

Simulation of the Multivariate Generalized Hyperbolic Distribution using Adaptive Importance Sampling

Marco Bee and Roberto Benedetti

Abstract In this paper we use Adaptive Importance Sampling for simulating the multivariate Generalized Hyperbolic Distribution and computing tail probabilities. Adaptive Importance Sampling is an extension of classical Importance Sampling that updates sequentially the instrumental density at each iteration. Under the only condition that the density is known in closed form, the method can be used for sampling multivariate distributions and estimate quantities of interest. Some simulations experiments and a real-data application show that, for the problem at hand, the method has an excellent performance, which makes this technique an appealing alternative to more traditional sampling algorithms, whose implementation for the simulation of the Generalized Hyperbolic Distribution is not straightforward.

Key words: Statistical computing, Statistical analysis of financial data, Importance sampling

1 Introduction

The Family of Generalized Hyperbolic Distributions (GHD), first introduced by Barndorff-Nielsen (1977), has recently been the object of considerable interest. From the theoretical point of view, the main reason is that each GHD is the marginal distribution of a Lévy process; in other words, the value of the increment of the process over a fixed time interval has that distribution. This result is particularly relevant for financial applications, because Lévy processes are often used as continuous-time models of asset prices. On the empirical side, the interest is mostly related to the ability at modelling fat-tailed distributions. Maximum likelihood estimation of the parameters can be performed by means of the EM algorithm (Protasov, 2004).

Marco Bee, Department of Economics, University of Trento, via Inama 5, e-mail: marco.bee@unitn.it · Roberto Benedetti, Department of Business, Statistical, Technological and Environmental Sciences, University 'G. d'Annunzio' of Chieti-Pescara, e-mail: benedett@unich.it

Generalized Hyperbolic Distributions can be represented as elliptically symmetric variance-mean normal mixtures: if x is a random vector with the p -variate GHD, denoted as $x \sim \text{GHD}_p(\lambda, \gamma, \beta, \mu, \delta, \Delta)$, the conditional distribution of $x|z$ is $N_p(\mu + z\beta\Delta, z\Delta)$, where $z \sim \text{GIG}(\lambda, \delta, \gamma)$ and GIG stands for Generalized Inverse Gaussian. According to this representation, the standard way of simulating the p -variate GHD is based on the following two steps: (i) simulate an observation z from the GIG distribution; (ii) simulate a $N_p(\mu + z\beta\Delta, z\Delta)$ variate. However, simulating the GIG is not straightforward (see Dagpunar, 1989), so that this method is rather cumbersome.

In this paper we use a recent extension of Importance Sampling, called Adaptive Importance Sampling (IS) for simulating the multivariate GHD. Adaptive IS was first introduced by Cappé *et al.* (2004). Cappé *et al.* (2008) extended it to general mixture classes, and Wraith *et al.* (2009) propose an interesting application. The method is an extension of standard IS because, at each iteration, a sample is simulated from the instrumental density and used to improve the IS density itself. So, ‘adaptivity’ means that the instrumental density sequentially provides a progressively better approximation of the target density. Basically, the only requirement is that the density can be written in closed form.

The paper is organized as follows. Section 2 introduces the statistical methods used in the paper, namely standard Importance Sampling and Adaptive Importance Sampling, and reviews the GHD. Section 3 presents the results of the simulation experiments as well as a real-data application. Finally, Section 4 concludes.

2 Methodology

Importance Sampling (see Casella and Robert, 2004, Sect. 3.3 for a review) is a variance reduction technique. Instead of simulating observations from the distribution of interest f as in crude Monte Carlo (MC), IS samples an instrumental distribution g that assigns “more weight” to the event of interest. The features of g are crucial for the properties of the estimator, but unfortunately there is no general, readily available procedure for finding the optimal instrumental density.

2.1 Adaptive Importance Sampling

Adaptive Importance Sampling (or Population Monte Carlo) is a sampling algorithm first proposed in a Bayesian setup by Cappé *et al.* (2004). The idea is that, instead of using a fixed IS density g , one can use a sequence of importance densities $g^{(t)}$ ($t = 1, \dots, T$), aimed at approximating the target density f . The general formulation of the algorithm is as follows:

1. Simulate a first sample $x_n = (x_1, \dots, x_n)' \sim g^{(1)}$ by means of standard IS; compute the IS weights $w_j^{(1)} = f(x_j^{(1)})/g^{(1)}(x_j^{(1)})$ ($j = 1, \dots, n$).
2. Use this sample to approximate the moments of f and construct the updated importance function $g^{(2)}$.
3. Measure the goodness of the approximation by means of the relative entropy (or Kullback-Leibler divergence; see Rubinstein and Kroese, 2004, Sect. 3.3) from the target:

$$K(f||g^{(2)}) = \int \log \left(\frac{f(x)}{g^{(2)}(x)} \right) f(x) dx. \quad (1)$$

4. ‘Adjust’ $g^{(2)}$ so that $K(f||g^{(2)}) \leq K(f||g^{(1)})$. Formally, compute $\min_{\theta} K(f||g^{(2)})$, where θ is the vector of parameters of $g^{(2)}$.
5. Repeat the preceding steps until some convergence criterion is met.

The first important issue is the choice of the functional form of g , which should be flexible enough to allow for a close match of the target and a reasonably easy minimization of (1). According to Cappé *et al.* (2004, 2008), a convenient choice is a finite mixture of normal distributions, namely

$$g^{(t)}(x) = g(x; \pi^{(t)}, \theta^{(t)}) = \sum_{d=1}^D \pi_d^{(t)} \phi_d(x; \theta_d^{(t)}), \quad (2)$$

where $\pi^{(t)} = (\pi_1^{(t)}, \dots, \pi_D^{(t)})'$ is the vector of weights of the D mixture components and $\theta^{(t)}$ is a $D \times 2$ matrix whose i -th row contains the parameters of the i -th normal density, i.e. $\theta_i^{(t)} = (\mu_i^{(t)}, \sigma_i^{(t)2})$. If the target density is p -variate ($p \geq 2$), (2) is a p -variate normal mixture and the dimensions of θ change accordingly.

For the present application, the most interesting result is that an approximate sample from the target density can be obtained by sampling with replacement from $g^{(t)}$ and weighting the observations by means of $w^{(t)}$.

A convenient convergence diagnostic is the so-called normalized perplexity $\text{perp}_n = \exp(H_n)/n$, where $H_n = -\sum_{j=1}^n \bar{\omega}_j \log(\bar{\omega}_j)$ is the Shannon entropy of the normalized weights $\bar{\omega}_j$. It can be shown (Cappé *et al.*, 2008, sect. 2.1) that perp_n is an estimator of $\exp(K(f||g))$ and that is included in the interval $[0, 1]$. Moreover, Cappé *et al.* (2008) prove that, as $n \rightarrow \infty$, the algorithm monotonically decreases the cross-entropy (1). Thus, it seems plausible to stop the algorithm when the normalized perplexity cannot be further increased (so that the entropy cannot be further decreased). Therefore, monitoring the perplexity over some (say 5) successive iterations, we can stop the algorithm when the normalized perplexity does not change significantly or a predefined ‘large’ number of iterations M is reached.

From the technical point of view, the fundamental issue is the update of the importance function g at each iteration. When g is a multivariate normal mixture, the relevant equations are derived by Cappé *et al.* (2004). At iteration t , the importance weights associated to the sample $(x_1^{(t)}, \dots, x_n^{(t)})$ are given by

$$w_j^{(t)} = \frac{f(x_j^{(t)})}{\sum_{d=1}^D \pi_d^{(t)} \phi(x_j^{(t)}; \mu_d^{(t)}, \Sigma_d^{(t)})}, \quad j = 1, \dots, n,$$

and the normalized weights are $\bar{w}_j^{(t)} = w_j^{(t)} / \sum_{j=1}^n w_j^{(t)}$. The update is performed iterating, for any $i = 1, \dots, D$, the following equations:

$$\pi_i^{(t+1)} = \sum_{j=1}^n \bar{w}_j^{(t)} \tau_i(x_j^{(t)}; \pi^{(t)}, \mu^{(t)}, \Sigma^{(t)}), \quad (3)$$

$$\mu_i^{(t+1)} = \frac{\sum_{j=1}^n \bar{w}_j^{(t)} x_j^{(t)} \tau_i(x_j^{(t)}; \pi^{(t)}, \mu^{(t)}, \Sigma^{(t)})}{\pi_i^{(t+1)}}, \quad (4)$$

$$\Sigma_i^{(t+1)} = \frac{\sum_{j=1}^n \bar{w}_j^{(t)} (x_j^{(t)} - \mu_i^{(t+1)})(x_j^{(t)} - \mu_i^{(t+1)})' \tau_i(x_j^{(t)}; \pi^{(t)}, \mu^{(t)}, \Sigma^{(t)})}{\pi_i^{(t+1)}}, \quad (5)$$

where

$$\tau_i(x; \pi, \mu, \Sigma) = \frac{\pi_i \phi(x; \mu_i, \Sigma_i)}{\sum_{d=1}^D \pi_d \phi(x; \mu_d, \Sigma_d)}. \quad (6)$$

2.2 The Multivariate Generalized Hyperbolic Distribution

A random vector X is said to have a multivariate GHD($\lambda, \alpha, \beta, \mu, \delta, \Delta$) if its density is given by:

$$f(x; \lambda, \gamma, \beta, \mu, \delta, \Delta) = (2\pi)^{-p/2} \left(\frac{\gamma}{\delta} \right)^\lambda \left(\frac{\sqrt{\delta^2 + (x - \mu)' \Delta^{-1} (x - \mu)}}{\gamma^2 + \beta' \Delta \beta} \right)^{\lambda - p/2} \cdot K_{\lambda - p/2} \left(\sqrt{(\delta^2 + (x - \mu)' \Delta^{-1} (x - \mu))(\gamma^2 + \beta' \Delta \beta)} \right) \cdot \frac{\exp\{(x - \mu)' \beta\}}{K_\lambda(\delta \gamma)},$$

where $\delta > 0$ and $\gamma = \sqrt{\alpha^2 - \beta' \Delta \beta} \geq 0$ are scalars, $\beta, \mu \in \mathbb{R}^p$, $\alpha^2 \geq \beta' \Delta \beta$, Δ is a $p \times p$ positive definite symmetric matrix with $\det(\Delta) = 1$ (to avoid identifiability problems) and $K_\lambda(x)$ is the modified Bessel function of the third kind of order λ .

The distribution is generalized because, according to the values taken by λ , one gets several distributions known by different names: see McNeil *et al.* (2005, p. 80) for details. A major step towards the implementation of the GHD in practical applications was the development of Maximum Likelihood Estimation by means of the EM algorithm (Protassov, 2004).

2.3 Adaptive Importance Sampling for the simulation of the GHD

The implementation of Adaptive IS for the simulation of the GHD requires to derive the updating equations resulting from the minimization of the cross-entropy (1). The algorithm based on Gaussian mixtures works as follows.

1. Choose the instrumental density $g^{(1)} = \sum_{d=1}^D \pi_d \phi(\mu_d, \Sigma_d)$, i.e. the parameters D , π , μ_i , Σ_i ($i = 1, \dots, D$).
2. Simulate $x_1^{(1)}, \dots, x_n^{(1)}$ independently from $g^{(1)}$.
3. Compute the importance weights $w_j^{(1)} = f(x_{1j}^{(1)}, \dots, x_{pj}^{(1)}) / g(x_{1j}^{(1)}, \dots, x_{pj}^{(1)})$ ($j = 1, \dots, n$), where f is the GHD _{p} density function.

For $t \geq 1$:

4. Update the importance function to $g^{(t+1)}$ using equations (3) to (6) and the previous weighted sample $(x_1^{(t)}, w_1^{(t)}), \dots, (x_N^{(t)}, w_N^{(t)})$.
5. Simulate $x_1^{(t+1)}, \dots, x_N^{(t+1)}$ independently from $g^{(t+1)}$;
6. Compute the importance weights $w_j^{(t+1)}$ ($j = 1, \dots, n$);
7. Go back to step 4.

3 Simulation and application

The choice of the mixture parameters D , π , μ_i and Σ_i ($i = 1, \dots, D$), of the sample size n , and of the maximum number of iterations M depends on several factors. To assess their impact in the present setup, we perform some simulation experiments.

First we fit a five-variate GHD to a sample of log-returns of five financial time series. The data are the exchange rates of the Swiss Franc (CHF), US Dollar (USD), English Pound (GBP), Canadian Dollar (CAD) and Japanese Yen (JPY) against the Euro. The time horizon is from January 1, 1990 to December 31, 2009. The MLEs obtained by means of the EM algorithm (Protassov, 2004) are as follows:

$$\begin{aligned}
 \hat{\mu} &= (0.0003, -7.89 \cdot 10^{-5}, -0.0003, -0.0007, 0.0006)', \\
 \hat{\beta} &= (-35.0977, -2.6307, 17.2315, 12.1250, -12.5377)', \\
 \hat{\delta} &= 0.0043, \quad \hat{\gamma} = 300.8946, \\
 \hat{\Delta} &= \begin{pmatrix} 0.299 & 0.023 & 0.021 & -0.038 & 0.252 \\ 0.023 & 1.990 & 0.742 & 1.309 & 1.274 \\ 0.021 & 0.742 & 1.082 & 0.646 & 0.511 \\ -0.038 & 1.309 & 0.646 & 2.202 & 0.760 \\ 0.252 & 1.274 & 0.511 & 0.760 & 2.613 \end{pmatrix}.
 \end{aligned} \tag{7}$$

Using only the first two variables, namely CHF and USD, we get:

$$\begin{aligned}\hat{\mu} &= (0.0002, -0.0002)', \quad \hat{\beta} = (-35.716, 6.094)', \\ \hat{\delta} &= 0.0029, \quad \hat{\gamma} = 301.4950, \quad \hat{\Delta} = \begin{pmatrix} 0.385 & 0.047 \\ 0.047 & 2.602 \end{pmatrix}.\end{aligned}\quad (8)$$

Fig. 1(a) displays the contour plot of the GHD_2 density with the parameters in (8) and of the instrumental density; Fig. 1(b) displays 1000 simulated observations from the GHD_2 . The IS density is the best-fitting bivariate normal mixture, obtained by means of the Adaptive IS procedure, with $n = 10000$; the algorithm stops after a number of iterations equal to $\min\{20, t^*\}$, where $t^* > 5$ is such that $\max_{i,j \in \{t^*, t^*-1, \dots, t^*-5\}} |\text{perp}_n^{(i)} - \text{perp}_n^{(j)}| < \varepsilon$, with $\varepsilon = 0.001$. With $D = 7$, the quality of the approximation is very good, with a maximum perplexity equal to 0.9943.

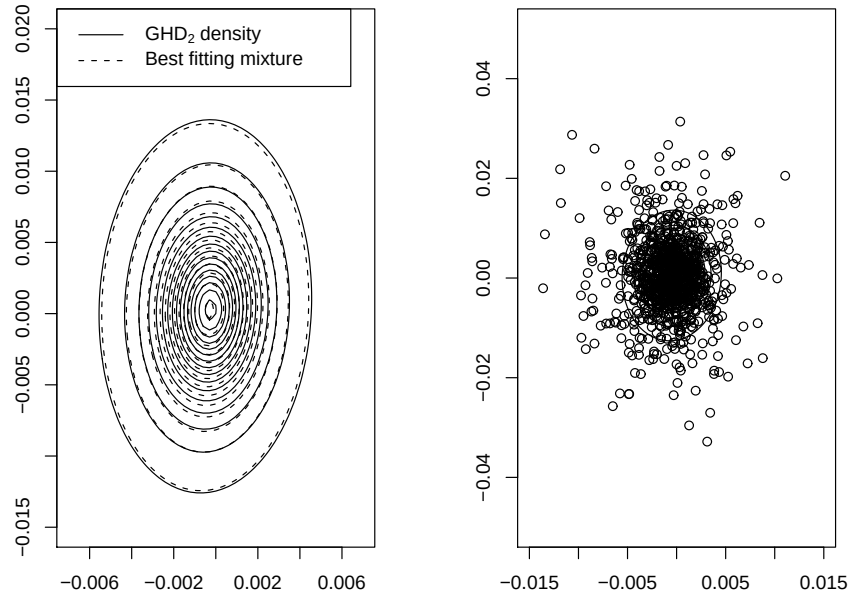


Fig. 1 (a) Contours of the GHD_2 density and of the best fitting Gaussian mixture; (b) 1000 observations sampled from the best-fitting mixture.

Next we apply the Adaptive IS procedure to the 5-variate distribution with the parameters in (7). In this case, we perform a small simulation study, consisting in replicating the procedure $B = 50$ times.

Fig. 2 displays the boxplots of the perplexities obtained in the first 10 iterations; ‘Failure rate’ is the ratio ‘{number of replications in which the algorithm did not converge}/{total number of replications}’. The initialization is the same in all cases. Whereas the four graphs are approximately similar, there is a considerable difference in the failure rate. We can therefore conclude that, when p increases, a large sample size is of crucial importance. When n is small, it often happens that the algorithm breaks down, typically because the covariance matrix of one of the populations becomes singular.

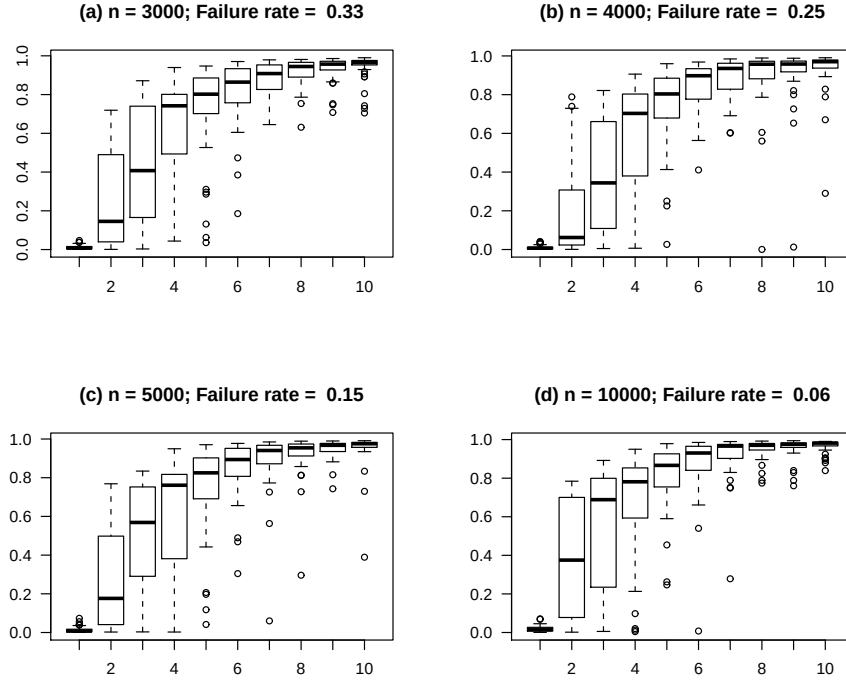


Fig. 2 Normalized perplexity estimates for the first 10 iterations of the algorithm over 50 simulation replications. The distributions are shown as boxplots. The thick horizontal line is the median, the box is the interquartile range (IQR), containing 50% of the points; the whiskers correspond to the interval $1.5 \cdot IQR$ from either Q1 (lower) or Q3 (upper); points outside the interval $[Q1, Q3]$ (outliers) are represented as circles. ‘Failure rate’ is the ratio ‘{number of replications in which the algorithm did not converge}/{total number of replications}’.

Finally, we compute the joint tail probability $p = P(r_1 < c_1, r_2 < c_2, r_3 < c_3, r_4 < c_4, r_5 < c_5)$ for the logarithmic returns of the exchange rates at hand, using $n = 10\,000$ and various thresholds c_i ($i = 1, \dots, 5$). The results are displayed in Table 1.

Table 1 Joint tail probability $p = P(r_1 < c_1, r_2 < c_2, r_3 < c_3, r_4 < c_4, r_5 < c_5)$ for the 5-dimensional data and various thresholds.

$\mathbf{c}$	$(0, 0, 0, 0, 0)$	$-(2, 2, 2, 2, 2) \cdot 10^{-3}$	$-(5, 2, 2, 2, 5) \cdot 10^{-3}$	$-(5, 2, 5, 2, 10) \cdot 10^{-3}$
p	0.0348	0.0062	0.0024	$4 \cdot 10^{-4}$

4 Conclusion

In this paper we have used Adaptive Importance Sampling to simulate the multivariate Generalized Hyperbolic Distribution and compute joint tail probabilities for some exchange rates. Both the simulation experiments and the empirical application showed that the procedure guarantees a very precise approximation. For small-dimensional problems, even a rather small sample size produces good results; when the dimension increases, a larger sample size is necessary. This approach seems to be a convenient and general solution for the simulation of non-standard multivariate distributions with known density.

References

1. Barndorff-Nielsen O.E.: Exponentially decreasing distributions for the logarithm of the particle size. Proceedings of the Royal Society. London. Series A. Mathematical and Physical Sciences 353, 401-419 (1977).
2. Cappé, O., Guillin, A., Marin, J., Robert, C.P.: Population Monte Carlo. Journal of Computational and Graphical Statistics. **13**, 907-929 (2004).
3. Cappé, O., Douc, R., Guillin, A., Robert, C.P.: Adaptive Importance Sampling in general mixture classes. Statistics and Computing. **18**, 447-459 (2008).
4. Dagpunar, J.S.: An easily implemented Generalized Inverse Gaussian generator. Communications in Statistics - Simulations. **18**, 703-710 (1989).
5. Protasov, R.S.: EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ . Statistics and Computing. **14**, 67-77 (2004).
6. Wraith, D., Kilbinger, M., Benabed, K., Cappé, O., Cardoso, J.-F., Fort, G., Prunet, S., Robert, C.P.: Estimation of cosmological parameters using adaptive importance sampling. Physical Review D. **80**, 023502. (2009).