# Esscher Approximations for Maximum Likelihood Estimates

## Exploratory Ideas

Jim Bridgeman

University of Connecticut Actuarial Seminar

September 9, 2011

- A series expansion for any probability density function with finite moments
  - possible convergence questions but manageable in practice
- Known to actuaries by Esscher's name (1932)
- Known to statisticians as the saddlepoint approximation (Daniels 1954)
- Integrate the series to get approximate probability values under the density
- A location parameter in the expansion can be chosen arbitrarily
- Choose a value for it that speeds up the convergence of the integrated series

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

- Try to approximate the point where the derivative of the probability density function is 0
- Either: take the derivative of the series expansion for the density
- Or: make a series expansion for the derivative of the density
- Or: take a weighted average of the two
- If the limits exist they will be same in all cases but the partial sums will not be the same! Maybe one will converge faster than another
- Find the value for the random variable that minimizes the absolute value of the partial sum (or sums)
- Assume that the arbitrary location parameter is the unknown point of maximum likelihood
    - Vastly simplifies the minimization problem

- What Does the Esscher Look Like?
- Why Is the Esscher So Good?
- Where Does the Esscher Come From?
- How To Use the Esscher for Maximum Likelihood

# What Does the Esscher Look Like?

For a random variable X and an arbitrary location parameter a the density of X can be represented as

$$f_X(x) =$$

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

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 $X - a$; the characteristic function at $-t$

so $\widehat{f_{X-a}}(ih)$ is the moment generating function of $X - a$ evaluated at $h$

$\varphi(z)$ is the standard normal density

$\widehat{f_{X-a}}^{(j)}(t)$ is the $j$th derivative of the Fourier transform for $X - a$

so $i^j\ \widehat{f_{X-a}}^{(j)}(ih)$ is the $j$th derivative of the moment generating function of $X - a$, evaluated at $h$

# What Does the Esscher Look Like?

For a random variable X and an arbitrary location parameter a the density of X can be represented as

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

where $j? = 0$ for odd $j$ and $j? = (j-1)(j-3)\cdots(1)$ for even $j$

$h$ is chosen so that $i \ \widehat{f_{X-a}}^{(1)}(ih) = 0$ (eliminating the $j = 1$ term)

$c$ is chosen so that $\frac{i^2 \ \widehat{f_{X-a}}^{(2)}(ih)}{c^2 \ \widehat{f_{X-a}}(ih)} - 1 = 0$ (eliminating the $j = 2$ term)

- if $a = \mu_X$ then $h = 0$ and $c = \sigma_X$ (called the Edgeworth expansion)

$$H_m(z) = \sum_{k=0}^{\left\lfloor \frac{m}{2} \right\rfloor} (-1)^k \frac{m!(2k)?}{(m-2k)!(2k)!} z^{m-2k} = m\text{th Hermite polynomial}$$

# What Does the Esscher Look Like?

In the literature the order of summation is n first, then j

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

The ? notation makes $n$ and $j$ both odd or both even, so result is real

- To get to our way, change the order of summation, change variables so $2n + j$ replaces $n$, and simplify

$$f_X(x) =$$

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

- Need to use $\lim_{N\to\infty}$ or else you won't know where the new $n$ stops

# What Does the Esscher Look Like?

## To find the probability that u<X<v just integrate

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

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

- The integrals have been codified as "Esscher functions" and can be handled numerically
- It turns out that this integrated series has far faster convergence when the location parameter $a$ is chosen to be either $u$ or $v$
  - Even better when the other limit is $\infty$, i.e. in the tail.
- The proper choices for $h$ and $c$ allow any choice needed for the location parameter $a$
- For tail moments (CTE, option pricing) you get a similar integral

# Why is the Esscher So Good?

Summing the integrals over n first, then j, suggests one reason convergence is good

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

$$\frac{\widehat{f_{X-a}}(ih)}{c}\left\{\begin{array}{c}\int\limits_u^v e^{h(x-a)}\varphi\left(\frac{x-a}{c}\right)dx+\sum\limits_{n=3}^{\infty}\frac{1}{n!}\int\limits_u^v e^{h(x-a)}H_n\left(\frac{x-a}{c}\right)\varphi\left(\frac{x-a}{c}\right)dx\cdot\\ \left(\sum\limits_{j=3}^{n}\frac{i^{n-j}n!(n-j)?}{j!(n-j)!}\left[\frac{i^j\ \widehat{f_{X-a}}^{(j)}(ih)}{c^j\ \widehat{f_{X-a}}(ih)}-j?\right]\right)\end{array}\right\}$$

- The choice of $h$ and $c$ has achieved two convergence-enhancing steps simultaneously (also true even prior to integrating)

1. eliminated the $n = 1$ and 2 terms of a typical series expansion
2. reduced all further error terms ($n > 2$) by eliminating the $j = 0$, 1 and 2 terms of the coefficient for each remaining term $n > 2$, with greatest relative effect on the most important terms (those divided by the smallest $n!$)

# Why is the Esscher So Good?

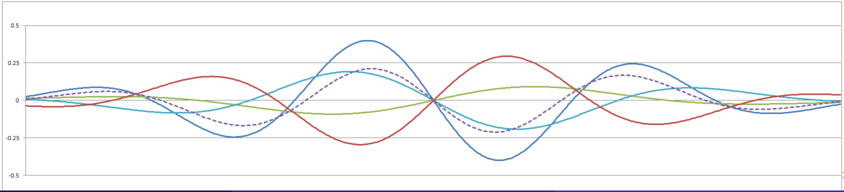
Our summation (j first, then n) suggests another reason convergence is good

- $H_m(z)\,\varphi(z) = (-1)^m\,\varphi^{(m)}(z)$ so

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

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

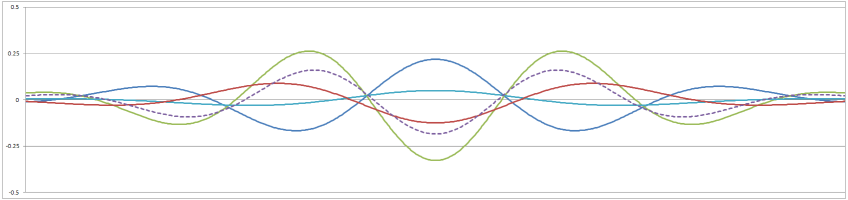
- For example, here are $\frac{1}{3!}\frac{(2n)?}{(2n)!}\varphi^{(2n+3)}(z)$ and their sum for successive odd values 3, 5, 7, 9 as would appear in $N = 10$, $j = 3$, $-3 < z < 3$



# Why is the Esscher So Good?

## Our summation (j first, then n) suggests another reason convergence is good

- On the same scale, here are $\frac{1}{4!}\frac{(2n)?}{(2n)!}\varphi^{(2n+4)}(z)$ and their sum for even values 4, 6, 8, 10 as would appear in $N = 10$, $j = 4$, $-3 < z < 3$




- Even before integrating, at each point the terms dampen each other a bit. They will shrink even more as $j!$ gets larger
- The terms oscillate over $z$ and decay (exponentially as $O\left(e^{-\frac{1}{2}z^2}\right)$) for large $|z|$ (important b/c they will be multiplied by $e^{hz}$)
- Oscillations will tend to zero out when integrated over entire cycles
- Best offsetting when integrated from/to 0; especially to/from $\infty$

# Why is the Esscher So Good?

## Looked at together

↑ $j$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | *X* |
| | | | | | | | | | **X** | *0* |
| | | | | | | | | **X** | **0** | *X* |
| | | | | | | | **X** | **0** | **X** | *0* |
| | | | | | | **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 | 2 | $n$ | $+$ | $j$ | $\rightarrow$ | | |

- Each new column is sparse compared to its theoretical weight, especially when the $n!$ dividing it is small
- Each new column dampens the oscillations of half the prior columns

First use just some algebra and the usual rules for Fourier Transforms

$\widehat{f_X}(t) = e^{-iat} \frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi\left(\frac{x}{c}\right)}(t-ih)} \widehat{\varphi\left(\frac{x}{c}\right)}(t-ih)$ by translation$\sim_{FT}$ $\times$exp

$= \frac{1}{c} e^{-iat} \left\{ \frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))} \right\} \widehat{\varphi\left(\frac{x}{c}\right)}(t-ih)$ by reciprocal scaling. Now expand

$= \frac{1}{c} e^{-iat} \left\{ \sum_{n=0}^{\infty} \frac{1}{n!} \left[ \frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))} \right]^{(n)}_{t=ih} (t-ih)^n \right\} \widehat{\varphi\left(\frac{x}{c}\right)}(t-ih)$ by Taylor's

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

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

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

# Where Does the Esscher Come From?

## Invert the Fourier Transform

- Back in density space

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

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

because

$$\varphi^{(n)}\left(\frac{x-a}{c}\right) = (-1)^n \varphi\left(\frac{x-a}{c}\right) H_n\left(\frac{x-a}{c}\right)$$

- Now use Leibniz's product rule creatively to unravel the coefficient

# Where Does the Esscher Come From?

## Use Leibniz's Product Rule to get the Coefficient

For $n > 0$

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

$$0 = \left[\frac{\widehat{\varphi}(ct)}{\widehat{\varphi}(ct)}\right]^{(n)}_{t=0} = \widehat{\varphi}(0)\left[\frac{1}{\widehat{\varphi}(ct)}\right]^{(n)}_{t=0} + \sum_{j=1}^{n}\frac{n!}{j!(n-j)!}c^j\widehat{\varphi}^{(j)}(ct)\,|_{t=0}\left[\frac{1}{\widehat{\varphi}(ct)}\right]^{(n-j)}_{t=0}$$

Now multiply by $\widehat{f_{X-a}}(ih)$ and subtract, noting that $\widehat{\varphi}(0) = 1$

$$\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]^{(n)}_{t=0} =$$

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

but now using $\widehat{\varphi}^{(j)}(0) = i^{-j}j?$and $\left[\frac{1}{\widehat{\varphi}(ct)}\right]^{(n-j)}_{t=0} = c^{n-j}(n-j)?$ get

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

# Where Does the Esscher Come From?

## Substitute Back into the Expression for the Density

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

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

- Choose $h$ and $c$ to kill $j = 1$ and 2, change the order of summation, change variables so $2n + j$ replaces $n$, and simplify

$$f_X(x) =$$

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

- Take the Talylor's series for $\frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))}$ around $ih$ in Fourier space
- Expand $\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]^{(n)}_{t=0}$ by Leibniz's rule, using the trick $\left[\frac{\widehat{\varphi}(ct)}{\widehat{\varphi}(ct)}\right]^{(n)}_{t=0} = 0$ to kill the first term and make the rest of the terms into differences
- Given $a$ choose $h$ and then $c$ to kill the first two difference terms
- $e^{-ia}...\widehat{\left(\frac{x}{c}\right)}(t-ih)$ becomes $e^{h(x-a)}...\left(\frac{x-a}{c}\right)$ back in density space.
- The $e^{h(x-a)}$ is called "exponential tilting" in the literature, so exponential tilting comes from a Taylor's series around $ih$ in Fourier space.
- Changing the order of summation to $j$ first, then $n$ seems most natural to me

## (1) Derivative of the Esscher

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 (leaving $j = 1$ and 2 still in the picture for the moment)

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

- The $h$ term comes from the derivative of $e^{h(x-a)}$ and the $\frac{1}{c}$ term from the derivative of $\varphi\left(\frac{x-a}{c}\right) H_{2n+j}\left(\frac{x-a}{c}\right) = (-1)^j \varphi^{(2n+j)}\left(\frac{x-a}{c}\right)$
- The series does not necessarily converge! Derivative of approx maybe $\neq$ approx for derivative when oscillations are involved

**(2) Esscher for the Derivative**

Doing for $f_X^{(1)}(x)$ exactly what we did for $f_X(x)$ it's easy to get to

$f_X^{(1)}(x) =$

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

which does converge. But how to deal with $\widehat{f_{X-a}^{(1)}}^{(j)}(ih)$?

$\widehat{f_{X-a}^{(1)}}(t+ih) = i(t+ih)\,\widehat{f_{X-a}}(t+ih)$ is a basic Fourier property so

$$\widehat{f_{X-a}^{(1)}}^{(j)}(t+ih)\big|_{t=0} = -h\;\widehat{f_{X-a}}^{(j)}(ih) + ji\;\widehat{f_{X-a}}^{(j-1)}(ih) \text{ by Leibniz's rule}$$

$$= -h\;\widehat{f_{X-a}}^{(j)}(ih) + \tfrac{1}{c} jc\,(-i)^{-1}\;\widehat{f_{X-a}}^{(j-1)}(ih), \text{ including } j = 0$$

**(2) Esscher for the Derivative - continued** - plug into the expansion:

- The $-h$ term is exactly $-h$ times the original Esscher.
- The $\frac{1}{c}$ term is like $\frac{1}{c}$ times the original Esscher except $j$ is lowered by 1 and there're no $j?$ terms (they went with the $-h$)

$$f_X^{(1)}(x) = \frac{\widehat{f_{X-a}}(ih)}{c} e^{h(x-a)} \varphi\left(\frac{x-a}{c}\right) \cdot$$

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

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

**(3) Weighted Average**

- If (1) converges then any weighted average $\theta(1) + (1-\theta)(2)$ also will converge.

# How To Use Esscher for Maximum Likelihood

## Whichever Way: For a given N, minimize over a, h, and c

Maximum Likelihood occurs at a value $x_m$ where $f_X^{(1)}(x_m) = 0$
Try to approximate $x_m$ given only $N$ terms in the sums:

- Try to minimize $|(1)|$, $|(2)|$, or $|\theta(1) + (1-\theta)(2)|$ over $x_m$, $a$, $h$, $c$, and (maybe) $\theta$ using a numerical tool such as SOLVER
  - But with so many variables it might not be stable or fast
- Try to minimize $|(1)|$ over $x_m$ and $a$ using the usual Esscher values for $h$ and $c$ corresponding to each trial value of $a$
  - But this may be unstable, slow, or wrong because the derivative of an approximation may not converge, or not quickly, to the derivative when the approximation is oscillatory as ours is (coming from Fourier space).
- Try to minimize $|(2)|$ over $x_m$ and $a$ using the usual Esscher values for $h$ and $c$ corresponding to each trial value of $a$
  - But this may be slow because $i^2\,\widehat{f_{X-a}}^{(2)}(ih)$ hasn't been eliminated in the $\frac{1}{c}$ term

# How To Use Esscher for Maximum Likelihood

## Instead, Choose a to be the Unknown Point of Maximum Likelihood

There is a vast simplification if we set $a = x_m$ (so $f_X^{(1)}(a) = 0$) because $H_{2m}(0) = (-1)^m (2m)?$, $H_{2m+1}(0) = 0$

For simplicity take the limits through even integers $2N$

(1) $0 \sim \frac{\widehat{f_{X-a}}(ih)}{c} \varphi(0) \cdot$

$$\{ h [ 1 + \lim_{N\to\infty} \sum_{j=1}^{N} \frac{(-1)^j}{(2j)!} [ \frac{i^{2j} \widehat{f_{X-a}}^{(2j)}(ih)}{c^{2j} \widehat{f_{X-a}}(ih)} - (2j)? ] \sum_{n=0}^{N-j} \frac{(2n)?}{(2n)!} (2(n+j))? ]$$

$$- \frac{1}{c} [ - \lim_{N\to\infty} \sum_{j=1}^{N} \frac{(-1)^j}{(2j-1)!} \frac{i^{2j-1} \widehat{f_{X-a}}^{(2j-1)}(ih)}{c^{2j-1} \widehat{f_{X-a}}(ih)} \sum_{n=0}^{N-j} \frac{(2n)?}{(2n)!} (2(n+j))? ] \}$$

(2) $0 = \frac{\widehat{f_{X-a}}(ih)}{c} \varphi(0) \cdot$

$$\{ -h [ 1 + \lim_{N\to\infty} \sum_{j=1}^{N} \frac{(-1)^j}{(2j)!} [ \frac{i^{2j} \widehat{f_{X-a}}^{(2j)}(ih)}{c^{2j} \widehat{f_{X-a}}(ih)} - (2j)? ] \sum_{n=0}^{N-j} \frac{(2n)?}{(2n)!} (2(n+j))? ]$$

$$- \frac{1}{c} [ - \lim_{N\to\infty} \sum_{j=1}^{N} \frac{(-1)^j}{(2j-1)!} \frac{i^{2j-1} \widehat{f_{X-a}}^{(2j-1)}(ih)}{c^{2j-1} \widehat{f_{X-a}}(ih)} \sum_{n=0}^{N-j} \frac{(2n)?}{(2n)!} (2(n+j))? ] \}$$

(3) The weighted average with $\theta = \frac{1}{2}$ is particularly simple

$$0 \sim \frac{\widehat{f_{X-a}}(ih)}{c^2} \varphi(0) \lim_{N\to\infty} \sum_{j=1}^{N} \frac{(-1)^j}{(2j-1)!} \frac{i^{2j-1} \widehat{f_{X-a}}^{(2j-1)}(ih)}{c^{2j-1} \widehat{f_{X-a}}(ih)} \sum_{n=0}^{N-j} \frac{(2n)?}{(2n)!} (2(n+j))?$$

# How To Use Esscher for Maximum Likelihood

## Choose h and c to Eliminate the First Two Derivatives of Moment Generating Function

If (1) converges then so does (3) but looking at (2) that is impossible since the $h$ term in (2) equals $-hf_X(a) \neq 0$.

Focusing on (2), the even terms of the $\frac{1}{c}$ piece have been eliminated so now the usual choices for $h$ and $c$ given a trial value for $a$ completely eliminate $i^2 \widehat{f_{X-a}}^{(2)}(ih)$ and $i\, \widehat{f_{X-a}}^{(1)}(ih)$, producing:

(2)

$$0 = \frac{\widehat{f_{X-a}}(ih)}{c} \varphi(0) \cdot$$

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

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

- Given $N$ use a numerical tool to find a value of $a$ that minimizes $|(2)|$
- That value for $a$ is the approximate point of maximum likelihood in a $2N^{\text{th}}$ order approximation.

# What if No/No Known Moment Generating Function?

## Approximate it Using a Taylor's Series Involving Moments as Coefficients

- The method needs derivatives of the moment generating function.
- What if the moment generating function is unknown?
  - Approximate any derivative of the moment generating function by expanding it in a Taylor's series around $h = 0$

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

$$i^{j+m} \widehat{f_{X-a}}^{(j+m)}(0) \text{ is the } (j+m)^{\text{th}} \text{ moment of } X - a$$

- But what if that Taylor's series doesn't converge?
  - This would be the case when there is no moment-generating function
  - In terms of Fourier transforms this means that the Fourier transform is not an analytic function and its Taylor expansion doesn't exist off the real axis
  - The lognormal distribution would be an example

# What if No/No Known Moment Generating Function?

## Use the Series Expansion with Moments Anyway (or Go To Log-Likelihood)

- As long as you know the moments themselves, use the same series up to a value $m = M$ representing the order of approximation you want (and moments you know)

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

$i^{j+m} \widehat{f_{X-a}}^{(j+m)}(0)$ is the $(j+m)^{\text{th}}$ moment of $X - a$

- To any order there is a new density that has a moment generating function and moments matching $X - a$'s moments to that order
  - Just add arbitrary higher moments that give convergence
- You will be approximating maximum likelihood for that new density
  - For a maximum likelihood estimate, far from the tails, error introduced by discrepancies at higher moments should be tolerable?
  - Approximates the non-oscillatory density with given moments?
- Alternatively, do the entire Esscher for the log of the density.

# When Might You Use The Esscher?

There are many situations when it is easier to know the moment-generating function, or just a lot of moments, than to know the probability density:

- Sums of random variables (the typical statistical applications)
- Compound random variables
- Compound random process (Esscher's application)
- More general random processes (maybe not "easy" but still perhaps "less difficult")
- Monte Carlo simulations (a lot of moments, at least)
- Computationally intense? Perhaps,
  - but we are in a world of actuaries willing to devote entire CPU farms to "stochastic within stochastic" simulations
  - why not devote some CPU to computationally intense analytic approaches?

# References

**For Esscher/Saddlepoint Aproximation**

Butler (2007) Saddlepoint Approximations with Applications (Cambridge)
Daniels (1954) "Saddlepoint approximations in statistics," Ann. Math. Statist. 25, 631-650
Esscher (1932) "On the probability function in the collective theory of risk," Skand. Act. Tidskr. 175-195
Esscher (1963) "On approximate computations when the corresponding characteristic functions are known," Skand. Act. Tidskr. 78-86
Jensen (1995) Saddlepoint Approximations (Oxford)

**For Fourier Transforms**

Howell (2001) Principles of Fourier Analysis (CRC)
Meikle (2004) A New Twist to Fourier Transforms (Wiley-VCH)
Rudin (1966) Real and Complex Analysis (McGraw-Hill)
Strichartz (2003) A Guide To Distribution Theory and Fourier Transforms (World Scientific)