A Appendix A: The variational algorithm

Given an observed matrix of user behavior y, we would like to compute the posterior distribution of user preferences θ_{uk}, item attributes β_{ik}, user activity ξ_u and item popularity η_i, $p(\theta, \beta, \xi, \eta \,|\, y)$. Our derivation of the variational algorithm for HPF makes use of general results about the class of *conditionally conjugate* models [11, 16]. We define the class, show that HPF is in the class, and then derive the variational inference algorithm.

Complete conditionals. Variational inference fits the variational parameters to minimize their KL divergence to the posterior. For the large class of conditionally conjugate models, we can easily perform this optimization with a coordinate-ascent algorithm, one in which we iteratively optimize each variational parameter while holding the others fixed. A *complete conditional* is the conditional distribution of a latent variable given the observations and the other latent variables in the model. A conditionally conjugate model is one where each complete conditional is in an exponential family.

HPF, with the z_{ui} variables described in Section 3.2, is a conditionally conjugate model. (Without the auxiliary variables, it is not conditionally conjugate.) For the user weights θ_{uk}, the complete conditional is a Gamma,

$$\theta_{uk} \,|\, \beta, \xi, z, y \sim \text{Gamma}(a + \textstyle\sum_i z_{uik}, \xi_u + \sum_i \beta_{ik}). \quad (4)$$

The complete conditional for item weights β_{ik} is symmetric,

$$\beta_{ik} \,|\, \theta, \eta, z, y \sim \text{Gamma}(a + \textstyle\sum_u z_{uik}, \eta_i + \sum_i \theta_{uk}). \quad (5)$$

These distributions stem from conjugacy properties between the Gamma and Poisson. In the user weight distribution, for example, the item weights β_{ik} act as "exposure" variables [9]. (The roles are reversed in the item weight distribution.) We can similarly write down the complete conditionals for the user activity ξ_u and the item popularity η_i.

$$\xi_u \,|\, \theta \sim \text{Gamma}(a' + Ka, b' + \textstyle\sum_k \theta_{uk}).$$
$$\eta_i \,|\, \beta \sim \text{Gamma}(c' + Kc, d' + \textstyle\sum_k \beta_{ik}).$$

The final latent variables are the auxiliary variables. Recall that each z_{ui} is a K-vector of Poisson counts that sum to the observation y_{ui}. The complete conditional for this vector is

$$z_{ui} \,|\, \beta, \theta, y \sim \text{Mult}\left(y_{ui}, \frac{\theta_u \beta_i}{\sum_k \theta_{uk}\beta_{ik}}\right). \quad (6)$$

Though these variables are Poisson in the model, their complete conditional is multinomial. The reason is that the conditional distribution of a set of Poisson variables, given their sum, is a multinomial for which the parameter is their normalized set of rates. (See [20, 5].)

Deriving the algorithm. We now derive variational inference for HPF. First, we set each factor in the mean-field family (Equation 4) to be the same type of distribution as its complete conditional. The complete conditionals for the item weights β_{ik} and user weights θ_{uk} are Gamma distributions (Equations 4 and 5); thus the variational parameters λ_{ik} and γ_{uk} are Gamma parameters, each containing a shape and a rate. Similarly, the variational user activity parameters κ_u and the variational item popularity parameter τ_i are Gamma parameters, each containing a shape and a rate. The complete conditional of the auxiliary variables z_{uik} is a multinomial (Equation 6); thus the variational parameter ϕ_{ui} is a multinomial parameter, a point on the K-simplex, and the variational distribution for z_{ui} is $\text{Mult}(y_{ui}, \phi_{ui})$.

In coordinate ascent we iteratively optimize each variational parameter while holding the others fixed. In conditionally conjugate models, this amounts to setting each variational parameter equal to the expected parameter (under q) of the complete conditional. [4] The parameter to each complete conditional is a function of the other latent variables and the mean-field family sets all the variables to be independent. These facts guarantee that the parameter we are optimizing will not appear in the expected parameter.

For the user and item weights, we update the variational shape and rate parameters. The updates are

$$\gamma_{uk} = \langle a + \textstyle\sum_i y_{ui}\phi_{uik}, b + \sum_i \lambda_{ik}^{\text{shp}} / \lambda_{ik}^{\text{rte}} \rangle \quad (7)$$
$$\lambda_{ik} = \langle c + \textstyle\sum_u y_{ui}\phi_{uik}, d + \sum_u \gamma_{ik}^{\text{shp}} / \gamma_{ik}^{\text{rte}} \rangle. \quad (8)$$

These are expectations of the complete conditionals in Equations 4 and 5. In the shape parameter, we use that the expected count of the kth item in the multinomial is $\text{E}_q[z_{uik}] = y_{ui}\phi_{uik}$. In the rate parameter, we use that the expectation of a Gamma variable is the shape divided by the rate.

For the variational multinomial the update is

$$\phi_{ui} \propto \exp\{\Psi(\gamma_{uk}^{\text{shp}}) - \log \gamma_{uk}^{\text{rte}} + \Psi(\lambda_{ik}^{\text{shp}}) - \log \lambda_{ik}^{\text{rte}}\}, \quad (9)$$

where $\Psi(\cdot)$ is the digamma function (the first derivative of the log Γ function). This update comes from the expectation of the log of a Gamma variable, for example $\text{E}_q[\log \theta_{nk}] = \Psi(\gamma_{nk}^{\text{shp}}) - \log \gamma_{nk}^{\text{rte}}$.

[4] It is a little more complex then this. For details, see [16].