Sample size determination to evaluate ballast water standards: a decision-theoretic approach

Journal:	*Communications in Statistics – Theory and Methods*
Manuscript ID	LSTA-2019-0360
Manuscript Type:	Original Paper
Date Submitted by the Author:	04-Apr-2019
Complete List of Authors:	Costa, Eliardo; Universidade Federal do Rio Grande do Norte, Departamento de Estatística Paulino, Carlos Daniel; Universidade de Lisboa Instituto Superior Tecnico, Departamento de Matemática, CEAUL, FCUL Singer, J.M.; Universidade de São Paulo, Statistics
Keywords:	Bayes risk, Poisson distribution, negative binomial distribution
Abstract:	With the objective of verifying compliance with international standards, we employ a Bayesian decision approach to compute minimum sample sizes for estimating organism concentration in ballast water. To obtain the minimum sample size, we use a total cost minimization criterion defined as the sum of the sampling cost and the Bayes risk either under a Poisson/gamma model with a gamma prior distribution or under a negative binomial distribution with a Pearson Type VI prior distribution. We also conduct a simulation study to evaluate credible interval lengths associated with the proposed minimum sample sizes.

Sample size determination to evaluate ballast water standards: a decision-theoretic approach

Eliardo G. Costa*[1], Carlos Daniel Paulino[2], and Julio M. Singer[3]

[1]Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Brazil
[2]Departamento de Matemática, IST and CEAUL, FCUL, Universidade de Lisboa, Portugal
[3]Departamento de Estatística, Universidade de São Paulo, Brazil

Abstract

With the objective of verifying compliance with international standards, we employ a Bayesian decision approach to compute minimum sample sizes for estimating organism concentration in ballast water. To obtain the minimum sample size, we use a total cost minimization criterion defined as the sum of the sampling cost and the Bayes risk either under a Poisson/gamma model with a gamma prior distribution or under a negative binomial distribution with a Pearson Type VI prior distribution. We also conduct a simulation study to evaluate credible interval lengths associated with the proposed minimum sample sizes.

Keywords: Bayes risk, Poisson distribution, negative binomial distribution.

*Corresponding author. *E-mail address:* eliardocosta@ccet.ufrn.br

1 Introduction

With the expansion of maritime traffic, ballast water has become the leading dispersing agent of invasive organisms with serious environmental, public health and economic consequences as indicated in Strayer (2010), McCarthy *et al.* (1992) and Marbuah *et al.* (2014). An overview of research in this field in the last thirty years is presented in Bailey (2015). In order to reduce these effects the International Maritime Organization (IMO) proposed the D-2 standard which sets upper limits on the organism concentration in ballast water discharged by ships. The D-2 standard requires that deballasted water should contain no more than 10 living organisms (referred to simply as organisms in the remainder) per mL, for organisms with maximum dimension between 10 μm and 50μm among other restrictions. Recently, Cohen *et al.* (2017) suggested that the standards must be re-evaluated and the limits must be even smaller. Given the large amount of ballast water carried by some vessels, it is impractical to analyze the whole water volume and an alternative is to rely on sampling methods that guarantee some pre-specified acceptable error rates associated to the decision of whether a given deballasting process complies with the D-2 standard.

Adopting a frequentist approach, Costa *et al.* (2015, 2016) addressed the problem of determining the appropriate sample size either from the point of view of hypothesis testing or interval inference, respectively. Costa *et al.* (2019), on the other hand, adopted a Bayesian approach based on credible intervals to determine the required sample size using average coverage and average length criteria. All these approaches take the inherent heterogeneity

distribution of the organism concentration in ballast tanks into account.

Here we consider a Bayesian decision approach, with a total cost minimization criterion which minimizes the sum of the sampling cost and the Bayesian risk. An advantage of this approach is that the cost of collecting the sample is explicitly taken into account.

For the Bayes risk, we must specify a loss function based on the specification of two quantities, namely, the lower [say, $a(\boldsymbol{x}_n)$] and the upper [say, $b(\boldsymbol{x}_n)$] limits of a credible interval for the mean organism concentration obtained from a sample $\boldsymbol{x}_n$ of organisms in n aliquots with a given volume w collected from a ballast water tank.

Once the required minimum sample size, say n_{m}, has been determined, a real dataset $\boldsymbol{x}_{n_{\mathrm{m}}}$ is collected and the ship is declared not compliant with the D-2 standard if $a(\boldsymbol{x}_{n_{\mathrm{m}}}) > 10$ or compliant, if $b(\boldsymbol{x}_{n_{\mathrm{m}}}) < 10$. Otherwise, if $a(\boldsymbol{x}_{n_{\mathrm{m}}}) < 10 < b(\boldsymbol{x}_{n_{\mathrm{m}}})$, more data are needed to make a decision.

In a different setup, Etzioni & Kadane (1993) use a similar criterion with quadratic and logarithmic loss functions under a normal model. Sahu & Smith (2006) consider a loss function for the hypothesis testing problem of the parameter of a normal model. Islam (2011) and Islam & Pettit (2012, 2014) consider quadratic, linex and bounded linex loss functions for point estimation of the mean and the variance of a normal model with normal prior distributions or exponential and Poisson models both with a gamma prior distribution for point estimation of their respective parameters. Following the same approach or the one in which a loss function must be specified, we may cite Pham-Gia & Turkkan (1992), Bernardo (1997), Lindley (1997), Brutti *et al.* (2008, 2009), Parmigiani & Inoue (2009), Brutti *et al.* (2014), De Santis

& Gubbiotti (2017), among others.

In Section 2 we describe the Bayesian models. Sample size determination along with a simulation study is presented in Section 3. We conclude with a discussion in Section 4.

2 Bayesian models

2.1 Poisson model with a gamma prior distribution

Let X be the number of organisms in an aliquot of volume w collected from a ballast tank with organism concentration λ. The expected number of organisms in this aliquot is $w\lambda$, *i.e.*, $\mathbb{E}[X|\lambda] = w\lambda$. Suppose that, given λ, X follows a Poisson distribution with mean $w\lambda$; this essentially corresponds to the assumption that the organisms are homogeneously distributed in the ballast tank. A possible and first natural choice for a prior distribution is the conjugate gamma distribution in which density,

$$f(\lambda) \propto \lambda^{\theta_0 - 1} \exp(-\theta_0 \lambda / \lambda_0),$$

where λ_0 and θ_0 are positive and known fixed constants (hyperparameters), respectively interpreted as a prior estimate for the mean concentration and the corresponding variability. The larger (smaller) θ_0, the smaller (larger) the prior uncertainty about λ.

Considering a random sample of size n of $X|\lambda$ and a gamma prior distri-

bution for λ, we may write the model hierarchically as follows

$$X_i|\lambda \overset{\text{iid}}{\sim} \text{Poisson}(w\lambda), \quad i = 1, 2, \ldots, n; \tag{1}$$

$$\lambda \sim \text{Gamma}(\theta_0, \theta_0/\lambda_0). \tag{2}$$

In this context, the posterior distribution of λ is a gamma distribution with parameters $\theta_0 + s_n$ and $nw + \theta_0/\lambda_0$, where $s_n = \sum_{i=1}^{n} x_i$, *i.e.*, $\lambda|\boldsymbol{x}_n \sim \text{Gamma}(\theta_0 + s_n, nw + \theta_0/\lambda_0)$, where $\boldsymbol{x}_n = (x_1, \ldots, x_n)$. Details are presented in the Supplementary Material.

2.2 Negative binomial model with Pearson Type VI prior distribution

Suppose that the organism concentration in the i-th aliquot is λ_i and the corresponding number of organisms is X_i, $i = 1, \ldots, n$. The expected number of organisms in the i-th aliquot is $w\lambda_i$, *i.e.*, $\mathbb{E}\left[X_i|\lambda_i\right] = w\lambda_i$. For $i = 1, \ldots, n$, suppose that, given λ_i, X_i follows a Poisson distribution with mean $w\lambda_i$ and that given a mean concentration λ in the tank, $\lambda_i \sim \text{Gamma}(\phi, \phi/\lambda)$, so that $\mathbb{E}\left[\lambda_i|\lambda\right] = \lambda$ and $\text{Var}\left[\lambda_i|\lambda\right] = \lambda^2/\phi$. Thus, given λ and ϕ, X_i follows a negative binomial distribution with $\mathbb{E}\left[X_i|\lambda, \phi\right] = w\lambda$ and $\text{Var}\left[X_i|\lambda, \phi\right] = w\lambda + (w\lambda)^2/\phi$, where ϕ is a shape (or agglomeration) parameter assumed known; we use the notation $X_i|\lambda, \phi \sim \text{NB}(w\lambda, \phi)$.

A natural conjugate prior distribution for the negative binomial distribu-

tion is the Pearson Type VI distribution (Johnson *et al.*, 1994a,b), *i.e.*,

$$f(\lambda) \propto \left(\frac{w}{\phi}\lambda\right)^{\theta_0-1}\left(1+\frac{w}{\phi}\lambda\right)^{-\theta_0-(\theta_0/\lambda_0+1)},$$

with location parameter 0, scale parameter ϕ/w and shape parameters θ_0 and θ_0/λ_0+1, where λ_0 and θ_0 are known positive fixed constants (hyper-parameters). We use the notation $\lambda \sim \text{PVI}(0, \phi/w, \theta_0, \theta_0/\lambda_0+1)$. In this case, $\mathbb{E}[\lambda] = (\phi/w)\lambda_0$ and $\text{Var}[\lambda] = (\lambda_0^2/\theta_0)[\phi^2(\lambda_0+1)/(w^2(1-\lambda_0/\theta_0))]$, for $\lambda_0 < \theta_0$.

Considering a random sample of size n from $X|(\lambda,\phi)$ and a Pearson Type VI prior distribution for λ; then we may write the model hierarchically as

$$X_i|\lambda,\phi \overset{\text{iid}}{\sim} \text{NB}(w\lambda,\phi), \quad i=1,2,\ldots,n; \tag{3}$$

$$\lambda \sim \text{PVI}(0,\phi/w,\theta_0,\theta_0/\lambda_0+1). \tag{4}$$

In this context, the posterior distribution of λ is also a Pearson Type VI distribution, with the same location and scale parameters of the prior distribution, and shape parameters θ_0+s_n and $\theta_0/\lambda_0+n\phi+1$, *i.e.*, $\lambda|\boldsymbol{x}_n \sim \text{PVI}(0,\phi/w,\theta_0+s_n,\theta_0/\lambda_0+n\phi+1)$. Details are presented in the Supplementary Material.

3 Sample size determination

3.1 Total cost minimization

A way to approach the problem of sample size determination is to consider it as a decision problem (Lindley, 1997; Parmigiani & Inoue, 2009; Islam & Pettit, 2014). For this purpose, given that λ is the parameter of interest, it is necessary to specify a loss function $L(\lambda, d_n)$ based on a sample $X_1, \ldots, X_n$ and a decision d_n. For interval inference, a decision consists of the specification of two quantities, the lower [say, $a(\boldsymbol{x}_n)$] and the upper [say, $b(\boldsymbol{x}_n)$] limits, limits of a credible interval for the parameter of interest λ. For simplicity of notation, we drop the argument $\boldsymbol{x}_n$.

Let f be a prior distribution for the unknown parameter λ and d_n a decision function; then the Bayes risk is (see Parmigiani & Inoue, 2009)

$$r(f, d_n) = \int_{\Lambda} \int_{\mathcal{X}^n} L(\lambda, d_n) g(\boldsymbol{x}_n|\lambda) f(\lambda) d\boldsymbol{x}_n d\lambda \quad (5)$$

where Λ is the parameter space. The Bayes risk $r(f, d_n)$ may be viewed as a mean of the sampling expected loss (as a function of the parameter of interest) weighted by the prior distribution; this is a way to summarize the sampling expected loss over all possible values of the parameter of interest (here, the mean concentration λ).

The decision d_n^* that minimizes $r(f, d_n)$ among all the possible decisions d_n is called a Bayes rule. Note that if the order of the integration may be

inverted, we have

$$\begin{aligned} r(f, d_n) &= \int_{\mathcal{X}^n} \left[\int_{\Lambda} L(\lambda, d_n) f(\lambda|\boldsymbol{x}_n) d\lambda \right] g(\boldsymbol{x}_n) d\boldsymbol{x}_n \\ &= \int_{\mathcal{X}^n} \mathbb{E}\left[L(\lambda, d_n) \middle| \boldsymbol{x}_n \right] g(\boldsymbol{x}_n) d\boldsymbol{x}_n, \end{aligned} \tag{6}$$

so that the decision d_n^* that minimizes $r(f, d_n)$ is the same that minimizes the posterior expected value of the loss function, namely, $\mathbb{E}\left[L(\lambda, d_n) \middle| \boldsymbol{x_n} \right]$. In this context, the required sample size minimizes the total cost

$$\mathrm{TC}(n) = r(f, d_n^*) + cn,$$

where c is the cost of sampling an aliquot. Often it is not possible to compute $r(f, d_n^*)$ analytically. In such cases, we may use Monte Carlo simulations as an alternative. Since simulation methods are used, the estimates of $\mathrm{TC}(n)$ may show a variation around its true value. We may reduce this variation in the following ways: (i) taking the number of Monte Carlo replicates as large as possible and/or, (ii) fitting a curve by least squares or some other method to the estimates of $\mathrm{TC}(n)$ as a function of n. Müller & Parmigiani (1995) propose to fit the following curve to the estimates of $\mathrm{TC}(n)$,

$$\mathrm{TC}(n) = \frac{E}{(1 + Hn)^G} + cn,$$

where E, H, e G are parameters to be estimated. The numerical methods required to estimate these parameters sometimes do not reach convergence depending on the initial values for the corresponding algorithms. Because

(i) the parameters H and G play similar roles and essentially represent a kind of decreasing rate of the Bayes risk and (ii) taking into account that the expression $(1+n)^G$ offers more alternatives to a decreasing rate than $1+Hn$, we propose to fit the function

$$\mathrm{TC}(n) = \frac{E}{(1+n)^G} + cn,$$

that may be linearized as

$$\log[\mathrm{TC}(n) - cn] = \log E - G\log(1+n), \tag{7}$$

where the term $-\log(1+n)$ may be interpreted as an explanatory variable and $\log[\mathrm{TC}(n) - cn]$, as a dependent variable like in linear regression. Assuming that an error is added, the estimates of E and G may be computed by least squares. Then, the required sample size is the largest integer closest to

$$\left(\frac{\widehat{E}\,\widehat{G}}{c}\right)^{1/(\widehat{G}+1)} - 1, \tag{8}$$

where $\widehat{E}$ and $\widehat{G}$ are, respectively, the estimates of E and G obtained by least squares.

3.2 Loss functions

Loss function 1

Firstly, we consider the following loss function

$$L(\lambda, d_n) = \rho\tau + (a - \lambda)^+ + (\lambda - b)^+, \tag{9}$$

where $0 < \rho < 1$ is a weight, $\tau = (b - a)/2$ is the half-length of the desired interval, the function x^+ is equal to x if $x > 0$ and equal to zero, otherwise. The smaller is τ, the narrower the interval. The terms $(a - \lambda)^+$ and $(\lambda - b)^+$ are included to penalize intervals that do not contain the parameter of interest (λ). These terms are equal to zero if $\lambda \in [a, b]$ and increase as λ moves away from the interval. Note that the loss function (9) is a weighted sum of two terms, τ and $(a-\lambda)^+ + (\lambda-b)^+$, where the weights are ρ and 1, respectively. In this context, Rice *et al.* (2008) argue that the second term of the loss function must receive the greatest weight, *i.e.*, $\rho < 1$. The Bayes rule corresponds to taking a and b as the quantiles of probabilities $\rho/2$ and $1 - \rho/2$ of the posterior distribution of λ.

An algorithm to obtain the minimum sample size satisfying the total cost minimization for this loss function is outlined in the Supplementary Material. Sample sizes computed under either Poisson/gamma or negative binomial/Pearson Type VI distributions via loss function 1 are displayed in Tables 1 and 2, respectively. In Figure 1 we depict a curve fitted to the estimated Bayes risk as a function of n for the negative binomial/Pearson Type VI model with $\theta_0 = 11$, $\phi = 10$, $w = 0.5$, $c = 0.005$ and $\lambda_0 = 10(w/\phi)$.

The vertical line indicates the minimum n at 23. Note that for each n the estimates (ten for each n) of the Bayes risk do not vary much.

3.2.1 Loss function 2

A second loss function is

$$L(\lambda, d_n) = \gamma\tau + (\lambda - m)^2/\tau,$$

where $\gamma > 0$ is a fixed constant and $m = (a + b)/2$ is the center of the credible interval. The first term involves the half-width of the interval and the second, the square of the distance between the parameter of interest and the center of the interval, which it is divided by the half-width to maintain the same measurement unit of the first term. The weights attributed to each term are γ and 1, respectively. If $\gamma < 1$, we attribute the greatest weight to the second term; if $\gamma > 1$, the situation is reversed and if $\gamma = 1$ the two terms have the same weight. In this case, the Bayes rule corresponds to the quantities which define the interval $[a^*, b^*] = [m - \mathrm{SD}_\gamma, m + \mathrm{SD}_\gamma]$, where $(m, \mathrm{SD}_\gamma) = (\mathbb{E}\left[\lambda|\boldsymbol{x}_n\right], \gamma^{-1/2}\sqrt{\mathrm{Var}\left[\lambda|\boldsymbol{x}_n\right]})$. For more details see Parmigiani & Inoue (2009), Rice *et al.* (2008) and Schervish (1995).

An algorithm to obtain the minimum sample size satisfying the the total cost minimization for this loss function is also outlined in the Supplementary Material. Sample sizes computed under either Poisson/gamma or negative binomial/Pearson Type VI distributions via loss function 2 are displayed in Tables 3 and 4, respectively.

3.3 Simulation study

We carried out a simulation study to evaluate the lengths of the credible intervals associated with the proposed minimum sample sizes. For such purposes, we considered the sample sizes obtained previously for each w, c, γ, θ_0 (and ϕ in the negative binomial model with Pearson Type VI prior distribution). For each scenario and given n, we drew a sample of size 100 of $\boldsymbol{x}_n$, obtained the corresponding posterior credible intervals and computed the mean of their lengths. The results are displayed in Tables 1-4.

4 Discussion

The results in Table 1 obtained under the Poisson/gamma model indicate that n does not decrease much when θ_0 increases. Costa *et al.* (2019) also observed the same behavior with a different approach to compute the minimum sample size. This feature is also visible when we compute n under the same model via loss function 2. Under the negative binomial/Pearson Type VI model, the sample size is directly affected when we vary θ_0 with fixed ϕ, or vary ϕ with fixed θ_0 (see Tables 2 and 4). This is also observed in Costa *et al.* (2019).

In general, the sample sizes computed via loss function 1 are smaller than those obtained via loss function 2 (see Tables 1-4). This may be justified by the fact that when we consider $\rho = 0.05$ in loss function 1, we are giving a larger weight (equal to 1) to the event in which the parameter λ lies in the credible interval. Consequently, the length of the interval may be larger, leading to smaller sample sizes. On the other hand, when we consider loss function 2 with $\gamma = 1$, we are giving the same weight to both the length of

the credible interval and to the distance between λ and the interval center; even when $\gamma = 1/4$ and the second term receives the greatest weight, this loss function seems to provide more conservative intervals. Hence, larger sample sizes are required.

The results in Table 2, obtained under model (3)-(4) via loss function 1 show that the sample sizes increase with the parameter ϕ. This may be justified by the fact that ϕ is a scaling parameter in the prior distribution; the larger ϕ the greater the uncertainty about λ. Additionally, the loss function 1 providing the quantiles of the posterior distribution for inference may be an additional explanation. Note that with loss function 2 this does not happen, maybe because of the conservative credible intervals generated (posterior mean $\mp$ posterior standard deviation), see Table 4.

All the features pointed out previously may also be observed via the estimated posterior credible intervals lengths obtained via simulation (see Tables 1-4). From the practical point of view, loss function 2 is preferred because in general, it seems to provide posterior credible intervals with smaller lengths than those from loss function 1, specially for $\gamma = 1$ in which the same weight is given for both components that compose the loss function. In this context, for the negative binomial model, we have the same problem of considering the ϕ parameter known as in Costa *et al.* (2019). However, since the minimum n is not heavily affected by the increase in ϕ, and if we want to be conservative we may consider ϕ as the smallest possible ($\phi = 1$, for example) and collect no more than 10 additional aliquots compared to the case where ϕ is large. In this sense, we recommend the use of the loss function 2, and for the negative binomial/Pearson Type VI model setting

$\gamma = 1$, which provide credible intervals with lengths not greater than 3.

Although the focus of this study is ballast water sampling, similar results may be applied to other problems in which the Poisson or the negative binomial models underlie the data generating process.

Acknowledgements

This research received financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grants 153526/2014-9 and 3304126/2015-2) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant 2013/21728-2), Brazil.

References

Bailey, S. A. (2015). An overview of thirty years of research on ballast water as a vector for aquatic invasive species to freshwater and marine environments. *Aquatic Ecosystem Health and & Management* **April**, 37–41.

Bernardo, J. M. (1997). Statistical inference as a decision problem: the choice of sample size. *Journal of the Royal Statistical Society. Series D (The Statistician)* **46**, 151–153.

Brutti, P., De Santis, F. & Gubbiotti, S. (2008). Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine* **27**, 2290–2306.

Brutti, P., De Santis, F. & Gubbiotti, S. (2009). Mixtures of prior distributions for predictive Bayesian sample size calculations in clinical trials. *Statistics in Medicine* **28**, 2185–2201.

Brutti, P., De Santis, F. & Gubbiotti, S. (2014). Predictive measures

of the conflict between frequentist and Bayesian estimators. *Journal of Statistical Planning and Inference* **148**, 111–122.

Cohen, A. N., Dobbs, F. C. & Chapman, P. M. (2017). Revisiting the basis for us ballast water regulations. *Marine Pollution Bulletin* **118**, 348–353.

Costa, E. G., Lopes, R. M. & Singer, J. M. (2015). Implications of heterogeneous distributions of organisms on ballast water sampling. *Marine Pollution Bulletin* **91**, 280–287.

Costa, E. G., Lopes, R. M. & Singer, J. M. (2016). Sample size for estimating the mean concentration of organisms in ballast water. *Journal of Environmental Management* **180**, 433–438.

Costa, E. G., Paulino, C. D. & Singer, J. M. (2019). Sample size for estimating organism concentration in ballast water: a bayesian approach. *Submitted* **-**, –.

De Santis, F. & Gubbiotti, S. (2017). A decision-theoretic approach to sample size determination under several priors. *Applied Stochastic Models in Business and Industry* **33**, 282–295.

Etzioni, R. & Kadane, J. B. (1993). Optimal experimental design for another's analysis. *Journal of the American Statistical Association* **88**, 1404–1411.

Islam, A. F. M. (2011). *Loss functions, utility functions and Bayesian*

sample size determination. Tese de doutorado. Queen Mary, University of London.

Islam, A. F. M. S. & Pettit, L. I. (2012). Bayesian sample size determination using linex loss and linear cost. *Communications in Statistics-Theory and Methods* **41**, 223–240.

Islam, A. F. M. S. & Pettit, L. I. (2014). Bayesian sample size determination for the bounded linex loss function. *Journal of Statistical Computation and Simulation* **84**, 1644–1653.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994a). *Continuous univariate distributions*, 2 ed., vol. 1. New York: John Wiley & Sons.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994b). *Continuous univariate distributions*, 2 ed., vol. 2. New York: John Wiley & Sons.

Lindley, D. V. (1997). The choice of sample size. *Journal of the Royal Statistical Society: Series D (The Statistician)* **46**, 129–138.

Marbuah, G., Gren, I.-M. & McKie, B. (2014). Economics of harmful invasive species: a review. *Diversity* **6**, 500–523.

McCarthy, S. A., McPhearson, R. M., Guarino, A. & Gaines, J. (1992). Toxigenic *Vibrio cholerae* 01 and cargo ships entering Gulf of Mexico. *Lancet* **339**, 624–625.

Müller, P. & Parmigiani, G. (1995). Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association* **90**, 1322–1330.

Parmigiani, G. & Inoue, L. Y. T. (2009). *Decision theory: principles and approaches*. New York: John Wiley & Sons.

Pham-Gia, T. & Turkkan, N. (1992). Sample size determination in Bayesian analysis. *Journal of the Royal Statistical Society: Series D (The Statistician)* **41**, 389–392.

Rice, K. M., Lumley, T. & Szpiro, A. A. (2008). Trading bias for precision: decision theory for intervals and sets. `http://www.bepress.com/uwbiostat/paper336`. Working Paper 336, UW Biostatistics.

Sahu, S. K. & Smith, T. M. F. (2006). A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**, 235–253.

Schervish, M. (1995). *Theory of Statistics*. New York: Springer-Verlag.

Strayer, D. L. (2010). Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology* **55**, 152–174.

Table 1: Sample sizes n and estimated mean posterior credible lengths via simulation (within parentheses), both computed under a Poisson/gamma (1)-(2) model with $\lambda_0 = 10$, $\rho = 0.05$ via loss function 1.

Aliquot volume (w)	Aliquot cost (c)	Shape parameter (θ_0)				
		1.0	2.5	5.0	7.5	10.0
0.5	0.005	14 (4.47)	14 (4.57)	13 (4.70)	13 (4.66)	12 (4.67)
	0.010	9 (5.73)	9 (5.53)	8 (5.83)	8 (5.63)	8 (5.54)
1.0	0.005	11 (3.61)	11 (3.61)	11 (3.62)	11 (3.64)	11 (3.53)
	0.010	7 (4.55)	7 (4.63)	7 (4.47)	7 (4.46)	6 (4.66)

Table 2: Sample sizes n and estimated mean posterior credible lengths via simulation (between parenthesis), both computed under a negative binomial/Pearson Type VI (3)-(4) model with $\lambda_0 = 10(w/\phi)$, $\rho = 0.05$ via loss function 1.

Aliquot volume (w)	Aliquot cost (c)	ϕ	Shape parameter (θ_0)			
			11	25	50	75
0.5	0.005	1.0	9 (14.05)	7 (12.75)	5 (11.13)	4 (10.05)
		2.5	12 (8.09)	10 (7.99)	7 (7.61)	6 (6.65)
		5.0	16 (5.74)	14 (5.68)	11 (5.37)	9 (5.11)
		7.5	20 (4.80)	17 (4.76)	14 (4.63)	11 (4.41)
		10.0	23 (4.21)	20 (4.26)	16 (4.23)	14 (4.00)
	0.010	1.0	5 (16.95)	4 (15.05)	3 (12.14)	2 (10.58)
		2.5	7 (10.24)	6 (9.43)	4 (8.02)	3 (7.21)
		5.0	10 (7.10)	8 (6.99)	6 (6.20)	5 (5.54)
		7.5	12 (6.18)	10 (5.83)	8 (5.32)	6 (4.93)
		10.0	15 (5.22)	12 (5.21)	10 (4.76)	8 (4.47)
1.0	0.005	1.0	6 (17.13)	5 (15.60)	4 (13.93)	3 (12.66)
		2.5	8 (9.68)	7 (8.77)	5 (8.89)	5 (7.96)
		5.0	10 (6.66)	9 (6.51)	7 (6.20)	6 (5.90)
		7.5	12 (5.09)	11 (5.14)	9 (5.14)	8 (4.80)
		10.0	14 (4.59)	12 (4.62)	10 (4.48)	9 (4.28)
	0.010	1.0	4 (18.22)	3 (19.36)	2 (16.37)	2 (13.38)
		2.5	5 (11.30)	4 (10.94)	3 (10.33)	3 (8.66)
		5.0	6 (8.14)	5 (7.87)	4 (7.20)	4 (6.47)
		7.5	7 (6.44)	6 (6.44)	5 (6.10)	5 (5.37)
		10.0	8 (5.79)	7 (5.54)	6 (5.34)	5 (4.99)

Table 3: Sample sizes n and estimated mean posterior credible lengths via simulation (within parentheses), both computed under the Poisson/gamma (1)-(2) model with $\lambda_0 = 10$ via loss function 2.

Aliquot volume (w)	Aliquot cost (c)	γ	Shape parameter (θ_0)				
			1.0	2.5	5.0	7.5	10.0
0.5	0.005	1	94 (0.93)	94 (0.92)	92 (0.93)	90 (0.94)	89 (0.94)
		1/4	60 (2.30)	59 (2.32)	58 (2.33)	56 (2.35)	55 (2.37)
	0.010	1	60 (1.14)	59 (1.15)	58 (1.15)	56 (1.18)	55 (1.18)
		1/4	38 (2.89)	37 (2.90)	36 (2.94)	35 (2.95)	34 (3.00)
1.0	0.005	1	75 (0.73)	75 (0.73)	75 (0.72)	74 (0.73)	73 (0.74)
		1/4	48 (1.81)	48 (1.81)	47 (1.84)	46 (1.85)	46 (1.85)
	0.010	1	48 (0.91)	48 (0.91)	47 (0.92)	46 (0.93)	46 (0.92)
		1/4	30 (2.27)	30 (2.28)	30 (2.30)	29 (2.32)	29 (2.32)

Table 4: Sample sizes n and estimated mean posterior credible lengths via simulation (within parentheses), both computed under the negative binomial/Pearson Type VI (3)-(4) model with $\lambda_0 = 10(w/\phi)$ via loss function 2.

Aliquot volume (w)	Aliquot cost (c)	γ	ϕ	Shape parameter (θ_0) 11	25	50	75
0.5	0.005	1	1.0	205 (2.05)	193 (2.09)	175 (2.20)	158 (2.28)
			2.5	202 (2.08)	191 (2.09)	173 (2.20)	157 (2.27)
			5.0	201 (2.09)	190 (2.11)	172 (2.22)	156 (2.28)
			7.5	201 (2.10)	190 (2.11)	172 (2.21)	156 (2.27)
			10.0	201 (2.09)	189 (2.13)	172 (2.19)	156 (2.26)
		1/4	1.0	128 (5.09)	118 (5.32)	104 (5.59)	93 (5.75)
			2.5	126 (5.32)	117 (5.46)	103 (5.63)	92 (5.76)
			5.0	126 (5.31)	116 (5.42)	102 (5.64)	92 (5.77)
			7.5	125 (5.20)	116 (5.44)	102 (5.57)	91 (5.76)
			10.0	125 (5.22)	116 (5.37)	102 (5.58)	91 (5.76)
	0.010	1	1.0	129 (2.66)	119 (2.66)	105 (2.73)	93 (2.89)
			2.5	126 (2.62)	117 (2.68)	103 (2.81)	92 (2.90)
			5.0	125 (2.62)	116 (2.70)	102 (2.78)	92 (2.89)
			7.5	125 (2.61)	116 (2.69)	102 (2.80)	91 (2.87)
			10.0	125 (2.63)	116 (2.70)	102 (2.83)	91 (2.90)
		1/4	1.0	81 (6.48)	73 (6.84)	63 (6.87)	55 (7.18)
			2.5	79 (6.61)	71 (6.77)	61 (7.05)	54 (7.15)
			5.0	78 (6.60)	71 (6.79)	61 (6.95)	54 (7.15)
			7.5	78 (6.64)	71 (6.77)	61 (7.08)	54 (7.16)
			10.0	78 (6.62)	71 (6.82)	61 (6.94)	53 (7.20)
1.0	0.005	1	1.0	167 (1.61)	161 (1.66)	151 (1.71)	143 (1.72)
			2.5	164 (1.61)	154 (1.67)	149 (1.68)	141 (1.73)
			5.0	163 (1.64)	158 (1.65)	149 (1.70)	140 (1.73)
			7.5	163 (1.66)	158 (1.65)	148 (1.68)	140 (1.73)
			10.0	163 (1.64)	157 (1.66)	148 (1.70)	140 (1.73)
		1/4	1.0	106 (4.03)	101 (4.16)	98 (4.17)	87 (4.17)
			2.5	104 (4.11)	99 (4.24)	92 (4.26)	85 (4.39)
			5.0	103 (4.12)	98 (4.24)	91 (4.33)	84 (4.41)
			7.5	103 (4.13)	98 (4.19)	91 (4.28)	84 (4.38)
			10.0	103 (4.17)	98 (4.22)	91 (4.26)	84 (4.39)
	0.010	1	1.0	107 (2.08)	101 (2.08)	93 (2.17)	86 (2.24)
			2.5	104 (2.05)	99 (2.09)	91 (2.14)	85 (2.18)
			5.0	103 (2.07)	98 (2.08)	91 (2.15)	84 (2.18)
			7.5	103 (2.06)	98 (2.09)	91 (2.13)	84 (2.22)
			10.0	103 (2.06)	98 (2.11)	90 (2.17)	84 (2.19)
		1/4	1.0	67 (5.17)	63 (5.16)	57 (5.24)	52 (5.44)
			2.5	66 (5.07)	62 (5.36)	56 (5.36)	51 (5.51)
			5.0	65 (5.17)	61 (5.32)	55 (5.40)	51 (5.55)
			7.5	65 (5.20)	61 (5.27)	55 (5.42)	51 (5.47)
			10.0	65 (5.14)	61 (5.25)	55 (5.45)	51 (5.51)

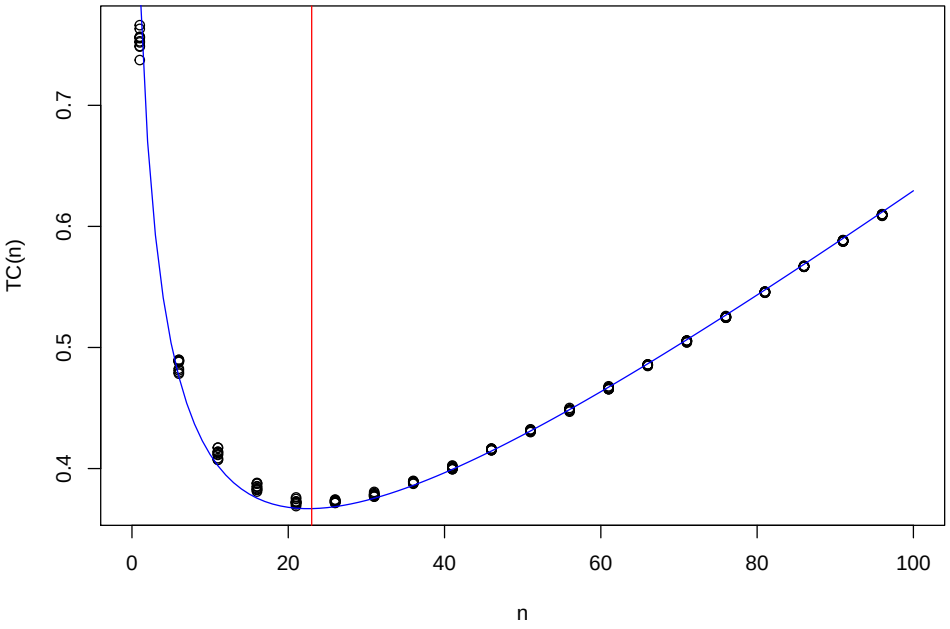

Figure 1: Curve fitting to the estimated Bayes risk as a function of n for the negative binomial/Pearson Type VI model with $\theta_0 = 11$, $\phi = 10$, $w = 0.5$, $c = 0.005$ and $\lambda_0 = 10(w/\phi)$; a vertical mark line is in the minimum $n = 23$

List of Figures

Supplementary Material

1 Poisson/gamma model

For simplicity of notation, we drop the argument $\boldsymbol{x_n}$ in the limits of the credible interval, $a(\boldsymbol{x}_n)$ and $b(\boldsymbol{x}_n)$, for the posterior distribution throughout the text

1.1 Posterior distribution and properties

Under the Poisson/gamma model each X_i follows marginally a negative binomial distribution with mean $w\lambda_0$ and parameter θ_0, *i.e.*, $X_i \sim \text{NB}(w\lambda_0, \theta_0)$. Furthermore, $S_n = \sum_{i=1}^n X_i \sim \text{NB}(nw\lambda_0, n\theta_0)$. The corresponding likelihood function is

$$\mathcal{L}(\lambda; \boldsymbol{x}_n) = \prod_{i=1}^n \frac{e^{-w\lambda}(w\lambda)^{x_i}}{x_i!} = \frac{e^{-nw\lambda}(w\lambda)^{s_n}}{\prod_{i=1}^n x_i!},$$

where $s_n = \sum_{i=1}^n x_i$ and $\boldsymbol{x}_n = (x_1, \ldots, x_n)$. If we consider a prior gamma distribution for λ, the posterior distribution is

$$\begin{aligned} f(\lambda|\boldsymbol{x}_n) &\propto \lambda^{s_n} e^{-nw\lambda} \times \lambda^{\theta_0 - 1} e^{-(\theta_0/\lambda_0)\lambda} \\ &= \lambda^{\theta_0 + s - 1} e^{-(nw + \theta_0/\lambda_0)\lambda}, \end{aligned} \tag{1}$$

which is a gamma distribution with parameters $\theta_0 + s_n$ and $nw + \theta_0/\lambda_0$. The corresponding mean and variance are, respectively,

$$\mathbb{E}\left[\lambda \middle| \boldsymbol{x}_n\right] = \frac{\theta_0 + s_n}{\theta_0/\lambda_0 + nw} \quad \text{and} \quad \text{Var}\left[\lambda \middle| \boldsymbol{x}_n\right] = \frac{\theta_0 + s_n}{(\theta_0/\lambda_0 + nw)^2}. \tag{2}$$

1.2 Obtaining the Bayes rule

1.2.1 Loss function 1

To obtain the decision d_n^* that minimizes the Bayes risk $r(f, d_n)$ is equivalent to obtaining the one that minimizes the posterior expected value of the loss function, namely $\mathbb{E}\left[L(\lambda, d_n)\middle|\boldsymbol{x_n}\right]$. For the loss function 1, we have

$$\begin{aligned}
\mathbb{E}\left[L(\lambda, d_n)\middle|\boldsymbol{x}_n\right] &= \rho\tau + \int_0^\infty (a-\lambda)^+ f(\lambda|\boldsymbol{x}_n)d\lambda + \int_0^\infty (\lambda - b)^+ f(\lambda|\boldsymbol{x}_n)d\lambda \\
&= \int_b^\infty \lambda f(\lambda|\boldsymbol{x}_n)d\lambda - \int_0^a \lambda f(\lambda|\boldsymbol{x}_n)d\lambda \\
&+ \frac{\rho(b-a)}{2} + a\int_0^a f(\lambda|\boldsymbol{x}_n)d\lambda - b\int_b^\infty f(\lambda|\boldsymbol{x}_n)d\lambda. \qquad (3)
\end{aligned}$$

Note that if a and b are the quantiles of probabilities $\rho/2$ and $1-\rho/2$ of the posterior distribution of λ, respectively, the sum of the three last terms in (3) are equal to zero, and the minimum has been reached. For more details see Rice et al. (2008). Then,

$$\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right] = \int_{b^*}^\infty \lambda f(\lambda|\boldsymbol{x}_n)d\lambda - \int_0^{a^*} \lambda f(\lambda|\boldsymbol{x}_n)d\lambda. \tag{4}$$

If we consider the Poisson/gamma model, the posterior distribution is a gamma distribution with parameters $\kappa = \theta_0 + s_n$ and $\psi = nw + \theta_0/\lambda_0$; we

have

$$\int_0^a \lambda f(\lambda|\boldsymbol{x}_n)d\lambda = \int_0^a \lambda \frac{\psi^\kappa}{\Gamma(\kappa)}\lambda^{\kappa-1}e^{-\psi\lambda}d\lambda = \frac{\kappa}{\psi}\int_0^a \frac{\psi^{\kappa+1}}{\Gamma(\kappa+1)}\lambda^\kappa e^{-\psi\lambda}d\lambda.$$

Note that the last integral is the cumulative probability until the point a of a gamma distribution with parameters $\kappa+1$ and ψ; similarly, we may obtain the other integral in $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$, and we have

$$\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right] = \frac{\kappa}{\psi}\left[\int_{b^*}^\infty \frac{\psi^{\kappa+1}}{\Gamma(\kappa+1)}\lambda^\kappa e^{-\psi\lambda}d\lambda - \int_0^{a^*} \frac{\psi^{\kappa+1}}{\Gamma(\kappa+1)}\lambda^\kappa e^{-\psi\lambda}d\lambda\right]. \tag{5}$$

1.2.2 Loss function 2

For the loss function 2, the posterior expected value is

$$\mathbb{E}\left[L(\lambda, d_n)\middle|\boldsymbol{x}_n\right] = \gamma\tau + \int_0^\infty \frac{(\lambda-m)^2}{\tau} f(\lambda|\boldsymbol{x}_n)d\lambda.$$

The minimum of the integral in m is reached when $m = \mathbb{E}\left[\lambda\middle|\boldsymbol{x}_n\right]$; then,

$$\mathbb{E}\left[L(\lambda, d_n)\middle|\boldsymbol{x}_n\right] = \gamma\tau + \frac{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]}{\tau}. \tag{6}$$

To minimize the posterior expected value in τ, consider $\tau = th(\gamma)\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]}$, for $t > 0$ and $h(\gamma)$ a positive function in γ (Rice et al., 2008), then

$$\mathbb{E}\left[L(\lambda, d_n)\middle|\boldsymbol{x}_n\right] = \sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]}\left[th(\gamma)\gamma + \frac{1}{th(\gamma)}\right].$$

Differentiating in t the expression in brackets we obtain the minimum when

$t = 1/[\gamma^{1/2}h(\gamma)]$, and replacing this value in τ we obtain

$$\tau = th(\gamma)\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]} = \frac{1}{\gamma^{1/2}h(\gamma)}f(\gamma)\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]} = \gamma^{-1/2}\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]}.$$

Another way to obtain the minimum is to differentiate (6) in τ directly, set the derivative equal to zero and solve it in τ. In this case, the Bayes rule is the interval $[a^*, b^*] = [m-\text{SD}_\gamma, m+\text{SD}_\gamma]$, where $(m, \text{SD}_\gamma) = (\mathbb{E}\left[\lambda\middle|\boldsymbol{x}_n\right], \gamma^{-1/2}\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]})$. Then, the posterior expected value is

$$\begin{aligned}\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right] &= \gamma\gamma^{-1/2}\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]} + \frac{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]}{\gamma^{-1/2}\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]}} \\ &= 2\gamma^{1/2}\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right]}. \qquad (7)\end{aligned}$$

Given the posterior of the Poisson/gamma model and (2), we may easily compute $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$.

1.3 Algorithm to obtain n

The choice of the set in which n varies is arbitrary, in this way we consider $n = 1, 5, 10, \ldots, 95, 100$, and for each value of n in this set the estimate of TC(n) is computed 10 times (also arbitrary), *i.e.*, we obtain 10 estimates for TC(n). A possible algorithm to obtain the minimum sample size n is described as follows for loss functions 1 or 2.

Step 1. Set values for λ_0, θ_0, w, c, and ρ (loss function 1) or γ (loss function 2), in addition choose a set in which n may vary;

Step 2. For each, n draw a sample of size M (*e.g.*, $M = 1000$) of s_n from a

negative binomial distribution with mean $nw\lambda_0$ and shape parameter $n\theta_0$, compute the respective interval limits a^* e b^* (in loss function 1 case), then compute the respective $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$ using (5) or (7), and finally compute the average of the M posterior expected values. This value is the estimate of the minimized Bayes risk for the respective n;

Step 3. For each estimated Bayes risk, add the cost cn, keep these values;

Step 4. With the values obtained in Step 3 and the respective values of n, fit a regression model as stated in equation (7) of the article and compute the minimum n using expression (8) of the article.

2 Negative binomial/Pearson Type VI model

2.1 Posterior distribution and properties

For the negative binomial model., the corresponding likelihood function is

$$\begin{aligned}\mathcal{L}(\lambda;\boldsymbol{x}_n) &= \prod_{i=1}^{n}\frac{\Gamma(\phi+x_i)}{\Gamma(x_i+1)\Gamma(\phi)}\left(\frac{w\lambda}{w\lambda+\phi}\right)^{x_i}\left(\frac{\phi}{w\lambda+\phi}\right)^{\phi}\\ &= \left[\prod_{i=1}^{n}\frac{\Gamma(\phi+x_i)}{\Gamma(x_i+1)\Gamma(\phi)}\right]\left(\frac{w}{\phi}\lambda\right)^{s_n}\left(1+\frac{w}{\phi}\lambda\right)^{-s_n-n\phi},\end{aligned}$$

where $s_n=\sum_{i=1}^{n}x_i$ and $\boldsymbol{x}_n=(x_1,\ldots,x_n)$. If we consider a Pearson Type VI prior distribution for λ, the posterior distribution is

$$\begin{aligned}f(\lambda|\boldsymbol{x}_n) &\propto \left(\frac{w}{\phi}\lambda\right)^{s_n}\left(1+\frac{w}{\phi}\lambda\right)^{-s_n-n\phi}\times\left(\frac{w}{\phi}\lambda\right)^{\theta_0-1}\left(1+\frac{w}{\phi}\lambda\right)^{-\theta_0-(\theta_0/\lambda_0+1)}\\ &= \left(\frac{w}{\phi}\lambda\right)^{\theta_0+s_n-1}\left(1+\frac{w}{\phi}\lambda\right)^{-(\theta_0+s_n)-(\theta_0/\lambda_0+n\phi+1)},\end{aligned}$$

which corresponds to a Pearson Type VI distribution with location and scale parameters equal to 0 and ϕ/w, respectively, and shape parameters $\theta_0 + s_n$ and $\theta_0/\lambda_0 + n\phi + 1$. The corresponding mean and variance is

$$\mathbb{E}\left[\lambda\middle|\boldsymbol{x}_n\right] = \frac{\phi}{w}\frac{\theta_0 + s_n}{\theta_0/\lambda_0 + n\phi}, \tag{8}$$

and the variance is

$$\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right] = \left(\frac{\phi}{w}\right)^2 \frac{\theta_0 + s_n}{(\theta_0/\lambda_0 + n\phi)^2}\left(\frac{\mathbb{E}\left[\lambda\middle|\boldsymbol{x}_n\right] + 1}{1-q}\right), \tag{9}$$

where $q = (\theta_0/\lambda_0 + n\phi)^{-1}$.

2.2 Obtaining the Bayes rule

2.2.1 Loss function 1

Consider the loss function 1 and according to (4), we have

$$\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right] = \int_{b^*}^{\infty} \lambda f(\lambda|\boldsymbol{x}_n)d\lambda - \int_0^{a^*} \lambda(\lambda|\boldsymbol{x}_n)d\lambda,$$

where a^* and b^* are the quantiles of probabilities $\rho/2$ and $1-\rho/2$ of the posterior distribution, respectively. Given the negative binomial/Pearson Type VI model, the posterior distribution is a PVI$(0, \phi/w, \kappa, \psi)$, where $\kappa = \theta_0 + s_n$ and $\psi = \theta_0/\lambda_0 + n\phi + 1$; then,

$$\begin{aligned}\int_0^a \lambda f(\lambda|\boldsymbol{x}_n)d\lambda &= \int_0^a \lambda\left(\frac{w}{\phi}\lambda\right)^{\kappa-1}\left(1+\frac{w}{\phi}\lambda\right)^{-\kappa-\psi} d\lambda \\ &= \frac{\phi}{w}\int_0^a \left(\frac{w}{\phi}\lambda\right)^{\kappa}\left(1+\frac{w}{\phi}\lambda\right)^{-(\kappa+1)-(\psi-1)} d\lambda.\end{aligned}$$

Note that the last integral is the cumulative probability until the point a of a distribution PVI$(0, \phi/w, \kappa+1, \psi-1)$. Similarly, we may obtain the other integral in $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$, and we have

$$\begin{aligned}\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right] &= \frac{\phi}{w}\left[\int_{b^*}^{\infty}\left(\frac{w}{\phi}\lambda\right)^{\kappa}\left(1+\frac{w}{\phi}\lambda\right)^{-(\kappa+1)-(\psi-1)}d\lambda\right] \\ &- \frac{\phi}{w}\left[\int_{0}^{a^*}\left(\frac{w}{\phi}\lambda\right)^{\kappa}\left(1+\frac{w}{\phi}\lambda\right)^{-(\kappa+1)-(\psi-1)}d\lambda\right] \quad (10)\end{aligned}$$

2.2.2 Loss function 2

If we consider the loss function 2 and according to (7), we have

$$\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right] = 2\gamma^{1/2}\sqrt{\text{Var}\left[\lambda\middle|\boldsymbol{x}_n\right)]}. \quad (11)$$

Given the posterior distribution of the negative binomial/Pearson Type VI model and (9), we may easily compute $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$.

2.3 Algorithm to obtain n

A possible algorithm to obtain the minimum sample size n is described as follows for loss functions 1 or 2.

Step 1. Set values for λ_0, θ_0, ϕ, w, c, and ρ (loss function 1) or γ (loss function 2), in addition choose a set in which n may vary;

Step 2. For each n, draw a sample of size M (*e.g.*, $M = 1000$) of s_n (s_n may be drawn as follows: drawn a sample of size n of λ from the prior distribution PVI$(0, \phi/w, \theta_0, \theta_0/\lambda_0 + 1)$, with these n values drawn a

sample of size n of $X_i, i = 1, 2, \ldots, n$ from a negative binomial distribution with mean $w\lambda$ and parameter ϕ, then sum the simulated values of X_i), compute the respective interval limits a^* and b^* (in loss function 1 case), then compute the respective $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$ using (10) or (11), and finally compute the average of the M posterior expected values, keep these values;

Step 3. For each estimated Bayes risk, add the respective cost cn, keep these values;

Step 4. With the values obtained in Step 3 and the respective values of n, fit a regression model as stated in equation (7) of the article and compute the minimum n using expression (8) of the article.

References

Rice, K. M., Lumley, T., and Szpiro, A. A. (2008). Trading bias for precision: decision theory for intervals and sets. `http://www.bepress.com/uwbiostat/paper336`. Working Paper 336, UW Biostatistics.