Lost Relatives of the Gumbel Trick

Matej Balog1,2, Nilesh Tripuraneni3, Zoubin Ghahramani1,4, Adrian Weller1,5
1University of Cambridge 2MPI-IS Tübingen 3UC Berkeley 4Uber AI Labs 5Alan Turing Institute

Gumbel trick

Given:
\[ \hat{p} = (\hat{p}_1, \hat{p}_2, ..., \hat{p}_N) = (e^{\theta_1}, e^{\theta_2}, ..., e^{\theta_N}) \]

Trick to sample from \( p \) and estimate \( Z = \sum_i \hat{p}_i \):

- Perturb log-probs \( \theta_i \) with Gumbel distribution and find MAP:
  \[ \theta_i = \text{log}(\hat{p}_i) \sim \text{Gumbel}(\ln Z) \]

What’s new?

- New, related tricks that yield lower MSE estimators of \( Z \)
- New bounds on \( Z \) in discrete graphical models with negligible additional computational cost.

Competing exponential clocks

\[ \text{~Exp}(\lambda_i) \]

- \( \text{Pr}(\text{clock } i \text{ rings first}) \propto \lambda_i \)
- Time of first ring:
  \[ \min T_i \sim \text{Exp} \left( \sum_i \lambda_i \right) \]

\[ \text{Gumbel} \] trick: apply \( g(x) = -\ln x - c \) to the clocks

- \( \text{~Exp}(\lambda_i) \)
- \( \text{~Gumbel}(0) \)

\[ \text{~Exp}(\lambda_i) \] trick bounds:

\[ U := \max_{x \in \mathbb{R}} \left\{ \theta(x) + \sum_{i=1}^n \gamma_i(x) \right\} \]

\[ L := \max_{x \in \mathbb{R}} \left\{ \theta(x) + \frac{1}{n} \sum_{i=1}^n \gamma_i(x) \right\} \]

New bounds from \text{Weibull} and \text{Fréchet} tricks

For any \( \alpha \in (-1, 0) \cup (0, \infty) \):

\[ \ln Z \leq \frac{\ln \Gamma(1 + \alpha)}{\alpha} + nc - \frac{1}{\alpha} \ln \mathbb{E}_{\gamma} \left[ e^{-\alpha L} \right] \]

\[ \ln Z \geq c + \frac{\ln \Gamma(1 + \alpha)}{\alpha} - \frac{1}{\alpha n} \ln \mathbb{E} \left[ \exp \left(-\alpha L \right) \right] \]

Takeaway:

- Save up to 40% samples by using the Exponential trick instead of the Gumbel trick. Applications:
  - \( A^* \) sampling of Maddison et al. [2014]
  - Large-scale sampling of Chen and Ghahramani [2016]

Gumbel trick bounds:

\[ U := \max_{x \in \mathbb{R}} \left\{ \theta(x) + \sum_{i=1}^n \gamma_i(x) \right\} \]

\[ L := \max_{x \in \mathbb{R}} \left\{ \theta(x) + \frac{1}{n} \sum_{i=1}^n \gamma_i(x) \right\} \]

\[ \mathbb{E}[L] \leq \ln Z \leq \mathbb{E}[U] \]

Full-rank perturbations

- \( \text{Exp} \) trick bounds:
  - \( M = 1 \) to 6: \( L \leq U \)
  - \( M = 1 \) to 128: \( L \leq U \)

- Gumbel trick bounds:
  - \( M = 1 \) to 6: \( \text{Exp} \) trick bounds
  - \( M = 1 \) to 128: \( \text{Gumbel} \) trick bounds

Low-rank perturbations (discrete graphical models):

A* sampling:

MAB approximate MAP:

Points:

- Contact: matej.balog@gmail.com
- Slides: http://matejbalog.eu/en/research/
- Code: https://github.com/matejbalog/gumbel-relatives