A156 Statistical Methods for Determination of Sample Sizes: The Bayesian Method

Elizabeth A. Gardner, PhD, University of Alabama at Birmingham, Department of Justice, UBOB 210, 1530 3rd Avenue, South, Birmingham, AL 35294-4562; and Shannon L. Crock, MSFS, University of Alabama at Birmingham, 1201 University Boulevard, Birmingham, AL 35205*

After attending this presentation, attendees will understand the principles of the Bayesian method for determining the size of a sample from a population in the context of drug analysis. The drug analyst will be able to explain each of the variables in the Bayesian distribution, calculate the sample size required for a given confidence level and percent of positive samples, and relate the sample size calculated with the Bayesian method to those calculated by the binomial and hypergeometric distributions.

This presentation will impact the forensic science community by presenting an explanation of the Bayesian method for calculating the sample's size at a level appropriate for expert witness testimony in court, providing an understanding that will make it more easily accessible for use in laboratories.

The Bayesian Approach incorporates subjective logic, in which probabilities take into account subjective beliefs about the probability of events. The subjectivity aspect of the Bayesian approach is an advantage; smaller samples are calculated when the analyst's experience indicates that positives are likely. And the opposite is also possible. Larger sample sizes can be calculated when experience leads the analyst to believe that the unknown may not be an illicit drug. However, subjective methods in forensic science have lately come under fire and, finally, the relative complexity of the method itself may provide a barrier to use in the crime lab.

The basic equation for Bayes theorem is: $P(A|B) = P(B|A)P(A)$

$P(B)$

Bayes theory uses condition probabilities, denoted by a vertical line. Therefore, the above equation states the probability that A is true, given that B is true, can be calculated if the probability that A is true, the probability that B is true, and the probability that B is true given that A is true are all known.

This is often rewritten as: $P(\theta|\chi)\ \alpha\ L(\theta|\ \chi)p(\theta)$ where $L(\theta|\ \chi)$ is the likelihood function (the binomial distribution in the case of drug sampling) and $p(\theta)$ is the prior distribution . The prior distribution represents uncertainty about knowledge of θ, the proportion of positives in the population. For the prior distribution, a beta function with parameters *a* and *b* are used. The beta distribution is used in subjective logic, in which probabilities take into account subjective beliefs about the probability of events. The parameters *a* and *b* represent the likelihood of "positive" (*a*) or "negative" (*b*) outcomes.

Guidelines set by the European Network of Forensic Science Institutes (EMFSI) Drugs Working Group recommend that the lower parameter should be set to one. If experience has shown that all 1.15 lb. bricks of plant material are marijuana, then *b* is set to one and *a* is given the value of three or even 10. When there is no prior knowledge, *a* and *b* are both set to one and the calculation equals the binomial distribution.

Plots of the sample size as *a* is held constant and *b* increases and as *b* is held constant and *a* increases, shows that the decrease in sample size with increasing *a* is linear, while an increase in sample size for increasing *b* is not. Finally, the sample sizes calculated by the Bayesian approach will be compared to those calculated by the frequentist methods: the binomial distribution and the hypergeometric distribution.

Bayesian Approach, Statistical Sampling, Illicit Drugs

 ** Presenting Author*