What defines the lognormal distribution?

F.S. and J.H.

January 16, 2016

Abstract: The lognormal distribution appears to be the probability law of many interesting natural phenomena. However, due to its heavy-tailed density, it cannot be concluded from information about its moments, alone, whether it is the cause for a given set of observations.

1 Introduction

A lognormal random variable X is one for which the (natural and therefore any) logarithm of X is normally distributed, i.e. for which $\mathbb{P}[(\log(X) - \mu)/\sigma \leq t] \quad = \quad \Phi(t)$, where $t \in \mathbb{R}$ and Φ is the standard normal distribution function.
Taking the derivative of $\Phi(t)$ shows us that the density function of this lognormal variable has to be

$$\frac{d}{dt}\mathbb{P}[X \leq \exp(t\sigma + \mu)] \quad = \quad \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(\log t \ - \ \mu)^2}{\sigma^2}\right)\frac{1}{t} \tag{1}$$

Which properties of the normal distribution are inherited?

The decay of the density for large t is much weaker than that of a Gaussian curve, which means that *fewer* expected values exist. In fact, for every $a > 0$ the integral

$$\mathbb{E}[\exp(aX)] \quad = \quad \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(a \cdot t - \frac{1}{2}\frac{(\log t \ - \ \mu)^2}{\sigma^2}\right) dt$$

diverges since the linear growth of $a \cdot t$ weighs heavier than the (square of) logarithmic one, in the exponent. However, even though this moment generating function doesn't exist, all moments (still) do

$$\mathbb{E}[X^n] \quad = \quad \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma t} t^n \exp\left(-\frac{1}{2}\frac{(\log t \ - \ \mu)^2}{\sigma^2}\right) dt \quad = \quad e^{n\mu + \frac{1}{2}n^2\sigma^2}.$$

In order to see how much less concentrated a lognormal variable is consider a normally distributed variable with mean μ and variance σ^2. Then, in comparison, the mean of the lognormal variable is much larger than just e^μ, and its variance is also exponentially growing in σ^2:

	Mean	Var
$\log X$	μ	σ^2
X	$e^{\mu + \frac{1}{2}\sigma^2}$	$(e^{\sigma^2} - 1)\mathbb{E}[X]^2$

This shows that a lognormal variable is much more 'elusive' than a normal one. But what consequences does this have? What are the 'drawbacks'?

2 Determining the Distribution from the Moments

The so called **Stieltjes Moment Problem** asks whether the moments of a positive random variable determines its distribution and whether such a distribution is *unique* (`https://en.wikipedia.org/wiki/Hamburger_moment_problem`).

In other words, estimating the moments (maybe through a clever way of sampling, and using large datasets), can we conclude that the variable has (1) as its distribution?

The answer for the lognormal distribution is **No**.

The Moment Problem is **uniquely solvable** if the Carleman condition (which is justly termed: `https://en.wikipedia.org/wiki/Torsten_Carleman`) is fulfilled:

$$\sum_{n=0}^{\infty} \mathbb{E}[X^n]^{-\frac{1}{2n}} = +\infty$$

In the case of a lognormal variable $\sum_n^\infty \exp\left(n\mu + \frac{1}{2}n^2\sigma^2\right)^{-\frac{1}{2n}} = \sum_n^\infty \exp\left(-\frac{1}{2}\mu - \frac{1}{4}n\sigma^2\right) = e^{-\frac{1}{2}\mu}\frac{1}{1-\exp(-\frac{\sigma^2}{4})} < \infty$. Carleman's condition is only sufficient. Therefore, we 'only' *don't know* whether the LogNormal distribution is uniquely determined by its moments.

However, it has been proven in a paper from C. Kleiber (`https://www.researchgate.net/publication/234060305_The_Generalized_Lognormal_Distribution_and_the_Stieltjes_Moment_Problem`) that it is indeterminate (together with a whole class of other similar so called generalized lognormal distributions): `https://www.researchgate.net/publication/234060305_The_Generalized_Lognormal_Distribution_and_the_Stieltjes_Moment_Problem`.
The paper also identifies sets of distributions with the same moments (called Stieltjes classes, there). They are constructed by looking for functions (=:perturbations) $p(t)$, such that $f_\epsilon(t) = f(t) + \epsilon p(t)$ is a probability density and $\int_{\mathbb{R}_+} t^n p(t)dt = 0$, for all $n > 0$.

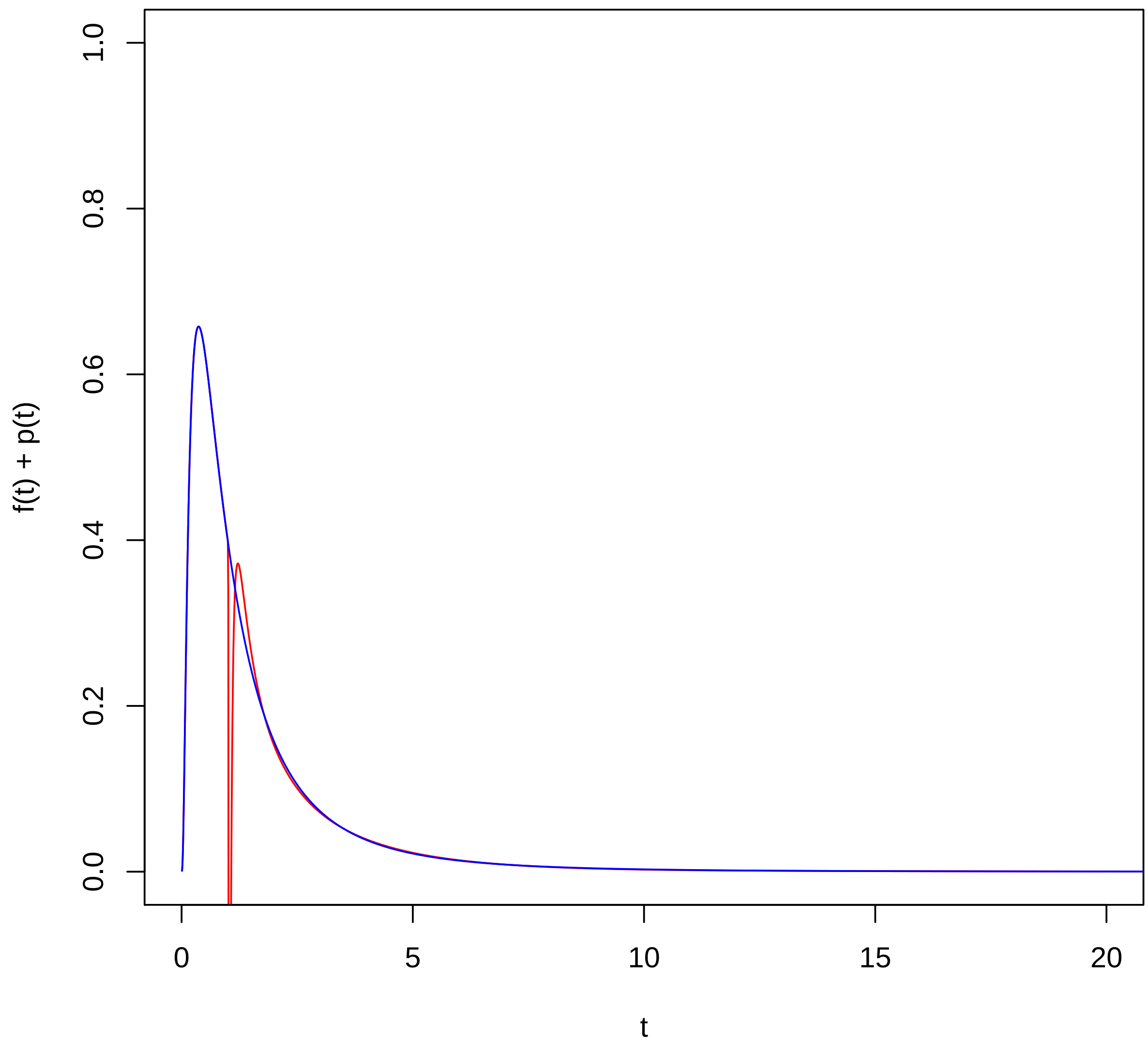
Using the construction of Theorem 4 of that paper, we check whether we get the first few moments **equal** from two different distributions, one of which is the lognormal distribution. We assume $\mu = 0$ and $\sigma = 1$:

```
t <- (1:100000)/100 # Gives us a uniform set of points `t' in [0,1000000].
f <- function(t){(1/(sqrt(2*pi)*t))*exp(-0.5*log(t)^2)}
g <- function(t){ifelse(t>1,sin(10*(t-1)^0.25)*exp(-10*(t-1)^0.25)/f(t),0.0)}
H <- max(abs(g(t)),na.rm=TRUE)     # If na.rm=FALSE, then H becomes `NA' (happens at 0-->log0)
p <- function(t){ g(t)/H }    # Normalization
f2 <- function(t){ f(t)+p(t) }
plot(t,f(t)+p(t), type="line", col="red",ylim=c(0,1), xlim=c(0,20))

## Warning in plot.xy(xy, type, ...):  plot type 'line' will be truncated to first character

lines(t,f(t), col="blue")
integral<-function(F,G){ sum(F*G,na.rm=TRUE)/100000 }# Riemann-Integral Approx. over [0,10]
qf<-rep(0,6);
qp<-rep(0,6);
for(i in 0:5) {qf[i+1] <- integral(f(t),t^i)}
for(i in 0:5) {qp[i+1] <- integral(f2(t),t^i)}
print(cbind(1:6-1,qf,qp));

##                  qf            qp
## [1,] 0 1.000000e-03  9.574996e-04
## [2,] 1 1.648721e-03  1.606100e-03
## [3,] 2 7.389053e-03  7.308890e-03
## [4,] 3 9.001294e-02  5.028129e-02
## [5,] 4 2.975532e+00 -3.862111e+01
## [6,] 5 2.607672e+02 -4.287795e+04
```



The differences we see are in the ith moment are growing in i and are due to numerical errors. The integration becomes more and more difficult for increasing moments, as the underlying relation $\int\limits_0^\infty x^k \sin(x^{\frac{1}{4}}) exp(-x^{\frac{1}{4}}) = 0$ involves larger and larger subsets of $\mathbb{R}_+$ to verify, numerically.

3 Conclusion

When we find evidence in data that suggests that the lognormal distribution may be the underlying probability law, we should check if this information is ’only’ equivalent to the moments of the distribution. In this case, we have to take into account the possibility of a whole classes of other possible distributions to be responsible for the given measurements.

Open Questions: It would be an interesting question to check to what degree such other examples also fulfill the nice geometric moment properties of the lognormal distribution (see `https://en.wikipedia.org/wiki/Log-normal_distribution`).