Latent variables

- What to do when a variable z is unobserved?
- If we never condition on z when in the inference problem, then we can just integrate it out.
- However, in certain cases, we are interested in the latent variables themselves, e.g. the clustering problems.
- More on latent variables when we cover Gaussian mixtures.

The TrueSkill latent variable model

- TrueSkill model is a player ranking system for competitive games.
- The goal is to infer the skill of a set of players in a competitive game, based on observing who beats who.
- In the TrueSkill model, each player has a fixed level of skill, denoted z_i.
- We initially don't know anything about anyone's skill, but we assume everyone's skill is independent (e.g. an independent Gaussian prior).
- We never get to observe the players' skills directly, which makes this a latent variable model.

TrueSkill model

- Instead, we observe the outcome of a series of matches between different players.
- For each game, the probability that player i beats player j is given by

$$p(i \text{ beats } j | z_i, z_j) = \sigma(z_i - z_j)$$

  where sigma is the logistic function: $\sigma(y) = \frac{1}{1+\exp(-y)}$.
- We can write the entire joint likelihood of a set of players and games as:

$$p(z_1, z_2, \ldots z_N, \text{game 1, game 2, .. game T})$$
$$= \left[\prod_{i=1}^{N} p(z_i)\right] \left[\prod_{\text{games}} p(\text{i beats j}|z_i, z_j)\right]$$

Posterior

- Given the outcome of some matches, the players' skills are no longer independent, even if they've never played each other.
- Computing the posterior over even two players' skills requires integrating over all the other players' skills:

$$p(z_1, z_2|\text{game 1, game 2, ... game T})$$
$$= \int \cdots \int p(z_1, z_2, z_3 \ldots z_N|x) dz_3 \ldots dz_N$$

- **Message passing** can be used to compute posteriors!
- More on this model in Assignment 2.