Reverse quantiles for log-Normal

David L Miller

September 7, 2020

Sometimes we have, say, posterior samples from a distribution. We want to know what the standard deviation of that distribution (or the CV or somesuch) but only have the samples and their mean. How can we get to the standard deviation from there? Based on calculations in `https://www.johndcook.com/quantiles_parameters.pdf`, but fully-derived here for the log-Normal case.

First, define the cumulative distribution function for a random variable X as:

$$F_X(x) = \mathbb{P}[X \leq x].$$

So a quantile, we want to find x for which

$$F_X(x) = \mathbb{P}[X \leq x] = q,$$

where $q$is the appropriate quantile, for example $q = 0.975$ or 0.025. We might typically have these two quantiles to construct a 95% interval. For generality, let's say we have two quantiles such that $q_1 < q_2$ and $x_1 < x_2$.

For a normally distributed X, we can standardize X, such that $Z = \frac{X-\mu}{\sigma}$ and then $F_Z(z) = \Phi(z)$ (i.e., the Normal$(0, 1)$ cumulative distribution function, so

$$\mathbb{P}\left[Z \leq \frac{x-\mu}{\sigma}\right] = \Phi\left(\frac{x-\mu}{\sigma}\right) = q.$$

Now for a log-Normally distributed X, we have that:

$$X \sim \text{logNormal}(\mu, \sigma^2) \quad \Rightarrow \quad Y = \log X \sim \text{Normal}(\mu, \sigma^2),$$

so we need to calulate

$$\mathbb{P}[Y \leq \log x] = \Phi\left(\frac{\log x - \mu}{\sigma}\right) = q, \tag{1}$$

and if we need to find μ and σ then we can use two quantiles we know say $q_1 < q_2$ and $x_1 < x_2$.

So, first we re-write (1) as:

$$\begin{aligned}
\mu &= \log x_i - \Phi^{-1}(q_i)\,\sigma \quad \text{and}\\
\sigma &= \frac{\log x_i - \mu}{\Phi^{-1}(q_i)}
\end{aligned}$$

where i indexes the quantiles we know (i.e., $i = 1, 2$). We can then solve the above by equating the $i = 1$ and $i = 2$ versions of the above (since although the quantiles change, the mean and standard deviation do not).

To obtain μ:

$$\begin{aligned}
\frac{\log x_1 - \mu}{\Phi^{-1}(q_1)} &= \frac{\log x_2 - \mu}{\Phi^{-1}(q_2)}\\
\Phi^{-1}(q_2)(\log x_1 - \mu) &= \Phi^{-1}(q_1)(\log x_2 - \mu)\\
\Phi^{-1}(q_2)\log x_1 - \Phi^{-1}(q_2)\,\mu &= \Phi^{-1}(q_1)\log x_2 - \Phi^{-1}(q_1)\,\mu\\
\Phi^{-1}(q_2)\log x_1 - \Phi^{-1}(q_1)\log x_2 &= \mu\left[\Phi^{-1}(q_2) - \Phi^{-1}(q_1)\right]\\
\mu &= \frac{\Phi^{-1}(q_2)\log x_1 - \Phi^{-1}(q_1)\log x_2}{\Phi^{-1}(q_2) - \Phi^{-1}(q_1)}.
\end{aligned}$$

and for σ:

$$\begin{aligned}
\log x_1 - \Phi^{-1}(q_1)\,\sigma &= \log x_2 - \Phi^{-1}(q_2)\,\sigma \\
\log x_1 - \log x_2 &= \left[\Phi^{-1}(q_1) - \Phi^{-1}(q_2)\right]\sigma \\
\sigma &= \frac{\log x_1 - \log x_2}{\Phi^{-1}(q_1) - \Phi^{-1}(q_2)}.
\end{aligned}$$

Appendix: R implementation

Below is a quick R implementation of the above equations.

```r
# go from quantiles to mean/sd for log-Normal distribution
# based on working at https://www.johndcook.com/quantiles_parameters.pdf

# takes as input all of your samples, computes the quantiles
# and then returns the mean/sd of the variable, assuming it's
# log-normally distributed
ln_get_mean_var <- function(samples, q1=0.025, q2=0.975){

  # first get where the quantiles lie, x1, x2
  x1 <- quantile(samples, p=q1)
  x2 <- quantile(samples, p=q2)

  # now the inverse CDF ("quantile function", Phi^-1) for
  # Normal(0,1) at the two points defined above
  iCDF1 <- qnorm(q1)
  iCDF2 <- qnorm(q2)

  # now calculate the mean and sd
  mu <- (iCDF2 * log(x1) - iCDF1 * log(x2))/(iCDF2 - iCDF1)
  sigma <- (log(x1)-log(x2))/(iCDF1-iCDF2)

  # use unname() to remove names from the results
  return(list(mu=unname(mu), sigma=unname(sigma)))
}


# test
rvs <- rlnorm(10000, 100, 10)

# should be close to parameters above
ln_get_mean_var(rvs)

## $mu
## [1] 100.0189
##
## $sigma
## [1] 10.03689
```