Lost Relatives of the Gumbel Trick

Matej Balog[1,2], Nilesh Tripuraneni[3], Zoubin Ghahramani[1,4], Adrian Weller[1,5]

[1]University of Cambridge [2]MPI-IS Tübingen [3]UC Berkeley [4]Uber AI Labs [5]Alan Turing Institute

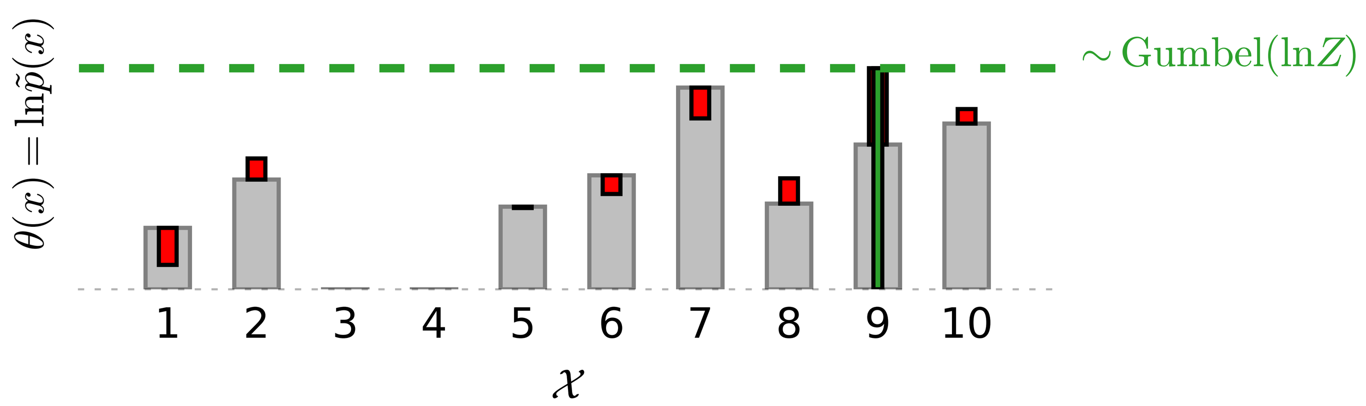
Gumbel trick

Given:

$$\tilde{\boldsymbol{p}} = (\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_N) = (e^{\theta_1}, e^{\theta_2}, \dots, e^{\theta_N})$$

Trick to sample from $\boldsymbol{p}$ and estimate $Z = \sum_i \tilde{p}_i$:

Perturb log-probs θ_i with Gumbel distribution and find MAP:



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

$\sim\mathrm{Exp}(\lambda_1)$

$\sim\mathrm{Exp}(\lambda_2)$

…

$\sim\mathrm{Exp}(\lambda_N)$

(1) $\Pr(\text{clock } i \text{ rings first}) \propto \lambda_i$

(2) Time of first ring:

$$\min_i T_i \sim \mathrm{Exp}\left(\sum_i \lambda_i\right)$$

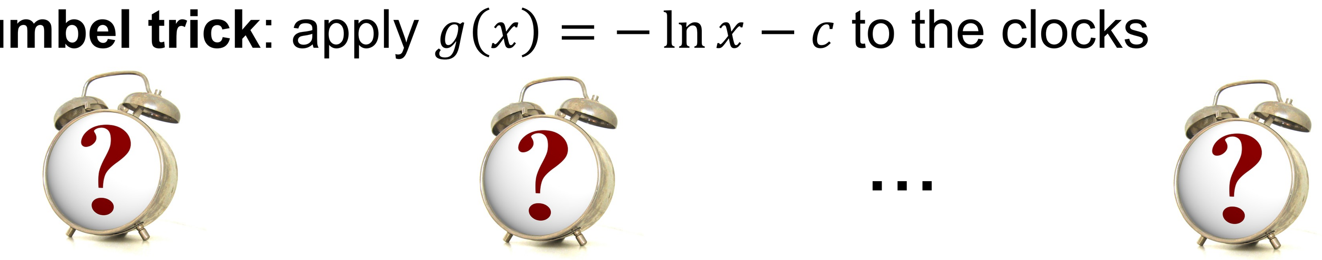
Taking $\lambda_i = \tilde{p}_i$:

(1) $\Pr(\text{clock } i \text{ rings first}) = p_i$

(2) Time of first ring:

$$\min_i T_i \sim \mathrm{Exp}(Z)$$

Gumbel trick: apply $g(x) = -\ln x - c$ to the clocks



$\sim\theta_1 + \mathrm{Gumbel}(0)$ $\quad\sim\theta_2 + \mathrm{Gumbel}(0)$ $\quad\dots\quad$ $\sim\theta_N + \mathrm{Gumbel}(0)$

(1) $\Pr(\text{maximum at } i) = p_i$

(2) Distribution of maximum value:

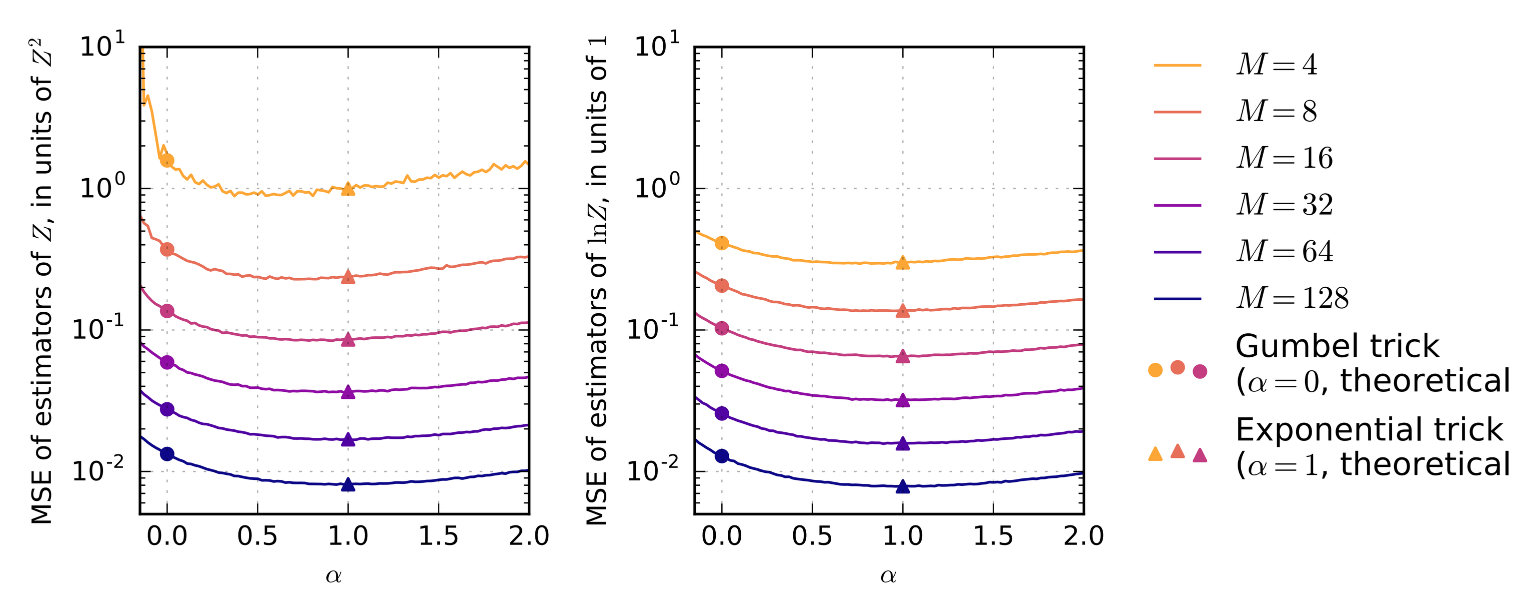
$$-\ln\left(\min_i T_i\right) - c \sim \ln Z + \mathrm{Gumbel}(0)$$

Relatives of the Gumbel trick

Obtain them by applying different functions g to the clocks.

Trick	$g(x)$	Mean $f(Z)$	Asymptotic var. of $\hat{Z}$
Gumbel	$-\ln x - c$	$\ln Z$	$\frac{\pi^2}{6} Z^2$
Exponential	x	$\frac{1}{Z}$	Z^2
Weibull α	$x^\alpha,\ \alpha > 0$	$Z^{-\alpha}\Gamma(1+\alpha)$	$\frac{1}{\alpha^2}\left(\frac{\Gamma(1+2\alpha)}{\Gamma(1+\alpha)^2} - 1\right) Z^2$
Fréchet α	$x^\alpha,\ \alpha \in (-1, 0)$	$Z^{-\alpha}\Gamma(1+\alpha)$	$\frac{1}{\alpha^2}\left(\frac{\Gamma(1+2\alpha)}{\Gamma(1+\alpha)^2} - 1\right) Z^2$
Pareto	e^x	$\frac{Z}{Z-1}$ for $Z > 1$	$\frac{Z^2}{(Z-2)^2}$
Tail t	$\mathbb{1}_{\{x>t\}}$	e^{-tZ}	$\frac{(1-e^{-tZ})^2}{t^2}$

Full-rank perturbations

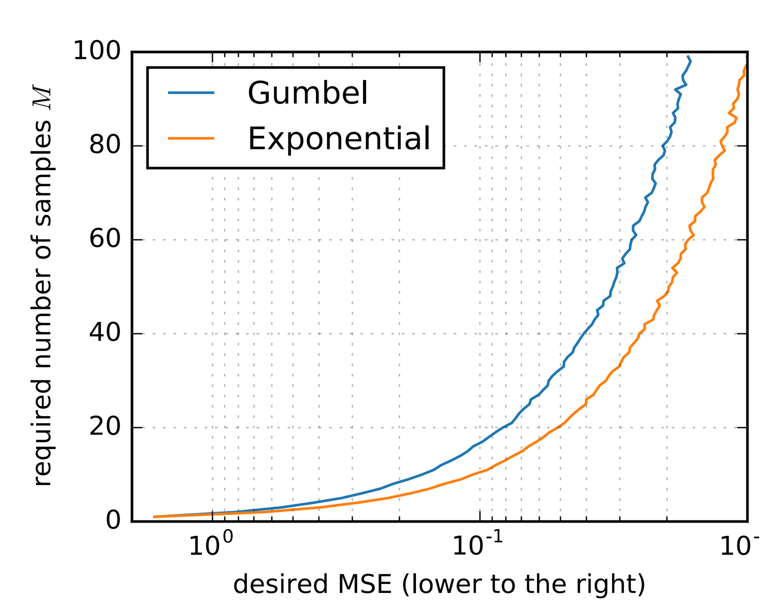


Takeaway:

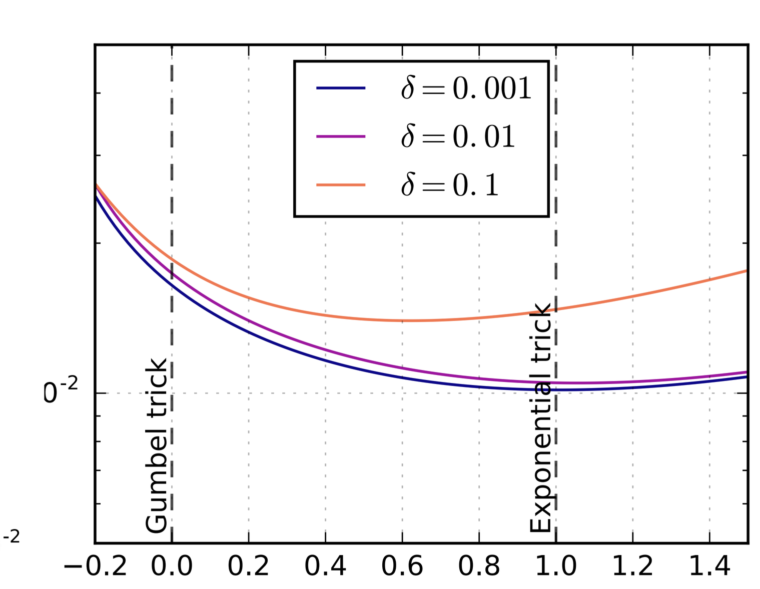
Save up to 40% samples by using the Exponential trick instead of the Gumbel trick. Applications:

- A* sampling of *Maddison et al. [2014]*
- Large-scale sampling of *Chen and Ghahramani [2016]*

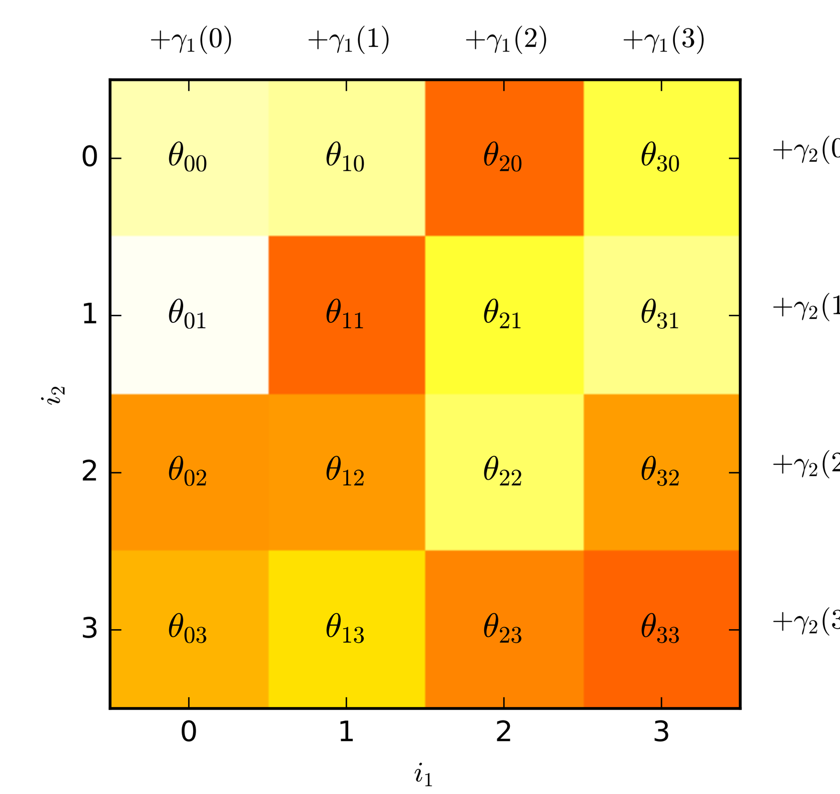
A* sampling:



MAB approximate MAP:



Discrete graphical models



Gumbel trick bounds:

$$U := \max_{\mathbf{x}\in\mathcal{X}} \left\{\theta(\mathbf{x}) + \sum_{i=1}^{n} \gamma_i(x_i)\right\}$$

$$L := \max_{\mathbf{x}\in\mathcal{X}} \left\{\theta(\mathbf{x}) + \frac{1}{n}\sum_{i=1}^{n} \gamma_i(x_i)\right\}$$

$$\mathbb{E}[L] \le \ln Z \le \mathbb{E}[U]$$

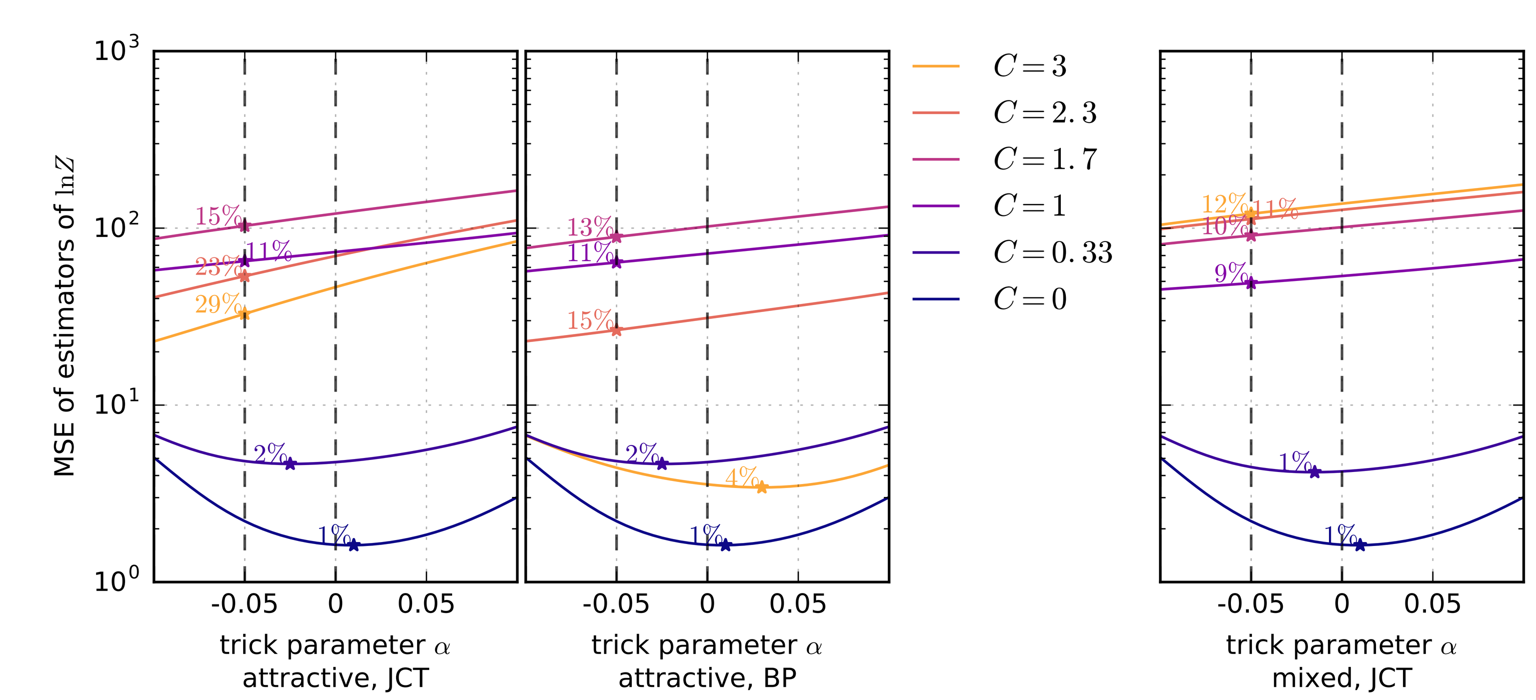
New bounds from *Weibull* and *Fréchet* tricks

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

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

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

Low-rank perturbations (discrete graphical models):



Pointers

Contact: matej.balog@gmail.com

Slides: http://matejbalog.eu/en/research/

Code: https://github.com/matejbalog/gumbel-relatives