

## Handout 18: Bayesian Statistics II

**Bayesian Statistics** Let's start by defining the setup from last in a formal, generic way. Bayesian statistics first considers a random variable  $\vec{\theta}$ , called the **prior**, with a pdf/pmf equal to  $f_{\vec{\theta}}$ . Then, we have a sequence  $X_1, \dots, X_n$  of i.i.d. random variables, each of which is defined through the conditional pdf/pmf  $f_{x_j|\vec{\theta}}(x_j|\vec{\theta})$ . Finally, we can compute the **posterior** distribution of  $\vec{\theta}|\vec{x}$  by using a form of Bayes rule:<sup>1</sup>

$$f_{\vec{\theta}|\vec{x}}(\vec{\theta}|\vec{x}) \propto \left[ \prod_j f_{x_j|\vec{\theta}}(x_j) \right] \times f_{\vec{\theta}}.$$

<sup>1</sup> I am using the notation that  $\vec{X} = (X_1, \dots, X_n)$ , and likewise for  $\vec{x}$ .

Generally, the main task in Bayesian statistics is to determine the shape of the posterior when given a prior and likelihood.

**Point Estimation** While the entire posterior distribution is the clearest picture of our knowledge of the parameter  $\vec{\theta}$ , sometimes we need to convert our knowledge into a single best guess point estimator. The Bayesian point estimator is the expected value of the posterior distribution. In general, this will be somewhere between the expected value of the prior distribution and the mean of the data  $\vec{X}$ .

**Credible Intervals** We can represent a version of a confidence interval for Bayesian statistics. For a univariate  $\theta$ , a **credible interval** with credibility  $1 - \alpha$  is simply a set of bounds  $l$  and  $u$  such that:

$$\mathbb{P}[l \leq \theta \leq u | \vec{X}] \geq 1 - \alpha.$$

A similar interval can also be constructed for any component of a multivariate parameter  $\vec{\theta}$ . We can compute these values using R. A key difference from the frequentist confidence intervals is that the probabilistic interpretation of the interval is not lost even after we collect data.

**Conjugate Priors** In many cases, as we will see on today's worksheet, we can choose a prior distribution that aligns with the likelihood function such that the prior and the posterior come from the same family. For example, the Beta and the Binomial, as we saw on the last worksheet. The prior for a likelihood is called the **conjugate prior**.