$$p_Y(y;\theta)=\frac{-\theta^y}{y\log(1-\theta)}$$

$$\sum_{i=1}^{\infty} p_Y(y;\theta)=\sum_{i=1}^{\infty}\frac{-\theta^y}{y\log(1-\theta)}=\frac{1}{\log(1-\theta)}\sum_{i=1}^{\infty}\frac{-\theta^y}{y}=\frac{\log(1-\theta)}{\log(1-\theta)}=1$$

$$E[Y]=\sum_{1}^{\infty}\frac{-y\theta^y}{y\log(1-\theta)}=\sum_{1}^{\infty}\frac{\theta^y}{\log(1-\theta)}=\frac{-1}{\log(1-\theta)}\sum_{1}^{\infty}\theta^y=\frac{-\theta}{(1-\theta)\log(1-\theta)}$$

$$V(Y)=E[Y^2]-E[Y]^2=\sum_{i=1}^{\infty}\frac{-y\theta^y}{\log(1-\theta)}-E[Y]^2=\frac{-\theta}{\log(1-\theta)}\sum_{y=1}^{\infty}y\theta^{y-1}-E[Y]^2=$$
$$\frac{-\theta}{\log(1-\theta)}\frac{d}{d\theta}\sum_{y=1}^{\infty}\theta^y-E[Y]^2=\frac{-\theta}{\log(1-\theta)}\frac{d}{d\theta}\frac{\theta}{1-\theta}-E[Y]^2=\frac{-\theta}{\log(1-\theta)}\frac{((1-\theta)+\theta)}{(1-\theta)^2}-E[Y]^2=$$
$$\frac{-\theta}{\log(1-\theta)(1-\theta)^2}-\frac{\theta^2}{\log(1-\theta)^2(1-\theta)^2}=\frac{-\theta\log(1-\theta)-\theta^2}{\log(1-\theta)^2(1-\theta)^2}$$

Distributions in the exponential family have the form: $p_y(y|\theta)=g(y)h(\theta)e^{\eta(\theta)\tau(\theta)}$

Let $g(y)=\frac{-1}{\log(1-\theta)}, h(\theta)=\frac{1}{y}, \eta(\theta)=\log(\theta), \tau(\theta)=y$

It is clear the log-series is a member of the exponential family.

Define $T(y)=\sum_{n=1}^{N}\tau(y_n) \;=\; \sum_{n=1}^{N}y_n$ This is the sufficient statistic for our distribution.

We may also define: $\bar{Y}_n=\frac{1}{N}\sum_{n=1}^{N}y_n$ this is just the sufficient statistic divided by N.

$$L(y_1,y_2,y_3,\dots y_n|\theta)=\prod_{i=1}^{N}p_Y(y_i|\theta)=\prod_{i=1}^{N}\left(\frac{-\theta^{y_i}}{y_i\log(1-\theta)}\right)$$

$$\log(L(y_1,y_2,y_3,\dots y_n|\theta))=\sum_{i=1}^{N}\log(p_Y(y_i|\theta))=\sum_{i=1}^{N}(y_i\log(\theta)-\log(y_i)-\log(-\log(1-\theta)))$$

Maximum occurs where the derivative is zero : $\sum_{i=1}^{N}\left(\frac{y_i}{\theta}-\frac{1}{(\theta-1)\log(1-\theta)}\right)=0$

$\sum_{i=1}^{N}\frac{y_i}{\theta}=\sum_{i=1}^{N}\frac{1}{(\theta-1)\log(1-\theta)}$ thus: $\sum_{i=1}^{N}y_i=\sum_{i=1}^{N}\frac{\theta}{(\theta-1)\log(1-\theta)}$ and $\bar{Y}_N=\frac{\theta}{(\theta-1)\log(1-\theta)}$

This can be stated as $\frac{\theta}{(\theta-1)\log(1-\theta)}-\bar{Y}_N=0$

The condition that defines $\hat{\theta}$: $\frac{\theta}{(\theta-1)\log(1-\theta)}-\bar{Y}_N$ has a derivative of: $-\frac{\theta+\log(1-\theta)}{(\theta-1)^2\log(1-\theta)^2}$

Since $\theta+\log(1-\theta)<0\,\forall x\in(0,1)$, and $(\theta-1)^2\geqslant 0$ as well as $\log(1-\theta)^2\geqslant 0$

It is clear that the derivative is strictly positive, and thus the condition is monotonic increasing.

Applying Newton's method for the zeros of a derivative: $\theta_{k+1}=\theta_k-\frac{f'(\theta_k)}{f''(\theta_k)}$

$f'(\theta)=\frac{\bar{Y}_N}{\theta} - \frac{1}{(\theta-1)\log(1-\theta)}$ $\quad f''(\theta)=\frac{-\bar{Y}_N}{\theta^2}+\frac{1+\log(1-\theta)}{(\theta-1)^2\log(1-\theta)^2}$

so our sequence is: $\theta_{k+1}=\theta_k-\dfrac{\frac{\bar{Y}_N}{\theta_k} - \frac{1}{(\theta_k-1)\log(1-\theta_k)}}{\frac{-\bar{Y}_N}{\theta_k^2}+\frac{1+\log(1-\theta_k)}{(\theta_k-1)^2\log(1-\theta_k)^2}}$

We may note that since $f''(\theta)\neq 0 \quad \forall\theta\in(0,1)$, This recursion will always yield a step
Provided we have sufficiently good initialization, it will converge to the maximum liklihood estimator.

Since $f''(\theta)\neq 0 \quad \forall\,\theta\in(0,1)$, We may apply the Implicit Function Theorem to $f'(\hat{\theta})=0$
so $\exists l(.) s/t \ \hat{\theta}=l(\bar{Y}_N)$ which is continious and differientiable in a neighborhood around $\bar{Y}_N$
Applying slutsky's theorem to this continous function: $plim\,\hat{\theta}=plim\,l(\bar{Y}_N)=l(plim\,\bar{Y}_N)=l(\mu^0)=\theta$

A consistent estimator of variance of $\hat{\theta}$ using Fisher's Information is the negative inverse Hessian.

$$V(\hat{\theta})\sim\ -[f''(\hat{\theta})]^{-1}=\frac{-1}{\frac{-\bar{Y}_N}{\theta^2}+\frac{1+\log(1-\theta)}{(\theta-1)^2\log(1-\theta)^2}}$$

Since $\hat{\theta}$ is obtained by solving for the zero of the derivative of the log-liklihood function,
Take a taylor expansion at θ^0. $0=f'(\theta^0)+f''(\theta^0)(\hat{\theta}-\theta^0)+R_{2.}$ disregarding the remainder term.

$$\hat{\theta}-\theta^0=\frac{-f'(\theta^0)}{f''(\theta^0)} \text{ and } \sqrt{(n)}(\hat{\theta}-\theta^0)=\frac{-\sqrt{(n)}\,f'(\theta^0)}{f''(\theta^0)}=\frac{-\sqrt{(n)}}{f''(\theta^0)}\sum_{i=1}^{n} y_i\log(\theta)-\log(y_i)-\log(-\log(1-\theta))$$

Since the sum times $\sqrt{(n)}$ converges in distribution to $N(0,-f''(\theta^0))$, $\sqrt{(n)}(\hat{\theta}-\theta^0)\to N(0,\frac{-1}{f''(\theta^0)})$

By assuming that the single measurement in the 9+ bucket has value 9. $\bar{Y}_n=1.479 \ \ \hat{\theta}=.5217389$
We can compute the negative inverse hessian at this point as well:

$$\frac{-1}{\frac{-1.479}{.5217389^2}+\frac{1+\log(1-.5217389)}{(.5217389-1)^2\log(1-.5217389)^2}}=.3007821$$

Using the Wald test, it is known that: $N\frac{(\hat{\theta}-\theta^0)^2}{V(\hat{\theta})}\sim\chi^2(1)$ as $\frac{\sqrt{(n)}(\hat{\theta}-\theta^0)}{-f''(\theta^0)}\to N(0,1)$.
Considering the null hypothesis: $H_0:\theta=.50$ against $H_a\theta\neq.50$
After assuming the null hypothesis we obtain a test statistic of: $1000\frac{(.0217389)^2}{0.300782}=1.571174$
By testing this statistic we obtain a p-value of 0.2100366, and fail to reject the null hypothesis.

Using the Liklihood-Ratio test, as the Wald test is not-invariant to non-linear transformations
We may Consider the null hypothesis: $H_0:\log(\theta)=-.70$ which is equivalent to: $H_0:\theta=e^{-.70}$
Our test statistic is: $2(f(\hat{\theta})-f(\theta_0))\sim\chi^2(1)$ and TS =2.110686.
This leads to a p-value of: .146282 and we fail to reject our null hypothesis at a 90% confidence level.
By testing with a Wald Statistic, a p-value of 0.1469593 is obtained, and we still fail to reject the Null.

Applying Pearson's Chi-Squared Test: $\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(n-1)$

Where O_i is the number of observations in bin i, and E_i is the expected number in each bin.
After calculating the test statistic we arrive at TS = 2.317605 with p-value: .9853933
So we fail to reject the Null Hypothesis that the data was not taken from a Log-Series distribution and conclude that the data is consistent with that distribution.

Code:

```
> LogData <- c( 710, 175, 74, 23, 10, 4, 2, 1, 1)
> barY <- 0
> numMeasures <- 0
> for( i in 1:9){
+   barY <- barY + LogData[i]*i
+   numMeasures <- numMeasures + LogData[i]
+ }
> barY <- barY / numMeasures
>
> #Populate the data from the frequency table. The C code is showing
> datapoints <- numeric( numMeasures)
> datapoints[1:710] <- 1
> datapoints[(710+1):(710+175)] <- 2
> datapoints[(710+175+1):(710+175+74)] <- 3
> datapoints[(710+175+74+1):(710+175+74+23)] <- 4
> datapoints[((710+175+74+23+1)):(710+175+74+23+10)] <- 5
> datapoints[(710+175+74+23+10+1):(710+175+74+23+10+4)] <- 6
> datapoints[(710+175+74+23+10+4+1):(710+175+74+23+10+4+2)] <- 7
> datapoints[(710+175+74+23+10+4+2+1):(710+175+74+23+10+4+2+1)] <- 8
> datapoints[(710+175+74+23+10+4+2+1+1):(710+175+74+23+10+4+2+1+1)] <- 9
>
> #Now Solve barY = theta / log( 1- theta )
> #Should make the epsilon a variable or something
> hatTheta <- .5
> fprime <- 1
> while( abs( fprime ) > .000002) {
+   fprime <- (barY / hatTheta) - (1 / ((hatTheta-1)*log( 1- hatTheta)))
+   fdubprime <- ( -barY / hatTheta^2 )+ 1 / ((1-hatTheta)^2 * log( 1 - hatTheta )
^2 ) + 1 / ((1-hatTheta)^2 * log( 1- hatTheta ) )
+   hatTheta <- hatTheta - fprime / fdubprime
+
+ }
> barY
[1] 1.479
> hatTheta
[1] 0.5217389
>
> vTheta <- -1 / fdubprime
> vTheta
[1] 0.3007821
>
> #Test the Wald Statistic:
>
> nullTheta <- .5
> 1000*(hatTheta - nullTheta)^2 / vTheta
[1] 1.571174
> 1- pchisq( 1000*(hatTheta - nullTheta)^2 / vTheta, 1 )
[1] 0.2100366
>
>
> #Since we're looking at a non-linear function we should use the liklihood ratio
test instead of the Wald Test
> nullTheta <- exp( -.7 )
>
> liklihoodThetaHat <- 0
> liklihoodNULL <- 0
```

```
> for( i in 1:1000 ){
+   liklihoodThetaHat <- liklihoodThetaHat + datapoints[i]*log( hatTheta) -
log( datapoints[i]) - log( - log( 1 - hatTheta ) )
+   liklihoodNULL <- liklihoodNULL + datapoints[i]*log( nullTheta) -
log( datapoints[i]) - log( - log( 1 - nullTheta ) )
+ }
>
> #Compare the Wald Statistic against the Liklihood Ratio Statistic
> 1- pchisq( 1000*(hatTheta - nullTheta)^2 / vTheta, 1 )
[1] 0.1469593
> 1- pchisq(2*( liklihoodThetaHat - liklihoodNULL ),1)
[1] 0.1462737
>
> Probs <- numeric( 9 )
> Probs[9] <- 1
> for( i in 1:8) {
+   Probs[i] <- -hatTheta^i / ( i*log( 1- hatTheta))
+   Probs[9] <- Probs[9] - Probs[i]
+ }
> Expected <- Probs*1000
>
> Expected
[1] 707.348120 184.525526  64.182767  25.114986  10.482773   4.557726   2.038237
0.930499   0.819366
>
> ChiSum <- 0
> for( i in 1:9 ){
+   ChiSum <- ChiSum + ((LogData[i]-Expected[i])^2 / Expected[i])
+ }
>
> ChiSum
[1] 2.317605
> 1- pchisq(ChiSum,8)
[1] 0.9696944
```