Lecture 6a: Dirichlet distribution; drawing from the posterior

Christopher A. Sims
Princeton University
sims@princeton.edu

September 15, 2020

Likelihood for population probabilities of bins

- In an example in the last lecture, we observed that, conditional on the sample and using a flat prior, the probability distribution of the population probability of a single bin with n observations and a sample size of N is $\mathrm{Beta}(n+1, N-n+1)$

- This result generalizes: If we have a set of k non-overlapping bins, their population probabiities given the sample are distributed as $\mathrm{Dirichlet}(n_1, \ldots, n_k, n_0)$, where n_i is the population count in bin i for $i > 0$ and $n_0 = N - \sum_1^k n_k$.

- The Dirichlet pdf, with parameter vector $\vec{\alpha}$ is proportional to

$$\prod_{i=1}^{k-1} p_i^{\alpha_i - 1} \cdot \Bigl(1 - \sum_{1}^{k-1} p_i\Bigr)^{\alpha_k}.$$

Note that, though it has a parameter vector of length k, it is a pdf only for $k-1$ p_i's. Or, you can think of it as a pdf for a k-dimensional $\vec{p}$ on the unit simplex, the set of $\vec{p}$'s satisfying $\sum p_i = 1$.

More on the Dirichlet

- The conditional distribution of any collection of elements of $\vec{p}$, divided by their sum, conditional on the remaining elements of $\vec{p}$, is again Dirichlet, with parameter vector given by the corresponding elements of the original $\vec{\alpha}$.

- Since this conditional distribution does not depend on the other elements of $\vec{p}$, the normalized subvector is independent of the remaining elements of $\vec{p}$. Therefore the marginal distribution of the normalized subvector is the same Dirichlet.

Using Monte Carlo draws from the posterior

- When the posterior density is complicated, or when we want to calculate the distribution or expectation of a function of the parameter that would require evaluating a nasty integral, it can be helpful to generate random draws from the posterior distribution.

- For example, suppose we have two bins A and B with modest and similar counts in the sample. A has more observations than B in the sample, but we're not sure whether that's strong evidence that the underlying population probability of A is bigger than that of B.

- The results above on the Dirichlet imply that the joint distribution of the two bin probabilities, p_A and p_B, normalized to sum to one, is.

under a flat prior, proportional to $p_A^{n_A} p_B^{n_B}$, which is proportional to a $\mathrm{Beta}(n_A + 1, n_B + 1)$ density on p_A.

- To estimate the probability, given the data, that $p_A > p_B$, we can make, say, 1000 draws from a $\mathrm{Beta}(n_A + 1, n_B + 1)$ distribution (e.g. using `rbeta()` in R) and count the proportion of the draws that have $p_A > .5$.

- Of course in this case we could also just look up the answer as `qbeta(.5, nA + 1, nB + 1)`, but on the exercise for this week you'll check a probability that's more difficult to calculate analytically.