Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

Estimation of the Conformance Fraction in a Presence of Misclassification Errors: a Bayesian Analysis in an Absence of Expert's Knowledge

Roberto da Costa Quinino[a]

Magda Carvalho Pires[a]

Emilio Suyama[a]
[a]Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil

Linda Lee Ho[b]
[b]Universidade de São Paulo (USP), São Paulo, Brazil

Abstract

This paper discusses the problem of the estimation of the proportion p when the inspection system is imperfect (subject to diagnosis errors) and the sampled items are classified repeatedly m times. One assumes that no relevant information of the prior distributions of these errors is available and consequently a posterior distribution for the proportion p with high variability is generated due to non-informative prior distributions for those errors. In this paper, the authors suggest to split randomly the sample into two subsamples. Parameters of prior distributions are estimated by the first sample and a Bayesian inferential procedure is proceeded by a second sample. Numerical results indicate that such procedure yields better performance (lower variance for the posteriori distribution) rather than a single sample of size $n= n_1+n_2$ and non-informative prior distributions for the classification errors.

Key words: *Quality control, Proportion estimation, Bayesian analysis, Binomial distribution, Classification errors, Repeated classifications.*

Introduction

To implement quality control for attributes, one needs to take into account the efficiency of the system to classify the manufactured items as conforming or non-conforming. Two types of errors may occur during a diagnostic test: the first, known as type I, occurs when a conforming item is classified as non- conforming; the second, called type II, refers to the scenario when an actually non-conforming item is considered conforming.

In a pioneering paper, Bross (1954) has shown that, in the presence

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

of classification errors, the estimators obtained by means of a classical statistics approach are biased. Other authors, such as Johnson and Kotz (1988), Johnson *et al.* (1991), Evans *et al.* (1996), Viana (1994), Gustafson (2003), Quinino and Ho (2004), Stamey and Gerlach (2007) have emphasized that, if ignored, classification errors may jeopardize the entire process of inference and, consequently, the proportion estimative.

Let us suppose that in a random sample of n items, X items are declared as conforming for some determined characteristic. The random variable X has binomial distribution with parameters (n, p), that is, X~Bin (*n,p*) and p is the probability that an item be classified as conforming. However, the occurrence of classification errors in the system implies a modification in this probability function. Let e_1, $0 < e_1 < 1$, be the probability that a conforming item be wrongly classified as non-conforming, and e_2, $0 < e_2 < 1$, be the probability that a non-conforming item be classified as conforming (false-non-conforming and false-conforming error probabilities). So, the probability of an item be classified as conforming is $q = p(1 - e_1) + (1 - p)e_2$, which yields a random variable X whose binomial distribution has parameter q instead of p.

The difficulty in such analysis is better grasped as one tries to establish the maximum likelihood estimator. The likelihood function for the case that presents classification errors may be written as $L(x|n_1, q) = q^x (1-q)^{n-x}$. This is maximized to all points (p, e_1, e_2) so that $p(1-e_1)+(1-p)e_2 = x/n$ (Gaba and Winkler, 1992). Therefore, the maximum likelihood estimator is not unique.

To solve this problem, various classical methods have been suggested, a review of this subject can be found in Johnson *et al.* (1991). In general, the methods proposed rely on alternative sampling plans for a preliminary estimation of classification errors. From a Bayesian point of view, Gaba and Winkler (1992) considered an approach that requires the use of an informative prior distribution. This may pose a problem, since in many cases this information is not available. They have also verified that independent uniform prior distributions for parameters (p, e_1, e_2) yield a posterior mean for p equal to 1/2, regardless of the sample result; moreover, the likelihood attains its maximum value at all points (p, e_1, e_2) such that $p(1-e_1)+(1-p)e_2 = x/n$.

To minimize the problem discussed here, one may gather information by repeated binary responses. This approach is found in the literature under classical and Bayesian statistical view. Fujisawa and Izumi (2000) studied this problem under classical statistical point of view and they concluded that the likelihood parameters are identifiable when (i) the sum of $e_1 + e_2 < 1$ and (ii) the number of repeated classifications must be at most three times. Evans *et al*. (1996) analyzed this problem under a Bayesian perspective. They observed that the identifiability problem is reduced but not eliminated when a Beta prior for p, an independent Dirichlet for (e_1, e_2), $e_1 + e_2 < 1$,are used. This result signs the great important performed by the prior

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

distributions. Thus, to eradicate the non-identifiability problem, more information is necessary.

The question raised here is what may be proposed since it is reasonable the existence of the misclassification errors but no additional informal is available to elucidate its prior distributions. Inspired by Johnson and Kotz (1988) and Tenenbein (1970), one solution is split the sample in two subsamples; the first used to obtain the prior distributions and make inferential process with the second sample.

The present article proposes a model in which the process of Bayesian inference for the proportion in the presence of classification errors includes making the repeated classifications. The estimation process is split into two phases. In the first, a sample of size $n_2 < n_1$ is collected to estimate the prior distributions. In the second phase, a second sample of size n_1 is used to make the Bayesian inferential process. In practical terms, we believe that the methodology developed here will be useful insofar as it makes repeated classifications easier and more operational than informative prior distributions.

The paper is organized as follows: in Section 2, the original likelihood function of the proposed model is described. The Bayesian method and a numerical example are provided in Sections 3 and 4, respectively. The conclusions are outlined in Section 5.

Likelihood Function

Let each item in a random sample of size n_1 be independently classified m times as conforming or non-conforming, with m being an odd number; C_{ij} (i = 1, 2, … , n_1; j = 1, 2, . . . ,m) denotes a Bernoulli random variable corresponding to the j-th classification of the i-th item. So, $C_{2,3} = 1$ indicates that the second item was classified as conforming in the third classification. Let F_i be a Bernoulli random variable that denotes the final classification of the i-th item after m classifications. Consider that F_i = 1 if, and only if, $\sum_{j=1}^{m} C_{ij} > m/2$. The choice of an odd number for m avoids a tie and consequently avoids difficulty in reaching a final classification for an item. Table 1 illustrates this classification procedure.

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

Table 1 – Repeated classifications of n1 items (each one classified m times)

Item	Classifications (C_{ij})					Final Classification
	1	2	3	…	M	
1	C_{11}	C_{12}	C_{13}	$\cdots$	C_{1m}	F_1
2	C_{21}	C_{22}	C_{23}	$\cdots$	C_{2m}	F_2
3	C_{31}	C_{32}	C_{33}	$\cdots$	C_{3m}	F_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
n_1	$C_{n_1 1}$	$C_{n_1 2}$	$C_{n_1 3}$	$\cdots$	$C_{n_1 m}$	F_{n_1}

Let E_i be also another Bernoulli random variable that denotes the real state of the i-th item, so that the interest is to estimate $P(E_i = 1) = p$. Moreover, we have $e_1 = P(C_{ij} = 0 \mid E_i = 1)$ and $e_2 = P(C_{ij} = 1 \mid E_i = 0)$. So the probability of an item be classified as conforming is equal to

$$P(F_i = 1) = p\bar{B}i\left(m/2; m, 1-e_1\right) + (1-p)\bar{B}i\left(m/2; m, e_2\right) \tag{1}$$

where $Bi\left(m/2; m, 1-e_k\right)$ denotes the cumulative binomial distribution function defined at m/2 and $\bar{B}i\left(m/2; m, 1-e_k\right) = 1 - Bi\left(m/2; m, 1-e_k\right)$. Considering that m odd, equation (1) can be expressed as

$$P(F_i = 1) = pBi\left(m/2; m, e_1\right) + (1-p)\bar{B}i\left(m/2; m, e_2\right) \tag{2}$$

Now let us suppose a random sample of n_1 items, r items are considered conforming ($r = \sum_{i=1}^{n_1} F_i$), so the likelihood function can be written as

$$\begin{aligned} L(r \mid n_1, m, p, e_1, e_2) = & \left[pBi\left(m/2; m, e_1\right) + (1-p)\bar{B}i\left(m/2; m, e_2\right)\right]^r \times \\ & \left[1 - pBi\left(m/2; m, e_1\right) - (1-p)\bar{B}i\left(m/2; m, e_2\right)\right]^{n_1 - r} \end{aligned} \tag{3}$$

Note that if m=1, then (3) equals

$$L(r \mid n_1, m, p, e_1, e_2) = \left[p(1-e_1) + (1-p)e_2\right]^r \left[pe_1 - (1-p)(1-e_2)\right]^{n_1 - r} \tag{4}$$

Expression (4) is precisely the likelihood function used by Gaba and

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

Winkler (1992) and Viana *et al.* (1993), which indicates that expression (3) is a generalization of these models obtained through the use of repeated classifications.

Bayesian Analysis

Consider a joint prior distribution of (p, e_1, e_2) given by:

$$f(p,e_1,e_2)=f_\beta(p\,|\,\alpha,\beta)f_\beta(e_1\,|\,\alpha_1,\beta_1)f_\beta(e_2\,|\,\alpha_2,\beta_2) \tag{5}$$

where $f_\beta(a\,|\,b,c)$ denotes a Beta density function for the random variable a with parameters b and c. Beta distributions are widely used in Bayesian models to describe information concerning proportions (Gupta and Nadarajah, 2004). In this article, we consider the random variables (p, e1, e2) independent. As in Rahme *et al.* (2000) and Stamey *et al.* (2004) a natural way to obtain the posterior distribution for p could be the use of MCMC methods. Another possibility would be to use the Sampling/Importance Resampling (SIR) technique or Bayesian weighted bootstrap (Rubin, 1988). However, we have chosen an approach based on numerical integration because the posterior distribution of p can be made explicit despite the fact that it does not have a closed form (a program using the software Matlab is developed for this purpose and it is available upon request). For this, the equation (3) can be rewritten as

$$\begin{aligned}L(r\,|\,p,e_1,e_2n_1,m)=\sum_{j=0}^{r}\sum_{t=0}^{n-r}\binom{r}{j}\binom{n-r}{t}p^{n-j-t}(1-p)^{j+t}\times\\ \left[Bi(m/2;m;e_1)\right]^{r-j}\left[\bar{B}i(m/2;m;e_1)\right]^{n-r-t}\left[Bi(m/2;m;e_2)\right]^{t}\left[\bar{B}i(m/2;m;e_2)\right]^{j}\end{aligned} \tag{6}$$

The joint posterior density of (p, e_1, e_2) is obtained multiplying the joint prior distribution (5) by the likelihood function (6) and normalizing it as requires Bayes' theorem (Winkler, 2003 and Gelman *et al.,* 2004). Integrating it in respect to e_1 and e_2, one finds the marginal posterior density function for p, which can be written as:

$$f(p\,|\,r,n_1,m)=\sum_{j=0}^{r}\sum_{t=0}^{n_1-r}w_{jt}^{*}f_\beta\left(p\,|\,\alpha^{*},\beta^{*}\right) \tag{7}$$

where, $w_{jt}^{*}=\dfrac{a_{jt}^{*}}{\sum_{j=0}^{r}\sum_{t=0}^{n_1-r}a_{jt}^{*}}$, $a_{jt}^{*}=\binom{r}{j}\binom{n_1-r}{t}B\left(\alpha^{*},\beta^{*}\right)k_1(j,t)k_2(j,t)$;

$$k_1(j,t)=\int_0^1 e_1^{\alpha_1-1}(1-e_1)^{\beta_1-1}\left[Bi\left(m/2;m;e_1\right)\right]^{r-j}\left[\bar{B}i\left(m/2;m;e_1\right)\right]^{n_1-r-t}de_1\,;$$

$$k_2(j,t)=\int_0^1 e_2^{\alpha_2-1}(1-e_2)^{\beta_2-1}\left[Bi\left(m/2;m;e_2\right)\right]^{t}\left[\bar{B}i\left(m/2;m;e_2\right)\right]^{j}de_2\,;$$

and $B\left(\alpha^{*},\beta^{*}\right)$ denotes the value of the Beta function calculated at $\left(\alpha^{*},\beta^{*}\right)$ with $\alpha^{*}=\alpha+n-j-t$ and $\beta^{*}=\beta+j+t$.

The fact that the information needed to define informative prior distributions for misclassification errors and the interest proportion is insufficient

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

implies, for instance, no usage of U(0, 1) distributions, which is a particular case of the Beta distribution when both parameters are equal to one $[f_\beta(e_i \mid 1,1), i = 1,2]$. The employment of these distributions yields a multimodal posterior distribution for p with a high variability.

The use of repeated classifications extenuates the problem, but its results are still unsatisfactory. Thus, the posterior distribution may be useless not providing the necessary information about the proportion of interest. This evidences the need to obtain additional information from classification errors. Many authors such as Gaba and Winkler (1992), Gustafson (2003), Stamey and Gerlach (2007), Swartz *et al.* (2004), Paulino *et al.* (2005), Katsis (2005), Mwalili *et al.* (2007), have used very informative prior distributions to solve the problem. In practice, however, such information may not be available and, consequently, the problem to solve is: which prior distribution should be used. Efron *et al.* (2003) observed that the main problem of the Bayesian analysis is the choice of the prior distribution. An objective method to obtain additional information about the misclassification errors is the subject of the next subsection.

Elicitation of an Informative Prior Distribution

An alternative when we definitively do not have an informative prior distribution is to use the results of the m repeated classifications.

The sample is randomly split into two subsamples of sizes n_1 and n_2 ($n_2 < n_1$). The subsample of size n_2 is used to estimate the hyper-parameters (α_1, β_1) and (α_2, β_2) of the Beta prior distributions of misclassification errors e_1 and e_2 which is compounded by two sets: one made up of items which final classification is conforming ($F_i = 1$) and the other with items which final classification is non-conforming ($F_i = 0$).

For each item of the first set we calculate the proportion of non-conforming repeated classifications, bearing in mind that the mean and the variance of this proportion estimate, respectively, the mean and variance of the Beta prior distribution for e_1.

In the second subsample, we calculated the proportion of conforming classifications for each item. Similarly, the mean and the variance of that proportion estimate are, respectively, the mean and variance of the Beta prior distribution for e_2.

Finally, we are able to estimate (α_1, β_1) and (α_2, β_2), by solving the system of equations equaling the estimates of the mean and the variance and the closed expressions of the mean and variance of Beta distribution which are respectively:

$$\hat{\alpha}_1 = \frac{k_3}{k_4}(k_3 - k_3^2 - k_4) \tag{8}$$

$$\hat{\beta}_1 = \frac{1-k_3}{k_4}(k_3 - k_3^2 - k_4) \tag{9}$$

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

$$\hat{\alpha}_2 = \frac{k_5}{k_6}(k_5 - k_5^2 - k_6) \quad (10)$$

$$\hat{\beta}_2 = \frac{1-k_5}{k_6}(k_5 - k_5^2 - k_6) \quad (11)$$

where ;

and $k_3 = \frac{1}{\sum_{s=1}^{n_2} I_{\{F_s=1\}}} \sum_{i=1}^{n_2}\sum_{j=1}^{m} \frac{(1-C_{ij})I_{\{F_i=1\}}}{m}$. $k_4 = \frac{1}{\sum_{s=1}^{n_2} I_{\{F_s=1\}} - 1} \sum_{i=1}^{n_2}\left[\sum_{j=1}^{m} \frac{(1-C_{ij})I_{\{F_i=1\}}}{m} - k_3\right]^2$

$$k_5 = \frac{1}{\sum_{s=1}^{n_2} I_{\{F_s=0\}}} \sum_{i=1}^{n_2}\sum_{j=1}^{m} \frac{C_{ij}I_{\{F_i=0\}}}{m} \qquad k_6 = \frac{1}{\sum_{s=1}^{n_2} I_{\{F_s=0\}} - 1} \sum_{i=1}^{n_2}\left[\sum_{j=1}^{m} \frac{C_{ij}I_{\{F_i=0\}}}{m} - k_5\right]^2$$

The expressions k_3 and k_4 express respectively the estimates of the mean and the variance of the prior distribution of the misclassification error e_1; similarly k_5 and k_6, respectively estimates of the mean and the variance of the prior distribution of the misclassification error e_2. The prior distribution for m = 1 is not viable in the method proposed here, given the impossibility to estimate (α_1, β_1) and (α_2, β_2) from proportions of mistaken classifications.

To determine the sample size n_2 needed to apply the proposal presented here, the next algorithm may be adopted:

i) Split the sample n into two subsamples: one with size n_2 =2 and other with complementary size n_1 =n- n_2.

ii) Calculate k_3, k_5, $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_2$ and $\hat{\beta}_2$ employing the elements from the sample of size n2. If $0.5 < k_3$ and $k_5 < 1$ and $\hat{\alpha}_i > 0; \hat{\beta}_i > 0, i = 1,2$, then the sizes of the subsamples are determined.

iii) Otherwise, a new item is randomly chosen from the complementary subsample and jointed in the first subsample. Repeat the step ii employing the size $n_2 \leftarrow n_2 + 1$.The procedure continues until the stoppage criterion is reached or if $n_2 > 0.5n - 1$.

If no solution is found, one needs to increase n or m so as to assess the effects of classification errors allowing for the estimation of (α_1, β_1) and (α_2, β_2) according to the method of moments.

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

Numerical Examples and Discussions

The example described in this section is adapted from Ding *et al*. (1998) and Fujisawa and Izumi (2000). A company needs to estimate the number of conforming semiconductors bought from a new production line recently implanted in a supplier. This task is very important mainly if the statistical control process of the supplier is not yet certified and price of their product is lower when compared with the main competitors. To carry out the business and consequently to state the final price, one needs to know the conformance fraction of the semiconductors. For that, a random sample of 357 semiconductors is collected and sent by the supplier.

All sampled units are tested to get a more reliable result. Unfortunately, it is usually impossible to design tests without errors. Items that are classified as nonconforming may be conforming and those that are classified as conforming may be nonconforming. As classification system is new one, the quality engineers rather adopt the inexistence of relevant information about the level of the classification errors. The results of the three repeated classifications from 357 units are in Table 2: 117 semiconductors with zero conforming classification in three repeated classifications; 5 units with one conforming result (in three repeated classifications); 10 with two conforming results and 225 with three conforming classifications.

Table 2 - Frequency of observed conforming responses (k) among repeated diagnostic tests (m) for the semiconductors

k(m=3) and n=357			
0	1	2	3
117	5	10	225

To apply the proposed method, the sample of 357 units (n) described in Table 2 is randomly split into two subsamples according to the algorithm described in the end of the section 3. One compounded of 50 cases (n_2) is used to elicit the prior distributions and the remainder 307 cases (n_1) are employed for the inferential process. Table 3 describes the composition of two subsamples.

Table 3 - Frequency of observed conforming responses (k) among repeated diagnostic tests (m) for the semiconductors: split samples

k(m=3) and n_1=307				k(m=3) and n_2=50			
0	1	2	3	0	1	2	3
101	4	8	194	16	1	2	31

With the sample of 50 cases, the hyperparameters of the joint prior distribution expressed in (5) are estimated employing the expressions (8)–(11). With such, we get $\hat{\alpha}_1 = 0.041$, $\hat{\beta}_1 = 1.991$, $\hat{\alpha}_2 = 0.038$ and $\hat{\beta}_2 = 1.905$. Following Evans

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

et al. (1996), a non-informative prior distribution U (0, 1) for p is used, that is, so $\hat{\alpha}=1$ and $\hat{\beta}=1$ are adopted. With these values plus the data of 307 cases, a posterior distribution of p expressed in (7) is found numerically. Figure 1 presents the plot of the posterior distribution of p and some descriptive statistics (Mean, Mode, Median and Credibility Interval at 95%).

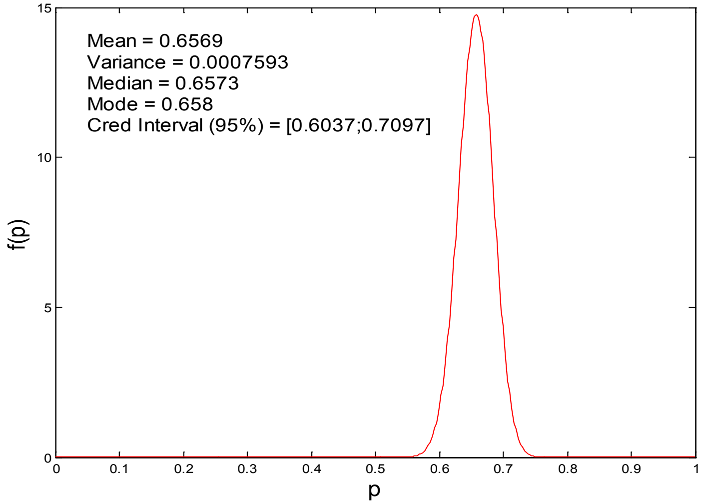

Figure 1 - Plot of the posterior distribution of p

To illustrate the importance of the estimation of the hyperparameters of prior distributions on the posterior distribution, a multimodal posterior distribution for p (and high variance also) as shown in Figure 2 is the result if non-informative prior distributions for the misclassification errors (that is, $\hat{\alpha}_1=1$, $\hat{\beta}_1=1$, $\hat{\alpha}_2=1$ and $\hat{\beta}_2=1$) and the full sample of 357 cases are used. Note that in this case the credibility interval is wide providing a poor and non-informative result.

Conclusions

This article presents a Bayesian methodology to estimate a proportion. The evaluations are subject to misclassification errors and repeated classifications are performed. As we have little or no information about the classification errors, we propose to split the sample into two subsamples. One is used to estimate prior distributions and the other to make the inferential procedures. A numerical study revealed that the methodology presents satisfactory performance.

Non-informative and useless results (in practical sense) are obtained if non-informative prior distributions are used. The posterior distribution is multimodal with high variability. This observation points out that prior distributions estimated

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

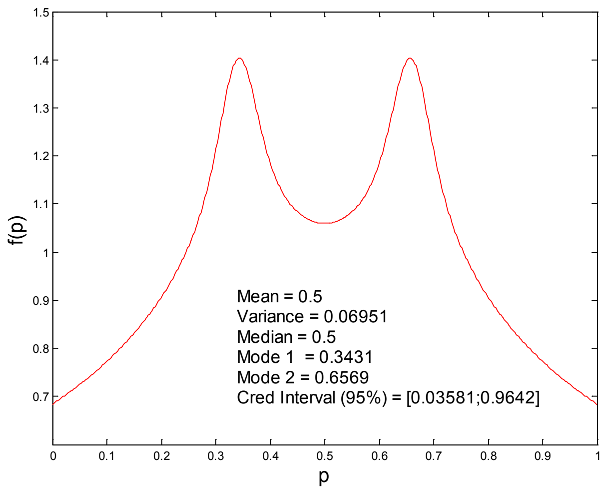

Figure 2 - Plot of the posterior distribution for p (non-information prior distributions)

by a subsample can improve sensibly the results when we are able to make repetitive classifications but they are subject to misclassification errors. The program (using software Matlab) used in the analysis of binary data subject to misclassification errors is available from the authors upon request.

For future studies, a natural extension of the present proposal is the determination of the sample sizes n and n_2 such that satisfies prior specifications of posterior variance or the range of the highest posterior density intervals for p.

Acknowledgments

The authors would like to acknowledge CNPq for the partial financial support.

References

Bross, I. (1954), "Misclassification in 2×2 tables", Biometrics 10, pp. 478–486.

Ding, J.; Greenberg, B. S. and Matsuo, H. (1998), "Repetitive testing strategies when the testing process is imperfect", Management Science, Vol. 44, No. 10, pp.1367-1378.

Efron, B.; Holmes, S.; Morris, C. and Tibshirani, R. (2003), "Bradley Efron: A conversation with good friends", Statistical Science, Vol. 18, No. 2, pp. 268–281.

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

Evans, M.; Guttman, I.; Haitovsky, Y. and Swartz, T. (1996), Bayesian analysis of binary data subject to misclassification. In: Berry, D.; Chaloner, K. and Geweke, J. (Eds.), Bayesian Analysis In Statistics and Econometrics: Essays In Honor of Arnold Zellner. New York: North Holland, pp. 66–77.

Fujisawa, H. and Izumi, I. (2000), "Inference about Misclassification Probabilities from Repeated Binary Responses", Biometrics, Vol. 56, No. 3, pp. 706–711.

Gaba, A. and Winkler, R. L. (1992), "Implications of errors in survey data: A bayesian model", Management Science, Vol. 38, No. 7, pp. 913–925.

Gelman, A.; Carlin, J. B.; Stern, H. S. and Rubin, D. B. (2004), Bayesian data analysis, 2nd Edition. London: Chapman & Hall.

Gupta, A. K. and Nadarajah, S. (2004), Handbook of Beta distribution and its applications, New York: Marcel Dekker.

Gustafson, P. (2003), Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments, New York: Chapman & Hall.

Johnson, N. L., Kotz, S. (1988), "Estimation from binomial data with classifiers of known and unknown imperfections", Naval Research Logistics, Vol. 35, pp. 147–156.

Johnson, N. L., Kotz, S., Wu, X. (1991), Inspection errors for attributes in quality control. London: Chapman & Hall.

Katsis, A. (2005), "Sample size determination of binomial data with the presence of misclassification", Metrika, Vol. 62, No. 2, pp. 323–329.

Mwalili, S. M.; Lesaffre, E. and Declerck, D. (2007), "The zero-inflated non-conforming binomial regression model with correction for misclassification: an example in caries research", Statistical Methods in Medical Research, Vol. 78, pp. 47–65.

Paulino, C. D.; Silva, G. L. and Achcar, J. A. (2005), "Bayesian analysis of correlated misclassified binary data", Computational Statistics and Data Analysis, Vol. 49, No. 4, pp. 1120–1131.

Quinino, R. C. and Ho, L. L. (2004), "Repetitive tests as an economic alternative procedure to control attributes with diagnosis errors", European Journal of Operational Research, Vol. 155, No. 1, pp. 209–225.

Rahme, E.; Joseph, L. and Gyorkos, T. W. (2000), "Bayesian sample size determination for estimating binomial parameters from data subject to misclassification", Applied Statistics, Vol. 49, No. 1, pp. 119–128.

Rubin, D. B. (1988), Using the SIR algorithm to simulate posterior distributions. In: Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M. (Eds.), Bayesian Statistics 3. Oxford University Press, Cambridge, pp. 395–402.

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

Stamey, J. D. and Gerlach, R. (2007), "Bayesian sample size determination for case-control studies with misclassification", Computational Statistics and Data Analysis, Vol. 51, No. 6, pp. 2982–2992.

Stamey, J. D.; Seaman, J. W. and Young, D. M. (2004), "Bayesian analysis of complementary Poisson rate parameters with data subject to misclassification", Journal of Statistical Planning and Inference, Vol. 134, No. 1, pp. 36–48.

Swartz, T.; Haitovsky, Y.; Vexler, A. and Yang, T. (2004), "Bayesian identifiability and misclassification in multinomial data", The Canadian Journal of Statistics, Vol. 32, No. 3, pp. 1–18.

Tenenbein, A. (1970), "A double sampling scheme for estimating from binomial data with misclassifications", Journal of the American Statistical Association, Vol. 65, No. 331, pp. 1350–1361.

Viana, M. A. G. (1994), "Bayesian small-sample estimation of misclassification multinomial data", Biometrics, Vol. 50, No. 1, pp. 237–243.

Viana, M. A. G., Ramakrishnan, V., Levy, P. S. (1993), "Bayesian analysis of proportion from the results of small screening samples", Communications in Statistics - Theory and methods, Vol. 22, No. 2, pp. 575–585.

Winkler, R. L. (2003), Bayesian inference and decisions, 2nd Edition. London: Probabilistic Publishing.

Biography

Roberto da Costa Quinino received his D.Sc. (1998) from the Department of Production Engineering at Universidade de São Paulo. Since then he has been serving on the Department of Statistics at Universidade Federal de Minas Gerais, Brazil. His major interests are classification errors and statistical quality control.
Contact: roberto@est.ufmg.br

Magda Carvalho Pires received her D.Sc. (2010) from the Department of Statistics at Universidade Federal de Minas Gerais. Since then she has been serving on the faculty of the same department. Her major interests are classification errors, reliability and statistical quality control.
Contact: magda@est.ufmg.br

Emílio Suyama received his D.Sc. (1995) from the Department of Statistics at Universidade de São Paulo. Since then he has been serving on the Department of Statistics at Universidade Federal de Minas Gerais, Brazil. His major research interests are sampling survey and statistical quality control.
Contact: suyama@est.ufmg.br

Linda Lee Ho received her D.Sc. (1995) from the Department of Production

Brazilian Journal of Operations & Production Management
Volume 7, Number 1, 2010, pp. 181-193

Engineering at Universidade de São Paulo. Since 1990 she has been serving on the faculty of the same department. Her major research interests are reliability, capability indices, and statistical quality control.
Contact: lindalee@usp.br

Article Info:

Received: February, 2010

Accepted: June, 2010