Derivation
of the Dirichlet-Categorical predictive posterior

Notation:

- D is the seen dataset, consisting of a sequence of items y.
- Each of the items y are identified by their category membership, so we can write $y = k \in K$.
- N_k is the number of (count) of items of category k in D.
- $[y = k]$ is equal to 1 if true (if $y = k$) and 0 otherwise. (This is an Iverson bracket.)
- Useful identity: $\Gamma(n+1) = n\Gamma(n)$. This follows from the definition of the Gamma function for integers: $\Gamma(n) = (n-1)!$.
- pdf: Probability Density Function. The pdf of a Dirichlet is: $f(x;\alpha) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k^K x_k^{\alpha_k - 1}$

We seek to find the predictive posterior, i.e., the probability of the next item y belonging to category j, given prior α and seen data D.

$$p(y = j \in K | D, \alpha) = \int_\theta P(y = j|\theta) P(\theta | D, \alpha) d\theta$$

We use the posterior Dirichlet ($\mathrm{Dir}(\alpha + N)$) pdf:

$$= \int_\theta \theta_j \prod_{k=1}^K \theta^{N_k} \frac{\Gamma\left(\sum_{k=1}^K N_k + \alpha_k\right)}{\prod_{k=1}^K \Gamma(N_k + \alpha_k)} \prod_{k=1}^K \theta_k^{N_k + \alpha_k - 1} d\theta$$

Rearranging terms; moving products into exponent:

$$= \frac{\Gamma\left(\sum_{k=1}^K N_k + \alpha_k\right)}{\prod_{k=1}^K \Gamma(N_k + \alpha_k)} \int_\theta \theta_j \prod_{k=1}^K \theta_k^{N_k + \alpha_k - 1} d\theta$$

Adding predicted count into exponent:

$$= \frac{\Gamma\left(\sum_{k=1}^K N_k + \alpha_k\right)}{\prod_{k=1}^K \Gamma(N_k + \alpha_k)} \int_\theta \prod_{k=1}^K \theta_k^{[j=k] + N_k + \alpha_k - 1} d\theta$$

The integral has a closed form corresponding to the Dirichlet pdf normalisation constant (Note it's the reciprocal here):

$$= \frac{\Gamma\left(\sum_{k=1}^K N_k + \alpha_k\right)}{\prod_{k=1}^K \Gamma(N_k + \alpha_k)} \frac{\prod_{k=1}^K \Gamma([j = k] + N_k + \alpha_k)}{\Gamma\left(1 + \sum_{k=1}^K + N_k + \alpha_k\right)}$$

Use $\Gamma(n+1) = n\Gamma(n)$:

$$= \frac{\Gamma\left(\sum_{k=1}^K N_k + \alpha_k\right)}{\prod_{k=1}^K \Gamma(N_k + \alpha_k)} \frac{(N_j + \alpha_j)\Gamma(N_j + \alpha_j) \prod_{k=1, k \neq j}^K \Gamma(N_k + \alpha_k)}{\sum_{k=1}^K N_k + \alpha_k \Gamma\left(\sum_{k=1}^K N_k + \alpha_k\right)}$$

Finding the solution from the lecture is left to the reader: see what cancels out.