A. Synthetic Experiments

In this section we provide synthetic experiments to demonstrate that (11) is a good approximation to (7) when the variational Dirichlet posterior has a single mode, which is easily achieved in practice when $\boldsymbol{\beta} > 1$. We use the SPN shown in Fig. 3 as a

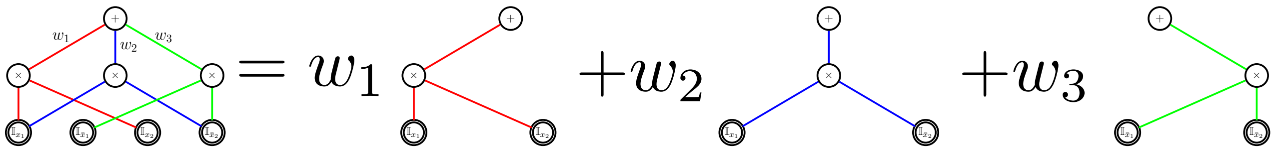
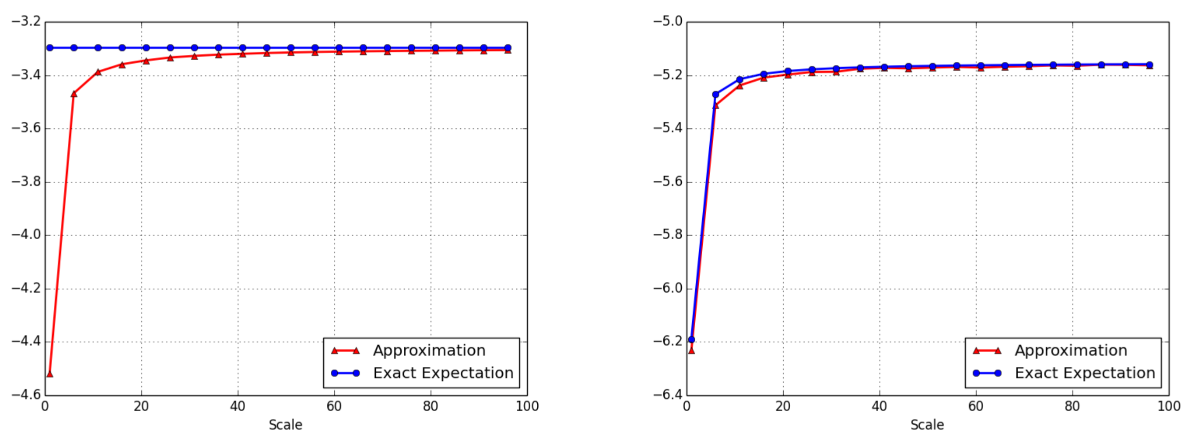


Figure 3. A simple SPN over two binary variables X_1 and X_2 that can be decomposed into three components.

synthetic example where each leaf univariate distribution is fixed to be an indicator variable, i.e., a point mass distribution. The joint distribution is given by $p(\mathbf{x}|\mathbf{w}) = w_1\mathbb{I}_{x_1}\mathbb{I}_{x_2} + w_2\mathbb{I}_{x_1}\mathbb{I}_{\bar{x}_2} + w_3\mathbb{I}_{\bar{x}_1}\mathbb{I}_{\bar{x}_2}$ with $w_1 + w_2 + w_3 = 1$.

Since the KL divergence $\mathbb{KL}(q(\mathbf{w}|\boldsymbol{\beta}) \parallel p(\mathbf{w}|\boldsymbol{\alpha}))$ is the same in both (7) and (11), to show the difference, we only need to compare $\mathbb{E}_{q(\mathbf{w}|\boldsymbol{\beta})}[\log p(\mathbf{x}|\mathbf{w})]$ and $\log p(\mathbf{x}|\exp(\mathbb{E}_{q'(\mathbf{w}'|\boldsymbol{\beta})}[\mathbf{w}']))$. Recall that we have $\mathbf{w}' = \log \mathbf{w}$ by definition.

We use a data set $\{(1,1),(1,0),(0,0)\}$ that contains all the instances of $\mathbf{x}$ in the support of the distribution. To show the approximation accuracy of (11), we test with both symmetric and skewed Dirichlet distributions $q(\mathbf{w}|\boldsymbol{\beta})$. To be more specific, we use $\boldsymbol{\beta}_1 = [1.0, 1.0, 1.0]^T$ and $\boldsymbol{\beta}_2 = [10.0, 1.0, 1.0]^T$ as the hyperparameters of the variational posterior and compute both $\mathbb{E}_{q(\mathbf{w}|\boldsymbol{\beta})}[\log p(\mathbf{x}|\mathbf{w})]$ and $\log p(\mathbf{x}|\exp(\mathbb{E}_{q'(\mathbf{w}'|\boldsymbol{\beta})}[\mathbf{w}']))$ as a function of $\boldsymbol{\beta}$. To demonstrate the quality of the approximation and the scale factor of the Dirichlet, for each $\boldsymbol{\beta}$, we fixed the shape of the Dirichlet distribution but gradually increase its scale factor, i.e., $\boldsymbol{\beta}' = s\boldsymbol{\beta}$, where s ranges from 1 to 100. $\mathbb{E}_{q(\mathbf{w}|\boldsymbol{\beta})}[\log p(\mathbf{x}|\mathbf{w})]$ does not admit a closed form computation, hence we sample 10000 weight vectors from $q(\mathbf{w}|\boldsymbol{\beta})$ and use the empirical average to approximate the expectation.



(a) Symmetric Dirichlet. (b) Skewed Dirichlet.

Figure 4. Synthetic experiments to compare the approximation of $\log p(\mathbf{x}|\exp(\mathbb{E}_{q'(\mathbf{w}'|\boldsymbol{\beta})}[\mathbf{w}']))$ (red, lower bound) to $\mathbb{E}_{q(\mathbf{w}|\boldsymbol{\beta})}[\log p(\mathbf{x}|\mathbf{w})]$ (blue, exact expectation). Fig. 4(a) corresponds to symmetric $\boldsymbol{\beta}_1 = [1.0, 1.0, 1.0]^T$ and Fig. 4(b) corresponds to skewed $\boldsymbol{\beta} = [10.0, 1.0, 1.0]^T$.

It can be observed from Fig. 4 that the approximation is actually very close to the true exact expectation, especially when the Dirichlet distribution has a high concentration around its single mode, which corresponds to a large scale factor. This holds no matter whether the variational posterior is symmetric or not.