Collapsed Inference Algorithms

- **If your priors are conjugate**, they can be integrated out!
- **Result:** a simpler distribution (less variables) from which to build inference methods

- Let $\sigma_{j,k} = \sum_{t=1}^{N_j} \mathbf{1}(z_{j,t} = k)$ be the number of times topic k is observed in document j
- Let $\delta_{i,r} = \sum_{j=1}^{M} \sum_{t=1}^{N_j} \mathbf{1}(w_{j,t} = r)$ be the number of times word type r is assigned topic k

- $P(\boldsymbol{W}, \boldsymbol{Z} \mid \alpha, \beta) = P(\boldsymbol{Z} \mid \alpha) P(\boldsymbol{W} \mid \boldsymbol{Z}, \beta)$

$$P(\boldsymbol{Z} \mid \alpha) = \int P(\Theta \mid \alpha) P(\boldsymbol{Z} \mid \Theta) d\Theta = \prod_{j=1}^{M} \int \frac{1}{B(\alpha)} \prod_{i=1}^{K} \theta_{j,i}^{\sigma_{j,i} + \alpha_i - 1} d\theta_j = \prod_{j=1}^{M} \frac{B(\alpha + \sigma_j)}{B(\alpha)}$$

$$P(\boldsymbol{W} \mid \boldsymbol{Z}, \beta) = \int P(\Phi \mid \beta) P(\boldsymbol{W} \mid \boldsymbol{Z}, \Phi) d\Phi = \prod_{i=1}^{K} \int \frac{1}{B(\beta)} \prod_{r=1}^{V} \phi_{i,r}^{\delta_{i,r} + \beta_r - 1} d\phi_i = \prod_{i=1}^{K} \frac{B(\beta + \delta_i)}{B(\beta)}$$

so then

$$P(\boldsymbol{W}, \boldsymbol{Z} \mid \alpha, \beta) = \prod_{j=1}^{M} \frac{B(\alpha + \sigma_j)}{B(\alpha)} \prod_{i=1}^{K} \frac{B(\beta + \delta_i)}{B(\beta)}$$

where $B(\cdot)$ is the multivariate Beta function.