Metric-Predicted Variable on One or Two Groups

Shubham Gupta

August 3, 2019

1 Introduction

- We have a metric predicted variable measured from two groups.
- **Aim**: Find out how different/same are these two measurements.

2 Estimating the mean and standard deviation of a normal distribution

- Normal distribution is given as:

$$p(y|\mu, \sigma) = \frac{1}{Z} exp\bigg(- \frac{1}{2} \frac{(y-\mu)^2}{\sigma^2} \bigg).$$

$$Z = \sigma(2\pi), (Zisthenormalizer).$$

- Probability density of a dataset D containing values y_1, y_2, y_3 is:

$$p(D|\mu, \sigma).$$

- How to allocate probability across combinations of μ and σ?

$$p(\mu, \sigma|D) = \frac{p(D|\mu, \sigma)p(\mu, \sigma)}{\int\int d\mu d\sigma p(D|\mu, \sigma)p(\mu, \sigma)}.$$

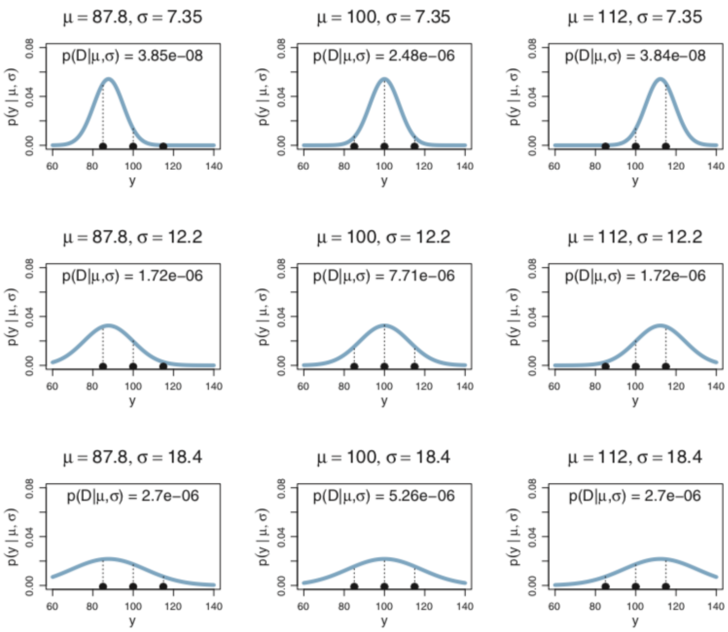

Figure 16.1 The likelihood $p(D|\mu,\sigma)$ for three data points, $D = \{85, 100, 115\}$, according to a normal likelihood function with different values of μ and σ. Columns show different values of μ, and rows show different values of σ. The probability density of an individual datum is the height of the dotted line over the point. The probability of the set of data is the product of the individual probabilities. The middle panel shows the μ and σ that maximize the probability of the data. (For another example, see Figure 2.4, p. 23.)

Figure 1: Combinations of μ and σ for different y values

3 The t distribution

- **t** disitrbution has heavy tails.

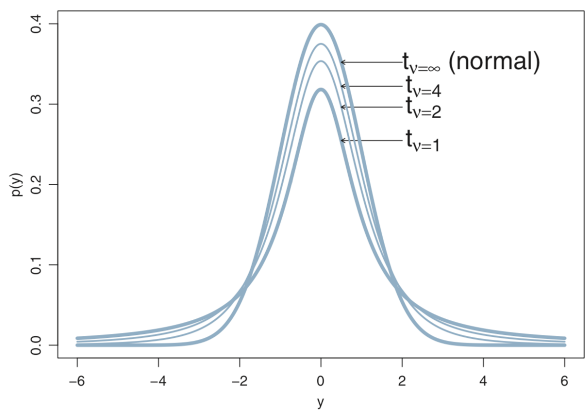

Figure 16.4 Examples of t distributions. In all cases, $\mu = 0$ and $\sigma = 1$. The normality parameter, ν, controls the heaviness of the tails. Curves for different values of ν are superimposed for easy comparison. The abscissa is labeled as y (not x) because the distribution is intended to describe predicted data.

Figure 2: Student's T disitribution

- It has 3 parameters:
  - μ controls the mean
  - σ controls the width
  - ν controls heaviness of tails. Also called **normality** parameter. When $\nu = 0$, heavy tails. When $\nu = \infty$ normal distribution shape.
  - We will use the t distribution as a descriptive model of data with outliers. t distributions are robust to outliers.

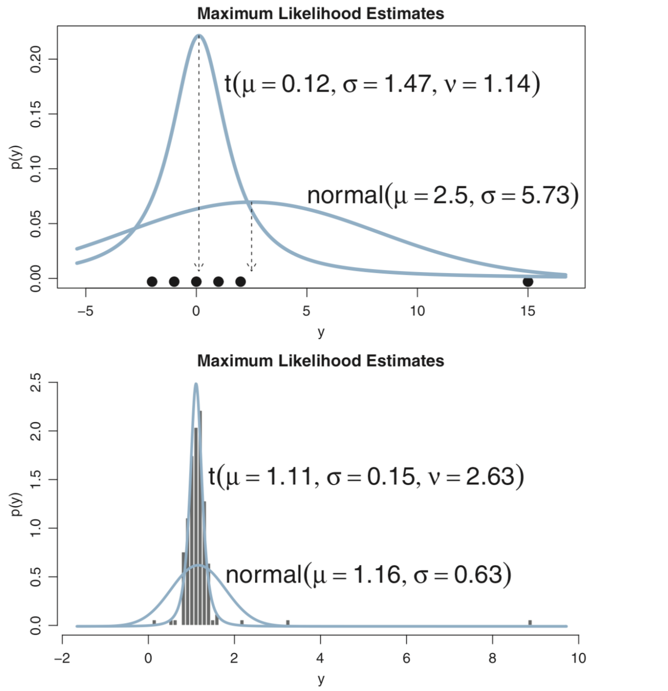

Figure 16.5 The maximum likelihood estimates of normal and *t* distributions fit to the data shown. Upper panel shows "toy" data to illustrate that the normal accommodates an outlier only by enlarging its standard deviation and, in this case, by shifting its mean. Lower panel shows actual data (Holcomb & Spalsbury, 2005) to illustrate realistic effect of outliers on estimates of the normal.

Figure 3: MLE of t distribution and normal distributions

  - We can see that the μ for the t distribution is much closer to the data points than the normal distribution. The last data point is accomodated by setting the **normality** to a **very small value**.
  - σ is **not the standard deviation** for t distribution. Standard deviation will be more than σ because of the large tails.
  - Use of tailed distributions is called **robust estimation** as it helps ensure our models ar erobust to outliers.
  - When we pick the prior distribution for ν, we need to pick a distribution that gives higher chance to values under 30 and lower chance to value above 30(since above 30, t distribution will be similar to normal distribution)

- **Exponential distribution** is used. It has one paramter: reciprocal of it's mean.
- Generally in JAGS/PyMC3, the exponential distribution will range from 0 to ∞. We will add 1 to this value to make the distribution range from 1 to ∞ with $mean = 1 + oldmean$

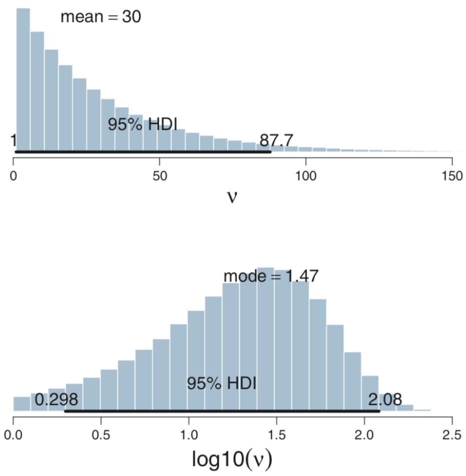

Figure 16.7 The prior on the normality parameter. Upper panel shows the shifted exponential distribution on the original scale of ν. Lower panel shows the same distribution on a logarithmic scale.

Figure 4: Prior for nommality parameter

- Above figure shows that the exponential distribution indeed gives more preference to values under 30.

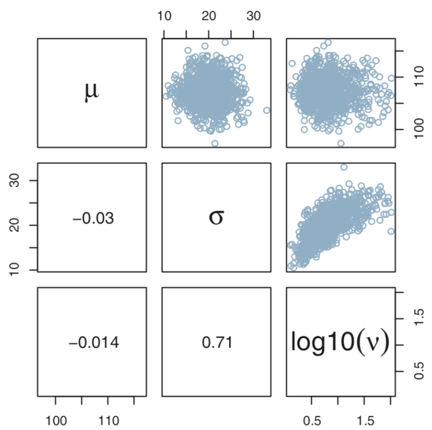

Figure 16.8 Posterior distribution of `Jags-Ymet-Xnom1grp-Mrobust-Example.R` applied to fictitious IQ data from a "smart drug" group. Off-diagonal cells show scatter plots and correlations of parameters indicated in the corresponding diagonal cells. Notice the strong positive correlation of σ and log10(ν).

Figure 5: Correlation of features

- Above plot shows that ν is **positively** related to σ. This implies as the distribution becomes more normal, the width of the distribution also increases. This indicates that the data **contains outliers**.

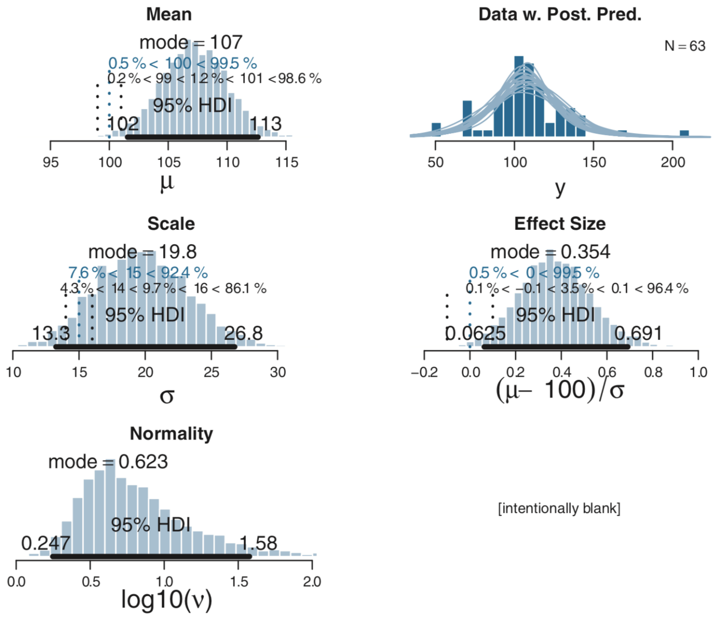

Figure 16.9 Posterior distribution of `Jags-Ymet-Xnom1grp-Mrobust` applied to fictitious IQ data from a "smart drug" group. Compare with Figure 16.3.

Figure 6: Posterior Distribution

- Above plot shows that the mode for the $log_{10}\nu$ distribution is around 0.68, compared to the initial prior value. Hence, the data is better fit with smaller ν values suggesting again that there are outliers in the data.
- For ν, any value above 30 represents a near normal distribution(or above 1.47 for $log_{10}\nu$).
- Upper right panel shows t-distribution can describe the data better than the normal distribution.

3.0.1 T-distribution in stan

- The t-distribution is represented using the $student_t(nu, mu, sigma)$.
- Stan model will converge faster because of the HMC method.

3.1 Two Groups

- Generally used when we want to compare two groups.
- Estimate mean and scale for each group. Since there will be lesser outliers, we can use a single normaility parameter for both groups.

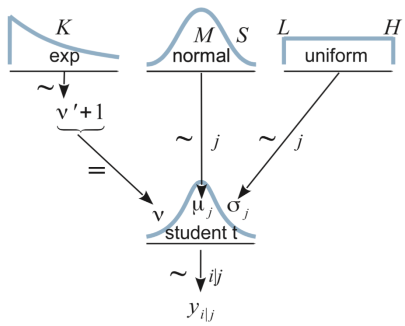

Figure 16.11 Dependency diagram for robust estimation of two groups. At the bottom of the diagram, $y_{i|j}$ is the ith datum within the jth group.

Figure 7: Two groups dependency graph

– We perform MCMC for the same IQ problem, and get the following:

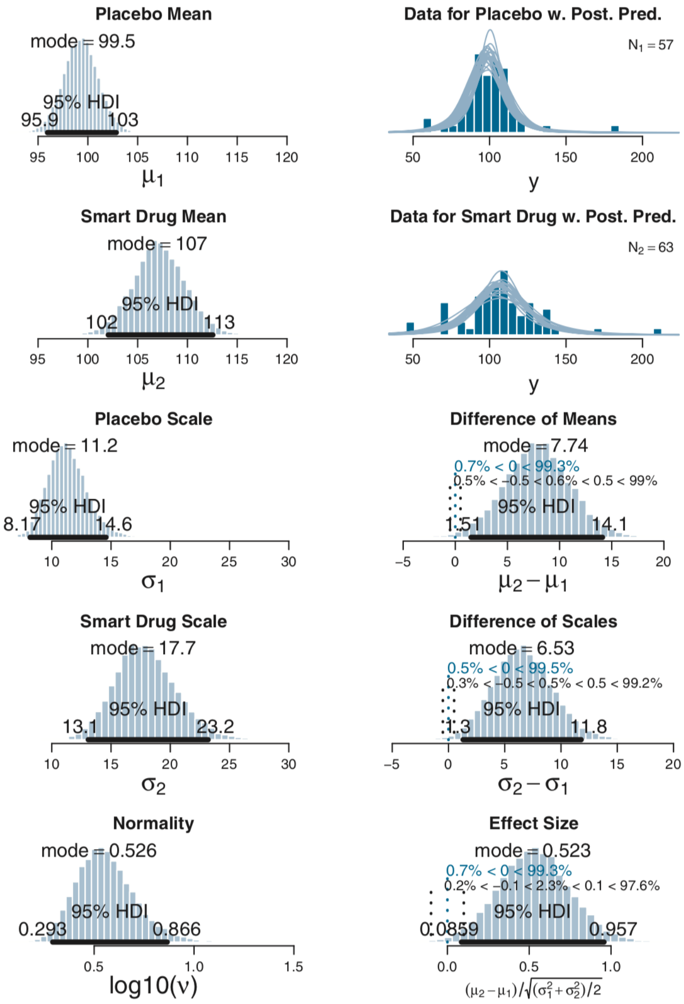

Figure 16.12 Posterior distribution for two groups.

Figure 8: Two groups posterior plot

- Problems with t test:
  - Provides only test of equality of means without test of equality of variances.
  - For equality of variances, run **F Test**. But this would inflate p values for both groups.

3.2 Noise Distribution

- If noise distribution does not match the data distribution, we can try the following:
  1. Use better noise distribution(LOL)
  2. Transform the data so that they match the shape of the assumed noise distribution.