The Normal, or Gaussian, distribution

Christopher A. Sims
Princeton University
sims@princeton.edu

September 22, 2020

©2020 by Christopher A. Sims. ©2020. This document is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.
http://creativecommons.org/licenses/by-nc-sa/3.0/

Likelihood for an i.i.d. Gaussian sample

- $x_i \sim N(\mu, \sigma^2)$
- Likelihood for the sample, size N:

$$\sigma^{-N}(2\pi)^{-N/2}\exp\left(\frac{-1}{2\sigma^2}\sum_i (x_i-\mu)^2\right).$$

Rewriting in terms of sufficient statistics

- Rewrite likelihood:

$$\sigma^{-N}(2\pi)^{-N/2}\exp\left(\frac{-1}{2\sigma^2}\Big(\sum_i x_i^2 - 2\mu\sum_i x_i + N\mu^2\Big)\right).$$

- If we write $\bar{x} = \sum_i x_i/N$ and $s^2 = \sum_i (x_i - \bar{x})^2/N$, this can be in turn rewritten as

$$\sigma^{-N}(2\pi)^{-N/2}\exp\Big(\frac{-N}{2\sigma^2}(s^2 + (\bar{x}-\mu)^2)\Big)$$

Sufficient statistic

- A lower-dimensional function of the data that does not depend on the parameters such that we can write the likelihood as a function of the sufficient statistic and the parameters alone.

- Here, $\bar{x}$ and s^2, or $\sum x_i$ and $\sum x_i^2$ form sufficient statistics.

- Why is this important? In a large sample, when there is no sufficient statistic, likelihood evaluation may involve a separate calculation for each observation in a long, computationally expensive, loop.

- When we are exploring the posterior, maximizing it or drawing from the distribution it describes, this has to be repeated for many values of the parameter vector. A sufficient statistic can make things much faster.

Flat-prior conditional pdf for $\mu \mid \sigma$

- Notice that the likelihood, as a function of μ with the data and σ^2 held fixed, is proportional to

$$\exp\left(\frac{-N(\mu-\bar{x})^2}{2\sigma^2}\right) ,$$

  which is proportional to a $N(\bar{x}, \sigma^2/N)$ distribution.

- In other words, the conditional distribution of μ is normal with mean at the sample mean and variance $1/N$ times the population variance.

Marginal for σ^2

- To get the marginal pdf for σ^2, we have to integrate μ out of the posterior.

- Since we know that μ enters in a term proportional to a normal pdf, integrating with respect to μ will just give us the inverse of the normalizing factor that makes that term match the normal.

- That is

$$\int \exp\left(\frac{-N(\mu-\bar{x})^2}{2\sigma^2}\right)\, d\mu = \sqrt{2\pi}\frac{\sigma}{\sqrt{N}}\,.$$

Marginal for σ^2

- So the likelihood function in σ^2, with μ integrated out, is

$$\sigma^{-(N-1)}(2\pi)^{-(N-1)/2}\exp\left(\frac{-Ns^2}{2\sigma^2}\right).$$

- As a function of $1/\sigma^2$, this is proportional to a Gamma distribution, which means that it is itself proportional to an **inverse gamma** distribution.

- As a function of $1/\sigma^2$ it is proportional to a $\mathrm{Gamma}((N+1)/2, Ns^2/2)$ pdf, but because of the Jacobian term arising out of the transformation from σ^2 to $1/\sigma^2$, it is an $\mathrm{IG}((N-3)/2, Ns^2/2)$.

- Often people use, instead of a flat prior on σ^2, a flat prior on $\log\sigma^2$, which is proportional to $1/\sigma^2$. This makes the marginal for σ^2 $\mathrm{IG}((N-1)/2, Ns^2/2)$.

Marginal for μ

- The joint posterior distribution for σ^2 and μ is called normal-inverse-gamma, for obvious reasons.

- Usually our main interest is μ, and we don't know σ^2, so we are interested in the marginal distribution for μ.

- Looking again at the likelihood in terms of sufficient statistics,

$$\sigma^{-N}(2\pi)^{-N/2}\exp\left(\frac{-N}{2\sigma^2}(s^2+(\bar{x}-\mu)^2)\right),$$

  we can see that it also is in the form of a Gamma as a function of $1/\sigma^2$ with μ held fixed. This lets us it integrate it out analytically.

Integration details

Let $u = 1/\sigma^2$. We rewrite the likelihood again, so that the Gamma density for u is grouped together:

$$u^{N/2} \exp\Big(\frac{N(s^2 + \bar{x} - \mu^2)}{2} \cdot u\Big) \frac{du}{u^2} (2\pi)^(N/2) \, .$$

We can drop the $(2\pi)^{N/2}$ constant, as it will cancel out when we normalize. The du/u^2 term arises from converting a flat prior on σ^2 to a flat prior on u. Combining the u^{-2} from du/u^2 with the initial term, we get

$$u^{N/2-2} \exp\Big(\frac{N(s^2 + \bar{x} - \mu^2)}{2} \cdot u\Big) \cdot du \, .$$

This is recognizable as the **kernel** of (i.e. is proportional to) a $\mathrm{Gamma}(N/2-1, .5N(s^2+(\bar{x}-\mu)^2))$ pdf. Inserting the normalizing constant for that distribution, and its inverse, we get

$$\left(\frac{N(s^2+(\bar{x}-\mu)^2)}{2}\right)^{(N/2-1)} \Gamma(N/2-1)^{-1}$$
$$\cdot\, u^{N/2-2} \exp\left(\frac{N(s^2+\bar{x}-\mu^2)}{2}\cdot u\right)\cdot du$$
$$\cdot \left(\frac{N(s^2+(\bar{x}-\mu)^2)}{2}\right)^{(-N/2+1)} \Gamma(N/2-1)\,.$$

The first two lines of this expression are, as a function of u, a properly normalized Gamma distribution and hence integrate to one in u. What's left is the third line, which cancels out terms we introduced to get a normalizing factor. It is what appears on the next slide.

Marginal for μ, cont.

The result, after removing multiplicative factors that don't depend on μ, is

$$\frac{1}{\left(1 + \frac{(\mu - \bar{x})^2}{s^2}\right)^{N/2-1}}\,.$$

This is proportional to a t-distribution density with $N-3$ degrees of freedom and scale factor $s/\sqrt{N-3}$.

Marginal for μ, cont., cont.

- The d.f. and scale parameter are found by matching to a standard t with ν degrees of freedom and scale factor σ, which has pdf proportional to

$$\frac{1}{\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{(\nu+1)/2}}$$

- Again, if we made our prior proportional to $1/\sigma^2$, the degrees of freedom would change, here to $N-1$, with the corresponding change in the scale factor.

About the t distribution

- As the degrees of freedom in the t increases, it becomes closer and closer to a normal distribution.

- Both are symmetric distributions around their means μ, but the tails of the t decline as one over a polynomial, much slower than the rate of decline in the tails of a normal.

- The degrees of freedom (d.f.) have to be larger than 0. With d.f. 1 or less, it has no well defined expected value. With d.f. 2 or less it has infinite variance.

Where t vs normal matters

- Marty Weitzman, in the paper cited on the reading list (and John Geweke, an econometrician, earlier) pointed out that standard calculations in quantitative finance can lead to strange conclusions if we recognize that the posterior distributions for parameters are likely to be t-distributed.
- A standard simple model has an optimizing agent with CRRA (constant relative risk aversion) utility whose income and consumption are the random yield Y on a security.

Expected utility

- The agent owns a quantity A of the asset, which she knows, and it will deliver Ae^Y in consumption goods. Expected utility is then

$$E\left[\frac{A^{1-\gamma}e^{(1-\gamma)Y}}{1-\gamma}\right]$$

Generally Y is assumed to be normally distributed, with mean μ and variance σ^2 estimated from historical data. For a $N(\mu,\sigma^2)$ Y, the expectation above evaluates to

$$\frac{A^{1-\gamma}e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2\sigma^2}}{1-\gamma}$$

The anomaly

- However, Weitzman and Geweke pointed out, if μ and σ^2 are unknown, this expression is only the conditional expected utility given the μ and σ^2.

- If we have estimated them from data, they have a normal-inverse-gamma joint posterior distribution.

- Because the inverse-gamma has a polynomially declining tail, when we integrate with respect to σ^2 to take the expectation, we are integrating an exponentially increasing term divided by a polynomially increasing term.

- This will be plus or minus infinity, depending on the sign of $1 - \gamma$.

Conclusion

CRRA uility, in other words, is extremely sensitive to fat tails in the distribution of the log of consumption. Uncertainty about variances of yields might explain very large risk premia.