(ROS) textbook

Please take a look at Chp.9, 10, (and 6 if time permits) and check

1. using simulation as inference summary takes care of uncertainty propagation
2. three levels of uncertainty (esp. blown uc for prediction)
3. “For linear regression, the expected value, E(y|xnew), is the same as the linear predictor, Xnew β, but as we discuss in Chapter 13, these two quantities differ for nonlinear model”

Using the matrix of posterior simulations to express uncertainty about a parameter estimate or function of parameter estimates

The real advantage of summarizing inference by simulations is that we can directly use these to propagate uncertainty. For a simple example, suppose we want an estimate and standard error for some combination of parameters, for example a/b (not that there would be any good reason to do this for this problem, just to demonstrate it as an example). We can work with the simulations directly:


```
a <- sims[,1]
b <- sims[,2]
z <- a/b
print(c(median(z), mad(z)))
```

9.2 Prediction and uncertainty: `predict`, `posterior_linpred`, and `posterior_predict`

After fitting a regression, $y = a + bx + \text{error}$, we can use it to predict a new data point, or a set of new data points, with predictors x^{new}. We can make three sorts of predictions, corresponding to increasing levels of uncertainty:

- The *point prediction*, $\hat{a} + \hat{b}x^{\text{new}}$: Based on the fitted model, this is the best point estimate of the average value of y for new data points with this new value of x. We use $\hat{a}$ and $\hat{b}$ here because the point prediction ignores uncertainty.
- The *linear predictor with uncertainty*, $a + bx^{\text{new}}$, propagating the inferential uncertainty in (a, b): This represents the distribution of uncertainty about the expected or average value of y for new data points with predictors x^{new}.
- The *predictive distribution for a new observation*, $a + bx^{\text{new}} + \text{error}$: This represents uncertainty about a new observation y with predictors x^{new}.

For example, consider a study in which blood pressure, y, is predicted from the dose, x, of a drug. For any given x^{new}, the point prediction is the best estimate of the average blood pressure in the population, conditional on dose x^{new}; the linear predictor is the modeled average blood pressure of people with dose x^{new} in the population, with uncertainty corresponding to inferential uncertainty in the coefficients a and b; and the predictive distribution represents the blood pressure of a single person drawn at random drawn from this population, under the model conditional on the specified value of x^{new}.

For example, let’s predict the weight of a person who is 70 inches tall, so that `c_height = height − 66 = 4`:

$$\text{linear predictor: } a + 4.0\,b,$$
$$\text{predicted value: } a + 4.0\,b + \epsilon.$$

Sometimes we are interested in the predicted expectation or linear predictor, as it represents the predicted average weight for everyone of this height in the population; in other settings we want to predict the weight of an individual person. It depends on the context, and we need to be able to assess the uncertainty for each.

The above model gives a point prediction for a 70-inch-tall person to weigh 153.2+4.0∗4.9 = 172.8 pounds. If this equation represented the true model, rather than an estimated model, then we could use $\hat{\sigma} = 29.1$ as an estimate of the standard deviation for the predicted value. Actually, though, the estimated error standard deviation is slightly higher than $\hat{\sigma}$, because of uncertainty in the estimate of the regression parameters—a complication that gives rise to those special prediction standard deviation formulas seen in some regression texts. For example, in the regression model $y = a + bx + \text{error}$, the standard deviation for the linear predictor $a + bx$ is

$$\hat{\sigma}_{\text{linpred}} = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^{\text{new}} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, \tag{9.1}$$

and the standard deviation for the predicted value $a + bx + \epsilon$ is

$$\hat{\sigma}_{\text{prediction}} = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x^{\text{new}} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}. \tag{9.2}$$

Rather than trying to use such formulas (which do not even work once we move to more complicated models and nonlinear predictions), we simply compute all predictive uncertainties using simulations in R.

For a linear model, the two sorts of predictive distributions have the same mean, the point prediction $\hat{a} + \hat{b}x$.

The difference comes in the uncertainties: as the sample size n in the fitted regression increases, the standard deviation of the linear predictor goes to zero (it scales like $1/\sqrt{n}$), while the standard deviation of the predicted value approaches the nonzero σ, as illustrated in formula (9.2). Even if the parameters of the fitted model are perfectly known, there is uncertainty in predicting a new data point.

s.e., qoi, PRIOR

4. feature noise as extra source of uncertainty increases standard error i.e. estimated uncertainty of qoi (quantity or interest).

5. Don’t be confused by the general application of Bayes for every uncertain things (prediction, hidden state, model parameter). Just stick to prior and dgp.

Propagating uncertainty

In the above calculations we expressed uncertainty in the election outcome conditional on various pre-set values of economic growth. But growth is estimated only approximately in the lead-up to the election, and later the figures are often revised by the government. Hence it makes sense when applying our model to account for uncertainty in this predictor.

Let us say that, in advance of the election, our best estimate of economic growth was 2.0% but with some uncertainty that we shall express as a normal distribution with standard deviation 0.3%. We can then *propagate the uncertainty* in this predictor to obtain a forecast distribution that more completely expresses our uncertainty.

We just need to add a line to our R code to simulate the distribution of the prediction:

```
x_new <- rnorm(n_sims, 2.0, 0.3)
y_pred <- rnorm(n_sims, a + b*x_new, sigma)
```

Following this, we summarize as before, obtaining this result:

```
Predicted Clinton percentage of 2-party vote: 52.3, with s.e. 4.1
Pr (Clinton win) = 0.71.
```

The point prediction is unchanged at 52.3% of the two-party vote, but the standard deviation has increased slightly to reflect this extra uncertainty.

Another complication is that Bayesian inference can be applied to any uncertain quantity. In the above example, the uncertain quantity is an observable outcome, the incumbent party’s vote share in the next election. In Section 9.4, the uncertain quantity is the difference in the probability of a baby being a girl, comparing two different categories of parents. In Section 9.5 we discuss the more general problem where the uncertain quantity is a regression coefficient or set of regression coefficients.

The prior distribution will be set up differently in these different situations. For the voting example, a regression model fit on past data gives predictive inference for a single observable outcome that can be used as prior information when combining with a poll on the current election. In the births example, we need prior information on a population difference. In the general example of regression modeling, we must specify prior information on all the coefficients, which in practice often entails weak priors on coefficients for which we have little knowledge or about which we do not want to make any strong assumptions.