First-Year Exam (June 2014): Solutions to 205B Question

(1) You can see, both from rough sketches of the likelihood function $\ell(\theta|y)$ for several values of y (which will look like the curves in Figure 1 below) and from the repeated-sampling mean, that all three of (a)–(c) are desirable qualitative behaviors for a good estimator in this problem.

(2) Solving $E_{RS}(Y|\theta) = \frac{\theta}{\theta+1} = y$ for θ yields $\hat{\theta}_{MoM} = \frac{y}{1-y}$. This estimator behaves sensibly at $y = 0$ and is monotonically increasing in y — thus properties (1)(a) and (1)(b) are satisfied — but $\hat{\theta}_{MoM}$ goes to $+\infty$ as $y \uparrow 1$, and in fact $\hat{\theta}_{MoM} \geq 1$ (thereby violating the basic range restriction for θ) for all $y \geq \frac{1}{2}$.

(3) The log-likelihood function is $\ell\ell(\theta|y) = \log\theta + (\theta - 1)\log y$, from which $\frac{\partial}{\partial\theta}\ell\ell(\theta|y) = \frac{1}{\theta} + \log y$; this first partial derivative has a unique zero at $-\frac{1}{\log y}$, which is a monotonically increasing function of y. This is the global maximum of $\ell\ell(\theta|y)$ in $(0, 1)$ as long as $-\frac{1}{\log y} < 1$, which is true only for $0 < y < \frac{1}{e} \doteq 0.37$; for any $y \geq \frac{1}{e}$, the maximum occurs at the boundary $\theta = 1$, so in that case $\hat{\theta}_{MLE} = 1$. Thus $\hat{\theta}_{MLE} = \min\left(-\frac{1}{\log y}, 1\right)$. In this setting $\hat{\theta}_{MLE}$ satisfies all three of the desirable qualitative behaviors in (1).

(4) Evidently $\frac{\partial^2}{\partial\theta^2}\ell\ell(\theta|y) = -\frac{1}{\theta^2}$, from which the observed information is

$$\hat{I}\left(\hat{\theta}_{MLE}\right) = \left[-\frac{\partial^2}{\partial\theta^2}\ell\ell(\theta|y)\right]_{\theta=\hat{\theta}_{MLE}} = (\log y)^2 \tag{1}$$

for $0 < y < \frac{1}{e}$; for $y \geq \frac{1}{e}$, $\hat{I}\left(\hat{\theta}_{MLE}\right) = 1$. For $0 < y < \frac{1}{e}$, the usual standard error (SE) associated with the MLE is then $\widehat{SE}\left(\hat{\theta}_{MLE}\right) = \sqrt{\hat{I}^{-1}\left(\hat{\theta}_{MLE}\right)} = \frac{1}{|\log y|}$, with associated approximate 95% confidence interval $-\frac{1}{\log y} \pm \frac{1.96}{|\log y|} = \frac{1 \pm 1.96}{|\log y|}$; for $y \geq \frac{1}{e}$ the corresponding interval is 1 ± 1.96. This method assumes that the sample size n is big enough (a) for observed information to provide an accurate SE and (b) for the repeated-sampling distribution of the MLE (for the actual n in the problem under study) to be close to Gaussian; here, with $n = 1$, neither of these assumptions is anywhere near correct. Moreover, you can readily see that (i) the left endpoint of the interval $\frac{1 \pm 1.96}{|\log y|}$ is guaranteed to be negative, (ii) the right endpoint of that interval will only be less than 1 iff $\log y < -2.96$, i.e., iff y is less than about 0.05, and (iii) the interval 1 ± 1.96 runs from -0.96 to 2.96. Fisher did not have $n = 1$ in mind when he proposed this method.

(5) The fact that $V = -\theta\log Y$ has the Exponential(1) distribution can be shown directly using the change-of-variables formula, or (if You forget it) You can derive it as follows: the CDF $F_V(v)$ of V is $F_V(v) = P(V \leq v) = P(-\theta\log Y \leq v) = P(\log Y \geq -\frac{v}{\theta}) = 1 - P(Y \leq e^{-\frac{v}{\theta}}) = 1 - F_Y\left(e^{-\frac{v}{\theta}}\right)$, where $F_Y(\cdot)$ is the CDF of Y. Thus the density of V is (chain rule) $p(v) = -p_Y\left(e^{-\frac{v}{\theta}}\right)e^{-\frac{v}{\theta}}\left(-\frac{1}{\theta}\right)$, where $p_Y(\cdot)$ is the density of Y; this reduces after simplification to $p(v) = e^{-v}$, as desired. All of this means that in repeated sampling $1 - \alpha = P_{RS}\left[F_{\mathcal{E}}^{-1}\left(\frac{\alpha}{2}\right) < -\theta\log Y < F_{\mathcal{E}}^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$, where $F_{\mathcal{E}}^{-1}(\cdot)$ is the inverse CDF of the standard exponential distribution; after some rearrangement this becomes $P_{RS}\left[-\frac{F_{\mathcal{E}}^{-1}\left(\frac{\alpha}{2}\right)}{\log Y} < \theta < -\frac{F_{\mathcal{E}}^{-1}\left(1-\frac{\alpha}{2}\right)}{\log Y}\right] = 1 - \alpha$, so that $\left[-\frac{F_{\mathcal{E}}^{-1}\left(\frac{\alpha}{2}\right)}{\log Y}, -\frac{F_{\mathcal{E}}^{-1}\left(1-\frac{\alpha}{2}\right)}{\log Y}\right]$ is a frequentist-valid $100(1 - \alpha)$% confidence interval for θ in this model. The left and right endpoints of this interval cannot go negative for any $Y \in (0, 1)$, which is the good news; but it's easy to show (based on the CDF of the Exponential(1) distribution, which is $F_{\mathcal{E}}^{-1}(p) = -\log(1 - p)$) that the right endpoint will be bigger than 1 whenever $Y > \frac{\alpha}{2}$ and the left endpoint will also be bigger than 1 for all $Y > 1 - \frac{\alpha}{2}$. Valid and sensible frequentist inference is difficult when n is small.

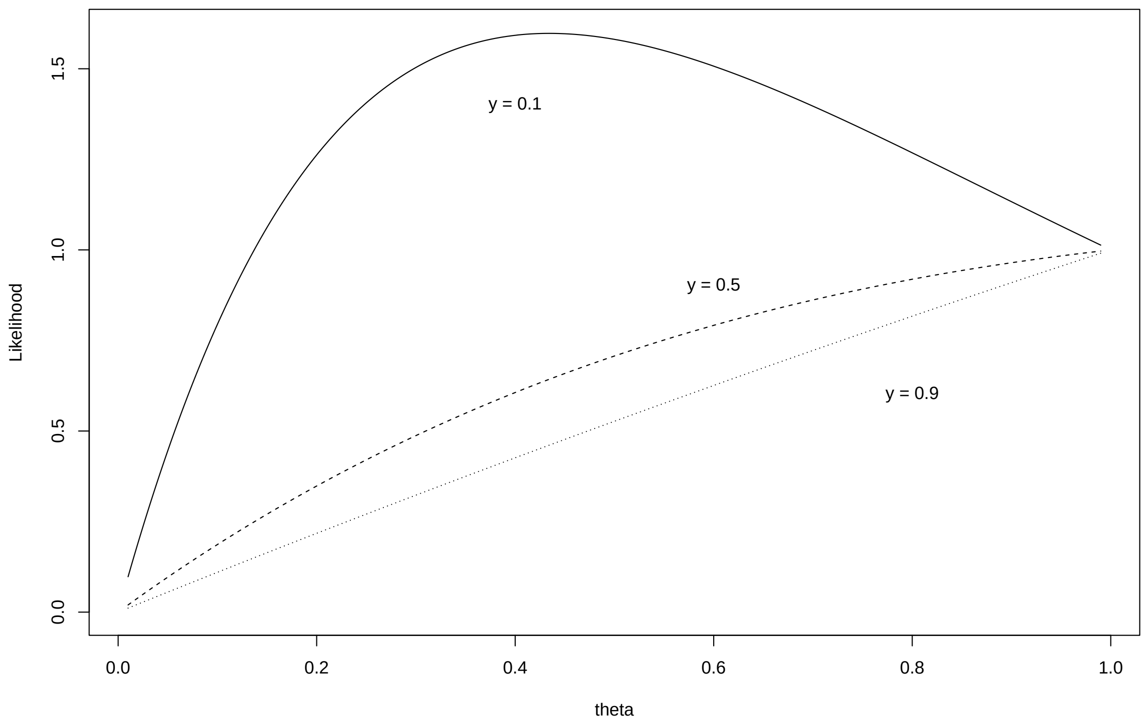

Figure 1: *Un-normalized likelihood function, for* $y = (0.1, 0.5, 0.9)$.