Local Sensitivity Analysis

- Unfortunately, it may be too difficult to examine a large class of prior specifications, especially when the target parameter $\boldsymbol{\theta}$ is multidimensional.
- **Local** sensitivity analysis simply focuses on how changes in the hyperparameter value(s) affect the posterior.
- **Example 1(a)**: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, σ^2 known.
- Conjugate prior for μ: $\mu \sim N(\delta, \tau^2)$
- Compare resulting posterior (the plot and/or quantiles) to the posterior from these priors:

$$\mu \sim N(\delta - \tau, \tau^2)$$
$$\mu \sim N(\delta + \tau, \tau^2)$$
$$\mu \sim N(\delta, 0.5\tau^2)$$
$$\mu \sim N(\delta, 2\tau^2)$$

See `R` example.

Local Sensitivity Analysis

- **Example 1(b)**: $X_1, \ldots, X_{200}$ are annual deaths from horse kicks for 10 Prussian cavalry corps for each of 20 years.
- Let $X_i \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$, and let $\lambda \sim \text{Gamma}(\alpha, \beta)$ be the prior.
- Compare posteriors from these priors for λ:

$$
\begin{aligned}
\lambda &\sim \text{Gamma}(2, 4) \\
\lambda &\sim \text{Gamma}(4, 8) \\
\lambda &\sim \text{Gamma}(1, 2) \\
\lambda &\sim \text{Gamma}(0.1 \times 2, \sqrt{0.1} \times 4) \\
\lambda &\sim \text{Gamma}(3 \times 2, \sqrt{3} \times 4)
\end{aligned}
$$

See `R` example with Prussian horse kick data.
General recommendation when the posterior is highly sensitive to changes in prior specification:Choose a more "objective" prior (or be prepared to defend your prior knowledge!).

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**:

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

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:

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

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

(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: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$,

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

Posterior predictive distribution is:

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

Posterior Predictive Distribution

So

$$p(x_{new}|\mathbf{x}) = \frac{(n+\beta)^{\sum x_i+\alpha}}{\Gamma(\sum x_i+\alpha)\Gamma(x_{new}+1)} \int\limits_0^\infty \lambda^{x_{new}+\sum x_i+\alpha-1} e^{-(n+\beta+1)\lambda}\, \mathrm{d}\lambda$$

$$= \frac{(n+\beta)^{\sum x_i+\alpha}}{\Gamma(\sum x_i+\alpha)\Gamma(x_{new}+1)} \frac{\Gamma(x_{new}+\sum x_i+\alpha)}{(n+\beta+1)^{x_{new}+\sum x_i+\alpha}}$$

$$= \frac{\Gamma(x_{new}+\sum x_i+\alpha)}{\Gamma(\sum x_i+\alpha)\Gamma(x_{new}+1)} \left(\frac{n+\beta}{n+\beta+1}\right)^{\sum x_i+\alpha} \left(\frac{1}{n+\beta+1}\right)^{x_{new}}$$

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**: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, 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

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

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

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

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 $j = 1, \ldots, J$, sample
  1. $\mu^{[j]}$ from $p(\mu|\mathbf{x})$ and
  2. $x^{*[j]}$ from $p(x_{new}|\mu^{[j]})$
- Then $x^{*[1]}, \ldots, x^{*[J]}$ are an iid sample from $p(x_{new}|\mathbf{x})$.
- See `R` example with lead data.