SIMULATION OF POSITIVE NORMAL VARIABLES USING SEVERAL PROPOSAL DISTRIBUTIONS

Vincent Mazet, David Brie

Université Henri Poincaré, Nancy 1,
CRAN, CNRS UMR 7039, BP 239,
54506 Vandœuvre-lès-Nancy Cedex, France
Tel: +33 3 83 68 44 74, fax: +33 3 83 68 44 62,
`firstname.lastname@cran.uhp-nancy.fr`

Jérôme Idier

École Centrale de Nantes,
IRCCyN, CNRS UMR 6597,
BP 92101, 44321 Nantes Cedex 3, France
Tel: +33 2 40 37 69 09, fax: +33 2 40 37 69 30,
`jerome.idier@irccyn.ec-nantes.fr`

ABSTRACT

In this paper, we propose a new methodology to generate random variables distributed according to a Gaussian with positive support. We narrow the study to the univariate case. The method consists in an accept-reject algorithm in which a previous step is added consisting in choosing among several proposal distributions the one which gives the highest average probability of acceptance for given parameters of the target distribution. This results in a very fast method since it generates low reject.

1. INTRODUCTION

We propose an accept-reject algorithm to simulate positive normal variables in the univariate case. The target distribution (of parameters μ and σ^2) is given by:

$$f(x) = \frac{1}{C} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathbb{1}_{\mathbb{R}^+} \tag{1}$$

where C is a normalization constant allowing $f(x)$ to have an integral equal to one.

$$C = \sqrt{\frac{\pi\sigma^2}{2}} \left[1 + \operatorname{erf}\left(\frac{\mu}{\sqrt{2\sigma^2}}\right)\right].$$

Note that the value of C is not needed for the implementation of the method. The shape of the distribution is varying with respect to μ and σ^2. The mean and the variance of the normal distribution truncated at zero are:

$$\mathbb{E}[x] = \mu + \sqrt{\frac{2\sigma^2}{\pi}} \frac{\exp\left(-\mu^2/2\sigma^2\right)}{1 + \operatorname{erf}\left(\mu/\sqrt{2\sigma^2}\right)},$$

$$\operatorname{Var}[x] = \sigma^2 + \frac{\mu^2}{4} - \left[\frac{\mu}{2} + \sqrt{\frac{2\sigma^2}{\pi}} \frac{\exp\left(-\mu^2/2\sigma^2\right)}{1 + \operatorname{erf}\left(\mu/\sqrt{2\sigma^2}\right)}\right]^2.$$

This work originates from the problem of (blind) deconvolution of positive sparse spikes arising in applications such as optical spectroscopy [1, 2] or DNA sequencing [3]. In such a case, the signal to restore may be modeled as a Bernoulli-positive Gaussian process. Using an MCMC method requires to generate samples following a positive normal distribution. Another example is the blind separation of positive sources with positive mixing coefficients arising in chemical mixture analysis applications [4].

In [5, 6], it is proposed to use the inversion method, which generates $u \sim \mathcal{U}_{[0,1]}$, then computes:

$$x = \mu + \sqrt{2\sigma^2} \operatorname{erf}^{-1}\left(u + \operatorname{erf}(\mu/\sqrt{2\sigma^2})(u-1)\right)$$

where erf is the error function. This method has the advantage to give an explicit expression and to be mathematically exact. However, in practical applications, the use of erf can be inefficient if $-\mu$ is too large since the precision of the approximation of erf strongly matters [7] (see also section 4).

Another approach consists in using an accept-reject method. Of course, the simplest proposal distribution is the normal distribution, but this method is only suited to the case where μ is large enough (see figure 2). Robert [7] presents an approach whose proposal distribution is an exponential. Contrary to the normal distribution, this one is suited to the case where μ tends toward $-\infty$ (see figure 2). Thus, it appears that depending on the shape of the target distribution, different proposal distributions have to be used to get a high APA (average probability of acceptation). This is the main idea of the proposed approach which consists in determining among different proposal distributions a priori chosen the one which is the best suited to the target distribution.

In section 2, we present the proposed approach in a general setting and apply it to the considered case of simulating positive normal variables in section 3. Section 4 presents some comparison with the inversion and other accept-reject methods and illustrates that the proposed approach performs well for all the possible situations which may occur in positive normal variable simulation. Finally, section 5 concludes the paper and gives some perspectives to this work.

2. PROPOSED APPROACH

The accept-reject algorithm needs a proposal distribution g and a constant M such as

$$\forall x \in S, \quad M \geq f(x)/g(x) \tag{2}$$

where S is the support of f. Necessary, the proposal distribution g is non-zero on S because M needs to be finite to have a non-zero APA (see equation (4)). The accept-reject algorithm results from the following lemma:

Lemma 1 [5, 7, 8] The random variable x resulting from the following algorithm is distributed according to f:

1. generate $z \sim g(z)$ and $u \sim \mathcal{U}_{[0,1]}$,
2. compute $\rho(z) = f(z)/Mg(z)$,
3. if $u \leq \rho(z)$: $x = z$ (accept),

   else: go back to step 1 (reject).

The choice of the proposal distribution is deciding for the method performances. First of all, the proposal distribution should be easily simulated, otherwise the method looses its interest. In particular, they should be simulated with a probability of acceptation of 1; otherwise, the APA of the proposal distribution has to be taken into account to evaluate the final APA. Then, one can choose common distributions (in our example: the normal and exponential distributions), or build particular distributions which can be easily simulated (in our example: a normal distribution coupled with a uniform one). Also the choice of the proposal distribution has to be made regarding to the complexity of the algorithm: some distributions, interesting at a first sight, turn out to be inadequate because the determination of M or the calculus of its parameters can be difficult, time consuming, or even impossible!

Considering the determination of M, any constant satisfying the equation (2) suits; however, M should be the smallest to have a high APA (see equation (4)). The optimal value for M is then:

$$M = \max_{x \in S} f(x)/g(x).$$

Unfortunately, M is not always computable. Knowing M, one can then compute the probability of acceptation ρ:

$$\rho(x) = f(x)/Mg(x). \tag{3}$$

At last, the APA $\overline{\rho} = \mathbb{E}[\rho(x)]$ allows to define a measure of the algorithm efficiency. The higher the APA is, the better the algorithm works.

$$\overline{\rho} \triangleq \int \rho(x)g(x)dx = \frac{1}{M}\int f(x)dx = \frac{1}{M}. \tag{4}$$

Note that if g is close to zero, then M increases and the APA decreases: the algorithm efficiency depends on the adequation between f and g. In particular, g has to have a heavier tail than f to keep M finite. However, the difference between f and g should not be too important, unless M becomes too high and then the APA becomes too low [8].

Sometimes, the expression of the probability of acceptance depends on the parameters of the proposal distribution (for example, the parameter α of the exponential distribution in our particular case: see section 3). In that case, one has to compute the parameters which maximise the probability of acceptance.

As mentioned before, the main idea of the proposed approach is to choose a priori some proposal distributions and then to determine the one which is the best suited to the target distribution. Among every proposal distributions, only one gives the best APA for some particular parameters of the target distribution. Then, one has to compute the parameter intervals on which the corresponding proposal distribution yields the best APA. So, the proposed algorithm is identical to the accept-reject algorithm but with a previous step added, consisting in selecting among a set of p different proposal distributions the best one:

1. determine the proposal distribution $g_i \in \{g_1, \ldots, g_p\}$ according to the parameters of the target distribution,
2. compute M_i,
3. generate $z \sim g_i(z)$ and $u \sim \mathcal{U}_{[0,1]}$,
4. compute $\rho(z) = f(z)/M_i g_i(z)$,
5. if $u \leq \rho(z)$: $x = z$ (accept),

   else: go back to step 3 (reject).

In section 3, this approach is applied to the simulation of variables following a positive normal distribution, that is a normal distribution truncated at $t = 0$. However, before going further, let us note that if f is truncated at $t \neq 0$, the method can be adapted by simply shifting the random variable. For example, in the case of simulating positive normal variables:

$$\begin{aligned} & X \sim \mathcal{N}(\mu, \sigma^2) \;/\; X \in [0, +\infty[\\ \Leftrightarrow \quad & Y = X + t \sim \mathcal{N}(\mu + t, \sigma^2) \;/\; Y \in [t, +\infty[. \end{aligned}$$

Without loss of generality, we may restrict our attention to the case $\sigma^2 = 1$, since the other cases result from a scale change:

$$\begin{aligned} & X \sim \mathcal{N}(\mu, 1) \;/\; X \in [0, +\infty[\\ \Leftrightarrow \quad & Y = X\sigma \sim \mathcal{N}(\mu\sigma, \sigma^2) \;/\; Y \in [0, +\infty[. \end{aligned}$$

3. SIMULATION OF POSITIVE NORMAL VARIABLES

Four proposal distributions, shown on figure 1, are considered:

① The normal distribution:

$$g_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \;;$$

② The normal distribution coupled with the uniform one (it is a distribution defined on $\mathbb{R}^+$, uniform on $[0, \mu[$ and distributed according to a normal law $\mathcal{N}(\mu, \sigma^2)$ on $[\mu, +\infty[$):

$$g_3(x) = \frac{\mathbb{1}_{\mathbb{R}^+}}{\mu + \sqrt{\frac{\pi\sigma^2}{2}}} \begin{cases} 1 & \text{if } 0 \le x < \mu \\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \text{if } x \ge \mu \end{cases}$$

with $\mu \ge 0$;

③ The normal distribution truncated at the mean:

$$g_2(x) = \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathbb{1}_{[\mu, +\infty]}$$

with $\mu \le 0$;

④ The exponential distribution [7]:

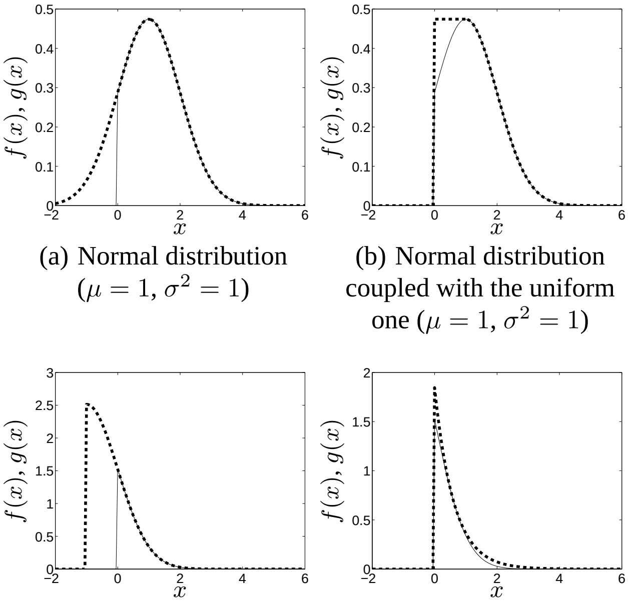
$$g_4(x) = \alpha \exp(-\alpha x) \mathbb{1}_{\mathbb{R}^+}$$

where the value of α corresponds to the one that maximises the APA (see appendix A):

$$\alpha = \left(\sqrt{\mu^2 + 4\sigma^2} - \mu\right) / 2\sigma^2. \qquad (5)$$

The techniques used to generate variables from these proposal distributions are discuted in the appendix B.

Our choice is motivated by the fact that the distributions ① and ④ are expected to yield a very high APA for $|\mu| \gg 0$, and the distributions ② and ③ are expected to improve the APA around zero (see figure 2). For each distributions, we have to compute the expression of the constant M and of the probability of acceptance ρ (table 1). Calculus are detailed in appendix A. The APA for each proposal distribution are drawn with respect to μ on figure 2.



(a) Normal distribution ($\mu = 1$, $\sigma^2 = 1$)

(b) Normal distribution coupled with the uniform one ($\mu = 1$, $\sigma^2 = 1$)

(c) Normal distribution truncated at the mean ($\mu = -1$, $\sigma^2 = 1$)

(d) Exponential distribution ($\mu = -1$, $\sigma^2 = 1$)

Fig. 1. Proposal distributions (dashed line) used to simulate a normal distribution truncated at zero (plain line).

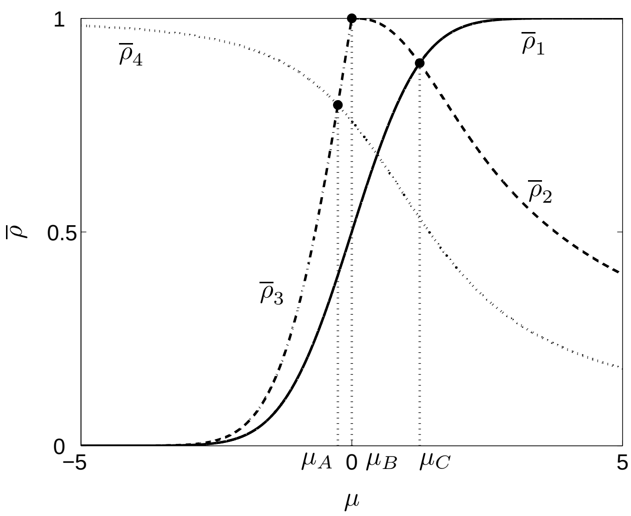


Fig. 2. Average probability of acceptance for the four proposal distribution ($\sigma^2 = 1$).

The three intersection points μ_A, μ_B, and μ_C allow us to define the best proposal distribution for a given μ. They are obtained by equaling the different APA:

$$\overline{\rho}_4(\mu_A) = \overline{\rho}_3(\mu_A) \;\; \overline{\rho}_3(\mu_B) = \overline{\rho}_2(\mu_B) \;\; \overline{\rho}_2(\mu_C) = \overline{\rho}_1(\mu_C)$$

proposal distrib.	M	$\rho(x)$
①	$\sqrt{2\pi\sigma^2}/C$	1 if $x \geq 0$, 0 unless
②	$\left(\mu + \sqrt{\pi\sigma^2/2}\right)/C$	$\begin{cases} \exp\left(-(x-\mu)^2/2\sigma^2\right) & \text{if } 0 \leq x < \mu \\ 1 & \text{if } x \geq \mu \end{cases}$
③	$\sqrt{2\pi\sigma^2}/2C$	1 if $x \geq 0$, 0 unless
④	$\exp\left(\frac{\alpha}{2}(2\mu + \alpha\sigma^2)\right)/\alpha C$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2} - \frac{\alpha}{2}(2\mu - 2x + \alpha\sigma^2)\right)$

Table 1. Expression of the constants M and the probabilities of acceptance ρ for the four proposal distributions.

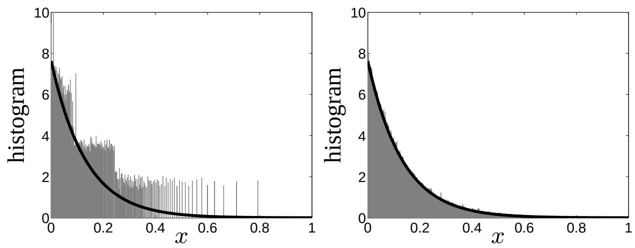
The calculus are straightforward for μ_B and μ_C and yields:

$$\mu_B = 0, \qquad \mu_C = \sqrt{\pi\sigma^2/2}.$$

Section C details the calculus for μ_A and yields an approximated solution:

$$\mu_A \approx -0.257\sigma.$$

The APA of the method corresponds to the top of the curves in figure 2. The lowest APA corresponds to $\mu = \mu_A$ and is equal to about 0.797, which remains a very good probability of acceptance: the proposed algorithm is then very fast since it generates low reject.



(a) With the inversion method

(b) With the proposed approach

Fig. 3. Histogram of 100,000 variables with $\mu = -7.5$ and $\sigma^2 = 1$, and the truncated normal distribution overlaid.

4. NUMERICAL EXPERIMENTS

In this section, we compare the proposed approach with the inversion method and with other accept-reject algorithms. All the simulations are performed with Matlab; in particular, the function erf is the one defined by [9], and the normal and uniform variable generation are made using functions `randn` and `rand` respectively.

As mentioned before, the inversion method [5, 6] is mathematically exact and has an APA of 1 since every generated value is accepted. But the method could be inefficient in practical applications because of the function erf which is known only through approximations, resulting in numerical problems. To illustrate this, consider the generations of 100,000 positive normal variables with parameters $\mu = -7.5$ and $\sigma^2 = 1$ *via* both methods. The histograms are shown on figure 3 (infinite values are not plotted), and the truncated normal distribution is overlaid for comparison. It is clear that the inversion method is inefficient while the proposed approach works well. The performances of the inversion method could be improved by using a better approximation of the erf function, yielding a greater computational burden.

Moreover, for some parameter values, the inversion method could give negative results (about 0.07 % of the generated variables are negative for $\mu = -8$ and $\sigma^2 = 1$). Practically, it does not matter since these values can be rejected and replaced by new ones (therefore the APA is less than one). But they are also cases where the inversion method gives only infinite values (for example with $\mu = -8.5$ and $\sigma^2 = 1$). It is clear that in such a case, the inversion method cannot be applied.

We now compare the proposed approach with three accept-reject algorithms having only one proposal distributions (distributions ①, ② and ④). We just compare for different μ the computation time, which, in fact, is directly linked with the APA. The results are presented in table 2 and correspond to the simulation of 10,000 variables with $\sigma^2 = 1$.

In these four cases, the computation times are coherent with the APA as shown in figure 2, except for $\mu = 0.5$ where the normal distribution is faster than the normal distribution coupled with the uniform one, while its APA is expecting to be worse. This is due to the fact that the computation time of generating a normal variable is faster than generating a variable distributed according to distribution ②, but this difference depends on the software. Then, it seems difficult to determine the range of the different proposal distribution using the computation time, that is why we preferred to determine it using the theoretical APA. Nevertheless, the difference is small and the computation time between the proposed approach and the accept-reject algorithm using distribution ① is negligible and particular to the case $\mu = 0.5$.

	PA	①	②	④
$\mu = -2$	0.36 s	11.094 s	($\mu < 0!$)	0.39 s
$\mu = 0$	0.406 s	0.485 s	0.422 s	0.454 s
$\mu = 0.5$	0.406 s	0.375 s	0.422 s	0.515 s
$\mu = 2$	0.281 s	0.297 s	0.531 s	0.813 s

Table 2. Comparison of computation time for the proposed approach (PA) and accept-reject algorithm with a normal distribution ①, a normal distribution coupled with the uniform one ② and an exponential distribution ④.

5. CONCLUSION

This paper presents a new approach to simulate variables having a distribution whose shape is significantly varying according to the values of its parameters. Basically, the idea is to use an accept-reject algorithm whose proposal distribution is selected among a set of different distributions chosen a priori. It is then applied to the generation of random variables distributed according to a normal distribution truncated in zero. Four different proposal distributions are considered and the intervals on which they give the highest APA (and consequently on which they should be used) are calculated. Numerical simulations have been used to illustrate that the method performs well and fast for all the possible situations.

Regarding the perspectives of this work, a first point that has to be addressed concerns the possible use of other proposal distributions such as a normal distribution coupled with a polynomial function, or a normal distribution truncated at the mean with parameters differing from μ and σ^2 and determined to maximize the APA. Also, future works could be directed at considering the case of the two-sided truncated normal distribution which may be useful in the processing of censored data and at investigating the multivariate case.

6. REFERENCES

[1] S. Gulam Razul, W.J. Fitzgerald, and C. Andrieu, “Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC,” *Nucl. Inst. Meth. A*, vol. 497, pp. 492–510, 2003.

[2] V. Mazet, J. Idier, and D. Brie, “Déconvolution impulsionnelle positive myope,” in *Proceedings of the 20th GRETSI*, Leuven, Belgium, september 6-9 2005.

[3] N.M. Haan and S.J. Godsill, “Bayesian model for DNA sequencing,” in *Proceedings of ICASSP*, Orlando, Florida, USA, 2002, vol. 4, pp. 4020–4023.

[4] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, “Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling,” 2005, submitted to *IEEE Trans. Signal Process.*

[5] L. Devroye, *Non-uniform random variate generation*, Springer-Verlag, New York, 1986, Available on the Internet: http://jeff.cs.mcgill.ca/~luc/rnbookindex.html.

[6] A.E. Gelfand, A.F.M. Smith, and T.-M. Lee, “Bayesian analysis of constrained parameter and truncated problems using Gibbs sampling,” *J. Amer. Stat. Assoc.*, vol. 87, pp. 523–532, 1992.

[7] C.P. Robert, “Simulation of truncated normal variables,” *Stat. Comp.*, vol. 5, pp. 121–125, 1995.

[8] C.P. Robert, *Méthodes de Monte Carlo par chaînes de Markov*, Economica, Paris, 1996.

[9] W.J. Cody, “Rational Chebyshev approximations for the error function,” *Math. Comp.*, pp. 631–638, 1969.

A. DETAILS OF CALCULUS OF M AND ρ

The calculus of the constant M and the probability of acceptance $\rho(x)$ are straightforward for distributions ①, ②, and ③. For distribution ④, we have for $x \geq 0$:

$$\frac{f(x)}{g(x)} = \frac{1}{C\alpha} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} + \alpha x\right),$$

for which the maximum is reached for $x = \alpha\sigma^2 + \mu$. We obtain the best constant M:

$$M = \frac{1}{C\alpha} \exp\left(\frac{\alpha}{2}\left(2\mu + \alpha\sigma^2\right)\right),$$

leading to the probability of acceptance:

$$\rho(x) = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} - \frac{\alpha}{2}\left(2\mu - 2x + \alpha\sigma^2\right)\right).$$

The best value of α is the one which maximises the APA $\overline{\rho} = 1/M$ whose derivative is:

$$\frac{\partial\overline{\rho}}{\partial\alpha} = C \exp\left(-\frac{\alpha}{2}\left(2\mu + \alpha\sigma^2\right)\right)\left[1 - \alpha\mu - \alpha^2\sigma^2\right].$$

So, the derivative is zero for (remember that $\alpha > 0$):

$$\alpha = \left(\sqrt{\mu^2 + 4\sigma^2} - \mu\right)/2\sigma^2.$$

B. GENERATING VARIABLES FROM THE PROPOSAL DISTRIBUTIONS

A random variable x distributed according to distribution ② sets either in the uniform part (of area $\mathcal{A}_u$), or in the normal part (of area $\mathcal{A}_g$):

$$\mathcal{A}_u = \frac{\mu}{\mu + \sqrt{\pi\sigma^2/2}} \qquad \mathcal{A}_g = \frac{\sqrt{\pi\sigma^2/2}}{\mu + \sqrt{\pi\sigma^2/2}}.$$

By remarking that $\mathcal{A}_u + \mathcal{A}_g = 1$, the following algorithm generates a random variable x distributed according to distribution ②:

1. generate $u \sim \mathcal{U}_{[0,1]}$,
2. if $u < \mathcal{A}_u$, then x is in the uniform part, so:

   generate $v \sim \mathcal{U}_{[0,1]}$, and compute $x = \mu v$.

   otherwise x is in the normal part, so:

   generate $v \sim \mathcal{N}(0, \sigma^2)$ and compute $x = |v| + \mu$.

A random variable x distributed according to distribution ③ is build by adding μ to the absolute value of a centered normal variable y:

$$x = |y| + \mu \qquad \text{where} \qquad y \sim \mathcal{N}(0, \sigma^2).$$

The inversion method [5] is used to simulate the exponential distribution ④: generate $u \sim U_{[0,1]}$, then compute:

$$x = -\ln(1-u)/\alpha.$$

C. CALCULUS OF μ_A

Equation (5) can be rewritten as:

$$2\sigma^2\alpha + \mu = \sqrt{\mu^2 + 4\sigma^2},$$

from which we have:

$$\begin{aligned} & (2\sigma^2\alpha + \mu)^2 = \mu^2 + 4\sigma^2, \\ \Leftrightarrow \quad & 4\sigma^4\alpha^2 + 4\sigma^2\alpha\mu + \mu^2 = \mu^2 + 4\sigma^2, \end{aligned}$$

that is:

$$\mu = \frac{1 - \sigma^2\alpha^2}{\alpha}$$

By equaling the APA of distributions ③ and ④ and replacing μ given above, we obtain:

$$\alpha\sigma \exp\left(-1 + \sigma^2\alpha^2/2\right) = \sqrt{2/\pi}$$

which simplifies to:

$$\alpha\sigma \exp\left(\sigma^2\alpha^2/2\right) = e\sqrt{2/\pi}$$

We now obtain the following system:

$$\begin{cases} \alpha\sigma \exp\left(\sigma^2\alpha^2/2\right) = e\sqrt{2/\pi}, \\ \mu = (1 - \sigma^2\alpha^2)/\alpha, \end{cases}$$

We did not find an explicit solution to the first equation, but using a zero finding algorithm yields the approximated solution $\alpha\sigma \approx 1.137$, from which we get $\mu_A \approx -0.257\sigma$.

The Matlab code of this method is available free at:
http://mtde.cran.uhp-nancy.fr/Personnes/Perso_Mazet/rpnorm-en.htm