

Posterior Predictive Distribution

- ▶ Recall that for a fixed value of θ , our data \mathbf{X} follow the distribution $p(\mathbf{X}|\theta)$.
- ▶ However, the true value of θ is uncertain, so we should average over the possible values of θ to get a better idea of the distribution of \mathbf{X} .
- ▶ **Before** taking the sample, the uncertainty in θ is represented by the prior distribution $p(\theta)$. So for some new data value x_{new} , averaging over $p(\theta)$ gives the **prior predictive distribution**:

$$p(x_{new}) = \int_{\Theta} p(x_{new}, \theta) d\theta = \int_{\Theta} p(x_{new}|\theta)p(\theta) d\theta$$

Posterior Predictive Distribution

- ▶ **After** taking the sample, we have a **better representation** of the uncertainty in θ via our posterior $p(\theta|\mathbf{x})$. So the **posterior predictive distribution** for a new data point x_{new} is:

$$\begin{aligned} p(x_{new}|\mathbf{x}) &= \int_{\Theta} p(x_{new}|\theta, \mathbf{x})p(\theta|\mathbf{x}) d\theta \\ &= \int_{\Theta} p(x_{new}|\theta)p(\theta|\mathbf{x}) d\theta \end{aligned}$$

(since x_{new} is independent of the sample data \mathbf{x})

- ▶ This reflects how we would predict new data to behave / vary.
- ▶ If the data we **did observe** follow this pattern closely, it indicates we have chosen our model and prior well.

Posterior Predictive Distribution

Example 2 again: $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$,

$$\lambda \sim \text{Gamma}(\alpha, \beta)$$

$$\lambda | \mathbf{x} = \text{Gamma}(\sum x_i + \alpha, n + \beta)$$

Posterior predictive distribution is:

$$\begin{aligned} p(x_{\text{new}} | \mathbf{x}) &= \int_0^{\infty} p(x_{\text{new}} | \lambda) p(\lambda | \mathbf{x}) d\lambda \\ &= \int_0^{\infty} \left[\frac{\lambda^{x_{\text{new}}} e^{-\lambda}}{(x_{\text{new}})!} \right] \left[\frac{(n + \beta)^{\sum x_i + \alpha}}{\Gamma(\sum x_i + \alpha)} \lambda^{\sum x_i + \alpha - 1} e^{-(n + \beta)\lambda} \right] d\lambda \end{aligned}$$

Posterior Predictive Distribution

So

$$\begin{aligned} p(x_{new}|\mathbf{x}) &= \frac{(n + \beta)^{\sum x_i + \alpha}}{\Gamma(\sum x_i + \alpha)\Gamma(x_{new} + 1)} \int_0^{\infty} \lambda^{x_{new} + \sum x_i + \alpha - 1} e^{-(n + \beta + 1)\lambda} d\lambda \\ &= \frac{(n + \beta)^{\sum x_i + \alpha}}{\Gamma(\sum x_i + \alpha)\Gamma(x_{new} + 1)} \frac{\Gamma(x_{new} + \sum x_i + \alpha)}{(n + \beta + 1)^{x_{new} + \sum x_i + \alpha}} \\ &= \frac{\Gamma(x_{new} + \sum x_i + \alpha)}{\Gamma(\sum x_i + \alpha)\Gamma(x_{new} + 1)} \left(\frac{n + \beta}{n + \beta + 1}\right)^{\sum x_i + \alpha} \left(\frac{1}{n + \beta + 1}\right)^{x_{new}} \end{aligned}$$

which is a negative binomial with mean $\frac{\sum x_i + \alpha}{n + \beta}$ and variance $\frac{\sum x_i + \alpha}{(n + \beta)^2} (n + \beta + 1)$.

Posterior Predictive Distribution

- ▶ \Rightarrow The posterior predictive distribution has the same mean as the posterior distribution, but a **greater** variance (additional “sampling uncertainty” since we are drawing a **new** data value).
- ▶ See R example (Prussian army data).

More about Posterior Predictive Distribution

- ▶ **Example 1(a) again:** $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, σ^2 known.
- ▶ Posterior for $\mu|\mathbf{x}$ is normal with mean

$$\mu_{\text{post}} = \frac{\delta/\tau^2 + n\bar{x}/\sigma^2}{1/\tau^2 + n/\sigma^2}$$

and variance

$$\sigma_{\text{post}}^2 = \frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2}.$$

- ▶ Note $x_{\text{new}}|\mu \sim N(\mu, \sigma^2)$, so the posterior predictive distribution is:

$$p(x_{\text{new}}|\mathbf{x}) = \int_{-\infty}^{\infty} p(x_{\text{new}}|\mu)p(\mu|\mathbf{x}) d\mu.$$

More about Posterior Predictive Distribution

- ▶ Sometimes the form of $p(x_{new}|\mathbf{x})$ can be derived directly, but it is often easier to sample from $p(x_{new}|\mathbf{x})$ using Monte Carlo methods:
- ▶ For $j = 1, \dots, J$, sample
 1. $\mu^{[j]}$ from $p(\mu|\mathbf{x})$ and
 2. $x^{*[j]}$ from $p(x_{new}|\mu^{[j]})$
- ▶ Then $x^{*[1]}, \dots, x^{*[J]}$ are an iid sample from $p(x_{new}|\mathbf{x})$.
- ▶ See R example with lead data.

Posterior Predictive Distribution in Regression

Example 3: In the regression setting, we have shown that the posterior predictive distribution for a new response vector \mathbf{y}^* is multivariate-t.

- ▶ To check model fit, we can generate samples from the posterior predictive distribution (letting \mathbf{X}^* = the observed sample \mathbf{X}) and plot the values against the y -values from the original sample.
- ▶ If an observed y_i falls far from the center of the posterior predictive distribution, this i -th observation is an outlier.
- ▶ If this occurs for many y -values, we would doubt the adequacy of the model.
- ▶ See R example (small automobile data set).