Radiation Protection Dosimetry
Vol. 94, No. 4, pp. 347–352 (2001)
Nuclear Technology Publishing

BAYESIAN PRIOR PROBABILITY DISTRIBUTIONS FOR INTERNAL DOSIMETRY

G. Miller, W. C. Inkret, T. T. Little, H. F. Martz and M. E. Schillaci
Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received January 7 2000, in final revised form October 31 2000, accepted November 9 2000

Abstract — The problem of choosing a prior distribution for the Bayesian interpretation of measurements (specifically internal dosimetry measurements) is considered using a theoretical analysis and by examining historical tritium and plutonium urine bioassay data from Los Alamos. Two models for the prior probability distribution are proposed: (1) the log-normal distribution, when there is some additional information to determine the scale of the true result, and (2) the 'alpha' distribution (a simplified variant of the gamma distribution) when there is not. These models have been incorporated into version 3 of the Bayesian internal dosimetry code in use at Los Alamos (downloadable from our web site). Plutonium internal dosimetry at Los Alamos is now being done using prior probability distribution parameters determined self-consistently from population averages of Los Alamos data.

INTRODUCTION

The Bayesian approach to interpretation of measurements in health physics has several important advantages. It directly addresses the questions of greatest interest; for example, 'Did I have an intake of plutonium?' and 'With what probability?' (as opposed to 'If I didn't have an intake of plutonium, what is the probability the measurement result would exceed the decision level?'). It properly includes the effect of rarity of true positives in the problem of distinguishing signal from noise. It allows inferences about the values of many parameters from little data (underdetermined problems). However, one of the main disadvantages has been the lack of guidance in the choice of the prior probability distribution, which is always necessary in the Bayesian approach. The Bayesian health physicist is allowed to choose the prior probability distribution subjectively. It seems important, nevertheless, that objective data be used to support educated guesses. The prior probability distribution has a small effect on the inferred result when a large amount of measurement data is available. In the opposite case, which is not unknown in health physics, the prior can influence the inference in an important way.

In this paper we apply some theoretical concepts and make use of historical data from tritium and plutonium internal dosimetry at Los Alamos to arrive at suitable, simple models for the prior probability distribution. We basically propose two models for the prior probability distribution: (1) the log-normal distribution, when there is some additional information to determine the scale of the true result, and (2) the 'alpha' distribution (a simplified variant of the gamma distribution) when there is not. Although we specifically consider urine bioassay measurements in the context of internal dosimetry, these concepts carry over to many other areas of measurement interpretation. A broader discussion of the use of Bayesian methods for internal dosimetry is contained in previous papers in this series[1–5].

At the practical level for internal dosimetry, we have incorporated these new models for the prior probability distribution into version 3 of our Bayesian internal dosimetry code (the Bayes II software package, downloadable from our web site, www.lanl.gov\bayesian). The Bayesian unfolding algorithm is described in detail in Reference 4. In order to carry out a Bayesian analysis of bioassay data using the new models for the prior probability distribution, one needs only to choose the value of a single parameter. When the worker has been involved in an incident or incidents, the prior parameter characterises the additional information on the possible magnitude of the intake (for example, nose swabs or air monitor readings). When no incidents have occurred, the prior parameter reflects the population average of the number of intakes (in a certain range of magnitude) that occur per unit time. From Los Alamos plutonium data in recent years, this number (the parameter α) is about 1 'intake' per 1000 workers per year or even less.

The generic problem of Bayesian interpretation of Gaussian measurements using the alpha prior probability distribution leads to a 'universal curve'. This single curve relates measurement result above background (in standard deviations) to posterior odds of 'positive' divided by the quantity $\alpha\Delta t$ (Δt is the time interval). For example, using this curve the Bayesian health physicist would determine that the odds are 20 to 1 in favour of 'positive' if the measurement is 4.7 standard deviations above background for $\alpha\Delta t = 0.001$. In contrast, the classical prescription of 1.645 standard deviations (for a false positive rate of 0.05)[6] leads to posterior odds of only 0.003 to 1 for an intake — or 300 to 1 against there having been an intake. This example

Contact author E-mail: guthrie@lanl.gov.

shows the importance of the Bayesian method when detecting rare events (α small).

CHOICE OF PRIOR PROBABILITY DISTRIBUTION

In general two cases exist: (1) where there is qualitative information or quantitative information from other measurements giving a non-zero, although perhaps very uncertain, estimate of the true intake amount, and (2) where there is not.

In the first case, the log-normal distribution is appropriate for the prior probability distribution. The log-normal distribution for intake amount ξ is given by

$$P(\xi) = \frac{1}{\sqrt{2\pi}\sigma_{LN}\xi} \exp\left[-\frac{1}{2\sigma_{LN}^2}\left(\ln\frac{\xi}{a}\right)^2\right] \quad (1)$$

When plotted on a log scale (versus ln ξ rather than ξ), the log-normal distribution is Gaussian with maximum and median probability occurring at ln a and standard deviation σ_{LN}. The log-normal distribution can be very broad, for example, with $\sigma_{LN} = 3$ the standard deviation of ln ξ is 3, and when ln ξ varies by 3, ξ varies by a factor of $e^3 \approx 20$. Two standard deviations correspond to a factor of 400. The value of a might be obtained from the other measurements, for example for internal dosimetry based on urine measurements, the additional measurements might be nose swabs, air monitoring data, faecal monitoring data, wound count data, or *in vivo* count data. The value of σ_{LN} would be chosen depending on the relevance and quality of the additional measurements. We normally use the additional measurements to define discrete categories, for example 'true air monitor alarm', or 'wound count greater than 7 Bq' that have the same prior probability distribution.

In the case of no additional measurements, the prior probability distribution is obtained using data from a population similar to that which is being measured. For example, Figure 1 shows tritium urine bioassay data for a worker population over the time period 1998 and 1999. Data are selected for which a preceding data point 14 days before (14 days ± 10%) was 'zero' (below a critical level), so that an elevated measured value corresponds to an intake occurring in the 14 day interval. It is apparent that the distribution is asymmetrical. The portion of the distribution for negative values reflects measurement uncertainty, while the portion of the distribution for positive values indicates the prior probability distribution for intakes occuring in a 14 day time period as well as measurement uncertainty. The observed distribution is in fact the convolution of the measurement uncertainty distribution with the prior probability distribution for the population[2]. The fit shown in Figure 1 will be discussed in what follows.

The data of Figure 1 can be fitted by varying parameters of a function representing the prior probability distribution. Reference 5 discusses fitting this data using a log-normal or Pareto prior with or without a delta-function component. However, we would like a simpler parameterisation of the prior probability distribution (fewer parameters) and one with a more satisfying theoretical justification.

We expect the distribution shown in Figure 1 to depend on the time interval Δt. As Δt increases there would be more time for intakes to occur. Figure 2 shows the same type of distribution as Figure 1 except that the time period before the preceding 'zero'' result was 28 rather than 14 days. The prior probability distribution clearly changes, becoming broader.

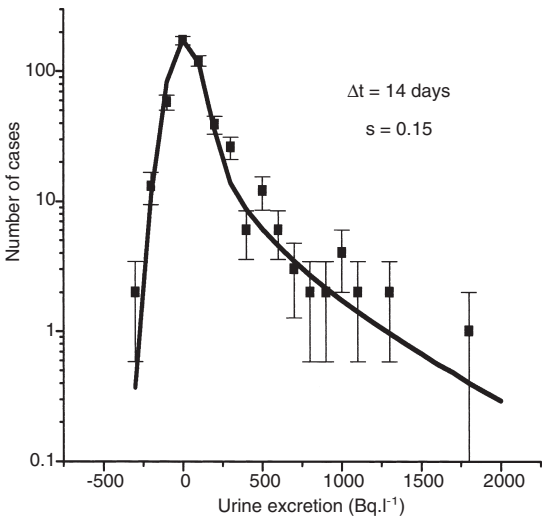

Figure 1. Distribution of tritium bioassay measurement results.

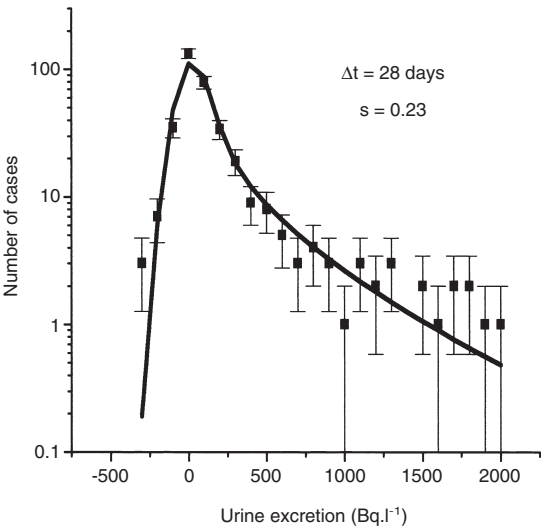

Figure 2. Distribution of tritium bioassay measurement results for a 28 day sampling period.

In Reference 4 we proposed that the prior probability distribution describing intakes occuring in a time interval Δt would have the form

$$P(\xi) = (1 - \lambda\Delta t)\delta(\xi) + \Delta t w(\xi) \tag{2}$$

where $\delta(\xi)$ is the delta function (the delta function, $\delta(\xi)$, is the limit of very narrow distributions peaked at $\xi = 0$ and having unit integral, $\int\delta(\xi)\, d\xi = 1$). In Equation 2, $w(\xi)$ when multiplied by dt is the probability that an intake of amount ξ in infinitesimal interval $d\xi$ occurs in the infinitesimal time interval dt. The integral

$$\int_0^\infty w(\xi)\, d\xi = \lambda \tag{3}$$

when multiplied by dt represents the total probability of an intake in time interval dt.

In this paper we generalise to cases where the normalisation integral given by Equation 3 diverges. We consider intake probability functions of the form

$$w(\xi) = \frac{\alpha}{\xi} \exp\left(-\frac{\xi}{A}\right) \tag{4}$$

where α is a probability per unit time, and the parameter A limits the distribution for large values of ξ. The total probability of an intake in time interval dt is now infinite because very small intakes have a very large probability.

The prior probability distribution we are seeking is the probability distribution of intakes in a finite time interval Δt. It turns out that the probability distribution of intakes in a finite time interval Δt is given by

$$P(\xi) = \frac{\alpha\Delta t}{\xi}\left(\frac{\xi}{A}\right)^{\alpha\Delta t} \frac{\exp\left(-\frac{\xi}{A}\right)}{\Gamma(1 + \alpha\Delta t)} \tag{5}$$

This distribution has normalisation integral unity as must be true for any probability distribution. The distribution given by Equation 5 is the gamma distribution, discussed in many reference books (e.g. see Reference 7). The function $\Gamma(1 + \alpha\Delta t)$ is the gamma function, given by

$$\Gamma(1 + x) \approx 1 - Cx \tag{6}$$

for $x \ll 1$, where C is Euler's constant, $C = 0.577$.

The relation between $w(\xi)$, the probability that an intake of amount ξ occurs in an infinitesimal time interval dt, and $P(\xi)$, the probability that the sum of all intakes in a finite time interval Δt equals ξ, is the kinetic equation ('the Fokker-Plank equation'[8]) of the process

$$\frac{\partial P}{\partial t} = \int_0^\infty w(\xi')[P(\xi - \xi') - P(\xi)]d\xi' \tag{7}$$

which equates the change in P occuring in time interval dt to the difference in probability events entering and leaving the interval $d\xi$. It is shown in Reference 8 that the gamma distribution satisfies Equation 7 with $w(\xi)$ given by Equation 4. When the normalisation integral given by Equation 3 is nondivergent, Equation 2 provides a solution of Equation 7 for Δt small.

The distributions of tritium bioassay measurements shown in Figures 1 and 2 are fit with gamma distributions having $A = 800$ Bq.l^{-1} and variable $s = \alpha\Delta t$. The fits have s increasing with Δt but not quite linearly. There are clearly significant uncertainties because of the small number of events in each histogram bin. Reasons for the non-linear scaling related to sample contamination will be discussed later. The measurement uncertainty distribution is assumed Gaussian with $\sigma = 80$ Bq.l^{-1}.

In order to have only a single parameter we will use the limit of Equation 5 for $A \to \infty$, an improper distribution with parameter α. We refer to this distribution as the 'alpha' distribution. The alpha distribution is truncated for $\xi > A$ and has the form

$$P(\xi) = \frac{\alpha\Delta t}{\xi}\left(\frac{\xi}{A}\right)^{\alpha\Delta t} \tag{8}$$

The alpha distribution with $A \to \infty$ is scale invariant, meaning that its form is the same no matter what the units or scale used. The log-normal and Pareto distributions used in Reference 5 can fit the data of Figures 1 and 2 equally well but they have scale parameters that are small compared with the measurement uncertainty. The assumption underlying the alpha distribution is that no matter what the measurement uncertainty, the distribution will always be decreasing and will never reveal a peak that would define a scale.

The alpha and gamma distributions describe situations where very small intakes are very probable, so that no matter how small the time interval Δt, intakes will have occurred. Nevertheless,

$$P(\xi) \to \delta(\xi)$$

as

$$\Delta t \to 0$$

as can be shown by integrating Equation 8 from $\xi = 0$ to some small positive value and taking the limit $\Delta t \to 0$.

When $\alpha\Delta t$ is small, the alpha distribution is approximated by

$$P(\xi_l < \xi < \xi_u) = \left(\frac{\xi}{A}\right)^{\alpha\Delta t}\Bigg]_{\xi_l}^{\xi_u} \approx \alpha\Delta t \ln\left(\frac{\xi_u}{\xi_l}\right) \tag{9}$$

Using Equation 9, we can identify $\alpha\Delta t$ approximately as the probability of an intake having ξ values that are between some lower value and $e = 2.718 \approx 3$ times that value. Because of scale invariance, this applies also to measured bioassay data values as well as intake amounts. The lower limit should be chosen large enough to eliminate false positives. We will find it convenient later to take these limits as $5\,\sigma$ and $15\,\sigma$, where σ is the measurement uncertainty.

EXAMPLE — PLUTONIUM INTERNAL DOSIMETRY — LOG-NORMAL PRIOR

At Los Alamos, plutonium intakes are divided into two categories: (1) incident-related intakes where there are indications from the workplace that some off-normal occurrence has taken place, and (2) non-incident-related intakes detected only from routine urine bioassay samples. The log-normal distribution is used for the prior probability distribution for the first category of intakes.

Incident-related intakes are further divided into sub-categories depending on the particular workplace indicators involved. These sub-categories have been defined and cover all plutonium work at Los Alamos since 1944. They are as shown in Table 1.

Table 1 shows the number of ^{239}Pu incidents of each type since 1980 and the average values of the intakes determined using the previous (UF2.5) and current (UF3.0) versions of the Bayesian unfolding code. The date 1980 was chosen based on two criteria: (1) the date is within the modern era where the same basic facility was being used, and (2) early enough to include a large amount of data. The current version of the Bayesian unfolding code uses a log-normal prior probability distribution as given by Equation 1 with $\sigma_{LN} = 3$, having a median value determined by the above data. The median of a log-normal distribution is given by $a = \bar{\xi} \exp(-\sigma_{LN}^2/2) = \bar{\xi}/90$, where $\bar{\xi}$ is the average value.

The following iterative process was used. The average intake values for the various incident categories were determined using the previous version of the code, which used a different prior (log-normal plus delta function with subjectively determined parameters). These average values were used to determine the rounded (to the nearest factor of 3) values of the median parameter a as shown in Table 1. These medians defined the log-normal prior probability distribution used in the new code. Using the new code the average intakes were recalculated and found to be consistent with the rounded medians found using the old code, so the log-normal medians were not adjusted further.

The data in Table 1 are for ^{239}Pu (for which most of the plutonium exposures at Los Alamos occurred). The α values obtained in this way were applied to situations involving ^{238}Pu or ^{241}Am using the factors shown in Table 2, which are based on nominal isotope ratios expected to be present.

Table 1. Incident types and average ^{239}Pu intakes since 1980.

Incident type	Number since 1980	Average intake (Bq) (UF2.5)	Average intake (Bq) UF(3.0)	α(Bq)*
Nose count >17 Bq (either side)	27	160	140	3
High nose count	33	100	80	1
High room air count	34	5	3	0.1
Wound count >7 Bq	36	5	4	0.1
Wound with excision	15	30	20	0.3
Unspecified incident type	38	120	80	0.1
Other incident type	50	6	2	0.1

*Median of the log-normal distribution — see text.

EXAMPLE — PLUTONIUM INTERNAL DOSIMETRY — ALPHA PRIOR

For plutonium intakes not related to known incidents, the alpha prior given by Equation 8 is used. The value of the parameter α was determined from historical data since 1980. The raw data are shown in Table 3.

The numbers denoted by $N(y_1 < y_c)$ appearing in Table 3 are the number of bioassay data where there are two measurements y_1 and y_2 separated by Δt such that no intakes have occurred preceding the first measurement (using the Bayesian unfolding code) and the first measured value is less than $y_c = 0.74$ mBq.d^{-1}. The numbers denoted by $N(y_l < y_2 < y_u)$ represent a subset of those cases where an intake has occurred in the monitoring interval (using the Bayesian unfolding code) and the second measured value is in the range $y_l < y_2 < y_u$ where $y_l = 1.85$ mBq $\approx 5\sigma$ and $y_u = 5.6$ mBq $\approx 15\sigma$, where σ is the measurement uncertainty standard deviation.

We use the data shown in Table 3 to determine α

Table 2. Adjustment factors for α values.

Dominant isotope	Factors ^{239}Pu	Factors ^{238}Pu	Factors ^{241}Am
^{239}Pu	1	0.1	0.2
^{238}Pu	0.05	1	0.025
^{241}Am	0.5	0.05	1

Table 3. Non-incident-related intakes.

Nuclide	Δt (years)	$N(y_1 < y_c)$* (UF2.5)	$N(y_1 < y_c)$* (UF3.0)	$N(y_l < y_2 < y_u)$** (UF2.5)	$N(y_l < y_2 < y_u)$** (UF3.0)
^{239}Pu	1	13,347	13,398	23	23
^{239}Pu	2	6915	6927	16	15
^{239}Pu	3	4789	4800	10	10
^{238}Pu	1	14,461	14,459	2	2
^{238}Pu	2	7490	7486	2	2
^{238}Pu	3	5250	5248	1	1

*Number of bioassay data pairs separated by Δt — see text.
**Number of intakes — see text.

using Equation 9. There is a complication in that sometimes high measured values at the end of a monitoring interval are the result of sample contamination rather than an intake occurring in the monitoring interval. Sample contamination most frequently occurs if urine samples are collected in the facility. A speck of dust with alpha activity as little as 0.3 mBq (0.02 dpm) is a significant contamination. Sample contamination has a fixed probability b per sample, thus

$$N = N_0\left[\alpha\Delta t \ln\left(\frac{y_u}{y_l}\right) + b\right] \tag{10}$$

where N_0 the total number of monitoring intervals of length Δt considered, and N is the number of cases where the second bioassay result was in the range $y_l < y < y_u$. Fitting the data in Table 3 using Equation 10, the results shown in Table 4 are obtained.

It is apparent that α for non-incident-related intakes is extraordinarily small, and we have only obtained an upper limit. Figures 3 and 4 show examples of ^{239}Pu bioassay data used in Table 3 for cases that we can fairly clearly identify as real intakes and contamination events, although the Bayesian unfolding code considers them both to be intakes.

Table 4. Values of the quantities α and b for non-incident-related intakes.

Nuclide	$\alpha(y^{-1})$*	$b(y^{-1})$**
^{239}Pu	$2 \times 10^{-4} \pm 3 \times 10^{-4}$	$1.5 \times 10^{-3} \pm 0.6 \times 10^{-3}$
^{238}Pu	$4 \times 10^{-5} \pm 9 \times 10^{-5}$	$1 \times 10^{-4} \pm 2 \times 10^{-4}$

*Probability of intake — see text.
**Probability of sample contamination — see text.

The current version of the Bayesian unfolding code (UF3.0) uses the alpha prior probability distribution with $\alpha = 0.001$ for ^{239}Pu, ^{238}Pu, and ^{241}Am. This value seems like an overestimate, particularly for ^{238}Pu, however, it is felt that the small values of α obtained empirically for ^{238}Pu reflect the small fraction of ^{238}Pu work for plutonium workers at Los Alamos. For those workers primarily working with ^{238}Pu, clearly a larger α is appropriate. The older version of the code (UF2.5) used a different form for the prior probability distribution (log-normal plus delta function with subjectively determined parameter values). As seen in Table 3, the numbers obtained using the current code are essentially unchanged, so the iteration process has converged.

DISCUSSION

We have examined historical urine bioassay data for ^{239}Pu and ^{3}H in order to determine more objectively prior probability distributions for applications in internal dosimetry. In situations where additional quantitative or semiquantitative information exists the log-normal distribution is used, with the median value and standard deviation determined by the additional information. In other situations we use the alpha distribution with the parameter α determined by population data. The parameter α might also be chosen subjectively. The quantity $\alpha\Delta t$ roughly has the meaning of the number of needles expected in this particular haystack.

The interpretation of measurement results using the alpha prior probability distribution is quite simple and natural. Figure 5 summarises the situation. Shown in

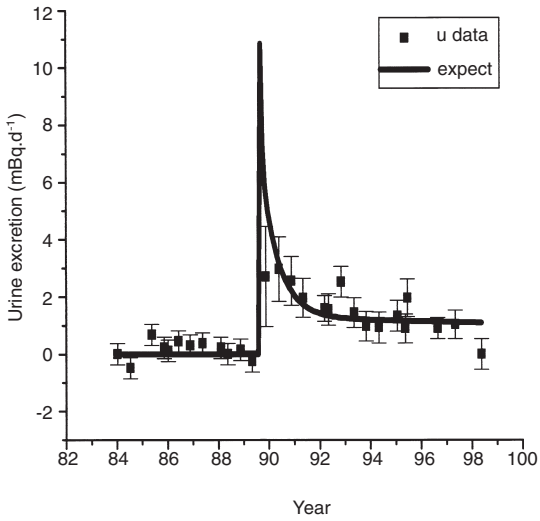

Figure 3. Bioassay data and UF code fit for a 'real' non-incident-related intake.

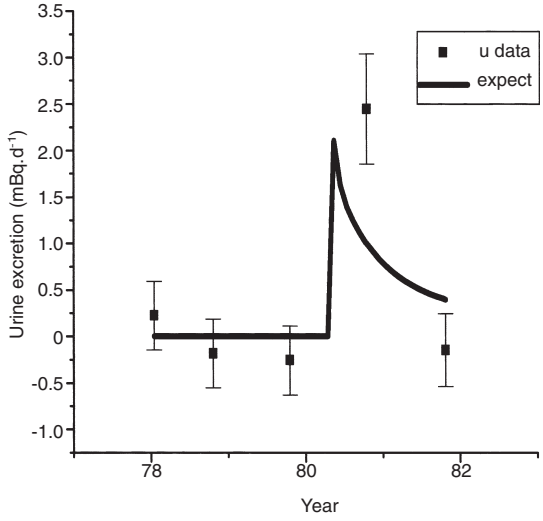

Figure 4. Bioassay data and UF code fit for what seems to be a spurious intake caused by sample contamination.

Figure 5 is the posterior odds ratio in favour of 'positive' relative to 'zero' (true amount greater than or less than $0.1\sigma_0$) normalised by the quantity $\alpha\Delta t$ versus the measurement result normalised by σ_0. The quantity σ_0 is the net measurement uncertainty standard deviation for zero true amount. It turns out that by normalising with $\alpha\Delta t$ a 'universal curve' independent of the value of $\alpha\Delta t$ is obtained for $\alpha\Delta t \ll 0.1$ (Figure 5 is an overplot for $\alpha\Delta t = 0.01$, 0.001, and 0.0001). As an example of the use of Figure 5, assume that we desire at least 10 to 1 odds in favour of 'positive' (posterior probability of 'positive' = 10/11) and want to know what decision level that requires. Assume that $\alpha\Delta t$ is estimated to have the value 0.001. The abscissa in Figure 5 is then $10/0.001 = 10000$. From the plotted curve this corresponds to an ordinate value of about 4.6, which means that the decision level must be $4.6\sigma_0$.

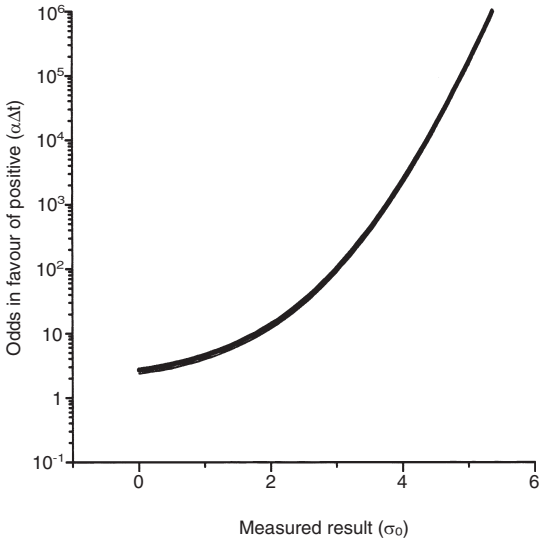

Figure 5. Universal curve of normalised posterior odds in favour of 'positive' as against measurement result.

Figure 5 allows a simple application of Bayesian statistics to the measurement decision process, where α is determined either using population data or subjectively. Mostly, however, one addresses more complex situations or one wishes more detailed information, requiring the use of complex numerical calculations. Computer codes for these calculations are available, for example our Bayesian software packages I and II for Windows 95, 98, and NT, downloadable from our Web site: www.lanl.gov\bayesian.

We use a Gaussian measurement uncertainty model. The measurement uncertainty standard deviation is assumed to be given by

$$\sigma = \sqrt{[\sigma_0^2 + B\psi + (B_{var}\psi)^2]} \tag{11}$$

where ψ is the true result, B enters for measurements based on counting (as discussed in Reference 5) and B_{var} is a multiplicative biological/sample collection protocol variability (typically 0.3 for urine samples with a specific gravity excretion time correction). Another universal feature of the curve plotted in Figure 5 is that it is essentially independent of the quantities B and B_{var}. Non-Gaussian measurement uncertainly can greatly change Figure 5 however. Non-Gaussian effects may be investigated numerically (for example by varying the parameter 'beta' in our Bayesian software package I).

For the purpose of interpreting data an improper prior probability distribution such as the alpha distribution is simple and useful (this distribution is improper because the normalisation integral diverges when $A \rightarrow \infty$). The posterior probability distribution is independent of A for $A \rightarrow \infty$, so this parameter drops out of the problem of data interpretation. However, in order to simulate data or calculate expectation values, a proper prior probability distribution is necessary, which means that a finite value of the parameter A must be chosen.

REFERENCES

1. Miller, G., Inkret, W. C. and Martz, H. F. *Bayesian Detection Analysis for Radiation Exposure*. Radiat. Prot. Dosim. **48**, 251–256 (1993).
2. Miller, G., Inkret, W. C. and Martz, H. F. *Bayesian Detection Analysis for Radiation Exposure, II*. Radiat. Prot. Dosim. **58**, 115–125 (1995).
3. Miller, G. and Inkret, W. C. *Bayesian Maximum Posterior Probability Method For Interpreting Plutonium Urinalysis Data*. Radiat. Prot. Dosim. **63**, 189–196 (1996).
4. Miller, G., Inkret, W. C. and Martz, H. F. *Internal Dosimetry Intake Estimation Using Bayesian Methods*. Radiat. Prot. Dosim. **82**, 5–17 (1999).
5. Miller, G., Inkret, W. C., Schillaci, M. E., Martz, H. F. and Little, T. T. *Analyzing Bioassay Data Using Bayesian Methods — A Primer*. Health Phys. **78**(6), 598–613 (2000).
6. Currie, L. A. *Limits for Qualitative Detection and Quantitative Determination, Application to Radiochemistry*. Anal. Chem. **40**, 586 (1968).
7. Rothschild, V. and Logothetis, N. *Probability Distributions* (New York: John Wiley and Sons) (1986).
8. Feller, W. *An Introduction to Probability Theory and It's Applications*. Vol. II (New York: John Wiley) (1966).