

# Biostatistics 234: Homework 1

Fernando Mora

## Problem 1: Normal Data with a Normal Prior

1. My measurement is the daily amount of table salt used in my household. I measured a small glass bowl of sea salt every morning before breakfast for 8 days, and recorded the difference in weight in grams.

```
x <- c(9, 8, 4, 0, 8, 8, 1, 11)
```

2. I usually measure out salt in grams when cooking pizza, so based on what a teaspoon looks like, I assumed I might use about 7 grams of salt per day. Given days we don't cook, or cook a lot, this could be 5 grams more or five grams less, so the standard deviation prior will be 5. My prior mean distribution would thus be  $\mu_0 \sim N(7, 5)$ .
- 3.

Table 1: Daily Salt Use in Grams

| Wed | Thu | Fri | Sat | Sun | Mon | Tue | Wed | Mean  | Variance |
|-----|-----|-----|-----|-----|-----|-----|-----|-------|----------|
| 9   | 8   | 4   | 0   | 8   | 8   | 1   | 11  | 6.125 | 15.83929 |

4. I decided to set my sampling standard deviation based on the observed s.d., which was 3.98

$$5. \bar{\mu} = \frac{\frac{8}{3.98^2} + \frac{1}{5^2}}{\frac{8}{3.98^2} + \frac{1}{5^2}} 6.125 + \frac{\frac{1}{5^2}}{\frac{8}{3.98^2} + \frac{1}{5^2}} 3.98 = 5.97.$$

$$var(\mu|data) = (\frac{8}{3.98^2} + \frac{1}{5^2})^{-1} = 1.83.$$

$$sd(\mu|data) = 1.35$$

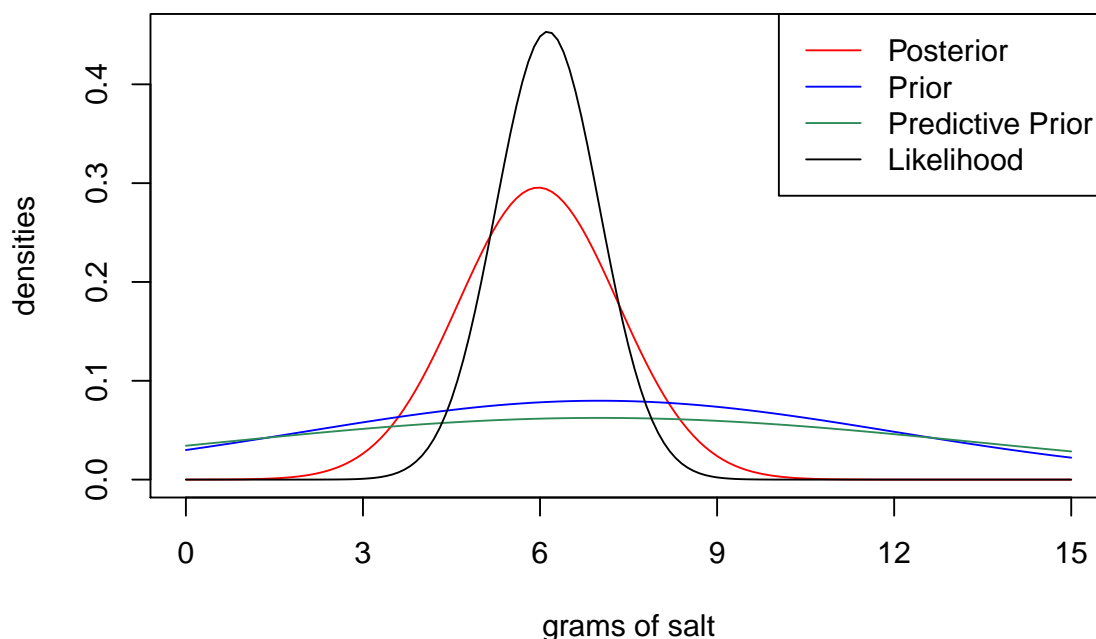
- 7.

Table 2: Summary of Posterior, prior, predictive, and likelihood densities

|                  | Mean | sd   | Variance |
|------------------|------|------|----------|
| Posterior        | 5.97 | 1.35 | 1.83     |
| Prior            | 7.00 | 5.00 | 25.00    |
| Prior Predictive | 7.00 | 6.39 | 40.84    |
| Likelihood       | 6.13 | 1.00 | 0.77     |

8. The plot of our four densities shows that the posterior estimate and the likelihood function are more accurate in terms of variance in the distribution of observations. Interestingly, to me, the likelihood function had the smallest overall variance and the shortest tails, due to my very generous prior specification for variance. The prior and prior predictive distributions had enormously wide tails and low certainty around the mean estimate.

### Plot of Posterior, prior, predictive, and likelihood densities



### Problem 2: Count Data with a Gamma Prior

1. The support for the prior and sampling densities would entail a Poisson distribution, so its support  $k$  will be  $k \in \mathbb{N}_0$ , or any positive natural number, starting at zero. For the Gamma distribution as conjugate prior and posterior, the support will be  $k \in (0, \infty)$ , or any number from zero to infinity. Really not clear about support for the likelihood! If it is to estimate the probability of our estimate being 0, then  $k \in 0 < x < 1$ ?
2. The parameter  $b$  will act like a prior sample size, because it can be used as the denominator of the distribution mean  $\frac{a}{b}$ , which makes me think of  $n$  in terms of arithmetic mean.
3. ...
4. To avoid unnecessary risk of infection, I did not go to Ralph's, but instead decided to count the number of daily emails sent from our director of student services every weekday. The 5 observations will be Monday-Friday from January 11 to January 15.
5. Priors
  - 5a) I will guess  $\mu_0 = 7$ ; the average number of emails Amy sends daily.
  - 5b) Prior  $s_0 = 5$ ; the standard deviation of emails Amy sends in a day.
  - 5c) I will guess  $n_0 = 1/5$ , that my prior estimate for  $\mu_0$  would need 5 actual observations before  $\bar{y}$  looked like my guess.
  - 5d) Assuming  $7 = a/b$  and  $5^2 = a/b^2$ :  $a_1 = 49/25 = 1.96$ .  $b_1 = 7/25 = 0.28$ .
  - 5e) Assuming  $7 = a/b$  and  $b = 1/5$ :  $a_2 = 7/5 = 1.4$ .  $b_2 = 1/5 = 0.2$ .

6. Since we saw above we can estimate priors  $(a, b)$  with a guess for mean and standard deviation, or alternatively mean and sample size, we could set prior based on our best guess as to the expected mean and its variance, or we can set it with our guess of the mean value and our guess as to the  $k$  worth of our prior relative to the number of observations.
- 7.

Table 3: Daily Emails from OSS

| Mon | Tue | Wed | Thu | Fri | Sum | Mean | Variance |
|-----|-----|-----|-----|-----|-----|------|----------|
| 8   | 3   | 2   | 7   | 6   | 26  | 5.2  | 6.7      |

8a. Posterior one:

$\text{Gamma}(1.96 + 26, 0.28 + 5)$

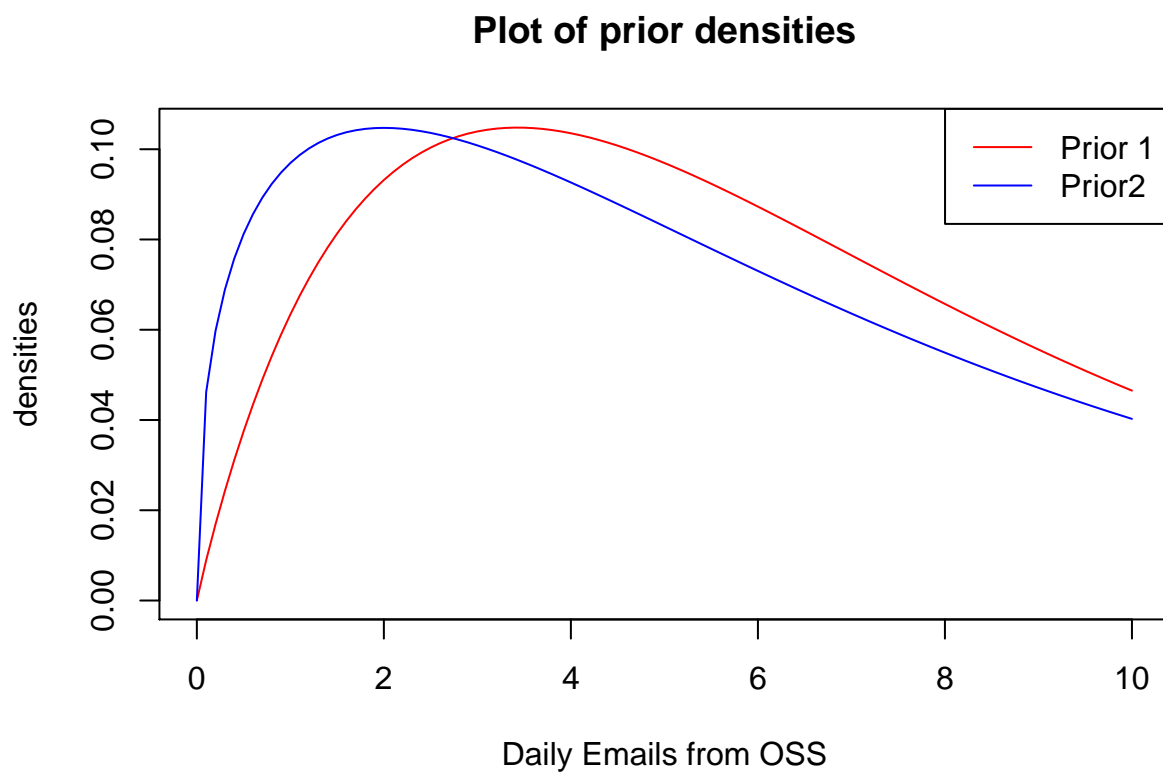
8b. Posterior two:

$\text{Gamma}(1.4 + 26, 0.2 + 5)$

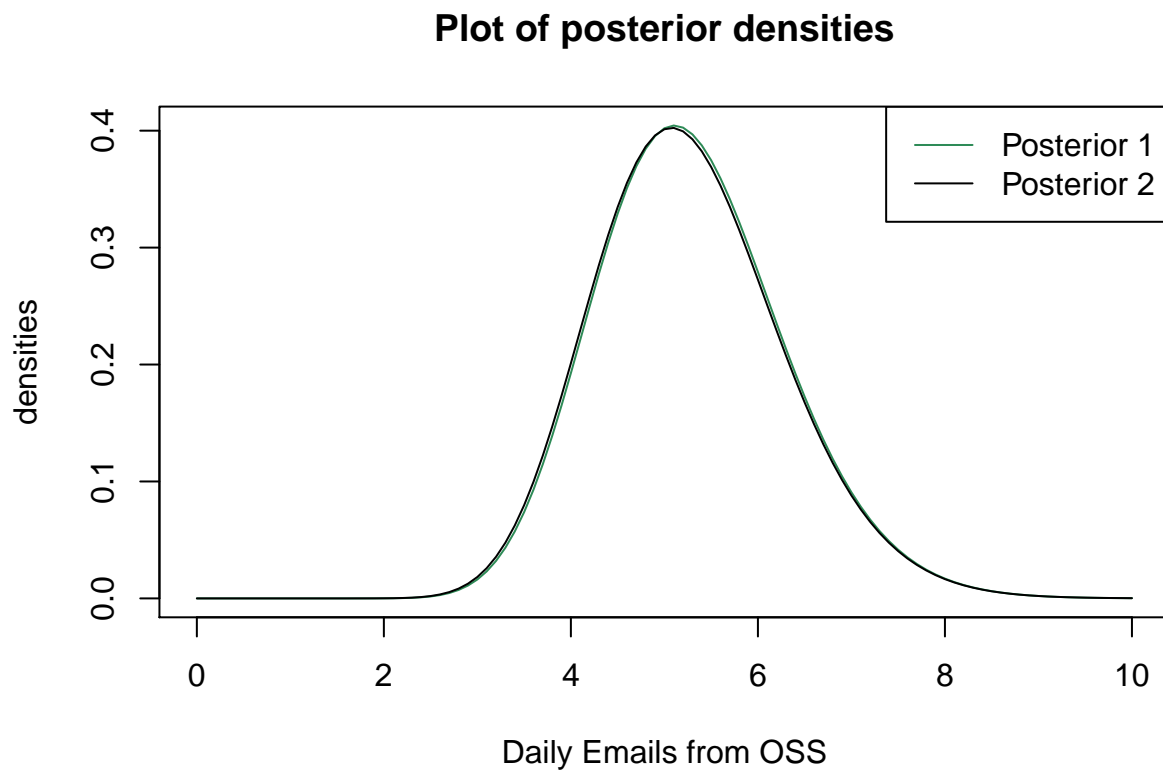
9a.  $\bar{\mu}|Y = 5.3, sd(\bar{\mu}|data) = 2.23$

9b.  $\bar{\mu}|Y = 5.27, sd(\bar{\mu}|data) = 2.28$

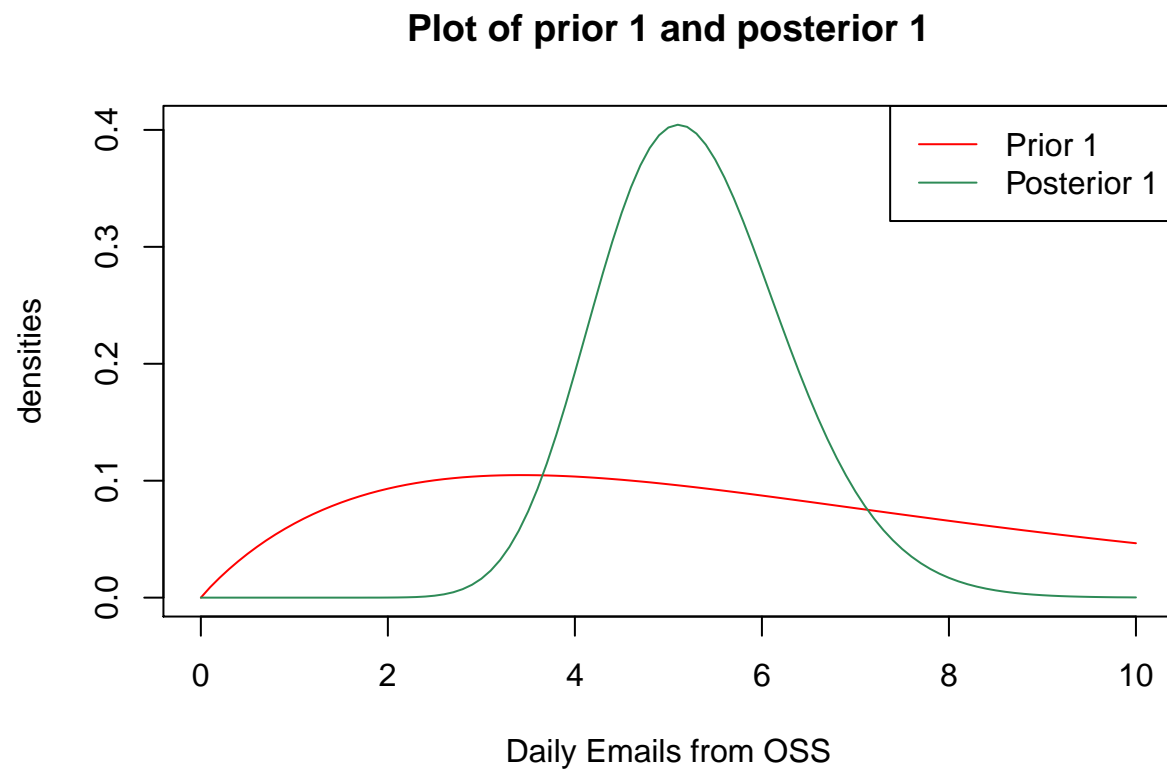
10a. Prior density 1 had a lower rate than prior 2 (5 vs. 7, respectively) which made the height of the second density appear left-skewed or closer to 0.



10b. The posterior means were only different by a few hundredths of a decimal, and their standard deviations were both virtually equal by a few thousands of a decimal, so their densities overlap even more closely and look like a nicely normal density.



11a. The posterior density has much smaller tails than its prior, and a sharply taller height density around  $\bar{\mu}$ .



11b. The same can be said here, where the second posterior density has a much higher density around  $\bar{\mu}$ , and much smaller tails than its corresponding prior, and no more apparent skewness.

**Plot of prior 2 and posterior 2**

