Stat Comput (2017) 27:439–448
DOI 10.1007/s11222-016-9631-8

A Laplace transform inversion method for probability distribution functions

Stephen G. Walker[1]

Received: 9 September 2015 / Accepted: 22 January 2016 / Published online: 3 February 2016

Abstract This paper introduces a new Laplace transform inversion method designed specifically for when the target function is a probability distribution function. In particular, we use fixed point theory and Mann type iterative algorithms to provide a means by which to estimate and sample from the target probability distribution.

Keywords Recursive estimation · Fixed point solution

1 Introduction

There are a number of settings in statistics and probability (e.g. queueing theory) in which a distribution function, say on $(0, \infty)$ and which we will write as $G(t)$, is not available in closed form and is therefore difficult to study and also, as a consequence, difficult to sample from. On the other hand, in many such cases the Laplace transform has a simple form (for example, infinitely divisible distributions). One such case in point is the positive stable distribution, for which the Laplace transform is given by

$$\mathcal{L}(G)(x) = \int_0^\infty e^{-tx}\, \mathrm{d}G(t) = F(x) = \exp(-\xi\, x^\alpha)$$

for some $\xi > 0$ and $0 < \alpha < 1$. This problem has motivated researchers to introduce strategies to sample from $G(t)$ with only knowledge of $F(x)$. See for example, Devroye (1981), Devroye (1986) and Ridout (2009). The article is further motivated by the fact that there appears no current Laplace inversion method which is suited to the target being a probability distribution function. This means that popular inversion methods, such as the Gaver–Stehfest, are not even guaranteed to provide a numerical estimate which is a distribution function.

On the other hand, the inverse representation found in this paper is based on the weighted sum of a series of gamma distribution functions, and hence will always be a distribution function, rather than the more usual sum involving orthonormal basis functions, which are not ideal ways of representing a probability distribution function. If our algorithm is started at a distribution function, the iterative estimation procedure preserves the estimate as a distribution function. Our method is rooted in fixed point theory and Mann type iterative algorithms.

To this end, let $g(t)$ be a real-valued piecewise continuous function on $[0, \infty)$ which is integrable and non-negative. Under such a scenario it can be assumed that $g(t)$ is a density function. The Laplace transform of $g(t)$ (also referred to as the Laplace–Stieltjes transform) is defined as

$$F(x) = \int_0^\infty e^{-xt}\, g(t)\, \mathrm{d}t$$

for $x \geq 0$. We will also write $\mathrm{d}G(t) = g(t)\,\mathrm{d}t$, where $G(t)$ is the distribution function corresponding to $g(t)$.

With $g(t)$ as a density function then $F(0) = 1$, $F(x)$ is non-increasing, and $F(\infty) = 0$. Furthermore, $(-1)^n F^{(n)}(x) \geq 0$ for all $n \geq 0$ and

$$p_n(x) = (-1)^n \frac{x^{n-1}}{\Gamma(n)} F^{(n)}(x)$$

is a density function for all $n \geq 1$. This can easily be seen to be

$$p_n(x) = \int_0^\infty \mathrm{gamma}(x|n, t)\, g(t)\, \mathrm{d}t.$$

✉ Stephen G. Walker
s.g.walker@math.utexas.edu

[1] Department of Mathematics, University of Texas, Austin, USA

We will be making use of $p_1(x)$, which from now on we will write as $p(x)$, and then we have and will write gamma$(x|n = 1, t) = k(x|t) = t\,e^{-xt}$.

The aim in this paper is to estimate $g(t)$ using samples from $p(x)$ and that the theoretical underpinning of the approach is based on a weak convergence argument developed in Sects. 3 and 5.

In Sect. 2 we give a review of current inversion methods; in Sect. 3 we show how the solution G can be seen as a fixed point for some function which leads to a random Mann type iterative algorithm. Section 4 contains some illustrations, where we compare with the Stehfest and trapezoidal Bromwich methods, and Sect. 5 contains the convergence theorem for the algorithm.

2 Current inversion methods

There are a number of key inversion procedures; one is the widely used Bromwich contour integral given by

$$g(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{tx}\, F(x)\, \mathrm{d}x \tag{1}$$

where c is such that $F(x)$ is analytic for $x \geq c$. Non-contour integral versions of this are given in Berberan-Santos (2005), for example, though non-trivial integrations are still required to get at $g(t)$. According to Abate and Whitt (2006), forms for $g_n(t)$ are given by

$$g_n(t) = \frac{1}{t} \sum_{k=0}^{n} \omega_k\, F\left(\frac{\alpha_k}{t}\right) \tag{2}$$

for parameters (ω_k, α_k) which depend on n but not on F, and Abate and Whitt (2006) describe a number of ideas for determining these parameters, including a Fourier series approach and the Talbot method (Talbot 1979). One of the most popular methods, and which we will use for comparison later, is the Stehfest method, which has $\alpha_k = k \log 2$, $\omega_0 = 0$ and for $k \geq 1$,

$$\omega_k = \log(2)\,(-1)^{k+n/2} \times \sum_{l=\lfloor (k+1)/2 \rfloor}^{\min\{k,n/2\}} \frac{l^{n/2}\,(2l)!}{(n/2-l)!\,l!\,(l-1)!\,(k-l)!\,(2l-k)!}. \tag{3}$$

Convergence is described in Kuznetsov (2103). Popularity arises since only real calculations are required.

A trapezoidal rule applied to (1) is described in detail in Abate et al. (2000). Here,

$$g_n(t) = \sum_{k=0}^{n} (-1)^k\, a_k(t) \tag{4}$$

where $a_k(t) = \frac{1}{2}\, b_k(t)\, e^{A/2}/t$ and

$$b_0(t) = F\left(\frac{1}{2}A/t\right) + 2\,\mathrm{Re}\left[F\left(\frac{1}{2}A/t + i\pi/t\right)\, e^{i\pi}\right]$$

with, for $k \geq 1$,

$$b_k(t) = 2\,\mathrm{Re}\left[F\left(\frac{1}{2}A/t + i\pi/t + ik\pi/t\right)\right].$$

The $A > 0$ is a tuning parameter for accuracy.

Another inversion idea (see Widder 1946) is based on the limits of derivatives of $F(x)$; i.e. if

$$g_n(t) = \frac{(-1)^n}{n!} \left(\frac{n+1}{t}\right)^{n+1} F^{(n)}((n+1)/t) \tag{5}$$

then $g_n(t) \longrightarrow g(t)$ pointwise. This is not useful in practice due to the need to derive the derivatives; however, one such algorithm has appeared in Abate and Whitt (1995).

Other ideas are based on representing $g(t)$ by a series of functions $(\phi_k(t))$; i.e.

$$g(t) = \sum_{k=0}^{\infty} \omega_k\, \phi_k(t),$$

and these methods include those of Weeks, which is based on the Laguerre polynomials. For example, the Papoulis method (Papoulis 1957) has

$$g_n(t) = \sum_{k=0}^{n} \omega_k\, P_{2k}(e^{-bt})$$

where P_k is the Legendre polynomial, and ω_k is obtained from the recursive formula

$$b\,F[(2k+1)b] = \sum_{m=0}^{k} \omega_m\, \frac{(k-m+1)_m}{2\left(k+\frac{1}{2}\right)_{m+1}},$$

and $(\gamma)_m$ is the Pochhammer symbol; i.e. $(\gamma)_0 = 1$ and, for $m > 0$, $(\gamma)_m = \gamma(\gamma+1)\cdots(\gamma+m-1)$, with $b > 0$ chosen arbitrarily.

In Abate and Whitt (1995), the authors are primarily concerned with $g(t)$ a density function, however none of these inversion methods is entirely suitable for $g(t)$ as a density function. The point, and we will observe this in the illustration section, is that there is no guarantee the inversion leads to a distribution function. For recent summaries, see Abate and Whitt (2006), Davis (2005), Kuhlman (2013) and Cohen (2007).

A related problem to Laplace inversion is concerned with sampling from $g(t)$, for which it is essential the inversion method provides a distribution function, and the first papers in this direction appear in Devroye (1981) and Devroye (1986). The paper Devroye (1981) has some restrictions put on $F(x)$ and the paper Devroye (1986) is somewhat complicated, relies on $g(t)$ to be bounded and uses an automated rejection algorithm for which acceptance probabilities would be unknown. More recently Ridout (2009) has proposed a sampling algorithm which also uses standard inversion methods.

On the other hand, in this paper we find a form for $g_n(t)$ which not only provides us with an interpretation as a mixture of densities, but also a means by which to sample approximately from $g(t)$. This is achieved without recourse to either (1) or (5) or one of the series expansions using basis functions. In this respect we are presenting a new inverse Laplace transform method. The only technical aspect of the algorithm is sampling from the density function

$$p(x) = -F'(x). \tag{6}$$

Since $F(x)$ is available explicitly, so will $p(x)$ be, and sampling a one dimensional density function with closed form should not cause any problems. There are by now many techniques; see, for example, Devroye (1986).

3 New inversion technique

Suppose $(x_n)_{n\geq 1}$ is an independent and identically distributed sample from $p(x)$. Now

$$p(x) = \int_0^\infty t\, e^{-xt}\, g(t)\,\mathrm{d}t, \tag{7}$$

which in statistics is known as a mixture model, since $k(x|t) = t\exp(-xt)$ is a density function on $(0,\infty)$ for all $t > 0$, and there are a number of estimation procedures for $g(t)$ which are consistent having observed a sample from $p(x)$. By consistent we mean that as the sample size n grows to infinity, so the estimator of $G(t)$; i.e. $G_n(t)$, converges to $G(t)$, in the weak sense.

The nonparametric maximum likelihood estimator of $g(t)$ based on a sample of size n is detailed in Jewell (1982). However, this is not fast since there is a need for an iterative algorithm to find $g_n(t)$, which is discrete with an unknown a priori number of atoms. The number of atoms needs to be checked off and so up to n algorithms need to be run in order to find the optimal number of atoms.

A fast algorithm which has a Bayesian interpretation, of no direct concern since it is the consistency property which is relevant, is as follows: for some arbitrary $g_0(t)$, obtain $g_n(t)$ from $g_{n-1}(t)$, with some deterministic sequence (w_n) for which $w_n \to 0$, as

$$g_n(t) = g_{n-1}(t)\left\{1 - w_n + w_n \frac{t\, e^{-x_n t}}{\int_0^\infty t e^{-x_n t}\, g_{n-1}(t)\,\mathrm{d}t}\right\}. \tag{8}$$

See Newton (2002) for the origins and motivation for this recursive estimation procedure. This is also a Mann type iterative algorithm (Mann 1953) but with stochastic terms. It is not connected with random fixed point theory (see, for example, Duan and Li 2005) since our fixed point $g(t)$ is non-random. In fact, (8) is, to my knowledge, only known in the statistics literature, yet the connection with fixed points has not been utilized though is known; see Eq. (7) in Newton (2002), Sect. 3.2 in Martin and Ghosh (2008), and Remark 4.2 in Martin and Tokdar (2009).

The implementation involves out of necessity evaluating $(g_n(t))$ at a set of fixed points $(t_j)_{j=1}^M$ for some large M; also necessary for all numerical computational strategies, and used by Newton (2002) and others. Hence, from (8), we have, for all $j = 1,\ldots,M$,

$$g_n(t_j) = g_{n-1}(t_j)\left\{1 - w_n + w_n \frac{t_j\, e^{-x_n t_j}}{\int_0^\infty t e^{-x_n t}\, g_{n-1}(t)\,\mathrm{d}t}\right\}.$$

To complete the algorithm we need to be able to evaluate (approximately) the integrals

$$I_n = \int_0^\infty t e^{-x_n t}\, g_{n-1}(t)\,\mathrm{d}t$$

using the values of $(g_{n-1}(t_j))$. There are possibilities; the one we will use is based on the linear interpolation between $\xi_{j-1,n}$ and $\xi_{j,n}$, where $\xi_{j,n} = t_j e^{-x_n t_j}\, g_{n-1}(t_j)$; i.e. the trapezoidal rule. The assumption is that t_M works as an upper limit to the integration, with $\int_{t_M}^\infty g(t)\,\mathrm{d}t$ being suitably small. This might need to be evaluated via some preliminary analyses. Thus we approximate

$$\widehat{I}_n = \frac{1}{2}\sum_{j=1}^M (t_j - t_{j-1})\,(\xi_{j-1,n} + \xi_{j,n}), \tag{9}$$

and hence we compute recursively

$$g_n(t_j) = g_{n-1}(t_j)\left\{1 - w_n + w_n \frac{t_j e^{-x_n t_j}}{\widehat{I}_n}\right\} \tag{10}$$

for $j = 1,\ldots,M$.

We can avoid the density estimate $g_n(t)$ and go directly to the distribution $G_n(t)$ by updating

$$G_n(t_j) = (1 - w_n)G_{n-1}(t_i) + w_n \frac{\int_0^{t_j} s e^{-s x_n}\,\mathrm{d}G_{n-1}(s)}{\int_0^\infty s e^{-s x_n}\,\mathrm{d}G_{n-1}(s)}.$$

The integrals in the last term can be written, using integration by parts, as

$$\int_{t_{j-1}}^{t_j} s e^{-s x_n}\, \mathrm{d}G_{n-1}(s) = s e^{-s x_n} G_{n-1}(s)\Big|_{t_{j-1}}^{t_j} + \int_{t_{j-1}}^{t_j} (s x_n - 1) e^{-s x_n} G_{n-1}(s)\, \mathrm{d}s$$

so we can implement a trapezoidal rule using only $G_{n-1}(t_j)$ values.

The value of $g_n(t)$ for any t can be approximated via linear interpolation; i.e.

$$g_n(t) = g_n(t_{j-1}) + \frac{t - t_{j-1}}{t_j - t_{j-1}} \left[g_n(t_j) - g_n(t_{j-1})\right]$$

and so also can $G_n(t)$. This evaluation of $G_n(t)$ for all t allows us to sample approximately from $g(t)$ since we can take a uniform random variable u from $(0, 1)$ and take the sample random variable $t = G_n^{-1}(u)$ with $G_n^{-1}(u)$ being easy to compute since $G_n(t)$ is piecewise quadratic.

4 Illustrations

4.1 The algorithm

Before proceeding with the illustrations, we put in bullet point format the algorithm going from stage $n-1$ to n:

- Start with $(g_{n-1}(t_j))$
- Take x_n from $p(x)$
- Compute $\widehat{I}_n$ using (9)
- Compute

$$g_n(t_j) = g_{n-1}(t_j) \left\{1 - w_n + w_n \frac{t_j e^{-x_n t_j}}{\widehat{I}_n}\right\}$$

for $j = 1, \ldots, M$.

The algorithm runs in $n\,M$ time, while the process needs no more computing complexity than implementing a trapezoidal rule. We use a substantial amount of samples in the recursive method, of the order 10^4, though the running time is not overly long; typically a couple of minutes. All algorithms are coded using Scilab[1].

In the following examples we will be making a comparison with the Stehfest method, described in the Introduction, and using the recommended choice of $n = 18$; see (2) and (3).

[1] An open source software for numerical computation, www.scilab.org.

4.2 Illustrations

Example 1 We start by considering the case when $g(t)$ is the gamma density with density function given by

$$g(t) = \frac{b^a}{\Gamma(a)} t^{a-1} e^{-tb}$$

for $a > 0$ and $b > 0$. Then

$$F(x) = \left(\frac{b}{b+x}\right)^a$$

and a sample x from $p(x)$, a Pareto density function, can be obtained via $x = b\left(u^{-1/a} - 1\right)$ where u is a uniform random variable from the interval $(0, 1)$. For high accuracy we take $n = 10000$ and $M = 1000$ with $t_j = 0.003\,j$, $t_0 = 0$, $t_M = 3$ and $w_n = (n+1)^{-0.7}$. We take $g_0(t) = \exp(-t)$ and the true $g(t)$ has parameters $a = 3/2$ and $b = 3$.

We ran the algorithm and compute the $G_n(t)$ from the recursive algorithm. We plot this function in Fig. 1 and is compared with the true distribution function $G(t)$. In Table 1 we compare the estimated values of $G(t)$ with the correct values for a various selection of t values. We also include the estimated values from the Stehfest method. The recursive approach is clearly on this evidence competitive with a popular method; though it does take longer to run.

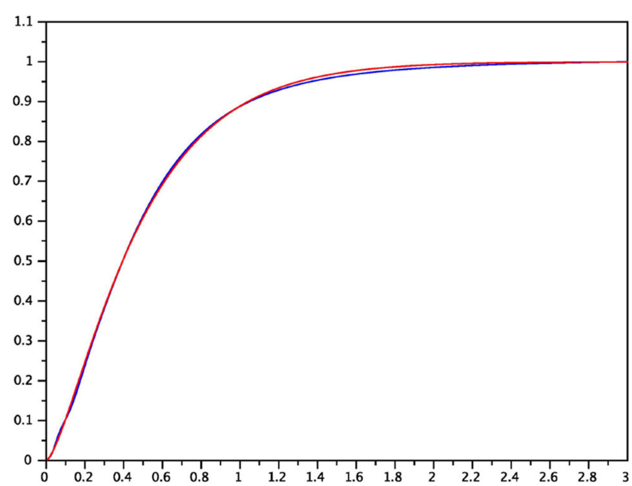

Fig. 1 *Example* 1 Comparison of $G_n(t)$ (*blue*) and true $G(t)$ which is the Ga(3/2, 3) distribution (*red*) . (Color figure online)

Table 1 *Example* 1 Estimated $G(t)$ for Stehfest and Recursive methods, alongside true values, for various values of t

	$t = 0.3$	t=0.6	$t = 0.75$	$t = 1.5$	$t = 2.1$
True	0.385	0.692	0.788	0.971	0.994
Stehfest	0.385	0.692	0.788	0.971	0.994
Recursive	0.388	0.689	0.787	0.963	0.989

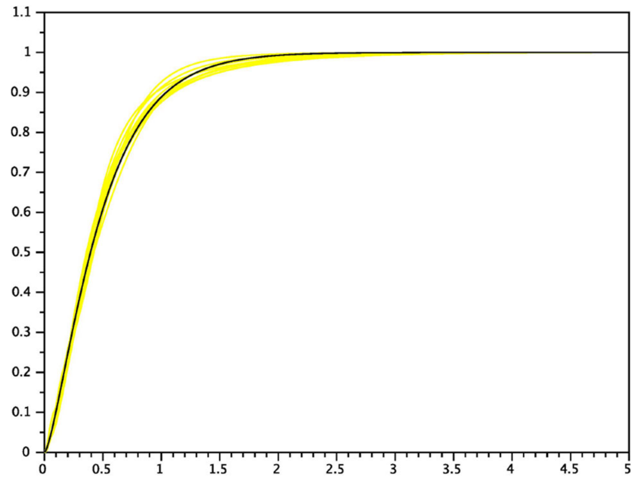

Fig. 2 *Example* 1 Comparison of 10 simulations of $G_n(t)$ (*light colour*) and true $G(t)$ which is the Ga(3/2, 3) distribution (*black*). (Color figure online)

As suggested by a referee, a representation of the variability of the estimate, based on a sample of size $n = 5000$ and repeated 10 times, is given in Fig. 2, the approximations given in the light colour and the true distribution in black.

In this example the density estimate of $g(t)$ is unstable and wobbles near the origin, though the tail is smooth and fairly accurate. A referee has suggested a solution to this might be to take the density estimate as an average over a number of runs.

Example 2 Next we consider a more challenging example used by Ridout (2009), namely the positive stable distribution. Here

$$F(x) = \exp\left(-bx^a\right)$$

with $0 < a < 1$ and $b = \gamma^a / \cos(a\pi/2) > 0$ for some $\gamma > 0$. So $p(x)$ is a Weibull density function and easy to sample; if u is uniform from (0, 1) then $x = \{(-\log u)/b\}^{1/a}$. We use similar settings as before; i.e. $n = 10,000$ and $M = 1000$ with (after some preliminary investigations on a suitable t_M) $t_j = 0.15j$. Since $g(t)$ has heavy tails, we start with $G_0(t) = t/(1+t)$. We use the same (w_n) as in Example 1 and the stable parameters are given by $a = 0.9$ and $\gamma = 1$, as in Ridout (2009).

We plot the estimate of the density $g(t)$ in Fig. 3. The median, as given in Ridout (2009), is 6.9; whereas the median of the recursive approximation $g_n(t)$ is 7.5. This is quite reasonable given the scale of the distribution; i.e. the 99 % percentile is 116. We also, to see the variability of the density estimate, simulate 10 copies, which are presented in Fig. 4.

Unfortunately, for this example, the Stehfest method gives an estimate for $g(t)$ which is not a density function, in that it takes negative values. See Fig. 5. This problem is not resolved

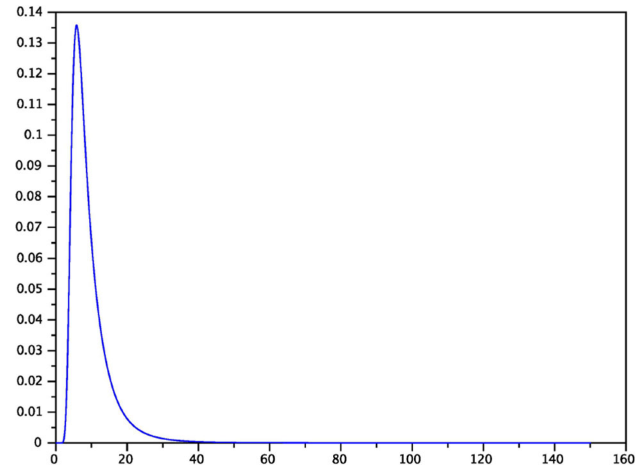

Fig. 3 *Example* 2 Recursive estimated $g(t)$ for the positive stable distribution

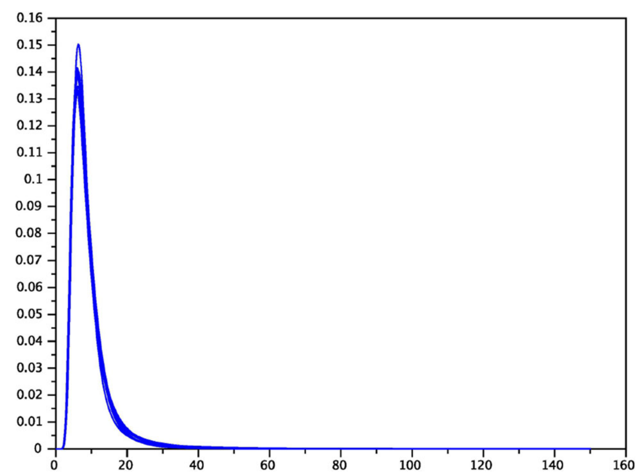

Fig. 4 *Example* 2 10 recursive estimated $g(t)$ for the positive stable distribution

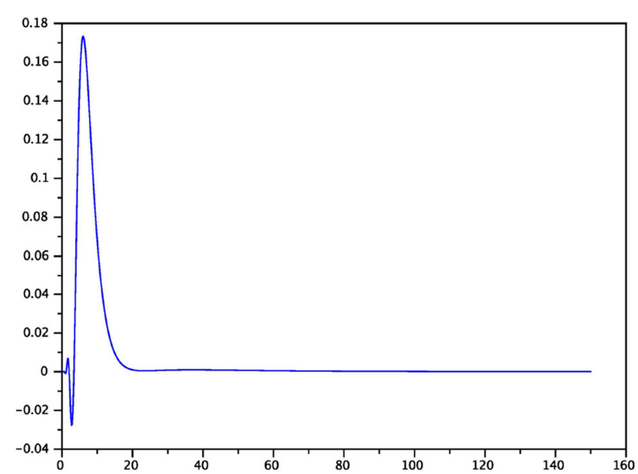

Fig. 5 *Example* 2 Stehfest estimated $g(t)$ for the positive stable distribution

by changing the $n = 18$ to other values; e.g. $n = 10$ and $n = 25$ were tried. The Bromwich trapezoidal method, (4), works well, with $A = 10$, and the median is estimated as 6.9, the correct value.

Example 3 Here we consider $g(t)$ as the uniform density on (0, 1) so

$$F(x) = \frac{1}{x}\left(1 - e^{-x}\right).$$

Then $p(x) = -F'(x)$ can be sampled using a rejection algorithm, for example. In this example we take $n = 10,000$, $M = 1000$ and $t_j = 0.001\,j$. The estimated $G(t) = t$ is given in Fig. 6, with the recursive estimate shown alongside the Stehfest estimate. As can be seen, the Stehfest method again fails to deliver a distribution function. Table 2 gives the estimates for the two methods for varying values of t and demonstrates the precision of the recursive approach. Also note the lack of the Stehfest method to provide the value 1 at $t = 1$.

The trapezoidal Bromwich approach (4) again solves the problem of the Stehfest method and provides a very accurate estimate of $G(t) = t$; see Table 2.

Example 4 To test the trapezoidal Bromwich method, we now take $g(t)$ to be a point mass at $t = 1$. Hence $F(x) = \exp(-x)$. We run the trapezoidal method with $n = 1000$ with $A = 10$ and we plot the estimate of $G(t) = 0$ for $t < 1$ and $G(t) = 1$ for $t > 1$ in Fig. 7. This is clearly not a distribution function and note the value of about 2 for $G_n(t)$ for $t > 1$. Other settings for A and n also yield problems for the method.

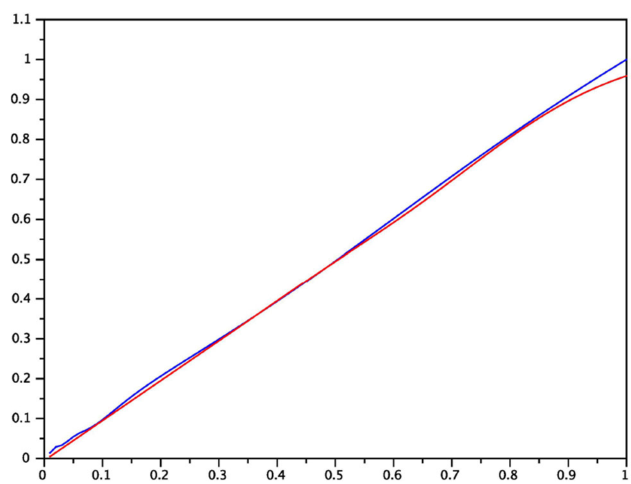

Fig. 6 *Example* 3 Comparison of recursive estimation (*blue*) and Stehfest estimation (*red*) of $G(t) = t$ distribution. (Color figure online)

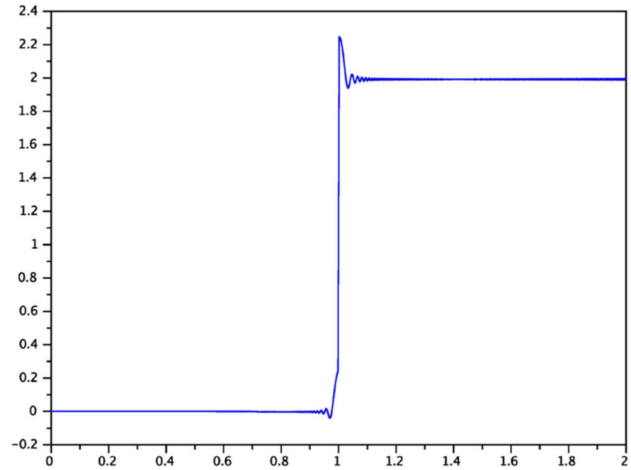

Fig. 7 *Example* 4 Trapezoidal Bromwich estimate of distribution function corresponding to $g(t)$ a point mass at 1

However, noting that there is a discontinuity at $t = 1$ from the trapezoidal Bromwich estimate, we ran the recursive algorithm with $n = 1000$, $M = 1000$ and starting at $G_0(t_j) = 0.4$ for $t_j < 1$ and $G_0(t_j) = 0.6$ for $t_j \geq 1$, with $t_j = 0.002\,j$. So we are acknowledging the information provided by the trapezoidal Bromwich approach to get a good starting point for the recursive method. It can not be deduced what the $G(t)$ is from the trapezoidal Bromwich estimate and hence the need to run the recursive algorithm, which converges to a distribution function. The estimate of $G(t)$ in this case is given in Fig. 8 and is exactly $G(t)$.

We are reliant on knowing the points of discontinuity. For if we start with a discontinuity at the wrong place, even close to the correct one, the estimate will not find the correct discontinuity. See also Sakurai (2004).

5 Convergence of (8)

The consistency of $g_n(t)$ to $g(t)$ with respect to weak convergence and the L_1 metric is given in Tokdar et al. (2009),

Table 2 *Example* 3 Estimated $G(t) = t$ for Stehfest, Trapezoidal Bromwich and Recursive methods, alongside true values, for various values of t

	$t = 0.1$	$t = 0.2$	$t = 0.3$	$t = 0.4$	$t = 0.5$	$t = 0.6$	$t = 0.7$	$t = 0.8$	$t = 0.9$	$t = 1$
True	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Stehfest	0.095	0.195	0.295	0.395	0.495	0.592	0.697	0.805	0.896	0.959
Trap. Bromwich	0.099	0.199	0.299	0.399	0.499	0.599	0.699	0.799	0.899	0.999
Recursive	0.097	0.206	0.299	0.394	0.496	0.602	0.708	0.811	0.908	1

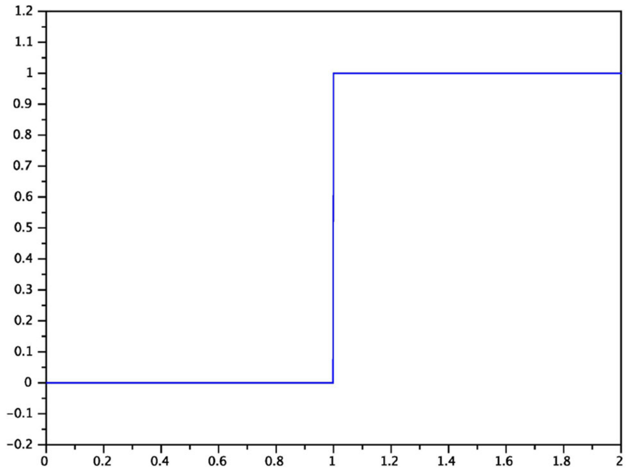

Fig. 8 *Example 4* Recursive estimate of distribution function corresponding to $g(t)$ a point mass at 1. Note the discontinuity at $t = 1$ is acknowledged from the trapezoidal Bromwich method and used as a starting point for the recursive algorithm

Martin and Ghosh (2008) and Martin and Tokdar (2009), where the key condition on the weights (w_n) is that

$$\sum_n w_n = \infty \quad \text{and} \quad \sum_n w_n^2 < \infty.$$

Moreover, for w_n behaving as $n^{-\gamma}$ for $\gamma \in (2/3, 1]$, Martin and Tokdar (2009) show that the L_1 distance between $g_n(t)$ and $g(t)$ goes to zero a.s. at the rate $n^{-\frac{1}{2}(1-\gamma)}$. That is,

$$\int_0^\infty |g_n(t) - g(t)|\, \mathrm{d}t \le c\, n^{-\frac{1}{2}(1-\gamma)} \quad \text{a.s.}$$

for all large n, for some constant $c > 0$. However, this result only holds for some particular $k(x|t)$ and the exponential density, which is the one we are effectively working with, is not included. Hence, in this section, we present a proof of weak convergence.

We start with some lemmas which are needed to complete the proof of convergence. In the following $\mathrm{Ga}(t|a, b)$ denotes the gamma density function with density proportional to $t^{a-1} \exp(-bt)$.

Lemma 1 *If $g(t)$ is a density function on $(0, \infty)$ and $g_0(t)$ is a gamma distribution with parameters $a_{1,0}$ and $b_{1,0}$, then, for $n \ge 1$,*

$$g_n(t) = \sum_{j=1}^{2^n} q_{j,n} Ga(t|a_{j,n}, b_{j,n}),$$

where, recursively, for $j = 1, \ldots, 2^{n-1}$,

$$q_{j,n} = (1 - w_n)\, q_{j,n-1}, \quad a_{j,n} = a_{j,n-1} \quad \textit{and} \quad b_{j,n} = b_{j,n-1}$$

with

$$a_{j+2^{n-1},n} = a_{j,n-1} + 1 \quad \textit{and} \quad b_{j+2^{n-1},n} = b_{j,n-1} + x_n$$

and

$$q_{j+2^{n-1},n} = w_n \frac{q_{j,n-1}\, a_{j,n-1}\, b_{j,n-1}^{a_{j,n-1}} / (b_{j,n-1} + x_n)^{a_{j,n-1}+1}}{\sum_{j=1}^{2^{n-1}} q_{j,n-1}\, a_{j,n-1}\, b_{j,n-1}^{a_{j,n-1}} / (b_{j,n-1} + x_n)^{a_{j,n-1}+1}}.$$

The proof of this is based on (8), and a similar expression appears in Ghosh and Tokdar (2005). Note the suitability for $g(t)$ as a density since $g_n(t)$ is a mixture of gamma distributions. For a further exposition of the gamma components, let S_n be the set of subsets of $\{1, \ldots, n\}$ including the empty set. So $|S_n| = 2^n$.

Lemma 2 *It is that*

$$g_n(t) = \sum_{A \in S_n} q_{n,A}\, Ga\left(t \,\middle|\, a_{1,0} + |A|, b_{1,0} + \sum_{i \in A} x_i\right), \tag{11}$$

for some probability $(q_{n,A})_{A \in S_n}$ which depends on the (x_i). Recursively,

$$q_{n+1,A} = (1 - w_{n+1}) q_{n,A} \quad \textit{and} \quad q_{n+1,A \cup \{n+1\}} = w_{n+1} \frac{q_{n,A}\, \phi(x_{n+1}, A)}{\sum_{A \in S_n} q_{n,A}\, \phi(x_{n+1}, A)},$$

where

$$\phi(x, A) = (a_{1,0} + |A|) \left(b_{1,0} + \sum_{i \in A} x_i\right)^{a_{1,0}+|A|} \times \left(b_{1,0} + \sum_{i \in A} x_i + x\right)^{-a_{1,0}-|A|-1}.$$

The aim now is to prove that G_n converges weakly to G^* with probability one based on the iterative algorithm given in (8). To this end, we first consider the sequence (y_n) given as

$$y_n = (1 - w_n) y_{n-1} + w_n \eta(y_{n-1}, x_n)$$

where the (x_n) are an i.i.d. sequence, $0 \le y_n \le 1$ and $0 \le \eta \le 1$ is a continuous function from $[0, 1] \times (0, \infty)$ to $[0, 1]$. Also

$$\sum_n w_n = \infty \quad \text{and} \quad \sum_n w_n^2 < \infty$$

and define $\eta(y) = E\,[\eta(y, x)]$ and assume $\eta(y^*) = y^*$ only at y^*.

Theorem 1 *It is that $y_n \to y^*$ almost surely.*

Proof Consider the martingale difference sequence

$$M_n = \eta(y_{n-1}, x_n) - \eta(y_{n-1}).$$

So by the martingale convergence theorem it is that $\sum_n w_n M_n$ converges to a finite random variable a.s. Now write

$$y_n = y_{n-1} + w_n[\eta(y_{n-1}) - y_{n-1}] + w_n M_n$$

so

$$y_N = y_0 + \sum_{n=1}^{N} w_n[\eta(y_{n-1}) - y_{n-1}] + \sum_{n=1}^{N} w_n M_n.$$

Since y_N and y_0 are bounded between 0 and 1 and that $\sum_{n=1}^{\infty} w_n M_n$ converges, it must be that

$$\sum_{n=1}^{N} w_n[\eta(y_{n-1}) - y_{n-1}]$$

does not diverge. Since $\sum_n w_n = \infty$ this implies that either $\eta(y_n) - y_n$ converges to 0; i.e. y_n converges to y^*, or $\eta(y_n) - y_n$ must change sign infinitely often. The former gives us $y_n \to y^*$ and the latter, since $y_n - y_{n-1} \to 0$, gives us that y^* must be an accumulation point of (y_n) and so there exists a sequence n' such that $y_{n'} \to y^*$.

Since $\sum_n w_n^2 < \infty$ we have

$$E\left[\sum_n (y_n - y_{n-1})^2\right] < \infty.$$

Hence,

$$\sum_n (y_n - y_{n-1})^2 < \infty \quad \text{a.s.}$$

and because

$$(y_l - y_m)^2 \leq \sum_{n=m+1}^{l} (y_n - y_{n-1})^2$$

it is that $(y_l - y_m)^2 < \epsilon$ a.s. for all large l, m for any $\epsilon > 0$. Hence, $y_n \to \widetilde{y}$ a.s. for some $\widetilde{y} \in [0, 1]$ (Cauchy) and since a subsequence converges to y^* it must be that $\widetilde{y} = y^*$. □

Now consider the problem when $y_n = G_n(t)$ and so we need to reconsider the above problem for every t, where

$$G_n(t) = (1 - w_n)\, G_{n-1}(t) + w_n \eta(t, G_{n-1}, x_n)$$

and

$$\eta(t, G, x) = \frac{\int_0^t s\, e^{-sx}\, \mathrm{d}G(s)}{\int_0^\infty s\, e^{-sx}\, \mathrm{d}G(s)}.$$

Also define

$$\eta(t, G) = \int_0^\infty \eta(t, G, x)\, p^*(x)\, \mathrm{d}x$$

and recall that

$$p^*(x) = \int_0^\infty t\, e^{-tx}\, \mathrm{d}G^*(t)$$

and note that

$$\int_0^\infty \int_0^t s\, e^{-sx}\, \mathrm{d}G^*(s)\, \mathrm{d}x = G^*(t).$$

Under the assumption that the family of density functions indexed by t,

$$p_{G^*}(x|t) = G^*(t)^{-1} \int_0^t s\, e^{-sx}\, \mathrm{d}G^*(s),$$

is complete, we have for $G^* \ll G$ (G^* is absolutely continuous with respect to G), that $\eta(t, G) = G(t)$ for all t only at $G \equiv G^*$. A referee has suggested an alternative derivation of this result avoiding the completeness of $p_{G^*}(x|t)$: Now, from the uniqueness of Laplace transforms,

$$\int_0^\infty e^{-tx}\, \alpha(x)\, \mathrm{d}x = 0 \quad \forall t \iff \alpha \equiv 0.$$

Thus, using $G^* \ll G$,

$$\int_0^\infty se^{-sx}\, \left[p^*(x)/p(x) - 1\right] \mathrm{d}x = 0 \quad \forall s \iff p^* = p.$$

That is,

$$\int_0^\infty se^{-sx}\, [p^*(x)/p(x)]\, \mathrm{d}x = 1 \quad \forall s \iff p^* = p.$$

Hence, using Fubini, allowed since $\eta(t, G)$ is finite for all G^*, is actually bounded by 1, and has non-negative integrand,

$$\eta(t, G) = \int_0^\infty \int_0^t se^{-sx} g(s)\, \mathrm{d}s [p^*(x)/p(x)]\, \mathrm{d}x = G(t) \quad \forall t \iff p^* = p.$$

Starting with $G^* \ll G_0$, the Cauchy part of the proof for theorem 1 gives us that $G_n(t) \to G_\infty(t)$ pointwise for some function $G_\infty(t)$. The first part of the proof of theorem 1 gives us that $\eta(t, G_n) - G_n(t)$ must change sign infinitely often or converge to 0 for all t. Consequently, $\eta(t, G_\infty) = G_\infty(t)$ for all t, which is the preserve of $G^*(t)$. Therefore, $G_n(t) \to G^*(t)$ pointwise a.s.

Finally, to complete the almost sure weak convergence of G_n to G^* we require the *tightness* of the probability measures governing G_n; write these as Π_n. Since $G_n(0) = 0$ for all n we are left with showing that

$$H = \lim_{\delta\to 0} \limsup_n \Pi_n(w_T(G_n, \delta) > \epsilon) = 0 \tag{12}$$

for all $\epsilon > 0$ and $T < \infty$, where

$$w_T(G, \delta) = \sup_{t,s<T,\, |t-s|<\delta} \{|G(t) - G(s)|\}.$$

See Billingsley (1999). From (11) we have that $w_T(G_n, \delta) \le c\,\delta \sup_{t<T}\{g_n(t)\}$ for some constant $c > 0$, and since

$$H \le c\epsilon^{-1} \lim_{\delta\to 0} \limsup_n \delta \,\mathrm{E} \sup_{t<T} g_n(t)$$

from the Markov inequality, we are left with showing that

$$\limsup_n \mathrm{E} \sup_t g_n(t) < \infty.$$

Also, from (11), we have the mode of a gamma density $\mathrm{Ga}(t|a, b)$ at $t = (a-1)/b$ for $a > 1$, and the modal value is

$$\frac{a(a-1)^{(a-1)} \exp(1-a)}{\Gamma(a)} \frac{b}{a}.$$

Now $a(a-1)^{(a-1)} \exp(1-a)/\Gamma(a)$ is bounded by 1 and so taking $a_{1,0} > 1$ we have

$$\begin{aligned} \mathrm{E} \sup_t g_n(t) &\le \sum_{A\in S_n} q_{n,A}\, \mathrm{E}\left\{\frac{b_{1,0} + \sum_{i\in A} x_i}{a_{1,0} - 1 + |A|}\right\} \\ &\le \max\left\{\frac{b_{1,0}}{a_{1,0}-1}, \mu\right\}, \end{aligned}$$

where $\mu = \mathrm{E}(x_i)$ and assumed to be finite. □

6 Discussion

We have provided an iterative scheme for estimating the inverse probability distribution function of a Laplace transform; namely (8). The iterative algorithm has been shown to converge to G^*, with the stochastic Mann type algorithm, summarized at the start of Sect. 4, being easy to implement.

Moreover, and importantly, the recursive method guarantees convergence to a distribution function. The popular Stefhest method does not guarantee a distribution function and this has been demonstrated in a couple of examples, where the deviance from a distribution function is quite noticeable. When the target distribution is smooth, the Stehfest method is fast and accurate. When discontinuities appear, or sharp turns, e.g. near the origin in Example 2, the Stehfest method struggles.

The trapezoidal Bromwich method works very well and the only instance we can find when it runs into trouble is when $g(t)$ is a point mass. Using the estimate it gives however we can start the recursive algorithm at a good place and obtain the required distribution.

Acknowledgements I am very grateful for the comments and suggestions of two anonymous referees on an earlier version of the paper.

References

Abate, J., Whitt, W.: Numerical inversion of Laplace transforms of probability distributions. ORSA J. Comput. **7**, 38–43 (1995)

Abate, J., Choudhury, G.L., Whitt, W.: An introduction to numerical transform inversion and its application to probability models. In: Grassmann, W. (ed.) Computational Probability, pp. 257–323. Kluwer, Dordrecht (2000)

Abate, J., Whitt, W.: A unified framework for numerically inverting Laplace transfroms. INFORMS J. Comput. **18**, 408–421 (2006)

Berberan-Santos, M.N.: Analytical inversion of the Laplace transform without contour integration: application to luminesence decay laws and other relaxation functions. J. Math. Chem. **38**, 165–173 (2005)

Billingsley, P.: Convergence of Probability Measures. Wiley, Chicago (1999)

Cohen, A.M.: Numerical Methods for Laplace Transform Inversion. Springer, New York (2007)

Davis, B.: Integral Transforms and their Applications. Springer, New York (2005)

Devroye, L.: On the computer generation of random variables with a given characteristic function. Comput. Math. Appl. **7**, 547–552 (1981)

Devroye, L.: An automatic method for generating random variables with a given characteristic function. SIAM J. Appl. Math. **46**, 698–719 (1986)

Devroye, L.: Non-uniform Random Variate Generation. Springer, New York (1986)

Duan, H., Li, G.: Random Mann iteration scheme and random fixed point theorems. Appl. Math. Lett. **18**, 109–115 (2005)

Ghosh, J.K., Tokdar, S.T.: Convergence and consistency of Newton's algorithm for estimating a mixing distribution. In: Fan, J., Koul, H.L. (eds.) Frontiers in Statistics, pp. 429–443. Imperial College Press, London (2005)

Jewell, N.P.: Mixtures of exponential distributions. Ann. Stat. **10**, 479–484 (1982)

Kuhlman, K.L.: Review of inverse Laplace transform algorithms for Laplace-space numerical approaches. Numer. Algorithms **63**, 339–355 (2013)

Kuznetsov, A.: On the convergence of the Gaver–Stehfest algorithm. SIAM J. Numer. Anal **51**, 2984–2998 (2103)

Mann, W.R.: Mean value methods in iterations. Proc. Am. Math. Soc. **4**, 506–510 (1953)

Martin, R., Ghosh, J.K.: Stochastic approximation and Newton's estimate of a mixing distribution. Stat. Sci. **23**, 365–382 (2008)

Martin, R., Tokdar, S.T.: Asymptotic properties of predictive recursion: robustness and rate of convergence. Electron. J. Stat. **3**, 1455–1472 (2009)
Newton, M.A.: On a nonparametric recursive estimator of the mixing distribution. Sankhya Ser. A **64**, 306–322 (2002)
Papoulis, A.: A new method of inversion of the Laplace transform. Q. Appl. Math. **14**, 405–414 (1957)
Ridout, M.S.: Generating random numbers from a distribution specified by its Laplace transform. Stat. Comput. **19**, 439–450 (2009)
Sakurai, T.: Numerical inversion for Laplace transforms of functions with discontinuities. Adv. Appl. Probab. **36**, 616–642 (2004)
Talbot, A.: The accurate numerical inversion of Laplace transforms. IMA J. Appl. Math. **23**, 97–120 (1979)
Tokdar, S.T., Martin, R., Ghosh, J.K.: Conistency of a recursive estimate of mixing distributions. Ann. Stat. **37**, 2502–2522 (2009)
Widder, D.V.: The Laplace Transform. Princeton University Press, Princeton (1946)