Testing Plan
For Esscher Maximum Likelihood Estimator

March 20, 2022

The goal is to test numerically the effectiveness of using sample values of an unknown probability distribution $f_X(x)$ to estimate the mode, or maximum probability value, of that unknown distribution, using formula (20) from the paper and trial values of an estimate a for the mode.

To use formula (20) requires values of the functions $i^j \hat{f}^{(j)}_{X-a}(ih)$ for the unknown distribution $f_X(x)$, the estimate a, and a parameter h that depends upon a.

Those values, in turn, can be estimated using formula (23) from the paper. The appendix to the paper provides two ways to estimate formula (23) using moments $i^j \hat{f}^{(j)}_X(0) = \mathbb{E}\left[X^j\right]$ = average value of X^j of the unknown distribution $f_X(x)$. The idea is to use moments of the sample values from the unknown distribution to estimate the moments of the unknown distribution.

The testing will be organized in several stages.

1. Using known distributions with known mode and moments, test the effectiveness of the overall program when sampling is not required.
   (a) Explore the degree the of accuracy achieved using varying numbers of moments
   (b) Explore the difference in accuracy between the two ways to estimate (23) in the appendix to the paper
   (c) Test whether there is a difference in accuracy depending upon whether the known distribution has a moment-generating function or not
2. Using the same known distributions, perform the same tests using sample moments rather than known ones
   (a) Explore the degree of accuracy achieved using varying sample sizes and varying numbers of moments.
   (b) Test whether the difference in accuracy between the two ways to estimate (23) in the appendix to the paper is affected by using varying sample sizes.
   (c) Test whether there is a difference in accuracy depending upon whether the known distribution has a moment-generating function or not
3. Using known distributions with known mode, but with a limited number of moments, or even no moments, test the effectivess (if any) of the overall program using sample moments

1 The algorithm to find the mode

1.1 Choosing successive values of a

The idea is to keep selecting new trial values of a until we find one that makes the formula in (20) equal zero, or very close. We can select new trial values of a in systematic fashion because of the following fact:

For any a the value of the formula in (20) is the slope of the unknown distribution $f_X(x)$ at the point $x = a$.

Proof. For any a, h, and c and any x the value of the formula in (19) is the slope of the unknown distribution $f_X(x)$ at the point x.

Choose $x = a$ so that the value of the formula in (19) becomes the slope of the unknown distribution $f_X(x)$ at the point $x = a$.

Now choose h so that $i\hat{f}^{(1)}_{X-a}(ih) = 0$, and choose c so that $\frac{i^2\hat{f}^{(2)}_{X-a}(ih)}{c^2} - 1 = 0$.

With all these choices, the formula in (19) becomes the formula in (20), so the formula in (20) is the slope of the unknown distribution $f_X(x)$ at the point $x = a$. ■

This means that if the unknown distribution $f_X(x)$ is unimodal, then when $a <$ the mode the slope is > 0 for the unknown distribution $f_X(x)$ at the point $x = a$, so the formula in (20) > 0. Similarly, when $a >$ the mode the formula in (20)< 0.

So the strategy is to pick a first value of a and calculate the formula in (20), If the result is negative then keep picking smaller values of a until the formula in (20) turns positive. Then pick a value for a in between the one that gave negative (20) and the one that finally gave positive (20). Then keep on picking new values of a in between the most recent negative and positive results. Eventually you'll squeeze in on a value for a which gives 0 or close to it for the formula in (20). (If the very first choice gives a positive result then just pick larger values until you find a negative result. Then go on picking in between values.)

If the unknown distribution $f_X(x)$ is multimodal, then this procedure will converge on one of the modes, but it won't tell us (without more detailed examination of the steps) that there are others or what the others are.

1.2 The inputs to the formula in (20)

Given a value for a, to compute the value of the formula in (20) requires as inputs values for h, c, and $i^j\hat{f}^{(j)}_{X-a}(ih)$ for $0 \le j \le M$, where M is the highest moment we plan to use in the approximation.

Start with the values $i^k\hat{f}^{(k)}_X(0) = \mathbb{E}\left[X^k\right]$ for $0 \le k \le M$, which are the moments or average values of X^k for the distribution $f_X(x)$. These will be the moments of a random sample from the distribution or, in some of the testing, we will have a formula for these moments.

1.2.1 Calculating h

Using the formula in the Appendix with $j = 1$, and either trial and error or a polynomial root finder, find the value of h closest to 0 for which $0 = i\hat{f}^{(1)}_{X-a}(ih)$ Using the formula in the Appendix and the given value of a this is:

$$0 = \sum_{m=0}^{M-1} \frac{1}{m!} \left[\sum_{k=0}^{m+1} \frac{(m+1)!}{k!\,(m+1-k)!} (-1)^{m+1-k} a^{m+1-k} i^k \widehat{f_X}^{(k)}(0) \right] h^m$$

If there are no real solutions for h then take the real part of the complex solution with the smallest imaginary part.

(Using the alternative formula in the Appendix this would be

$$0 = \sum_{m=0}^{M-1} \frac{1}{m!} \left(i^{m+1} \widehat{f_X}^{(m+1)}(0) - a i^m \widehat{f_X}^{(m)}(0) \right) h^m - \frac{1}{M!} \left(a i^M \widehat{f_X}^{(M)}(0) \right) h^M$$

because $e^{-ah} \neq 0$ no matter what the values of a and h are.)

1.2.2 Calculating c

Using the given value for a and the calculated value for h, use the formula in the Appendix with $j = 2$ to calculate

$$\begin{aligned} i^2 \widehat{f_{X-a}}^{(2)}(ih) &= \sum_{l=0}^{2} \frac{1}{(2-l)!} i^{2-l} \widehat{f_X}^{(2-l)}(0) \sum_{m=0}^{M-2} \frac{(2+m)!}{m!\,(m+l)!} (-1)^{m+l} a^{m+l} h^m \\ &+ \sum_{l=1}^{M-2} \frac{1}{(2+l)!} i^{2+l} \widehat{f_X}^{(2+l)}(0) \sum_{m=l}^{M-2} \frac{(2+m)!}{m!\,(m-l)!} (-1)^{m-l} a^{m-l} h^m \end{aligned}$$

Now using this result, calculate $c = \sqrt{i^2 \widehat{f_{X-a}}^{(2)}(ih)}$. Use the positive value of the square root.

(Using the alternative formula in the Appendix this would be

$$\begin{aligned} i^2 \widehat{f_{X-a}}^{(2)}(ih) &= e^{-ah} \left[\sum_{l=0}^{2} i^{j-l} \widehat{f_X}^{(2-l)}(0) \sum_{m=0}^{2-l} \frac{2!}{(2-m-l)!\,(m+l)!m!} (-1)^{m+l} a^{m+l} h^m \right. \\ &\left. + \sum_{l=1}^{M-2} i^{2+l} \widehat{f_X}^{(2+l)}(0) \sum_{m=l}^{2+l} \frac{2!}{(2-m+l)!\,(m-l)!m!} (-1)^{m-l} a^{m-l} h^m \right] \end{aligned}$$

followed by $c = \sqrt{i^2 \widehat{f_{X-a}}^{(2)}(ih)}$)

1.2.3 Calculating $i^j \hat{f}^{(j)}_{X-a}(ih)$ for $0 \le j \le M$

For $j = 1$ we already know that $i \hat{f}^{(1)}_{X-a}(ih) = 0$ because that's how we got h. We also know that $i^2 \widehat{f_{X-a}}^{(2)}(ih) = c^2$ because that's how we got c. But for $j = 0, 3, \cdots, M$ use the formula in the appendix with the given value for a and the calculated values for h and c to calculate

$$\begin{aligned} i^j \widehat{f_{X-a}}^{(j)}(ih) &= \sum_{l=0}^{j} \frac{1}{(j-l)!} i^{j-l} \widehat{f_X}^{(j-l)}(0) \sum_{m=0}^{M-j} \frac{(j+m)!}{m!\,(m+l)!} (-1)^{m+l} a^{m+l} h^m \\ &+ \sum_{l=1}^{M-j} \frac{1}{(j+l)!} i^{j+l} \widehat{f_X}^{(j+l)}(0) \sum_{m=l}^{M-j} \frac{(j+m)!}{m!\,(m-l)!} (-1)^{m-l} a^{m-l} h^m \end{aligned}$$

(Using the alternative formula in the appendix this would be

$$\begin{aligned} i^j \widehat{f_{X-a}}^{(j)}(ih) &= e^{-ah} \left[\sum_{l=0}^{j} i^{j-l} \widehat{f_X}^{(j-l)}(0) \sum_{m=0}^{j-l} \frac{j!}{(j-m-l)!\,(m+l)!m!} (-1)^{m+l} a^{m+l} h^m \right. \\ &\left. + \sum_{l=1}^{M-j} i^{j+l} \widehat{f_X}^{(j+l)}(0) \sum_{m=l}^{j+l} \frac{j!}{(j-m+l)!\,(m-l)!m!} (-1)^{m-l} a^{m-l} h^m \right] \end{aligned}$$

)

1.3 Calculate the value of the formula in (20)

We now have all the inputs needed to calculate the value of the formula in (20). Put the given value for a and the calculated values for h, c, and $i^j \widehat{f_{X-a}}^{(j)}(ih)$ for $j = 0, 3, \cdots, M$ into the formula in (20)

$$\begin{aligned} &\frac{\widehat{f_{X-a}}(ih)}{\sqrt{2\pi}c} \{ -h\,[\,1 + \sum_{j=2}^{\frac{M}{2}} \frac{(-1)^j}{(2j)!} [\, \frac{i^{2j} \widehat{f_{X-a}}^{(2j)}(ih)}{c^{2j}\, \widehat{f_{X-a}}(ih)} - (2j)?\,] \sum_{n=0}^{\frac{M}{2}-j} \frac{(2n)?}{(2n)!} (2\,(n+j))?\,] \\ &-\frac{1}{c} [\, \sum_{j=2}^{\frac{M+1}{2}} \frac{(-1)^j}{(2j-1)!} \frac{i^{2j-1} \widehat{f_{X-a}}^{(2j-1)}(ih)}{c^{2j-1}\, \widehat{f_{X-a}}(ih)} \sum_{n=0}^{\frac{M+1}{2}-j} \frac{(2n)?}{(2n)!} (2\,(n+j))?\,]\, \} \end{aligned}$$

where the $\sqrt{2\pi}$ came from the appearance in (20) of $\varphi(0)$, the value of the normal density at 0.

1.4 Choose a new trial value for a

After calculating the value of the formula in (20) choose a new trial value for a somewhere in between the last value that gave a positive value of the fomula in (20) and the last value that gave negative. The easiest strategy is probably just to pick a new trial value halfway between those two, although you could do fancier things that might speed up the convergence a little.

1.5 Conclude on an estimate for the mode

When the trial values for a start agreeing closely, or when the value of the formula in (20) gets close enough to 0, then the final trial value for a is the algorithm's estimate for the mode of the original probability distribution $f_X(x)$. Try declaring "close enough" to be about 0.001 times the standard deviation of the distribution. (You can calculate the standard deviation using the first two moments: $\sqrt{i^2 \hat{f}_X^{(2)}(0) - \left(i\hat{f}_X^{(1)}(0)\right)^2} = \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}$).

2 The testing program

2.1 Testing with known mode and moments

The first distributions we will test will be ones where we already know the mode and all the moments so that we can see how well the algorithm works with known inputs rather than random samples.

(a) For each test start by using four moments ($M = 4$) and record how far the estimated mode is from the actual, measured in standard deviations. (You can calculate the standard deviation using the first two moments: $\sqrt{i^2 \hat{f}_X^{(2)}(0) - \left(i\hat{f}_X^{(1)}(0)\right)^2} = \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}$).

(b) Then test how much accuracy is lost if only three moments ($M = 3$) are used.

(c) Then increase the number of moments, recording how many moments are required to reach accuracy of 0.1, 0.01, and 0.001 standard deviations.

(d) For each accuracy milestone, test the difference in accuracy if the alternative formulas from the Appendix are used to estimate all the values of formula (23) throughout the calculation.

2.1.1 normal μ, σ

The mode is μ.

Moments are $i^k \hat{f}_X^{(k)}(0) = \mathbb{E}\left[X^k\right] = \sum_{n=0}^{\frac{k}{2}} \frac{k!(2n)?}{(k-2n)!(2n)!} \mu^{k-2n} \sigma^{2n}$

Make tests with $\mu = -1$, 0, and 1 and $\sigma = \frac{1}{2}$, 1, and 2

2.1.2 gamma $\alpha > 1$

The mode is α when $\alpha > 1$

Moments are $i^k \hat{f}_X^{(k)}(0) = \mathbb{E}\left[X^k\right] = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}$, or $= (\alpha + k - 1)(\alpha + k - 2) \cdots \alpha$ if α is an integer > 1

Make tests with $\alpha = 2$, 5, 10, 1.5, and 9.5

2.1.3 weibull (on positive x) $\tau > 1$

The mode is $\left(\frac{\tau-1}{\tau}\right)^{\frac{1}{\tau}}$ when $\tau > 1$.
Moments are $i^k \hat{f}_X^{(k)}(0) = \mathbb{E}\left[X^k\right] = \Gamma\left(\frac{\tau+k}{\tau}\right)$ when $\tau > 1$.
Make tests with $\tau = 1.5$, 5, and 10.

2.1.4 weibull (on negative x) $\alpha > 1$

The mode is $-\left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}}$
Moments are $i^k \hat{f}_X^{(k)}(0) = \mathbb{E}\left[X^k\right] = (-1)^k\,\Gamma\left(\frac{\alpha+k}{\alpha}\right)$
Make tests with $\alpha = 1.5$, 5, and 10

2.1.5 generalized gamma α, τ. with $\alpha\tau > 1$

The mode is $\left(\frac{\alpha\tau-1}{\tau}\right)^{\frac{1}{\tau}}$ when $\alpha\tau > 1$
Moments are $i^k \hat{f}_X^{(k)}(0) = \mathbb{E}\left[X^k\right] = \frac{\Gamma\left(\frac{\alpha\tau+k}{\tau}\right)}{\Gamma(\alpha)}$
Make tests with $\alpha = 5$ and 10 and $\tau = 0.5$, 1, and 2

2.1.6 inverse gaussian μ, θ

The mode is $\mu\left[\left(1+\frac{9\mu^2}{4\theta^2}\right)^{\frac{1}{2}} - \frac{3\mu}{2\theta}\right]$
Moments are $i^k \hat{f}_X^{(k)}(0) = \mathbb{E}\left[X^k\right] = \sum_{n=0}^{k-1} \frac{(k+n-1)!}{(k-n-1)!} \cdot \frac{\mu^{n+k}}{(2\theta)^n}$
Make tests with $\mu = -1$, 1, 2 and $\theta = 0.5$, 1 and 2

2.1.7 lognormal μ, σ

The mode is $e^{\mu-\sigma^2}$
Moments are $i^k \hat{f}_X^{(k)}(0) = \mathbb{E}\left[X^k\right] = e^{k\mu+\frac{1}{2}k^2\sigma^2}$
Make tests with $\mu = -1$, 1, 2 and $\theta = 0.5$, 1 and 2
This is the first test of a distribution that has no moment generating function so look for any departure from patterns observed in tests of the other distributions.