QÜESTIIÓ, vol. 22, 2, p. 231-242, 1998

ON MAXIMUM ENTROPY PRIORS AND A MOST LIKELY LIKELIHOOD IN AUDITING*

A. HERNÁNDEZ BASTIDA⋆
Universidad de Granada

M.C. MARTEL ESCOBAR⋆
F.J. VÁZQUEZ POLO⋆
Universidad de Las Palmas de G.C.

There are two basic questions auditors and accountants must consider when developing testing and estimation applications using Bayes' Theorem: What prior probability function should be used and what likelihood function should be used. In this paper we propose to use a maximum entropy prior probability function MEP with the most likely likelihood function MLL in the Quasi-Bayesian QB model introduced by McCray (1984). It is defined on a adequate parameter. Thus procedure only needs an expected value of θ_0 *known (in this paper the expected tainting) to obtain a MEP all an auditor or accountant need to suply are the range, as with any other prior, and the expected tainting,* θ_0. *We will see some practical applications of the methodology proposed about internal control evaluation in auditing.*

Keywords: Maximum Entropy Prior, Partial Prior Information, Auditing.

AMS Classification: 62A15, 62C10

---

* A draft version of this paper was presented by authors at 4th World Congress of the Bernoulli Society. Vienna, August, 1996. We would like to thank, without implicating, Nicanor Guerra, for computation assistance.

⋆ Authors in alphabetical order. All three authors have made major contributions to this work.
Research partially supported by «Dirección General de Universidades e Investigación del Gobierno Autónomo de Canarias», under grant 2705. *Address for correspondance*: Departamento de Métodos Cuantitativos en Economía y Gestión. Facultad de CC. Económicas. Módulo D. 35017 Las Palmas de G.C. *e-mail*: polo@empresariales.ulpgc.es

– Received September 1996.
– Accepted June de 1997.

1. INTRODUCTION

Bayes' Theorem has not been widely used in auditing and accounting because accountants find some dificulties in assessing a prior probability density function about the average tainting (or total error in an account or file). There are several reasons for these inconveniences. Firstly, it is difficult for accounting firms to define standard specific procedures to obtain the prior distribution using auditing evidence accumulated in the normal course of the evaluation of internal accounting control. Therefore it may be difficult for the auditor to justify the parameters in the prior probability density function to be used since they have not an intuitive meaning. Procedures involved may be also quite complicated and cumbersome to implement without significant staff training.

This paper suggests the usage of the class of prior probability density functions known as maximum entropy priors MEP's because they do not have any of the problems mentioned above such a class of distributions includes not only those aspects about the prior density which are unquestionable, but they are also the most noninformative priors available.

Moreover, this is a very simple procedure to use. For instance, if an auditing firm is about using MEP in internal accounting control evaluation and analytical reviews, two steps should be considered:

The former one includes the specification by the auditing firm of a matrix relating any of the possible evaluations to one or more of the statistical measures considered to be relevant in the prior distribution, i.e. the mean value, one or two quantile values. This matrix will be specific for this particular auditing firm and will be reflecting its own criteria.

In the second step, once that matrix has been established, the MEP distribution is completely defined and it can be used in bayesian analysis for DUS (Dollar Unit Sampling) or phisical unit sampling.

This paper introduces a new method to specify the posterior distribution that uses the MEP. This method is called MLPC (most likely posterior curves) and does not need to know the probability distribution for the tainting in the population.

The remainder of this article is organized as follows: Section 2 describes maximum entropy approach used later. Section 3 shows how maximum entropy priors and most likely likelihood have been used to obtain the posterior distribution of total amount of error in an accounting population. By illustrations, Section 4 describes how a firm could establish a relationship between the evaluation of internal accounting control and analytical review and the MEP characteristics. Finally, Section 5 contains a summary and some concluding remarks.

2. MAXIMUM ENTROPY PRIORS

Let the parameter space Θ be a continuous and bounded subset of the real line. In practice Θ will be a closed and bounded interval of the real line, $\Theta = [a,b]$, and represents the taint in an accounting population.

Even, if the parameter space is continuous and unbounded there is not a natural definition of entropy. We will use the definition proposed by *Jaynes (1968)*, for a probability distribution π as follows:

$$(1) \qquad Ent(\pi) = -E\left\{\log \frac{\pi(\theta)}{\pi_0(\theta)}\right\} = -\int_\Theta \pi(\theta)\log\left(\frac{\pi(\theta)}{\pi_0(\theta)}\right)\cdot d\theta$$

where π_0 is the natural «invariant» noninformative prior for the problem, usually we use the natural noninformative prior uniform.

It is well known (see *Berger (1985)*, pp. 92-93) that if partial prior information is given by:

$$(2) \qquad E^\pi[g_k(\theta)] = \int_\Theta g_k(\theta)\cdot\pi(\theta)\cdot d\theta = \mu_k, \;\; k = 1,\ldots,m,$$

and we try to solve:

$$\max \;\; Ent(\pi)$$

subject to:

$$E^\pi[g_k(\theta)] = \int_\Theta g_k(\theta)\cdot\pi(\theta)\cdot d\theta = \mu_k, \;\; k = 1,\ldots,m.$$

then, the solution is given by the expression:

$$(3) \qquad \pi(\theta) \propto \pi_0(\theta)\cdot exp\left\{\sum_k \lambda_k\cdot g_k(\theta)\right\}$$

where λ_k $(k = 1,2,\ldots,m)$ are constants to be determined from the constraints in (2).

Notice that:

- If $g_1(\theta) = \theta$ and, $g_k(\theta) = (\theta-\mu_1)^k$, $2 \le k \le m$, restrictions and hence partial information consists of specifying m central moments in the distribution,
- If $g_k(\theta) = I_{(-\infty,z_k]}(\theta)$, restrictions are now refered to the specification of m quantiles.

There are some interesting applications in statistical auditing involving the bayesian approach.

2.1. Partial Information given by the Mean

The maximum entropy prior for a location parameter specified by the mean θ_0 is given by

$$\pi(\theta) = \frac{\lambda e^{\lambda\theta}}{e^{\lambda b} - e^{\lambda a}} \tag{4}$$

where the unique restriction is $g_1(\theta) = \theta$, $\mu_1 = \theta_0$, and a and b are specified (in the domain of θ) and λ is obtained by solving the nonlinear equation

$$\frac{\lambda(ae^{-\lambda a} - be^{-\lambda b}) + e^{-\lambda a} - e^{-\lambda b}}{\lambda(e^{\lambda b} - e^{\lambda a})} = \theta_0 \tag{5}$$

if θ_0 is less than $(a+b)/2$, then λ is positive, λ is negative otherwise. If $\theta_0 = (a+b)/2$, equation (5) cannot be solved, but it can be shown that in the limiting shape of the MEP is that of the uniform prior between a and b. Equation (5) is not restricted to positive values for a and b as constraints above might suggest. If a is negative, it is a simple matter to perform a linear transformation to obtain a MEP satisfying the boundary constraints in equation (5).

2.2. Partial Information when one quantile is given

In this case the restriction is $g_1(\theta) = I_{(-\infty,z_1]}(\theta)$, where z_1 is the known α-quantile ($\alpha \in (0,1)$). We can obtain the maximum entropy prior through by easy algebraical manipulations:

$$\pi(\theta) = \begin{cases} \frac{e^{\kappa}}{e^{\kappa}(z_1-a)+(b-z_1)} & , \text{ if } \quad a \leq \theta \leq z_1 \\ \frac{1}{e^{\kappa}(z_1-a)+(b-z_1)} & , \text{ if } \quad z_1 < \theta \leq b \end{cases} \tag{6}$$

where κ is a constant to be determined from the constraint $\alpha = \int_a^{z_1} \pi(\theta)d\theta$. Easy computation gives us: $\kappa = \log\left((b-z_1)\alpha/(1-\alpha)(z_1-a)\right)$. Finally the MEP obtained from one quantile given is the two piece uniform distribution,

$$\pi(\theta) = \begin{cases} \frac{\alpha}{z_1-a} & , \text{ if } \quad a \leq \theta \leq z_1 \\ \frac{1-\alpha}{b-z_1} & , \text{ if } \quad z_1 < \theta \leq b \end{cases} \tag{7}$$

2.3. Partial Information when two quantiles are given

Now the restrictions are $g_1(\theta) = I_{(-\infty,z_1]}(\theta)$; $g_2(\theta) = I_{(-\infty,z_2]}(\theta)$, where z_1 is the known α_1-quantile and z_2 is the known α_2-quantile ($\alpha_1, \alpha_2 \in (0,1)$), $\alpha_1 < \alpha_2$, $z_1 < z_2$, then

$$\pi(\theta) \propto \exp\{\kappa_1 g_1(\theta) + \kappa_2 g_2(\theta)\}$$

where κ_1, κ_2 are constants to be determined from the constraints,

$$\alpha_1 = \int_a^{z_1} \pi(\theta) d\theta, \text{ and } \alpha_2 = \int_a^{z_2} \pi(\theta) d\theta.$$

Calculations involved can be easily performed obtaining the three piece uniform distribution MEP given by,

$$\pi(\theta) = \begin{cases} \frac{\alpha_1}{z_1 - a} & , \text{ if } \quad a \le \theta \le z_1 \\ \frac{\alpha_2 - \alpha_1}{z_2 - z_1} & , \text{ if } \quad z_1 < \theta \le z_2 \\ \frac{1 - \alpha_2}{b - z_2} & , \text{ if } \quad z_2 < \theta \le b \end{cases} \tag{8}$$

Obviously, if we consider three o more quantile the procedure yields four or more piece uniform distributions MEP. The three cases considered above are specially atractive to use in auditing and accounting settings as we will show in the next section. Furthermore, it should be used with the most likely likelihood in the Quasi-Bayesian Model (*McCray (1984,1986)*).

3. MEP AND MLPC IN AUDITING

A magnitude of prime interest in accounting auditing is the total amount of error since it has an intuitive meaning for auditors and it is also something those have prior relevant information about. Suppose a range of equally spaced possible total amount of error is defined (for example 500 or 1000).

There are various sampling techniques in auditing. Dollar unit sampling (DUS) maybe particularly appealing to auditors. Roughly speaking, this is a method to select sample items such that the probability of any given item being selected is directly proportional to its record value in the book. If the auditor is interested in including zero and/or small recorded value (wich have a small chance of being selected), then he/she could design specific audit test about them.

Suppose DUS sampling is used (see *Felix and Grimlund (1977)*, *Cox and Snell (1979)*, and *Godfrey and Neter (1982)*, among others) and we are interested in combining our sample observations to prior information to get a posterior probability distribution over all possible states of nature, the possible total amounts of error previously specified.

In DUS, the population size is the known Recorded Book Value (*RBV*) and the sample plan consists of selecting dollar units with equal chance of being selected. The amount of error for each dollar selected is the difference between its two associated values: its book value and its audited value (presumed to be correct). The fraction (error / book value) is called **taint** of the dollar-unit randomly selected. Taintings in a dollar unit sample are recorded and used to make inference about the total error amount in the population. In a empirical situation, most of these tainting values are zero. We assume that no amount can be overestimated or underestimated by a quantity bigger than its book value, therefore the variation range of taintings goes from -100 to +100 per cent.

We have then 201 categories of taintings: $T_{-100},\dots,T_{-1},T_0,T_{+1},\dots,T_{+100}$, associated to different taintings: $-100\%,\dots,+100\%$. When the error tainting in the sample is zero, the sample dollar is counted in category T_0, and when it is in between the category T_{i-1} and T_i it is counted in category T_i. Let θ_i $(i=-100,\dots,+100)$ denote the population proportion of dollar-units with i percent error. For a random sample of dollar units of size n, let n_i $(i=-100,\dots,+100)$ be the observed frequencies in category T_i $(i=-100,\dots,+100)$. The counts in categories follow a multinomial model[1]

$$M(\theta_{-100},\dots,\theta_{100}) = \frac{n!}{n_{-100}!\cdot\dots\cdot n_{100}!}\Pi_{i=-100}^{100}\theta_i^{n_i} \tag{9}$$

The relevant magnitude for auditors is the total amount of error, λ, and the prior knowledge is assessed on λ, say $\xi(\lambda)$. This parameter is given by:

$$\lambda = \frac{\text{RBV}}{100}\cdot\sum_{i=-100}^{100} i\cdot\theta_i$$

Bayes' Theorem asserts that the logical way to modify prior beliefs about a unknown parameter is to combine prior and likelihood distributions resulting in a posterior one. Prior and likelihood function must be referred to same unknown parameter. However in our model, likelihood function refers to proportions of errors, θ_i's, and prior refers to total amount of error (a linear combination of the proportion of error). Thus, a traditional Bayesian would require that auditors supply a prior probability mass function for each possible proportion of errors θ_i's. However, there has been a

[1]Exactly, the multinomial model appears when the random sample is chosen with replacement. In other case, because in practice the total book value is very large in relation to sample size, the multinomial model is a good approximation of the likelihood function.

development which can be considered to be a modification of the usual likelihood for handling data from a QB scheme. Briefly, the approach defines the likelihood function for the unknown λ as the likelihood induced by $M(\theta_{-100},\ldots,\theta_{100})$ in *Zehna (1966)* notation. An intuitive approach of that function is in *McCray (1984)*. For a complete development of that function see *Hernández et al. (1996)* and other references therein.

The posterior mass function, called Most Likely Posterior Curve (MLPC)[2], could be calculate by:

$$\xi(\lambda \mid \text{data}) = \frac{M^*(\lambda)\cdot\xi(\lambda)}{\sum_\lambda M^*(\lambda)\cdot\xi(\lambda)}$$

where:

$$M^*(\lambda) = \sum_\lambda \left(\sup_{\phi^{-1}(\{\lambda\})} M(\theta_{-100},\ldots,\theta_{+100}) \right) \cdot I_{\phi^{-1}(\{\lambda\})}(\theta_{-100},\ldots,\theta_{+100})$$

$$\phi^{-1}(\{\lambda\}) = \left\{ (\theta_{-100},\ldots,\theta_{+100}) \;\middle|\; \sum_{i=-100}^{+100} \theta_i = 1 \text{ and } \frac{RBV}{1000} \cdot \sum_{i=-100}^{+100} i\cdot\theta_i = \lambda \right\}$$

Therefore the QB formulation can be summarized via Bayes'Theorem with a maximized likelihood function. Any prior can be used.

This posterior distribution has two advantages,

1. it does no require modelling the accounting population taintig to obtain the likelihood (as in the most of the audit models, *Cox and Snell (1979)*, *Godfrey and Neter (1984)*, and *Felix and Grimlund (1977)*),

2. the user doesn't have to asses neither prior distribution on few intuitive magnitudes nor a prior distribution on a high dimensional parameter space (*Tsui et al. (1985)*).

Despite this two advantages, this model needs to elicit a complet prior distribution and it can be a very difficul task for auditors, though this is a prior distribution on one parameter, with a strongly intuitive meaning, the total amount of error. But this assignment is allways a subjective matter, and it could cause problems to justify its functional form.

It should be very useful a more objective procedure such that the prior distribution could include those issues with the most certainty, and outside of those, it could be the least informative prior distribution.

[2]The Quasi-Bayesian educational PC share ware **MLPC v.3.02** and manual are available through the Internet at the anonymous ftp site: **csg.uwaterloo.ca** in the directory: **pub/dmg/mlpc**.

This procedure even could include qualitative judgements like «excellent», «very good», ... about the evaluation of internal accounting control of the firm. It could be possible to identify each of the previous judgements with some aspects of the prior distribution, like the mean, one or more quantiles,... (of course, depending on the policy of the firm). This method can be possible using the Maximum Entropy Prior distribution, and this is the method we propose in this work, clearly more advantageous than the other usual estimation procedures of the total amount of error in auditing.

4. ILLUSTRATIONS

The following examples illustrate how above MEP's might be used in audit situations and the effect of different mean and/or quantiles on a few descriptives posterior upper bounds: $qb_{50}, qb_{80}, qb_{90}, qb_{95}, qb_{99}$. Consider the following DUS data from an inventory with a reported book value of $\$1,000,000$, a sample size of 100 items, and taints observed of 0, 10, 90 and -25 with number of cases of 94, 1, 1, 4, respectively. As an example, we give in Table 1 below a matrix of possible relationship between Internal Control and/or Analitical Review Evaluation (IC/AR) and prior information about the total amount of error in the inventory (usually, overstatement error).

Table 1. *Matrix of Possible Relationships between IC/AR Evaluation and Prior Information*

IC/AR Evaluation	Prior Information			
	Expected average tainting θ_0 (percent)			
Excellent	3			
Very Good	5			
Good	10			
Poor	15			
Very Poor	30			
	One Quantile			
	Maximum Tainting Percent (MTP)		Credibility (CR)	
Excellent	3		0.95	
Very Good	5		0.95	
Good	10		0.95	
Poor	15		0.80	
Very Poor	30		0.70	
	Two Quantiles			
	MTP	CR	MTP	CR
Excellent	3	0.95	5	0.99
Very Good	5	0.95	10	0.99
Good	10	0.95	15	0.99
Poor	15	0.80	30	0.95
Very Poor	30	0.70	50	0.80

For instance, in the one quantile case if IC/AR evaluation yields an output of Very Good, then auditors can feel comfortable accepting that the probability that total amount of error be minor than 5% is 0.95. The following table shows the behaviour of posterior probability in these situations. MEP refers to MEP with mean known, and MEP1 and MEP2 refer to MEP with one and two quantiles given, respectively.

Table 2. *Posterior Descriptive Quantities of Total Amount of Error*[3]

Distribution		IC/AR Evaluation				
		Excellent	V. Good	Good	Poor	V. Poor
Post. Mean	MEP	27872.07	26690.69	31237.37	31802.57	32488.20
	MEP1	32863.98	32972.65	32972.65	32972.65	32972.65
	MEP2	32873.64	32972.65	32972.65	32972.65	32972.65
Post. Mode	MEP	21050.04	28476.98	23682.89	24121.46	24651.69
	MEP1	25177.08	25040.71	25040.71	25040.71	25040.71
	MEP2	25164.96	25040.71	25040.71	25040.71	25040.71
Post. Median	MEP	25598.06	27286.12	28719.21	29242.20	29876.03
	MEP1	30301.68	30328.67	30328.67	30328.67	30328.67
	MEP2	30304.08	30328.67	30328.67	30328.67	30328.67
qb_{80}	MEP	35876.41	38302.43	40367.56	41122.54	42038.42
	MEP1	42618.63	42694.82	42694.82	42694.82	42694.82
	MEP2	42625.43	42694.82	42694.82	42694.82	42694.82
qb_{90}	MEP	42263.74	45155.88	47620.98	48522.93	49617.84
	MEP1	50253.08	50407.74	50407.74	50407.94	50407.74
	MEP2	50266.88	50407.74	50407.74	50407.94	50407.74
qb_{95}	MEP	48095.48	51417.08	54252.19	55290.94	56553.19
	MEP1	57152.31	57450.34	57450.34	57450.34	57450.34
	MEP2	57178.19	57450.34	57450.34	57450.34	57450.34
qb_{99}	MEP	60458.79	64718.94	68390.80	69748.42	71409.98
	MEP1	71221.28	72615.84	72615.84	72615.84	72615.84
	MEP2	71336.30	72615.84	72615.84	72615.84	72615.84

[3]These calculations were made using the educational freeware program MLPC, a Quasi-Bayesian software package. Number of data points was set at 500.

These results are shown in Table 2. For example, if there is a good evaluation of IC/AR, the qb_{80} is $40367. This means the most likely probability the actual overstatement in the inventory balance is less than $40367 is 0.80. Also for qb_{95}, an excellent evaluation of IC/AR results in almost a 15% reduction in the upper bound compare with a very poor evaluation of IC/AR.

5. COMMENTS AND CONCLUSIONS

In the proposed model in this paper, we must bear in mind that the magnitude λ is commonly more familiar to the auditor, the simplification is clear. Moreover, if we compare this model with other models based in the multinomial likelihood (see *Tsui et al. (1985)* or *Byekwaso (1994)*) we may take out some conclusions. All this models have the inconvenience of a high dimensional parametric space. Therefore, the choice of prior distribution, in practice, may only be carried out, if Dirichlet distribution is selected, requiring also a big effort to get the posterior.

The proposed model in this paper can be seen as a more simplified methodology compared to most of proposed models in the literature. This simplification is clear both is the conceptual level and for practical applications. Finally, since we have insisted on the advantage of this model for practical purposes, it is essential to optimize some nonlinear restricted mathematical programs. This equations can be solved by finding the minimum of an unconstrained function. These calculations are included in the MLPC software.

The combination of a maximum entropy prior and the most likely likelihood function appears to be well suited for audit and accounting applications of Bayesian analysis because it is easy to defend and support. The above example suggests that the resulting upper bounds are consistent with the prior and likelihood used.

In all situations, the fact that the mean is greater than the mode reflects that the posterior distribution has a moderate skew towards higher amount of error. Perhaps because one taint of 90% has been observed and the model is very sensible to it.

Respect to IC/AR initial classification, more conservatives upper bounds are obtained when quantiles are given than mean is given.

The use of the mean as prior information yields differents posterior distributions according to different IC/AR evaluation, note a reduction in the upper bounds from Very Poor to Excellent classification.

Using quantiles no differences are appreciated in the posterior distribution respect to IC/AR evaluation, only between Excellent and the remainder situations. There isn't difference between this situations because all priors are identical in the support of the most likely likelihood and differ where the likelihood is almost zero. Also, this fact occurs between MEP1 and MEP2.

Thus, when the minimum amount of information is required prior mean is a good election. It maybe better than quantiles, because in auditing context usually 95 and 99 quantiles are very close. Possibly, 50 and 95 quantiles yields better results.

This work could be extended incorporating another intuitive magnitude like the mode (see *Brockett et al. (1984)*).

REFERENCES

[1] **Berger, J.O.** (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd edition). Springer - Verlag, New York.

[2] **Brockett, P., Charnes, A.** and **Paick, K.** (1984). «A Method for Constructing a Unimodal Inferential or Prior Distribution». *Research Report* #473. Center for Cybernetic Studies, The University of Texas at Austin.

[3] **Byekwaso, S.** (1994). «Bayesian Sequential Inference for Error Rate and Error Amounts in Accounting Data». *Ph.D. Thesis*, The American University, Dpt. of Mathematics and Statistics. Washington.

[4] **Cox, D.R.** and **Snell, E.J.** (1979). «On Sampling and the Estimation for Rare Errors». *Biometrika*, **66, 1**, 125–132. Amendments and Corrections in vol. 69, page 491, 1982.

[5] **Felix, W.L. (Jr.)** and **Grimlund, R.A.** (1977). «A Sampling Model for Audit Tests of Composite Accounts». *Journal of Accounting Research*, Spring, 23–41.

[6] **Godfrey, J.T.** and **Neter, J.** (1984). «Bayesian Bounds for Monetary — Unit — Sampling in Accounting and Auditing». *Journal of Accounting Research*, **XX, 2**, Part 1, 304–315.

[7] **Hernández, A., McCray, J.H.** and **Vázquez-Polo, F.J.** (1996). «Bayesian Inference on a Subparameter. An Application in Auditing». Submitted to *Journal of Applied Statistics.*

[8] **Jaynes, E.T.** (1968). «Prior Probabilities». *IEEE Trans. Systems, Science and Cybernetics*, **4**, 227–291.

[9] **McCray, J.H.** (1984). «A Quasi-Bayesian Audit Risk Model for Dollar Unit Sampling». *The Accounting Review*, **LIX, 1**, 35–51.

[10] **McCray, J.H.** (1986). «A General Bayesian Risk Model for Dollar Unit Sampling and Multiple Populations».*Proceedings of VII Biennial Audit Research Symposium* University of Illinois, 181–252.

[11] **Tsui, K., Matsumura, E.M.** and **Tsui, K.L.** (1985). «Multinomial — Dirichlet Bounds for Dollar — Unit Sampling in Auditing». *The Accounting Review*, **LX, 1**, 76–96.

[12] **Zehna, P.W.** (1966) «Invariance of Maximum Likelihood Estimation». *Annals of Mathematical Statistics*, **37**, 755.