LA-UR- 98-1927

Approved for public release; distribution is unlimited.

Title: ANALYZING BIOASSAY DATA USING BAYESIAN METHODS-A PRIMER

CONF-971207--

RECEIVED
OCT 05 1998
OSTI

Author(s): GUTHRIE MILLER, ESH-12, LANL
WILLIAM C. INKRET, ESH-12, LANL
MARIO E. SCHILLACI, ESH-12, LANL

Submitted to: Proceedings of 1997 Workshop on Bayesian Statistical Methods for Bioassay, Radiochemistry, and Internal Dosimetry. Charleston, SC November 13-14, 1997

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. 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. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. The Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

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.

DISCLAIMER

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

Analyzing Bioassay Data Using Bayesian Methods – A Primer

by

Guthrie Miller, William C. Inkret,
and Mario E. Schillaci
Los Alamos National Laboratory
Los Alamos, NM 87545, U. S. A.

October 16, 1997

Abstract

The classical statistics approach used in health physics for the interpretation of measurements is deficient in that it does not allow for the consideration of "needle in a haystack" effects, where events that are rare in a population are being detected. In fact, this is often the case in health physics measurements, and the false positive fraction is often very large using the prescriptions of classical statistics. Bayesian statistics provides an objective methodology to ensure acceptably small false positive fractions. We present the basic methodology and a heuristic discussion. Examples are given using numerically generated and real bioassay data (Tritium). Various analytical models are used to fit the prior probability distribution, in order to test the sensitivity to choice of model. Parametric studies show that the normalized Bayesian decision level $k_\alpha - L_c/\sigma_0$, where σ_0 is the measurement uncertainty for zero true amount, is usually in the range from 3 to 5 depending on the true positive rate. Four times σ_0 rather than approximately two times σ_0, as in classical statistics, would often seem a better choice for the decision level.

1 Introduction

The goal of internal dosimetry is the determination of intakes of radioactive materials into the body from limited bioassay data showing the amount excreted from the body, for example, in urine. Objective and analytical interpretation of urine concentration measurement results is difficult when the magnitude of the measurement result approaches the magnitude of the measurement process uncertainty. In these cases, when may one conclude, with a reasonable degree of certainty, that the measurement results indicates a real signal and not statistical fluctuation in background noise?

A classical approach to this problem has been well developed by Altschuler and Pasternack (1963) and Currie(1968) . The application of this classical theory has been boiled down to widely used (and misused) recipes and rules of thumb [2]. A weakness of the classical approach is its inability to adequately address the case of false-positives. In this paper we will show that to evaluate false-positive fractions requires a Bayesian framework, which involves the use of a prior probability distribution.

Bayesian methods are used in other fields (see for example [6]) and have recently been applied to problems encountered in the health physics profession [8] (Harvel). The purpose of this paper is to more broadly communicate the basic ideas and concepts of Bayesian statistics and to increase the professional awareness of these powerful and applicable concepts. Although the Bayesian approach is more mathematically complicated than the classical approach, with the use of available software tools on desktop computers Bayesian concepts will become increasingly accessible for routine health physics work.

2 Formulation of the Problem

We discuss a measurement process to determine a contaminant at the lowest possible level (where statistical uncertainty is significant). For example, we might consider the measurement of tritium, uranium, or plutonium in urine in order to determine a possible internal dose. We denote the measured result by y, and its statistical uncertainty standard deviation by σ. The (unknown) intake of the contaminant is denoted by x. The forward, theoretical calculation of measured amount y from x is denoted by the function $Y(x)$. In most cases, $Y(x)$ is a simple linear function of x, so intake x and true (but unknown) excretion Y are for many purposes interchangeable. In this paper it is assumed that the measurement process is well understood and that measurement uncertainties have a Gaussian distribution. The conditional probability of measuring result y given true amount x is given by

$$P(y|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left[-\frac{(y - Y(x))^2}{2\sigma^2(x)}\right]. \qquad (1)$$

(Conditional probabilities are written in the form $P(A|B)$, which means the probability of event A given event B. In general, $P(A)$ denotes the probability

of event A. However, if y is a continuous variable, the probability that y is in a small interval dy is $P(y)dy$.) When considered as a function of x, $P(y|x)$ is called the likelihood function, when considered as a function of y, it is called the sampling function.

The quantity x represents the quantity of greatest interest, for example, an occupational intake of uranium from inhalation as opposed to an environmental intake from drinking water. The quantity y represents the measured bioassay quantity. The measurement error σ summarizes all the uncertainties in going from measurement of y to determination of intake x, including uncertainties in background subtraction. In the case of a subtracted background, $Y(x) = y - y_B$, with y_B the background, for example an environmental background subtraction, or subtraction of urine excretion coming from earlier intakes. Thus a component of σ^2 is σ_B^2 coming from the uncertainty of the background subtraction.

For the measurement process itself, the uncertainty standard deviation $\sigma(Y)$ usually varies only moderately with true amount Y and can be represented as a Taylor expansion in powers of Y up to quadratic terms, as follows:

$$\sigma^2(Y) = \sigma_0^2 + bY + cY^2. \tag{2}$$

The quantity σ_0 will be important in the remainder of the paper. It is the measurement uncertainty standard deviation for zero true amount. The representation of Eq. 2 will be the starting point in this paper for the discussion of measurement error uncertainties. As discussed in Appendix A, for measurement processes using counting, the coefficient b in Eq. 2 is approximately the number of physical units (say dpm) per count, while c is usually small.

Besides measurement error uncertainty, another important source of uncertainty is the variation caused by sample collection protocols and biological variability, which contribute to the coefficient c. If the actual excretion rate is fixed, these effects cause a certain distribution of measured results around the true amount, with a certain standard deviation. This standard deviation divided by the mean of the distribution (the coefficient of variation), contributes to the square root of the coefficient c. In an experimental investigation of variations of the excretion of plutonium in urine, Moss [10] found a 10% coefficient of variation for true 24 hr samples, and a 70% variation for spot samples. The use of creatinine excretion to normalize urine excretion rates, as is commonly done in routine medical practice, may be advantageous. Anderson et. al. [4] have measured the average coefficient of variation of creatinine excretion using commercially available equipment for seven subjects and found it to be 7%.

3 Decision level and detection limit

When the measured value y exceeds a critical level L_c, called the decision level, the result is termed "detected" or "positive". The decision level scaled to the measurement error uncertainty for zero true amount, σ_0, defines a quantity

$$k = \frac{L_c}{\sigma_0}. \tag{3}$$

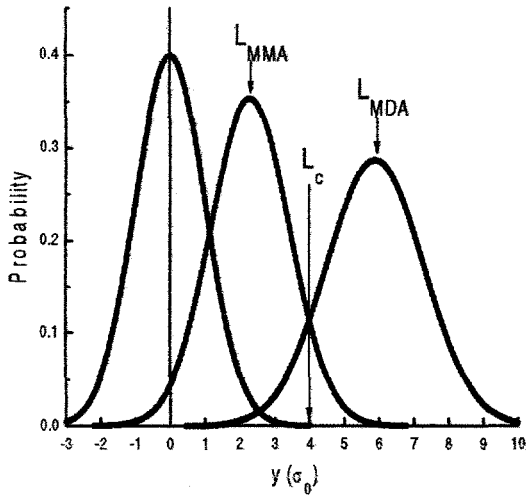

Figure 1: Distributions of measured results for $Y = 0$, L_{MMA} and L_{MDA}, showing the decision level L_c.

It is useful to associate another quantity α with k, where α is the fraction of measured results that exceed L_c when the true amount is 0 (see Fig. 1). A rough, but usually sufficiently accurate, approximation for α is

$$\alpha = \int_{L_c}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(\frac{y^2}{2\sigma_0^2} \right) dy \approx \frac{1}{\sqrt{2\pi}k_\alpha} \exp\left[\frac{k_\alpha^2}{2} \right]. \qquad (4)$$

More accurate values of α given k_α may be obtained from tables of single-tail areas of a normal distribution, or by numerical integration. The quantity α is useful in that it directly gives the minimum fraction of "positives" to be expected in the measurement process.

In classical statistics, the decision level is determined by specifying α to be a small number (say 0.05). In Bayesian statistics the decision level and k_α are determined by specifying the desired maximum false positive fraction, as will be discussed, which often results in a much smaller value of α.

The detection limits $Y = L_{MDA}$ and L_{MMA} are graphically depicted in Fig. 1. The detection limit L_{MDA} is the minimum true amount that is detected with high probability $1 - \beta$ (β is a given small number, say 0.05), while L_{MMA} is the maximum true amount that is missed with the same high probability. For example, stating the same thing for L_{MDA} in slightly different words, if the true result exceeds the detection limit L_{MDA}, the measured result will exceed the decision level L_c "most of the time", where "most of the time" means with probability at least $1 - \beta$. The quantities L_{MDA} and L_{MMA} are obtained as

the two roots Y of the (quadratic) equation

$$L_c \pm k_\beta \sigma(Y) = Y, \tag{5}$$

which are

$$Y = \frac{2L_c + k_\beta^2 b}{1 - k_\beta^2 c} \frac{1}{2} \left(1 \pm \sqrt{1 - \frac{4(1 - k_\beta^2 c)(L_c^2 - (k_\beta \sigma_0)^2)}{(2L_c + k_\beta^2 b)^2}} \right). \tag{6}$$

Equation 5 uses a quantity k_β which, in analogy with k_α defined in Eq. 4, gives the number of standard deviations σ one must be away from the peak value of the Gaussian distribution of measured results in order to have the single-tail integral be β. Equation 5 expresses the fact that if the true amount exceeds the decision level by $k_\beta \sigma$, there will be only a small chance β that the measured result is less than the decision level, and a similar statement for the case when the true amount is less than the decision level.

Equation 6 applies for both classical and Bayesian decision levels L_c. One finds that for $L_c = k_\beta \sigma_0$, as would be the case for classical statistics, the detection limit L_{MMA} is zero (usually $k_\alpha = k_\beta$ in classical decision theory, with $\alpha = \beta = 0.05$). The Bayesian decision level has k_α greater than k_β.

Although not related to Bayesian statistics, biological variability and bioassay sample collection uncertainty are important sources of uncertainty, often overlooked. There is a singularity in the expression given by Eq. 6, when $1 - k_\beta^2 c = 0$. This singularity means that k_β is limited to be less than $1/\sqrt{c}$, and therefore for bioassay procedures with very large biological variability uncertainty, for example fecal sampling, it is impossible to achieve small values of β.

Note that decision levels refer to measured quantities and detection limits refer to true amounts. Detection limits are usually used in theoretical calculations, while decision levels are used to interpret data.

4 Heuristic example of Bayesian decision theory

Imagine a measurement process with only two outcomes, $y = m_+$ (measure plus), and $y = m_0$ (measure zero). The true condition x likewise has only two values, + and 0. This model problem is unrealistic, since the measured values and true amounts are discrete and take only two values, nevertheless it is similar enough to the real problem to be instructive, and the mathematics are simplified so that everything can be explicitly calculated. The sampling function $P(y|x)$ is known and given by the following matrix (the sampling function is the probability of measuring result y given true condition x considered as a function of y):

$$\begin{bmatrix} P(m_+|+) & P(m_+|0) \\ P(m_0|+) & P(m_0|0) \end{bmatrix} = \begin{bmatrix} 1-\beta & \alpha \\ \beta & 1-\alpha \end{bmatrix}, \tag{7}$$

in terms of two parameters α and β, which are similar, but not identical to, the parameters α and β introduced in the previous section. Note that

$$\sum_y P(y|x) - 1, \tag{8}$$

since conditional probabilities are normalized. For example, given that the true condition is +, there are only two possibilities, measure + and measure 0, with probabilities $1-\beta$ and β, so that

$$P(m_+|+) + P(m_0|+) = 1 - \beta + \beta = 1 \tag{9}$$

Assume that a fairly large number N of measurements are taken for similar cases. In that population of cases, the number of measured positives and zeros are given by the following matrix equation:

$$\begin{aligned} \begin{bmatrix} n(+) \\ n(0) \end{bmatrix} &= N \begin{bmatrix} P(m_+|+) & P(m_+|0) \\ P(m_0|+) & P(m_0|0) \end{bmatrix} \begin{bmatrix} P(+) \\ P(0) \end{bmatrix} \\ &= N \begin{bmatrix} 1-\beta & \alpha \\ \beta & 1-\alpha \end{bmatrix} \begin{bmatrix} P(+) \\ P(0) \end{bmatrix}, \end{aligned} \tag{10}$$

in terms of the prior probability distribution $P(+)$ and $P(0)$ giving the probabilities of true + and 0 in the measured population. Note that we can solve Eq. 10 for the prior probabilities, giving the result

$$\begin{bmatrix} P(+) \\ P(0) \end{bmatrix} = \frac{1}{N(1-\alpha-\beta)} \begin{bmatrix} 1-\alpha & -\alpha \\ -\beta & 1-\beta \end{bmatrix} \begin{bmatrix} n(+) \\ n(0) \end{bmatrix}, \tag{11}$$

which shows how the prior probability distribution can be determined from population data. In what follows we assume that the prior probability distribution is a given fixed, unknown quantity that may be estimated from population averages using Eq. 11.

Now return to Eq. 10 and express $n(+)$, the number of positive measured results, in terms of the number $n_+(+)$ coming from true positives and $n_0(+)$ coming from true zeros. From Eq. 10,

$$\begin{aligned} n_+(+) &= NP(m_+|+)P(+) \\ n_0(+) &= NP(m_+|0)P(0). \end{aligned} \tag{12}$$

Similarly,

$$\begin{aligned} n_+(0) &= NP(m_0|+)P(+) \\ n_0(0) &= NP(m_0|0)P(0). \end{aligned} \tag{13}$$

We now ask some general questions about a measurement process involving uncertainty, and answer them within the model. The errors of the measurement process result in the cross terms $n_0(+)$, the number of true zeros that measure +, and $n_1(0)$, the number of true positives that measure 0. The smallness of

these cross terms can be judged by comparing them with the total number of true zeros or positives, or the total number that measure zero or positive. There are four natural fractions that should be small for the measurement process to be effective. These are:

1. The false positive fraction, the fraction of all results that measure positive that come from true zeros, given by

$$f_1 = \frac{n_0(+)}{n(+)}, \tag{14}$$

2. The false negative fraction, the fraction of all results that measure zero that come from true positives, given by

$$f_2 = \frac{n_+(0)}{n(0)}, \tag{15}$$

3. The missed positives, the fraction of all true positives that are measured as zero, given by

$$f_3 = \frac{n_+(0)}{n_+}, \tag{16}$$

4. The missed negatives, the fraction of all true zeros that are measured as positive, given by

$$f_4 = \frac{n_0(+)}{n_0}. \tag{17}$$

Substituting from Eqs. 12 and 13, these fractions are as follows:

1. false positive fraction

$$f_1 = \frac{P(m_+|0)P(0)}{P(m_+|+)P(+) + P(m_+|0)P(0)} \tag{18}$$

$$= \frac{\alpha P(0)}{(1-\beta)P(+) + \alpha P(0)} = \frac{\alpha}{\alpha + (1-\beta)\delta}, \tag{19}$$

where

$$\delta = \frac{P(+)}{P(0)}. \tag{20}$$

Equation 18, which actually derives Bayes theorem, gives the probability $P(0|m_+)$ of true result 0 given that the measured result is +. Bayes theorem states that

$$P(x|y) = \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)}. \tag{21}$$

2. false negative fraction

$$f_2 = \frac{P(m_0|+)P(+)}{P(m_0|+)P(+) + P(m_0|0)P(0)} = \frac{\beta P(0)}{\beta P(+) + (1-\alpha)P(0)} = \frac{\beta}{\beta\delta + (1-\alpha)}. \quad (22)$$

3. missed positive fraction

$$f_3 = \frac{P(m_0|+)P(+)}{P(+)} = \beta \quad (23)$$

4. missed negative fraction

$$f_4 = \frac{P(m_+|0)P(0)}{P(0)} = \alpha \quad (24)$$

In classical statistics, the quantities $\alpha = P(m_+|0)$ and $\beta = P(m_0|+)$, giving errors of the first and second kind, are specified as small numbers. In Bayesian statistics, we require in addition that the false positive and false negative fractions be small. This means, from Eq. 19, that if want the false positive fraction to be a given small number ϵ, then α must be

$$\alpha = \frac{\epsilon\delta(1-\beta)}{1-\epsilon}. \quad (25)$$

Therefore if δ is small (true positives rare in the population), α must be very small. Similarly, if δ is large (true negatives rare in the population), β must be very small in order to have a small false negative fraction. In other words, it is hard to find a needle in a haystack!

This model shows how the prior probability distribution can be determined from population averages, and how Bayesian statistics is nothing more than a more inclusive modeling of the measurement process, which brings in "needle in a haystack" effects. The parameters α and β in classical statistics are constrained by considering these Bayesian effects and can be determined by requiring acceptably small false positive and false negative fractions. The usual case in analysis of bioassay data for internal dosimetry is that true positives are rare in the population rather than true negatives being rare. In this situation, the α of classical statistics is further constrained and is determined by specifying an acceptable false positive fraction.

5 Review of Bayesian Statistics

In Bayesian statistics the problem is formulated in terms of the quantity of greatest interest x, even though this quantity is only indirectly related to the actual measured quantity y. So the mathematical problem becomes an inverse

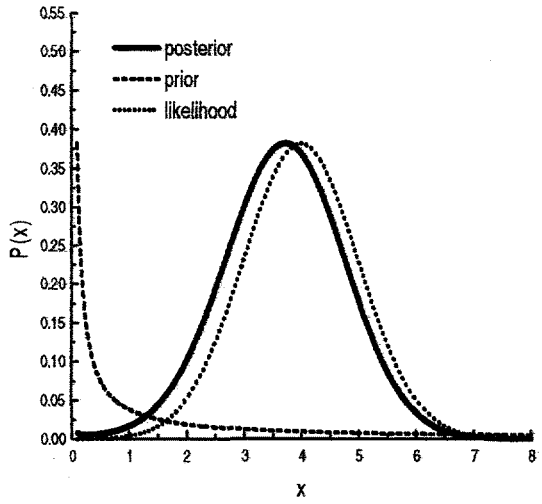

Figure 2: Posterior, prior and likelihood distributions when the measurement uncertainty $\sigma = 1$ and the measured result $y = 4$.

problem, to determine x from y. We wish to determine the posterior probability distribution $P(x|y)$, the probability distribution of x given that we have measured y. In order to do this, we need an estimate of the prior probability distribution $P(x)$, the probability distribution of x in the population we are measuring. By Bayes theorem,

$$P(x|y) = \frac{P(y|x)P(x)}{\int_0^\infty P(y|x)P(x)\,dx}. \tag{26}$$

Consider an example where the measurement error uncertainty σ is taken as equal to 1 and the forward biokinetic model as $Y(x) = x$. The prior probability distribution is taken as the scale invariant distribution

$$P(x)\,dx \propto \frac{dx}{x}. \tag{27}$$

This distribution is uniform when plotted on a logarithmic scale, and is a natural "uniform" distribution. Note that this distribution must be truncated at small and large values to obtain a normalizable probability distribution.

Assume that the measurement is several standard deviations from zero, say $y = 4$. The prior, likelihood, and posterior distributions are shown in Fig. 2. The posterior distribution is, except for a normalization factor, just the product of the likelihood function and the prior distribution. The normalization factor ensures that the integral of the posterior distribution is 1, as must be true for

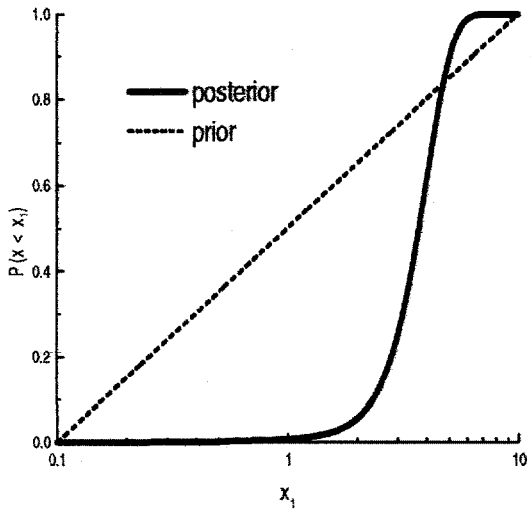

Figure 3: Cumulative posterior and prior probability distributions when the measurement uncertainty $\sigma = 1$ and the measured result $y = 4$.

any probability distribution. Because of the slope of the prior distribution in Fig. 2, the posterior distribution differs somewhat from the likelihood function, which, in this example, is a simple Gaussian peaked at the measured value.

We may also consider the integrated probability from 0 up to some value x_1. Fig. 3 shows integrated, or cumulative probabilities for the case shown in Fig. 2. In this figure a logarithmic scale is used for the x_1 axis so that the cumulative prior probability distribution is a straight line. In the measured population, x is distributed from 0.1 to 10. with uniform (log scale) probability. As a result of the measurement $y = 4$, we have better knowledge of the actual value of x, and can assign a probability interval from x_l, the lower probability limit, to x_u, the upper probability limit, corresponding say to the 10% and 90% probability points in Fig. 3. In general, we define x_l so that the probability is only ϵ_l that x is smaller than x_l, and x_u so that the probability is only ϵ_u that x is larger than x_u. As a result of the measurement we have reasonable certainty (probability $1 - \epsilon_l - \epsilon_u$) that the quantity of interest x lies somewhere between x_l and x_u.

6 Interpretation of Bayesian results–"detected"

The Bayesian methodology allows us to calculate the posterior probability distribution of the quantity x, and from that to obtain a confidence interval x_l to x_u. For many purposes, this is sufficient, for example, when x_u/x_l is near 1, it is reasonable to state that x has been measured to be in the range x_l to x_u. We wish, however, to define the concept of "detected" at the lowest possible level. In some cases, x_u/x_l may be much larger than 1, and the confidence interval by itself is not sufficient. By "detected" we mean that the measurement definitely reveals the presence of the contaminant at some, possibly very small, level, and we would expect this to be borne out by subsequent measurements.

The definition of "detected" obviously is that x_l exceeds some critical level x_c. There are two considerations affecting the choice of x_c: 1) that x_c be in the range where we have some direct empirical knowledge of the prior distribution based on real data, and 2) that x_c be much greater than the lower probability limit of the prior distribution. To satisfy consideration 1) x_c should be no smaller than some moderate fraction (say 0.1) of $Y^{-1}(\sigma_0)$ (Y is the function, usually linear, relating intake x to excretion Y, Y^{-1} is its inverse function, and σ_0 is the measurement error uncertainty for $x = 0$. To simplify the subsequent discussion, $Y(x) = x$ will be assumed). The measurement error occludes any knowledge of the structure of the prior distribution for smaller values of x. Regarding consideration 2), if x_c were comparable or smaller than the lower probability limit of the prior distribution, we might call a case "detected" when the measurement result was zero, or without any measurements, which would be counterintuitive. Consideration 1) is ensured by choosing $x_c = 0.1\sigma_0$. Consideration 2) is ensured by choosing

$$\epsilon_l = \epsilon \int_0^{x_c} P(x)dx, \tag{28}$$

where ϵ is some small number (say 0.1). Equation 28 ensures that for a case that is "detected", the posterior probability curve will be shifted toward larger values (to the right) relative to the prior probability curve in a plot like that shown in Fig. 4. Figure 4 shows a marginally "detected" case where $x_c = 0.1\sigma_0$. The cumulative prior probability up to $x_1 = x_c$ is 0.5, while the cumulative posterior probability up to $x_1 = x_c$ is $\epsilon_l = 0.05$, which is $\epsilon = 0.1$ times the cumulative prior probability, satisfying Eq. 28.

In many cases of interest, the cumulative prior probability up to $0.1\sigma_0$ is near 1, so consideration 2) above is not critical. In these cases the measured population contains mostly very small (relative to σ_0) values of x.

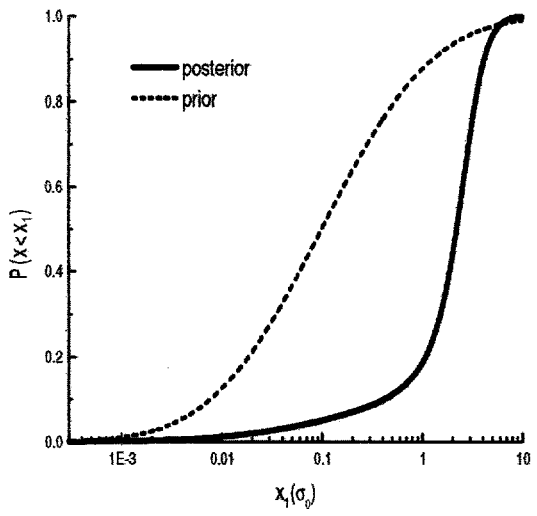

Figure 4: Cumulative posterior and prior probability distributions for a marginally "detected" case.

7 Discussion of measurement objectives, types of errors

Up until now, the objective of the measurement has been to determine "detected" by having the measured value exceed a critical level L_c. The objective of the measurement could, on the other hand, be to determine that the intake is less than a certain amount (for example, an intake amount corresponding to a committed effective dose equivalent of 1 mSv). We might then define "small" to mean that the upper Bayesian confidence limit x_u be less than a specified amount x_{SMALL}. This would require that the measured value be less than a certain critical level. It is natural to have this latter critical level be the same as L_c, which means that "not detected" is the same as "small", and conversely. This determines x_{SMALL} to be the upper Bayesian confidence limit when the measured value is at the detection limit L_c, which we denote by $x_u(L_c)$ Implicit in this construction are the quantities ϵ_l and ϵ_u that define the Bayesian confidence interval x_l, x_u. The condition "detected" has false positive fraction ϵ_l, and "small" has false negative fraction ϵ_u, both of which need to be small numbers for the determinations to be meaningful. Note that "detected" means that the intake amount is likely to be greater than a quantity we have called x_c in Section 6, and which we might denote by $x_l(L_c)$, while "small" means that the intake is likely to be smaller than a larger quantity $x_u(L_c)$.

Similar to $x_u(L_c)$ but slightly different is the detection limit L_{MDA}, which

is a simpler and easier to evaluate quantity, since it does not involve the prior probability distribution other than having a dependence on L_c. If the true amount exceeds L_{MDA} the measured result is likely to be positive. The statement relating $x_u(L_c)$ to this is: if the measured amount is not positive ("small") the true amount is likely to be less than $x_u(L_c)$. These concepts would be useful in setting up an internal dosimetry program to meet legal requirements, for example "internal dose evaluation programs shall be adequate to demonstrate compliance with .." [1], in which case L_{MDA} or possibly $x_u(L_c)$ would correspond to some amount less than the legal dose limit (to allow, say for external dose).

The detection limit L_{MMA}, the maximum missed amount, is of interest in quantifying missed doses. An intake of this amount or less is likely to be missed.

8 Form of the prior distribution

The prior probability distribution is the probability distribution of x in the measured population, denoted by $P(x)$. Our approach to the prior distribution is to determine it as much as possible empirically from data. In practice this means fitting data with analytical forms having variable parameters. We will use analytical distributions that allow a wide range of variation. Certainly the value and the slope of the prior distribution at a point x corresponding to y several standard deviations from zero should be independently variable. The mean and standard deviation of a probability distribution are perhaps it's most basic parameters. These should be independently variable. So the analytical forms used should have at least two parameters. The other constraint is that the distribution apply to a positive quantity x.

Two analytical distributions that we will use are 1) the log-normal distribution, and 2) the Pareto distribution. These are defined in Table 1 below.[11]

Table 1: Analytical forms used for the prior probability distribution

distribution	analytical form	mean	standard deviation
log-normal	$\frac{1}{\sqrt{2\pi}\sigma_g x}\exp[\frac{-\left(\log\frac{x}{x_0}\right)^2}{2\sigma_g^2}]$	$x_0\exp(\sigma_g^2/2)$	$x_0\exp(\sigma_g^2/2)\sqrt{\exp(\sigma_g^2)-1}$
Pareto	$\frac{1}{px}\left(\frac{x_0}{x}\right)^p$	$\frac{px_0}{p-1}$	$\frac{px_0}{p-1}\sqrt{\frac{p-1}{p(p-2)}}-1$

Each of these distributions has two positive parameters. For the log-normal distribution, x_0 is the maximum probability point of the distribution when plotted on a log scale, and σ_g, the geometric standard deviation, defines the width of the distribution on a log scale. The maximum of the Pareto distribution occurs at x_0 and the distribution is defined as zero for $x < x_0$. When the power p in the Pareto distribution becomes small, the distribution approaches the uniform distribution already discussed, although in the limit $p \to 0$ the integral of the distribution diverges over the range $x_0 < x < \infty$ and the distribution

must be truncated to zero at large values of x. For p less than 1, the formula for the mean of the Pareto distribution is more complicated, depending on the maximum value of x in the integration. The same applies to the formula for the standard deviation of the Pareto distribution, when p is less than 2.

When the quantity x is an internal dosimetry intake occurring during a bioassay sampling interval, there are some additional general considerations [9]. Let Δt represent the time interval between bioassay samples. Consider the case where intakes are infrequent, and imagine the limit $\Delta t \to 0$. In this case the prior distribution has the limit

$$P(x) \to \delta(x), \tag{29}$$

where $\delta(x)$ is the delta function. (The delta function, $\delta(x)$, is the limit of very narrow distributions peaked at $x = 0$ and having unit integral, $\int \delta(x)\,dx = 1$). Equation 29 implies that if the bioassay sampling interval is very short, there will not have been time for any intakes to occur. For small time interval Δt, the prior has the form

$$P(x) = \delta(x)(1 - \lambda\Delta t) + \lambda\Delta t w(x), \tag{30}$$

where $w(x)$ is some normalized distribution of positive quantity x, and $\lambda\Delta t$ is the average number of intakes occurring in time interval Δt (λ is the average number of intakes occurring per unit time). As a result of these considerations, it may be reasonable to include a delta-function component in the analytical form representing the prior distribution.

9 Numerical experiments using simulated data

In investigating questions such as what effect the assumption of a particular form of the prior distribution makes, it is useful to have a controlled experiment, where we know the correct answer. So, we have performed a series of numerical experiments using Monte-Carlo generated data. The data were generated from a log-normal prior distribution, with normal measurement errors. Let z_1 and z_2 be random variables from normal distributions with zero mean and unit standard deviation. Then the intake x was generated as

$$x = x_0 \exp(\sigma_g z_1), \tag{31}$$

and the measured value was generated as

$$y = Y(x) + \sigma(x) z_2, \tag{32}$$

where $Y(x)$ is the forward biokinetic model that relates intake to excretion.

The distribution of measured results for a population of workers in a routine sampling program might be as shown in Fig. 5, using numerically generated data. Such a distribution would be made by including only cases where the preceding bioassay result was zero by some criteria (say $y < 1.645\sigma_0$), and for

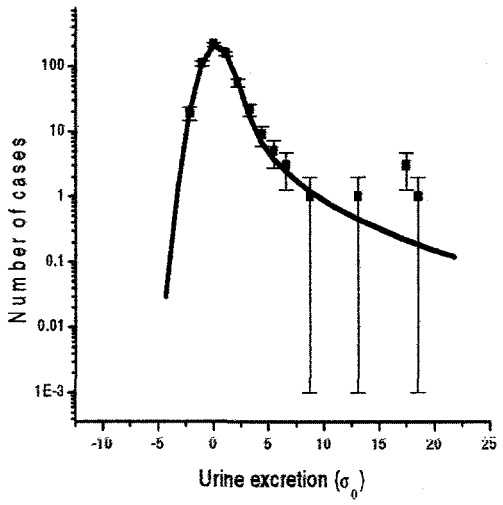

Figure 5: Distribution of simulated measured urine bioassay results.

long biological half-life materials like plutonium would exclude persons involved in known incidents and persons who were historically positive. The distribution thus represents intakes occurring during the sampling interval. The fitting curve shown in Fig. 5 will be discussed in the next section.

10 Bayesian decision theory

We will discuss the interpretation of numerically generated bioassay data. Shown in Fig. 5 is a fit based on the equation

$$n(y) = N\Delta y \int_0^\infty P(y|x)P(x)dx, \tag{33}$$

where $n(y)$ is the observed number of cases with measured values in the bin Δy around y, N is the total number of cases, $P(y|x)$ is the known probability of measuring result y if the true result is x, and $P(x)dx$ is the unknown prior probability that the true amount x is in interval dx. The fit shown in Fig. 5 is obtained by varying parameters in a representation of $P(x)$, as discussed in Ref. [8]. In this case the fit, like the generated data, assumes a log-normal form for the prior distribution, and successfully reproduces the parameters used to generate the data, as long as a sufficiently large data set is available, as shown in Table 2 below.

In Fig. 6 the quantities $n(+)(y)$ and $n_0(y)$, defined by

Table 2: Fits to the distribution of measured bioassay results for different data set sizes.

N	σ_0	Y_0	σ_g
60	0.905	0.084	2.16
600	0.92	0.137	1.75
6000	0.973	0.104	1.95
60000	0.979	0.103	1.97
actual	1.	0.1	2.

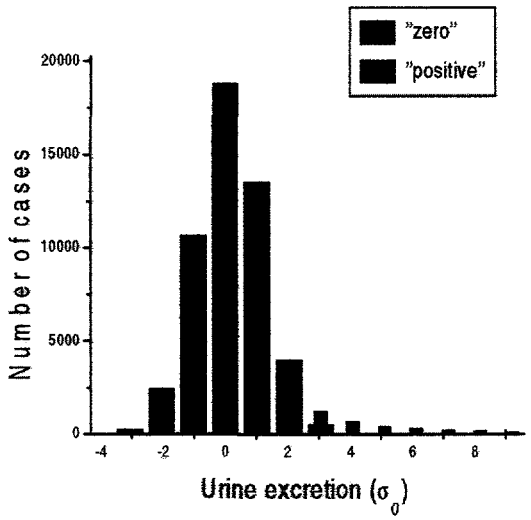

Figure 6: "Positive" and "zero" components of the distribution of simulated measured bioassay results.

$$n(+)(y) = N\Delta y \int_{x_c}^{\infty} P(y|x)P(x)dx$$
$$n_0(y) = N\Delta y \int_{0}^{x_c} P(y|x)P(x)dx \quad (34)$$

are displayed. This shows how the fitted model representing the observed distribution can be broken down into contributions from the two classes "positive", with $x > x_c$, and "zero", with $x < x_c$, where as discussed in Sec. 6

$$x_c = 0.1\sigma_0, \quad (35)$$

where σ_0 is the measurement uncertainty for zero true amount. Under these conditions ($x_c << \sigma_0$),

$$n_0(y) \approx N\Delta y P_0 P(y|0), \quad (36)$$

where

$$P_0 = \int_0^{x_c} P(x)dx \quad (37)$$

is the prior probability of "zero".

The fraction of cases coming from "zeros", given by

$$f_0(y) = \frac{n_0(y)}{n_0(y) + n(+)(y)} \quad (38)$$

is shown in Fig. 7. The fraction $f_0(y)$ can be written as

$$f_0(y) = \int_0^{x_c} P(x|y)dx \quad (39)$$

in terms of $P(x|y)$, the probability that the true amount is x given measured value y. Using Bayes theorem to calculate $P(x|y)$ gives Eq. 38.

The quantity $f_0(y)$ is the false positive fraction, which should be bounded by a small quantity in order to have a valid determination of "positive", that is

$$f_0(y) < \epsilon. \quad (40)$$

For example, in Fig. 6 for a bin somewhere between $y = 2.$ and $3\sigma_0$ there are equal numbers of cases from "positives" and "zeros". So, with that as the decision level, the true result may be zero 50% of the time. In internal dosimetry, where having a small amount of a radioactive substance in one's body is often emotionally disturbing but probably not physically harmful, it is reasonable to report an intake only if we are quite certain that the intake has actually occurred. Thus a smaller false positive fraction of $\epsilon = 5\%$, as obtained from Fig. 7, is more appropriate than over 70%, which goes along with the usual classical decision level of $1.645\sigma_0$.

The Bayesian analysis allows us to determine a meaningful decision level y_c, such that if the measured result y exceeds y_c, the true result is very likely positive. For the example shown in Figs. 5, 6, and 7, the normalized decision

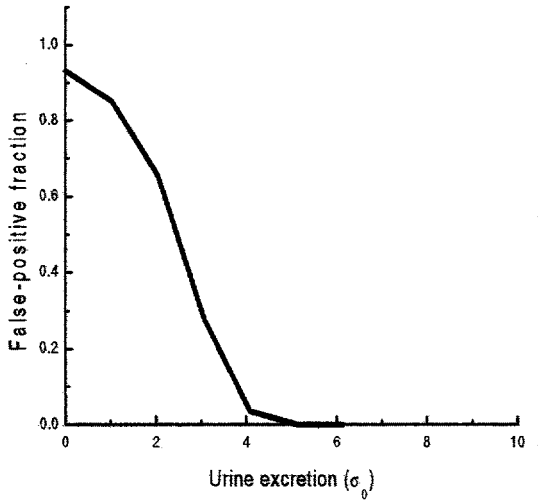

Figure 7: False positive fraction.

level is approximately $k_\alpha = y_c/\sigma_0 \approx 4$ for a maximum false positive fraction of $\epsilon = 5\%$.

The Bayesian decision level may be expressed in terms of an α just as the classical decision level, where α gives the fraction of the time the measured result will exceed the decision level, if the true result is zero. There is basically little connection between α and the false positive fraction, however the following can be proven (see Ref. [8]). Let ϵ be the maximum false positive fraction and P_0 be the prior probability of "zero" in the Bayesian analysis. Then the Bayesian decision level always exceeds $k_{\alpha_c}\sigma_0$, where k_{α_c} corresponds to

$$\alpha_c = \frac{\epsilon}{P_0} \tag{41}$$

using Eq. 4. The quantity P_0 is the false positive fraction without additional information provided by measurements, so the normalization in Eq. 41 is natural. Normally P_0 is near 1 so it is not an important numerical factor. If we define an α that goes along with the Bayesian decision level using Eq. 4, this α will be less than α_c, usually much less. In the example, the normalized Bayesian decision level is $k_\alpha = 4$ which gives

$$\alpha = 3.3\,10^{-5}, \tag{42}$$

much less than the false positive fraction $\epsilon = 0.05 \approx \alpha_c$.

The analysis in the case of remeasurement is the same, except that the measurement uncertainty standard deviation given by Eq. 2 is decreased by a

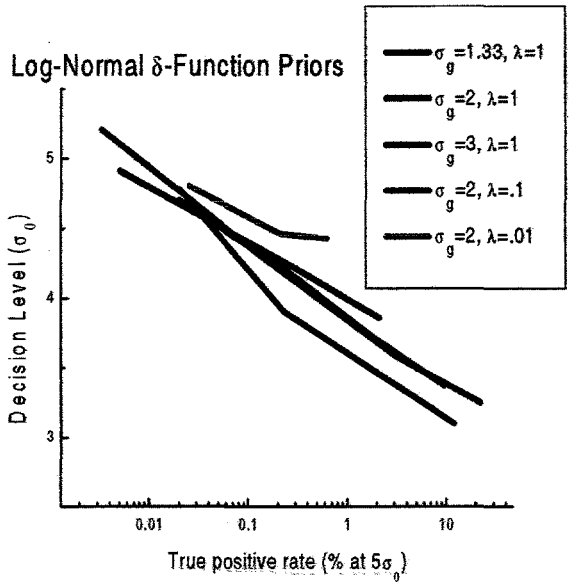

Figure 8: Bayesian decision level versus the "true" positive rate, for various log-normal plus δ-function prior distributions.

factor of $1/\sqrt{N}$, where N is the number of measurements that are averaged to make the final determination.

11 Parametric study of Bayesian decision levels

The Bayesian decision level depends on the prior distribution, so it might seem that it could take practically any value. In fact, for a broad range of prior distributions, the Bayesian decision level is approximately correlated with the true positive rate in the population. This is illustrated in Fig. 8. The "true" positive rate is the rate at which measured cases exceed 5σ. For the prior distributions shown, the false positive fraction at $5\sigma_0$ is very small, so these events are all "true" positives. The log-normal plus δ-function prior distribution has the parameters σ_g, Y_0, and λ. The strength of the δ-function component of the prior distribution is $1-\lambda$. The values of σ_g and λ are shown in Fig. 8, and the value of Y_0 is varied to trace out each curve shown. The Bayesian decision level (for 0.05 false positive fraction) is seen to range from about $3\sigma_0$ to about $5\sigma_0$, much larger than the usual classical decision level of $1.645\sigma_0$.

12 Bayesian analysis of Tritium bioassay data

As an illustration of the method, we consider the determination of a decision level for a group of about 80 workers on bi-weekly bioassay for Tritium. A 200 μL aliquot of urine is counted for 10 minutes using a liquid scintillator. All bioassay results that are preceded by a "zero" result are assembled into a data set. These bioassay results represent urine excretion from intakes that may have occurred in the preceding monitoring interval, and not before, and therefore reflect the probability that an intake occurred during the monitoring interval.

The definition of the preceding "zero" is that the measured result is below a certain critical level. This critical level is as small as possible as long as a sufficiently large number of cases are produced. The choice $1.645\,\sigma_0$, which allows 95 % of true zeros to measure "zero", is an upper limit. Using this choice for the Tritium data, we obtain about 350 cases, using data from the past year.

This data set is histogramed into bins, so an experimental distribution of measured results is obtained. That experimental distribution is then fit using various theoretical models of the prior distribution convoluted with a Gaussian measurement error distribution. The results of 4 such fits assuming a log normal prior or a Pareto prior, with or without a δ-function component are shown in Table 3 below.

Table 3: Fits to the distribution of measured Tritium bioassay results using different theoretical prior distributions

prior	χ^2/NDF	σ_0 (nCi/L)	Y_0 (nCi/L)	σ_g or p	λ	P_0	L_c (nCi/L)	L_c/σ_0
l+d	0.792	5.37	24.6	0.959	0.15	0.85	18.8	3.51
l	0.932	5.42	$3.9\,10^{-2}$	5.45	1.	0.7	19.6	3.61
p+d	0.925	5.42	0.7	0.217	0.34	0.64	19.6	3.61
p	0.95	5.47	$1.5\,10^{-3}$	0.177	1.0	0.66	20.	3.65

There are 14 histogram bins, and the number of fit parameters is either 4 for the log normal plus δ-function case, or 3 for the other cases. The quantity NDF, the number of degrees of freedom, is the number of data values (histogram bins) minus the number of fit parameters, it is 10 for the log normal plus δ-function case and 11 for the other cases. The value of χ^2/NDF should be about 1 for a satisfactory fit, which it is. The quantity λ gives the prior probability in the continuous portion of the distribution (not the δ-function part). It is automatically 1 for fits that do not have a δ-function component. The quantity P_0 is the cumulative prior probability up to 0.1 σ_0. The quantity L_c is the Bayesian decision level for a 5 % false positive fraction. As seen the Bayesian decision level lies between 3.51 and 3.65 σ_0, depending on the form of the prior distribution assumed.

In contrast, the usual classical decision level is $1.645\ \sigma_0$ or 8.83 nCi/L. This decision level results in a false positive fraction of 85%.

13 Unfolding data to determine multiple intakes

Data unfolding depends on causality, that excretion follows intake in time, so that intakes in a given sampling interval affect only subsequent urine data values and not urine data values preceding the intake. Thus, starting with the first sampling interval, the intake in each sampling interval and its standard deviation is determined by the right-hand urine data value of that interval as the expectation value and standard deviation of the Bayesian posterior probability distribution. The excretion expected from previous intakes and its propagated uncertainty is subtracted to determine if a new intake has occurred. If, as is usually the case, the sampling intervals are sufficiently short so that the average number of intakes in the interval is small, then it is likely that either 0 or 1 intakes occur in the interval (the probability of multiple intakes is small). If we assume the time of intake is the center time of the sampling interval, unfolding then becomes a sequence of one-dimensional integrals over intake amounts to determine expectation values and standard deviations of the posterior distribution (Miller and Inkret 1996). In contrast, the simple unfolding techniques used up until now have not been probabilistic. The intake in each sampling interval has been determined to match the right-hand urine excretion data value exactly, even if the required intake quantity was negative. Also there was no method to calculate uncertainties in the excretion expected from prior intakes, which is an important background that must be subtracted to determine if new intakes have occurred.

It is simplest to discuss data unfolding using prior probability models that contain a δ-function, representing "no intake". Then the probability of an intake can be defined as the integral of the posterior probability over all positive values of intake quantity x, without having to define x_c.

After carrying out the data unfolding as described, we obtain a set of intake quantities x occurring in each bioassay sampling interval, and we can calculate a probability associated with each intake. If the probability for a certain intake is low, it is reasonable to drop it entirely, and assume no intake took place in that sampling interval. The bioassay data value on the right side of the interval can then be used in combination with that on the left hand side of the interval to determine the previous intake.

A computer code (UF) has been developed based on this algorithm. Data unfolding and calculation of intake probabilities is first done, then the least probable intake is dropped, and the process is repeated, until all the remaining intake probabilities are at least 90% or some such high value. This method results in an intake scenario with relatively few, well determined, intakes, well suited to the regulatory requirement that all intakes be reported and justifiable.[1] An intake scenario with relatively few intakes results in higher dose estimates for cases with nonzero dose, as discussed in ref. [8]

14 Discussion

We have shown how a Bayesian analysis determines the decision level by specifying the maximum allowed false positive fraction. The prior probability distribution must be known. The prior probability distribution can be determined from data if a suitably large population of results is available.

Many recipes and descriptions of statistical methods have been written that start with a given classical decision level, for example $k_\alpha = 1.645$ ($\alpha = 0.05$) is often used. Guidance at the rule-of thumb level needs to be corrected to allow larger and variable k_α's, with k_α determined by proper Bayesian arguments. The health physicist may run computer codes to analyze results from large populations to directly determine decision levels (these codes will be downloadable from our Web site), or k_α results from similar facilities and work environments can be used based on professional judgment that the situations are comparable. Choosing $k_\alpha = 4$ is a better guess than $k_\alpha = 1.645$, which produces in most cases a very high false-positive fraction. It is interesting that experienced health physicists have at times used the classical MDA ($k_\alpha \approx 4$) rather than the classical decision level as a decision level, because it seemed to produce a more reasonable rate of "positives".

A Appendix

We assume a measurement process involving counting. The average value of the measured result y is given by

$$Y = f(N - B), \tag{43}$$

where f is a numerical factor, N is the average number of measured counts, and B is the average number of background counts. The variance of Y is given by

$$\sigma_Y^2 = \sigma_f^2(N - B)^2 + f^2(\sigma_N^2 + \sigma_B^2). \tag{44}$$

Since counts are usually governed by Poisson statistics (if the physical half life is long compared to the count time),

$$\sigma_N^2 = N. \tag{45}$$

Therefore,

$$\sigma_Y^2 = f^2(B + \sigma_B^2) + f^2(N - B) + \sigma_f^2(N - B)^2. \tag{46}$$

Equation 46 may be written in the form

$$\sigma^2(Y) = \sigma_0^2 + bY + cY^2, \tag{47}$$

and we thus find that

$$\begin{aligned} \sigma_0^2 &= f^2(B + \sigma_B^2) \\ b &= f \\ c &= \frac{\sigma_f^2}{f^2}. \end{aligned} \tag{48}$$

Using tracer methodology (where the sample is spiked with a known amount of a tracer isotope, whose activity can be distinguished by energy spectroscopy from the activity of the isotope of interest),

$$f = \frac{A_t}{N_t - B_t}, \tag{49}$$

where A_t is the known tracer activity and N_t and B_t are the measured tracer counts and background counts. Neglecting the usually small error in A_t,

$$\frac{\sigma_f}{f} = \sqrt{c} = \frac{\sqrt{N_t + \sigma_{B_t}^2}}{N_t - B_t} \approx \frac{1}{\sqrt{N_t}}, \tag{50}$$

for a large number of tracer counts (say $N_t \approx 1000$) at full recovery. Thus using tracer methodology, σ_f/f is small (say a few percent) so that c is very small, and b is approximately the factor relating net counts to physical units.

References

[1] Code of federal regulations 10, part 835. Office of the Federal Register, 1994.

[2] Performance criteria for radiobioassay. Technical Report HPS N13.30-1996, American National Standards Institute, Inc., 1313 Dolley Madison Blvd, Suite 402, McLean, VA, 22101, 1996.

[3] B. Altshuler and B. Pasternack. Statistical measures of the lower limit of detection of a radioactivity counter. *Health Physics*, 9:293, 1963.

[4] J. Anderson, B. Kahn, R. Kosson, E. Kim, and T. Labone. Use of creatinine measurement to confirm the time interval of a urine sample. *"Radiation Protection Management"*, 12:51–64, 1995.

[5] L. A. Currie. Limits for qualitative detection and quantitative determination, application to radiochemistry. *Analytical Chemistry*, 40:586, 1968.

[6] Kennth M. Hanson and Silver Richard N., editors. *Maximum Entropy and Bayesian Methods, Santa Fe 1995, Proceedings of the Fifteenth International Workshop*, Dordrecht, 1996. Kluwer Academic Publishers.

[7] G. Miller and W. C. Inkret. Bayesian methods for interpreting plutonium urinalysis data. In *Maximum Entropy and Bayesian Methods, Santa Fe, New Mexico, 1995*, pages 367–373, Boston, 1996. Kluwer Academic Publishers.

[8] G. Miller, W. C. Inkret, and H. F. Martz. Bayesian Detection Analysis for Radiation Exposure, II. *Radiation Protection Dosimetry*, 58:115–125, 1995.

[9] G. Miller, W. C. Inkret, and H. F. Martz. Internal dosimetry intake estimation using bayesian methods. *Radiation Protection Dosimetry*, 1997. LA-UR-96-2178, submitted for publication.

[10] W. D. Moss, Campbell E. E., Schulte H. F., and Tietjen G. L. A study of the variations found in plutonium urinary data. *Health Physics*, 17:571, 1969.

[11] V. Rothschild and N. Logothetis. *"Probability Distributions"*. John Wiley and Sons, New York, 1986.