We only have to compute entropies over multinomial distributions.

**HOMF− NoB**

**Summary**

**The Task**

In collaborative filtering it is hard to make recommendations with new users or items for which we have no data. This is the cold-start problem. To alleviate this problem, we elicit the most useful initial ratings.

**Challenges**

- **No data** available for new users or items: parameter uncertainty is high.
- Rating matrices have variable noise levels across users and items.
- Typically, we have a small budget for collecting ratings.

**We propose**

- A new Heteroskedastic Ordinal Matrix Factorization model (HOMF) that captures variable noise levels across users and items.
- An efficient active learning strategy (BALD) that distinguishes between model uncertainty and the intrinsic noiseness of the data.

**We show that**

- HOMF yields state-of-the-art predictive performance.
- BALD significantly reduces the amount of data needed for learning.
- Bayesian active learning works best with heteroskedastic noise models.

**Heteroskedastic Ordinal Matrix Factorization**

**Key Components:**

- An ordinal regression likelihood, $p(r_{ij}|a_i, b_j) = \prod_{k=1}^{K} \text{sign}(r_{ij} - k - 0.5 | a_i - b_j)$.
- User and item specific noise levels: heteroskedasticity, $p(\alpha_i, \beta_j, \gamma_{ij}, \gamma_{i'}, \gamma_{j'}) = \mathcal{N}(\alpha_i, \beta_j, \gamma_{ij}, \gamma_{i'}, \gamma_{j'})$.
- We learn the boundary variables in the ordinal likelihood.
- We use hyper-priors on all parameters $\Theta$ for robustness to parameter fixing.

**Generative Process**

**Bayesian Active Learning by Disagreement (BALD)**

The symmetry in mutual information allows us to rearrange Equation (1) to obtain

$$\arg\max_{r_{ij}} \text{H}(p(\Theta | r_{ij})| - \text{E}_{\Theta \sim p(\Theta)} \text{H}(p(\Theta | r_{ij}))).$$

However, this is infeasible in practice because

- The posterior has to be updated for every possible value of $r_{ij}$.
- We must compute intractable entropies: $\text{H}(p(\Theta | R^C))$ and $\text{H}(p(\Theta | r_{ij}, R^C))$.

**Predictive Performance of our Heteroskedastic Model**

**Cold-start Experimental Setup**

**Value in the First Active Sample**

**Heteroskedastic vs. Constant Noise**

Bayesian active learning improves substantially with an accurate noise model.