Supplemental material for "Verifying compliance with ballast water standards: a decision-theoretic approach"

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

June 2021

The material contained herein is supplementary to the article named in the title and published in SORT-Statistics and Operations Research Transactions Volume 45(1).

[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.

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

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

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

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

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

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.*, $X_i \sim \mathrm{NB}(w\lambda_0, \theta_0)$. Furthermore, $S_n = \sum_{i=1}^{n} X_i \sim \mathrm{NB}(nw\lambda_0, n\theta_0)$.

1.2 Obtaining the Bayes rule

Obtaining the decision d_n^* that minimizes the Bayes risk $r(h, d_n)$ is equivalent to obtaining the one that minimizes the posterior expected loss, namely $\mathbb{E}\left[L(\lambda, d_n)|\boldsymbol{x_n}\right]$. In our case, the posterior expected loss is

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

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

$$\mathbb{E}\left[L(\lambda, d_n)|\boldsymbol{x}_n\right] = \gamma\tau + \frac{\mathrm{Var}\left[\lambda|\boldsymbol{x}_n\right]}{\tau}. \tag{3}$$

To minimize the posterior expected loss in τ, consider

$$\tau = t g(\gamma) \sqrt{\mathrm{Var}\left[\lambda|\boldsymbol{x}_n\right]}, \tag{4}$$

for $t > 0$ and $g(\gamma)$ a positive function in γ (Rice et al., 2008), then

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

Differentiating the expression in brackets with respect to t we obtain the minimum when $t = 1/[\gamma^{1/2} g(\gamma)]$, and replacing this value in (4) we obtain

$$\tau = t g(\gamma) \sqrt{\mathrm{Var}\left[\lambda|\boldsymbol{x}_n\right]} = \frac{1}{\gamma^{1/2} g(\gamma)} g(\gamma) \sqrt{\mathrm{Var}\left[\lambda|\boldsymbol{x}_n\right]} = \gamma^{-1/2} \sqrt{\mathrm{Var}\left[\lambda|\boldsymbol{x}_n\right]}.$$

Another way to obtain the minimum is to differentiate (3) with respect to τ directly, set the derivative equal to zero and solve it in τ. Thus, the Bayes rule corresponds to the quantities which define the interval $[a^*, b^*] = [m^* - \mathrm{SV}_\gamma, m^* + \mathrm{SV}_\gamma]$, where $m^* = \mathbb{E}\left[\lambda \middle| \boldsymbol{x}_n\right]$ and $\mathrm{SV}_\gamma = \gamma^{-1/2}(\mathrm{Var}\left[\lambda \middle| \boldsymbol{x}_n\right])^{1/2}$. Then, the minimized posterior expected loss is

$$\begin{aligned}
\mathbb{E}\left[L(\lambda, d_n^*) \middle| \boldsymbol{x}_n\right] &= \gamma\gamma^{-1/2}(\mathrm{Var}\left[\lambda \middle| \boldsymbol{x}_n\right])^{1/2} + \frac{\mathrm{Var}\left[\lambda \middle| \boldsymbol{x}_n\right]}{\gamma^{-1/2}(\mathrm{Var}\left[\lambda \middle| \boldsymbol{x}_n\right])^{1/2}} \\
&= 2(\gamma \mathrm{Var}\left[\lambda \middle| \boldsymbol{x}_n\right])^{1/2}. \qquad (5)
\end{aligned}$$

Given the posterior distribution of the Poisson/gamma model and (2), we obtain a closed expression for the minimized posterior expected loss. Also, since the posterior distribution is gamma, we may easily compute the Bayesian coverage probability for a given interval.

1.3 Algorithm to obtain n

The choice of the set in which n varies is arbitrary. In general we consider $n = 1, 6, \ldots, 491, 496$, and for each value of n in this set the estimate of $TC(n)$ is computed 10 times (also arbitrarily), *i.e.*, we obtain 10 estimates for $TC(n)$. A possible algorithm to obtain the optimal sample size n is

Step 1. Set values for λ_0, θ_0, w, c and γ, 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 $n\theta_0$, then compute the respective $\mathbb{E}\left[L(\lambda, d_n^*) \middle| \boldsymbol{x}_n\right]$ using (5), and finally compute the average of the M minimized posterior expected losses. This value is the estimate of the minimized Bayes risk for the respective n.
Step 3. For each estimated Bayes risk, add the respective cost cn and 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 optimal n using expression (8).

2 Negative binomial/gamma 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 gamma prior distribution for λ, the posterior distribution is

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

which do not correspond to a known distribution. To bypass this problem, we use the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) to draw samples from the posterior distribution of λ. Specifically, we use the random walk Metropolis-Hastings algorithm through the function `rwmetrop` of the R package `LearnBayes` (R Core Team, 2016). In each case, we consider a burn-in of 100 iterations and a thinning of 10 or 15 with a final sample of size 900. We inspect trace and autocorrelation plots for a fixed $n = 5$ when computing the optimal sample size and we expect the same or better behavior for $n > 5$. The trace plots showed a random behavior around a value and in the autocorrelation plots the autocorrelations for almost every lag are zero (see Figure 1 for an example).

Given the posterior distribution of the negative binomial/gamma model and (5), we may easily compute an estimate for $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$ using a sample obtained via the Metropolis-Hastings algorithm. Also, using this sample we may obtain an estimate for the Bayesian coverage probability associated to an interval as the proportion of values of the sample within the given interval.

2.2 Algorithm to obtain n

The choice of the set in which n varies is arbitrary. In general we consider $n = 1, 5, 10, \ldots, 590, 600$ for $c = 0.001$ and $n = 1, 3, 5, \ldots, 197, 199$ for $c = 0.01$. For each value of n in the respective set the estimate of $TC(n)$ is computed 3 times. In some cases where the fitted curve is not visually satisfactory we consider a refined and/or smaller set for n in order to compute the optimal n. A possible algorithm to obtain the optimal sample size n is described as follows.

Step 1. Set values for λ_0, θ_0, ϕ, w, c and γ, 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 $\boldsymbol{x}_n$. The sample $\boldsymbol{x}_n$ may be drawn as follows: draw one value of λ from the prior distribution, given this value draw a sample of size n of $X_i, i = 1, 2, \ldots, n$ from a negative binomial distribution with mean $w\lambda$ and shape parameter ϕ, then compute an estimate for $\mathbb{E}\left[L(\lambda, d_n^*)\middle|\boldsymbol{x}_n\right]$, and finally compute the average of the M minimized posterior expected loss estimates and keep these values.

Step 3. For each estimated Bayes risk obtained in Step 2, add the respective cost cn and 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 optimal n using expression (8).

In Step 2 we only inspect the trace and autocorrelation plots for $n = 5$. We expect that the behavior is the same or better for $n > 5$. In the cases where the fitted curve is not visually satisfactory and the change of the set in which n varies did not improve the fitting, we consider an estimate for the total cost as the mean of the 3 estimates obtained for a fixed n, then obtain the optimal sample size computationally verifying the n which corresponds to the minimum total cost. The acceptance rates for the Metropolis-Hastings algorithm used to obtain the optimal sample sizes ranged between 31% and 71%.

We implemented the above algorithms in R (R Core Team, 2016). The code is available in `https://github.com/eliardocosta/ssdet`, or may be obtained from the first author via e-mail.

3 Figures

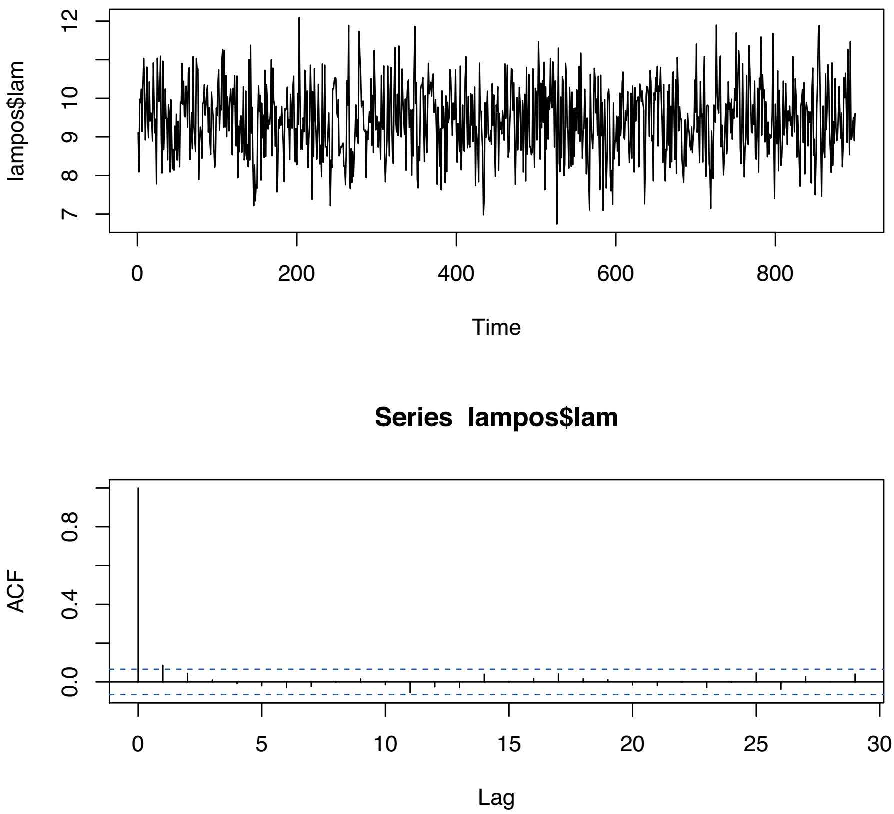

Figure 1: *Trace and autocorrelation plots for a sample drawn from the posterior distribution of λ under the negative binomial/gamma model.*

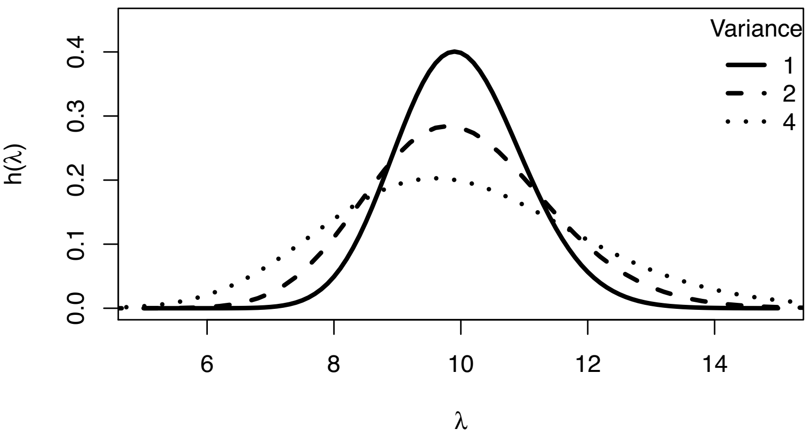

Figure 2: *Prior gamma distribution for different variance.*

References

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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.