Probability Distributions: Continuous

Data Science: Jordan Boyd-Graber
University of Maryland
FEBRUARY 11, 2018

Combining Discrete and Continuous Distributions

- We can chain together two distributions
- E.g., imagine your multinomial distribution came from a Dirichlet
- Often called "Bayesian Data Analysis"
- This why explain why "add one" Laplace smoothing isn't crazy

Why Bayesian

- Imagine you have vector of counts $\vec{n}$ that come from multinomial $\vec{\theta}$. This multinomial comes from a Dirichlet with parameter $\vec{\alpha}$. (Chain rule)

$$p(\vec{n}) = p(\vec{n} \,|\, \theta) p(\vec{\theta} \,|\, \vec{\alpha}) \tag{1}$$

- Now let's assume that you see some counts $\vec{n}$. You want to know what the multinomial distribution parameter looks like.

$$p(\vec{\theta} \,|\, \vec{n}, \vec{\alpha}) \tag{2}$$

Conjugacy

- If $\vec{\theta} \sim \mathrm{Dir}(()\alpha)$, $\vec{w} \sim \mathrm{Mult}(()\theta)$, and $n_k = |\{w_i : w_i = k\}|$ then

$$
\begin{aligned}
p(\theta|\alpha,\vec{w}) &\propto p(\vec{w}|\theta)p(\theta|\alpha) && (3)\\
&\propto \prod_k \theta^{n_k} \prod_k \theta^{\alpha_k - 1} && (4)\\
&\propto \prod_k \theta^{\alpha_k + n_k - 1} && (5)
\end{aligned}
$$

- Conjugacy: this **posterior** has the same form as the **prior**
- In fact, it looks like you're just adding counts!

Conjugacy

- If $\vec{\theta} \sim \text{Dir}(()\alpha)$, $\vec{w} \sim \text{Mult}(()\theta)$, and $n_k = |\{w_i : w_i = k\}|$ then

$$p(\theta|\alpha, \vec{w}) \propto p(\vec{w}|\theta)p(\theta|\alpha) \tag{3}$$

$$\propto \prod_k \theta^{n_k} \prod_k \theta^{\alpha_k - 1} \tag{4}$$

$$\propto \prod_k \theta^{\alpha_k + n_k - 1} \tag{5}$$

- Conjugacy: this **posterior** has the same form as the **prior**
- In fact, it looks like you're just adding counts!

Why add one?

- The count that we add is equivalent to the Dirichlet parameter
- What does this mean in the case of Dirichlet distribution?

$$f(\theta) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_i\right)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Why add one?

- The count that we add is equivalent to the Dirichlet parameter
- What does this mean in the case of Dirichlet distribution?

$$f(\theta) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_i\right)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

- Uniform distribution! Doesn't matter what x is.

Next time

- Drawing from and plotting various distributions
- Be sure to bring laptops