



HAL
open science

Noisy radioactivity data analysis using parametric Poisson models

Guillaume Manificat, Salima Helali, Kevin Galliez, Miriam Basso, Maxime Morin, Raja Ben Hajria

► **To cite this version:**

Guillaume Manificat, Salima Helali, Kevin Galliez, Miriam Basso, Maxime Morin, et al.. Noisy radioactivity data analysis using parametric Poisson models. 2024. hal-04550550v2

HAL Id: hal-04550550

<https://hal.science/hal-04550550v2>

Preprint submitted on 14 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Noisy radioactivity data analysis using parametric Poisson models

Guillaume MANIFICAT^{1*}; Salima HELALI^{1†}; Kévin GALLIEZ, Miriam BASSO,
Maxime MORIN, Raja BEN HAJRIA

¹ Institute for Radiation Protection and Nuclear Safety (IRSN) Paris, France

Abstract

In metrology, the use of decision threshold and detection limit concepts often creates many problems for metrologists in radioactivity analysis laboratories. It is usual to censor data when it becomes difficult to discern the presence or absence of the activity, due to noise in the measurement data. This implies that if the measurement results are non significant which are below critical value of the test statistic which is called the decision threshold (DT) in metrology. The analysis simply states that the true value (signal) of the radioactivity is below a certain limit called the detection limit (DL). These problems are frequently related to the incorrect understanding of the DT formulas or the wrong choice between several formulas whose numerical results are significantly different. Moreover, it is often unclear how to generate an appropriate and justified DT. In the current research paper, we elaborate a statistical method of DT determination, capable of providing DT with a high statistical power, using a smaller number of repeated measurements. The method is then applied to a real test case. Next, statistical approaches methods are adopted to estimate the density, the expectation and the variance of the radioactivity. Some of its asymptotic properties are also discussed. Efficiency and feasibility of these approaches are corroborated through applications on simulated real data sets.

Key Words: Decision threshold, Likelihood ratio test, Radioactivity, Noise data, Nuisance parameters.

1 Introduction

The central aim of radiological environmental monitoring is to contribute to the knowledge of the radiological state of the environment as well as spatial and temporal evolution, in order to reduce radiation exposure from contaminated soil, waste storage facilities or other contaminated infrastructure, groundwater or surface water. Radioactivity measurement rests on characteristic limits (decision thresholds and detection limits). The decision thresholds is the name given in physics to the critical value of the hypothesis test (Currie (1968); Lehmann et al (2005)). The detection limit is the lowest possible value of the parameter that will on average be declared significant. When an observation is below the decision threshold, the estimation is usually left censored and this observation is considered as useless. Due to ever-lower levels of environmental activity, the number of radiological analyses for which metrologists are unable to provide significant results is increasing. Current standards steadily propose uniquely DT formulas without much justification

*e-mail adress: guillaume.manificat@irsn.fr

†e-mail adress: helali.salima@gmail.com

and without taking profit of the statistical literature. For instance, the DT, as defined by the standard [ISO 11929 \(2020\)](#), is calculated as

$$y^* = k_{1-\alpha} w u(n_n = 0),$$

where $k_{1-\alpha}$ denotes the quantile of the probability density of the measurement result y , for a null parameter that exceeds the DT with the probability α , w refers to the value of a conversion factor and $u(n_n = 0)$ stands for the null measurement uncertainty of the net indication n_n . It is unclear in the standard [ISO 11929 \(2020\)](#) how the DT is calculated and under which statistical test it is obtained. Currently, there is no clear method for the determination of DT of the radioactivity in literature. There is equally an imperious need in metrology for a clear method to determine an optimal DT due to ever-lower levels of environmental activity. The Institute for Radiological Protection and Nuclear Safety (IRSN) is thus highly interested in improving DT for radioactivity analyses. The basic aim of this contribution is to present an efficient statistical method of DT determination, capable of providing an optimal DT in a particular measurement of the radioactivity, using a smaller number of repeated measurements. The perfect method of DT determination would involve an infinite number of real-world measurements, but such a method is in practice impossible, and some realistic methods need to be used so as to approximate the infinite-measurement cases. As such, large numbers of repeated measurements will increase the statistical confidence in the determined DT, but there will always be a trade-off between the number of measurements and the time allocated to acquire the data, with excessive repetitions of measurements being unnecessarily expensive and slow. Therefore, determining the DT with minimal repeated measurements is desirable and requires significant statistical analysis.

The paper is organized as follows. In the next section, we recall the main concepts and definitions considered throughout this work. In [Section 3](#), we identify the model and the notations. In [section 4](#), we display the optimal threshold of the radioactivity signal. In [section 5](#), we propose an estimation of the radioactivity signal. In [section 6](#) we present a non parametric estimator of the radioactivity and parametric estimators of its expectation and variance. The asymptotic normality properties of these estimators and their associated confidence intervals are determined. [Sections 7 and 8](#) are devoted to the application results obtained through simulations and real data analysis. Several pertinent concluding remarks and potential future perspectives are provided in [Section 9](#). [Section 10](#) exhibits the proofs of the theoretical results.

2 Background

2.1 *Statistical test, decision threshold and confidence interval:*

A statistical test corresponds to a procedure for deciding whether a hypothesis about a quantitative feature of a sample is true or false. The decision threshold of the hypothesis test represents the fixed value of the radioactivity such that, when the measurement result is greater than it, it is decided that the radioactivity is present. Grounded on [Lovric \(2011\)](#), a $100(1 - \alpha)\%$ confidence interval is an interval estimate around a population parameter ϑ that, under repeated random samples of size N , is expected to include ϑ 's true value $100(1 - \alpha)\%$ of time. The detection limit is the largest true value that would have a non-negligible probability of being considered insignificant by the measurement. It can therefore give rise to observations equal to the decision threshold.

2.2 *Distribution estimation:*

A set of repeated measurement results represents a sample of the radioactivity that contains all possible measurement results, with their distribution being a radioactivity distribution. A sum-

mary statistics of the sample (e.g. mean, maximum or variance) stands for an estimator of the radioactivity summary statistics. The distribution of the values of estimators calculated on repeated samples is a sampling distribution. Due to the central limit theorem (JCGM 100 (2008)), it is possible for the sampling distribution to be considered normal even if the radioactivity distribution is not. When using an estimator, there are three main properties that describe their behaviour: bias, consistency and uncertainty (see Dodge (2003)). Bias corresponds to the difference between an estimator's expected value and the true value of the parameter being estimated. Consistency is the tendency, as the number of sampled data points increases, for an estimator to converge to the true value of the parameter being estimated. A stated measurement uncertainty is a non-negative parameter characterising the dispersion of the quantity values being attributed to a measurement, based on the used information (see De Bièvre (2012)).

2.3 Re-sampling and Bootstrapping:

When attempting to improve the confidence in a measurement, taking more measurements is the obvious first step. However, there are practical and economic limits on the possible number of repeated measurements. Therefore, when data have been collected, and it would still be desirable to reduce the error on the estimator, re-sampling the data can be a solution. Re-sampling is a method of statistical analysis that uses a fixed number of measurements in order to simulate what would be expected to happen, if more measurements had been taken. Bootstrapping is a common re-sampling technique that has been applied to settle metrology problems in literature (for e.g. see Hiller et al (2014) and Ciarlini and Regoliosi (2000)).

2.4 Uncertainty

Consider a measurement system used to quantify the activity contained in various samples. The counting procedure is as follows: two measurements of two samples are taken: a 'blank' measurement (no radioactivity) and a 'gross' (possibility of radioactivity). Departing from the value of the net count: $\mathbf{net} = \mathbf{gross} - \mathbf{blank}$, we calculate the measured activity by: $\mathbf{A} = \mathbf{net}/c$, where c is a calibration factor. As for any measurement, it is necessary to estimate the uncertainty of the activity based on the measured activity A . If we assume that nuclear counting has a Poisson behaviour, we can write the uncertainties of *gross* and *blank* as $\mu_{gross} = \sqrt{gross}$ and $\mu_{blank} = \sqrt{blank}$, respectively. If **gross** and **blank** are two independent variables, then the uncertainty of the net count is provided by $\mu_{net} = \sqrt{\mu_{blank}^2 + \mu_{gross}^2}$. If we neglect the uncertainty of the calibration factor c , then the final uncertainty on the measured activity is equal to: $\mu_A = \mu_{net}/c$. The relative uncertainty on the activity is equal to $\frac{\mu_A}{A}$.

3 Model and notations

To report the method devised for DT determination and the proposed estimators of radioactivity, we need first to clarify a number of notations and assumptions considered throughout this work. These concepts are identified in this section. For each given point data of radioactivity A , two measurements of equal duration are taken; namely a blank measurement C_{blank} and a gross measurement C_{gross} , where C_{blank} is as similar as possible to C_{gross} but ensures the absence of the signal S . The Noisy radioactivity data are defined as follows

$$A = Y (C_{gross} - C_{blank}),$$

where

- $C_{blank} := C_{bl}$: is a random variable that models the noise of the counter: $C_{bl} \sim \mathcal{P}(\lambda)$.
- $C_{gross} := C_{gr} = S + B'$: is a random variable that is equal to the sum of two independent random variables, measured by counter, with S being the random variable of interest that models the signal and B' being a random variable that models the noise of the measurement of C_{gr} . $S \sim \mathcal{P}(\theta)$, $B' \sim \mathcal{P}(\lambda')$, then $C_{gr} \sim \mathcal{P}(\mu := \theta + \lambda')$.
- Y : is a Gaussian random variable that models the calibration factor (measurement error).

Throughout this paper, we will equally use the following notations:

- $X := C_{gr} - C_{bl}$: is the net count.
- DT : is the decision threshold based on the hypothesis test of θ : $H_0 : \theta = 0$ (no signal) / $H_1 : \theta > 0$.
- DL : is the detection limit.
- $R := X\mathbb{1}_{\{X \geq DT\}} + DL\mathbb{1}_{\{X < DT\}}$.
- $D := YR$: models the radioactivity used by IRSN which has been so far based on [ISO 11929 \(2020\)](#). There is a loss of information because non significant results have been left-censored.
- n : is the size of data.
- ϕ : is the Gaussian distribution of $\mathcal{N}(0, 1)$.

Let (A_1, \dots, A_n) and (D_1, \dots, D_n) be two samples of i.i.d. of random variables defined as follows:

$$(A_i = Y_i X_i = Y_i (C_{gr,i} - C_{bl,i}))_{1 \leq i \leq n},$$

vs

$$(D_i = Y_i R_i = Y_i (X_i \mathbb{1}_{\{X_i > DT_i\}} + DL_i \mathbb{1}_{\{X_i \leq DT_i\}}))_{1 \leq i \leq n},$$

where $X_i = \frac{1}{N_i} \sum_{k=1}^{N_i} X_i^k$, DT_i , DL_i are constructed by $X_i^1, \dots, X_i^{N_i}$ and N_i is the number of repetition of the measurements of the observed point data X_i . Noting that Y_i and $C_{bl,i}$ are constants for some i , all the parameters θ , λ , λ' , μ are unknown and θ is the interest parameter. Throughout this research work, the following assumption will be considered:

$$\mathcal{A} : \lambda = \lambda'.$$

This assumption is usually justified by the fact that metrologists put a great emphasis on ensuring that the measurements are undertaken in very similar physical conditions (temperature, background radioactivity, apparatus, etc).

4 A method for the determination of a DT of θ

The main target of this section is to provide an optimal DT of the parameter θ for each point data X_i for $i = 1 \dots n$ based on the likelihood ratio test, in order to make a decision about the presence of the signal. Let us consider the following data for $i = 1 \dots n$:

- Let $C_{bl,i}^1, \dots, C_{bl,i}^{N_i}$ be an N_i i.i.d. sample drawn from the random variable $C_{bl,i} \sim \mathcal{P}(\lambda_i)$.
- Let $C_{gr,i}^1, \dots, C_{gr,i}^{N_i}$ be an N_i i.i.d. sample drawn from the random variable $C_{gr,i} \sim \mathcal{P}(\mu_i = \theta_i + \lambda_i)$.

It is assumed that $N_i \leq 5$ is insufficient to the asymptotic normality of an empirical estimator of θ . In this section, in order to check the presence of the signal, we need to test the following pair of hypotheses

$$H_0 : \theta = 0 \text{ vs } H_1 : \theta > 0,$$

with a fixed risk α . The following proposition provides the optimal DT_i for $i = 1 \dots n$. For a point data X_i , a DT_i is constructed based on the repeated measurements $(C_{bl,i}^j, C_{gr,i}^j)_{1 \leq j \leq N_i}$.

Proposition 1 (DT of θ). Under the hypothesis test $H_0 - H_1$ and the assumption \mathcal{A} , an optimal DT of the parameter θ for $i = 1 \dots n$, denoted by DT_i , is generated by the resolution of the following equation:

$$\alpha = I_{1/2}(DT_i + 1, \overline{C_{bl,i}} + 1),$$

where I is the regularized beta incomplete function. An approximation of DT_i is determined by

$$DT_i = \frac{k_{1-\alpha}^2 + \sqrt{k_{1-\alpha}^4 + 8k_{1-\alpha}^2 \overline{C_{bl,i}}}}{2},$$

where $k_{1-\alpha} = \phi^{-1}(1 - \alpha)$.

The proof of Proposition 1 relies upon the likelihood ratio test devoted to detect the presence of the signal in context, where the assumption \mathcal{A} is held. More details of this proof are supplied in section 10. The following proposition specifies a decision threshold DT_i for $i = 1 \dots n$ in the case where C_{bl} is large. We consider that C_{bl} is large when $C_{bl} > 10$.

Proposition 2 (DT when C_{bl} is large). Consider that C_{bl} is large. Under the hypothesis test $H_0 - H_1$ and the assumption \mathcal{A} , an optimal DT of the parameter θ for $i = 1 \dots n$, is obtained by the resolution of the following equation:

$$\alpha = 1 - \frac{\Gamma(\lfloor DT_i + 1 \rfloor, 2\overline{C_{bl,i}})}{\lfloor DT_i \rfloor},$$

where $\Gamma(., .)$ is the upper incomplete gamma function and $\lfloor . \rfloor$ is the floor function.

Remark 1. A confidence interval of the signal S is of the form: $\left[x_i - DT_i \frac{\sqrt{X_i}}{N_i}, x_i + DT_i \frac{\sqrt{X_i}}{N_i} \right]$.

5 Estimation of θ based on the count term X

Let us consider the following data

- $C_{gr,1} \dots, C_{gr,n}$ n iid sample drawn from the random variable $C_{gr} \sim \mathcal{P}(\mu = \lambda + \theta)$.
- $C_{bl,1} \dots, C_{bl,n}$ n iid sample drawn from the random variable $C_{bl} \sim \mathcal{P}(\lambda)$.

Proposition 3 (Estimator of θ). If C_{bl} and C_{gr} are large, an estimation of the parameter θ is expressed in terms of

$$\theta_n = \frac{\sum_{i=1}^n C_{gr,i} - C_{bl,i}}{n}.$$

Based on the central limit theorem, the estimator θ_n is asymptotically unbiased with a variance going to zero when $n \rightarrow \infty$.

6 Estimation of the radioactivity

Consider that $A = XY$, where $Y \sim \mathcal{N}(a, b^2)$. Noting that, in the count term $X = C_{gr} - C_{bl}$, we are basically interested in the law of $C_{gr} = c_{gr} | \{C_{bl} = c_{bl}\}$. Since X and Y are independent, then

$$\mathbb{E}(A) = \mathbb{E}(X)\mathbb{E}(Y) = \theta a.$$

$$\text{VAR}(A) = \text{VAR}(X)\text{VAR}(Y) + \text{VAR}(X)\mathbb{E}(Y)^2 + \text{VAR}(Y)\mathbb{E}(X)^2 = b^2(\theta + 2\lambda) + a^2(\theta + 2\lambda) + b^2\theta^2.$$

The following proposition provides estimators of the density of A ; f_A , $\mathbb{E}(A)$ and $\text{VAR}(A)$ based on the iid data $C_{gr,1}, \dots, C_{gr,n}$ and $C_{bl,1}, \dots, C_{bl,n}$.

Non parametric estimation of f_A : Let z be the observation of $A = XY$. We can consider the kernel non parametric estimator of the density $f_A(z)$ defined by [Rosenblatt \(1956\)](#) as follows:

$$\hat{f}_{A,n}(z) = \frac{1}{nh_n} \sum_{p=1}^n K\left(\frac{z - Z_p}{h_n}\right), \forall z \in \mathbb{R},$$

where (h_n) is a sequence of positive real numbers that goes to zero dependent, which is on the size n (smoothing parameter) called the bandwidth and K is the kernel function.

Proposition 4 (Estimation of the radioactivity). The considered estimators of the expectation $\mathbb{E}(A) := \mu_A$ and the variance $\text{VAR}(A) := \sigma_A^2$ are indicated by $\mu_{A,n}$ and $\sigma_{A,n}^2$ as follows

$$\mu_{A,n} = \bar{X} \times \bar{Y},$$

$$\text{and } \sigma_{A,n}^2 = \bar{Y}^2 \sigma_{X,n}^2 + \bar{X}^2 \sigma_{Y,n}^2,$$

$$\text{where } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \sigma_{X,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ and } \sigma_{Y,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

These approaches are asymptotically unbiased with a variance that goes to zero when $n \rightarrow \infty$. The following proposition demonstrates the asymptotic normality of the previous estimators. It is worth noting here that the mean estimator $\mu_{A,n}$ is a function of the mean estimator of $\mathbb{E}(X)$ and the one of $\mathbb{E}(Y)$. Likewise, the uncertainty estimator $\sigma_{A,n}^2$ is a function of the uncertainty estimator of σ_X^2 and the one of σ_Y^2 , which is similar to the uncertainty approach in [BIPM et al \(2008\)](#) in Equation (10). The following proposition provides the asymptotic normality of the estimators in Proposition 4 and the associated confidence intervals. A confidence interval is constructed based on the repeated measurements $(X_i, Y_i)_{1 \leq i \leq n}$.

Proposition 5 (Asymptotic normality and confidences intervals of the radioactivity): The asymptotic normality of the estimators $f_{A,n}(z)$ and $\mu_{A,n}$ are expressed by

$$\sqrt{nh_n}(\hat{f}_{A,n}(z) - f_A(z)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, f_A(z) \int K(u)du\right) \text{ and } \sqrt{n}(\mu_{A,n} - \mu_A) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \mu_X^2 \sigma_X^2 + \mu_Y^2 \sigma_Y^2$. Then the confidence intervals of $f_A(z)$ and μ_A are provided by

$$I_{f_A} = \left(\hat{f}_{A,n}(z) \pm \frac{z_{1-\frac{\delta}{2}} \sigma_{f,n}(z)}{\sqrt{n}} \right) \text{ and } I_{\mu_A} = \left(\mu_{A,n} \pm \frac{z_{1-\frac{\delta}{2}} \sigma_n}{\sqrt{n}} \right),$$

where $z_{1-\frac{\delta}{2}}$ is the normal $(1 - \frac{\delta}{2})$ quantile, $\sigma_{f,n}(x) = \sqrt{f_{n,A}(z) \int K(u)du}$ and $\sigma_n = \sqrt{\mu_X^2 \sigma_{Y,n}^2 + \mu_Y^2 \sigma_{X,n}^2}$.

7 Simulations

In this section, we investigate the performance of the theoretical results. In the simulation study, three sample sizes are considered, $n = 50$, $n = 100$, and $n = 200$, as well as the following models:

- a) $S \sim \mathcal{P}(n, 2)$, $B \sim \mathcal{P}(n, 50)$, $C_{blank} \sim \mathcal{P}(n, 50)$ and $N = 5$.
- b) $S \sim \mathcal{P}(n, 15)$, $B \sim \mathcal{P}(n, 10)$, $C_{blank} \sim \mathcal{P}(n, 10)$ and $N = 5$.
- c) $S \sim \mathcal{P}(n, 15)$, $B \sim \mathcal{P}(n, 9)$, $C_{blank} \sim \mathcal{P}(n, 10)$ and $N = 5$.

7.1 Comparison between the estimations of θ using X and R

In this subsection of the simulation study, three proposed estimators of θ are considered, with reference to Proposition 3. They are defined as follows:

- Based on the sample X_1, \dots, X_n , we consider $\tilde{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$.
- Based on the sample $R_{1,irs n}, \dots, R_{n,irs n}$, we consider $\bar{\theta}_n = \frac{1}{n} \sum_{i=1}^n R_{i,irs n}$, where for $i = 1 \dots n$,

$$R_{i,irs n} = X_i \mathbf{1}_{\{X_i \geq DT_{i,irs n}\}} + DL_{i,irs n} \mathbf{1}_{\{X_i < DT_{i,irs n}\}},$$

$$DT_{i,irs n} = k_{1-\alpha} \sqrt{2C_{bl,i}}, \text{ and } DL_{i,irs n} = DT_{i,irs n} + 0.5(k_{1-\alpha})^2 \left[1 + \sqrt{1 + \frac{4DT_{i,irs n}^2}{k_{1-\alpha}^4}} \right].$$

- Based on the sample $R_{1,opt}, \dots, R_{n,opt}$, we consider $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n R_{i,opt}$, where for $i = 1 \dots n$,

$$R_{i,opt} = X_i \mathbf{1}_{\{X_i \geq DT_{i,opt}\}} + DL_{i,opt} \mathbf{1}_{\{X_i < DT_{i,opt}\}},$$

$$\alpha = 1 - \frac{\Gamma(\lfloor DT_{i,opt} \rfloor + 1, 2C_{bl,i})}{\lfloor DT_{i,opt} \rfloor}, \text{ and } DL_{i,opt} = DT_{i,opt} + 0.5(k_{1-\alpha})^2 \left[1 + \sqrt{1 + \frac{4DT_{i,opt}^2}{k_{1-\alpha}^4}} \right].$$

For each model of *a)*, *b)* and *c)*, we compute the Integrated Squared Error (ISE) and the Integrated Absolute Error (IAE) of $\tilde{\theta}_n$, $\bar{\theta}_n$ and $\hat{\theta}_n$:

$$ISE(\theta_n) = (\theta_n - \theta)^2 \text{ and } IAE(\theta_n) = |\theta_n - \theta|,$$

where $\theta_n \in \{\tilde{\theta}_n, \bar{\theta}_n, \hat{\theta}_n\}$.

		n			
		Model	50	100	1000
a)	ISE(θ_n)	X	0.0196	0.0067	0.0002
		R_{irs_n}	55.587	0.1799	1.1104
	R_{opt}	0.2979	0.0599	0.1015	
IAE(θ_n)	X	0.1400	0.0820	0.0148	
	R_{irs_n}	7.4556	0.4242	1.0537	
	R_{opt}	0.5458	0.2449	0.3186	
b)	ISE(θ_n)	X	0.0023	0.0032	0.0014
		R_{irs_n}	0.04812	0.0055	0.0082
	R_{opt}	0.0023	0.0032	0.0014	
IAE(θ_n)	X	0.0485	0.0566	0.0378	
	R_{irs_n}	0.2193	0.0744	0.0910	
	R_{opt}	0.0485	0.0566	0.0378	
c)	ISE(θ_n)	X	1.7742	2.0563	0.8273
		R_{irs_n}	1.7742	2.0563	0.9367
	R_{opt}	1.7742	2.0563	0.8273	
IAE(θ_n)	X	1.3320	1.4340	0.9096	
	R_{irs_n}	1.3320	1.4340	0.9678	
	R_{opt}	1.3320	1.4340	0.9096	

Table 1: Results for ISE and IAE for $N = 5$ trials of the proposed estimators of θ ($\tilde{\theta}_n, \overline{\theta}_n, \hat{\theta}_n$) for $n = 50, n = 100, n = 1000$.

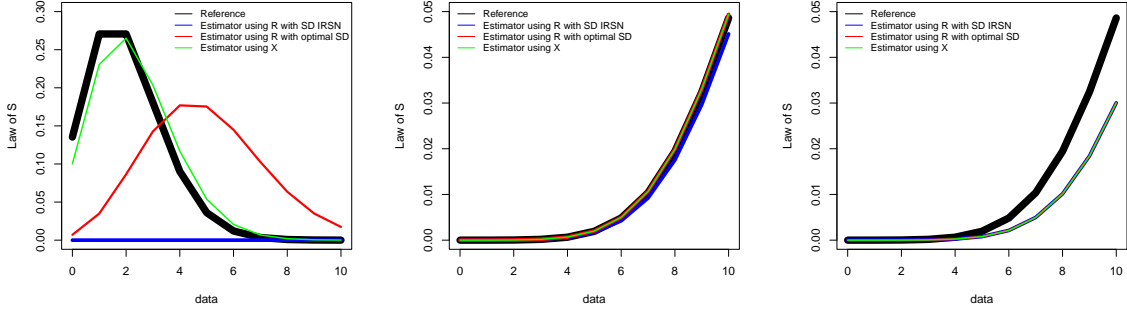


Figure 1: Results for the Signal distribution estimators of $\mathcal{P}(\theta)$ using X, R_{irs_n} and R_{opt} for $n = 50, N = 5$ for high values of C_{bl} and C_{gr} for model a) (left) b) (center) and c) (right).

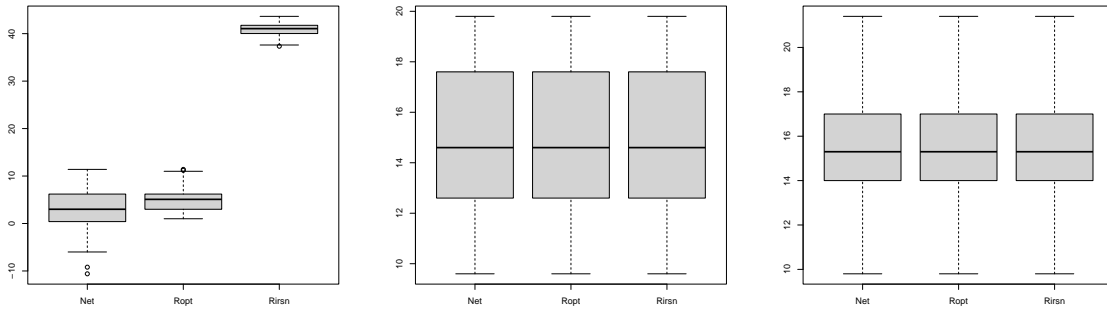


Figure 2: Results for the boxplots using X , R_{irs_n} and R_{opt} for $n = 50$, $N = 5$ for high values of C_{bl} and C_{gr} for model a) (left) b) (center) and c) (right).

Based on Table 1 and Figures 1 and 2, we infer that:

- For all cases, $ISE(\theta_n)$ and $IAE(\theta_n)$ based on the model X and R_{opt} are smaller than those obtained by the model R_{irs_n} .
- The ISE and the IAE decrease as the sample size increases.
- The Poisson law estimators $\mathcal{P}(\tilde{\theta}_n)$ and $\mathcal{P}(\hat{\theta}_n)$ based on the models X and R_{opt} respectively, can yield better results compared to the Poisson law estimator $\mathcal{P}(\bar{\theta}_n)$ grounded on the R_{irs_n} model.
- The average of R_{opt} is closer to the average of X compared to that of R_{irs_n} .

In addition, the performed simulations reveal that the proposed DT defined in Proposition 1 can entail satisfactory estimates of the signal θ , better than the DT defined in ISO 11929 (2020).

7.2 Performance of the Kernel density estimator

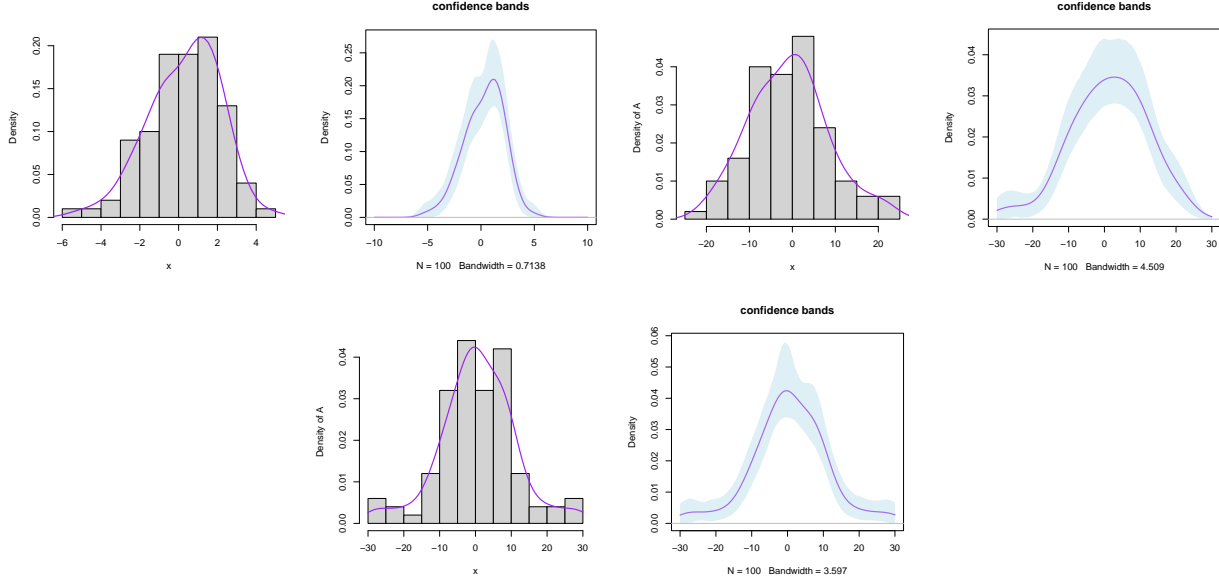


Figure 3: Results of the non parametric kernel density estimator of the f_A density and its confidence interval for $n = 50$, $N = 100$ for high values of C_{bl} for model a) (left up) b) (right up) and c) (down).

	n		
	50	100	1000
a) $MSE(f_{A,n}(z))$	0.003040	0.001463	0.000131
b) $MSE(f_{A,n}(z))$	0.000561	0.000287	$2.237e^{-05}$
c) $MSE(f_{A,n}(z))$	0.000277	0.000230	$2.415e^{-05}$

Table 2: Results for ISE and IAE for $N = 5$ trials of the kernel density estimator of $f_A(z)$ for $n = 50$, $n = 100$, $n = 1000$.

The density curves in Figure 3 are closer to the histogram of data. This figure shows that kernel estimator is a good approach of radioactivity density. Based on Table 7.2, we deduce that MSE of $f_{A,n}(z)$ are closer to zero and decrease as the sample size increases.

8 Real data analysis

The monitoring of the environment is carried out by measuring the main radionuclides released into water, air and soil. In our research work, in order to illustrate the proposed theoretical results, we used a data set of measurements of radionuclides released into water. In particular, we considered the data set consisting of $n = 360$ observations within the period 2019 – 2022 of the following variables

- S : the measurement of tritium in water by Bq/L.
- C_{bl} : pure water (deep water)

- $C_{gr} = S + B'$, where $B' \sim C_{bl}$.
- Y : calibration factor.

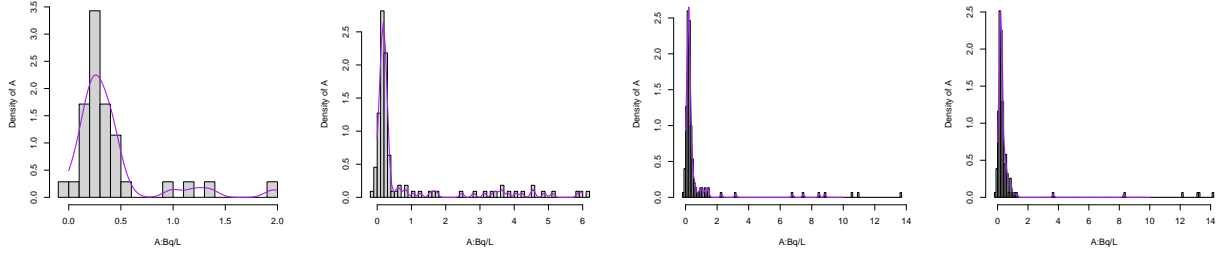


Figure 4: Results for the non parametric kernel density estimator of the density of f_A for the radioactivity data within 2019-2022 (left \rightarrow right)

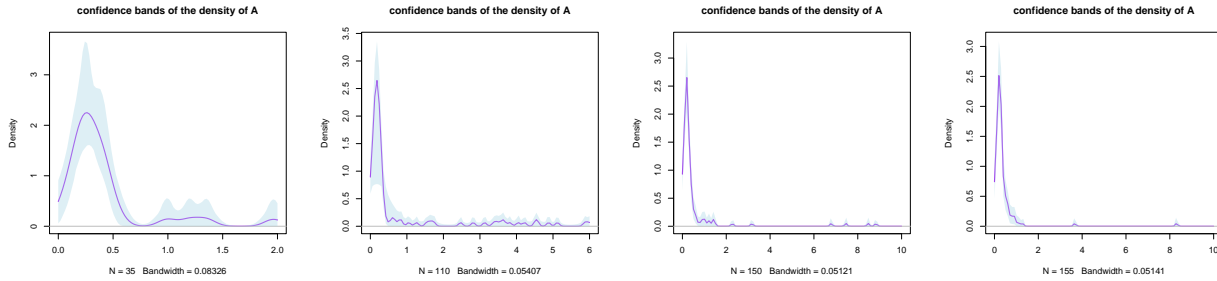


Figure 5: Results for the non parametric confidence bands of the density f_A , based on the kernel density estimator for the radioactivity data within 2019-2022 (left \rightarrow right)

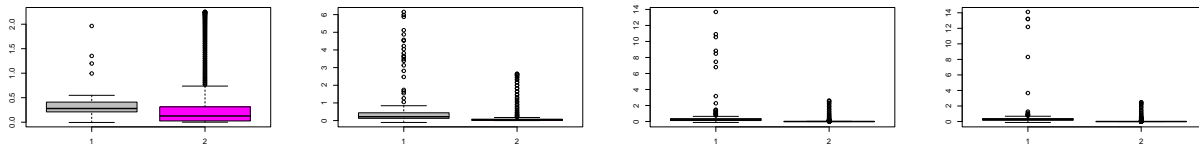


Figure 6: Results of boxplot for the radioactivity data within 2019-2022 (left \rightarrow right)

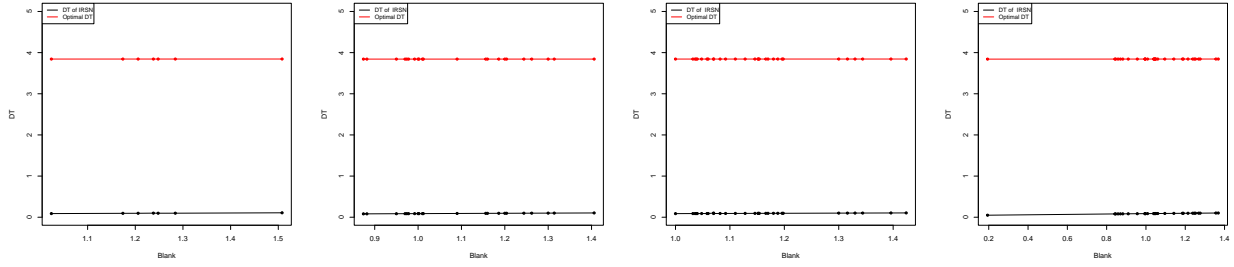


Figure 7: Results for DT of the count part for the radioactivity data within 2019-2022 (left \rightarrow right)

	2019	2020	2021	2022
$\text{ISE}(\mu_{A,n})$	0.021049	0.121666	0.092424	0.105236
$\text{IAE}(\mu_{A,n})$	0.145083	0.348807	0.304014	0.324401

Table 3: Results for ISE and IAE of the proposed estimator of $\mathbb{E}(A)$ (Taylor method) for the radioactivity data (reference: \bar{A}).

	2019	2020	2021	2022
$\text{ISE}(\hat{f}_{A,n}(z))$	0.000910	0.001543	0.000146	$7.175e^{-06}$
$\text{IAE}(\hat{f}_{A,n}(z))$	0.026128	0.034026	0.010487	0.002319

Table 4: Results for ISE and IAE of the kernel estimator $\hat{f}_{A,n}(z)$ for the radioactivity data (reference: histogram).

	Result of ISO	Proposed method
Activity	6.08	6.08
DT	0.735	3.84
CI	$[6.08 \pm 0.94] = [5.13, 7.02]$	$[6.08 \pm 1.89] = [4.18, 7.97]$
Conclusion	$A > DT$	$A > DT$

Table 5: Example 1

	Result of ISO	Proposed method
Activity	1.08	1.08
DT	0.737	3.32
CI	$[1.08 \pm 0.77] = [0.30, 1.85]$	$[1.08 \pm 1.22] = [-0.14, 2.30]$
Conclusion	$A > DT$	$A < DT$

Table 6: Example 2

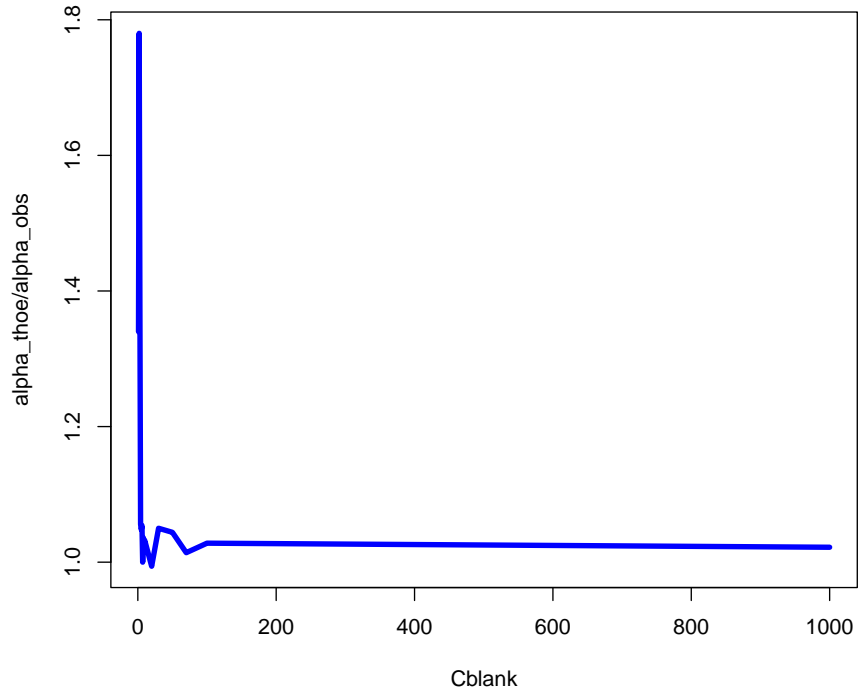


Figure 8: Comparison of the ratio of real/theoretical false positives..

Figure 4 displays the results of density kernel estimators within 2019 – 2022. The density curves in Figure 4 are closer to the histogram of data. This figure shows that kernel estimator is a good approach of radioactivity density. Grounded on Table 4, we deduce that ISE and IAE of $f_{A,n}(z)$ are closer to zero and decrease as the sample size increases. Hence, it can be concluded that the kernel estimators are robust and perform well in terms of estimating the density of the radioactivity, as it is theoretically expected. Table 3 shows that ISE and IAE of the proposed estimator of $\mathbb{E}(A)$ are closer to zero and decrease as the sample size increases. Figure 7 demonstrates that the proposed DT is higher than the DT defined in ISO 11929 (2020), which implies that certain radioactivity values are declared significant with the standard when in fact they are not (see for instance Tables 5 and 6). With reference to Tables 5 and 6, we infer that the confidence interval of ISO 11929 (2020) converge to the proposed confidence interval which is based on the proposed DT. In figure 8, we tested the DT determined in Proposition 1. We determined the ratio of the false positive rate between the observed risk $\alpha_{practice}$ and the theoretical risk test $\alpha_{theoretical} = 0.05$ for a given value of parameter λ . We have made 1000 draws of the random variable Y . For each draw, we determine the DT, then we make a new draw which we compare to the previously DT calculated. Figure 8 shows that ratio of real/theoretical false positives $\alpha_{thoeretical}/\alpha_{practice}$ is closer to 1 especially when C_{blank} is large, which implies the Decision threshold is optimal, as it is theoretically expected.

9 Conclusion

In this research paper, we proposed an optimal decision threshold based on Neyman-Pearson Lemma in order to detect the signal of the radioactivity. Several theoretical results related to the model of the radioactivity; in particular to the density estimator and the confidence interval of the activity signal, are presented. The application of our method was illustrated with the study on real datasets of tritium in water. In this respect, we would assert that this synthesis can be regarded as a preliminary study for further investigations on the decision threshold of the activity signal. Indeed, our work provides a theoretical foundation of the optimal decision theshold in the case of the alpha/beta/gamma spectrometry.

10 Proofs

Proof of Proposition 1 : For $i = 1 \dots n$, the predictive likelihood ratio for testing H_0 vs H_1 is determined by

$$\begin{aligned}
\frac{l(\theta_1, (C_{gr,i}|C_{bl,i}))}{l(\theta_0, (C_{gr,i}, |C_{bl,i}))} &= \frac{l(\theta, (C_{gr,i}|C_{bl,i}))}{l(0, (C_{gr,i}|C_{bl,i}))} \\
&= \frac{\lambda^{C_{bl,i}} \exp(-\lambda)}{C_{bl,i}!} \times \frac{(\lambda+\theta)^{C_{gr,i}} \exp(-\lambda-\theta)}{C_{gr,i}!} \\
&= \frac{\lambda^{C_{bl,i}} \exp(-\lambda)}{C_{bl,i}!} \times \frac{\lambda^{C_{gr,i}} \exp(-\lambda)}{C_{gr,i}!} \\
&= \exp(\theta) \left(1 + \frac{\theta}{\lambda}\right)^{C_{gr,i}}.
\end{aligned}$$

Since $\theta > 0$, the likelihood ratio is increasing in a monotonic way with $C_{gr,i}$. Then based on Lemma of Neyman-Pearson in Lehmann et al (2005), the test using variable $C_{gr,i}$ will be the uniformly most powerful test, and the set of $C_{gr,i}$ where $\frac{l(\theta_1, (C_{gr,i}|C_{bl,i}))}{l(\theta_0, (C_{gr,i}|C_{bl,i}))} > k$ is equivalent to the set of $C_{gr,i}$ where $C_{gr,i} > DT_i$. The best test to determine whether $\theta > 0$ will be in the form of $C_{gr,i} > DT_i$.

In order to determine DT_i , the level constraint

$$P_{H_0}(C_{gr,i} > DT_i | C_{bl,i} = c_{bl,i}) = \alpha$$

must be satisfied. Or, knowing the measure of $C_{bl,i}$, based on [Manificat \(2024\)](#), the probability of having $C_{bl,i}$ in $C_{gr,i}$ is a negative binomial $\mathcal{NB}(c_{gr,i} + 1, 1/2, 1/2)$. Then

$$\mathbb{P}_{H_0}(C_{gr,i} = c_{gr,i} | C_{bl,i} = c_{bl,i}) = \mathbb{P}(C_{gr,i} = c_{gr,i} | C_{bl,i} = c_{bl,i}, \theta = 0) = \frac{(c_{gr,i} + c_{bl,i})!}{c_{gr,i}! c_{bl,i}!} \left(\frac{1}{2}\right)^{c_{gr,i} + c_{bl,i} + 1}.$$

Since we have N_i observation of $C_{gr,i}$ and $C_{bl,i}$, then DT_i is defined as follows:

$$\begin{aligned} \mathbb{P}(\overline{C_{gr,i}} > DT_i | \overline{C_{bl,i}} = \overline{c_{bl,i}}) &= \alpha \\ &= \mathbb{P}(\overline{C_{gr,i}} > DT_i | \overline{C_{bl,i}} = \overline{c_{bl,i}}) \\ &= 1 - \mathbb{P}(\overline{C_{gr,i}} \leq DT_i | \overline{C_{bl,i}} = \overline{c_{bl,i}}) \\ &= 1 - F(DT_i) \\ &= 1 - [1 - I_{1/2}(\overline{C_{bl,i}} + 1, DT_i + 1)] \\ &= I_{1/2}(\overline{C_{bl,i}} + 1, DT_i + 1), \end{aligned}$$

where I is the regularized beta incomplete function and F is the cumulative distribution of the negative binomial. Based on [Abramowitz and Stegun \(1988\)](#), the approximation of I for large $C_{gr,i}$ and $C_{bl,i}$, is expressed as

$$I_x(a, b) = \phi \left[3 \frac{(bx)^{1/3} \left(1 - \frac{1}{9b}\right) - (a(1-x))^{1/3} \left(1 - \frac{1}{9a}\right)}{\sqrt{\frac{(a(1-x))^{2/3}}{a} + \frac{(bx)^{2/3}}{b}}} \right] + O\left(\frac{1}{\min(a, b)}\right),$$

for $a, b > 0$ and $x \in [0, 1]$. Therefore, in our context, we obtain

$$1 - \alpha \sim \phi \left(\frac{DT_i - \overline{C_{bl,i}}}{\sqrt{\overline{C_{bl,i}} + DT_i}} \right).$$

Hence,

$$\phi^{-1}(1 - \alpha) \sqrt{\overline{C_{bl,i}} + DT_i} = DT_i - \overline{C_{bl,i}}.$$

Through solving the second degree equation,

$$-DT_i^2 + DT_i (k_{1-\alpha}^2 + 2\overline{C_{bl,i}}) + k_{1-\alpha}^2 \overline{C_{bl,i}} - \overline{C_{bl,i}}^2,$$

where $k_{1-\alpha} = \phi^{-1}(1 - \alpha)$, we obtain $\Delta > 0$ and we can choose

$$DT_i = \overline{C_{bl,i}} + \frac{k_{1-\alpha}^2 + \sqrt{k_{1-\alpha}^4 + 8k_{1-\alpha}^2 \overline{C_{bl,i}}}}{2}.$$

We equally find DT_i based on numerical Newton Raphson method using package *rootSolve* with the *R* software.

Proof of proposition 2 : It is be noted, knowing the measure of $C_{bl,i}$, the probability of having $C_{bl,i}$ in $C_{gr,i}$ is a negative binomial $\mathcal{NB}(c_{gr,i} + 1, 1/2, 1/2)$.

$$\mathbb{P}(C_{gr,i} = c_{gr,i} | C_{bl,i} = c_{bl,i}, \theta = 0) = \frac{(c_{gr,i} + c_{bl,i})!}{c_{gr,i}! c_{bl,i}!} \left(\frac{1}{2}\right)^{c_{gr,i} + c_{bl,i} + 1}.$$

If $C_{bl,i} \rightarrow \infty$, then

$$\mathbb{P}(C_{gr,i} = c_{gr,i} | C_{bl,i} = c_{bl,i}, \theta = 0) \rightarrow \frac{1}{\sqrt{4\pi c_{bl,i}}} \exp\left[-\frac{(c_{gr,i} - c_{bl,i})^2}{4c_{bl,i}}\right] = \frac{1}{\sqrt{4\pi c_{bl,i}}} \exp\left[-\frac{(c_{gr,i} + c_{bl,i} - 2c_{bl,i})^2}{4c_{bl,i}}\right].$$

In addition, we have

$$\lim_{n \rightarrow \infty} \mathcal{P}(\alpha) = \mathcal{N}(\alpha, \alpha).$$

Thus, when $C_{bl} \rightarrow \infty$, we get

$$\frac{\alpha^x \exp(-x)}{x!} \sim \frac{1}{\sqrt{2\pi\alpha}} \frac{\exp(-(x - \alpha)^2)}{2\alpha}.$$

Then, we have in our context,

$$\frac{1}{\sqrt{4\pi c_{bl,i}}} \exp\left[-\frac{(c_{gr,i} + c_{bl,i} - 2c_{bl,i})^2}{4c_{bl,i}}\right] \sim \frac{(2c_{bl,i})^{c_{gr,i} + c_{bl,i}} \exp(-2c_{bl,i})}{(c_{gr,i} + c_{bl,i})!},$$

which is a Poisson distribution of parameter $2c_{bl,i}$. Then DT_i is defined as follows:

$$\begin{aligned} \mathbb{P}(\overline{C_{gr,i}} > DT_i | \overline{C_{bl,i}}) &= \alpha \\ &= \mathbb{P}(\overline{C_{gr,i}} > DT_i | \overline{C_{bl,i}}) \\ &= 1 - \mathbb{P}(\overline{C_{gr,i}} \leq DT_i | \overline{C_{bl,i}}) \\ &= 1 - F(DT_i) \\ &= 1 - \text{cumulative distribution of } \mathcal{P}(2c_{bl,i}) \\ &= 1 - \frac{\Gamma(\lfloor DT_i + 1 \rfloor, 2\overline{C_{bl,i}})}{\lfloor DT_i \rfloor}, \end{aligned}$$

where $\Gamma(.,.)$ is the upper incomplete gamma function and $\lfloor . \rfloor$ is the floor function (see [Barbour et al \(1992\)](#)). We equally find DT_i based on numerical methods using package *rootSolve* with the *R* software.

Proof of Proposition 3 : It is to be noted that the Delaporte law is a convolution between the negative binomial law and the Poisson.

$$\begin{aligned} \mathbb{P}(C_{gr} = d | C_{bl} = c, \theta) &= \sum_{c' \in \mathbb{N}} \frac{\theta^{d-c'} \exp(-\theta) (c' + c)!}{(d - c')! c'! c!} \left(\frac{1}{2}\right)^{c' + c} \\ &= \sum_{c'=0}^d \mathbb{P}(N_1 = d - c') \mathbb{P}(N_2 = c') \\ &= \mathbb{P}(N_1 + N_2 = d) \\ &= g_1 \star g_2(d), \end{aligned}$$

where $N_1 \sim \mathcal{P}(\theta)$, $N_2 \sim \mathcal{NB}(c, \frac{1}{2}, \frac{1}{2})$ and g_1 stands for the Poisson law and g_2 corresponds to the negative binomial law.

The negative binomial law is a sum of geometric laws. Therefore according to the TLC theorem, the negative binomial law converges to the normal distribution of mean c and variance $2c$, when $C_{blank}(c) \rightarrow \infty$:

$$g_2(c') = \frac{(c' + c)!}{c'!c!} \left(\frac{1}{2}\right)^{c'+c} \rightarrow \frac{\exp\left(\frac{-(c'-c)^2}{4c}\right)}{\sqrt{4\pi c}} = \frac{\exp\left(\frac{-(c'+c-2c)^2}{4c}\right)}{\sqrt{4\pi c}}.$$

It is noteworthy that, if a Gaussian distribution has a large variance, it can be considered equivalent to a Poisson distribution $\mathcal{P}(2c)$:

$$\frac{\exp\left(\frac{-(c'+c-2c)^2}{4c}\right)}{\sqrt{4\pi c}} \sim \frac{(2c)^{c'+c} \exp(-2c)}{(c' + c)!}.$$

As a matter of fact, we can write the Delaporte law as a convolution of two Poisson distributions with a parameter given by the sum of the Poisson parameters. Thus, we obtain

$$\begin{aligned} \mathbb{P}(C_{gr} = d | C_{bl} = c, \theta) &= \sum_{c'=0}^d \mathbb{P}(N_1 = d - c') \mathbb{P}(N_2 = c' + c) \\ &= \frac{(\theta + 2c)^{d+c}}{(d + c)!} \exp(-\theta - 2c). \end{aligned}$$

In order to determine the estimator of θ , we calculate the predictive likelihood function

$$l(C_{gr,1}, \dots, C_{gr,n} | C_{bl,1}, \dots, C_{bl,n}, \theta) = \prod_{i=1}^n \frac{(\theta + 2c_i)^{d_i+c_i}}{(d_i + c_i)!} \exp(-\theta - 2c_i).$$

Since we need to use the logarithm in order to turn multiplication into addition, the log-likelihood L is provided by

$$\begin{aligned} L(C_{gr,1}, \dots, C_{gr,n} | C_{bl,1}, \dots, C_{bl,n}, \theta) &= \log l(C_{gr,1}, \dots, C_{gr,n} | C_{bl,1}, \dots, C_{bl,n}, \theta) \\ &= \sum_{i=1}^n \log \left(\frac{(\theta + 2c_i)^{d_i+c_i}}{(d_i + c_i)!} \exp(-\theta - 2c_i) \right). \end{aligned}$$

Therefore, we need to resolve the following equation

$$\frac{\partial L(C_{gr,1}, \dots, C_{gr,n} | C_{bl,1}, \dots, C_{bl,n}, \theta)}{\partial \theta} = 0.$$

Hence,

$$\sum_{i=1}^n \frac{d_i - c_i - \theta}{\theta + 2C_{bl,i}} = 0. \quad (10.1)$$

Since $C_{gr,i}$ is large, then (10.1) implies $\sum_{i=1}^n C_{bl,i} - \sum_{i=1}^n C_{gr,i} + n\theta = 0$. Thus,

$$\theta_n = \frac{\sum_{i=1}^n C_{gr,i} - C_{bl,i}}{n}.$$

Proof of Proposition 4 : We first state the radioactivity as follows:

$$\begin{aligned} A &= \frac{f}{\epsilon t V} (C_{gross} - C_{blank}) \\ &= YX \\ &= g(Y, X) \end{aligned}$$

where $X = C_{gr} - C_{bl}$, $Y = \frac{f}{\epsilon t V} \sim \mathcal{N}(a, b^2)$ and $g(u, v) = uv$. For any $g(x; y)$, the bi-variate first order Taylor expansion about any $a = (a_x, a_y)$ is

$$g(x, y) = g(a) + \frac{\partial g}{\partial x}(a)(x - a_x) + \frac{\partial g}{\partial y}(a)(y - a_y) + Rest.$$

Switching to random variables with finite means $\mathbb{E}(X) := \mu_x$ and $\mathbb{E}(Y) := \mu_y$, we can choose the expansion point to be $a = (\mu_x, \mu_y)$. In that case, the first order Taylor series approximation for $g(X, Y)$ is

$$g(X, Y) = g(a) + \frac{\partial g}{\partial x}(a)(X - \mu_x) + \frac{\partial g}{\partial y}(a)(Y - \mu_y) + Rest.$$

The approximation for $\mathbb{E}(g(X, Y))$ is therefore

$$\begin{aligned} \mathbb{E}(g(X, Y)) &\sim \mathbb{E}(g(a)) + \frac{\partial g}{\partial x}(a)\mathbb{E}((X - \mu_x)) + \frac{\partial g}{\partial y}(a)\mathbb{E}((Y - \mu_y)) \\ &= g(\mu_x, \mu_y) + 0 + 0. \end{aligned}$$

Relying on the definition of variance, the variance of $g(X, Y)$ is

$$\text{VAR}(g(X, Y)) = \mathbb{E} \left[(g(X, Y) - \mathbb{E}(g(X, Y)))^2 \right] \sim \mathbb{E} \left[(g(X, Y) - g(\mu_x, \mu_y))^2 \right].$$

In this respect, using the first order Taylor expansion for $g(X, Y)$ expanded around a :

$$\begin{aligned} \text{VAR}(g(X, Y)) &\sim \mathbb{E} \left[\left(g(\mu_x, \mu_y) + \frac{\partial g}{\partial x}(\mu_x, \mu_y)(X - \mu_x) + \frac{\partial g}{\partial y}(\mu_x, \mu_y)(Y - \mu_y) - g(\mu_x, \mu_y) \right)^2 \right] \\ &= \mathbb{E} \left[\left(\frac{\partial g}{\partial x}(\mu_x, \mu_y)(X - \mu_x) + \frac{\partial g}{\partial y}(\mu_x, \mu_y)(Y - \mu_y) \right)^2 \right] \\ &= \mathbb{E} \left[\left(\frac{\partial g}{\partial x}(\mu_x, \mu_y) \right)^2 (X - \mu_x)^2 + 2 \frac{\partial g}{\partial x}(\mu_x, \mu_y)(X - \mu_x) \frac{\partial g}{\partial y}(\mu_x, \mu_y)(Y - \mu_y) + \left(\frac{\partial g}{\partial y}(\mu_x, \mu_y) \right)^2 (Y - \mu_y)^2 \right] \\ &= \left(\frac{\partial g}{\partial x}(\mu_x, \mu_y) \right)^2 \text{VAR}(X) + 2 \frac{\partial g}{\partial x}(\mu_x, \mu_y) \frac{\partial g}{\partial y}(\mu_x, \mu_y) \text{COV}(X, Y) + \left(\frac{\partial g}{\partial y}(\mu_x, \mu_y) \right)^2 \text{VAR}(Y). \end{aligned}$$

For our case $g(x, y) = xy$, the Taylor approximation is indicated as follows:

$$\begin{aligned} A &= g(X, Y) \\ &= \mathbb{E}(X)\mathbb{E}(Y) + \frac{\partial g}{\partial x}(\mathbb{E}(X), \mathbb{E}(Y))[X - \mathbb{E}(X)] + \frac{\partial g}{\partial y}(\mathbb{E}(X), \mathbb{E}(Y))[Y - \mathbb{E}(Y)] + Rest. \end{aligned}$$

The approximation of $\mathbb{E}(A) = \mathbb{E}(g(X, Y))$ is:

$$\mathbb{E}(A) \sim \mathbb{E}(X)\mathbb{E}(Y) \sim \overline{XY}.$$

We therefore have

$$\frac{\partial g}{\partial X} = Y, \quad \frac{\partial g}{\partial Y} = X,$$

$$\left(\frac{\partial g}{\partial X}(\mathbb{E}(X), \mathbb{E}(Y)) \right)^2 = \mathbb{E}(Y)^2,$$

$$\frac{\partial g}{\partial X}(\mathbb{E}(X), \mathbb{E}(Y)) \frac{\partial g}{\partial Y}(\mathbb{E}(X), \mathbb{E}(Y)) = \mathbb{E}(X)\mathbb{E}(Y)$$

$$\left(\frac{\partial g}{\partial Y}(\mathbb{E}(X), \mathbb{E}(Y)) \right)^2 = \mathbb{E}(X)^2.$$

Then, as X and Y are independent, the approximation of $\text{VAR}(A) = \text{VAR}[g(X, Y)]$ is:

$$\text{VAR}(A) \sim \mathbb{E}(Y)^2 \text{VAR}(X) + \mathbb{E}(X)^2 \text{VAR}(Y).$$

Hence,

$$\mu_{A,n} = \overline{XY} = \frac{1}{n} \sum_{i=1}^n X_i \frac{1}{n} \sum_{i=1}^n Y_i.$$

$$\sigma_{A,n}^2 = \left(\frac{1}{n} \sum_{i=1}^n Y_i \right)^2 \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_{X,n})^2 + \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu_{Y,n})^2.$$

Proof of Proposition 5 : *Asymptotic properties and confidence interval of $\mu_{A,n}$:* Based on TLC Theorem, we get

$$\sqrt{n}(\overline{X} - \mu_X) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_X^2) \text{ and } \sqrt{n}(\overline{Y} - \mu_Y) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_Y^2).$$

Since \overline{X} and \overline{Y} are independent, then

$$\sqrt{n} \left[\begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix} - \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \right] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right),$$

where Σ the variance covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix}.$$

Based on Delta method, we obtain

$$\sqrt{n}(\mu_{A,n} - \mu_A) = \sqrt{n}(g(\overline{X}, \overline{Y}) - g(\mu_X, \mu_Y)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, Dg(\mu_X, \mu_Y) \Sigma Dg(\mu_X, \mu_Y)^t),$$

where $Dg(\mu_X, \mu_Y) = \left(\frac{\partial g}{\partial \mu_X}, \frac{\partial g}{\partial \mu_Y} \right)^t = (\mu_Y, \mu_X)^t$ stands for the Jacobian matrix of $g(\mu_X, \mu_Y)$. Since

$$Dg(\mu_X, \mu_Y) \Sigma Dg(\mu_X, \mu_Y)^t = \mu_X^2 \sigma_Y^2 + \mu_Y^2 \sigma_X^2,$$

then the previous result allows us to build up the confidence interval of $\mathbb{E}(A) = \mu_A$ which has the following form

$$I_{\mu_A} = \left(\overline{XY} \pm \frac{z_{1-\frac{\delta}{2}} \sigma_n(x)}{\sqrt{n}} \right),$$

where $z_{1-\frac{\delta}{2}}$ is the normal $(1 - \frac{\delta}{2})$ quantile and $\sigma_n(x) = \sqrt{\overline{X}^2 \sigma_{Y,n}^2 + \overline{Y}^2 \sigma_{X,n}^2}$. Based on [Rosenblatt \(1956\)](#), it follows that

$$\sqrt{nh_n}(\hat{f}_{A,n}(z) - f_A(z)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, f_A(z) \int K(u)du\right).$$

Then, the confidence interval based on kernel density estimator is given by

$$I_{f_A(z)} = \left(\hat{f}_{A,n}(z) \pm \frac{z_{1-\frac{\delta}{2}} \sigma_f(z)}{\sqrt{n}} \right),$$

where $z_{1-\frac{\delta}{2}}$ is the normal $(1 - \frac{\delta}{2})$ quantile and $\sigma_f(z) = \sqrt{\hat{f}_{n,A}(z) \int K(u)du}$.

Acknowledgement

The authors gratefully acknowledge the support of Anouar EN NOUGAOUI and all the member nuclear measurement laboratory of IRSN.

References

- Abramowitz, M., and Stegun, I. A., (1988). Handbook of mathematical functions with formulas, graphs, and mathematical tables.
- Bagui, S., and Mehra, K. L., (2022). Convergence of binomial to normal: multiple proofs. *American Journal of Mathematics and Statistics*, **9**, 44-50.
- Barbour, A. D., Holst, L., and Janson, S., (1992). Poisson approximation. *Oxford University Press*.
- BIPM, I., IFCC, I., and ISO, I., (2008). Evaluation des données de mesure-Guide pour l'expression de l'incertitude de mesure. *JCGM. Genève, Switzerland*.
- Ciarlini, P., and Regoliosi, G., (2000). In Advanced Mathematical and Computational Tools in Metrology IV. , p. 36-44.
- Currie, L. A., (1968). Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Analytical Chemistry*. **40** 586-593.
- Dodge, Y., (2003). The Oxford Dictionary of Statistical Terms. *OUP Oxford*.
- De Bièvre, P., (2012). The 2012 International Vocabulary of Metrology:"VIM". *Chemistry International-News magazine for IUPAC*. **34** 26-27.
- JCGM 100, (2008). Evaluation of Measurement Data. *Guide to The Expression of Uncertainty in Measurement (Saint-Cloud: BIPM)*.

- Hiller, J., Genta, G., Barbato, G., De Chiffre, L., and Levi, R., (2014). Measurement uncertainty evaluation in dimensional X-ray computed tomography using the bootstrap method. *International journal of precision engineering and manufacturing*, **15** 617-622.
- Lehmann, E. L., Romano, J. P., and Casella, G. , (2005). Testing statistical hypotheses. *New York: springer.*, **3**.
- Lovric, M., (2011). International encyclopedia of statistical science. *Springer Berlin Heidelberg*.
- Manificat, G., (2024). Classical and Bayesian statistical methods for low-level metrology. *hal-04483964*.
- ISO 11929, (2020). Determination of the characteristic limits (decision threshold, detection limit and limit of the confidence interval) for measurements of ionizing radiations - Fundamentals and Application. *International Organization for Standardization, Geneva*.
- Rosenblatt, M.,, (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The annals of mathematical statistics*, **27** 832—83.