

Alternatively, the user can simply type in the grid of values. Because the data were log transformed, the variable names ty0CA125 and ty0CA199 were used for the transformed CA-125 and CA-19-9 biomarker data for the controls. Similar variable names were used in WinBUGS for the transformed data for the cases.

```
model{ # Model for the outcomes for both tests
for(i in 1:51){
  ty0CA199[i] <- log(y0CA199[i]) # Transform controls
  ty0CA125[i] <- log(y0CA125[i])
  ty0CA199[i] ~ dnorm(mu0199,tau0199) # Log normal model
  ty0CA125[i] ~ dnorm(mu0125,tau0125) # for controls
}
for(i in 1:90){
```

```

ty1CA199[i] <- log(y1CA199[i]) # Transform cases
ty1CA125[i] <- log(y1CA125[i])
ty1CA199[i] ~ dnorm(mu1199,tau1199) # Log normal model
ty1CA125[i] ~ dnorm(mu1125,tau1125) # for cases
} # Diffuse priors
mu0199~dnorm(0,0.01)
tau0199~dgamma(0.01,0.01)
mu0125~dnorm(0,0.01)
tau0125~dgamma(0.01,0.01)
mu1199~dnorm(0,0.01)
tau1199~dgamma(0.01,0.01)
mu1125~dnorm(0,0.01)
tau1125~dgamma(0.01,0.01)
AUC199 <- phi( (mu1199 - mu0199)/sqrt(1/tau0199 + 1/tau1199))
AUC125 <- phi( (mu1125 - mu0125)/sqrt(1/tau0125 + 1/tau1125))
for(i in 1:101){
# Get ROC values over the grid x[] = phi^{-1}(t[])
ROC199[i]<- phi(sqrt(tau1199)*(mu1199-mu0199)+sqrt(tau1199/tau0199)*x[i])
ROC125[i]<- phi(sqrt(tau1125)*(mu1125-mu0125)+sqrt(tau1125/tau0125)*x[i])
}# Predictive probabilities of infection over the grid y[]
for(i in 1:321){
f1199[i] <- sqrt(tau1199)*exp((-tau1199/2)*pow(y[i]-mu1199,2))
f0199[i] <- sqrt(tau0199)*exp((-tau0199/2)*pow(y[i]-mu0199,2))
f1125[i] <- sqrt(tau1125)*exp((-tau1125/2)*pow(y[i]-mu1125,2))
f0125[i] <- sqrt(tau0125)*exp((-tau0125/2)*pow(y[i]-mu0125,2))
}

```

```

    probMale125[i] <- f1125[i]*pm/(f1125[i]*pm + f0125[i]*(1-pm))
    probFemale125[i] <- f1125[i]*pf/(f1125[i]*pf + f0125[i]*(1-pf))
    probMale199[i] <- f1199[i]*pm/(f1199[i]*pm + f0199[i]*(1-pm))
    probFemale199[i] <- f1199[i]*pf/(f1199[i]*pf + f0199[i]*(1-pf))
  }
} # The data and grid points
list(pm = 0.028, pf = 0.005,
x = c(-100, -2.32634787404084, -2.05374891063182, ...,
2.05374891063182, 2.32634787404084, 100),
y = c(2, 2.025, 2.05, 2.075, ..., 9.95, 9.975, 10) )
list(mu0199=0,mu1199=0,mu0125=0,mu1125=0,tau0199=1,
tau1199=1,tau0125=1,tau1125=1)
y0CA199[] y0CA125[] y1CA199[] y1CA125[]
28          13.3          2.4          79.1
15.5         11.1          719          31.4
8.2          16.7         2106.6         15
3.4          12.6         24000          77.8
...
END

```

There are two grids in the data, one is  $x$ , which corresponds to the grid of  $t$  values discussed above. The other is  $y$ , which is used to get values of predictive probabilities of disease over the range of log outcomes in the data, namely from 2 to 10. For each grid point, WinBUGS output includes mean (or median), and a 95% probability interval for the posterior distribution of the proportion of diseased individuals,  $\Pr(D|y, \text{data})$ , and the ROC values

$ROC(t)$ . In the code, there are predictive probabilities for both types of tests and for males and females. Results are plotted in Figure 3, but we wait to discuss them until we finish discussing how we handled the ROC curve output.

Summary output from WinBUGS was saved as a text file with column headings named “PostMean”, “L025”, and “U975”, and read into R. The R code that produced Figure 3 is presented below.

```
t <- c(0,seq(0.01,0.99,0.01),1)
ROC125 <- PostMean[1:101]
ROC199 <- PostMean[102:202]
ROC125L <- L025[1:101]
ROC125U <- U975[1:101]
ROC199L <- L025[102:202]
ROC199U <- U975[102:202]
plot(t,ROC199,type="l",ylim=c(0,1),xlim=c(0,1),
axes=F,xlab="",ylab="",lwd=2)
axis(side=1,at=c(0,0.2,0.4,0.6,0.8,1),
labels=c("0","0.2","0.4","0.6","0.8","1"))
axis(side=2,at=c(0,0.2,0.4,0.6,0.8,1),
labels=c("0","0.2","0.4","0.6","0.8","1"),line=0)
lines(t,ROC125,lty=1,lwd=2)
lines(t,ROC199L,lty=2,lwd=2)
lines(t,ROC199U,lty=2,lwd=2)
lines(t,ROC125L,lty=2,lwd=2)
lines(t,ROC125U,lty=2,lwd=2)
```



```
text(.1,.95, "CA-19-9") text(.3,.35, "CA-125")
```

*Exercise.* Revise the above code to give results for only the CA199 marker. Run the code and compare results with those obtained here. Make your own plot of the ROC curve for this single marker.

EXAMPLE 2.2. *Diagnosing pancreatic cancer continued.*

We again consider the Wieand data but now examine the use of the CA-19-9 and CA-125 markers to diagnose cancer of the pancreas for individual patients. We note that all individuals associated with the data in the example have already been diagnosed. But we can imagine that an individual has just been tested and that either or both of the marker outcomes have been obtained. Moreover, for this example, we have some additional information. We have a large sample of extraneous data that indicated that the prevalence for males is 2.8% and that the prevalence for females is 0.5%. Thus, we decide to calculate predictive probabilities of pancreatic cancer as a function of the log of the marker values for this individual, and why not at the same time, calculate these probabilities over a range of values as already discussed. We do this for both males and females, separately. Note that this really has little to do with males and females, but rather it has to do with the prevalence. The predictive probability always depends heavily on the prevalence. This particular analysis presumes *known* prevalences of 2.8% and 0.5%, respectively. These values correspond to the lifetime risk for pancreatic cancer in Australian males and females (Australian Institute of Health and Welfare National Mortality Database, Australia's Health 2004; <http://www.aihw.gov.au>). However, in the absence of known values, it is

simple to modify the code to have two independent beta priors for the two prevalences. If there were only one group of individuals involved, there would be a single beta prior for it.

To perform this analysis, we assume that the operating characteristics of the tests are the same in males and females, and in Australia if we are to make the best use of the prevalence information. Figure 4 presents a comparison of the predicted probability of pancreatic cancer across a range of (log transformed) biomarker values for male and female patients. Specific values are given below. The last *for loop* in the WinBUGS code above was used to generate these results. The vector  $y$  used in that code contains log transformed biomarker values ranging from 2 to 10 by increments of 0.025. This vector was created in R using

```
dput(seq(2,10,0.025), file="C:\\Temp\\y.txt")
```

and then pasted into WinBUGS. Predictive probabilities are given in the table below.

		Marker Value			
		4	5	6	7
CA-19-9	Male	0.043	0.437	0.943	0.998
	Female	0.008	0.145	0.800	0.991
CA-125	Male	0.072	0.324	0.717	0.891
	Female	0.013	0.094	0.453	0.766

*Exercise.* Revise the above code to obtain only predictive probabilities of pancreatic cancer using only the CA-19-9 marker. Obtain a plot of these probabilities over the same grid.

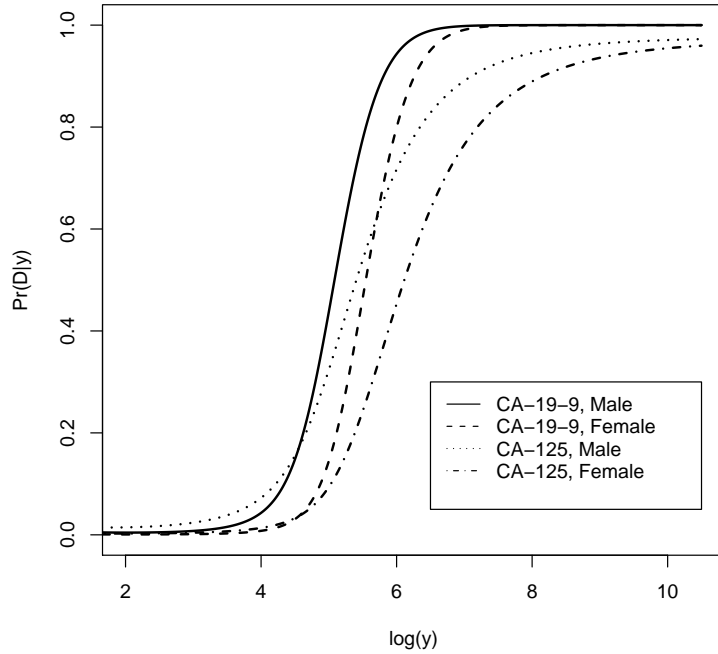


Figure 4: Predicted probability of pancreatic cancer across log transformed biomarker scores ( $y$ ).

*Exercise.* Augment the above code so that you can calculate the posterior probability that the AUC based on CA-19-9 is larger than the AUC based on CA-125, assuming that the two markers are conditionally independent. Also give a 95% posterior interval for the difference.

## 2.1 Posterior Calculations

With gold-standard data, inferences are based on a posterior sample

$$\{(\pi^{(j)}, \theta_0^{(j)}, \theta_1^{(j)}) : j = 1, \dots, MC\}.$$

For each sample point define

$$f_0^{(j)}(y) \equiv f_0(y|\theta_0^{(j)}), \quad F_0^{(j)}(y) \equiv F_0(y|\theta_0^{(j)}), \quad S_0^{(j)}(y) = 1 - F_0^{(j)}(y)$$

and

$$f_1^{(j)}(y) \equiv f_1(y|\theta_1^{(j)}), \quad F_1^{(j)}(y) \equiv F_1(y|\theta_1^{(j)}), \quad S_1^{(j)}(y) = 1 - F_1^{(j)}(y).$$

We can now compute posterior samples from the ROC curve. Let

$$\text{ROC}^{(j)}(t) = S_1^{(j)}\left(S_0^{-1(j)}(t)\right).$$

These are computed across a fine grid for  $t$ , such as  $t = 0, 0.01, 0.02, \dots, 1$ .

We also obtain posterior samples

$$\text{AUC}^{(j)} = \int_0^1 \text{ROC}^{(j)}(t) dt \quad \text{or} \quad \text{pAUC}^{(j)} = \int_{t_0}^{t_1} \text{ROC}^{(j)}(t) dt$$

which are computed using a numerical integration procedure if they do not have closed-form expressions. The ROC curve is estimated as

$$\begin{aligned} \text{ROC}(t) &\doteq \frac{1}{MC} \sum_{j=1}^{MC} \text{ROC}^{(j)}(t) \\ &= \frac{1}{MC} \sum_{j=1}^{MC} S_1^{(j)}\left(S_0^{-1(j)}(t)\right) \end{aligned}$$

with interval estimates derived from the percentiles of the posterior sample. Point and interval estimates for AUC (or pAUC) are found similarly. Note that computations related to the ROC curve do not involve the prevalence.

Disease diagnosis is based on the predictive probability of disease conditional on an individuals test data  $y$ , which is approximated using Monte

Carlo sampling as follows:

$$\begin{aligned}
\Pr(D|y, \text{data}) &= \int \Pr(D|y, \theta_0, \theta_1, \pi, \text{data}) p(\theta_0, \theta_1, \pi | \text{data}) d\theta_0 d\theta_1 d\pi \\
&= \int \frac{f_1(y)\pi}{f_1(y)\pi + f_0(y)[1-\pi]} p(\theta_0, \theta_1, \pi | \text{data}) d\theta_0 d\theta_1 d\pi \\
&\doteq \frac{1}{MC} \sum_{j=1}^{MC} \frac{f_1^{(j)}(y)\pi^{(j)}}{f_1^{(j)}(y)\pi^{(j)} + f_0^{(j)}(y)[1-\pi^{(j)}]}.
\end{aligned}$$

In the pancreatic cancer example where prevalences were assumed known for males and females, the “known” values were substituted. This would generally not be the case.

*Exercise.* Treat the pancreatic cancer data as if there were no GS test, that is, regard the  $51+90 = 141$  observations as if they came from the mixture  $\pi f_1(\cdot) + (1 - \pi)f_0(\cdot)$ . Using only the CA-19-9 marker data, obtain the predictive probabilities of pancreatic cancer. Use  $N(\log(20), 100)$  and  $N(\log(200), 100)$  priors for the means in the no cancer and cancer cases respectively, and the same priors for precisions. Also use a prior on  $\text{beta}(25, 75)$  prior for  $\pi$ , and then repeat using a  $\text{beta}(75, 25)$  prior on  $\pi$  and compare results. Be careful to check for convergence. You should use starting values that have the mean log marker value for the cancer group larger than for the non cancer group. Finally, modify the code so that you can also obtain an ROC curve estimate based on this marker. Give 95% confidence intervals for the  $\text{ROC}(t)$  values for several values of  $t$ .

*Exercise.* Microarray technology and statistical methods for microarray data analysis allow for identification of specific genes that are associated with disease (e.g. cancer) occurrence. The following case-control data from Pepe

(2003, page 98) (see also Pepe et al., 2003) are gene expression levels for a specific gene from 23 non-diseased individuals and from 30 ovarian cancer patients. Obtain graphical (e.g. side-by-side boxplots, histograms) and numerical (e.g. mean, standard deviation, range), summaries of the data to investigate the distribution of gene expression levels for the two groups. Using a binormal analysis with diffuse prior distributions, (i) estimate the ROC curve and obtain a pointwise probability band for it, (ii) obtain the posterior mean and median AUC, and (iii) obtain and interpret a 95% probability interval for AUC. Comment on the discriminatory ability of this gene as a potential diagnostic test for ovarian cancer.

	Gene expression level
Controls	0.442, 0.500, 0.510, 0.568, 0.571, 0.574, 0.588, 0.595, 0.595, 0.595, 0.598, 0.606, 0.617, 0.628, 0.641, 0.641, 0.680, 0.699, 0.746, 0.793, 0.884, 1.149, 1.785
Cases	0.543, 0.571, 0.602, 0.609, 0.628, 0.641, 0.666, 0.694, 0.769, 0.800, 0.800, 0.847, 0.877, 0.892, 0.925, 0.943, 1.041, 1.075, 1.086, 1.123, 1.136, 1.190, 1.234, 1.315, 1.428, 1.562, 1.612, 1.666, 1.666, 2.127

*Exercise.* Check the normality assumption for the cases and for the controls in the data above. If the data don't look normal, find a transformation that makes them look better in this sense.

*Exercise:* Using the ovarian cancer gene expression data from the previous exercise, plot the predicted probability across a grid of transformed gene expression levels. Also compute 95% pointwise bands for the prediction curve.

Do this separately for women in the general population where disease prevalence is 0.04%, and for women in the at-risk population (women 50 years of age or older, or with a family history of ovarian cancer) where disease prevalence is 0.25%. These prevalence values were derived from Bell et al. (1998). How would you classify women from the general population with expression levels of 0.50, 1.0, and 1.5? How would you classify women from the increased risk population with expression levels of 0.50, 1.0, and 1.5?

### 3 Regression

We now consider the situation where there is additional information in the form of a covariate or covariate vector, say  $x_1, \dots, x_{p-1}$ . A typical covariate that is related to disease status might be age. Older people or animals might be more likely to have a particular disease/infection/affliction than younger ones. We would like to incorporate such information if deemed relevant. This information can help diagnosis in both the dichotomous and continuous test cases. We first consider the continuous case. We make a crucial assumption that, the test outcome data are independent of the covariate data given disease status. So, for example, we would assume that a test outcome for HIV (human immunovirus) infection would be independent of age, conditional on actual HIV status. This translates to mean  $f_1(y|x, HIV^+) = f(y|HIV^+)$  and  $f_0(y|x, HIV^-) = f_0(y|HIV^-)$ , where  $y$  is HIV marker outcome. The distribution of marker outcomes among  $HIV^+$  individuals is not affected by  $x$ , and similarly for  $HIV^-$  individuals. In general, we replace  $HIV^+$  with  $D$  and  $HIV^-$  with  $\bar{D}$ .

On the other hand, we do assume that  $x$  is related to disease status. So now, prevalence depends on  $x$ . We thus define the prevalence of  $D$  among individuals who share covariate vector  $x$  as  $\Pr(D|x) = \pi_x$ . There are many models that might relate prevalence to covariate values. The most common is one that we discussed at length in Chapter 8, namely the logistic regression model. This model is

$$\text{logit}(\pi_x) = x'\beta, \quad \text{where } x' = (1, x_1, \dots, x_{p-1}), \quad \beta' = (\beta_1, \dots, \beta_p).$$

We focus on the non gold-standard (NGS) case, so we sample individuals with unknown disease status. Consider those in this population who have covariate values  $x$ , and then randomly sample one of them. From this individual, obtain a test outcome  $y$ . Then the marginal pdf for the test outcome is

$$f(y|x) = \pi_x f_1(y|D) + (1 - \pi_x) f_0(y|\bar{D}),$$

where we have assumed conditional independence, which was discussed above. Then the predictive probability of disease for an individual with test value  $y$  and covariate  $x$  is obtained using Bayes Theorem as:

$$\Pr(D|x, y) = \frac{\pi_x f_1(y|D)}{\pi_x f_1(y|D) + (1 - \pi_x) f_0(y|\bar{D})},$$

which looks just like what we had before only now  $\pi$  is replaced with  $\pi_x$ . We are essentially done except for the fact that the above pdf's are not known since they generally depend on unknown parameters as discussed in the previous section. Moreover,  $\beta$  is not known either if we use the logistic regression model for  $\pi_x$ . But these are all objects that we have dealt with before.



Our approach from here is to simply treat  $\beta$  as we did in chapter 8 on logistic regression, and to treat  $f_0$  and  $f_1$  as we did in the previous section of this chapter. Suppose that we believe that the test values  $y$  can reasonably be assumed to be log normal as discussed for the example in the previous section. Let's thus assume that the data have already been transformed so that  $y$  is now the logarithm of the original test value. We have a single sample of size  $n$  independently obtained values in the form  $\{y_i, x_i : i = 1, \dots, n\}$ . Using our previous notation, the likelihood function is obtained as

$$L(\beta, \theta_0, \theta_1 | Y) = \prod_{i=1}^n \{\pi_i f_1(y_i | \theta_1) + (1 - \pi_i) f_0(y_i | \theta_0)\}$$

where the pdf's are now normal with means  $\mu_i$  and variances  $\sigma_i^2$  for  $i = 0, 1$ , and  $\text{logit}(\pi_i) = x_i' \beta$ .

We use the same priors that were used before, from Chapter 8 and the previous section. If you think about it for just a second, you realize that the posterior is not tractable, that is, it's not in a nice recognizable form. Moreover, it's not immediately obvious how to program this model into WinBUGS in its current form. Thus, we consider augmenting the data in a way that makes everything much easier.

Let  $z_i$  be an indicator of the disease status for the  $i$ th individual in the data. So we have  $z_i \sim \text{Bernoulli}(\pi_i)$ . If we knew all of the  $z_i$ 's, the likelihood would be in a nice form, namely, the augmented data likelihood is

$$\begin{aligned} L(\beta, \theta_0, \theta_1 | Y, Z) &= \prod_{i=1}^n \{\pi_i f_1(y_i | \theta_1)\}^{z_i} \{(1 - \pi_i) f_0(y_i | \theta_0)\}^{1-z_i} \\ &= \prod_{\{y_i: z_i=1\}} f_1(y_i | \theta_1) \prod_{\{y_i: z_i=0\}} f_0(y_i | \theta_0) \prod_{i=1}^n \pi_i^{z_i} (1 - \pi_i)^{1-z_i}. \end{aligned}$$

Note that this likelihood is in the same form as if we had training data, as discussed in the previous section, and in addition, an independent binomial (logistic) regression sample. Thus if we put the same types of priors on  $(\theta_0, \theta_1)$  and  $\beta$ , all independent, then the corresponding augmented data posterior,  $p(\theta_0, \theta_1, \beta|Y, Z) = p(\theta_0|Y, Z)p(\theta_1|Y, Z)p(\beta|Z)$ , where each component is in exactly the same form as what was previously discussed.

The WinBUGS code is now especially easy to write and we give code below with  $p = 2$ . We assume that the single covariate is dichotomous. The data given were simulated with the first 100 observations from a  $N(50,100)$  pdf and the next 200 observations were from a  $N(40,100)$  pdf. Then for the first 100 observations, we let the Bernoulli covariate  $x$  take on the value 1 for 19 values and 0 for the 20th and continuing in this pattern. For the second hundred, we reversed the patterns. Thus, 95% of the first 100  $x$  values were ones and 95% of the last 200 are zeros. Let  $\pi_0 = \Pr(D|x = 0)$  and  $\pi_1 = \Pr(D|x = 1)$ . We have  $\text{logit}(\pi_0) = \beta_1$  and  $\text{logit}(\pi_1) = \beta_1 + \beta_2$ . We let  $\tilde{\pi}_i$  be the probability of disease when  $x = i$ , and let  $\tilde{\pi}_i$ 's have independent beta priors with  $\tilde{\pi}_1 \sim \text{Beta}(19, 1)$  and  $\tilde{\pi}_2 \sim \text{Beta}(1, 19)$ . This prior information is consistent with the truth so we are cheating quite a bit here, but mainly for the purpose of illustration. We then induce a joint prior on  $\beta$  the same way we did in Chapter 8. The WinBUGS code below makes this clear.

We used mildly informative priors for the  $\mu$ 's in that we picked prior guesses for them, but we placed large variances on the corresponding normal priors. We used diffuse priors for the precisions. Observe that the code specifies a distribution for the latent  $z$ 's first, and then specifies a distribution for  $y$ 's that is conditional on the  $z$ 's. This is akin to writing  $f(z, y) =$

$f(z)f(y|z)$  for the joint pdf.

Note that the variables denoted by  $v$  will give numerical approximations to the predictive densities  $f(y|z = i, \text{data}) = \int f(y|z = i, \theta_i)p(\theta_i | \text{data})d\theta_i$  for  $i = 0, 1$ . Note that in the code, we index these with 1 and 2 rather than 1 and 0. This is because zeros cannot be used in WinBUGS arrays. So in the code, we meant 1 to be 1, and 2 to be 0. Finally, note that we have initialized the  $z$ 's by simply defining the initial  $z$ 's to be the observed  $x$ 's. If you don't initialize them yourself, WinBUGS will try to do that by sampling from it's best guess for the Bernoulli( $\pi_i$ )'s. But this will give non-integral initial values, and WinBUGS will choke. So you need to start somewhere reasonable. You should monitor the  $z$ 's and try different starting values to check for convergence.

Finally note that we have also calculated the predictive probabilities of disease for each individual in the sample in the code below. You should be aware that the mean the  $z$ 's is theoretically equivalent to the posterior mean of the predictive probabilities. However, in monitoring these probabilities, defined as "prob" in the code, you will also be able to get posterior probability intervals for the proportions of individuals in the population with either  $x = 1$  or  $x = 0$  that will be diseased.

```
model{
  for(i in 1:n){
    z[i] ~ dbern(pi[i])
    y[i] ~ dnorm(mmu[i], ttau[i])
    mmu[i] <- z[i]*mu[1] + (1-z[i])*mu[2]
    ttau[i] <- z[i]*tau[1] + (1-z[i])*tau[2]
```

```

    logit(pi[i]) <- beta[1] + beta[2]*x[i]
  }
  tildepi[1] ~ dbeta(a[1],b[1])
  tildepi[2] ~ dbeta(a[2],b[2])
  beta[1] <- logit(tildepi[1])
  beta[2] <- logit(tildepi[2]) - logit(tildepi[1])
  for(i in 1:2){
    mu[i] ~ dnorm(d[i],0.001)
    tau[i] ~ dgamma(0.001,0.001)
  }
  v[1] ~ dnorm(mu[1],tau[1]) #Pred Dens Popn 1
  v[2] ~ dnorm(mu[2],tau[2]) #Pred Dens Popn 0
  # Get predictive probabilities of $$$
  for(i in 1:n){
    f1[i] <- sqrt(tau[1])*exp((-tau[1]/2)*pow(y[i]-mu[1],2))
    f2[i] <- sqrt(tau[2])*exp((-tau[2]/2)*pow(y[i]-mu[2],2))
    prob[i] <- f1[i]*pi[i]/(f1[i]*pi[i] + f2[i]*(1-pi[i]))
  }
}
initial values and data available from our website

```

*Exercise.* Run the above code and monitor the predictive probabilities and the  $z$ 's. Observe that the posterior means are approximately the same for each case in the data. Also give 95% probability intervals for the probabilities of disease in the population corresponding to cases 1 and 201 in the data. Finally, modify the code so that you can get plots of estimated probabilities

of disease as a function of  $y$  over a suitable range of  $y$  values, for both types of individual, with  $x = 0$  and with  $x = 1$ .

## 4 Regression: Dichotomous Case

Our final exploration in this section is to consider the dichotomous case version of the previous section. This simply means that we still have covariate information  $x$  about the prevalence, so the prevalence is still modeled as  $\text{logit}(\pi_x) = x'\beta$ . But now instead of a continuous response,  $y$ , we have a dichotomous response, which we also term as  $y$ . This response may be for a test outcome that is naturally dichotomous, or it may result from dichotomizing a continuous response by picking a cutoff. Either way, we have a sensitivity,  $\eta = \Pr(y = 1|D)$ , and a specificity,  $\theta = \Pr(y = 0|\bar{D})$ .

We will also have a sample of (independent) values,  $\{y_i, x_i\}$ , just as before. Only now, the likelihood looks like

$$L(\eta, \theta, \beta|Y) = \prod_{i=1}^n \{\pi_i \eta + (1 - \pi_i)(1 - \theta)\}^{y_i} \{\pi_i(1 - \eta) + (1 - \pi_i)\theta\}^{1-y_i}.$$

Priors are the same as before for  $(\eta, \theta, \beta)$ . This model can be handled easily with or without latent  $z_i$ 's. We leave the latent case to the reader as an exercise. The WinBUGS code below is for smoking cessation data that were analyzed by McInturff et al (2004).

The response variable corresponds to successful cessation of smoking or not. Covariates are age level (20-29;  $\geq 30$ , or  $< 20$ ), education level (HS graduate; some college, less than HS graduate), and smoking history (less than one pack per day or more). We set baseline categories to be “under 20”

years of age, “less than HS grad”, and “a pack or more a day”. Thus we have a 3 category age variable with two binary variables indicating age groups 20-29 ( $x_2$ ) and  $\geq 30$  ( $x_3$ ), a 3 category education variable with two variables indicating HS grad ( $x_4$ ) and some college ( $x_5$ ), and a smoking history variable indicating less than a pack a day ( $x_6$ ). There are thus  $1 + 2 + 2 + 1$  regression coefficients in the model. We elicited our prior on probabilities of smoking cessation for six covariate combinations, which are given in the table below.

This specification was then used to induce a prior on the six dimensional vector  $\beta$ , using the technique discussed in Chapter 8. In addition, an expert gave information about the sensitivity, which in this instance means the probability of saying you quit smoking when you did, and the specificity, which means the probability of saying you didn’t quit when in fact you didn’t quit. Since it would be unlikely that someone would say that they didn’t quit when they did, a rather strong prior was placed on the sensitivity, a Beta(99, 1). However, there was much less certainty about the specificity, which is reflected by the Beta(14, 2) prior, which has much less weight associated with it.

Covariate combinations considered for the prior elicitation were:

Case	Int	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	Distn	Prior mode
1	1	1	0	1	0	1	Be(8, 15)	0.33
2	1	0	1	1	0	1	Be(10, 15)	0.39
3	1	1	0	1	0	0	Be(3, 13)	0.14
4	1	1	0	0	1	1	Be(8, 10)	0.44
5	1	1	0	0	0	1	Be(4, 15)	0.18
6	1	0	0	1	0	0	Be(6, 15)	0.26

```

model{
  for(i in 1:N){
    y[i] ~ dbern(q[i]) # specification of the LR model with error
    q[i] <- pi[i]*Se+(1-pi[i])*(1-Sp)
    logit(pi[i]) <- beta[1]+x2[i]*beta[2]+x3[i]*beta[3]
                    +x4[i]*beta[4]+x5[i]*beta[5]+x6[i]*beta[6]
  }
  Se ~ dbeta(99, 1) # the priors are specified for Se and Sp
  Sp ~ dbeta(14, 2)
  p[1] ~ dbeta(8, 15) # spec of the prior on 6 probs of smoking
  p[2] ~ dbeta(10, 15) # cessation for 6 covariate combinations
  p[3] ~ dbeta(3, 13)
  p[4] ~ dbeta(8, 10)
  p[5] ~ dbeta(4, 15)
  p[6] ~ dbeta(6, 15)

  # relate the regression coefficients to the
  # probabilities on which prior was specified
  # this results in an induced informative prior
  # on the regression coefficients
  beta[1] <- xinv[1,1]*logit(p[1]) + xinv[1,2]*logit(p[2])
            + xinv[1,3]*logit(p[3]) + xinv[1,4]*logit(p[4])
            + xinv[1,5]*logit(p[5]) + xinv[1,6]*logit(p[6])
  beta[2] <- xinv[2,1]*logit(p[1]) + xinv[2,2]*logit(p[2])
            + xinv[2,3]*logit(p[3]) + xinv[2,4]*logit(p[4])

```

```

+ xinv[2,5]*logit(p[5]) + xinv[2,6]*logit(p[6])
beta[3] <- xinv[3,1]*logit(p[1]) + xinv[3,2]*logit(p[2])
+ xinv[3,3]*logit(p[3]) + xinv[3,4]*logit(p[4])
+ xinv[3,5]*logit(p[5]) + xinv[3,6]*logit(p[6])
beta[4] <- xinv[4,1]*logit(p[1]) + xinv[4,2]*logit(p[2])
+ xinv[4,3]*logit(p[3]) + xinv[4,4]*logit(p[4])
+ xinv[4,5]*logit(p[5]) + xinv[4,6]*logit(p[6])
beta[5] <- xinv[5,1]*logit(p[1]) + xinv[5,2]*logit(p[2])
+ xinv[5,3]*logit(p[3]) + xinv[5,4]*logit(p[4])
+ xinv[5,5]*logit(p[5]) + xinv[5,6]*logit(p[6])
beta[6] <- xinv[6,1]*logit(p[1]) + xinv[6,2]*logit(p[2])
+ xinv[6,3]*logit(p[3]) + xinv[6,4]*logit(p[4])
+ xinv[6,5]*logit(p[5]) + xinv[6,6]*logit(p[6])
}
initial values and data available from our web site

```

*Exercise.* Modify the above code to make inferences about the probability of quitting smoking over a range of covariate combinations. Make a table of estimated probabilities with probability intervals so that it is easy to see what is the effect of age, education and smoking history on the probability of quitting. Make inferences for sensitivity and specificity. Now run the code assuming perfect sensitivity and specificity and comment on the change in inferences. Next assume uniform priors for the last three  $p$ 's and see what effect this has on inferences. Finally, revise the code to specify  $N(0, 1)$  priors on all regression coefficients and compare results.



## 5 Dependent Dichotomous Tests

If two tests have the same biological basis for detecting disease/infection, for example, if both tests are ELISA tests only made by different manufacturers, then we would expect them to be dependent, even conditional on disease status. This problem is then handled by expanding the model discussed for two tests in the previous section. Using the same notation from there, we realize that

$$\Pr(T_1^+ | T_2^+, D) \neq \Pr(T_1^+ | D), \quad \Pr(T_1^- | T_2^-, \bar{D}) \neq \Pr(T_1^- | \bar{D}).$$

Equivalently, the corresponding joint probabilities don't factor as the product of their marginal probabilities.

However, the representation of the data remains the same. That is, we have, in the case of two populations and two tests, a  $2 \times 2 \times 2$  table of observed data. We don't discuss the augmented data case here where the latent counts of diseased individuals are also modeled. For this situation, our data are simply the two  $2 \times 2$  tables,  $y_1$  and  $y_2$ , where  $y_k \sim \text{Mult}(n_i, \{p_{ijk} : i, j = 1, 2\})$ , as in the case from chapter 14, except that now we have different  $p_{ij}$ 's. There are at least two ways to model the cell probabilities. The easiest to explain and program was introduced by Dendukuri and Joseph (2001) (DJ), who considered the two-tests and one population problem under dependence. An alternative that has some nice features was introduced by Georgiadis et al (2003). In either case, we still assume that test accuracies are the same in each population.

Define

$$\begin{aligned} CovDp &= \Pr(T_1^+, T_2^+ | D, \text{popn } k) - \eta_1 \eta_2 \\ CovDn &= \Pr(T_1^-, T_2^- | \bar{D}, \text{popn } k) - \theta_1 \theta_2. \end{aligned}$$

Then we observe that tests 1 and 2 are conditionally independent if and only if these two quantities are both zero. Define  $\eta_{ij} = \Pr(T_1^i, T_2^j | D)$  where  $i, j = 1$  corresponds to  $+$  and  $i, j = 0$  corresponds to  $-$ . So  $\eta_{10} = \Pr(T_1^+, T_2^- | D)$ . Define  $\theta_{ij}$ 's analogously, so that, for example,  $\Pr(T_1^+, T_2^+ | \bar{D}) = \theta_{11}$ . We then have the marginal probabilities

$$p_{ijk} = \Pr(T_1^i, T_2^j | \text{popn } k) = \pi_k \eta_{ij} + (1 - \pi_k) \theta_{ij}.$$

We then observe that

$$\begin{aligned} \eta_{11} &= \eta_1 \eta_2 + CovDp \\ \theta_{00} &= \theta_1 \theta_2 + CovDn \end{aligned}$$

etc. There are 6 more expressions like these, and they can be surmised from the WinBUGS code below.

Thus, our model now has two additional parameters that reflect two different kinds of conditional dependence. The total number of parameters is now 8, while we still only have 6 degrees of freedom in the two-test, two-population case. In the one population case, we have 7 parameters and only 3 degrees of freedom. So our model lacks identifiability and specification of the priors becomes even more important, again. When one of the tests has been in use for some time, it's properties may be reasonably well known. In this instance, there may be accurate and precise information for say  $(\eta_1, \theta_1)$ .

In this instance, we can proceed with the model for the two-population case discussed here with the expectation that the data in conjunction with this prior information will result in reasonable inferences for the test accuracies for the other test, which is likely the main goal of interest. Moreover, if the two population prevalences are known reasonably well, we would then be comfortable making inferences about the accuracy of both tests. In the one population, we should believe that our priors are reasonably precise and accurate for 4 parameters.

It would be rare in our experience that we would be particularly interested in estimating the above covariance parameters. The important point is that our model accommodates non zero covariances. There is one issue that we still need to discuss related to these parameters. It is straightforward to establish that  $CovD_p$  and  $CovD_n$  are bounded above and below, for example,

$$(\eta_1 - 1)(1 - \eta_2) \leq CovD_p \leq \min(\eta_1, \eta_2) - \eta_1\eta_2.$$

A similar inequality holds for  $CovD_n$ , which can be surmised from the code below. These inequalities are important since, if violated, the consequence is that various probabilistic quantities could be smaller than zero or above one, which could result in embarrassing results. These inequalities guarantee that we will obey probability laws in making our inferences.

Because we would rarely have very precise information about these covariances, we recommend placing uniform priors on them, over the range of possible values specified by the inequalities. If it were believed with certainty that there could not be a negative association, for example, the priors could be uniform starting at 0 and stopping at the appropriate upper bound.

We consider data that were collected by Bouma et. al (2004), for a single population. Two fluorescent-antibody tests (FAT) were used to detect classical swine fever virus. Expert opinion was elicited for the sensitivities and specificities for both tests as well as for the prevalences. Uniform priors were used for the covariances. Observe that we also make inferences for the correlations between the Bernoulli tests; definitions are given in the code.

```

model{
x[1:4] ~ dmulti(p[1:4], n)
p[1] <- pi*(Sefat1*Sefat2+covDp) + (1-pi)*((1-Spfat1)*(1-Spfat2)+covDn)
p[2] <- pi*(Sefat1*(1-Sefat2)-covDp) + (1-pi)*((1-Spfat1)*Spfat2-covDn)
p[3] <- pi*((1-Sefat1)*Sefat2-covDp) + (1-pi)*(Spfat1*(1-Spfat2)-covDn)
p[4] <- pi*((1-Sefat1)*(1-Sefat2)+covDp) + (1-pi)*(Spfat1*Spfat2+covDn)
ls <- (Sefat1-1)*(1-Sefat2)
us <- min(Sefat1,Sefat2) - Sefat1*Sefat2
lc <- (Spfat1-1)*(1-Spfat2)
uc <- min(Spfat1,Spfat2) - Spfat1*Spfat2
pi ~ dbeta(13.322, 6.281) ### Mode=0.70, 95% sure > 0.50
Sefat1 ~ dbeta(9.628,3.876) ### Mode=0.75, 95% sure > 0.50
Spfat1 ~ dbeta(15.034, 2.559) ### Mode=0.90, 95% sure > 0.70
Sefat2 ~ dbeta(9.628, 3.876) ### Mode=0.75, 95% sure > 0.50
Spfat2 ~ dbeta(15.034, 2.559) ### Mode=0.90, 95% sure > 0.70
covDn ~ dunif(lc, uc)
covDp ~ dunif(ls, us)
rhoD <- covDp / sqrt(Sefat1*(1-Sefat1)*Sefat2*(1-Sefat2))
rhoDc <- covDn / sqrt(Spfat1*(1-Spfat1)*Spfat2*(1-Spfat2))

```

```

}
list(n=214, x=c(121,6,16,71))
list(pi=0.7, Sefat1=0.75, Spfat1=0.90, Sefat2=0.75, Spfat2=0.90)

```

*Exercise.* Run the above code. Try different priors, for example, try  $\text{Unif}[0,uc]$  and  $\text{Unif}[0,us]$  priors for  $CovDn$  and  $CovDp$  to see what is the impact on the analysis. Also run modified code with the covariances set to zero and compare output. Does it seem necessary to model the covariances?

*Exercise.* Modify the above code to handle two population data. Then, analyze the Trout data from chapter 14 using this code and the priors discussed there. Compare results with what was obtained there. Does it seem necessary to model the dependence for those data?

Now consider a different example, where two tests are applied sequentially to samples of animals from multiple populations to detect brucellosis in cattle. We model the prevalences exchangeably as we did in chapter 14. Test 1 is BAPA, so  $Sebapa$  and  $Spbapa$  are the sensitivity and specificity for BAPA. Test 2 is Rivanol, so  $Seriv$  and  $Spriv$  are the sensitivity and specificity of Rivanol. If test 1 is positive, test 2 is applied. Otherwise, test 2 is not applied. So there are only 3 possible categories for the two test outcomes, namely  $(+, +)$ ,  $(+, -)$ , and  $(-, NA)$ . The code below thus models the three category responses within each population as a 3 category multinomial. The code makes clear precisely what the model is.

```

model{
  for(i in 1:K){

```

```

x[i, 1:3] ~ dmulti(p[i, 1:3], n[i])
p[i,1] <- pi[i]*(Sebapa*Seriv+covDp) + (1-pi[i])*((1-Spbapa)*(1-Spriv)+covDn)
p[i,2] <- pi[i]*(Sebapa*(1-Seriv)-covDp) + (1-pi[i])*((1-Spbapa)*Spriv-covDn)
p[i,3] <- 1-p[i,1]-p[i,2]
pi[i] ~ dbeta(alpha, beta)
}
Sebapa ~ dbeta(88.3,1.9) ## Mode=0.99, 95% sure > 0.95
Spbapa ~ dbeta(13.3,6.3) ## Mode=0.70, 95% sure > 0.50
Seriv ~ dbeta(130.7,15.4) ## Mode=0.90, 95% sure > 0.85
Spriv ~ dbeta(99.7,6.2) ## Mode=0.95, 95% sure > 0.90
alpha <- mu*psi
beta <- psi*(1-mu)
mu ~ dbeta(16.9,48.6) ## Mode=0.25; 95% sure < 0.35
psi ~ dgamma(7.23, 1.28)
ls <- (Sebapa-1)*(1-Seriv)
lc <- (Spbapa-1)*(1-Spriv)
us <- min(Sebapa,Seriv)-Sebapa*Seriv
uc <- min(Spbapa, Spriv) - Spbapa*Spriv
covDn ~ dunif(lc, uc)
covDp ~ dunif(ls, us)
rhoDp <- covDp / sqrt(Sebapa*(1-Sebapa)*Seriv*(1-Seriv))
rhoDn <- covDn / sqrt(Spbapa*(1-Spbapa)*Spriv*(1-Spriv))
pi21 ~ dbeta(alpha,beta)
a[1] <- 1-step(pi21-0.50)
a[2] <- 1-step(pi21-0.10)

```

```
a[3] <- step(pi21-0.90)
a[4] <- step(pi21-0.75)
}
```

initial values and data are available on our web site

*Exercise.* Analyze these data using the above code. Then modify the code to handle independent tests and run it. Comment on the necessity to model the dependence.

## References

- Bell, R., Petticrew, M., Luengo, S., and Sheldon, T. A. (1998). Screening for ovarian cancer: a systematic review. *Health Technology Assessment* (Winchester, England), 2: i-iv, 1-84.
- Bouma A, Stegeman JA, Engel B, de Kluijver EP, Elbers AR, De Jong MC. (2002). Evaluation of diagnostic tests for the detection of classical swine fever in the field without a gold standard. *J Vet Diagn Invest.* 13, 383-8.
- Branscum, A.J., Gardner, I.A., and Johnson, W.O. (2004). Bayesian modeling of animal and herd-level prevalences. *Preventive Veterinary Medicine*, 66, 101–112.
- Branscum, A.J., Gardner, I.A., and Johnson, W.O. (2005). Estimation of diagnostic test sensitivity and specificity through Bayesian modeling. *Preventive Veterinary Medicine*, 68, 145-163.
- Broemeling, L.D. (2007). *Bayesian Biostatistics and Diagnostic Medicine*. Chapman & Hall, Boca Raton, FL.
- Choi, Y.K., Johnson, W.O., and Thurmond, M.C. (2006). Diagnosis using predictive probabilities without cutoffs. *Statistics in Medicine*, 25, 699-717.
- Choi, Y.K., Johnson, W.O., Collins, M.T., and Gardner, I.A. (2006). Bayesian estimation of ROC curves in the absence of a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics*, 11, 210-229.



- Dendukuri, N., Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 57, 158-167.
- Enøe, C., Georgiadis, M.P., Johnson, W.O. (2000). Evaluation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease status is unknown. *Preventive Veterinary Medicine* 45, 61-81.
- Georgiadis, M.P., Gardner, I.A., Hedrick, R.P. (1998). Field evaluation of sensitivity and specificity of a polymerase chain reaction (PCR) for detection of *N. salmonis* in rainbow trout. *Journal of Aquatic Animal Health* 10, 372-380.
- Georgiadis, M.P., Johnson, W.O., Gardner, I.A., Singh, R. (2003). Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Applied Statistics* 52, 63-76.
- Hanson, T.E., Branscum, A.J., Gardner, I.A. (2008). Multivariate Mixtures of Polya Trees for Modeling ROC Data. *Statistical Modelling*, to appear.
- Hanson, T.E., Gardner, I.A., Johnson, W.O. (2000). Log-linear and logistic modeling of dependence among diagnostic tests. *Preventive Veterinary Medicine*, 45, 123-137.
- Hanson, T.E., Johnson, W.O., Gardner, I.A. (2003). Hierarchical models for estimating herd prevalence and test accuracy in the absence of a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics*, 8, 223-239.

- Hanson, T.E., Johnson, W.O., Gardner, I.A., Georgiadis, M.P. (2003). Determining the infection status of a herd. *Journal of Agricultural, Biological, and Environmental Statistics*, 8, 469-485.
- Hui, S.L., Walter, S.D. (1980). Estimating the error rates of diagnostic tests. *Biometrics* 36, 167-171.
- Johnson, W.O., Gastwirth, J.L., Pearson, L.M. (2001). Screening without a gold standard: the Hui-Walter paradigm revisited. *American Journal of Epidemiology* 153, 921-924.
- Johnson, W.O., Su, C-L., Gardner, I.A., Christensen, R., 2004. Sample size calculations for surveys to substantiate freedom of populations from infectious agents. *Biometrics*, 60, 165-71.
- Joseph, L., Gyorkos, T.W., Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* 141, 263-272.
- McInturff P, Johnson WO, Cowling D, Gardner IA. (2004). Modelling risk when binary outcomes are subject to error. *Stat Med.* 23, 1095-1107.
- Pepe, M.S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Statistical Science Series: Oxford University Press, New York.
- Pepe, M.S., Longton, G.M., Anderson, G.L., Schummer, M. (2003). Selecting differentially expressed genes from microarray experiments. *Biometrics* 59(1), 133-42.

- Su, C.L., Gardner, I.A., Johnson, W.O. (2004). Diagnostic test accuracy and prevalence inferences based on joint and sequential testing with finite population sampling. *Statistics in Medicine*, 23, 2237-2255.
- Weiner, D.A., Ryan, T.J., McCabe, C.H., Kennedy, J.W., Schloss, M., Tristani, F., Chaitman, B.R., and Fisher, L.D. (1979). Correlations among history of angina, ST-segment response and prevalence of coronary-artery disease in the Coronary Artery Aurgery Study (CASS). *New England Journal of Medicine*, 301, 230-235.
- Wieand, S., Gail, M.H., James, B.R., James, K.L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, 76, 585-592.