The Dirichlet Distributions

[*Prerequisite probability background:* Univariate gamma and beta distributions multivariate change of variables formulas, calculus of conditioning.]

For any positive integer k, and any $a_1 > 0, \ldots, a_k > 0$, the Dirichlet distribution $Dir(a_1, \ldots, a_k)$ denotes the distribution associated with the pdf

$$f(\omega) = \frac{1}{D(a_1, \ldots, a_k)} \omega_1^{a_1 - 1} \cdots \omega_k^{a_k - 1}, \quad \omega \in \Delta_k,$$

where Δ_k is the k-dimensional simplex: $\{\omega = (\omega_1, \ldots, \omega_k) \in \mathbb{R}^k : \omega_l \geq 0, \sum_{l=1}^k \omega_l = 1\}$. The normalizing constant, derived first by Dirichlet, equals $\frac{\Gamma(a_1)\cdots\Gamma(a_k)}{\Gamma(a_1+\cdots+a_k)}$. It is also possible to define the Dirichlet distribution through the pdf on the first $k-1$ variables. A $W = (W_1, \ldots, W_k) \in \Delta_k$ has the $Dir(a_1, \ldots, a_k)$ distribution if and only if the pdf of $(W_1, \ldots, W_{k-1})$ is proportional to $w_1^{a_1-1} \cdots w_{k-1}^{a_{k-1}-1}(1 - w_1 - \cdots - w_{k-1})^{a_k - 1}$.

Clearly, the Dirichlet distribution is an extension of the beta distribution to explain probabilities of two or more disjoint events. And in particular, $W = (W_1, W_2) \sim Dir(a, b)$ is same as saying $W_1 \sim Be(a, b)$, $W_2 = 1 - W_1$.

Below are some interesting connections with gamma and beta distributions, which lead to a better understanding of a Dirichlet random vector. Lemma 5 shows a very interesting "regenerative property".

Lemma 1. *If $X_l \sim Ga(a_l, 1)$, $l = 1, \ldots, k$, then $S := X_1 + \cdots + X_k \sim Ga(\sum_{l=1}^k a_l, 1)$, $W := (X_1/S, \ldots, X_k/S) \sim Dir(a_1, \ldots, a_k)$ and $S \perp\!\!\!\perp W$.*

Proof. Follows from a change of variable formula to the transformation $(X_1, \ldots, X_k) \mapsto (X_1, \ldots, X_{k-1}, S)$. □

Corollary 2. *If $(W_1, \ldots, W_k) \sim Dir(a_1, \ldots, a_k)$ and the indices $\{1, \ldots, k\}$ are partitioned into two disjoint subsets A and B then $(\sum_{l \in A} W_l, \sum_{l \in B} W_l) \sim Be(\sum_{l \in A} a_l, \sum_{l \in B} a_l)$.*

Proof. Follows directly from the gamma representation result of Lemma 1. □

Like beta is conjugate to binomial, the Dirichlet distributions are conjugate to the multinomial models:

Lemma 3. *If $Z|W \sim Mult(k, W)$, $W \sim Dir(a_1, \ldots, a_k)$ then $Z \sim Mult(k, (\frac{a_1}{\sum_l a_l}, \ldots, \frac{a_k}{\sum_l a_l}))$ and $W|Z \sim Dir(a_1 + I(Z = 1), \ldots, a_k + I(Z = k))$.*

Proof. The conditional distribution results follows easily since $f(w|z = j) \propto f(w)w_j$. For the marginal property, notice that $P(Z = j) = EP(Z = j|W) = EW_j = a_j / \sum_l a_l$ by Corollary 2. □

Lemma 4. *If $W \sim Dir(a_1, \ldots, a_k)$ and $Y \sim Be(b, a_1 + \cdots + a_k)$ then $((1 - Y)W, Y) \sim Dir(a_1, \ldots, a_k, b)$.*

Proof. By the previous lemma, we can identify $W = (X_1/S, \ldots, X_k/S)$, $Y = V/(V + S)$ where $X_l \sim Ga(a_l, 1)$, $S = X_1 + \cdots + X_k$ and $V \sim Ga(b, 1)$, independently of X. Hence $((1 - Y)W, Y) = (X_1/(V + S), \ldots, X_k/(V + S), V/(V + S)) \sim Dir(a_1, \ldots, a_k, b)$, again by Lemma 1. □

Lemma 5. *Let* $a_1, \ldots, a_k > 0$ *and* $a := a_1 + \cdots + a_k$. *If* $W \sim Dir(a_1, \ldots, a_k)$, $Z \sim Mult(k, (a_1, \ldots, a_k)/a)$ *and* $Y \sim Be(1, a)$, $W \perp\!\!\!\perp Y \perp\!\!\!\perp Z$, *then* $(1-Y)W + Ye_Z \sim Dir(a_1, \ldots, a_k)$ *where* $e_j \in \Delta_k$ *denotes the canonical vector with all zeros except a one on the* j*-th coordinate.*

Proof. By Lemma 4, $((1-Y)W, Y) \sim Dir(a_1, \ldots, a_k, 1)$. So by Corollary 2, given $Z = j$ the conditional distribution of $(1-Y)W + Ye_Z$ is $Dir(a_1, \ldots, a_{j-1}, a_j + 1, a_{j+1}, \ldots, a_k)$. Also, by definition $Z \sim Mult(k, (a_1/\sum_l a_l, \ldots, a_k/\sum_l a_l))$. Since the marginal of Z and the conditional of $\{(1-Y)W + Ye_Z\}|Z$ uniquely defines the distribution of W, we must have $(1-Y)W + Ye_Z \sim Dir(a_1, \ldots, a_k)$ by Lemma 3. $\square$