Esscher Approximations for Maximum Likelihood Estimates

46th Actuarial Research Conference - Storrs CT

James G. Bridgeman FSA CERA MAAA

University of Connecticut

August 12, 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

- 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?

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 $X - a$

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 jth derivative of the Fourier transform for $X - a$

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

$$f_X(x) =$$

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

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)

(note that if $a = \mu_X$ then the choices are $h = 0$ and $c = \sigma_X$)

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

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

What Does the Esscher Look Like?

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

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

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

- 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 ∞
- The proper choices for *h* and *c* allow any choice needed for the location parameter *a*

Where Does the Esscher Come From?

Work in Fourier Transform Space and Use Taylor's Series

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

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

$$= \frac{1}{c} e^{-iat} \left\{ \frac{\widehat{f_{X-a}}(t)}{\widehat{\varphi}(c(t-ih))} \right\} \widehat{\varphi\left(\frac{x}{c}\right)}(t-ih) \text{ and now use Taylor's Series}$$

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

Now use the usual Fourier Transform rules and more algebra to get

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

And just invert the Fourier Transform to get $f_X(x)$ back in density space

- 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 $\widehat{f_{X-a}}(ih)$, subtract and simplify

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

$$\left\{\text{note that } \widehat{\varphi}^{(j)}(0) = i^{-j}j?\text{and } \left[\frac{1}{\widehat{\varphi}(ct)}\right]^{(n-j)}_{t=0} = c^{n-j}(n-j)?\right\}$$

If you plug this expression for $\left[\frac{\widehat{f_{X-a}}(t+ih)}{\widehat{\varphi}(ct)}\right]^{(n)}_{t=0}$ into the formula for $f_X(x)$, change the order of summation and simplify then you get the Esscher expansion we were looking for.

To begin, for the moment leave the choice of h and c open

(1) The derivative of the Esscher expansion for $f_X(x)$ is (might not converge)

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

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

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

(2) The Esscher expansion for $f_X^{(1)}(x)$ is (and will converge)
(using all the same steps as for the Esscher of $f_X(x)$)

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

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

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

If (1) converges the two have the same limit but different partial sums.
(3) So would any weighted average $\theta(1) + (1-\theta)(2)$

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) θ 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

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

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

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.

- 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

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

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

- What if that Taylor's series doesn't converge (i.e. what if the Fourier Transform is not analytic so there is no moment generating function)?
  - 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.
  - By Carleman's Condition, to any order there is a density with a moment generating function and moments matching $X-a$ to that order
  - For a maximum likelihood estimate, far from the tails, error introduced by discrepancies at higher moments should be tolerable
  - Maximum likelihood for the non-oscillatory density with given moments