True Prevalence from Apparent Prevalence: Obtaining the posterior distribution for true prevalence given diagnostic test results (apparent prevalence) and priors for sensitivity and specificity.

In many cases, interpreting serological surveys of disease is difficult because most diagnostic (or screening) tests have imperfect sensitivity and specificity. Thus, there is a distinction between true prevalence (the proportion of a population that is actually infected) and apparent prevalence (the proportion of the population that tests positive for the disease). Given point estimates for sensitivity (se), specificity (sp), and apparent prevalence (AP), one may calculate true prevalence using the following expression:

$$\text{true prevalence} = (AP+sp-1)/(se+sp-1).$$

Obtaining estimates of true prevalence when sensitivity and specificity are known with uncertainty is more challenging. Given the outcome of a binomial experiment and prior distributions for sensitivity and specificity, the following code can be used to obtain point estimates and probability intervals for true prevalence.

Consider the following example, motivated by hypothetical data for sampling for *Salmonella enteriditis* (SE). Assume that interest centers on estimating true prevalence (**pi**), the predictive value positive (**pvp**), and 1 - the predictive value negative (**OneMinusPVN**).

Let us assume that we randomly sample 100 broilers using fecal culture for SE. Further, let us assume that of the n=100 individuals tested, y=0 test positive. That is, SE was not successfully cultured from any of the 100 birds.

The following model can be used to obtain posterior probabilities of SE shedding, given prior probabilities for the sensitivity (se), specificity (sp), and prevalence (pi) of the test.

let us assume that specificity is almost certainly 1.000. So, we model sp using the following prior:

sp ~ beta(9999,1).

Let us assume that sensitivity is well modeled by a prior where a 90% prior probability interval is (0.30, 0.70), with prior mode (best guess) of 0.50. Such a probability statement corresponds to the following distribution:

se ~ beta(8,8).

Assume that there is effectively no prior information for true prevalence (pi), so the prior for pi is uniform, namely:

pi ~ beta(1,1)

The following model can then be used to obtain posterior distributions of PVP, pi, and 1-PVN:

MODEL

```
Model{
 for(i in 1:1){
  y[i] ~ dbin(ap[i],n[i])
  ap[i] <- se*pi+(1-sp)*(1-pi)
  }
 se ~ dbeta(8, 8)
 sp ~ dbeta(9999, 1)
 pi ~ dbeta(1, 1)
 pvn <- sp*(1-pi)/((1-se)*pi+sp*(1-pi))
 pvp <- se*pi/(se*pi+(1-sp)*(1-pi))
 OneMinusPVN <- 1-pvn
}
```

DATA

```
list(y=c(0),n=c(100))
```

RESULTS

Estimates with 95% central credibility intervals

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
pi	0.02238	0.02437	2.058E-4	**5.185E-4**	**0.01468**	**0.08842**	10000	50001
se	0.4679	0.1249	6.486E-4	**0.2308**	**0.4663**	**0.7123**	10000	50001
sp	0.9999	9.814E-5	8.214E-7	**0.9996**	**0.9999**	**1.0**	10000	50001
pvp	0.9632	0.09053	5.082E-4	**0.7125**	**0.9903**	**0.9998**	10000	50001
OneMinusPVN	0.013	0.01687	1.427E-4	**2.418E-4**	**0.007532**	**0.05828**	10000	50001

These are the posterior distributions