Esscher Approximations for Maximum Likelihood Estimates - Exploratory Ideas

James G. Bridgeman, FSA, CERA, MAAA
Department of Mathematics, University of Connecticut

September 10, 2018

Abstract

The series expansion of a probability density function known to actuaries by Esscher's name and to statisticians as the saddlepoint approximation typically gets integrated to approximate probability values under the density, with a location parameter in the expansion chosen to optimize convergence of the integrated series. We review the general form, the accuracy and the derivation of the Esscher approximation. Then we propose using related series expansions for the derivative of a probability density function to approximate the value of the point of maximum likelihood for the density. The key is to make the value of the location parameter an unknown and assume it to be the point of maximum likelihood, which leads to a tractable optimization problem to find its value.

1 Introduction

Any probability density function with finite moments admits a series expansion known to actuaries since the 1930's by Esscher's name. See Esscher (1932, 1963), Kahn (1962), Woody (1973), and Gerber and Shiu (1994). Stated in this generality, convergence questions can arise about the expansion for some densities with finite moments that fail to have a moment-generating function. The convergence questions can be addressed for cases involved in practical applications (see section 6.)

Some 20 years after Esscher, statisticians rediscovered the same series expansion independently and called it the saddlepoint approximation. See Daniels (1954), Jensen (1995), and Butler (2007). This paper will call it the Esscher approximation for chauvinistic reasons.

In typical applications, the Esscher expansion for a probability density gets integrated in order to approximate probability values under the density or moments of the density. We can call such an integrated series an Esscher approximation also.

There is a location parameter in the Esscher expansion that can be chosen arbitrarily. One can choose a specific value for the location parameter that greatly improves the convergence of the integrated series, especially for probabilities or moments in the tail of the distribution. This specific choice for the location parameter characterizes most appearances of the Esscher (saddlepoint) approximation in the literature.

A high degree of accuracy, including high relative accuracy in the tails of a distribution, motivates use of the Esscher approximation:

> Saddlepoint approximations, for both density/mass functions and CDF's, are usually extremely accurate over a wide range of x-values and maintain this accuracy far into the tails of the distributions. Often an accuracy of 2 or 3 significant digits in relative error is obtained. (Butler, 2007)

Statistical practitioners find the level of accuracy achieved to be compelling:

> Accordingly, one should always use [the saddlepoint approximation] if it is available. (Jens, 1995)

They also find the level of accuracy achieved to be somewhat unaccountable, even all these years after its introduction:

> Among the various tools that have been developed for use in statistics and probability over the years, perhaps the least understood and most most remarkable tool is the saddlepoint approximation ... remarkable because [accuracy usually is] much greater than current suppporting theory would suggest ... least understood because of the difficulty of the subject itself and ... the research papers and books that have been written about it. (Butler, 2007)

This paper proposes to apply the Esscher expansion in a new direction, namely, to find appproximate values for the point of maximum likelihood on a probability density with finite moments. On the face of it this goal appears to be quixotic if not totally foolish. Intuitively, moments do not point toward likelihoods. On this issue the paper becomes an exploration of possibilities.

We follow two or three trails:

First, we try taking the derivative of the Esscher series expansion for the density. There is no guarantee that the resulting series will converge to the derivative of the original density.

Alternatively, we try to apply the logic of the Esscher expansion to the derivative of the density rather than to the density itself. The resulting series will converge to the derivative of the original density when moments of the original density are finite, at least up to the same practical considerations (see section 6) as for the Esscher approximation for the original density.

If both methods happen to result in convergent expansions we have the opportunity to try a weighted average of the two as a third convergent expansion for the derivative of the density. In this case we have an opportunity to try the question of which of the three expansions (including what weights in the third one) gives optimal convergence.

For each available convergent expansion, a value for the random variable that minimizes the absolute value of a partial sum of the expansion for the derivative of the original density will approximate the maximum likelihood value for the original density. Making the assumption that the arbitrary location parameter (now unkown) in the Esscher expansion equals the unknown maximum likelihood value turns out vastly to simplify this strategy to approximate its value by minimizing the absolute value of a partial sum. A seemingly tractable minimization problem results. We have not yet done the numerical work to validate this assertion.

To lay this all out clearly requires first laying out clearly the underlying Esscher approximation technique, in very general form, and in detail.

- Section 2 will describe the Esscher approximation in very general form
- Section 3 will offer some heuristic insight, without proof, into the impressive accuracy of the Esscher approximation
- Section 4 will derive the general Esscher approximation for a probability density in some detail as a template for later deriving the the corresponding approximation for the derivative of a probability density
- Section 5 will implement the strategy just described to approximate the maximum likelihood value for a probability density
- Section 6 will discuss some practical matters, including convergence issues
- Section 7 will conclude the paper with comments about the circumstances under which these ideas might prove useful

2 General Form of the Esscher Approximation

The Esscher expansion for a probability density f_X of a random variable X can be expressed in two different ways. The first, which we will emphasize, expands primarily around the departure of higher moments (≥ 3) of a distorted version of f_X from the corresponding moments of the standard normal density φ, modulating the difference by Hermite polynomials. The second (equivalent) expression, more usual in the literature, expands primarily in Hermite polynomials, modulated by higher moment differences.

Either form of the Esscher expansion can be used to approximate probability values under the density and moments of the density. To do so requires as a practical matter a way to evaluate certain integrals. The simplest forms for the needed integrals occur in the tails of the density (which also offer the most accurate approximations, see Section 3).

2.1 Esscher Expansion for a Density - Our Form

For any random variable X with finite moments and for an arbitrary location parameter a, the density f_X of X can be expressed as

$$
\begin{aligned}
f_X(x) \quad = \quad & \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \\
& \left\{ 1 + \lim_{N\to\infty} \sum_{j=3}^{N} \frac{1}{j!} \left[\frac{i^j \, \widehat{f_{X-a}}^{(j)}(ih)}{c^j \, \widehat{f_{X-a}}(ih)} - j? \right] \sum_{n=0}^{\left\lfloor \frac{N-j}{2} \right\rfloor} \frac{(-1)^n (2n)?}{(2n)!} H_{2n+j}\left(\frac{x-a}{c}\right) \right\}
\end{aligned} \tag{1}
$$

where:

- $\widehat{f_{X-a}}(t)$ is the Fourier transform $\mathbb{E}\left[e^{-it(X-a))}\right]$ of the density for the random variable $X-a$, its characteristic function evaluated at $-t$
- so $\widehat{f_{X-a}}(ih)$ is the moment generating function of $X-a$ evaluated at a point h described below (what this means if the moment generating function fails to exist will be discussed in Section 6)
- $\varphi(z)$ is the standard normal density
- $\widehat{f_{X-a}}^{(j)}(t)$ is the jth derivative of the Fourier transform of the density for the random variable $X-a$
- so $i^j \, \widehat{f_{X-a}}^{(j)}(ih)$ is the jth derivative of the moment generating function of $X-a$, evaluated at the point h (what this means if the moment generating function fails to exist will be discussed in Section 6)

- we denote $j? = 0$ for odd j and $j? = (j-1)(j-3)\cdots(1)$ for even j (this is not standard notation but will prove much easier to work with than the more usual $(j-1)!! = (j-1)(j-3)\cdots(1)$ for even j with no consensus what $(j-1)!!$ means for odd j). Note that $j?$ is the j-th moment of the standard normal random variable.

- h is a solution to $i\,\widehat{f_{X-a}}^{(1)}(ih) = 0$ (killing the $j = 1$ term in the sum), in practice found by a numerical tool such as EXCEL SOLVER

- c is a solution to $\frac{i^2\,\widehat{f_{X-a}}^{(2)}(ih)}{c^2\,\widehat{f_{X-a}}(ih)} - 1 = 0$ (killing the $j = 2$ term in the sum), in practice found by a numerical tool such as EXCEL SOLVER

- if it so happens that $a = \mu_X$ (which is not the usual Esscher choice) then the solutions are $h = 0$, $c = \sigma_X$, and the expression is called the Edgeworth expansion

- $H_m(z) = \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} (-1)^k \frac{m!(2k)?}{(m-2k)!(2k)!} z^{m-2k}$ = the mth Hermite polynomial. Note that on this definition the literature is consistent only up to $(\pm 1)^m$. Note also the easy calculation $\varphi^{(m)}(z) = (-1)^m H_m(z)\,\varphi(z)$.

2.2 Esscher Expansion for a Density - The Usual Form

The literature usually expresses the Esscher expansion with the opposite order of summation: first n then j:

$$
\begin{aligned}
f_X(x) \quad &= \quad \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \\
&\left\{ 1 + \sum_{n=3}^{\infty} \frac{1}{n!} H_n\left(\frac{x-a}{c}\right) \sum_{j=3}^{n} \frac{i^{n-j} n!\,(n-j)?}{j!\,(n-j)!} \left[\frac{i^j\,\widehat{f_{X-a}}^{(j)}(ih)}{c^j\,\widehat{f_{X-a}}(ih)} - j? \right] \right\}
\end{aligned}
\tag{2}
$$

The ? notation forces $n - j$ to be even so Equation (2) is real despite the suspicious appearance of i in a coefficient. To get from Equation (2) to Equation (1) change the order of summation, change variables so $2n + j$ replaces n, and simplify. Equation (1) requires use of an explicit upper limit N on the sum over j in order to provide the correct upper limit on the sum over n corresponding to each j.

2.3 Approximation for a Probability

To find the probability that the random variable X is in the range $u \leq X \leq v$ just integrate Equation (1)

$$\int_u^v f_X(x)\,dx =$$

$$= \frac{\widehat{f_{X-a}}(ih)}{c} \left\{ \begin{array}{c} \int_u^v e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) dx + \lim_{N\to\infty} \sum_{j=3}^{N} \frac{1}{j!} \left[\frac{i^j\, \widehat{f_{X-a}}^{(j)}(ih)}{c^j\, \widehat{f_{X-a}}(ih)} - j? \right] \cdot \\ \sum_{n=0}^{\left\lfloor \frac{N-j}{2} \right\rfloor} \frac{(-1)^n (2n)?}{(2n)!} \int_u^v e^{h(x-a)} H_{2n+j}\left(\frac{x-a}{c}\right) \varphi\left(\frac{x-a}{c}\right) dx \end{array} \right\} \tag{3}$$

2.4 Evaluation of the Probability Integrals

To evaluate the integrals in Equation (3) note that $H_m(z)\,\varphi(z) = (-1)^m \varphi^{(m)}(z)$ and integrate by parts repeatedly, using $e^{h(x-a)}\varphi\left(\frac{x-a}{c}\right) = e^{\frac{(hc)^2}{2}}\varphi\left(hc - \frac{x-a}{c}\right)$ at the last step, resulting in

$$\int_u^v e^{h(x-a)} H_m\left(\frac{x-a}{c}\right) \varphi\left(\frac{x-a}{c}\right) dx =$$

$$= c \left\{ \begin{array}{c} e^{hc(u-a)} \varphi\left(\frac{u-a}{c}\right) \sum_{k=0}^{m-1} (hc)^{m-1-k} H_k\left(\frac{u-a}{c}\right) \\ -e^{hc(v-a)} \varphi\left(\frac{v-a}{c}\right) \sum_{k=0}^{m-1} (hc)^{m-1-k} H_k\left(\frac{v-a}{c}\right) \\ +(hc)^m e^{\frac{(hc)^2}{2}} \left(\Phi\left(\frac{v-a}{c} - hc\right) - \Phi\left(\frac{u-a}{c} - hc\right)\right) \end{array} \right\} \tag{4}$$

where Φ is the standard normal cdf.

Equation (4) simplifies considerably when $v = \infty$ and with the choice $a = u$. This is easy to arrange because a can be chosen arbitrarily. With h and c chosen according to the choice $a = u$:

$$\int_a^\infty e^{h(x-a)} H_m\left(\frac{x-a}{c}\right) \varphi\left(\frac{x-a}{c}\right) dx =$$

$$= c \left[\frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\left\lfloor \frac{m-1}{2} \right\rfloor} (hc)^{m-1-2k} (-1)^k (2k)? + (hc)^m e^{\frac{(hc)^2}{2}} \Phi(hc) \right] \tag{5}$$

2.5 Approximate Tail Probabilities

Esscher (1932) expressed the right hand side of Equation (5) recursively in m and later authors (e.g. Woody (1973), Jens (1995)) called the resulting expressions "Esscher functions." In addition to the simplification, it turns out that Equation (3) converges much more rapidly when $v = \infty$ and $a = u$ (see Section 3.) With these choices, substituting Equation (5) into Equation (3) reduces it to:

$$\int_a^{\infty} f_X(x)\,dx =$$

$$= \widehat{f_{X-a}}(ih) \left\{ \begin{array}{c} e^{\frac{(hc)^2}{2}} \Phi(hc) + \lim_{N\to\infty} \sum_{j=3}^{N} \frac{1}{j!} \left[\frac{i^j\, \widehat{f_{X-a}}^{(j)}(ih)}{c^j\, \widehat{f_{X-a}}(ih)} - j? \right] \cdot \\ \sum_{n=0}^{\lfloor \frac{N-j}{2} \rfloor} \frac{(-1)^n (2n)?}{(2n)!} \left[\begin{array}{c} \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\lfloor \frac{2n+j-1}{2} \rfloor} (hc)^{2n+j-1-2k} (-1)^k (2k)? \\ + (hc)^{2n+j} \; e^{\frac{(hc)^2}{2}} \Phi(hc) \end{array} \right] \end{array} \right\} \tag{6}$$

2.6 Approximation for Moments

In similar fashion we can use the Esscher expansion Equation (1) to approximate moments of the random variable X conditional upon limits $u \leq X \leq v$. It's most convenient to look at moments about the location parameter a (from these some algebra will give moments about any other point of interest.)

$$\int_u^v (x-a)^w f_X(x)\,dx =$$

$$= \frac{\widehat{f_{X-a}}(ih)}{c} \left\{ \begin{array}{c} \int_u^v (x-a)^w e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) dx + \lim_{N\to\infty} \sum_{j=3}^{N} \frac{1}{j!} \left[\frac{i^j\, \widehat{f_{X-a}}^{(j)}(ih)}{c^j\, \widehat{f_{X-a}}(ih)} - j? \right] \cdot \\ \sum_{n=0}^{\lfloor \frac{N-j}{2} \rfloor} \frac{(-1)^n (2n)?}{(2n)!} \int_u^v (x-a)^w e^{h(x-a)} H_{2n+j}\left(\frac{x-a}{c}\right) \varphi\left(\frac{x-a}{c}\right) dx \end{array} \right\} \tag{7}$$

2.7 Evaluation of the Moment Integrals

To evaluate the integrals in Equation (7) again use $H_m(z)\varphi(z) = (-1)^m \varphi^{(m)}(z)$ and integrate by parts repeatedly, using $e^{h(x-a)}\varphi\left(\frac{x-a}{c}\right) = e^{\frac{(hc)^2}{2}}\varphi\left(hc - \frac{x-a}{c}\right)$ at the last step, resulting in

$$\int_u^v (x-a)^w e^{h(x-a)} H_m\left(\frac{x-a}{c}\right)\varphi\left(\frac{x-a}{c}\right)dx =$$

$$= c\left\{\begin{array}{c} (u-a)^w e^{hc(u-a)}\varphi\left(\frac{u-a}{c}\right)\sum_{k=0}^{m-1}(hc)^{m-1-k} H_k\left(\frac{u-a}{c}\right) \\ -(v-a)^w e^{hc(v-a)}\varphi\left(\frac{v-a}{c}\right)\sum_{k=0}^{m-1}(hc)^{m-1-k} H_k\left(\frac{v-a}{c}\right) \\ +\frac{1}{\sqrt{2\pi}}(hc)^m e^{\frac{(hc)^2}{2}} c^w \sum_{k=0}^{w}\frac{w!}{k!(w-k)!}(hc)^{w-k} 2^{\frac{k-1}{2}}\cdot \\ \Gamma\left(\frac{k+1}{2}\right)\left[\Gamma\left(\frac{k+1}{2};\frac{\left(\frac{v-a}{c}-hc\right)^2}{2}\right) - \Gamma\left(\frac{k+1}{2};\frac{\left(\frac{u-a}{c}-hc\right)^2}{2}\right)\right] \end{array}\right\}$$
$$+cw\sum_{k=0}^{m-1}(hc)^{m-1-k}\int_u^v (x-a)^{w-1} e^{h(x-a)} H_k\left(\frac{x-a}{c}\right)\varphi\left(\frac{x-a}{c}\right)dx \quad (8)$$

where $\Gamma(\cdot)$ and $\Gamma(\cdot;\cdot)$ are the complete and incomplete gamma functions, respectively. Evaluate the remaining integrals recursively in the same fashion. When $w = 0$ this agrees with Equation (4) by changing variables in the integral defining $\Gamma\left(\frac{1}{2};\cdot\right)$ to get an integral defining Φ.

With some painstaking calculation Equation (8) simplifies when $v = \infty$ and with the choice $a = u$. This is easy to arrange because a can be chosen arbitrarily. With h and c chosen according to the choice $a = u$:

$$\int_a^\infty (x-a)^w e^{h(x-a)} H_m\left(\frac{x-a}{c}\right)\varphi\left(\frac{x-a}{c}\right)dx = \frac{c^{w+1}}{\sqrt{2\pi}}\cdot$$

$$\sum_{t=0}^{w\wedge m}\left\{\begin{array}{c} \left[(hc)^{m-t} + \frac{\mathbf{1}_{(t>0)}}{(t-1)!}\sum_{k=0}^{m-t-1}\frac{(m-1-k)!}{(m-t-k)!}(hc)^k\right]\cdot \\ \left[\begin{array}{l} \mathbf{1}_{(t=w)}\cdot\sum_{k=0}^{\left\lfloor\frac{m-t-1}{2}\right\rfloor}(hc)^{m-t-1-2k}(-1)^k(2k)? + e^{\frac{(hc)^2}{2}}\cdot \\ \sum_{k=0}^{w-t}\frac{w!}{k!(w-t-k)!}(hc)^{w-t-k}2^{\frac{k-1}{2}}\Gamma\left(\frac{k+1}{2}\right)\left[1-\Gamma\left(\frac{k+1}{2};\frac{(hc)^2}{2}\right)\right] \end{array}\right] \end{array}\right\} \quad (9)$$

where $w \wedge m = \min(w, m)$ and the indicator functons $\mathbf{1}_{()} = 1$ when () is true and $= 0$ otherwise.

2.8 Approximate Tail Moments

Esscher (1932) also expressed the right hand side of Equation (9) recursively in w and m and later authors (e.g. Woody (1973)) recognized the resulting expressions as additional "Esscher functions." In addition to the simplification, it turns out that Equation (7) converges much more rapidly when $v = \infty$ and $a = u$ (see Section 3.) With these choices, substituting Equation (9) into Equation (7) gives:

$$\int_a^\infty (x-a)^w f_X(x)\,dx = \widehat{f_{X-a}}(ih)\frac{c^w}{\sqrt{2\pi}}\cdot$$

$$\left\{\begin{array}{c} e^{\frac{(hc)^2}{2}}\sum_{k=0}^{w}\frac{w!}{k!(w-k)!}(hc)^{w-k}2^{\frac{k-1}{2}}\Gamma\left(\frac{k+1}{2}\right)\left[1-\Gamma\left(\frac{k+1}{2};\frac{(hc)^2}{2}\right)\right] \\ +\lim_{N\to\infty}\sum_{j=3}^{N}\frac{1}{j!}\left[\frac{i^j\,\widehat{f_{X-a}}^{(j)}(ih)}{c^j\,\widehat{f_{X-a}}(ih)}-j?\right]\sum_{n=0}^{\left\lfloor\frac{N-j}{2}\right\rfloor}\frac{(-1)^n(2n)?}{(2n)!}\cdot \\ \sum_{t=0}^{w\wedge(2n+j)}\left\{\left[(hc)^{2n+j-t}+\frac{\mathbf{1}_{(t>0)}}{(t-1)!}\sum_{k=0}^{2n+j-t-1}\frac{(2n+j-1-k)!}{(2n+j-t-k)!}(hc)^k\right]\cdot\right. \\ \left.\left[\begin{array}{c}\mathbf{1}_{(t=w)}\cdot\sum_{k=0}^{\left\lfloor\frac{2n+j-t-1}{2}\right\rfloor}(hc)^{2n+j-t-1-2k}(-1)^k(2k)?+e^{\frac{(hc)^2}{2}}\cdot \\ \sum_{k=0}^{w-t}\frac{w!}{k!(w-t-k)!}(hc)^{w-t-k}2^{\frac{k-1}{2}}\Gamma\left(\frac{k+1}{2}\right)\left[1-\Gamma\left(\frac{k+1}{2};\frac{(hc)^2}{2}\right)\right]\end{array}\right]\right\}\end{array}\right\} \tag{10}$$

It's worth noting that the incomplete gamma functions $\Gamma\left(\frac{k+1}{2}\right)\left[1-\Gamma\left(\frac{k+1}{2};\frac{(hc)^2}{2}\right)\right]$ in the formulae of this section all can be calculated using just exponentials, polynomials and the standard normal cdf Φ. See, for example, appendix A of Klugman-Panjer-Willmot (2008) to reduce $\Gamma\left(\frac{k+1}{2}\right)\left[1-\Gamma\left(\frac{k+1}{2};\frac{(hc)^2}{2}\right)\right]$ to an exponential and polynomial when $k+1$ is even and to expressions involving $\Gamma\left(\frac{1}{2}\right)$ and $\Gamma\left(\frac{1}{2};\frac{(hc)^2}{2}\right)$ when $k+1$ is odd. By a change of variable in the defining integrals, these latter reduce to expressions in Φ.

3 Heuristic View of the Esscher's Accuracy

The surprising accuracy of the Esscher approximation appears to stem from the confluence of three convergence-enhancing influences operating simultaneously.

3.1 Sparse Error Terms

To begin with, looking at the usual expression of the Esscher approximation, Equation (2),

$$\begin{aligned} f_X(x) &= \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \\ &\left\{ 1 + \sum_{n=3}^{\infty} \frac{1}{n!} H_n\left(\frac{x-a}{c}\right) \sum_{j=3}^{n} \frac{i^{n-j} n! (n-j)?}{j! (n-j)!} \left[\frac{i^j \widehat{f_{X-a}}^{(j)}(ih)}{c^j \widehat{f_{X-a}}(ih)} - j? \right] \right\} \end{aligned}$$

we can see two convergence enhancing influences at work.

First, a series expansion of some kind in n has the $n = 1$ and 2 terms supressed, meaning errors only begin to appear at the third order term.

Second, the coefficient of each remaining term involves a summation (in j) that has zeroes for many of its possible summands: the $j = 0$, 1, 2, and all subsequent summands of parity opposite to n are zero (remember that $(n-j)? = 0$ when $n - j$ is odd.) The relative number of zero summands weighs most significantly for smaller values of n, the errors terms of lower order. Thus, in a partial sum to any order $n = N$, the remaining error terms of order $n > N$ will be smaller owing to the zero summands in their coefficients having greatest relative weight in the lower order error terms.

3.2 Destructive Interference Among Error Terms

Finally, the third covergence-enhancing influence appears when looking at the order of summation we started with, j first then n, Equation (1), but making the substitution $H_m(z)\varphi(z) = (-1)^m \varphi^{(m)}(z)$

$$\begin{aligned} f_X(x) &= \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \cdot \\ &\left\{ \begin{array}{c} \varphi\left(\frac{x-a}{c}\right) + \lim_{N\to\infty} \sum_{j=3}^{N} \frac{(-1)^j}{j!} \left[\frac{i^j \widehat{f_{X-a}}^{(j)}(ih)}{c^j \widehat{f_{X-a}}(ih)} - j? \right] \cdot \\ \sum_{n=0}^{\left\lfloor \frac{N-j}{2} \right\rfloor} \frac{(-1)^n (2n)?}{(2n)!} \varphi^{(2n+j)}\left(\frac{x-a}{c}\right) \end{array} \right\} \end{aligned} \quad (11)$$

The terms $\varphi^{(2n+j)}\left(\frac{x-a}{c}\right)$ are dampened oscillations that interfere materially to lower (on average) the absolute value of the sum $\sum_{n=0}^{\left\lfloor \frac{N-j}{2} \right\rfloor} \frac{(-1)^n (2n)?}{(2n)!} \varphi^{(2n+j)}\left(\frac{x-a}{c}\right)$, thus reducing the error terms for $j > N$ that remain in Equation (11) after a partial sum to a given value $j = N$.

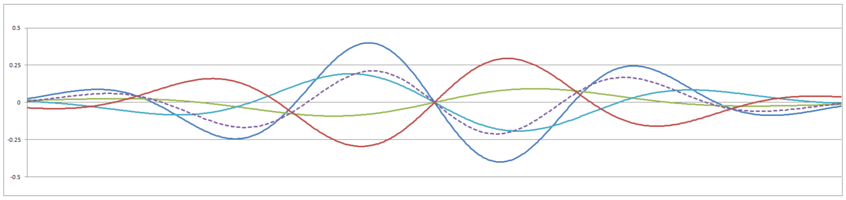

Figure 1: $\frac{(-1)^3}{3!}\frac{(-1)^n(2n)?}{(2n)!}\varphi^{(2n+3)}\left(\frac{x-a}{c}\right)$ for $n = 0, 1, 2, 3$ for $-3 \le \frac{x-a}{c} \le 3$, dotted line = their sum

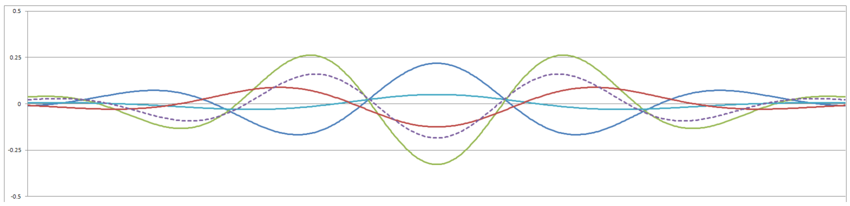

Figure 2: $\frac{(-1)^4}{4!}\frac{(-1)^n(2n)?}{(2n)!}\varphi^{(2n+4)}\left(\frac{x-a}{c}\right)$ for $n = 0, 1, 2, 3$ for $-3 \le \frac{x-a}{c} \le 3$, dotted line = their sum

For example, Figure 1 illustrates the case $N = 10$, $j = 3$, plotting the values of $\frac{(-1)^3}{3!}\frac{(-1)^n(2n)?}{(2n)!}\varphi^{(2n+3)}\left(\frac{x-a}{c}\right)$, and of their sum (the dotted line) for $n = 0$, 1, 2, 3 for $-3 \le \frac{x-a}{c} \le 3$.

Figure 2 illustrates the opposite parity $N = 10$, $j = 4$, plotting the values of $\frac{(-1)^4}{4!}\frac{(-1)^n(2n)?}{(2n)!}\varphi^{(2n+4)}\left(\frac{x-a}{c}\right)$, and of their sum (the dotted line) for $n = 0$, 1, 2, 3 for $-3 \le \frac{x-a}{c} \le 3$.

Integrating over an entire cycle or multiple entire cycles of each sum (the dotted lines) clearly will increase vastly the dampening effect, thus reducing significantly the the remaining error terms in an integration of Equation (11). Such an integration corresponds to choices of the limits of integration u and v in Equation (3) so that they encompass as much as possible a whole number of cycles in the sums (the dotted lines) that appear in the coefficients of the terms for each j.

The choice $a = u$ and $v = \infty$ in Equation (6) encompasses all cycles to the right of $\frac{x-a}{c} = 0$ in all of the sums that appear in the coefficients for all j, so Equation (6) represents the most dramatic rate of convergence.

Similarly dramatic destructive interference among error terms will occur in Equation (10) for moments in the tail of the distribution, again selecting $a = u$ and $v = \infty$ to bring in all cycles to the right of $\frac{x-a}{c} = 0$.

3.3 Combination of Convergence Enhancing Influences

											X
										X	*0*
									X	**0**	*X*
↑								**X**	**0**	**X**	*0*
j							**X**	**0**	**X**	**0**	*X*
						X	**0**	**X**	**0**	**X**	*0*
					X	**0**	**X**	**0**	**X**	**0**	*X*
				X	**0**	**X**	**0**	**X**	**0**	**X**	*0*
			0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	*0*
		0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	*0*
	X	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	*0*
			n	**or**	**2n**	+	**j**	→			

Figure 3: Coefficients of Error Terms

All three of the convergence-enhancing influences are illustrated schematically in Figure 3. The horizontal axis represents n in the usual form of the Esscher, Equation (2), or, equivalently, $2n + j$ in our form, Equation (1). In either form, each new main term brings in an entire column in Figure 3.

Error terms are represented by all columns (to infinity) to the right of the last column used. For all error terms we see that the first three coefficients are 0, half of the remaining coeffients are 0, and the non-zero error terms (the Xs), we can recall, are subject to significant destructive interference (running horizontally across the display) for all integrated versions of the Esscher (tail probabilities and tail moments, Equation (6) and Equation (10).)

Note particularly the preponderance of 0s in the lower order terms to the left in Figure 3. By the time the non-zero Xs come to contribute much, those remaining errors are in much higher order terms.

While not based on a comprehensive literature review, it appears that most theoretical analyses of the rate of convergence of the Esscher expansion take one or two but not all three of these convergence enhancing influences into account. If true, this would explain the observation by many analysts that the Esscher expansion converges in practice much more rapidly than existing theory suggests.

4 Derivation of the Esscher Expansion

Working entirely in Fourier transorm space provides the most general setting to derive the Esscher expansion and also provides a convenient base to extend its application into maximum likelihood approximations. Any typical reference on the Fourier transform supports the properties of the Fourier transform used in the following to derive the Esscher expansion. See, for example, Howell (2001), Meikle (2004), Rudin (1966), Strichartz (2003), or any of a large number of alternatives. Inessential differences in the choices for certain signs and constants can occur among different sources. We specify our choices below.

For any real (or complex) function $f(x)$ on the reals $\mathbb{R}$ we let $\widehat{f}(t)$ denote the Fourier transform of f evaluated at the value t.

$$\widehat{f}(t) = \int_{-\infty}^{\infty} e^{-itx} f(x) dx$$

If f happens to be the probability density function f_X for a random variable X then clearly

$$\widehat{f_X}(t) = \mathbb{E}\left[e^{-itX}\right]$$

and the probabilistic assertions in Section 2.1 make sense.

4.1 The Expansion Itself

To derive Equation (2) for X a random variable and a any arbitrary constant begin with the following typical property of a Fourier transform:

$$\begin{aligned}
\widehat{f_X}(t) &= e^{-iat}\widehat{f_{X-a}}(t) \text{ by the "translation transforms} \\
& \text{to exponential tilting" property} \\
&= e^{-iat}\frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih)}\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih) \text{ algebraically} \\
& \text{for any arbitrary constants } c \text{ and } h \text{ and for} \\
& \text{the standard normal pdf } \varphi
\end{aligned}$$

Now use the "scaling" property $\widehat{g\left(\frac{x}{c}\right)}(t) = c\,\widehat{g}(ct)$ to write $\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih) = c\widehat{\varphi}(c(t-ih))$ giving

$$\begin{aligned}
\widehat{f_X}(t) &= e^{-iat}\frac{\widehat{f_{X-a}}(t)}{c\widehat{\varphi}(c(t-ih))}\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih) \\
&= \frac{1}{c}e^{-iat}\left\{\frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))}\right\}\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih)
\end{aligned}$$

where the $\{\}$ is just to emphasize the next step. Make a Taylor's series expansion of the function (of t) inside the $\{\}$, expanding about the point $t = ih$.

$$\begin{aligned}
\widehat{f_X}(t) &= \frac{1}{c}e^{-iat}\left\{\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))}\right]_{t=ih}^{(n)}(t-ih)^n\right\}\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih) \\
&= \frac{1}{c}e^{-iat}\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]_{t=0}^{(n)}(t-ih)^n\,\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih)
\end{aligned}$$

after a change of variable in the derivatives.

(If the Taylor's series expansion does not exist it means that $\widehat{f_X}$ is not an analytic function on the imaginary axis or, equivalently, that f_X has no moment-generating function. See Section 6 for suggestions about how to proceed in this case provided at least that f_X has finite moments.)

At this point use the property that "derivatives transform into multiplication by a power" $\widehat{g^{(n)}}(t) = i^n t^n \widehat{g}(t)$ and the chain rule for derivatives to write $(t-ih)^n \widehat{\varphi\left(\frac{x}{c}\right)}(t-ih) = \frac{i^{-n}}{c^n} \overbrace{\widehat{\varphi^{(n)}\left(\frac{x}{c}\right)}}(t-ih)$ so

$$\widehat{f_X}(t) = \frac{1}{c}e^{-iat}\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]_{t=0}^{(n)}\frac{i^{-n}}{c^n}\overbrace{\widehat{\varphi^{(n)}\left(\frac{x}{c}\right)}}(t-ih)$$

The property "exponential tilting transforms to translation" gives $\overbrace{\widehat{\varphi^{(n)}\left(\frac{x}{c}\right)}}(t-ih) = \overbrace{\widehat{e^{hx}\varphi^{(n)}\left(\frac{x}{c}\right)}}(t)$ which means

$$\widehat{f_X}(t) = \frac{1}{c}e^{-iat}\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]_{t=0}^{(n)}\frac{i^{-n}}{c^n}\overbrace{\widehat{e^{hx}\varphi^{(n)}\left(\frac{x}{c}\right)}}(t)$$

Finally, the "translation transforms to exponential tilting" property gives $e^{-iat}\overbrace{\widehat{e^{hx}\varphi^{(n)}\left(\frac{x}{c}\right)}}(t) = \overbrace{\widehat{e^{h(x-a)}\varphi^{(n)}\left(\frac{x-a}{c}\right)}}(t)$ so

$$\widehat{f_X}(t) = \frac{1}{c}\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]_{t=0}^{(n)}\frac{i^{-n}}{c^n}\overbrace{\widehat{e^{h(x-a)}\varphi^{(n)}\left(\frac{x-a}{c}\right)}}(t)$$

Now the "linearity" and "inversion" properties mean that coming back from Fourier transform space into the original space of functions

$$\begin{aligned} f_X(x) &= \frac{1}{c}\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]_{t=0}^{(n)}\frac{i^{-n}}{c^n}e^{h(x-a)}\varphi^{(n)}\left(\frac{x-a}{c}\right) \\ &= \frac{1}{c}e^{h(x-a)}\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]_{t=0}^{(n)}\frac{i^{-n}}{c^n}\varphi^{(n)}\left(\frac{x-a}{c}\right) \end{aligned}$$

and the fact that $\varphi^{(n)}(z) = (-1)^n H_n(z)\varphi(z)$ for the Hermite polynomial $H_n(z)$ gives

$$f_X(x) = \frac{1}{c}e^{h(x-a)}\varphi\left(\frac{x-a}{c}\right)\sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]_{t=0}^{(n)}\frac{i^n}{c^n}H_n\left(\frac{x-a}{c}\right) \tag{12}$$

4.2 The Coefficients in the Expansion

If we can unravel the derivatives in the coefficients of Equation (12) we will establish Equation (2). To do so use Leibniz's product rule and a trick.

For $n > 0$ Leibniz's product rule gives

$$\left[\frac{\widehat{f_{X-a}}\,(t+ih)}{\widehat{\varphi}\,(ct)}\right]^{(n)}_{t=0} = \widehat{f_{X-a}}\,(ih)\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n)}_{t=0} + \sum_{j=1}^{n}\frac{n!}{j!\,(n-j)!}\widehat{f_{X-a}}^{\,(j)}\,(ih)\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n-j)}_{t=0} \tag{13}$$

The trick is to notice that $\left[\frac{\widehat{\varphi}(ct)}{\widehat{\varphi}(ct)}\right]^{(n)}_{t=0} = 0$ for $n > 0$ so Leibniz's product rule gives

$$\begin{aligned}0 &= \left[\frac{\widehat{\varphi}\,(ct)}{\widehat{\varphi}\,(ct)}\right]^{(n)}_{t=0} = \widehat{\varphi}\,(0)\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n)}_{t=0} + \sum_{j=1}^{n}\frac{n!}{j!\,(n-j)!}c^j\widehat{\varphi}^{(j)}\,(ct)\,|_{t=0}\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n-j)}_{t=0} \\ &= \widehat{f_{X-a}}\,(ih)\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n)}_{t=0} + \sum_{j=1}^{n}\frac{n!}{j!\,(n-j)!}c^j\widehat{\varphi}^{(j)}\,(0)\,\widehat{f_{X-a}}\,(ih)\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n-j)}_{t=0}\end{aligned} \tag{14}$$

after multiplying by $\widehat{f_{X-a}}\,(ih)$ and noting that $\widehat{\varphi}\,(0) = 1$ (an easy calculation from the definition of Fourier transform).

Now subtract Equation (14) from Equation (13) to get

$$\begin{aligned}\left[\frac{\widehat{f_{X-a}}\,(t+ih)}{\widehat{\varphi}\,(ct)}\right]^{(n)}_{t=0} &= \sum_{j=1}^{n}\frac{n!}{j!\,(n-j)!}\left[\widehat{f_{X-a}}^{\,(j)}\,(ih) - \widehat{f_{X-a}}\,(ih)\,c^j\widehat{\varphi}^{(j)}\,(0)\right]\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n-j)}_{t=0} \\ &= c^j\,\widehat{f_{X-a}}\,(ih)\sum_{j=1}^{n}\frac{n!}{j!\,(n-j)!}\left[\frac{\widehat{f_{X-a}}^{\,(j)}\,(ih)}{c^j\,\widehat{f_{X-a}}\,(ih)} - \widehat{\varphi}^{(j)}\,(0)\right]\left[\frac{1}{\widehat{\varphi}\,(ct)}\right]^{(n-j)}_{t=0}\end{aligned}$$

Finally a direct calculation gives $\widehat{\varphi}^{(j)}\,(0) = i^{-j}j?$ and $\left[\frac{1}{\widehat{\varphi}(ct)}\right]^{(n-j)}_{t=0} = c^{n-j}\,(n-j)?$ so the coefficients in Equation (12) are

$$\left[\frac{\widehat{f_{X-a}}\,(t+ih)}{\widehat{\varphi}\,(ct)}\right]^{(n)}_{t=0} = c^n\,\widehat{f_{X-a}}\,(ih)\sum_{j=1}^{n}\frac{i^{-j}n!\,(n-j)?}{j!\,(n-j)!}\left[\frac{i^j\,\widehat{f_{X-a}}^{\,(j)}\,(ih)}{c^j\,\widehat{f_{X-a}}\,(ih)} - j?\right] \tag{15}$$

4.3 The Completed Expansion

At last, substituting the coefficients from Equation (15) into Equation (12) gives

$$f_X(x) = \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \left\{1 + \sum_{n=1}^{\infty} \frac{1}{n!} H_n\left(\frac{x-a}{c}\right) \sum_{j=1}^{n} \frac{i^{n-j} n! (n-j)?}{j! (n-j)!} \left[\frac{i^j \widehat{f_{X-a}}^{(j)}(ih)}{c^j \widehat{f_{X-a}}(ih)} - j?\right]\right\}$$

Given the arbitrary constant a, solve $\widehat{f_{X-a}}^{(1)}(ih) = 0$ for a value h. With that choice for h, the $j = 1$ terms in the expansion all vanish (remember, $1? = 0$).

With those values for a and h, solve $\frac{i^2 \widehat{f_{X-a}}^{(2)}(ih)}{c^2 \widehat{f_{X-a}}(ih)} - 1 = 0$ for a value c. With that choice for c, the $j = 2$ terms in the expansion all vanish (remember, $2? = 1$).

Given those choices for h and c, with all the $j = 1$ and $j = 2$ terms in the expansion vanishing, the $n = 1$ and $n = 2$ terms vanish completely. The result is Equation (2)

$$f_X(x) = \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \left\{1 + \sum_{n=3}^{\infty} \frac{1}{n!} H_n\left(\frac{x-a}{c}\right) \sum_{j=3}^{n} \frac{i^{n-j} n! (n-j)?}{j! (n-j)!} \left[\frac{i^j \widehat{f_{X-a}}^{(j)}(ih)}{c^j \widehat{f_{X-a}}(ih)} - j?\right]\right\} \tag{2}$$

which, given our notation choices, corresponds to the usual presentation of the Esscher expansion in the literature.

To get Equation (1) change the order of summation in Equation (2) and replace the summation variable n by $2n + j$

$$f_X(x) = \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \left\{1 + \lim_{N\to\infty} \sum_{j=3}^{N} \frac{1}{j!} \left[\frac{i^j \widehat{f_{X-a}}^{(j)}(ih)}{c^j \widehat{f_{X-a}}(ih)} - j?\right] \sum_{n=0}^{\lfloor \frac{N-j}{2} \rfloor} \frac{(-1)^n (2n)?}{(2n)!} H_{2n+j}\left(\frac{x-a}{c}\right)\right\} \tag{1}$$

This is the presentation of the Esscher expansion that we began the paper with.

4.4 Characterization of the Expansion

Equation (1) displays the density f_X as an exponentially tilted, recentered and rescaled standard normal density, corrected by a series of factors $\frac{1}{j!}\left[\frac{i^j\,\widehat{f_{X-a}}^{(j)}(ih)}{c^j\,\widehat{f_{X-a}}(ih)} - j?\right]$ measuring the departure of the higher moments (≥ 3) of an (oppositely) exponentially tilted, recentered and rescaled version of f_X from the moments of the standard normal. Given a as the recentering point from f_X to f_{X-a}, our choice of h gives the tilting e^{-hx} of f_{X-a} required to zero the mean and our choice of c gives the rescaling from $e^{-hx}f_{X-a}(x)$ to $e^{-hcx}f_{\frac{X-a}{c}}(x)$ required to normalize the variance.

To see this requires the "exponential tilting transforms to translation" property followed by a change of variables and then by the "scaling" property: $\widehat{e^{-hcx}f_{\frac{X-a}{c}}(x)}(t) = \widehat{f_{\frac{X-a}{c}}(x)}(t+ihc) = \widehat{cf_{X-a}(cx)}(t+ihc) = \widehat{f_{X-a}(x)}\left(\frac{t}{c}+ih\right)$.

As the order of approximation (N) increases, the correction factors themselves (one for each $j \leq N$) take on increasingly fine-tuned oscillating modulation factors $\sum_{n=0}^{\left\lfloor \frac{N-j}{2} \right\rfloor} \frac{(-1)^n(2n)?}{(2n)!} H_{2n+j}\left(\frac{x-a}{c}\right)$. These represent a normalized prediction of the effects of unknown higher moments, given only the values for the lower moments. (This thought sets the foundation for some of the reasoning in Section 6 concerning what to do in the situation where f_X has moments, but no moment-generating function, so no Taylor's series by which to derive an Esscher expansion.)

4.5 Summary of the Derivation

To summarize the derivation,

- The Esscher expansion comes from a Taylor's series expansion in the space of Fourier transforms.

- Specifically, it comes from the Taylor's series expansion of a function $\frac{\widehat{f_{X-a}(t)}}{\widehat{\varphi}(c(t-ih))}$ that links f_X algebraically to the standard normal φ over in the space of Fourier transforms, after some translations and scaling.

- The powers in the expansion get multiplied in Fourier space by a version of $\widehat{\varphi}$, the transform of the normal density, modified by some scaling and translation

- That corresponds back in the original space, the densities, to taking derivatives of the normal φ

- That in turn corresponds to multiplying φ by Hermite polynomials

- The coefficients of the Taylor's expansion can be seen to compare moments of an exponentially tilted, recentered, and rescaled version of f_X with moments of φ
- Everything else just keeps meticulous track of the constants involved in the various scalings, translations and corresponding exponential tiltings
- Given a the other constants h and c are chosen to eliminate lower order error terms while minimizing higher order ones

It's worth noting that one could undertake the same program with some other density replacing the normal φ, making the moment structure of that other density the basis for the approximation.

5 Approximation for Maximum Likelihood Value

Instead of integrating the Esscher expansion (Equation (1)) of the density f_X to approximate tail probabilities (Equations (3) and (6)) or conditional tail moments (Equations (7) and (10)) what about using the Esscher expansion together with a derivative to approximate maximum likelihood values on f_X?

Three possible ways to do this suggest themselves:

1. Take the derivative of Equation (1) and set it equal to zero. There is no guarantee that the resulting expansion will converge to $f_X^{(1)}$, the derivative of f_X, especially since the Esscher expansion itself (Equation (1)) is oscillatory in nature (section 3).
2. Apply the Esscher expansion logic (Section 4) to $f_X^{(1)}$, the derivative of f_X, rather than to f_X, and set the resulting expansion equal to zero. We will see that the resulting expansion converges to $f_X^{(1)}$ when f_X has finite moments (see section 6.).
3. Take a weighted average of the expansions suggested in 1. and 2. and set it equal to zero. If f_X has finite moments and 1. converges then so will such a weighted average and perhaps the rate of convergence can be optimized by the right choice of weights in the average.

5.1 Derivative of the Esscher Expansion

Since $\varphi\left(\frac{x-a}{c}\right) H_{2n+j}\left(\frac{x-a}{c}\right) = (-1)^j \varphi^{(2n+j)}\left(\frac{x-a}{c}\right)$ a simple product rule calculation gives for the derivative of Equation (1):

$$
\begin{aligned}
f_X^{(1)}(x) \quad \sim \quad & \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \\
& \{\ h\ [\ 1 + \lim_{N\to\infty} \sum_{j=1}^{N} \frac{1}{j!} [\ \frac{i^j \widehat{f_{X-a}}^{(j)}(ih)}{c^j\ \widehat{f_{X-a}}(ih)} - j?\ ] \cdot \\
& \sum_{n=0}^{\lfloor \frac{N-j}{2} \rfloor} \frac{(-1)^n (2n)?}{(2n)!} H_{2n+j}\left(\frac{x-a}{c}\right)\ ] \\
& -\frac{1}{c} [\ H_1\left(\frac{x-a}{c}\right) + \lim_{N\to\infty} \sum_{j=1}^{N} \frac{1}{j!} [\ \frac{i^j \widehat{f_{X-a}}^{(j)}(ih)}{c^j\ \widehat{f_{X-a}}(ih)} - j?\ ] \cdot \\
& \sum_{n=0}^{\lfloor \frac{N-j}{2} \rfloor} \frac{(-1)^n (2n)?}{(2n)!} H_{2n+j+1}\left(\frac{x-a}{c}\right)\ ]\ \} \qquad \text{where}
\end{aligned} \tag{16}
$$

- We do not claim equality at all (or any) points; the derivative of an oscillatory approximation (see Section 3) may well not converge to the derivative of the original function at all points.
- For the moment we assume that h and c can take any values, so the $j = 1$ and $j = 2$ terms still remain.
- The h term comes from the derivative of $e^{h(x-a)}$.
- The $-\frac{1}{c}$ term comes from the derivative of $\varphi\left(\frac{x-a}{c}\right) H_{2n+j}\left(\frac{x-a}{c}\right) = (-1)^j \varphi^{(2n+j)}\left(\frac{x-a}{c}\right)$.

5.2 Esscher Expansion of the Derivative

Doing for $f_X^{(1)}$ exactly what we did in Section 4 for f_X it's easy to get to the analogue of Equation (1)

$$
\begin{aligned}
f_X^{(1)}(x) \quad = \quad & \frac{\widehat{f_{X-a}^{(1)}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot \\
& \left\{ 1 + \lim_{N\to\infty} \sum_{j=1}^{N} \frac{1}{j!} \left[\frac{i^j\ \widehat{f_{X-a}^{(1)}}^{(j)}(ih)}{c^j\ \widehat{f_{X-a}^{(1)}}(ih)} - j? \right] \sum_{n=0}^{\lfloor \frac{N-j}{2} \rfloor} \frac{(-1)^n (2n)?}{(2n)!} H_{2n+j}\left(\frac{x-a}{c}\right) \right\}
\end{aligned} \tag{17}
$$

where for the moment, again, we assume that h and c can take any values, so the $j=1$ and $j=2$ terms still remain. Equation (17) does converge when f_X has finite moments (see Section 6.) But how to deal with $\widehat{f_{X-a}^{(1)}}^{(j)}(ih)$?

The Fourier transform property "derivatives transform to multiplication by powers" produces $\widehat{f_{X-a}^{(1)}}(t+ih) = i(t+ih)\widehat{f_{X-a}}(t+ih)$. Now Leibniz's rule gives

$$\begin{aligned}\widehat{f_{X-a}^{(1)}}^{(j)}(t+ih)\,|_{t=0} &= -h\widehat{f_{X-a}}^{(j)}(ih) + ji\widehat{f_{X-a}}^{(j-1)}(ih) + \text{zeroes} \\ &= -h\widehat{f_{X-a}}^{(j)}(ih) - \frac{1}{c}jci^{-1}\widehat{f_{X-a}}^{(j-1)}(ih)\end{aligned} \tag{18}$$

where the last step is a preparation for separating into terms in h and in $\frac{1}{c}$ after substituting Equation (18) into Equation (17). Note that Equation (18) holds also for $j=0$.

Performing the substitution of Equation (18) into Equation (17)

$$\begin{aligned}f_X^{(1)}(x) &= \frac{\widehat{f_{X-a}}(ih)}{c}e^{h(x-a)}\varphi\left(\frac{x-a}{c}\right)\cdot \\ &\quad \{-h\,[\,1+\lim_{N\to\infty}\sum_{j=1}^{N}\frac{1}{j!}\,[\,\frac{i^j\widehat{f_{X-a}}^{(j)}(ih)}{c^j\,\widehat{f_{X-a}}(ih)} - j?\,]\cdot \\ &\quad \sum_{n=0}^{\lfloor\frac{N-j}{2}\rfloor}\frac{(-1)^n(2n)?}{(2n)!}H_{2n+j}\left(\frac{x-a}{c}\right)\,] \\ &\quad -\frac{1}{c}[\,\lim_{N\to\infty}\sum_{j=0}^{N-1}\frac{1}{j!}\frac{i^j\widehat{f_{X-a}}^{(j)}(ih)}{c^j\,\widehat{f_{X-a}}(ih)}\cdot \\ &\quad \sum_{n=0}^{\lfloor\frac{N-1-j}{2}\rfloor}\frac{(-1)^n(2n)?}{(2n)!}H_{2n+j+1}\left(\frac{x-a}{c}\right)\,]\,\}\qquad \text{where}\end{aligned} \tag{19}$$

- The $-h$ term is $-h$ times our original Esscher expansion Equation (1), with the $j=1$ and $j=2$ terms restored, and it matches the h term in the derivative of the Esscher expansion Equation (16).
- The $-\frac{1}{c}$ term is like $-\frac{1}{c}$ times our original Esscher expansion Equation (1), with the $j=1$ and $j=2$ terms restored, except that each j is lowered by 1 and the $j?$ terms have disappeared (they went with the $-h$.)

5.3 The Weighted Average of Equations (16) and (19)

Combining Equation (16) with Equation (19) using any weighting ϑ and $(1-\vartheta)$ will yield an expression that converges when Equation (16) converges and f_X

has finite moments. There is no need to write it out at this point since it turns out not to be needed at this level of detail in subsequent developments.

5.4 Possible Approaches to Maximum Likelihood

Points of maximum likelihood x_m occur when $f_X^{(1)}(x_m) = 0$. To finite order N in Equations (16) or (19), or a weighted average of the two, an approximate value for x_m will be determined by finding a value for x_m that minimizes the absolute value of the partial sum to N in Equation (16) or (19), or in a weighted average of the two. (Minimizing allows for the possibility that the partial sum has no zero.)

In principle, one could achieve this by using a numerical optimizer (EXCEL SOLVER, for example) to find a minimum for the absolute value of a partial sum to N in Equation (16) or Equation (19), or a weighted average of the two, ranging over trial values for x_m, a, h, c, and (possibly) the weighting variable ϑ. With so many variables, however, the numerial minimization may occur only unstably, unreliably, slowly, or not at all.

An obvious improvement would be for each trial value of a to calculate the usual Esscher values for h and c corresponding to the trial value of a, thus eliminating two dimensions in the minimization problem and also eliminating some of the $j = 1$ and $j = 2$ terms. (In the $-\frac{1}{c}$ section of Equation (19), however, note that the $j = 2$ term now still would remain.)

For Equation (16) or any weighted average involving Equation (16) the stability, reliability, or speed of the numerical minimization, as well as the accuracy of its convergence to a point of maximum likelihood, still might suffer owing to the oscillatory character of our original Esscher expansion, Equation (1).

Restricting a numerical minimization just to the absolute value of a partial sum to N in Equation (19), the Esscher expansion of the derivative, cures that objection by not taking derivatives of an oscillating approximation. It still may fail to deliver stability, reliability and convergence at the level we expect from an Esscher approximation, however, because Equation (19) retains a non-zero $j = 2$ term in its $-\frac{1}{c}$ section and because minimization still occurs over two dimensions, the trial values for a and x_m.

5.5 Approximate Points of Maximum Likelihood

Taking even more advantage of the flexibility in the Esscher approach provides the solution. Assume that the arbitrary paramenter a in the Esscher expansion takes the value $a = x_m$, the unknown point of maximum likelihood. This eliminates one more dimension in the minimization problem and, almost magically, kills off the remaining $j = 2$ terms.

When $a = x_m$ Equation (19) (and Equation (16) for that matter) evaluated at the point $x = x_m$ will contain myriad terms $H_{2n+j+1}\left(\frac{x_m-a}{c}\right) = H_{2n+j+1}(0)$ and $H_{2n+j}\left(\frac{x_m-a}{c}\right) = H_{2n+j}(0)$. But $H_m(0) = 0$ when m is odd and $H_m(0) = (-1)^{\frac{m}{2}} m?$ when m is even, so a vast simplification occurs.

In the h terms of Equation (19) (and also Equation (16)) only the terms with even j will survive non-zero when evaluating $f_X^{(1)}(x) = f_X^{(1)}(x_m) = f_X^{(1)}(a) = 0$. In the $\frac{1}{c}$ terms of Equation (19) (and also Equation (16)) only the terms with odd j will survive non-zero when evaluating $f_X^{(1)}(x) = f_X^{(1)}(x_m) = f_X^{(1)}(a) = 0$. This kills off the remaining $j = 2$ terms.

Beginning with Equation (19), the Esscher expansion for the derivative $f_X^{(1)}(x)$, assume that a is the unknown point of maximum likelihood $a = x_m$ so that $f_X^{(1)}(a) = 0$. Given a trial value for a, choose h and then c according to the usual Esscher specifications $i\,\widehat{f_{X-a}}^{(1)}(ih) = 0$ and $\frac{i^2\,\widehat{f_{X-a}}^{(2)}(ih)}{c^2\,\widehat{f_{X-a}}(ih)} - 1 = 0$. In Equation (19) replace the variable j by $2j$ or $2j-1$ according as it is even or odd. Take the limit through even values of N and replace the variable $2N$ by N, resulting in

$$
\begin{aligned}
0 \quad = \quad & \frac{\widehat{f_{X-a}}(ih)}{c}\varphi(0)\cdot \\
& \{ -h\,[\,1 + \lim_{N\to\infty}\sum_{j=2}^{N}\frac{(-1)^j}{(2j)!}\,[\,\frac{i^{2j}\widehat{f_{X-a}}^{(2j)}(ih)}{c^{2j}\,\widehat{f_{X-a}}(ih)} - (2j)?\,]\sum_{n=0}^{N-j}\frac{(2n)?}{(2n)!}(2(n+j))?\,] \\
& -\frac{1}{c}[\,\lim_{N\to\infty}\sum_{j=2}^{N}\frac{(-1)^j}{(2j-1)!}\frac{i^{2j-1}\widehat{f_{X-a}}^{(2j-1)}(ih)}{c^{2j-1}\,\widehat{f_{X-a}}(ih)}\sum_{n=0}^{N-j}\frac{(2n)?}{(2n)!}(2(n+j))?\,]\,\}
\end{aligned} \tag{20}
$$

For partial sums to N in Equation (20) use a numerical optimizer to minimize the absolute value of the right hand side by varying a. This is a one-dimensional minimization with a good chance to be stable and fast. Vanishing of terms involving $i\widehat{f_{X-a}}^{(1)}(ih)$ and $i^2\widehat{f_{X-a}}^{(2)}(ih) - 2?$ suggests that solutions a will show good convergence in N to the true value for x_m.

There is a calculational cost in solving for h and c for each new trial value for a, but this does not increase the dimensionality of the problem and should not cause instability.

5.6 What About the Alternative Approaches?

If a is assumed to be the unknown point x_m of maximum likelihood then we can conclude that Equation (16), the derivative of the our original Esscher expansion Equation (1), does not in fact converge at the point $x = a$.

To see this, begin with Equation (16), assume that it converges at the point $x = a$, and repeat the assumptions and steps in the derivation of Equation (20), resulting in

$$
\begin{aligned}
0 &= \frac{\widehat{f_{X-a}}(ih)}{c}\varphi(0)\cdot \\
&\{\ h\,[\ 1+\lim_{N\to\infty}\sum_{j=2}^{N}\frac{(-1)^j}{(2j)!}\,[\ \frac{i^{2j}\widehat{f_{X-a}}^{(2j)}(ih)}{c^{2j}\ \widehat{f_{X-a}}(ih)}-(2j)?\ ]\sum_{n=0}^{N-j}\frac{(2n)?}{(2n)!}\,(2\,(n+j))?\ ] \\
&-\frac{1}{c}\,[\ \lim_{N\to\infty}\sum_{j=2}^{N}\frac{(-1)^j}{(2j-1)!}\,\frac{i^{2j-1}\widehat{f_{X-a}}^{(2j-1)}(ih)}{c^{2j-1}\ \widehat{f_{X-a}}(ih)}\sum_{n=0}^{N-j}\frac{(2n)?}{(2n)!}\,(2\,(n+j))?\ ]\ \}
\end{aligned}
\tag{21}
$$

If Equation (21) converges then so does an equally weighted average of Equation (21) and Equation (20):

$$
\begin{aligned}
0 &= \frac{\widehat{f_{X-a}}(ih)}{c}\varphi(0)\cdot \\
&\{\ -\frac{1}{c}\,[\ \lim_{N\to\infty}\sum_{j=2}^{2N}\frac{(-1)^j}{(2j-1)!}\,\frac{i^{2j-1}\widehat{f_{X-a}}^{(2j-1)}(ih)}{c^{2j-1}\ \widehat{f_{X-a}}(ih)}\sum_{n=0}^{N-j}\frac{(2n)?}{(2n)!}\,(2\,(n+j))?\ ]\ \}
\end{aligned}
\tag{22}
$$

But then Equation (22) implies that the $-\frac{1}{c}$ term of Equation (20) vanishes leaving Equation (20) to read

$$
\begin{aligned}
0 &= \frac{\widehat{f_{X-a}}(ih)}{c}\varphi(0)\cdot \\
&\{\ -h\,[\ 1+\lim_{N\to\infty}\sum_{j=2}^{N}\frac{(-1)^j}{(2j)!}\,[\ \frac{i^{2j}\widehat{f_{X-a}}^{(2j)}(ih)}{c^{2j}\ \widehat{f_{X-a}}(ih)}-(2j)?\ ]\cdot \\
&\sum_{n=0}^{N-j}\frac{(2n)?}{(2n)!}\,(2\,(n+j))?\ ]\ \}\ \text{ which, at } x=a, \\
&= -hf_X(a)\ \text{ by Equation (1)}
\end{aligned}
$$

which contradicts the assumption that a is a point of maximum likelihood for f_X (unless $h=0$, which occurs only when $a=\mu_X$) .

Therefore the initial assumption that Equation (16), the derivative of the our original Esscher expansion Equation (1), converges at the point $x=a=x_m$ fails unless $a=\mu_X$. Whether it converges at other points remains open.

For the method we are proposing, therefore, taking a to be the unknown point of maximum likelihood, there remains only Equation (19) the Esscher expansion for the derivative $f_X^{(1)}$, leading to Equation (20) as the basis for numerical work.

6 Practicalities, Including Critical Convergence

On their face, the methods described here require knowledge of derivatives of the Fourier transform $\widehat{f_{X-a}}$ on the imaginary axis in the form $i^j \, \widehat{f_{X-a}}^{(j)}(ih)$. This equals the j-th derivative of the moment generating function of $X-a$ evaluated at h. The derivatives of the Fourier transform (or moment-generating function) are required: first, to solve for h and then c, given a, as defined in Section 2.1; and, second, to provide the coefficients in Equation (1) or Equation (2).

What if the moment-generating function is unknown (hence also the Fourier transform)?

Furthermore, we have asserted that the methods work whenever X has finite moments. But the convergence of the Esscher expansion Equation (1) or Equation (2) depends upon the derivation in Section 4. That derivation depended upon a convergent Taylor's series expansion involving the Fourier transform of $X-a$ at a point ih on the imaginary axis.

What if the Fourier transform is not analytic (has no convergent Taylor's series expansion) on the imaginary axis? Equivalently, what if X has finite moments but no moment-generating function?

6.1 Unknown Moment-Generating Function

With finite moments we can proceed to the Esscher approximations in this paper even without explicit knowledge of the moment-generating function, or Fourier transform.

Assuming that the convergent moment-generating function exists, i.e. that the Fourier transform extends analytically along the imaginary axis, make a Taylor's series expansion of $i^j \widehat{f_{X-a}}^{(j)}$ about the origin:

$$i^j \widehat{f_{X-a}}^{(j)}(ih) = \lim_{M\to\infty} \sum_{m=0}^{M-j} \frac{i^{j+m}}{m!} \widehat{f_{X-a}}^{(j+m)}(0)\, h^m \tag{23}$$

where $i^{j+m} \widehat{f_{X-a}}^{(j+m)}(0)$ is the $(j+m)^{\text{th}}$ moment of $X-a$. (This, of course, is determined easily from moments of X.)

Taking a large enough M (i.e. enough moments) provides as close an approximation as desired to the derivatives of the Fourier transform (or moment-generating function) required in the Esscher expansion and the approximation formulae in this paper.

6.2 Moments But No Moment-Generating Function

The derivation of the Esscher expansion in Section 4 requires the Taylor's series for $\frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))}$ around the point $t = ih$ to converge.

If the Fourier transform $\widehat{f_X}$, which always exists on the real axis, extends to an analytic function on the imaginary axis then $\frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))}$ is analytic on the imaginary axis and its Taylor's series converges. This follows from the fact that $\widehat{\varphi}$ has those properties. $\widehat{f_X}$ analytic on the imaginary axis also implies that X has a convergent moment-generating function, implying finite moments.

The converse implications do not hold. The lognormal random variable, for example, has finite moments but no moment-generating function.

Nevertheless, given finite values for moments of order $\leq M$ one can use Equation (23) to compute "as if" the results were to represent approximate values of derivatives of an actual moment-generating function or, equivalently, approximate values of derivatives on the imaginary axis of an actual Fourier transform. The resulting values would determine in turn, to any desired order $< M$, values for any of the Esscher approximations in this paper.

The question then would be, to what would those "as if" Esscher approximation values constitute an approximation?

6.3 Approximation to an Approximation

The answer is, they would approximate the density, tail probabilities, tail moments, or point of maximum likelihood (as the case may be) of some random variable possessing a moment generating function whose actual moments would match the given moments up to order $\leq M$.

The existence of random variables possessing moment generating functions whose actual moments match the given moments up to order $\leq M$ can be understood from a simple line of reasoning. Rigorous proof, however, of all the assertions in this section and Section 6.4 would require resort to Tauberian theory, which goes beyond the intended scope of this paper.

The simple line of reasoning goes as follows: Given a sequence of finite moments up to some order M, one could expand that sequence to a hypothetical sequence of finite moments of all orders by adding more moments arbitrarily. So long as one chose the arbitrary additional moments small enough in absolute value that the resulting moment-generating series converged absolutely on a neighborhood of zero then that series would define an analytic function on a neighborhood of zero in the complex plane, including a neighborhood on the imaginary axis.

By Fourier inversion the analytic function so created would define a density function (i.e. one that integrates to 1) with Fourier transform equal to the analytic function (hence, with moment-generating function corresponding to the hypothetical sequence of moments.)

If, further, one chose carefully (e.g. even moments must be positive, even and odd moments each should ultimately become monotonic in the tail of the sequence, etc.) then one could assure that the density so defined would be non-negative, i.e. would be the density of some random variable (call it X_M).

Esscher approximations based on M moments and Equation (23) would be approximations to the density, tail probabilities, tail moments, or point of maximum likelihood (as the case may be) of that random variable X_M. There would be many such X_M depending on the choices made. Esscher approximations based on M moments and Equation (23) would be approximations for any of them.

What can we say about the sequence of such X_M (actually, of sets of such X_M) as $M \to \infty$?

6.4 An Approximation to What?

When the given sequence of moments does not define a moment-generating function, then the X_M must themselves approximate as $M \to \infty$ a unique random variable X:

- whose moments are the given sequence of moments, and
- whose density f_X has oscillations, if any, that have bounded derivatives.

How to square this asserted uniqueness with the fact that for random variables without moment-generating functions an infinite sequence of moments may not determine a unique random variable? (See Durrett (2010) pp. 120-124 for counterexamples.)

In the counterexamples, the non-uniqueness always represents a set of densities all but one of which oscillates without a bound on the derivative.

In our case, destructive interference within oscillatory terms in Equation (1) as illustrated in Section 3.2 (the dotted lines in Figure 1 and Figure 2) explains the assertion that any oscillations in the limit f_X will have bounded derivatives. (The only exception could be offsetting constructive interference across terms in Equation (1) arising in the case of a discrete random variable with a moment-generating function.)

Thus, the Esscher approximations derived from Equation (23) and a sequence of finite moments that do not define a moment-generating function will approximate the density, tail probabilities, tail moments, or point of maximum likelihood (as the case may be) of the unique random variable characterized by the given sequence of moments and a density with oscillations, if any, that have bounded derivatives.

In particular, if a continuous and a discrete random variable have the same moments (so, perforce, no moment-generating function) then the Esscher expansion will converge to the continuous one.

To repeat, complete rigor in this subsection would require Tauberian theory.

7 When to Use an Esscher Approximation

Any application of the Esscher expansion will involve some complexity and most applications will be computationally intense, including numerical solutions, often at more than one stage. When is it likely to be worthwhile to invest in managing the complexity and setting up the numerical solutions?

The answer comes directly from the nature of the Esscher expansion. It is built around knowledge of the moment-generating function and its derivatives or, using Equation (23), around knowledge of a lot of moments.

So the situation in which to use an Esscher approximation is a situation in which it is easier to know or work with the moment-generating function, or just to know a lot of moments, than it is to know or work with a density or a distribution function. Such situations include:

- Sums of random variables (the typical use of the saddlepoint approximation in the statistics literature)
- Compound random variables
- Compound random processes (Esscher's original application)
- More general random processes, which after some work may lend themselves to an approximation of the moment-generating function or of moments
- Monte Carlo simulations, which can readily generate a lot of moments that can be warehoused for later use in Esscher approximations for quantities not anticipated to be needed at the time of the original Monte Carlo run

In any of these situations, an Esscher approximation may provide a way to calculate density values, tail probabilities, tail moments, or (now) points of maximum likelihood with higher accuracy than other available alternatives with the same effort.

Another feature of the Esscher approach is that it provides a way to avoid the bias (at least some of it) in using a pure "method of moments" approach for calculations involving tail probabilities or tail moments. This thought was the original motive to look in Section 5 for a way to use an Esscher approximation for a point of maximum likelihood.

It is true that an Esscher approximation can be difficult to organize and can be computationally intense. The application to point of maximum likelihood proposed in Section 5, for example, requires iterated numerical solutions for trial values of the constants.

But we are a world of actuaries willing to devote entire CPU farms, and the effort to organize their calculations, to brute things like "stochastic within stochastic" simulations. Perhaps we also can devote some CPU and some effort at organizing calculations for computationally intense analytic approaches.

References

[1] Butler, R. W. (2007). *Saddlepoint Approximations with Applications*. Cambridge University Press.

[2] Daniels, H. E. (1954). Saddlepoint approximations in statitstics. *Ann. Math. Statist.* **25**, 631-650.

[3] Durrett, R. (2010). *Probability Theory and Examples.* 4th Ed. Cambridge University Press

[4] Esscher, F. (1932). On the probability function in the collective theory of risk. *Skand. Act. Tidskr.* **15**, 175-195.

[5] Esscher, F. (1963). On approximate computations when the corresponding characteristic functions are known. *Skand. Act. Tidskr.* **65**, 78-86.

[6] Howell, K. B. (2001). *Principles of Fourier Analysis.* Chapman & Hall/CRC.

[7] Jensen, J. L. (1995) *Saddlepoint Approximations.* Oxford University Press.

[8] Klugman, Panjer, Willmot. (2008). *Loss Models - from data to decisions.* 3rd Ed. Wiley

[9] Meikle, H. D. (2004). *A New Twist to Fourier Transforms.* Wiley-VCH.

[10] Rudin, W. (1966). *Real and Complex Analysis.* McGraw-Hill

[11] Strichartz, R. S. *A Guide to Distribution Theory and Fourier Transforms.* World Scientific