NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

# The Multiple Roots Simulation Algorithm, the Inverse Gaussian Distribution, and the Sufficient Conditional Monte Carlo Method

by

Gunnar Taraldsen and Bo Henry Lindqvist

PREPRINT
STATISTICS NO. 4/2005

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL

http://www.math.ntnu.no/preprint/statistics/2005/S4-2005.pdf

# The Multiple Roots Simulation Algorithm, the Inverse Gaussian Distribution, and the Sufficient Conditional Monte Carlo Method

GUNNAR TARALDSEN and BO HENRY LINDQVIST

7th March 2005

**Abstract**

Michael, Schucany & Haas (1976) presented a method for the simulation of random variates using transformations with multiple roots. A generalization of this method to include vector-valued transformations is given here.

A short introduction to the inverse Gaussian distribution is given. The joint distribution of the sufficient statistic is well known, but the proof given here based on the Basu theorem is possibly new.

The multiple roots simulation algorithm is related to the method for doing Monte Carlo simulations conditioned on a sufficient statistic presented by Lindqvist & Taraldsen (2005). The method is explained here by application on the inverse Gaussian distribution. The result is in this case an academic alternative to the much simpler and preferable method discovered by Cheng (1984) for doing Monte Carlo simulations from the inverse Gaussian distribution conditioned on the sufficient statistic. This means that the inverse Gaussian can be used as a non-trivial test case for a general numerical implementation of the sufficient conditional Monte Carlo method.

*Keywords*: Inverse Gaussian distribution; conditional distribution; Monte Carlo simulation; nuisance parameters; point estimation; sufficiency

## 1 Introduction

Schrödinger (1915) and Smoluchowsky (1915) proved independently that the inverse Gaussian $IG(d/v, d^2/\beta)$ distribution is the distribution of the travel time over a distance $d$ for a linear Brownian motion which tends to move with speed $v$ with a diffusion constant $\beta$. The corresponding random distance $D$ has a Gaussian distribution $Normal(vt, \beta t)$ in an experiment with a fixed time $t$. In recent years the inverse Gaussian distribution has played versatile roles in models of stochastic processes including the theory of generalized linear models, reliability and lifetime data analysis, accelerated life testing, and repair time distributions (Johnsen et al. 1994, vol.1, p.291).

A summary of the properties of the inverse Gaussian distribution is given by Johnsen et al. (1994, vol.1, p.259-298). For further information the reader is referred to the three books by Chhikara & Folks (1989), Seshadri (1993), and Seshadri (1999). A brief self-contained introduction to some of the most important properties of the inverse Gaussian distribution is given below, but in the next section a result of independent interest is presented.

## 2 Generating Random Variates Using Transformations with Multiple Roots

It is well known that an arbitrary real random variable $X$ can be generated from a $U \sim Uniform(0,1)$ due to $X \sim F^{-1}(U)$, where $F$ is the cumulative distribution of $X$. This holds also when $F$ is not invertible, with a suitable interpretation of $F^{-1}$. More generally, a random variable $X$ can be generated from a random variable $Z$ if there is a function $g$ such that $X \sim g(Z)$. A classical example is given by Box & Muller (1958) who demonstrated how two independent uniform variables can be transformed in a simple way to give two independent Gaussian variables.

Consider the problem of generation of a random vector $X$ where

$$g(X) = Z \tag{1}$$

and where the random vector $Z$ is easy to generate. Michael et al. (1976) demonstrated how this can be used to generate $X$ for the case of real variables, and applied the method to equation (8) given below in order to generate samples from the inverse Gaussian distribution. A generalization of the main result of Michael et al. (1976) to the multidimensional case will be presented now. The proof can be considered to be simpler than the proof given by Michael et al. (1976).

It will be assumed that there is a countable number of roots of equation (1) for each $Z$, and the number of roots may depend on $Z$. Intuitively it could at first sight seem impossible to generate $X$ from generation of $Z$, but the solution turns out to be obvious:

**ALGORITHM 1**

1. Generate $Z$. Let $z$ denote the result.
2. Generate $X_z$ from the conditional distribution of $X$ given $Z = z$. Let $x$ denote the result.

The resulting $x$ is clearly a sample from the distribution of $X$ even without assuming any relation (1) between the variables $X$ and $Z$. It is here, as always, assumed that the random variables are all defined on a common underlying probability space.

The second observation is that equation (1) can be used to simplify the simulation from the conditional distribution. The conditional distribution is concentrated at $\{x \,|\, g(x) = z\}$, and is hence discrete when there is a countable number of roots $x_r$. Sufficient smoothness of $g$ gives

$$p_{r_0} = P(X = x_{r_0} \,|\, g(X) = z) = \frac{f(x_{r_0}) / \left|g'(x_{r_0})\right|}{\sum_r f(x_r) / \left|g'(x_r)\right|} \tag{2}$$

The smoothness of $g$ relatively to the probability density $f$ of $X$ must in particular be such that the absolute value $|g'(x_r)|$ of the Jacobi determinant is different from zero. A sufficient condition is that the domain of $g$ can be partitioned into a null set and a countable union of open sets for which the change-of-variables theorem (Rudin 1987, p.153) can be applied. Equation (2) is a generalization of equation 3 given by Michael et al. (1976).

Let $v_0 = 0$, $v_r = \sum_{r' \le r} p_{r'}$, and $V \sim Uniform(0,1)$. The conditional simulation required in Algorithm 1 can then be done by:

**ALGORITHM 2**

1. Generate $V$. Let $v$ denote the result.
2. Return $x = x_r$, where $x_r$ is the unique root such that $v_{r-1} < v \le v_r$.

The major difficulty in the combined algorithm is the determination of the roots, and the calculation of the corresponding weights. In favorable cases this can be done analytically. Both possibilities will be exemplified in the following.

The difficult part of the proof of equation (2) is to prove that the denominator $\sum_r f(x_r)/\left|g'(x_r)\right|$ is the density of $Z = g(X)$. In some cases this density is explicitly given, but more generally it can be found by simulation of $Z$. This information is useful in cases where the roots must be found numerically and the number of roots is unknown. It will typically not be necessary to determine all the roots.

# 3 The Inverse Gaussian Distribution

The one-parameter Wald probability density $f$ with shape parameter $\phi > 0$ is given by

$$f(x) = \sqrt{\frac{\phi}{2\pi}}\, e^{\phi} x^{-3/2} \exp\left[-\frac{\phi}{2}(x + x^{-1})\right], \quad x > 0 \tag{3}$$

The notation $X \sim Wald(\phi)$ means that the above $f$ is the density of the random variable $X$. If $X \sim Wald(\phi)$, then the corresponding scale family given by the variables $Y = \mu X$ gives the inverse Gaussian family. The notation $Y \sim IG(\mu, \lambda)$ is used, where in this case $\lambda = \phi\mu$. One may also write

$$IG(\mu, \mu\phi) = \mu\, IG(1, \phi) = \mu\, Wald(\phi) \tag{4}$$

Together with the Gaussian and gamma distributions, the inverse Gaussian completes the trio of families that are both an exponential and a group family of distributions (Lehmann & Casella 1998, p.68).

Equation (4) gives the properties of the $IG(\mu, \lambda)$ distribution from the properties of the $Wald(\phi) = IG(1, \phi)$ distribution. For most purposes it is hence sufficient to discuss the properties of the variable $X \sim IG(1, \phi)$. A short calculation gives the characteristic function

$$Ee^{itX} = \exp\left[(1 - \sqrt{1 - 2i\phi^{-1}t})\phi\right] \tag{5}$$

where $\sqrt{\cdot}$ denotes the principal value of the complex square root. From equation (5) the expectation and the variance are respectively

$$EX = 1, \quad \operatorname{Var} X = \phi^{-1} \tag{6}$$

The density of $X$ is unimodal with the mode at

$$x_{\text{mode}} = \sqrt{1 + (3\phi^{-1}/2)^2} - 3\phi^{-1}/2 \tag{7}$$

For small $\phi$ the distribution is localized close to 0, and for large $\phi$ the distribution is localized close to 1 and the shape is close to the Gaussian distribution. The log-normal distribution gives an even better approximation.

Shuster (1968) discovered the following fundamental property

$$g(X) := \phi\frac{(X-1)^2}{X} = Z \sim \chi^2_1 \tag{8}$$

The proof follows from calculus and equation (3). This property will be used more than once in the following.

# 4 Sufficient Statistics

Let $X_1, X_2, \ldots, X_n$ be a random independent sample from $IG(\mu, \lambda)$. The Halmos-Savage factorization theorem (Halmos & Savage 1949) together with equation (3) and (4) give that the sample mean $\overline{X} = (X_1 + \cdots + X_n)/n$ and the sample harmonic mean $\widetilde{X} = 1/(\overline{1/X}) = n/(1/X_1 + \cdots + 1/X_n)$ together is a sufficient statistic. An alternative sufficient statistic is given by $(\overline{X}, V)$, where

$$V = \overline{1/X} - 1/\overline{X} \tag{9}$$

and $V \sim (n\lambda)^{-1}\chi^2_{n-1}$ is independent of $\overline{X} \sim IG(\mu, n\lambda)$. A short and possibly new proof of this based on equation (8) and the Basu theorem will be given next. The characteristic function in (5) gives $\overline{X} \sim IG(\mu, n\lambda)$, and the rest of the proof follows from the key identity

$$\frac{\lambda}{\mu^2}\sum_i \frac{(X_i - \mu)^2}{X_i} = \lambda n V + \frac{\lambda n}{\mu^2}\frac{(\overline{X} - \mu)^2}{\overline{X}} \tag{10}$$

Equation (8) gives that left-hand side is a $\chi^2_n$ variable, and that the last term is a $\chi^2_1$ variable. A first consequence of this is that the distribution of $V$ only depends on the parameter $\lambda$, and is hence a pivotal statistic with respect to the parameter $\mu$. Since, for a fixed $\lambda$, $\overline{X}$ is a complete and sufficient statistic for $\mu$ the Basu theorem gives the stated independence of $V$ and $\overline{X}$. Finally, this independence gives the uniqueness result required for the conclusion $\lambda n V \sim \chi^2_{n-1}$.

The independence of $\overline{X}$ and $V$ for fixed $n \geq 2$ for positive non-constant i.i.d. variables $X_i$ determines the distribution of $X_i$ to be in the class $IG(\mu, \lambda)$, where the case $\mu = \infty$ is included. This characterization was proven by Letac & Seshadri (1985) and a proof is also found in the monograph by Seshadri (1993, p.97-100). This impressive result corresponds to the similar characterization of the Gaussian distribution in terms of independence of the sample mean and the sample variance and existence of the first moment. There is also a corresponding result for gamma variables: If $X_1$ and $X_2$ are independent nondegenerate random variables, then $X_1 + X_2$ is independent of $X_1/X_2$ if and only if there exist a constant $c$ such that $cX_1$ and $cX_2$ have standard gamma distributions (Johnsen et al. 1994, vol.1, p.350).

# 5 Simulation from the Inverse Gaussian Distribution

Simulation from $IG(\mu, \lambda)$ is reduced to simulation from $IG(1, \phi)$ due to $IG(\mu, \mu\phi) = \mu\, IG(1, \phi)$.

It seems possible to use rejection sampling (Ripley 1987, p.60) based on approximation by the log-normal distribution to sample from the Wald distribution. Other possibilities are given by adaptive rejection methods or Markov chain methods (Evans & Swartz 2000). Inversion based on the cumulative distribution (Johnsen et al. 1994, vol.1, p.262)

$$F(x) = \Phi\left[(x-1)\sqrt{\phi/x}\right] + e^{2\phi}\Phi\left[-(x+1)\sqrt{\phi/x}\right] \tag{11}$$

where $\Phi$ is the standard Gaussian cumulative distribution, is also possible. Neither of these possibilities will be used in the following.

Michael et al. (1976) demonstrated that Algorithm 1 gives a simple algorithm. The roots of $g(x) = z$ in equation (8) are

$$x_1 = 1 - \frac{1}{2\phi}(\sqrt{z^2 + 4\phi z} - z), \quad x_2 = 1/x_1 \tag{12}$$

and it can be observed that $x_1 < 1 < x_2$. The conditional probability of $X = x_1$ from equation (2) is

$$p_1 = P(X = x_1 \,|\, Z = z) = 1/(1 + x_1) \tag{13}$$

Note that $p_1 > 1/2$ since $x_1 < 1$. The resulting algorithm for sampling from $IG(1, \phi)$ is:

**ALGORITHM 3**

1. Generate independent $(U, Z)$ with $U \sim Uniform(0, 1)$ and $Z \sim \chi^2_1$. Let $(u, z)$ denote the result.
2. Let $x = x_1(z; \phi) = 1 - \frac{1}{2\phi}(\sqrt{z^2 + 4\phi z} - z)$.
3. If $(1 + x)u > 1$, let $x = 1/x_1$.

The algorithm can also be summarized by the single transformation

$$x = \psi(u, z; \phi) = [(1 + x_1)u \leq 1] \cdot x_1 + [(1 + x_1)u > 1] \cdot 1/x_1 \tag{14}$$

defined by indicator functions. A similar transformation results also in the more general case with an arbitrary number of roots.

# 6 Sufficient Conditional Simulation from the Inverse Gaussian Distribution

The concept of sufficiency is due to Fisher (1920) according to Savage (1976, p.453), Rao (1992, p.42), and Lehmann & Casella (1998, p.143). It is a part of the foundations of statistics through the sufficiency principle, and it is practically important with applications for example in construction of optimal estimators and nuisance parameter elimination (Halmos & Savage 1949, Welsh 1996, Lehmann 1997, Lehmann & Casella 1998). The required conditioning may however be difficult to implement in practical problems. Lindqvist & Taraldsen (2005) present the sufficient conditional Monte Carlo method as one alternative approach. This will be applied here to give an algorithm for the simulation of samples from the inverse Gaussian distribution conditioned on the sample mean and the sample harmonic mean.

Let $U$ be a $n \times 2$ matrix of independent random variables where the first column consist of $Uniform(0, 1)$ variables and the second column consist of $\chi^2_1$ variables. Let $\theta = (\mu, \phi)$ and define the column valued function $\chi(u, \theta)$ by

$$\chi_i(u, \theta) = \mu\psi(u_{i1}, u_{i2}; \phi) \tag{15}$$

where $\psi$ is defined in equation (14). Define also

$$\tau(u, \theta) = (\overline{\chi}(u, \theta), \widetilde{\chi}(u, \theta)), \tag{16}$$

where $\overline{\chi}$ is the arithmetic mean and $\widetilde{\chi}$ is the harmonic mean. It follows that $\chi(U, \theta)$ is an independent sample from $IG(\mu, \mu\phi)$ and $\tau(U, \theta)$ is a sufficient statistic.

Choose a possibly improper density $\pi$ on the first quadrant $\{(\mu, \phi) \,|\, \mu > 0, \phi > 0\}$, and let $\Theta$ have this density and be independent of $U$. Let $t \mapsto w_t(u)$ be the density of $\tau(u, \Theta)$. An algorithm for conditional simulation from $X \,|\, T = t$ is then (Lindqvist & Taraldsen 2005):

**ALGORITHM 4**

1. Generate V from a density proportional to $w_t$ times the density of $U$. Let $v$ be the result.
2. Generate $\Theta_t$ from the conditional distribution of $\Theta$ given $\tau(v, \Theta) = t$.
3. Return $X_t = \chi(V, \Theta_t)$.

Generally speaking Step 2 is comparable with the initial conditional problem, but this switch from $U$ to $\Theta$ leads to simplifications in the present case since there is a finite number of solutions, and simulation from the conditional distribution can be done as described in Algorithm 2.

The roots can be found by first determining $\phi$'s from the equation $\tau_1(u, \mu, \phi)t_2 = \tau_2(u, \mu, \phi)t_1$, which reduces to $\overline{\psi}(u_{i1}, u_{i2}; \phi)t_2 = \widetilde{\psi}(u_{i1}, u_{i2}; \phi)t_1$. Calculus gives that the function $\phi \mapsto x_1(z; \phi)$ is strictly increasing, $x_1(z; 0+) = 0$, and $x_1(z; \infty) = 1$. The equation for $\phi$ has therefore at most $n$ discontinuities, and these can be found explicitly. The roots $\phi$ can afterwards be found numerically in each of the $m + 1$ intervals given by the $m \leq n$ discontinuities. Substitution gives then the roots $\theta_r(u, t)$.

The weight is computed from $w_t(u) = \sum_r \pi(\theta_r)/J(u, \theta_r)$, and this is also needed for the determination of $p_r$. The density $\pi$ should be chosen such that the sampling of $V$ is simplified. The choice $\pi(\mu, \phi) = \mu$ removes the dependence on $\mu_r$. This follows since the required Jacobi determinant $J = |\partial_\theta \tau(u, \theta)|$ can be calculated analytically, and is given by

$$J = J(u, \theta) = \mu t_2 \left| t_2 t_1 \overline{\left( \frac{\psi_\phi}{\psi^2} \right)} - \overline{\psi_\phi} \right| \tag{17}$$

This describes the sufficient conditional Monte Carlo applied to the case of the inverse Gaussian distribution. Further details, implementation, and testing of the algorithm is not given since there exist a much simpler method for conditional simulation in this case. This simple method was discovered by Cheng (1984). It can be mentioned that in Science Citation Index there are only three references to this important discovery. The conditional density of $X_1$ is also explicitly known (Seshadri 1999, p.19)(Chhikara & Folks 1989), and this means in particular that most UMVU estimators can be computed easily numerically. In conclusion this means that the inverse Gaussian can be used as a non-trivial test case for a general numerical implementation of the sufficient conditional Monte Carlo method.

# References


Box, G. & Muller, M. (1958), 'A note on the generation of random normal deviates', *Ann. Math. Statist.* **29**, 610–11.

Cheng, R. (1984), 'Generation of inverse Gaussian variates with given sample mean and dispersion', *Applied Statistics* **33**(3), 309–16.

Chhikara, R. & Folks, J. (1989), *The inverse Gaussian distribution*, Marcel Dekker Inc.

Evans, M. & Swartz, T. (2000), *Approximating Integrals via Monte Carlo and Deterministic Methods*, Oxford.

Fisher, R. (1920), 'A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error', *Monthly Notices Roy. Astronom. Soc.* **80**, 758–770.

Halmos, P. & Savage, L. (1949), 'Application of the Radon-Nikodym theorem to the theory of sufficient statistics', *Ann. Math. Statist.* **20**, 225–241.

Johnsen, N. et al. (1994), *Continuous Univariate Discrete Distributions*, Wiley.

Lehmann, E. & Casella, G. (1998), *Theory of Point Estimation*, Springer.

Lehmann, E. L. (1997), *Testing statistical hypotheses*, Springer (second edition).

Letac, G. & Seshadri, V. (1985), 'On Khatri's characterization of the inverse Gaussian distribution', *Canadian Journal of Statistics* **13**(3), 249–52.

Lindqvist, B. & Taraldsen, G. (2005), 'Monte Carlo conditioning on a sufficient statistic', *Biometrika* **92**(2), xx–xx.

Michael, J., Schucany, W. & Haas, R. (1976), 'Generating random variates using transformations with multiple roots', *Amer.Statistician* **30**(2), 88–90.

Rao, C. (1992), 'R.A. Fisher: The Founder of Modern Statistics', *Statistical Science* **7**(1), 34–48.

Ripley, B. (1987), *Stochastic simulation*, Wiley.

Rudin, W. (1987), *Real and Complex Analysis*, third edn, McGraw-Hill, USA.

Savage, L. (1976), 'On Rereading R.A. Fisher', *Ann. Statist.* **4**(3), 441–500.

Schrödinger, E. (1915), 'Zür Theorie der Fall - und Steigversuche an Teilchenn mit Brownsche Bewegung', *Phys. Zeit.* **16**, 289–295.

Seshadri, V. (1993), *The Inverse Gaussian Distribution*, Oxford.

Seshadri, V. (1999), *The inverse Gaussian distribution*, Springer.

Shuster, J. (1968), 'On the inverse Gaussian distribution function', *J.Amer.Statist.Assoc.* .

Smoluchowsky, M. (1915), 'Notiz über die Berechning der Brownschen Molkular-bewegung bei des Ehrenhaft-millikanchen Versuchsanordnung', *Phys. Zeit.* **16**, 318–321.

Welsh, A. (1996), *Aspects of Statistical Inference*, Wiley.