

STAT 238 - Bayesian Statistics

Lecture Thirteen

Spring 2026, UC Berkeley

Aditya Guntuboyina

20 Feb 2026

1 Dirichlet Prior and Multinomial Likelihood

We saw the following fact in the last lecture:

$$(p_1, \dots, p_k) \sim \text{Dirichlet}(a_1, \dots, a_k) \text{ and } (X_1, \dots, X_k) \mid (p_1, \dots, p_k) \sim \text{Multinomial}(n; p_1, \dots, p_k) \\ \implies (p_1, \dots, p_k) \mid X_1 = x_1, \dots, X_k = x_k \sim \text{Dirichlet}(a_1 + x_1, a_2 + x_2, \dots, a_k + x_k).$$

Here the data are given by x_1, \dots, x_k where x_i represents count of the i -th outcome in n repeated trials (so $x_1 + \dots + x_k = n$). The probabilities p_1, \dots, p_k associated with k outcomes represent the unknown parameters which are to be estimated.

The posterior mean estimate of p_i is given by:

$$\mathbb{E}(p_i \mid X_1 = x_1, \dots, X_k = x_k) = \frac{a_i + x_i}{(a_1 + x_1) + \dots + (a_k + x_k)} = \frac{a_i + x_i}{n + a_1 + \dots + a_k}$$

where we used $x_1 + \dots + x_k = n$.

2 Special Case: the Dirichlet(0, ..., 0) prior

In the special case where all $a_i = 0$, then the posterior mean of p_i is simply x_i/n ; so we are estimating the probability of the i -th outcome by simply the proportion of times the i -th outcome appeared in the n trials. This is just the frequentist MLE. So the posterior mean estimate coincides with the frequentist MLE when the prior is Dirichlet(0, ..., 0).

More precisely, the posterior for the Dirichlet(0, ..., 0) prior is Dirichlet(x_1, \dots, x_k). A key property of this posterior is that if $x_i = 0$ for some i , then the posterior places all its mass on the set $\{p_i = 0\}$. In other words, any outcome that is not observed in the sample is assigned zero probability under the posterior and is therefore completely excluded from future consideration.

As an explicit example, suppose $k = 6$, $n = 4$, $x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 3, x_6 = 1$. In other words, we rolled a six-faced die 4 times; in 3 of the 4 rolls, we observed the outcome 5, and in the other roll, we observed a 6. After observing this data (note that we started with the prior Dirichlet(0, 0, 0, 0, 0, 0)), our posterior becomes Dirichlet(0, 0, 0, 0, 3, 1). This means that the posterior is concentrated on the probabilities $(0, 0, 0, 0, p_5, 1 - p_5)$ with $p_5 \sim \text{Beta}(3, 1)$.

3 Further special case: Dirichlet(0, ..., 0) prior and data in which no outcome is repeated

Suppose the actual outcomes observed in the n trials are y_1, \dots, y_n and assume that these are all distinct (which means that we did not observe any repeated outcome). As before, we work with the Dirichlet(0, ..., 0) prior. Then the posterior can be written as:

$$\sum_{i=1}^n w_i \delta_{\{y_i\}} \quad \text{where } (w_1, \dots, w_n) \sim \text{Dirichlet}(1, \dots, 1),$$

where $\delta_{\{y_i\}}$ denotes the point mass at the point y_i .

Note also that the density corresponding to the Dirichlet(1, ..., 1) distribution is constant, so the above posterior can be understood as the uniform distribution over all probability measures that are supported on the observed data points.

Posterior samples therefore can be generated by:

$$(w_1^{(j)}, \dots, w_n^{(j)}) \stackrel{\text{i.i.d.}}{\sim} \text{Dirichlet}(1, \dots, 1)$$

for $j = 1, \dots, M$ (for a large number M). Inference done with these samples is referred to as the **Bayesian Bootstrap** (see Rubin [1]). This is because the generation process for these samples is similar to the usual nonparametric bootstrap sample generation. Indeed, the usual bootstrap samples can be seen as being generated from:

$$\sum_{i=1}^n v_i \delta_{\{y_i\}} \quad \text{where } (v_1, \dots, v_n) \sim \frac{1}{n} \text{Multinomial}(n; 1/n, \dots, 1/n).$$

The distributions $(w_1, \dots, w_n) \sim \text{Dir}(1, \dots, 1)$ & $(v_1, \dots, v_n) \sim \text{Mult}(n; 1/n, \dots, 1/n)/n$ are quite close when n is large. For example,

$$\begin{aligned} \mathbb{E}w_i &= \mathbb{E}v_i = 1/n \\ \text{var}(w_i) &= \frac{n-1}{n^2(n+1)} \approx \text{var}(v_i) = \frac{n-1}{n^3} \\ \text{correlation}(w_i, w_j) &= \text{correlation}(v_i, v_j) = \frac{-1}{n-1} \quad \text{when } i \neq j. \end{aligned}$$

So operationally, bootstrap and the Bayesian bootstrap give very similar results, although conceptually they are quite different. For more information on the Bayesian bootstrap and its relation to the classical bootstrap, see the blog post by Rasmus Bååth: <https://www.sumsar.net/blog/2015/04/the-non-parametric-bootstrap-as-a-bayesian-model/>.

References

- [1] D. B. Rubin, "The Bayesian bootstrap," *The Annals of Statistics*, vol. 9, no. 1, pp. 130–134, 1981.