

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

TITLE: BAYESIAN METHODS FOR INTERPRETING PLUTONIUM URINALYSIS DATA

AUTHOR(S): Guthrie Miller and William C. Inkret

SUBMITTED TO: MAXIMUM ENTROPY & BAYESIAN METHODS — PROCEEDINGS OF THE FIFTEENTH INTERNATIONAL WORKSHOP

RECEIVED

AUG 29 1995

OSTI

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED ✓

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Los Alamos Los Alamos National Laboratory Los Alamos, New Mexico 87545

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

BAYESIAN METHODS FOR INTERPRETING PLUTONIUM URINALYSIS DATA

G. Miller and W. C. Inkret
Los Alamos National Laboratory

June 26, 1995

Abstract

We discuss an internal dosimetry problem, where measurements of plutonium in urine are used to calculate radiation doses. We have developed an algorithm using the MAXENT method. The method gives reasonable results, however the role of the entropy prior distribution is to effectively fit the urine data using intakes occurring close in time to each measured urine result, which is unrealistic. A better approximation for the actual prior is the log-normal distribution; however, with the log-normal distribution another calculational approach must be used. Instead of calculating the most probable values, we turn to calculating expectation values directly from the posterior probability, which is feasible for a small number of intakes.

1 Introduction

In the field of health physics, exposure to alpha-emitting radionuclides like plutonium is usually monitored by periodic urinalysis. We have developed a new algorithm to calculate the radiation dose to an individual from internal depositions of plutonium (by inhalation or via contaminated wounds) based on their urinalysis data. The code is described in more

detail in Ref. [1]. The mathematical method is to maximize the Bayesian posterior probability using an entropy function as the prior probability distribution. The MAXENT method as implemented in the MEMSYS software package is used. Some advantages of the new code are that it ensures positivity of the calculated doses, it smooths out fluctuating data, and it provides an estimate of the propagated error in the calculated doses. This method is generally applicable to the internal dosimetry problem, and we plan to implement it also for tritium and uranium, which are also monitored by urinalysis.

The MAXENT method, although an advance over previously used data unfolding methods, is still not definitive, since the entropy form of the prior probability distribution is not realistic. The entropy function has the property that its standard deviation divided by its mean is always less than $\sqrt{2}$, which is the limiting value for small α , as will be discussed. A more realistic prior would have a larger value of this quantity, as is the case, for example, with the log-normal distribution. We believe the log-normal distribution to be a better approximation to the actual prior distribution as determined by statistical studies of historical data. The relative narrowness of the entropy prior leads to an underestimate of

radiation doses in the internal dosimetry problem. We will discuss a Bayesian calculation for a log-normal prior distribution.

The current approach to internal dosimetry is to interpret bioassay measurements in terms of radionuclide intake quantities. Let x_i for $i = 1, M$ denote the intake that occurred during time interval i , where M is the total number of time intervals. For example, in plutonium dosimetry x_i is the activity of plutonium taken into the body on the i^{th} day by inhalation or via a contaminated wound.

The bioassay data are denoted by y_j for $j = 1, N$, with uncertainty estimates (standard deviations) σ_j . For example, y_j is the j^{th} measurement of plutonium activity excreted per day. The biokinetic response is assumed linear and known, so that

$$f_j = \sum_{i=1}^M x_i u_{ij}$$

is the predicted bioassay result at time j given intakes x_i , where u_{ij} is the biokinetic response at time j for unit intake at time i .

The problem is to determine the "best fit" values of $\{x_i\}$ given $\{y_j\}$ (read $\{x_i\}$ as "the set of x_i for all i "). The MEMSYS method is to define "best fit" as the x_i values that give a maximum of the Bayesian posterior probability of $\{x_i\}$, given the data and an entropy prior probability distribution.

2 MAXENT

The entropy form of the prior distribution is given by

$$P(x_i) = C \exp[\alpha S(x_i)], \quad (1)$$

where

$$S(x_i) = x_i - m_i - x_i \log \frac{x_i}{m_i},$$

α and m_i are parameters, and C is a normalization constant such that

$$\int P(x_i) \frac{dx_i}{x_i^{1/2}} = 1$$

(note the metric factor $1/x_i^{1/2}$).

In Ref. [1] a simplified approximation for the entropy function is obtained for the case of small α . For small α ,

$$P(x_i) = \sqrt{\frac{\lambda}{\pi}} \exp(-\lambda x_i), \quad (2)$$

where $\lambda = 2\alpha \log(\zeta)$. The quantity ζ is determined by solving the equation

$$\zeta = \sqrt{\frac{1}{2\alpha m_i \log \zeta}},$$

iteratively for ζ starting with $\zeta = e$ as the initial guess. This requires that α not be too large. Using this approximation, the mean and standard deviation about the mean of x_i are given by

$$\begin{aligned} \langle x_i \rangle &= \frac{1}{2\lambda} \\ \sqrt{\langle (\Delta x_i)^2 \rangle} &= \sqrt{2} \langle x_i \rangle \end{aligned} \quad (3)$$

The entropy distribution has the property that its standard deviation divided by mean is always less than $\sqrt{2}$, which is the limiting value for small α . A more realistic prior probability distribution would have a larger value of standard deviation divided by mean, as will be discussed.

A test of the MAXENT code using real ^{239}Pu urine excretion data is shown in Fig. 1. The MAXENT code reconstructed the intake scenario using a noninformative prior probability distribution (uniform small values of m_i).

Basically, the intake scenario calculated in Fig. 1 is reasonable. However, on closer inspection, one sees that an upward excursion

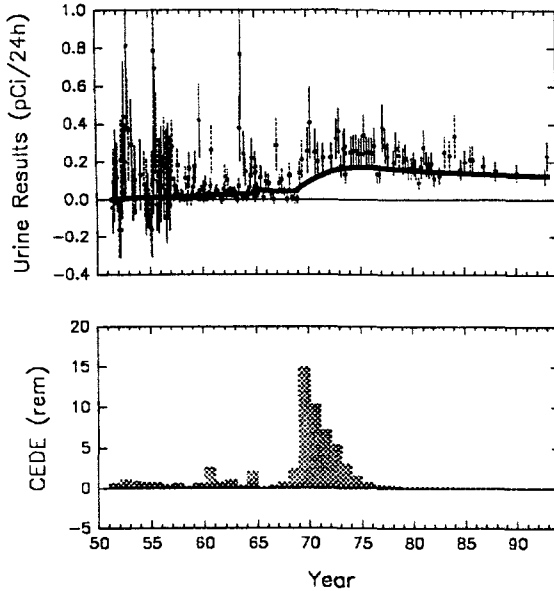


Figure 1: Example of a MAXENT code dose calculation.

of a urine excretion data point is fit by assuming intakes occurring very close in time to the time of the urine data. A “hand fit” for a simple case by a professional health physicist would usually assume an intake occurring further away in time, and giving a higher dose estimate.

3 Log-Normal Prior Distribution

The log-normal distribution has the form

$$P(x)dx = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(\log \frac{x}{m})^2}{\sigma_0^2}\right] \frac{dx}{x}. \quad (4)$$

The log-normal distribution is known empirically to describe many cases of interest[3]. We suspect, and have some empirical evidence, that it gives a reasonable description of the

distribution of radionuclide intakes for a population of exposed workers. The log-normal distribution has mean and standard deviation given by [2]

$$\begin{aligned} \langle x \rangle &= me^{\sigma_0^2/2} \\ \sqrt{\langle(\Delta x)^2\rangle} &= \langle x \rangle \sqrt{e^{\sigma_0^2} - 1} \end{aligned} \quad (5)$$

Thus, in contrast to the entropy distribution, the log-normal distribution can be arbitrarily broad for large σ_0 .

The log-normal distribution does not have the inverse property of the entropy distribution that allows solution of the maximum probability equations by data space mapping. That is, if

$$f(x) = \frac{\partial}{\partial x} [\log P(x)],$$

not all values of y allow solutions of $y = f(x)$. This means that the maximum posterior probability problem with M variables (M might be 10000) and N data points (N might be 100) will involve M equations in M unknowns rather than N equations in N unknowns—a very important practical difference.

Perhaps using the log-normal prior distribution fundamentally changes the mathematical character of the problem. To begin to understand whether this might be true, we consider a very simple case, with one intake x and one measurement y . It turns out that the problem easily generalizes to multiple measurements with one intake.

The posterior probability is then given by

$$P(x|y)dx \sim \exp\left[-\frac{(xu - y)^2}{2\sigma^2} - \frac{1}{2\sigma_0^2} \left(\log \frac{x}{m}\right)^2\right] \frac{dx}{x},$$

where u is the fraction of intake excreted at the sampling time after an intake, x is the intake amount, y is the measured activity of plutonium excreted per day, σ is the measurement error standard deviation, and σ_0 and m

are log-normal parameters that are considered known. Simplifying the notation,

$$P(x|y)dx \sim \exp [-\alpha(z - z_0)^2 - \beta(\log z)^2] \frac{dz}{z}, \quad (6)$$

where

$$\begin{aligned} z &= \frac{x}{m}, \\ z_0 &= \frac{y}{um}, \\ \alpha &= \frac{(um)^2}{2\sigma^2}, \\ \beta &= \frac{1}{2\sigma_0^2}. \end{aligned} \quad (7)$$

If we ask the question "What is the most probable value of x ?", the first problem is to decide "With respect to what variable?". If the probability density is plotted versus x , the maximum will occur at a different x than if plotted versus $\log x$. The second problem for Eq. 6 is that there are sometimes multiple local maxima.

The maxima of the exponential in Eq. 6 occur when

$$\frac{\partial}{\partial z} [\alpha(z - z_0)^2 + \beta(\log z)^2] = 0,$$

or, when

$$z - z_0 = -\frac{\beta \log z}{\alpha z} = -\gamma \frac{\log z}{z}, \quad (8)$$

where we have defined γ as

$$\gamma \equiv \frac{\beta}{\alpha}. \quad (9)$$

Figure 2 shows that Eq. 8 always has solutions. However, for large γ there are multiple solutions as shown in Fig. 3. The critical γ turns out to be

$$\gamma_{crit} = 2e^3 = 40.2, \quad (10)$$

and, for $\gamma > \gamma_{crit}$, there are multiple solutions.

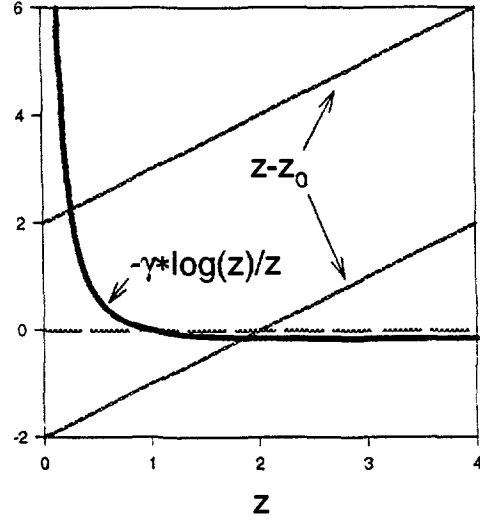


Figure 2: Graphical solution of the equation determining the most probable z values.

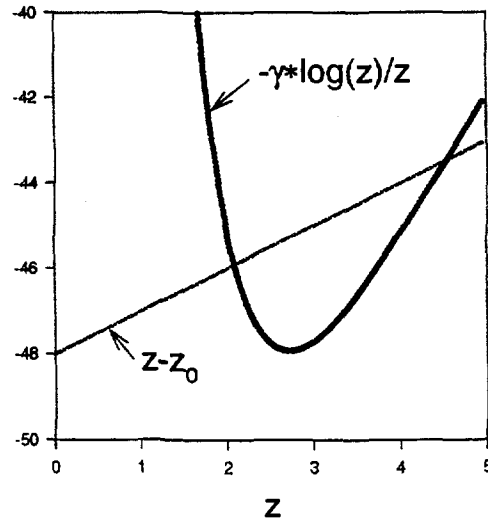


Figure 3: A case where multiple local maxima occur.

For these two reasons, 1) ambiguity about metric factors, and 2) possible multiple local maxima, we propose that finding the maximum of the posterior probability may not be the optimum approach.

Another approach is to directly use the posterior probability distribution to evaluate expectation values. This is conceptually simpler.

For the problem at hand,

$$\langle x^n \rangle = \frac{\int_0^\infty x^n P(x|y) dx}{\int_0^\infty P(x|y) dx}, \quad (11)$$

with $P(x|y)dx$ given by Eq. 6. We therefore can easily calculate $\langle x \rangle$, $\langle (\Delta x)^2 \rangle$, $\langle \chi^2 \rangle$ and other quantities of interest.

If there are multiple measurements, the generalizations of Eqs. 7 are

$$z_0 = \frac{1}{m} \frac{\sum_{j=1}^N w_j \frac{y_i}{u_j}}{\sum_{j=1}^N w_j}, \quad (12)$$

and

$$\alpha = m^2 \sum_{j=1}^N w_j, \quad (13)$$

where the weighting factor w_j is given by

$$w_j = \frac{u_j^2}{\sigma_j^2}.$$

4 Conclusions

The form of the prior distribution is important. We see this clearly in the internal dosimetry calculation where the MAXENT doses are smaller than those calculated "by hand" for simple cases. Unfortunately, the entropy form of the prior distribution is not able to approximate what we believe is the actual broad form of the prior for cases of interest. Thus, we are led to investigate application of the log-normal prior distribution.

With the log-normal prior, 1) data space mapping fails, and 2) there may be multiple local maxima of the posterior probability. Thus, we cannot simply change MAXENT to use a log-normal prior, and are led to consider alternate approaches.

An elegant approach is to directly evaluate average values using the posterior probability distribution. Straightforward application of this method is practical for determining small numbers of intakes and quite trivial for single intakes. The single-intake method can be used in a data unfolding procedure where multiple intakes are successively determined by one or more subsequent data points. The challenge for the future is to find methods to evaluate very large dimension integrals in order to be able to simultaneously determine large numbers of multiple intakes.

References

- [1] G. Miller and W. C. Inkret, "Bayesian Maximum Posterior Probability Method for Interpreting Urinalysis Data", Rad. Prot. Dosimetry, submitted, (1995).
- [2] V. Rothschild and N. Logothetis, *Probability Distributions*, (John Wiley, New York, 1986).
- [3] H. J. Gale, "The Lognormal Distribution and Some of Its Applications in the Field of Radiation Protection," AERE-R4736 (HMSO, London, 1965).