Estimating $\log(\lambda)$

Toby Kenney

May 3, 2018

Recall that if our objective function $f(\lambda)$ has a convergent Taylor series

$$e^{-\lambda} f(\lambda) = \sum \frac{a_n \lambda^n}{n!}$$

then we can obtain an unbiassed estimator by

$$g(x) = a_x$$

Unfortunately, $f(\lambda) = \log(\lambda)$ does not have a globally convergent Taylor series, so we cannot use this approach. For large λ, we have that $\log(x)$ is a good approximation for $\log(\lambda)$. Therefore, one possibility is to use a two-stage approach

$$g(x) = \begin{cases} a_x & \text{if } x < N \\ \log(x) & \text{if } x > N \end{cases}$$

where

$$e^{-\lambda} \log(\lambda) = \sum \frac{a_n \lambda^n}{n!}$$

is a Taylor series on the interval $[L, N]$. Unfortunately, since this Taylor series is expanded about 0, we cannot evaluate it here. Instead, we can perform a truncation of the Taylor series for $\log\left(1+\left(\frac{\lambda}{a}-1\right)\right)$. That is, we have

$$\begin{aligned}
\log(\lambda) &= \log(a) + \log\left(1+\left(\frac{\lambda}{a}-1\right)\right) \\
&= \log(a) + \sum \frac{(-1)^{n-1}}{n}\left(\frac{\lambda}{a}-1\right)^n \\
e^{-\lambda}\log(\lambda) &= \left(\sum \frac{(-\lambda)^n}{n!}\right)\left(\log(a) - \sum \frac{1}{n}\left(1-\frac{\lambda}{a}\right)^n\right) \\
&= \sum \frac{(-1)^n \log(a)\lambda^n}{n!} - \sum_{n,m} \frac{(-1)^n\lambda^n}{n!}\frac{1}{m}\left(1-\frac{\lambda}{a}\right)^m \\
&= \sum \frac{(-1)^n \log(a)\lambda^n}{n!} - \sum_{n,m} \frac{(-1)^n\lambda^n}{n!}\frac{1}{m}\sum_{k=0}^{m}(-1)^{n-k}\binom{m}{k}\left(\frac{\lambda}{a}\right)^k \\
&= \sum \frac{(-1)^n \log(a)\lambda^n}{n!} - \sum_{n,k} \frac{(-1)^{n+k}\lambda^{n+k}}{a^k n!}\sum_{m=k}^{\infty}\frac{1}{m}\binom{m}{k}
\end{aligned}$$

The inner sum does not converge, so we truncate it at a particular value N to get

$$\begin{aligned}
\log(\lambda) &\approx \log(a) + \sum_{n=0}^{N} \frac{(-1)^{n-1}}{n}\left(\frac{\lambda}{a}-1\right)^n \\
e^{-\lambda}\log(\lambda) &\approx \sum \frac{(-1)^n \log(a)\lambda^n}{n!} - \sum_{n=0}^{\infty}\sum_{m=1}^{N} \frac{(-1)^n\lambda^n}{n!}\frac{1}{m}\sum_{k=0}^{m}(-1)^{n-k}\binom{m}{k}\left(\frac{\lambda}{a}\right)^k \\
&= \sum \frac{(-1)^n \log(a)\lambda^n}{n!} - \sum_{n=0}^{\infty}\sum_{k=0}^{N} \frac{(-1)^k\lambda^{n+k}}{a^k n!}\sum_{m=k\vee 1}^{N}\frac{1}{m}\binom{m}{k} \\
&= \sum \frac{(-1)^n \log(a)\lambda^n}{n!} - \sum_{n=0}^{\infty}\sum_{k=0}^{N} \frac{(-1)^k\lambda^{n}}{a^k (n-k)!}\sum_{m=k\vee 1}^{N}\frac{1}{m}\binom{m}{k} \\
&= \sum \frac{(-1)^n\lambda^n}{n!}\left(\log(a) - \sum_{k=0}^{N}\frac{(-1)^k n!}{(n-k)!}a^{-k}\sum_{m=k\vee 1}^{N}\frac{1}{m}\binom{m}{k}\right)
\end{aligned}$$

For $k > 0$, we have

$$\begin{aligned}
\sum_{m=k}^{N}\frac{1}{m}\binom{m}{k} &= \frac{1}{k}\sum_{m=k}^{N}\binom{m-1}{k-1} \\
&= \frac{1}{k}\binom{N}{k}
\end{aligned}$$

This gives us

$$e^{-\lambda}\log(\lambda) \approx \sum \frac{(-1)^n\lambda^n}{n!}\left(\log(a) - \sum_{m=1}^{N}\frac{1}{m} - \sum_{k=1}^{N}\frac{(-1)^k n!}{(n-k)!k}a^{-k}\binom{N}{k}\right)$$

This gives us the estimator

$$g(x) = \log(a) - \sum_{m=1}^{N}\frac{1}{m} - \sum_{k=1}^{N}\frac{(-1)^k x!}{(x-k)!k}a^{-k}\binom{N}{k}$$

This can work well on an interval $[l, u]$, but not outside that interval. However, for larger values of x, $\log(x)$ can be approximately linear, so it will make a good estimator for $\log(\lambda)$. We therefore suggest the estimator

$$g(x) = \begin{cases} \log(a) - \sum_{m=1}^{N}\frac{1}{m} - \sum_{k=1}^{N}\frac{(-1)^k x!}{(x-k)!k}a^{-k}\binom{N}{k} & \text{if } x \leqslant c \\ \log(x) & \text{if } x > c \end{cases}$$

From some experimentation, it seems that $a = 3, N = 8$ gives a fairly reasonable approximation in typical cases. Ideally, we would adjust these slightly based on sample size.

We want to estimate the conditional variance of $g(x)$. For given λ, the variance of $g(x)$ is

$$\begin{aligned}\mathrm{Var}(g(x)|\lambda) &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n g(n)^2}{n!} - e^{-2\lambda}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\lambda^{n+m}g(n)g(m)}{n!m!} \\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n g(n)^2}{n!}\left(1 - e^{-\lambda}\sum_{m=0}^{n}\binom{n}{m}\frac{g(m)g(n-m)}{g(n)^2}\right)\end{aligned}$$

An easy unbiassed estimator of $e^{-\lambda}\frac{\lambda^n}{n!}$ is

$$z(x) = \begin{cases} 1 & \text{if } x = n \\ 0 & \text{otherwise} \end{cases}$$

However, this estimator has variance

$$e^{-\lambda}\frac{\lambda^n}{n!}\left(1 - e^{-\lambda}\frac{\lambda^n}{n!}\right)$$

which can be relatively high. An alternative estimator is the posterior mean estimator

$$z_n(x) = 2^{-(x+n+1)}\binom{x+k}{k}$$

This has mean

$$\mathbb{E}\left(z_n(X)\right) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n 2^{-(x+n+1)}}{n!} \binom{x+k}{k}$$

For the quantity $e^{-2\lambda}\frac{\lambda^n}{n!}$, the unbiassed estimator is

$$s_n(x) = (-1)^{x-n}\binom{x}{n}$$

but this is very high variance. We can attempt to reduce the variance of this estimator by taking an average of the values for adjacent values of x. In particular, we could set the estimator as

$$s_n(x) = \frac{1}{2}\binom{2\left\lfloor\frac{x}{2}\right\rfloor}{n-1}$$

This is approximately unbiassed, but is lower variance.

Note that since $g(n)$ grows slowly, we have that $\sum_{m=0}^{n}\binom{n}{m}g(m)g(n-m)$ can be approximated by 2^n, we let $a_n = \frac{\sum_{m=0}^{n}\binom{n}{m}g(m)g(n-m)}{2^n}$. Then we are trying to estimate

$$\sum_{n=0}^{\infty} e^{-2\lambda}\frac{2^n\lambda^n}{n!}a_n$$

We have that

$$\frac{d}{d\lambda}\left(e^{-2\lambda}\frac{(2\lambda)^n}{n!}\right) = 2e^{-2\lambda}\left(\frac{(2\lambda)^{n-1}}{(n-1)!} - \frac{(2\lambda)^n}{n!}\right)$$

This means that for the values of n for which this is largest, it is approximately a constant function of λ. We can therefore approximate the average value of this quantity by taking the average of the quantities

$$e^{-(\lambda_i+\lambda_j)}\frac{(\lambda_i+\lambda_j)^n}{n!}$$

These quantities can be estimated by

$$Z_n = \begin{cases} 1 & \text{if } X_i + X_j = n \\ 0 & \text{otherwise} \end{cases}$$

This estimator might be improved by restricting to cases where X_i and X_j are consistent with the $\lambda_i = \lambda_j$. We will take $(X_i - X_j)^2 < (X_i + X_j)$ as our test for this criterion.

Empirically, this estimator performs fairly well for larger λ, but less well for small λ. The strict criterion is not ideal. We replace it by a weighted criterion, where we give a certain weight to the hypothesis that X_i and X_j come from the same value of λ. The weight we use is the likelihood ratio:

$$L = \left(\frac{X_i + X_j}{2X_i}\right)^{X_i} \left(\frac{X_i + X_j}{2X_j}\right)^{X_j}$$

In the case where λ is the same for all X_i, this reweighted likelihood

---

which means that our estimator of $e^{-2\lambda}\frac{\lambda^n}{n!}$ needs to be most The posterior estimator

$$w_n(x) = 3^{-(x+n+1)}\binom{x+k}{k}$$

may be preferable. Indeed we have

$$\begin{aligned}
\mathbb{E}(s_n(x)^2) &= \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}\binom{x}{n}^2 \\
&= \frac{e^{-\lambda}}{(n!)^2}\sum_{x=0}^{\infty}\lambda^x \sum_{k=0}^{n}\binom{n}{k}^2 (n-k)!\frac{1}{(x-n-k)!} \\
&= \frac{e^{-\lambda}}{(n!)^2}\sum_{k=0}^{n}\binom{n}{k}^2 (n-k)!\sum_{x=0}^{\infty}\frac{\lambda^x}{(x-n-k)!} \\
&= \frac{e^{-\lambda}}{(n!)^2}\sum_{k=0}^{n}\binom{n}{k}^2 (n-k)!\lambda^{n+k}e^{\lambda} \\
&= \frac{\sum_{k=0}^{n}\binom{n}{k}^2 (n-k)!\lambda^{n+k}}{(n!)^2} \\
&= \frac{\sum_{k=0}^{n}\lambda^{n+k}}{(n-k)!(k!)^2}
\end{aligned}$$

If we use the posterior mean estimators, then our estimator for the conditional variance becomes:

$$\begin{aligned}
\mathrm{Var}\widehat{(g(x)}|\lambda) &= \sum_{n=0}^{\infty} 2^{-(x+n+1)}\binom{x+n}{n}g(n)^2 - \sum_{n=0}^{\infty}\sum_{m=0}^{n} 3^{-(x+n+1)}\binom{x+n}{n}\binom{n}{m}g(m)g(n-m) \\
&= \mathbb{E}_{N\sim NB(x+1,1)}g(N)^2 - \mathbb{E}_{M,N\sim NB(x+1,1,1)}g(M)g(N) \\
&= \frac{1}{2}\mathbb{E}_{M,N\sim NB(x+1,1,1)}\left(g(M) - g(N)\right)^2 \\
&= \frac{1}{2}\mathrm{Var}_{M,N\sim NB(x+1,1,1)}\left(g(M) - g(N)\right)
\end{aligned}$$

For the multivariate distribution of M and N, we can see that the marginal distribution is negative binomial with $r = x+1$ and $\beta = 1$, while the conditional distribution of N given $M = m$ is negative binomial with $r = x + m + 1$ and $\beta = \frac{1}{2}$. Now if we let $S = M + N$, then conditional on S, M and N have a binomial distribution with $p = \frac{1}{2}$, so by symmetry $\mathbb{E}(g(M) - g(N)|S) = 0$. Thus

$$\begin{aligned}
\mathrm{Var}\widehat{(g(x)}|\lambda) &= \frac{1}{2}\,\mathrm{Var}_{M,N}\left(g(M) - g(N)\right) \\
&= \frac{1}{2}\mathbb{E}_S\left(\mathrm{Var}_{M,N}\left(g(M) - g(N)|S\right)\right)
\end{aligned}$$

For large S, we have $g(M) = \log(M)$ with high probability, so this variance becomes

$$\begin{aligned}
&= \frac{1}{2}\mathbb{E}_S\left(\mathrm{Var}_N\left(\log(S - N) - \log(N)|S\right)\right) \\
&= \frac{1}{2}\mathbb{E}_S\left(\mathrm{Var}_N\left(\log\left(\frac{S}{N} - 1\right)\middle|S\right)\right) \\
&= \mathrm{Var}\left(g(M)\right) - \mathrm{Cov}\left(g(M), g(N)\right) \\
&= \mathbb{E}\left(g(M)^2\right) - \mathbb{E}\left(g(M)\mathbb{E}\left(g(N)|M\right)\right) \\
&= \frac{1}{2}\left(\mathrm{Var}_M\left(\mathbb{E}_{N|M}\left(g(M) - g(N)\right)\right) + \mathbb{E}_M\left(\mathrm{Var}_{N|M}\left(g(M) - g(N)\right)\right)\right) \\
&= \frac{1}{2}\left(\mathrm{Var}_M\left(g(M) - \mathbb{E}_{N|M}\left(g(N)\right)\right) + \mathbb{E}_M\left(\mathrm{Var}_{N|M}\left(g(N)\right)\right)\right)
\end{aligned}$$

$\binom{x+n+m}{n+m}\binom{n+m}{m} = \binom{x+n+m}{n}\binom{x+m}{m}$