Bayesian and Likelihood Methods in Statistics and Econometrics
S. Geisser, J.S. Hodges, S.J. Press and A. Zellner (Editors)

INFERENCE WITH IMPERFECT SAMPLING FROM A BERNOULLI PROCESS

Robert L. Winkler and Anil Gaba

Duke University

Abstract

When a sample is taken from a dichotomous process, various sources of noise may cause some observations to be classified incorrectly. In this paper, we consider imperfect sampling from a Bernoulli process with a noise parameter that is not known. A likelihood analysis reveals an identification problem, which is avoided under a Bayesian analysis with a joint prior distribution on the noise parameter and the Bernoulli proportion. An example is presented to illustrate the methodology and to provide some flavor of the implications of the noise and our uncertainty about the noise for inferences about a proportion.

1. Introduction

Data are frequently generated from dichotomous processes, with the parameter of interest being a proportion. In practice, inferences about a proportion p are almost always based on the assumption that the data-generating process is Bernoulli with parameter p. The maximum likelihood estimator is then the sample proportion, and a conjugate Bayesian analysis involves beta prior and posterior distributions.

When a sample is taken from a simple Bernoulli process, various sources of noise may cause some observations to be incorrect. For example, responses to dichotomous survey questions may be incorrect because some respondents lie or misinterpret the questions; observations of good and defective items in a quality control setting or positive and negative results in medical tests may be incorrect because of imperfect testing. Even with no lying, no misinterpretations, and perfect testing, errors in recording, coding, and handling the data provide sources of noise. As a result, the process that is actually observed may be somewhat different than the process that is of interest because of such contaminating factors.

The problem of making inferences about a Bernoulli parameter p based on imperfect sampling has received limited attention in the literature. Some work involving measurement error models (e.g., Fuller, 1987) is in the same spirit but emphasizes normal models. Research on classification models is relevant, although the focus is typically on classifying individual items instead of making inferences about proportions (however, see Press, 1968). The area of randomized response sampling (e.g., Warner, 1965) uses models similar to those considered here, but the noise is intentionally introduced to the process and is therefore known and carefully controlled (in an attempt to suppress the unknown and uncontrollable noise that might otherwise be present.)

Inferences for a noisy Bernoulli process are studied in Winkler (1985) under the assumption that the noise parameters are known. The results of likelihood and Bayesian analyses indicate that the noise can have considerable impact on inferences that are made about the Bernoulli parameter. For some values of the noise parameters, the noise leads to large shifts in point estimates and greatly increases the uncertainty about the parameter of interest. In a Bayesian analysis, the reduction in effective sample size due to the noise results in more weight being given to the prior distribution.

In this paper, we consider the case of imperfect sampling from a Bernoulli process with a single noise parameter under the assumption that the noise parameter is not known. The basic model is presented and a likelihood analysis is discussed in Section 2. The presence of the noise parameter and its interactions with the Bernoulli parameter lead to difficulties in the likelihood analysis. A full Bayesian analysis with a joint prior distribution on the noise parameter and the Bernoulli parameter is developed in Section 3. Joint and marginal posterior distributions can be expressed as mixtures and related to our uncertainty about the true status of the elements in the sample. Some examples given in Section 4 illustrate the methodology and provide some flavor of the implications of the noise and our uncertainty about the noise. The prior distribution assumes an even greater role than it does when the noise parameter is known. Section 5 concludes the paper with a brief summary and discussion.

2. A Likelihood Analysis

Suppose that each member of a large population is in either Group A or Group B, but not both, and let p denote the proportion of the population in Group A. For example, Group A might consist of items from a production process that are defective, individuals who will purchase a certain brand, individuals with a particular disease, or voters who will vote for a given candidate. A random sample is taken from the population, and each member of the sample is classified as being in Group A or Group B. The classification is imperfect. Testing of items from a production process or medical patients yields less than perfect information, as do market research surveys or pre-election polls. We denote the probability that a member of the sample is classified incorrectly by λ. Thus, p is the Bernoulli parameter of interest, and λ is the noise parameter in our model.

Our inferences are based on a random sample of size n, with r denoting the number of members of the sample classified as being in Group A. Let $x_i = 1$ if the ith member of the sample is classified as being in Group A and $x_i = 0$ if classified as Group B. Then the process we actually observe is not Bernoulli in p, but is Bernoulli in q, where

$$q = P(x_i = 1 \mid p, \lambda) = p(1-\lambda) + (1-p)\lambda \tag{1}$$

and

$$1 - q = P(x_i = 0 \mid p, \lambda) = p\lambda + (1-p)(1-\lambda). \tag{2}$$

The likelihood function is thus of the form

$$\ell(r, n \mid p, \lambda) = [p(1-\lambda) + (1-p)\lambda]^r [p\lambda + (1-p)(1-\lambda)]^{n-r}. \tag{3}$$

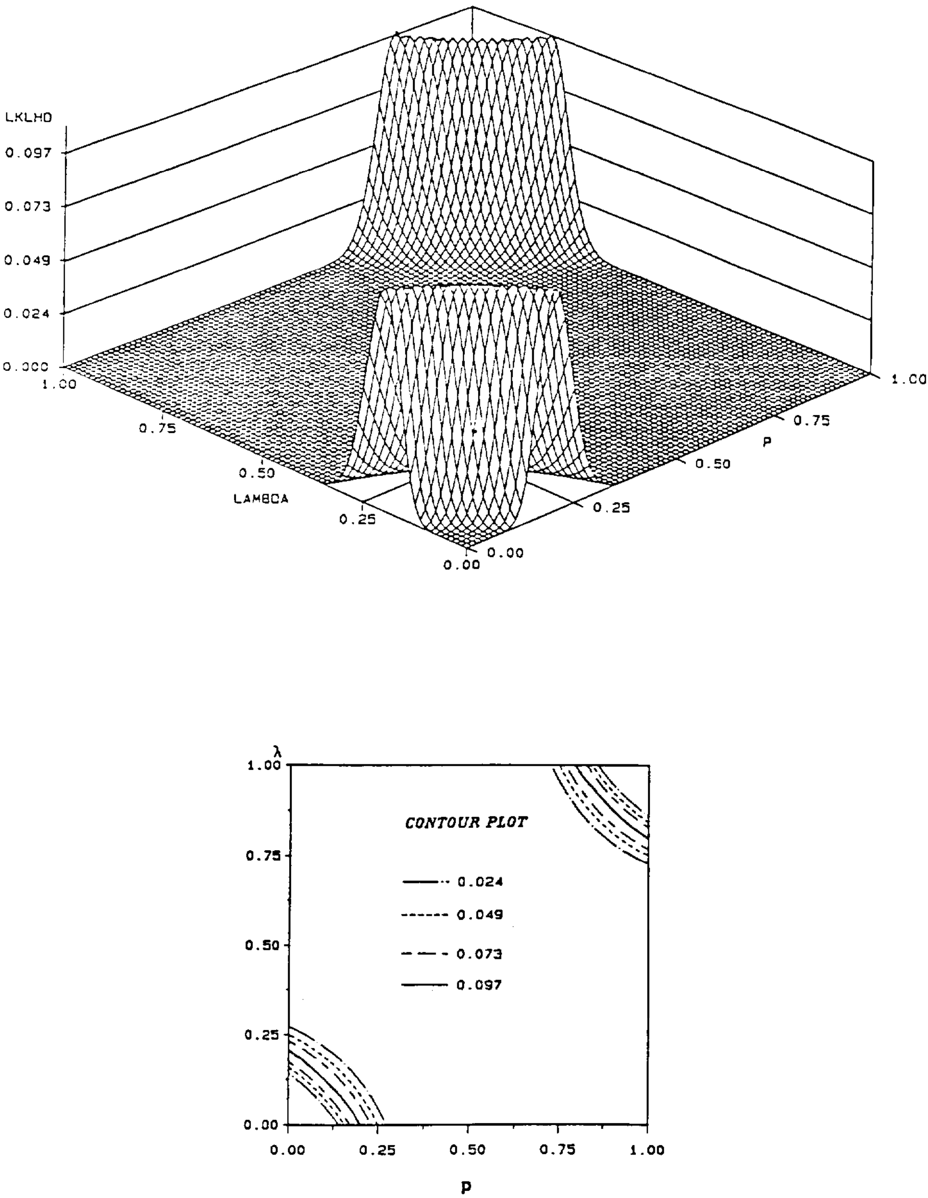
By expanding the terms in (3), we obtain an alternate representation for the likelihood function:

$$\ell(r, n \mid p, \lambda) = \sum_{j=0}^{r} \sum_{t=0}^{n-r} \binom{r}{j} \binom{n-r}{t} p^{n-j-t} (1-p)^{j+t} \lambda^{n+j-r-t} (1-\lambda)^{r-j+t}. \tag{4}$$

In this expression, j can be interpreted as the number of items classified in Group A but really in Group B, and t can be interpreted as the number of items classified in Group B and actually in Group B. Thus, j of the r classified in Group A and n-r-t of the n - r classified in Group B are misclassified. Of course, we do not know the values of j and t, and that is precisely what complicates our inferential problem. If we knew j and t, the likelihood function would be separable in p and λ: a product of Bernoulli likelihoods. With imperfect sampling, we lose this separability and wind up with a mixture of products of Bernoulli likelihoods.

The likelihood function is shown in Figure 1 for r = 21 and n = 104, the sample statistics for an example discussed in Section 4. This illustrates the identification problem that arises in a likelihood analysis of this noisy process. The likelihood function is unable to distinguish among all points (p, λ) with the same value of q from (1). As a result, the maximum likelihood estimate of (p, λ) is not unique, but

Figure 1: Likelihood Function and Likelihood Contour Plot with τ = 21 and n = 104



consists of those points (p, λ) such that $p(1-\lambda)+(1-p)\lambda = r/n$. For example, the set of maximum likelihood estimates for $r = 21$, $n = 104$ is shown in the contour plot in Figure 1. This set includes points such as $p = 0$ with a 20% misclassification rate, $p = 0.1$ with a 13% misclassification rate, $p = 0.2$ with no misclassification, $p = 0.8$ with a 100% misclassification rate, and $p = 1$ with an 80% misclassification rate. Without any further information about p or λ, alternative explanations for a given set of sample results seem equally compelling.

3. A Bayesian Analysis

With imperfect sampling from a Bernoulli process, we would expect to have some relevant prior information. In most cases, for instance, we would probably view high values of λ as quite unlikely. Thus, in a survey with noise due to possible lying on the part of the respondents, $(p, \lambda) = (0.8, 0.9)$ and (0.9,0.8) would have the same likelihood as (0.1,0.2) and (0.2,0.1), but we might dismiss the former two pairs because lying rates of 80-90% seem extremely unrealistic. A Bayesian analysis enables us to formalize our prior judgments and can provide the information necessary to avoid identification problems that arise in the likelihood analysis. Furthermore, it is not just a matter of incorporating additional information about p; information about λ is just as valuable in our model with imperfect sampling.

In the case of perfect sampling, the family of beta distributions is conjugate with respect to a Bernoulli process, and beta distributions are viewed as satisfactory approximations for a wide variety of types of prior information concerning a proportion. Both p and λ are proportions in our model with imperfect sampling, and we will assume that the prior density for (p, λ) is a product of beta densities for p and λ:

$$f(p, \lambda) = f_\beta(p) f_\beta(\lambda) \propto p^{\alpha_1 - 1}(1-p)^{\beta_1 - 1}\lambda^{\alpha_2 - 1}(1-\lambda)^{\beta_2 - 1}, \tag{5}$$

with α_1, α_2, β_1, and $\beta_2 > 0$. The independence assumption is not unreasonable for this symmetric model with a single noise parameter. Any dependence between p and λ might be expected to be negative, but the symmetry tempers this because a given value of λ is associated with both p and $1-p$.

Given the likelihood function in (3) or (4) and the prior distribution in (5), an application of Bayes' theorem yields the following posterior distribution:

$$f(p, \lambda \mid r, n) = \sum_{j=0}^{r}\sum_{t=0}^{n-r} w_{jt} f_\beta(\lambda \mid \alpha_2 + n - r - t + j, \beta_2 + r - j + t)$$

$$\cdot f_\beta(p \mid \alpha_1 + n - j - t, \beta_1 + j + t), \tag{6}$$

where

$$w_{jt} = a_{jt} \Big/ \sum_{j=0}^{r} \sum_{t=0}^{n-r} a_{jt} \tag{7}$$

and

$$a_{jt} = \binom{r}{j}\binom{n-r}{t}$$

$$\frac{(\alpha_2 + n - r - t + j - 1)!(\beta_2 + r - j + t - 1)!(\alpha_1 + n - j - t - 1)!(\beta_1 + j + t - 1)!}{(\alpha_2 + \beta_2 + n - 1)!(\alpha_1 + \beta_1 + n - 1)!}. \tag{8}$$

Thus, the posterior distribution is a mixture of products of beta distributions. The interpretation of the weight *Wjt* is that it is the posterior probability that *j* of the *r* classified in Group A and n - *r* - *t of the n* - *r* classified in Group B are misclassified.

Our primary interest focuses on p. If we integrate $f(p, \lambda \mid r, n)$ with respect to λ, we arrive at the following marginal posterior density for p:

$$f(p \mid r, n) = \sum_{s=0}^{n} v_s f_\beta(p \mid \alpha_1 + n - s, \beta_1 + s), \tag{9}$$

where

$$v_s = \sum_{(j,t)\ni j+t=s} w_{jt} = P(s \mid r, n) \tag{10}$$

is the posterior probability that there are really *s* members of Group B in the sample. The posterior mean of *p* is

$$E(p \mid r, n) = \sum_{s=0}^{n} v_s \left(\frac{\alpha_1 + n - s}{\alpha_1 + \beta_1 + n} \right), \tag{11}$$

and the posterior variance can be found from $E(p \mid r, n)$ and

$$E(p^2 \mid r, n) = \sum_{s=0}^{n} v_s \frac{(\alpha_1 + n - s + 1)(\alpha_1 + n - s)}{(\alpha_1 + \beta_1 + n + 1)(\alpha_1 + \beta_1 + n)}. \tag{12}$$

We can also determine how the sample information modifies the distribution of λ. The marginal posterior density for λ is

$$f(\lambda \mid r, n) = \sum_{q=0}^{n} u_q f_\beta(\lambda \mid \alpha_2 + n - q, \beta_2 + q), \tag{13}$$

where

$$u_q = \sum_{(j,t)\ni r+t-j=q} w_{jt} = P(q \mid r, n) \tag{14}$$

is the posterior probability that there are really q correctly classified elements and $n-q$ misclassifications in the sample. The posterior mean of λ and $E(\lambda^2 \mid r,n)$ are

$$E(\lambda \mid r,n) = \sum_{q=0}^{n} u_q \left(\frac{\alpha_2 + n - q}{\alpha_2 + \beta_2 + n} \right) \tag{15}$$

and

$$E(\lambda^2 \mid r,n) = \sum_{q=0}^{n} u_q \frac{(\alpha_2 + n - q + 1)(\alpha_2 + n - q)}{(\alpha_2 + \beta_2 + n + 1)(\alpha_2 + \beta_2 + n)}. \tag{16}$$

In general, then, the joint and marginal densities can be expressed as mixtures of the possible posterior densities that could arise under perfect knowledge of the exact number of misclassifications of each type. The shapes and moments of these mixtures will of course depend on the prior parameters α_1, β_1, α_2, and β_2 as well as the sample statistics r and n. Numerical examples in the following section illustrate some possible shapes for the posterior distributions and give an indication of possible implications of the noise parameter λ and our uncertainty about λ on inferences about p.

4. Examples

Self-reported incidents of delinquent behavior by youths provide examples of imperfect sampling. Clark and Tifft (1966) administered a questionnaire asking college students whether they had committed certain delinquent acts and then conducted extensive interviews and polygraph tests to check the validity of the answers to the questionnaire. Estimated misclassification probabilities ranged from zero to 0.625, with many questions resulting in equal or almost equal misclassification probabilities (suggesting that our symmetric model with one noise parameter might be appropriate in such cases) and many other questions resulting in substantial discrepancies between overreporting and underreporting (indicating that an asymmetric model with two noise parameters is sometimes more appropriate).

To illustrate the symmetric model, we chose a question with equal estimated misclassification probabilities: "Have you 'beaten up' on someone who hadn't done anything to you?" The same question was asked by Gould (1969) in a later study, and we use Gould's results ($r = 21, n = 104$) as our sample data. Considering the information from Clark and Tifft (1966) for this question and attempting to take into account differences in the populations being sampled in the two studies, we assessed prior distributions for p and λ with $\alpha_1 = 2$, $\beta_1 = 8$, $\alpha_2 = 2$, and $\beta_2 = 18$. The expected proportion who had beaten up someone is 0.20, and the expected misclassification probability is 0.10; the corresponding standard deviations are 0.12 and 0.07, respectively. The joint prior density is shown in Figure 2.

The joint posterior density determined from (6) is also shown in Figure 2. Note that in the posterior distribution, the second hump in the likelihood function (see Figure 1) disappears because most of the prior probability is concentrated on regions of low values of p and λ. The posterior distribution is unimodal despite the multimodality of the likelihood function. The posterior means for p and λ are 0.14 and 0.09, with standard deviations of 0.05 and 0.05. Thus, we estimate that 14% of the members of the population have beaten up on someone and 9% would lie when asked the question.

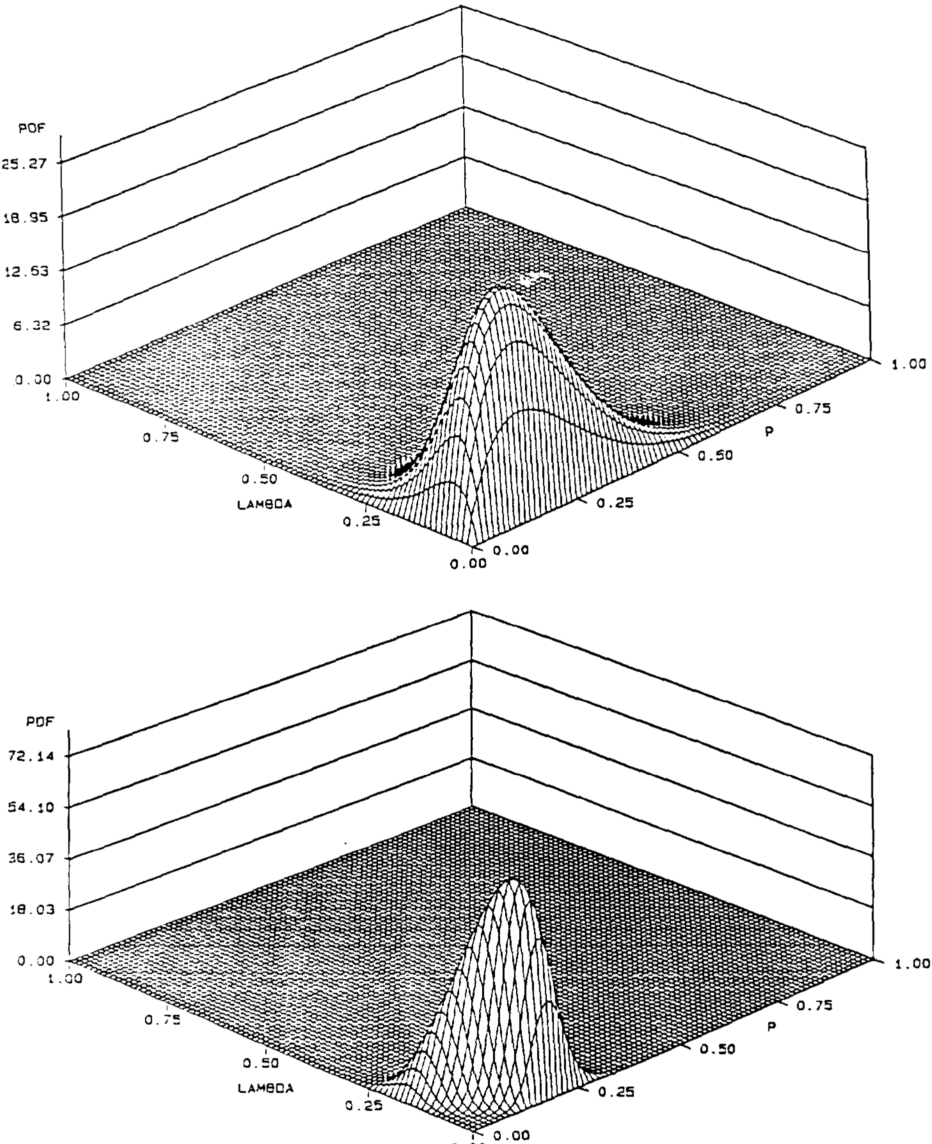
As noted in Section 1, inferences about a proportion p are usually based on the assumption of a Bernoulli process in p. An interesting question, therefore, is how inferences based on our model recognizing imperfect sampling differ from those based on an assumption of perfect sampling (i.e., $\lambda = 0$). With $\lambda = 0$ and the same prior distribution for p, the posterior distribution for p is a beta distribution with a mean that is 0.06 higher than our posterior mean under imperfect sampling and a standard deviation that is smaller than the posterior standard deviation under imperfect sampling by a factor of 1.45. The impact of the noise is to shift r/n toward 0.5. For a given r/n, then, the model with noise allows for this impact by placing more weight on more extreme values of p (lower values in our example because the prior mean and r/n are less than one-half), thereby shifting the posterior mean further away from 0.5.

The fact that the posterior standard deviation is larger under the imperfect-sampling model suggests that the noise causes a reduction in the information content of the sample regarding p. Another way to investigate the information content is to determine an “effective sample size” for the noisy sample. We generate this value by fitting beta distributions to the marginal posterior distributions of p and λ and defining the effective sample size as

$$n_i^* = (\alpha_i^* + \beta_i^*) - (\alpha_i + \beta_i), \tag{17}$$

where $i = 1$ for p, $i = 2$ for λ, and the parameters of the posterior beta fits are α_i^* and β_i^* for $i = 1, 2$. Such beta fits may be somewhat crude in some cases, since the marginal posterior densities cannot always be closely approximated by beta distributions, but n_i^* still seems useful as a rough measure of effective sample size. In our example, the effective sample sizes are 30.4 for p and 13.0 for λ. When compared with $n = 104$, the effective sample size for p indicates a considerable loss of information due to the noise, and the effective sample size for the noise parameter itself is even smaller.

Figure 2: Prior and Posterior Distributions with
$\alpha_1 = 2,\ \beta_1 = 8,\ \alpha_2 = 2,\ \beta_2 = 18,\ r = 21,$ **and** $n = 104$



The impact of the noise, then, is substantial in our example: a reduction of about 30% in the posterior mean and an increase of 45% in the standard deviation, as compared with an analysis assuming perfect sampling. Since the prior distribution plays a particularly important role in the imperfect-sampling analysis, the robustness of inferences to variations in the prior distribution is of interest. For p, we considered one tighter distribution ($\alpha_1 = 4$, $\beta_1 = 16$), and for those who prefer to use diffuse prior distributions, we obtained results for $\alpha_1 = \beta_1 = 1$, a uniform distribution. For λ, we varied the prior mean from 0.10 to 0.05 and 0.15 and $\alpha_2 + \beta_2$ from 20 to 10 and 40. (A diffuse prior distribution for λ seems highly unrealistic.)

TABLE 1: Posterior Means (Standard Deviations) for p and λ in "Have you beaten up..." Example

			(α_2, β_2)				
			(1,9)	(1,19)	(2,18)	(3, 17)	(4,36)
(α_1, β_1)	(1,1)	p	0.15 (0.06)	0.17 (0.06)	0.14 (0.06)	0.11 (0.06)	0.13 (0.06)
		λ	0.08 (0.06)	0.05 (0.05)	0.09 (0.05)	0.13 (0.05)	0.10 (0.04)
	(2,8)	p	0.15 (0.06)	0.17 (0.05)	0.14 (0.05)	0.12 (0.05)	0.14 (0.05)
		λ	0.08 (0.05)	0.05 (0.04)	0.09 (0.05)	0.12 (0.05)	0.10 (0.04)
	(4,16)	p	0.16 (0.05)	0.18 (0.04)	0.16 (0.05)	0.14 (0.05)	0.15 (0.05)
		λ	0.07 (0.05)	0.05 (0.04)	0.08 (0.05)	0.11 (0.05)	0.09 (0.04)

The results, which are presented in Tables 1-3, provide some (albeit limited) evidence about the behavior of the imperfect sampling model and the sensitivity of estimates of p and λ as well as the posterior level of uncertainty to variations in

TABLE 2: Differences in Posterior Means for p With and Without Noise (Ratios of Posterior Standard Deviations for p With and Without Noise) in "Have you beaten up.. ." Example

		(α_2, β_2)				
		(1,9)	(1,19)	(2,18)	(3,17)	(4,36)
	(1,1)	-0.06 (1.65)	-0.04 (1.42)	-0.07 (1.59)	-0.09 (1.59)	-0.07 (1.50)
(α_1, β_1)	(2,8)	-0.05 (1.50)	-0.03 (1.35)	-0.06 (1.45)	-0.08 (1.45)	-0.07 (1.38)
	(4,16)	-0.04 (1.33)	-0.03 (1.23)	-0.05 (1.31)	-0.06 (1.32)	-0.05 (1.27)

TABLE 3: Effective Sample Sizes for $p(\lambda)$ in "Have you beaten up..." Example

		(α_2, β_2)				
		(1,9)	(1,19)	(2,18)	(3,17)	(4,36)
	(1,1)	27.1 (10.1)	42.6 (3.2)	27.8 (10.1)	22.9 (19.4)	30.4 (8.6)
(α_1, β_1)	(2,8)	29.5 (12.1)	44.1 (5.1)	30.4 (13.0)	25.6 (22.0)	33.1 (12.7)
	(4,16)	39.2 (14.0)	53.1 (8.1)	39.1 (15.7)	33.5 (23.5)	40.9 (16.6)

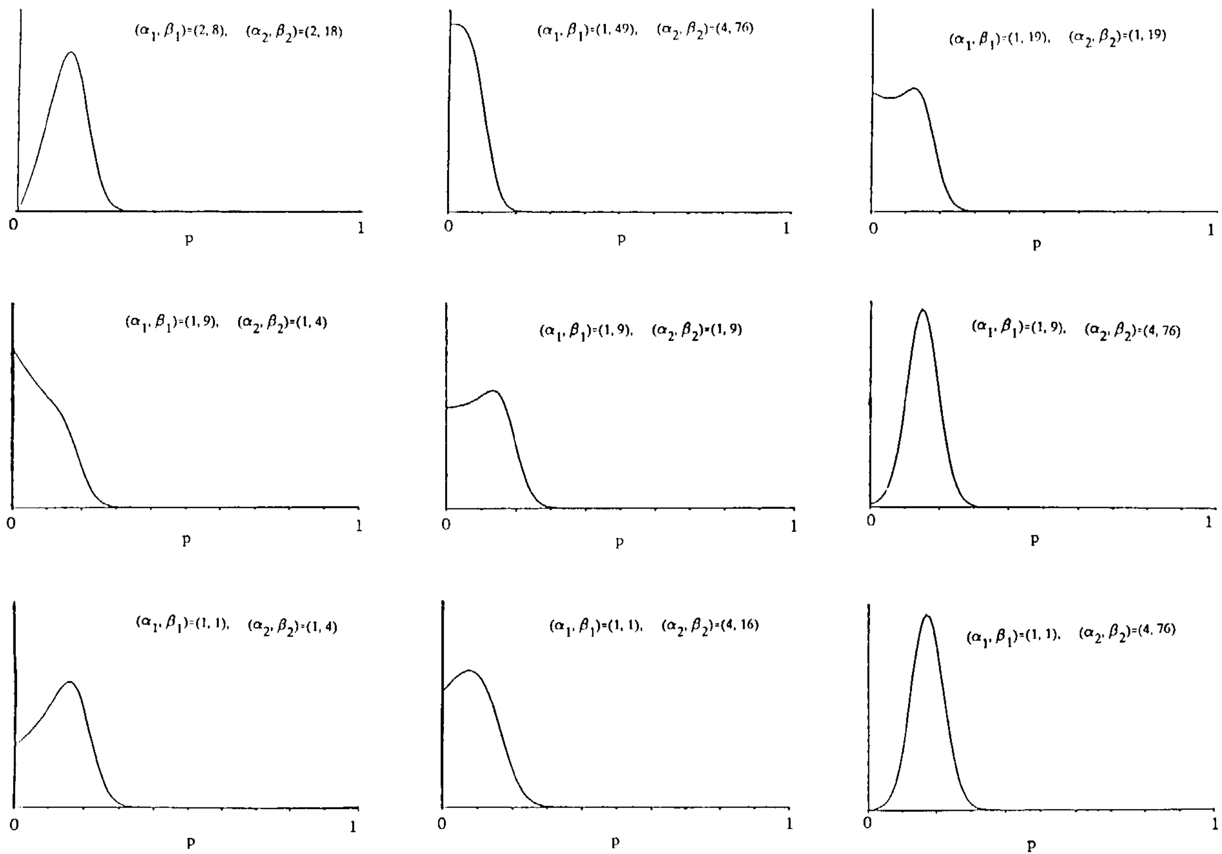
the prior distribution. The general nature of the basic results from the imperfect-sampling model (a reduction in the posterior mean of p and a loss of information as reflected by a larger standard deviation and an effective sample size considerably less than n) seems to hold consistently. The magnitude of the shifts and the balance between p and λ vary somewhat. For example, for a given prior distribution for p, note what happens as we move from $(\alpha_2, \beta_2) = (1, 19)$ through $(\alpha_2, \beta_2) = (2, 18)$ to $(\alpha_2, \beta_2) = (3, 17)$. The prior places more weight on higher values of λ, causing the posterior mean of λ to increase and the posterior mean of p to decrease because of the λ-p tradeoffs inherent in the likelihood function. At the same time, the effective sample size decreases for p and increases for λ.

Further insight into the imperfect-sampling model is provided by shapes of marginal posterior densities for p, which are given in Figure 3 for $r = 21$, $n = 104$, and a variety of prior distributions. Some of these densities are very similar to individual beta distributions, but others are very different. For example, when $\alpha_1 = \alpha_2 = 1$ and $\beta_1 = \beta_2 = 19$, there is a mode near 0.15 and the density flips up on the left to reach a local maximum at $p = 0$. Here the mode corresponds to the sample results being explained by a p near the sample proportion r/n and a small misclassification probability. The rise near $p = 0$ suggests an alternative explanation: a very small proportion combined with a relatively large λ. Explanations with intermediate values of p and λ apparently are less likely.

To demonstrate that the imperfect-sampling model can be extended to the case of unequal misclassification probabilities, we considered another question from Clark and Tifft (1966): "Have you skipped school without a legitimate excuse?" This question, which led to vastly different misclassification probabilities, was also asked by Gould (1969), who reports that $r = 12$ and $n = 104$. Once again considering the information from Clark and Tifft (1966) and attempting to judge differences in the populations used in the two studies, we assessed independent beta distributions for p, λ_1 (the probability that someone who skipped school answers no), and λ_2 (the probability that someone who never skipped school answers yes). The parameters for these distributions are, respectively, (3,7), (5,5), and (1,49). Note that λ_1 is judged to be considerably higher than λ_2, as might be anticipated. We expect half of those who have skipped school to lie, as opposed to only 2% of those who have not skipped school.

To conserve space, equations analogous to (6), (9), and (13) are not presented here, but the posterior distributions for p, λ_1, and λ_2 are mixtures of beta distributions. The posterior mean of p is 0.25, which is almost twice the posterior mean assuming perfect sampling, and the standard deviation of p is higher by a factor of 3.03 due to the noise. The effective sample size for inferences concerning p is 9.6, as

Figure 3: Marginal Posterior Distributions for p for $r = 21$, $n = 104$, and Selected Prior Parameters.



compared with $n = 104$. A uniform prior distribution for p, with the same distributions for λ_1 and λ_2, yields an even higher posterior mean for p, a smaller standard deviation, and an effective sample size of only 2.7.

The results in this section certainly do not cover all of the possible types of cases that might occur in the model with noise. Nonetheless, they illustrate the model and provide some flavor of the implications of noise for inferences about a proportion.

5. Summary and Discussion

We have considered inferences about a Bernoulli parameter p with imperfect, or noisy, sampling from the Bernoulli process with a single, unknown noise parameter λ. The likelihood function is Bernoulli not in p, but in $p(1-\lambda)+(1-p)\lambda$, and it can be represented as a mixture of products of Bernoulli likelihoods in p and λ. Because of the symmetry of the model, an identification problem arises in which the likelihood function is unable to distinguish among certain (p, λ) pairs. The incorporation of prior information via a Bayesian analysis avoids this identification problem. Results have been presented based on a joint prior density that is a product of beta densities for p and λ. The joint and marginal posterior densities can be expressed as mixtures of the possible posterior densities that could arise under perfect knowledge of the exact number of misclassifications of each type. Specific examples in Section 4 indicate some possible shapes for the joint posterior distribution and the marginal posterior distribution for p and provide an illustration involving a slightly more general model with two noise parameters. The examples reveal how inferences baaed on the imperfect-sampling model can differ (often substantially) from those based on an assumption of perfect sampling.

One important implication of the analysis is that ignoring the noise by treating the sampling process as Bernoulli in p can lead to misguided inferences. Point estimates of p will be shifted toward 0.5, and the degree of uncertainty about p often will be understated. Thus, unless λ is viewed as being very close to zero, it is important to include it in the model. Once it is included, however, the prior distribution of p and λ plays a very important role in the analysis, both in terms of identification and in terms of influencing the ultimate inferences. The assessment of $f(p)$ is identical in principle in the noisy and noise-free situations, although noise in past data may render that data less informative about p. As for the noise parameter, sometimes we may have considerable information about λ (e.g., errors in test results with well-known tests in quality control or medicine); other times our information about λ may be quite vague and imprecise (e.g., lying or misrepresentation in surveys, errors with new tests in quality control or medicine, errors in recording or handling data). In the latter case, the assessment of f(X) may be difficult, but this is exactly the situation in which care should be taken because inferences about p may not be highly robust with respect to variations in $f(\lambda)$.

The results presented here provide a model for dealing with a Bernoulli process with imperfect sampling and a flavor for the implications of the noise. Some generalizations being studied include a model with asymmetric noise involving two noise parameters, one for each type of misclassification (as illustrated briefly in Section 4), and a prior distribution allowing dependence between p and the noise parameters in the asymmetric model. We feel that models of this general type have great potential applicability and deserve more attention because of the number of real-world situations in which one or more sources of noise are present.

Acknowledgments

This research was supported in part by the National Science Foundation under Grant IST8600788. We are grateful for helpful comments from an editor and two referees.

References

Clark, J.P. and L.L. Tifft (1966). "Polygraph and Interview Validation of Self-Reported Deviant Behavior," *American Sociological Review,* Vol. 31, 516-523.

Fuller, W.A. (1987). Measurement *Error Models.* New York: Wiley.

Gould, L.C. (1969). "Who Defines Delinquency: A Comparison of Self-Reported and Officially Reported Indices of Delinquency for Three Racial Groups," Social *Problems,* Vol. 16, 325-336.

Press, S.J. (1968). "Estimating from Misclassified Data," *Journal of the American* Statistical *Association,* Vol. 63, 123-133.

Warner, S.L. (1965). "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical* Association, Vol. 60, 63-69.

Winkler, R.L. (1985). "Information Loss in Noisy and Dependent Processes," in *Bayesian Statistics* 2, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds. Amsterdam: North-Holland.