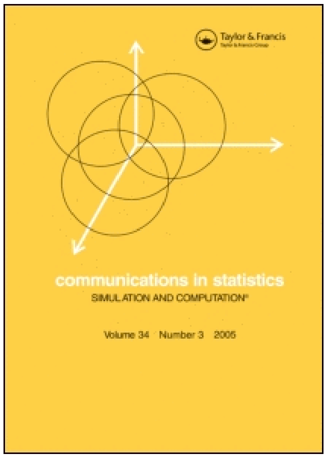

Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597237

Sample Size Determination for Comparing Two Poisson Rates with Underreported Counts

To cite this Article: Stamey, James and Katsis, Athanassios , 'Sample Size Determination for Comparing Two Poisson Rates with Underreported Counts', Communications in Statistics - Simulation and Computation, 36:3, 483 - 492
To link to this article: DOI: 10.1080/03610910701238392
URL: http://dx.doi.org/10.1080/03610910701238392

Communications in Statistics—Simulation and Computation®, 36: 483–492, 2007

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610910701238392

Inference

Sample Size Determination for Comparing Two Poisson Rates with Underreported Counts

JAMES STAMEY[1] AND ATHANASSIOS KATSIS[2]

[1]Department of Statistical Science, Baylor University, Waco, Texas, USA
[2]Department of Social and Educational Policy, University of Peloponnese, Korinthos, Greece

The optimal sample size comparing two Poisson rates when the counts are underreported is investigated. We consider two sampling scenarios. We first consider the case where only underreported data will be sampled and rely on informative prior distributions to obtain posterior identifiability. We also consider the case where an expensive infallible search method and a fallible method are available. An interval based sample size criterion is used in both sampling scenarios. Since the posterior distributions of the two rates are functions of confluent hypergeometric and hypergeometric functions simulation based methods are necessary to perform the sample size determination scheme.

Keywords Average length criterion; MCMC; Underreporting.

Mathematics Subject Classification Primary 62D05; Secondary 62F15.

1. Introduction

The use of the Poisson distribution has always been a valuable tool in statistical modeling due to its wide spectrum of applications ranging from economics and medicine to actuarial science. An informative, albeit not by any means exhaustive, list of relevant contributions can be found in Stamey et al. (2004) as well as in Ntzoufras et al. (2005) and in the references therein. Furthermore, one of the common features of a Poisson setup is the fact that data are often underreported. Whittemore and Gong (1991) consider an epidemiological example of Poisson counts with underreporting. Specifically, they estimate rates of cervical cancer deaths in four European countries accounting for the errors that often occur in death certificates. Similarly, counts are often underreported in environmental and biological data. For instance, Anderson et al. (1994) consider estimation of the rate of gallinule nests in a Southern Louisiana wetland. They account for the fact

Received June 5, 2006; Accepted December 8, 2006
Address correspondence to James Stamey, Department of Statistical Science, Baylor University, Waco, TX 76798-7140, USA; E-mail: james_stamey@baylor.edu

that a cursory airboat search is likely to miss some nests that a thorough search conducted by foot would not miss. Failing to account for this underreporting results in biased estimates and inaccurate sample sizes. This issue is illustrated in determining the required sample size to estimate a single Poisson rate utilizing a $(1-\alpha)100\%$ confidence interval. Assuming that no underreporting exists, a frequentist formula yields

$$n = \left(\frac{2z_{\alpha/2}}{\delta}\right)^2 \hat{\lambda} \tag{1}$$

where δ denotes the interval's width and $\hat{\lambda}$ is an estimate of the unknown rate. The above equation is crucially modified in the case where not all Poisson events are recorded. Practical situations where this may occur include absences from work or certain types of cancer as death causes. If p denotes the probability of reporting the event, where $p < 1$ then (1) becomes

$$n = \left(\frac{2z_{\alpha/2}}{\delta}\right)^2 \frac{\hat{\lambda}}{p}$$

which results in higher values for n. However, rarely is p known with certainty and the normality assumption for the above equations does not always hold.

In this article, the problem of obtaining the optimal sample size for comparing two Poisson rates with underreported counts is investigated. The cases of only fallible data and a double sampling scheme are analyzed separately. Since the reporting probability is usually unknown, the Bayesian approach seems like a natural choice for determining the sample size since it takes into account the uncertainty, which is inherent in any estimate of the unknown parameters. In a Bayesian setting, this uncertainty is expressed through the prior distribution on the parameters of interest. We have opted to use an interval based Bayesian sample size criterion. The reason being it offers a considerable practical advantage over a decision theoretic one. For more details on this issue, see Joseph and Wolfson (1997) and the references therein. Since analytic expressions are often intractable, we resort to Bayesian simulation methods using the R statistical software.

The remainder of the article is organized as follows. In Sec. 2, the complete Poisson modeling is presented along with the necessary computational details. In Sec. 3, the sample size criteria are developed whereas in Secs. 4 and 5, the proposed methodology is applied in specific examples while concluding remarks are provided in Sec. 6.

2. Poisson Model

The Poisson model we consider here is the following:

$$X_i \sim \text{Poisson}(n\lambda_i p_i), \quad i = 1, 2. \tag{2}$$

Here, n is the sample size, sometimes referred to as the opportunity size as it is often an area or length of time, λ_i is the Poisson rate of the ith population and p_i is the probability a particular occurrence is observed in the ith population, referred to as the reporting probability. Assuming the reporting probability is the same in both populations reduces the amount of variability in the estimators and generally would lead to a smaller required sample size, but this could be a strong assumption (for Bayesian binomial hypothesis testing see, among others, Pham Gia and Turkkan, 2003). We consider two sampling cases, one where only fallible data is available and another one where both fallible and more expensive infallible data are available. We note that the observations X_1 and X_2 provide only two degrees of freedom to estimate the four parameters, thus in order to identify all four parameters, either informative priors or validation data are required. We also allow for the second sample size to be a known constant times the first. For instance, expense or inconvenience in sampling may make it desirable to sample more heavily from one population than the other. Specifically, we let $n_2 = kn$.

2.1. *Only Fallible Data*

For the case of only fallible data, the procedure is an extension of Stamey et al. (2004) who determined the required fallible sample size for the one sample case. That is,

$$\lambda_i \sim \text{Gamma}(\alpha_i, \beta_i), \quad i = 1, 2$$
$$p_i \sim \text{beta}(a_i, b_i), \quad i = 1, 2.$$

As mentioned above, since the data only contain enough information to identify two of the four parameters, at least two of the above prior densities are required to be informative. The derived posterior distributions are not in a true "closed form" as they are functions of hypergeometric functions and confluent hypergeometric functions which are infinite sums. Specifically, the posteriors are

$$\pi(\lambda_i \mid x_i) = \frac{(n_i + \beta_i)^{x_i+\alpha_i} \lambda_i^{x_i+\alpha_i-1}}{\Gamma(x_i + \alpha_i)} \frac{{}_1F_1(x_i + a_i, x_i + a_i + b_i, -n_i\lambda_i)}{{}_2F_1(x_i + \alpha_i, b_i, x_i + a_i + b_i, n_i/(n_i + \beta_i))}$$

where ${}_1F_1$ is the confluent hypergeometric function, ${}_2F_1$ is the Gauss hypergeometric function, and $n_2 = kn$. These forms are not particularly useful in a simulation based sample size determination procedure, thus we use the Gibbs sampler to estimate the posterior densities. For the Gibbs sampler, we augment the observable data with the latent variables Z_i which are the unobserved underreported number of occurrences in population i. Combining the Poisson data in (2) with the above conjugate priors and the latent data yields the following joint posterior:

$$\pi(\lambda_1, \lambda_2 p_1, p_2, Z_1, Z_2 \mid x_1, x_2) \propto \prod_{i=1}^{2} p_i^{x_i+a_i-1}(1 - p_i)^{Z_i+b_i-1} e^{-(n_i+\beta_i)\lambda_i} \lambda_i^{x_i+Z_i+\alpha_i-1}. \qquad (3)$$

If the reporting probabilities are assumed to be the same the joint posterior simplifies slightly as there are only three unknown parameters instead of four. To implement the Gibbs sampler, the following full conditionals are required:

$$\lambda_i \mid Z_i, p_i \sim \text{gamma}(x_i + \alpha_i, n_i + \beta_i)$$
$$p_i \mid Z_i, \lambda_i \sim \text{beta}(x_i + a_i, Z_i + b_i)$$
$$Z_i \mid \nu, \tau, \lambda, p, \mathbf{x}, \sim \text{Poisson}(A_i\lambda_i(1 - p_i))$$

where $n_2 = kn$.

After a suitable burn-in, sampling iteratively from the above distributions yields an MCMC approximation to the posterior distribution. From this chain, quantities such as the ratio, λ_1/λ_2 or the difference, $\lambda_1 - \lambda_2$ may be approximated as well.

2.2. *Fallible and Infallible Data*

In some instances it may be possible to conduct both a fallible and infallible search over a small sample and strictly a fallible search over a larger sample. For the sample where both a fallible and infallible count are available we add the following data to the counts in (2):

$$T_i \sim \text{Poisson}(n_{0i}\lambda_i)$$
$$Y_i \mid T_i = t_i \sim \text{binomial}(t_i, p_i)$$

That is, T_i is the number of occurrences in a sample of size n_0 observed by the infallible search method. The variable Y_i is the number of the T_i occurrences observed by the fallible search method. This extra data adds the necessary degrees of freedom to estimate all four parameters without requiring any priors to be informative. Note we again allow the relationship between the sample sizes, $n_{02} = kn_0$, if unequal sample sizes are desired. Adding this data to the posterior distribution in (3) yields

$$\pi(\lambda_1, \lambda_2 p_1, p_2, Z_1, Z_2 \mid x_1, x_2, t_1, t_2, y_1, y_2)$$
$$\propto \prod_{i=1}^{2} p_i^{x_i+y_i+a_i-1}(1-p_i)^{Z_i+t_i-y_i+b_i-1} e^{-(n_i+n_{0i}+\beta_i)\lambda_i} \lambda_i^{x_i+Z_i+t_i+\alpha_i-1}.$$

The Gibbs sampler is used in a similar fashion as in Sec. 2.1 with the following changes in the full conditional distributions due to the additional data:

$$\lambda_i \mid Z_i, p_i, x_i, t_i, y_i \sim \text{gamma}(x_i + Z_i + t_i + \alpha_i, n_i + n_{0i} + \beta_i)$$
$$p_i \mid Z_i, \lambda_i, x_i, t_i, y_i \sim \text{beta}(x_i + y_i + a_i, Z_i + t_i - y_i + b_i)$$
$$Z_i \mid v, \tau, \lambda, p, \mathbf{x}, \sim \text{Poisson}(n_i\lambda_i(1-p_i)).$$

3. Sample Size Determination

Suppose interest lies in the ratio of the two rates, λ_1/λ_2. We now describe a simulation based procedure to determine the appropriate sample size to estimate this quantity subject to satisfying desired criteria. As in the previous section, we handle both the case where only fallible data is available and the case where fallible and infallible data obtain.

3.1. *Only Fallible Data*

For the case of only fallible data we determine the required sample size to obtain a $(1-\alpha)100\%$ posterior interval for the ratio, λ_1/λ_2. We employ an interval-based methodology, termed Average Length Criterion (ALC), that provides fixed coverage intervals. Originally developed in Joseph et al. (1995), is one of several criteria that would be straightforward to implement using our software. Under this technique, we are looking for the smallest sample size such that the average length of all fixed coverage equal tailed posterior intervals does not exceed a pre-specified length. The expectation is taken over the marginal distribution of the data. More specifically, let θ denote the parameter of interest. Thus, we seek n such that

$$\int_X l'(x \mid \theta) m(x) \leq l$$

where $m(x)$ is the marginal distribution of the data, l is the pre-specified length, and $l'(x \mid \theta)$ is the length of the $(1 - \alpha)100\%$ equal tailed interval $R(x)$ derived by finding the $\alpha/2 * 100$ and $(1 - \alpha/2)100$ percentiles such that

$$\int_{R(x)} f(\theta \mid x) d\theta = 1 - \alpha.$$

Though the highest posterior density (hpd) is the narrowest posterior interval, as Wang and Gelfand (2002) point out, the equal tailed interval is much simpler to compute and is invariant with respect to transformations while the hpd interval is not.

There are other interval-based criteria that are also used in practice such as the Average Coverage Criterion (ACC) where the coverage probability is averaged over many posterior intervals of a fixed length. The fact that the ALC deals with fixed coverage intervals (such as 90 or 95%), which is conceptually familiar and appealing to a practitioner, gives the ALC an edge over the ACC.

After choosing the appropriate sample size criterion, an algorithm to implement the procedure is presented below. It is based on a search over a range of possible sample sizes. Note that we set the sample size in the second population, $n_2 = kn$ where n is the sample from the first population for a known k. More specifically, the following steps are proposed:

1. Elicit priors for parameters and specify coverage and desired width.
2. Generate Q values of parameters from these priors.
3. For $j = 1, \ldots, Q$, generate data, $X_{1j} \sim \text{Poisson}(n\lambda_{1j}p_{1j})$, $X_{2j} \sim \text{Poisson}(kn\lambda_{2j}p_{2j})$.
4. For fixed coverage derive the $(1 - \alpha)100\%$ highest posterior density intervals.
5. Average the lengths of the intervals for the Q data sets.
6. Repeat Steps 1–5 for a suitable range of sample sizes.

A curve can then be fit using the sample sizes and corresponding average interval lengths. Fitting this curve with a logarithmic regression yields a formula to find the appropriate sample size. Specifically, the curve $\ln(ALC) = b_0 + b_1 \ln(n)$, with parameters estimated via least squares, fits exceptionally well in all cases, with coefficient of determination over 99.5% in the examples in Sec. 4. The regression equation can then be solved for n to determine the appropriate sample size, specifically,

$$n = \exp\left\{\frac{\ln(ALC) - b_0}{b_1}\right\}.$$

3.2. *Fallible and Infallible Data*

In some cases there may be an expensive "error free" search procedure that could be done over a small area along with the less expensive "error prone" method. For a fixed budget, we seek to minimize the cost while achieving the desired interval width. So here we have the constraint

$$B = 2 \cdot C_1 \cdot (n + n_0) + 2 \cdot C_2 \cdot n_0$$

where B is the total budget; C_1 is the cost of a fallible unit; C_2 is the cost of an infallible unit; n is the amount of area or time searched with the fallible method; n_0 is the amount of area or time searched with the infallible method and the fallible method.

So, given a fixed budget, D, we can solve for n in terms of n_0:

$$n = (D - 2C_2 n_0 - 2C_1 n_0)/(2C_1)$$

So now for the sample sizes that satisfy the cost constraint, we can perform a grid search to find the combination that minimizes the posterior variance. The procedure is similar for the case with just fallible data and is outlined as follows:

1. Elicit priors for parameters.
2. Generate Q values of parameters from these priors.
3. For $j = 1, \ldots, Q$, generate data;
   Fallible counts: $X_1 \sim \text{Poisson}(n\lambda_1 p_1)$, $X_2 \sim \text{Poisson}(kn\lambda_2 p_2)$,
   Infallible counts: $Y_1 \sim \text{Poisson}(n_0\lambda_1)$, $Y_2 \sim \text{Poisson}(kn_0\lambda_2)$
   Fallible counts in validation data: $Z_1 \sim \text{binomial}(Y_1, p_1)$, $Z_2 \sim \text{binomial}(Y_2, p_2)$.
4. For fixed coverage derive the $(1 - \alpha)100\%$ intervals.
5. Average the lengths of the intervals for the Q data sets. Perform search over possible sample sizes (given cost constraint) to find value that minimizes variance.

It is important to note that if only fallible data is used, informative priors are required for finding the posterior distributions. For the case with validation data the model is identifiable without informative priors, thus the elicited priors can be used for generating the parameters for the simulation but are not necessary for analyzing each data set which leads to a more "objective" analysis. Thus for this case, non informative priors, such as the Jeffreys prior, can be used in analyzing each generated data set (for practical considerations on this matter see also Joseph et al., 1997). In this instance we make the distinction between *sampling* and *fitting* priors discussed in Wang and Gelfand (2002). The sampling priors are those used to generate the parameters for the simulation study, which should generate a sufficiently rich range of values for each parameter while the fitting priors are actually used in the analysis. In most Bayesian sample size schemes for misclassified data, the sampling and fitting priors are the same, for instance, Dendukuri et al. (2004), Rahme et al. (2000), and Stamey et al. (2005), but with validation data we are not required to use informative priors for the analysis.

4. Cervical Cancer Example

Our first example is motivated by the data in Whittemore and Gong (1991). The data consists of the number of reported cervical cancer deaths in four European countries along with survey information about the reporting probability. The survey data consists of a sample of doctors from each country who were asked to complete a death certificate for a described case of someone known to have died from cervical cancer. Combining this survey data with the fallible counts from each country, Whittemore and Gong (1991) estimated the parameters of a Poisson regression with underreporting. Here, we use the survey information from Belgium and France as prior distributions for the reporting probabilities in order to determine what sample size, which in this case is the number of person-years, is required to obtain an interval estimate of the ratio of these two rates of a desired length.

From the survey results in Whittemore and Gong (1991), 43 out of 50 doctors from Belgium correctly classified the case as cervical cancer while 38 out of 53 doctors from France correctly classified the case. If we consider these two binomial likelihoods as functions of the parameters p_1 and p_2, beta kernels obtain yielding priors $p_1 \sim \text{beta}(43, 7)$ and $p_2 \sim \text{beta}(38, 15)$. For the Poisson rates we assume $\lambda_1 \sim \text{gamma}(3.3, 0.93)$ and $\lambda_2 \sim \text{gamma}(2.82, 1)$, where the units is in 10,000 person years. These priors are somewhat arbitrary, but both are based on information equivalent to a sample of approximately 10,000 person years where the expert suspects a little over 3 deaths per 10,000 person years in Belgium and just under 3 deaths per 10,0000 person years in France. The prior interval for λ_1/λ_2 is (0.23, 7.78). Suppose an interval of width 3 is desired. All results are based on $Q = 5{,}000$ generated parameter/data sets and each iteration is based on a Gibbs sampler with a 500 iteration burn in and 5,000 Gibbs iterations. Figure 1 has the average length for a range of sample sizes for the above prior specification (Prior I) and a second prior where the beta parameters are divided by two (Prior II), that is $p_1 \sim \text{beta}(21.5, 3.5)$ and $p_2 \sim \text{beta}(19, 7.5)$ (Prior II). For both priors we consider the case of equal sample sizes ($k = 1$) and the case where $k = 1.5$, thus the sample size from the second population will be 50% larger than the sample size from the first. As a specific example, for $k = 1$ and Prior I the resulting regression curve is $\ln(\text{ALC}) = -.361\ln(n) + 2.100$ with $r^2 = .999$. This provides a value of 16.02 for the estimated required sample size, which indicates sampling should be done for 160,200 person-years. Proceeding similarly, we find that for Prior II, the required sample size is 21.62, thus sampling should be done for 216,200 person-years. For Prior 1

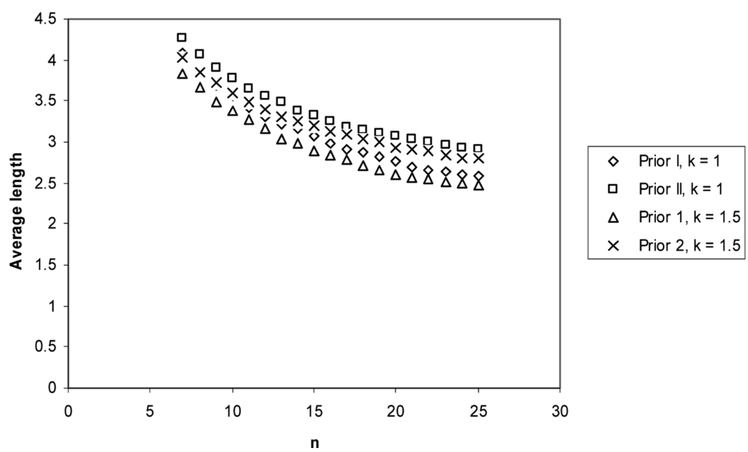

Figure 1. Average lengths for Example 1.

with $k = 1.5$ we find the required sampling time to be 138,400 person years from population 1 and 207,600 from population 2, while Prior II requires 189,000 person years from population 1 and 283,500 from population 2. It is interesting that for Prior I a total sampling time of 346,000 person years is required for $k = 1.5$ while the equal sample size case requires a total sampling time of 320,400. Of course, even though for the case of $k = 1.5$ the total sample size is larger, there are numerous reasons for not using equal sample sizes. For instance, in a rare disease it may be more difficult or expensive to sample from one of the populations.

5. Gallinule Example

Our second example is motivated by studies such as Anderson et al. (1994). Suppose interest lies in comparing the rates of Common Gallinule and Purple Gallinule nests in Southern Louisiana. These species are marsh birds generally found in the southeastern United States. Anderson et al. (1994) performed both a thorough search by foot and a fallible airboat "pass-through" over 500 linear feet while performing only the airboat search over an additional 4,300 feet to estimate these rates. Suppose in a future study, an expert believes the rate of Common Gallinule nests can be described by a gamma(14, 2) while the rate of Purple Gallinule nests can be described with a gamma(8, 2). Also, the expert believes the probability of spotting a Common Gallinule nest can be described with a beta(20, 5) while the Purple Gallinule visibility probability can be modeled with a beta(17, 8). Suppose also that $1,000 is available for the study and that a fallible search costs $2 per 100 feet while the error free search costs $25 per 100 feet. We used the priors only for parameter generation. For the analysis of each data set, we used beta(1, 1)'s for both reporting probabilities and the Jeffreys prior, $\pi(\lambda_j) \approx \frac{1}{\sqrt{\lambda_j}}$, $j = 1, 2$. We again based results on $Q = 5{,}000$ generated parameter/data sets and each iteration is based on a Gibbs sampler with a 500 iteration burn in and 5,000 Gibbs iterations. Figure 2 provides the average width for various sample sizes for this fixed cost problem for $k = 1$ and $k = 1.5$. Fitting the curve $\ln(ALC) = b_0 + b_1 n_0 + b_2 n_0^2$ we find that the minimum width occurs for 1,233 error free feet and an additional 8,354 fallible feet

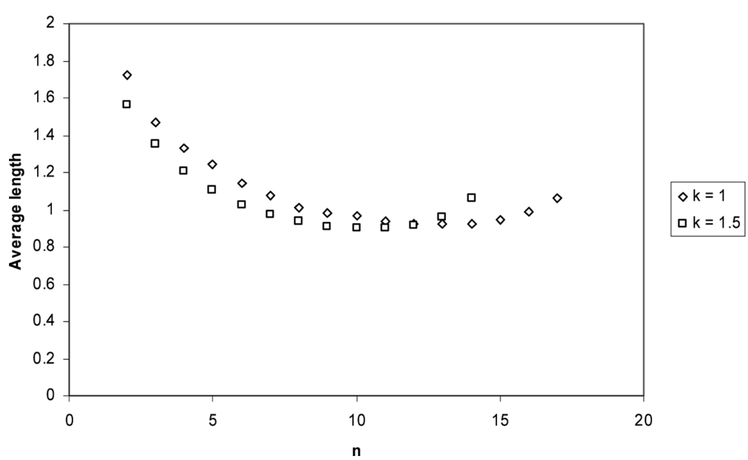

Figure 2. Average length for second example for $k = 1$ and 1.5.

which would yield an average width of 0.91 for $k = 1$. Similarly, we find the optimal allocation for $k = 1.5$ to be 993 error-free feet from the first population, 1,489 error-free feet from the second population, 6,594 fallible feet from the first population, and 9,891 from the second population. This would yield an average length of 0.90, slightly narrower than the equal sample size scenario average length of 0.91. We note that if all the money was put in the infallible data (2,000 feet for each sample), an average width of 1.29, demonstrating that in this case a mixture of fallible and infallible data is the optimal solution.

6. Conclusions

Epitomizing the above study consists of a fully Bayesian approach to the problem of obtaining the optimal sample size for comparing two Poisson rates with underreported counts. The cases of only fallible data and a double sampling scheme are developed separately using the ALC interval based criterion. The Gibbs sampler is used to estimate the posterior densities. The above-mentioned methodology uses equal sample sizes in both populations. The algorithms do not change when this fact does not hold, only the numerical search becomes more thorough. All the code used in the article was implemented in the R software and is available from the authors upon request.

Although we have considered biomedical examples, comparison of Poisson rates is present in such diverse areas as actuarial claim counts and detection of "spam" E-mail messages. In the former case, an insurance company wishes to examine the rates of accident claims between two or more of its branches (or types of customers) outlining a fairer premium policy. For the latter case, comparing Bayesian classifiers developed for the detection of "spam" E-mail (see Schechter, 2003) is of crucial importance to the manufacturer. Each classifier consists of a software program that declares an e-mail message as "spam", based on the frequency of certain words. In both cases, specifying satisfactory prior distributions will enable the practitioner to obtain the proper sample size in order to achieve the desired precision.

We firmly believe that the study of sample size determination with Poisson underreporting rates is an area with a great deal of applications that merits further research interest. Extending to the case of more than two Poisson rates and modeling the dependence among the reporting probabilities are uncharted research paths. It is imperative to aim at further advancing the sample size determination theory since additional statistical techniques will arm practitioners with a great arsenal of data analytic tools to accurately tackle this ever-important area of applied statistics.

References

Anderson, C., Bratcher, T., Kutran, K. (1994). Bayesian estimation of population density and visibility. *Texas Journal of Science* 46:7–12.

Dendukuri, N., Rahme, E., Bélisle, P., Joseph, L. (2004). Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test. *Biometrics* 60:388–397.

Evans, M., Guttman, I., Haitovsky, Y., Swartz, T. (1996). Bayesian analysis of binary data subject to misclassification in Bayesian analysis. In: Berry, D., Chaloner, K., Gweke, J., eds. *Statistics and Econometrics: Essays in Honor of Arnold Zellner*. New York: Wiley, pp. 67–77.

Joseph, L., Wolfson, D. (1997). Interval-based versus decision theoretic criteria for the choice of the sample size. *The Statistician* 46:145–149.

Joseph, L., Wolfson, D., du Berger, R. (1995). Sample size calculations for binomial proportions via hightest posterior density intervals. *The Statistician* 44:143–154.

Joseph, L., du Berger, R., Bélisle, P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statistics in Medicine* 16(7):769–781.

Ntzoufras, I., Katsis A., Karlis, A. (2005). Bayesian assessment of the distribution of insurance claim counts using the reversible jump algorithm. *North American Actuarial Journal* 9:90–105.

Pham Gia, T., Turkkan, N. (2003). Determination of exact sample sizes in the Bayesian estimation of the difference of two proportions. *The Statistician* 52:131–150.

Rahme, E., Joseph, L., Gyorkos, T. (2000). Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. *Applied Statistics* 49:119–128.

Schechter, B. (2003). Spambusters. *New Scientist* 2385:42.

Stamey, J. D., Seaman, J. W., Young, D. M. (2004). Bayesian sample size determination for estimating Poisson rate with underreported data. *Communcations in Statistics – Simulation and Computation* 33:341–354.

Stamey, J. D., Seaman, J. W., Young, D. M. (2005). Bayesian sample size determination for inference on two binomial populations with no gold standard. *Statistics in Medicine* 24:2963–2976.

Tenenbein, A. (1970). A double sampling scheme for estimating from binomial data with misclassification. *Journal of American Statistist Association* 65:1350–1361.

Wang, F., Gelfand, A. E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science* 17:193–208.

Whittemore, A. S., Gong, G. (1991). Poisson regression with misclassified counts: application to cervical cancer mortality rates. *Applied Statistics* 40:81–93.