Notes

WAIC: (widely available information criterion)
Bayesian approach for estimating out of sample expectation. Starts with the computed log pointwise posterior predictive density (lppd) and then adds a correction for effective number of parameters to adjust for overfitting.
Approach 1:

$$
\begin{aligned}
p_{WAIC1} &= 2\sum(log(E_{post}p(y_i|\theta)) - E_{post}(log(p(y_i|\theta)) \\
&= 2\sum(log(\frac{1}{S}\sum_{s=1}^{S}p(y_i|\theta^s)) - \frac{1}{S}\sum_{s=1}^{S}log(p(y_i|\theta^s))
\end{aligned}
$$

Approach 2: (variance of individual terms in the log predictive density summed over the n data points)

$$
p_{WAIC2} = \sum_{i=1}^{n} V_{s=1}^{S}(logp(y_i|\theta^s))
$$

$$
where V_{s=1}^{S}\alpha_s = \frac{1}{S-1}\sum_{s=1}^{S}(\alpha_s - \bar{\alpha})^2
$$

The second approach is more stable since it computes the variance separately for each point and then sums, which gives more stability.
WAIC evaluates the predictions that are actually being used for new data in a Bayesian context.
A cost of using WAIC is that it relies on a partitions of the data into n pieces which is not so easy to do in some structured data settings (such as times series, spatial and netword data).

Let y be the observed data and θ be the vector of parameters. y^{rep} is defined as the replicated data that could have been observed (if the experiment that produced y today was replicated with the same model and the same value of θ that produced the observed data).
$\tilde{y}$ is any future observable value or vector of observable quantities
y^{rep} is a replication like y
Posterior predictive distribution: $p(y^{rep}|y) = p(y^{rep}|\theta)p(\theta|y)d\theta$
The basic method of inference from iterative simulation is the same as for Bayesian simulation in general: use the collection of all the simulated draws from $p(\theta|y)$ to summarize the posterior density and to compute quantiles,

moments and other summaries of interest as needed. Posterior predictive simulations of unobserved outcomes $\tilde{y}$ can be obtained by simulation conditional on the drawn values of θ.
Source: Bayesian Data Analysis, Gelman, Carlin, Stern...

Calculations for different priors for the cluster assignments:
In general:

$$pr(\xi_v = h|Y, \{\eta_{k,l}\}, \{\xi_{-v}\}) = pr(\xi_v = h|\{\eta_{k,l}\}, \{\xi_{-v}\})\frac{P(Y|\{\eta_{k,l}\}, \{\xi_{-v}\}, \xi_v = h)}{P(Y|\{\eta_{k,l}\}, \{\xi_{-v}\}})$$

where $pr(\xi_v = h|\{\eta_{k,l}\}, \{\xi_{-v}\}) = pr(\xi_v = h|\{\xi_{-v}\})$ since $p(\{\eta_{k,l}\})$ doesn't depend on ξ_i (from Baye's theorem).

The second part of the right hand side of the equation doesn't depend on the distribution of the cluster assignments and equals to:

$$\frac{P(Y|\{\eta_{k,l}\}, \{\xi_{-v}\}, \xi_v = h)}{P(Y|\{\eta_{k,l}\}, \{\xi_{-v}\}} =$$

$$\frac{\prod_{i=1,\ldots,v-1}^{v+1,\ldots,I} \prod_{i'=i+1}^{I} (expit\eta_{\phi_{(\xi_i,\xi_{i'})}})^{y_{i,i'}}(1 - expit\eta_{\phi_{(\xi_i,\xi_{i'})}})^{(1-y_{i,i'})}}{\prod_{i=1,\ldots,v-1}^{v+1,\ldots,I} \prod_{i'=i+1,i'\neq v}^{I} (expit\eta_{\phi_{(\xi_i,\xi_{i'})}})^{y_{i,i'}}(1 - expit\eta_{\phi_{(\xi_i,\xi_{i'})}})^{(1-y_{i,i'})}} *$$

$$* \prod_{i'=v+1}^{I} (expit\eta_{\phi_{(h,\xi_{i'})}})^{y_{v,i'}}(1-expit\eta_{\phi_{(h,\xi_{i'})}})^{(1-y_{v,i'})} \prod_{i=1}^{v-1}(expit\eta_{\phi_{(\xi_i,h)}})^{y_{i,v}}(1-expit\eta_{\phi_{(\xi_i,h)}})^{(1-y_{i,v})} =$$

$$= \prod_{i'=v+1}^{I} (expit\eta_{\phi_{(h,\xi_{i'})}})^{y_{v,i'}}(1-expit\eta_{\phi_{(h,\xi_{i'})}})^{(1-y_{v,i'})} \prod_{i=1}^{v-1}(expit\eta_{\phi_{(\xi_i,h)}})^{y_{i,v}}(1-expit\eta_{\phi_{(\xi_i,h)}})^{(1-y_{i,v})}$$

The first part of the right hand side of the equation is the term that depends on the distribution of the cluster assignments.

For $\xi_i \sim DP(\alpha)$:

$$p(k,n) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)}\alpha^H \prod_{j=1}^{H} \Gamma(n_j)$$

later we will put a prior to α.
If $h = 1, ..., H$ (existing cluster) then:

$$pr(z_{v+1} = h|z) = \frac{pr(z_{v+1} = h \cap z)}{pr(z)} = \frac{\frac{\Gamma(\alpha)}{\Gamma(\alpha+n+1)}\alpha^H \prod_{j=1}^{H} \Gamma(n_j)}{\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)}\alpha^H \prod_{j=1}^{H} \Gamma(n_j)}$$

where in the top product $n'_h = n_h + 1$.
Hence

$$pr(z_{v+1} = h|z) = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha+n+1)}\frac{\Gamma(n_h+1)}{\Gamma(n_h)} = \frac{n_h}{\alpha+n}$$

Else if $h = H + 1$ (new cluster) then:

$$pr(z_{v+1} = h|z) = \frac{pr(z_{v+1} = h \cap z)}{pr(z)} = \frac{\frac{\Gamma(\alpha)}{\Gamma(\alpha+n+1)}\alpha^{H+1} \prod_{j=1}^{H+1} \Gamma(n_j)}{\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)}\alpha^H \prod_{j=1}^{H} \Gamma(n_j)} =$$

$$= \frac{\Gamma(\alpha+n)}{\Gamma(\alpha+n+1)}\alpha\Gamma(n_H) = \frac{\alpha}{\alpha+n}$$

To summarize:

$$pr(z_{v+1} = h|z) \propto \begin{cases} n_h, & \text{if h = 1,...,H} \\ \alpha, & \text{if h = H+1} \end{cases}$$

and

$$pr(\xi_v = h|Y, \{\eta_{k,l}\}, \{\xi_{-v}\}) \propto \begin{cases} n_h, & \text{if h = 1,...,H} \\ \alpha, & \text{if h = H+1} \end{cases}$$

For $\xi_i \sim PY(\alpha, \sigma)$:

$$p(k, n) = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+n)} \prod_{i=1}^{H} \frac{\Gamma(1-\sigma)}{\Gamma(n_i-\sigma)} \prod_{j=1}^{H-1} (\alpha + j\sigma)$$

If $h = 1, ..., H$ (existing cluster) then:

$$pr(z_{v+1} = h|z) = \frac{pr(z_{v+1} = h, z)}{pr(z)} = \frac{\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+n+1)} \prod_{i=1}^{H} \frac{\Gamma(1-\sigma)}{\Gamma(n_i-\sigma)} \prod_{j=1}^{H-1} (\alpha + j\sigma)}{\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+n)} \prod_{i=1}^{H} \frac{\Gamma(1-\sigma)}{\Gamma(n_i-\sigma)} \prod_{j=1}^{H-1} (\alpha + j\sigma)}$$

where in the top product $n'_h = n_h + 1$.
Hence

$$pr(z_{v+1} = h|z) = \frac{1}{\alpha+n} \frac{\Gamma(n_h - \sigma + 1)}{\Gamma(n_h - \sigma)} = \frac{n_h - \sigma}{\alpha + n}$$

Else if $h = H + 1$ (new cluster) then:

$$pr(z_{v+1} = h|z) = \frac{pr(z_{v+1} = h, z)}{pr(z)} = \frac{\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+n+1)} \prod_{i=1}^{H+1} \frac{\Gamma(1-\sigma)}{\Gamma(n_i-\sigma)} \prod_{j=1}^{H} (\alpha + j\sigma)}{\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+n)} \prod_{i=1}^{H} \frac{\Gamma(1-\sigma)}{\Gamma(n_i-\sigma)} \prod_{j=1}^{H-1} (\alpha + j\sigma)}$$

$$= \frac{1}{\Gamma(\alpha+n)} \frac{\Gamma(1-\sigma)}{\Gamma(1-\sigma)} (\alpha + H\sigma) = \frac{\alpha + H\sigma}{\alpha + n}$$

since $n_{H+1} = 1$ only the new node.
To summarize:

$$pr(z_{v+1} = h|z) \propto \begin{cases} n_h - \sigma, & \text{if h = 1,...,H} \\ \alpha + H\sigma, & \text{if h = H+1} \end{cases}$$

and

$$pr(\xi_v = h|Y, \{\eta_{k,l}\}, \{\xi_{-v}\}) \propto \begin{cases} n_h - `\sigma, & \text{if h = 1,...,H} \\ \alpha + H\sigma, & \text{if h = H+1} \end{cases}$$