Approximate Bayesian inference for the Rosenblatt distribution

Laura Rifo*
Unicamp, Campinas, Brazil - lramos@ime.unicamp.br

Plínio Andrade
Usp, São Paulo, Brazil - plinio@ime.usp.br

Abstract

The Rosenblatt distribution is a one-parameter family arising from a non-central limit theorem for long-range dependent random variables. The usual parametrization includes the normal distribution, the standardized chi-squared distribution and weighted sums of chi-squared variates. Although its density has not a manageable analytical form, its moments, cumulants and distribution function have been recently numerically obtained. We apply a Bayesian likelihood-free methodology to make inferences for that family, comparing the performance of diverse statistics derived from the empirical distribution. The estimates obtained by our methodology are remarkably precise.

Keywords: approximate Bayesian computation; Hellinger distance.

1. Introduction

The Rosenblatt distribution was first addressed as a non-trivial example of a non-Gaussian distribution appearing as a limit of sums of dependent variates Rosenblatt (1961). That example is constructed considering standard normal variates Y_k $(k = 1, 2, \ldots)$, with covariance structure $E(Y_0\ Y_k) \sim k^{-D}$, as $k \to \infty$, for some $0 < D < 1/2$. Then, given the normalizing constant $\sigma = \{(1-2D)(1-D)/2\}^{1/2}$, the sequence

$$Z_{D,n} = \frac{\sigma}{n^{1-D}} \sum_{k=1}^{n} (Y_k^2 - 1) \quad (n = 1, 2, \ldots)$$

converges to Z_D having a non-normal distribution, named Rosenblatt distribution after Taqqu (1975). If we let $0 \le D \le 1/2$, that parametrization includes the standard normal distribution, defining $Z_{1/2}$ as the limit when $D \to 1/2$, and a chi-squared distribution rescaled to have zero mean and unit variance, when $D = 0$. This distribution is related to the Rosenblatt process, as the distribution of the process at time $t = 1$, whose properties are well described in Taqqu (2011).

The distribution of Z_D can be given in terms of its characteristic function

$$\phi(t) = \exp\left(\frac{1}{2}\sum_{k=2}^{\infty}(2it\sigma)^k \frac{c_k}{k}\right),$$

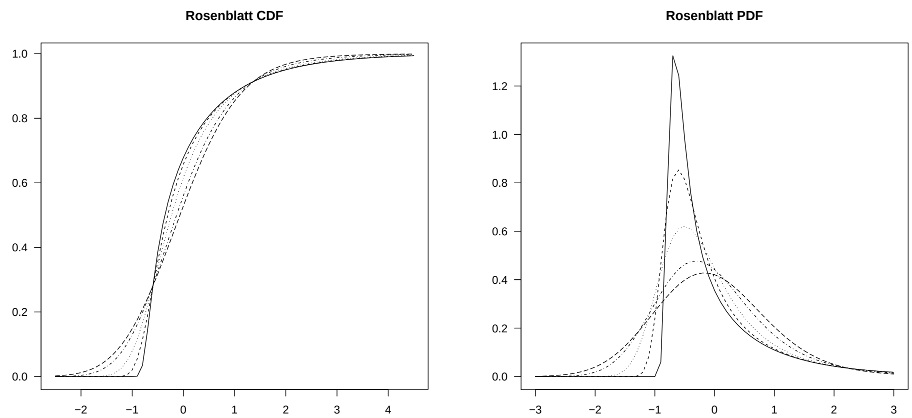
or as a weighted sum of standardized chi-squared variates

$$Z_D = \sum_{n=1}^{\infty} \lambda_n(\epsilon_n^2 - 1), \tag{1}$$

where the ϵ_n $(n = 1, 2, \ldots)$ are independent identically N(0,1) distributed, $\sum_{n=1}^{\infty} \lambda_n^k = \sigma^k\, c_k \quad (k = 2, 3, \ldots)$, and

$$c_k = \int_{x\in[0,1]^k} |x_k - x_1|^{-D} \prod_{j=2}^{k} |x_{j-1} - x_j|^{-D} dx\,.$$

The latter characterization provides an useful way to simulate random samples from Z_D, utilized in Veillette, M.S. & Taqqu, M.S. (2013) to derive numerical evaluations of the distribution function, cumulants and



(a) (b)

Figure 1: Plots of the (a) distribution function, and (b) density function of the Rosenblatt distribution for some values of D: 0·1 (solid), 0·2 (small dashes), 0·3 (dots, 0·4 (dot-dash), 0·45 (dashes)

moments. Figure 1 shows the plots of its density function and of its distribution function, for some values of D.

In this work, we propose a Bayesian likelihood-free approach to obtain an approximate sample from the posterior distribution for the parameter D, given an uniform prior knowledge on the interval $(0, 0 \cdot 5)$. As far as we know, there is no alternative method to estimate such parameter.

The results are simulation based and show a very efficient performance of this methodology, based on the approximate Bayesian computation developed in Grelaud et al. (2009). That algorithm strongly depends on the choice of a suitable statistics, as better as more informative in some sense. Among the available ones for our problem, we compare the estimates given by some discrepancies between distributions, namely, Kolmogorov-Smirnov, Cramér-von Mises, Kullback-Leibler and Hellinger.

2. Approximating a posterior distribution

Let us briefly recall the context of an estimation problem to set notation. We have an observation $X = (X_1, \ldots, X_n)$ in a sample space $\mathcal{X}$, modeled by an indexed family of probability distributions $\mathcal{P} = \{F(\cdot \mid \theta) : \theta \in \Theta\}$. Let us denote by $f(\cdot \mid \theta)$ the density function associated to $F(\cdot \mid \theta)$.

In the Bayesian framework, we take into account the previous knowledge about $\mathcal{P}$, expressed by a prior distribution π on the parameter space Θ. Also we assume that all the information given by an observed sample $x \in \mathcal{X}$ is contained in the likelihood function, $f(x \mid \theta)$. After observing x we can then upgrade our knowledge about θ by the posterior distribution

$$\pi(\theta \mid x) \propto \pi(\theta) f(x \mid \theta)\,, \tag{2}$$

as a compromise between both sources of information.

Clearly, this approach is strongly dependent on an analytical representation for the likelihood function. The approximate Bayesian computation, the so-called ABC algorithm, is a simulation-based methodology to obtain an approximate sample from a posterior distribution when the likelihood is not available or handled, supposing we are able to generate observations from X.

Its main idea is grounded on the comparison between the values of an informative statistic computed from the data and from simulated observations generated for different values of the parameter. It was presented in Pritchard et al. (1999) and well developed further, Sousa et al (2009), Toni, T. & Stumpf, M.P.H. (2010), Robert et al. (2011).

Given a discrepancy measure ρ in $\mathcal{X}$, for an observed x, the simplest ABC algorithm is a repetition of the following steps:

Step 1. generate θ from the prior π;

Step 2. generate $y = (y_1, \ldots, y_n)$ from $F(\cdot \mid \theta)$;

Step 3. compute $\rho(x, y)$;

Step 4. accept θ if $\rho(x, y) \leq \epsilon$, otherwise reject it.

That algorithm samples from the joint density,

$$\pi_{\epsilon,\rho}(\theta, y \mid x) = \frac{\pi(\theta) f(y \mid \theta) \mathbf{I}_{A_{\epsilon,x}}(y)}{\int_{A_{\epsilon,x} \times \Theta} \pi(\theta) f(y \mid \theta) dy d\theta},$$

where $A_{\epsilon,x} = \{y \in \mathcal{X}; \rho(y, x) \leq \epsilon\}$, and $\mathbf{I}_E$ is the indicator function of the event E.
In general, the discrepancy ρ is defined as a discrepancy or distance between $T(x)$ and $T(y)$, where T is a suitable summary statistic. The main idea is that this statistic coupled with a small value of ϵ should provide a good approximation of the posterior distribution for θ. The accuracy of that approximation will depend on the choice of the statistic T and of the value of ϵ, see Grelaud et al. (2009).

3. Results

As pointed out in the Introduction, we cannot handle directly with the likelihood associated to the Rosenblatt distribution to obtain likelihood based estimates nor an exact posterior distribution (2) for D.
In order to approximate a posterior distribution for the memory parameter D of the Rosenblatt distribution, we use the ABC algorithm measuring the discrepancy between x and y by means of diverse known discrepancies between the empirical distributions generated by x and y.
More precisely, let us consider the Kolmogorov-Smirnov distance, $d_{KS}(F_1, F_2) = \sup_x |F_1(x) - F_2(x)|$, the Cramér-von Mises discrepancy,

$$d_{CM}(F_1, F_2) = \int (F_1(x) - F_2(x))^2 dF_1(x)\,,$$

the Kullback-Leibler discrepancy,

$$d_{KL}(F_1, F_2) = \int \ln\left(\frac{f_1(x)}{f_2(x)}\right) f_1(x) dx,$$

and the Hellinger distance,

$$H^2(F_1, F_2) = 1 - \int \sqrt{f_1(x) f_2(x)} dx,$$

where f_i is the density associated to $F_i (i = 1, 2)$.
The simulations were carried on for four given values of $D = 0 \cdot 1, 0 \cdot 2, 0 \cdot 3, 0 \cdot 4$. For each value, we generate a sample from Z_D with size $n = 1000$, using the eigenvalue expansion of the Rosenblatt distribution as in Veillette, M.S. & Taqqu, M.S. (2013). That sample determines a posterior distribution for D from each of the proposed statistics. The Figure 2 shows the approximated posterior distribution for D from a typical realization of Z_D, for some of the values of D and statistics used.
We repeat the procedure described one thousand times and compute the mean and the standard deviation of the posterior mean in order to compare the performance of the four statistics statistics. The results are shown in Table 1.
The tabulated results show that the Hellinger distance has apparently a better performance among all, in the sense of the both bias and variation around the nominal value. They could be considered all together by using a modified algorithm called ABC of minimum entropy, in which we select the posterior sample with minimum entropy. However, regardless the statistic selected, the posterior distributions obtained in all the simulations performed are quite accurate.

5. Conclusions

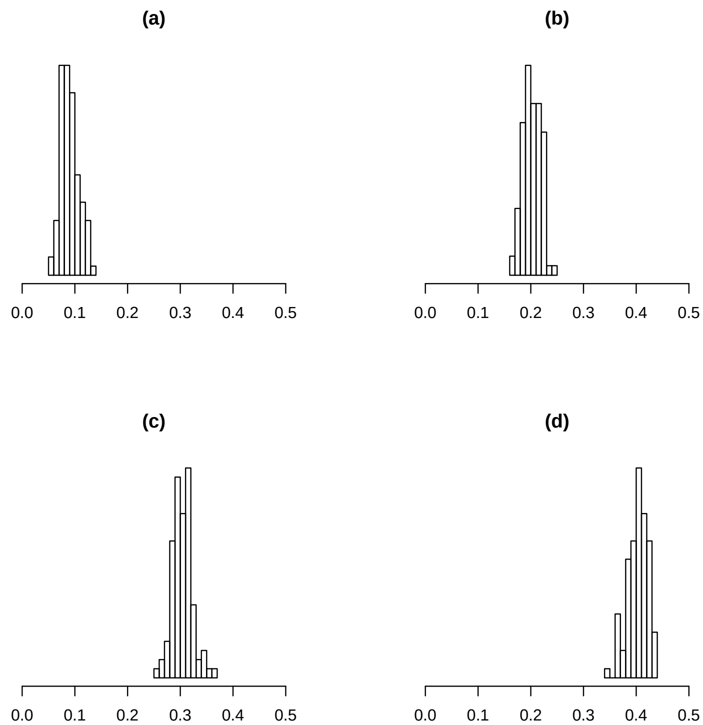


Figure 2: Posterior distribution for D given by the Hellinger distance when: (a) $D = 0{\cdot}1$, (b) $D = 0{\cdot}2$, (c) $D = 0{\cdot}3$, (d) $D = 0{\cdot}4$

Table 1: Sample mean and sample standard deviation of the approximated posterior mean for D, for each proposed statistic.

D	Kolmogorov-Smirnov	Cramér-von Mises	Kullback-Leibler	Hellinger
0.1	0.1016 (0.0065)	0.1008 (0.0068)	0.1062 (0.0046)	0.1001 (0.0054)
0.2	0.2027 (0.0109)	0.2002 (0.0100)	0.2091 (0.0094)	0.2006 (0.0066)
0.3	0.3110 (0.0209)	0.3070 (0.0186)	0.3259 (0.0105)	0.3027 (0.0130)
0.4	0.3952 (0.0165)	0.3935 (0.0125)	0.3708 (0.0111)	0.3958 (0.0110)

For the Rosenblatt distribution there is no likelihood-based inference method in the existing literature. In this work we obtained a remarkably precise posterior distribution for that family, based on the approximate Bayesian computation algorithm. In particular, the statistic defined by the Hellinger distance is the most informative one, among usual density discrepancies, to obtain that fine performance. Finally, although beyond the scope of this paper, it is worth to say that this family of distributions can be used to model one-dimensional distributions for long-memory stationary processes, as done in Andrade & Rifo (2015).

Acknowledgments

The authors are very grateful to Professors Murad S. Taqqu and Mark S. Veillette for providing the MATLAB code presented in Veillette, M.S. & Taqqu, M.S. (2013), used in this work.

This article was produced as part of the activities of FAPESP Center for Neuromathematics (grant N.2013 / 07699-0, São Paulo Research Foundation). The second author is a PhD student with CNPq grant at the University of São Paulo.

References

Andrade, P. & Rifo, L. (2015) Long-range dependence and approximate Bayesian computation. Forthcoming in *Comm. Statist. Simulation Comput.*

Grelaud, A., Robert, C. P., Marin, J. M., Rodolphe, F. & Taly, J. F. (2009) ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis*, **4**(2) 317–35.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, **16**, 1791–8.

Robert, C. P., Cornuet, J-M., Marin, J-M. & Pillai, N. S. (2011) Lack of confidence in approximate Bayesian computation model choice. *Proc. Natl. Acad. Sci. USA*, **108** (37) 15112-7.

Rosenblatt, M. (1961) Independence and dependence. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, **II**, 431–443.

Sousa, V. C., Fritz, M., Beaumont, M. A. & Chikhi, L. (2009) Approximate Bayesian computation without summary statistics: the case of admixture. *Genetics*, **181** (4) 1507–19.

Taqqu, M. S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **31**, 287–302.

Taqqu, M. S. (2011) The Rosenblatt process. *Selected works of Murray Rosenblatt*, 29–45.

Toni, T. & Stumpf, M. P. H. (2010) Simulation-based model selection for dynamical systems and population biology. *Bioinformatics*, **26**, 104–10.

Veillette, M. S. & Taqqu, M. S. (2013) Properties and numerical evaluation of the Rosenblatt distribution. *Bernoulli*, **19** (3) 982–1005.