LIKELIHOOD, POSTERIORS, DIAGNOSING NON-NORMALITY

(1) A distribution that allows asymmetry — different probabilities for negative and positive outliers — is the asymmetric double exponential, with pdf

$$p(x \mid \alpha, \beta) \propto \begin{cases} e^{-\alpha x} & \text{if } x \geq 0 \\ e^{\beta x} & \text{if } x < 0 \,. \end{cases}$$

The symbol $\propto$ stands for "is proportional to". To make this distribution integrate to one, we have to multiply the expressions above by $\alpha\beta/(\alpha + \beta)$. (And this is important for generating the likelihood function from the pdf).

Suppose we have an i.i.d. sample $\{x_1 ..., x_n\}$ from this distribution.

(a) Write down the likelihood for the sample. Show that the sum of the positive x_j's ($\sum x_j^+$) and the sum of the negative x_j's ($\sum x_j^-$), as a pair, form a sufficient statistic.

Because the sample is i.i.d., the pdf is the product of the n individual observation pdf's, i.e.

$$p(\vec{x}) = \prod_{j=1}^{n} \frac{\alpha\beta e^{-\alpha x_j^+ + \beta x_j^-}}{\alpha + \beta} = \left(\frac{\alpha\beta}{\alpha + \beta}\right)^n e^{-\alpha \sum x_j^+ \beta \sum (x_j^-} \,.$$

Since this sample pdf depends on the data $\vec{x}$ only via the two numbers $\sum x_j^+$ and $\sum x_j^-$, these are sufficient statistics. [Note that in general the pdf could depend on the data in other ways as well, so long as the pdf factors into one piece that depends on the sufficient statistics and the unknown parameters, and another that does not depend on the unknown parameters. The part that does not depend on the unknown parameters drops out when we consider the pdf as a function of the parameters (i.e. as a likelihood) and normalize it, or its product with a prior pdf, to integrate to one.]

(b) The mode of this distribution is obviously $x = 0$. Calculate the mean of the distribution as a function of α and β. [Hint: The mean of an exponentially distributed variable (with pdf $\alpha e^{-\alpha x}$ on $(0, \infty)$) is $1/\alpha$. The mean of x is the mean of the positive part of x times the probability of $x > 0$ plus the mean of the negative part of x times the probability of $x < 0$.]

The expected value of x conditioinal on $x > 0$ is the pdf over that region, normalized to integrate to one, i.e. the standard exponential with parameter α. So the mean conditional on $x > 0$ is $1/\alpha$. The mean conditional on $x < 0$ is, by the same argument with sign reversed, $-1/\beta$. The integral of $e^{-\alpha x}$ over $x > 0$ is $1/\alpha$, and the integral of $e^{\beta x}$ over $\beta < 0$ is $1/\beta$. Therefore $P[x > 0]/P[x < 0] = \beta/\alpha$, which in turn implies $P[x>0]=\beta/(\alpha+\beta)$. The unconditional mean of x is

$$P[x>0]E[x\mid x>0]+P[x<0]E[x\mid x<0]=\frac{\beta}{\alpha(\alpha+\beta)}-\frac{\alpha}{\beta(\alpha+\beta)}=\frac{\beta-\alpha}{\alpha\beta}:.$$

(c) Calculate the maximum likelihood estimates of α and β as a function of the sufficient statistics. This involves solving a pair of nonlinear equations in two unknowns, but they can be solved by hand. You get the equations to solve by finding first order conditions for a maximum of the likelihood (or the log likelihood, which is a little more convenient).
The log likelihood is

$$n\log(\alpha\beta)-n\log(\alpha+\beta)-\alpha\sum x_j^+=\beta\sum x_j^-\,.$$

The first order conditions with respect to α and β are

$$\partial\alpha: \qquad \frac{n}{\alpha}-\frac{n}{\alpha+\beta}-\sum x_j^+$$
$$\partial\beta: \qquad \frac{n}{\beta}-\frac{n}{\alpha+\beta}+\sum x_j^-\,.$$

If my algebra is right, the solution is

$$\alpha=\frac{n}{\Delta}\sqrt{1+\frac{\Delta}{S^+}}$$
$$\beta=\frac{n}{-\Delta}\sqrt{1+\frac{-\Delta}{S^-}},$$

where $S^+=\sum x_j^+$, $S^-=-\sum x_j^-$, and $\Delta=S^--S^+$.

(d) For the the "Bad day on Wall Street" daily per cent change in the DJIA data of Stock and Watson, the sum of the positive values is 3007.06 and the sum of the negative values is -2755.26. The number of observations is 7561. Find the corresponding maximum likelihood α and β, as well as the implied mean return. Are these estimates consistent with the idea that large negative values of the variable are more likely than large positive ones?
Plugging the given values into the equations for the MLE's gives us $\alpha=1.284515$, $\beta=1.342048$. They imply more rapid decay toward zero in the left (negative) tail of the distribution than in the right. So they contradict the notion that large negative values are more likely than large positive values. The sample data of largest absolute value are negative, but there are only a few of these in the distant tail. Of course this result might also reflect the fact that the sample mean of the data is positive, and if the model is to imply a positive mean return, it must have $\alpha<\beta$.

(e) Suppose we had a prior pdf on α and β that made them independent, and with the identical pdf's $10e^{-10\alpha}$, $10e^{-10\beta}$. What would the posterior pdf be? what would be the values of α and β that maximize the posterior pdf? [Hint:

The posterior pdf can be put into a form that looks just like the likelihood, but with altered values for $\sum x_j^+$ and $\sum x_j^-$, so the same formulas you used to maximize the likelihood can be re-used.]

With this prior pdf, the sample pdf times the prior pdf becomes

$$\left(\frac{\alpha\beta}{\alpha+\beta}\right)^n e^{-\alpha(\sum x_j^+ + 10) + \beta \sum (x_j^- - 10}\,.$$

As a function of α and β, this is just the original likelihood with S^+ and S^- both increased in absolute value by 10. Plugging these modified values of S^+ and S^- into the formulas for the MLE's, we get $\alpha = 1.280164$, $\beta = 1.337298$. Since the prior density is monotone decreasing away from zero, it is expected that the MLE"s get pulled toward zero, but despite the rather strong priors, the estimates are changed only slightly. The large sample makes the likelihood dominate ther prior even though the prior probability of, e.g., a β as large as the MLE is less than .000002.

(2) We might also consider a t distribution for the "Bad Day on Wall Street" data. These data are available as an R data file `bdws.RData` that can be loaded into R with the `load()` function, or as a csv file that can be loaded into R with `read.csv()` or used in another program. If it is loaded by `bdws <- load("bdws.RData")`, `bdws$PctChg` will be the percentage change time series we are interested in. As displayed in class, the t pdf with ν degrees of freedom, location parameter μ, and scale parameter σ, is given by

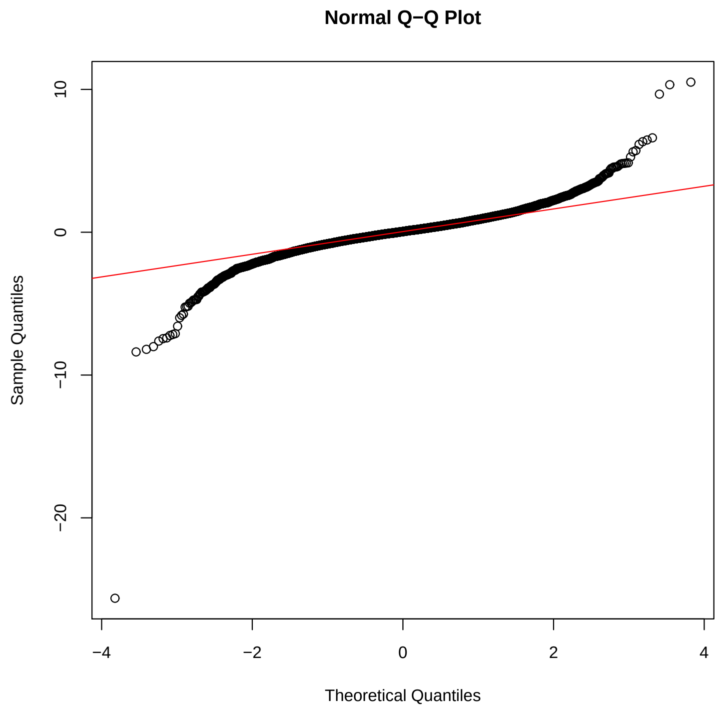
$$p(x \mid \mu, \sigma, \nu) = \frac{\Gamma((\nu+1))/2)}{\Gamma(\nu/2)\sigma\sqrt{\nu\pi}} \left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2} .$$

[Note: If you use the R built-in facilities for computing densities, cdf's, and quantiles of distributions, you don't need to make any use of the expression above for this problem set.] The pdf of an i.i.d. sample from this distribution does not have a sufficient statistic. The maximum likelihood values for the parameters of this distribution for the `bdws$PctChg` data are

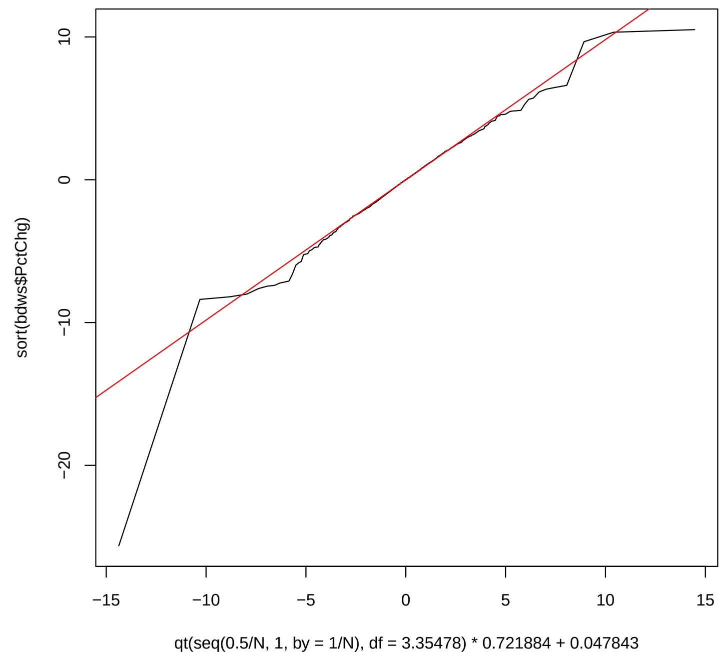
$$\mu = 0.04784383, \sigma = 0.72188439, \nu = 3.35478224 .$$

(a) Make a normal quantile-quantile plot (`qqnorm()` does this in R) for the data, including a reference line showing what the slope would be if the data were normal. The reference line can be produced with `qqline()` in R. Its default settings simply calculate and plot a straight line through the two points defined by the .25 and .75 quantiles of the sample distribution and reference distribution (which is by default normal).

(b) Make a quantile-quantile plot of the data against the maximum likelihood t distribution. Comment on whether it looks better than the normal q-q plot, and explain your conclusion. This plot requires using `qqplot()`, which requires that you specify the theoretical quantile function to be used in the plot. The R `qt()` function produces quantiles of the t, but its only parameter is the degrees of freedom ν. To account for the location and scale parameters, you need to give as the `distribution` argument in `qqline`
`function(p) qt(p, df=3.35478224)*sig + mu.`
The call to `qqplot()` should have `x = qt(ppoints(N), df=3.35478224)*sig + mu`, where N is the number of observations, `length(bdws$PctChg)`.
The examples at the bottom of the R help page on `qqplot` may be helpful.
Here's the normal qq plot, from `qqnorm(bdws$PctChg)` followed by `qqline(bdw$PctChg)`



Here's the *t* qqplot. The formula at the bottom shows how it was computed. The default `qqnorm()` and `qqplot()` plot individual points, not lines. I used `type="l"` to make it a line. But this requires using `sort(bdws$PctChg)` as the argument rather than `bdws$PctChg` itself, as otherwise the lines zigzag all over the place.

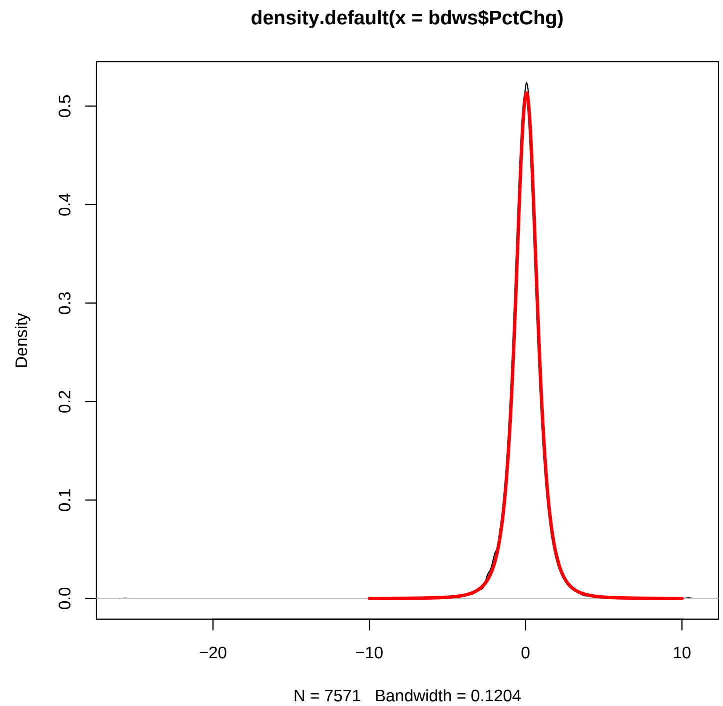


(c) [extra credit] Create a plot containing the estimated smoothed pdf for the data obtained with `density()`, the t density with maximum likelihood parameters for the sample, and the normal density with maximum likelihood parameters for the data (i.e., `mean(bdws$PctChg)` and `sd(bdwd$PctChg)`). Use different colors for the three lines. This is extra credit because getting these functions scaled and shifted so they all integrate to one requires some care and thought. Note that if $p(x)$ is the density of x, $p(y/\sigma)/\sigma$ is the density of $y = x\sigma$.

Turns out this is not that hard to set up. I used this R code:

```
plot(density(bdws$PctChg))
lines(seq(-10, 10, length=500), dt((seq(-10, 10, length=500) - .047843)/0.72188439,
                                    df=3.35478224)/.72188439, col="red", lwd=3)
```

The `lwd=3` is to make the t density plot a slightly fatter line, so one can see that it lies almost perfectly on top of the smoothed sample pdf for the data.



One final note. The largest negative value of the variable is -25.63, or a deviation from μ of 25.68. The probability of this big a deviation from μ in the negative direction is less than 1 in 100,000, under our estimated t distribution. But we have over 7,000 observations. It turns out that the probability of seeing a deviation this big in a sample this size is about .07. So the occurrence of this big negative value is slightly surprising, but, in contrast with what would be implied by a normal distribution, not at all impossible.