If only $J' = 1$ component is left, by construction its summary will be very close to the summary for the target component k'. We shouldn't expect adding this new component will improve the original model (since k' already exists unchanged). Thus, if the resulting number of components is $J' = 1$, we abort the birth process early and return to the original K component model.

Creation of combined model Here, we combine the K components from the existing model with the brand-new J' components. Working purely in terms of sufficient statistics, we find that it is easy to build a coherent combined model simply by *concatenating* the fresh components $S' = [\hat{N}' \; s(\mathbf{x}')]$ onto the existing global sufficient statistics $S^0 = [\hat{N} \; s(\mathbf{x})]$.

We now have an expanded model with $K + J'$ summaries, $S^* = [\hat{N}^* s^*]$:

$$\hat{N}^* = [\hat{N}_1 \; \hat{N}_2 \; \cdots \; \hat{N}_K \;\; \hat{N}'_{K+1} \; \hat{N}'_{K+2} \; \cdots \; \hat{N}'_{K+J'}] \tag{26}$$

$$s^* = [s_1(\mathbf{x}) \; s_2(\mathbf{x}) \; \cdots \; s_K(\mathbf{x}) \;\; s_{K+1}(\mathbf{x}') \; s_{K+2}(\mathbf{x}') \; \cdots \; s_{K+J'}(\mathbf{x}')] \tag{27}$$

This concatenation creates a valid set of sufficient statistics for an "expanded" dataset formed by the union of $\mathbf{x}$ and $\mathbf{x}'$. This set "double-counts" the subsample $\mathbf{x}'$, assigning these data items to both original components (mostly k') and new components $K + 1, \ldots K + J$". In the next phase (adoption), we pass through the entire dataset, and discover which interpretation (original or new components) is preferred by the model.

Using new, expanded sufficient statistics S^*, we can then expand both local and global factors. The resulting expanded model q^* remains valid due to our *nested* truncation of the variational posterior. At this stage, no local parameters have been assigned to the new components. For all n, we simply expand $q^*(z_n)$ to be a discrete distribution over $K + J'$ components, where only the first K have mass:

$$\text{Before: } q(z_n) = \text{Cat}(\hat{r}_{n1}, \ldots \hat{r}_{nK}) \tag{28}$$

$$\text{After: } q^*(z_n) = \text{Cat}(\hat{r}_{n1}, \ldots \hat{r}_{nK}, 0, 0, \ldots 0) \tag{29}$$

Crucially, all parameters $\hat{r}_{nk}$ are directly transfered from the previous model q, and no batches actually need to be visited at this stage (they can instead be lazily expanded during each visit of the adoption pass). Another consequence of this construction is that $q^*(\phi_k) = q(\phi_k)$ for all original components $k = 1, 2, \ldots K$, including the target component k'. For new components j, $q^*(\phi_j)$ are set to the resulting factors from the targeted analysis.

Only the stick-breaking factors $q^*(v)$ must be completely re-written after the expansion. Expansion forces these factors to shift probability mass onto newly inserted components. Given the counts from all $K + J'$ summaries S^*, the update equations become

$$q^*(v_k)|S^* = \text{Beta}(v_k|\alpha^*_{k1}, \alpha^*_{k0}), \qquad \hat{\alpha}^*_{k1} = 1 + \hat{N}_k, \quad \hat{\alpha}^*_{k0} = \alpha_0 + \sum_{\ell=k+1}^{K+J'} \hat{N}_k \tag{30}$$

The choice to insert new components last in the stick-breaking order (which is implicitly done by concatenation) is fairly principled. On average, freshly discovered components will be more "rare" than the original ones, and so will likely have smaller effective mass. Since the stick-breaking construction is "size-biased", inserting components with smaller mass later in the order makes sense.

2.3 Adoption of the new components

After expanding the model to have $K + J'$ components, we then proceed normally through the memoized variational inference E-step (local factors) and M-step (global factors) at each batch. Our goal in this pass is to have newborn components become "adopted" by the original dataset $\mathbf{x}$, attaining critical mass by actively explaining some data in $\mathbf{x}$.

At the start of this pass, we have the expanded set of global sufficient statistics S^* described earlier. Retaining the target summaries S' as well as the previous global summaries S^0 within S^* allows each brand-new component a chance to influence several batches of data.

To understand this necessity, consider the alternative: after creating an expanded model, we discard S' and keep only the original summaries S^0. To be consistent with local assignments, we must