Low-variance Gradient Estimates for the Plackett-Luce Distribution

Artyom Gadetsky*
NRU HSE†
Moscow, Russia
artygadetsky@yandex.ru

Kirill Struminsky*
NRU HSE†
Moscow, Russia
k.struminsky@gmail.com

Christopher Robinson
University of Sussex
Brighton, UK
cfr20@sussex.ac.uk

Novi Quadrianto
University of Sussex and NRU HSE†
Brighton, UK
n.quadrianto@sussex.ac.uk

Dmitry Vetrov
NRU HSE‡
Moscow, Russia
vetrovd@yandex.ru

Abstract

Learning models with discrete latent variables using stochastic gradient descent remains a challenge due to the high variance of gradients. Modern variance reduction techniques mostly consider categorical distributions and have limited applicability when the number of possible outcomes becomes large. In this work, we consider models with latent permutations and propose control variates for the Plackett-Luce distribution. Our proof-of-concept experiment recasts optimization over permutations as a variational optimization w.r.t. the Plackett-Luce distribution and solves it using stochastic gradient descent.

1 Introduction

Despite the recent breakthroughs in gradient estimation for continuous latent variables (Kingma & Welling, 2013; Rezende et al., 2014; Mohamed et al., 2019), gradient estimation for discrete latent variables remains a challenge. Currently, general-purpose estimators (Williams, 1992; Mnih & Gregor, 2014) remain unreliable and the state-of-the-art methods (Tucker et al., 2017; Grathwohl et al., 2018; Yin & Zhou, 2018) exclusively consider the categorical distribution. Although the reduction to the categorical case allows benefiting from gradient estimators for continuous relaxations, such solutions are hard to translate to discrete distributions with large support.

In this work, we consider a gradient estimator for the Plackett-Luce distribution, a distribution over permutations. Permutations naturally occur in various setting, such as ranking problems (Guiver & Snelson, 2009), optimal routing (Bello et al., 2016) and causal inference (Friedman & Koller, 2003). However, the support of the distribution is superexponential in the number of items k, which makes representing a distribution as a categorical distribution intractable even for dozens of items. At the same time, the Plackett-Luce distribution has $O(k)$ parameters and allows sampling in $O(k \log k)$.

We translate the recent variance reduction techniques introduced in Tucker et al. (2017); Grathwohl et al. (2018) to the case of Plackett-Luce distributions. Similarly to REBAR, we use the difference of the REINFORCE estimator and the reparametrized estimator for the relaxed model. In the experimental section we recast an optimization tasks over the discrete domain of permutations as a variational optimization for the Plackett-Luce distribution and then solve it using stochastic gradient

*Both authors contributed equally to this work.
†National Research University Higher School of Economics
‡Samsung-HSE Laboratory

4th workshop on Bayesian Deep Learning (NeurIPS 2019), Vancouver, Canada.

descent. *In future, our method could be used for end-to-end gradient training of deep models with latent permutations, routing problems and other combinatorial problems for permutations.*

1.1 A Brief Tour of Gradient Estimation

In this section, we consider a general optimization task $\min_\theta \mathbb{E}_{p(b|\theta)}[f(b)]$, where b is a discrete random variable parametrized by θ. The expectation can be intractable, for instance when b is a vector of categorical variables and the support of b is exponential in the vector length. The standard solution is to construct a stochastic estimate for the gradient $\hat{g}(f) := \frac{\partial}{\partial\theta}\mathbb{E}_{p(b|\theta)}[f(b)]$ without explicitly computing the expectation. In this section, we briefly review the gradient estimation algorithms.

1.2 REINFORCE

The REINFORCE estimator Williams (1992) gives us a widely-applicable unbiased estimate for the gradient

$$\hat{g}_{\textit{REINFORCE}}(f) = f(b)\frac{\partial}{\partial\theta}\log p(b \mid \theta), \;\; b \sim p(b \mid \theta). \tag{1}$$

Although an unbiased gradient estimate is sufficient to guarantee convergence of stochastic gradient descent, in practice, the algorithm may not converge due to the high variance of the estimate (Tucker et al., 2017). The variance of the REINFORCE estimator can be reduced using control variates. A Control variate is a function $c(b)$ with a zero mean $\mathbb{E}_{p(b|\theta)}[c(b)] = 0$ that can be used to define another unbiased estimator

$$\hat{g}_{\textit{CV}}(f) = \hat{g}_{\textit{REINFORCE}}(f) - c(b). \tag{2}$$

The variance of the new estimator $\hat{g}_{\textit{CV}}(f)$ is lower than the variance of $\hat{g}_{\textit{REINFORCE}}(f)$ if $c(b)$ is positively correlated with the random variable $f(b)$. As an illustration, the gradient of probability $\frac{\partial}{\partial\theta}\log p(b \mid \theta)$ has zero mean, therefore it can be used as a control variate Mnih & Gregor (2014).

1.3 Reparametrization Gradients for Continuous Relaxations

The reparametrization trick (Kingma & Welling, 2013; Rezende et al., 2014) is an alternative unbiased low-variance gradient estimator, applicable when f is differentiable and the latent variable $b_{\textit{cont}}$ is continuous. The estimator represents the latent variable as a differentiable determinisitc transformation $b_{\textit{cont}} = T(v, \theta)$ of a fixed distribution sample v and parameters θ and estimates the gradient as

$$\hat{g}_{\textit{reparam}}(f) = \frac{\partial}{\partial\theta}f(b_{\textit{cont}}) = \frac{\partial f}{\partial T}\frac{\partial T}{\partial\theta}, \; v_i \sim \text{uniform}[0,1], \; i = 1, \dots, k. \tag{3}$$

Although the reparametrization trick is not applicable when the latent variable b is discrete, Jang et al. (2016); Maddison et al. (2016) proposed the Gumbel-softmax estimator, a modification of the reparametrization trick for the relaxed categorical distribution.

To sample from a relaxed categorical distribution $p(b \mid \theta)$ with probabilities $\frac{\exp\theta_i}{\sum_j \exp\theta_j}$, Gumbel-Softmax first samples a vector of independent Gumbel random variables $z_i \sim \mathcal{G}(\theta_i, 1), i = 1, \dots, k$

$$z_i = T(\theta_i, v_i) = \theta_i - \log(-\log(v_i)), \; v_i \sim \text{uniform}[0,1], \; i = 1, \dots, k \tag{4}$$

with location parameter θ. According to the **Gumbel-max trick** (Maddison et al., 2014), the index of the maximal element $H(z) = \arg\max(z)$ is a categorical random variable with distribution $p(b \mid \theta)$. Then, to make the sampler differentiable, the Gumbel-softmax trick replaces $\arg\max(z)$ with a relaxation $\sigma(z) = \frac{1}{\sum \exp z_i}(\exp z_1, \dots, \exp z_k)$. The gradient estimate is the reparametrization gradient for the relaxed categorical distribution:

$$\hat{g}_{\textit{Gumbel}}(f) = \frac{\partial}{\partial\theta}f(b) = \frac{\partial f}{\partial b}\frac{\partial b}{\partial z}\frac{\partial z}{\partial\theta}, \; b = \sigma(z), \; z_i \sim \mathcal{G}(\theta_i, 1), i = 1, \dots, k. \tag{5}$$

The resulting reparametrization gradient $\hat{g}_{\textit{Gumbel}}(f)$ has much lower variance than $\hat{g}_{\textit{REINFORCE}}(f)$, but is generally biased due to the relaxation.

1.4 Relaxation-based Control Variates

Recently, Tucker et al. (2017) and Grathwohl et al. (2018) proposed control variates for REINFORCE estimator based on the relaxed conditional distribution. Both works use the REINFORCE gradient estimator for the relaxed categorical distribution as a control variate for the non-relaxed estimator. To eliminate the bias of the REINFORCE estimator, they subtract the low-variance reparametrization gradient estimator.

The key insight of Tucker et al. (2017) is the conditional marginalization step used to correlate the non-relaxed REINFORCE estimator and the control variate. Importantly, the conditional marginalization relies on reparametrization trick for the conditional distribution $p(z \mid b, \theta)$, obtained from the joint distribution $p(b, z \mid \theta) = p(b|z)p(z \mid \theta)$ of the Gumbel random vector z and the output of the Gumbel-max trick $b = H(z) = \arg\max(z)$. Tucker et al. (2017) derive a reparametrizable sampling scheme for $p(z \mid b, \theta)$

$$\tilde{z}_i = \begin{cases} -\log(-\log v_i) & i = b \\ -\log\left(-\frac{\log v_i}{\exp \theta_i} + \exp(-\tilde{z}_b)\right) & i \neq b \end{cases}, \tag{6}$$

where vector v is a uniform i.i.d. vector $v \sim \text{uniform}[0, 1]^k$. This gives a two-step generative process for the distribution $p(z \mid b, \theta)$. On the first step we sample the maximum variable v_b from the Gumbel distribtuion and on the second step we sample the other variables $v_i, i \neq b$ from the Gumbel distribution truncated at $\tilde{z}_b$ with location parameter θ_i.

The unbiased RELAX estimator from Grathwohl et al. (2018) is

$$\hat{g}_{RELAX}(f) =[f(b) - c_\phi(\tilde{z})]\frac{\partial}{\partial\theta}\log p(b \mid \theta) + \frac{\partial}{\partial\theta}c_\phi(z) - \frac{\partial}{\partial\theta}c_\phi(\tilde{z}) \tag{7}$$

$$b =H(z),\ z \sim p(z \mid \theta),\ \tilde{z} \sim p(z \mid b, \theta) \tag{8}$$

where $c_\phi(z)$ is a parametric function optimized to reduce the variance of the estimator.

Similarly, for a differentiable function f the REBAR estimator by Tucker et al. (2017) uses the function f with the relaxed argument $\sigma(z)$ and tunes the scalar parameter η

$$\hat{g}_{REBAR}(f) =[f(b) - \eta f(\sigma(\tilde{z}))]\frac{\partial}{\partial\theta}\log p(b \mid \theta) + \eta\frac{\partial}{\partial\theta}f(\sigma(z)) - \eta\frac{\partial}{\partial\theta}f(\sigma(\tilde{z})) \tag{9}$$

$$b =H(z),\ z \sim p(z \mid \theta),\ \tilde{z} \sim p(z \mid b, \theta) \tag{10}$$

2 Constructing Control Variates for the Plackett-Luce Distribution

In this paper, we extend the stochastic gradient estimators $\hat{g}_{REBAR}(f)$ and $\hat{g}_{RELAX}(f)$ from the categorical distribution to the Plackett-Luce distribution. With a slight abuse of notation, below we use letter b to denote an integer vector $b = (b_1, \ldots, b_k) \in S_k$ that represent a permutation, θ to denote the parameters of the Plackett-Luce distribution and $p(b \mid \theta)$ to denote the Plackett-Luce distribution.

The goal of this section is to define the two components required to apply the aforementioned gradient estimators: the mapping $b = H(z)$ and the two reparametrizable conditional distributions $p(z \mid \theta)$ and $p(z|b, \theta)$. After this we apply the estimators as defined in eq. 7 and eq. 9, but to emphasize the difference we refer to them as PL-RELAX and PL-REBAR.

Definition 1. *The Plackett-Luce distribution (Luce, 2005; Plackett, 1975) with scores $\theta = (\theta_1, \ldots, \theta_k)$ is a distribution over permutations S_k with the probability of outcome $b \in S_k$*

$$p(b|\theta) = \prod_{j=1}^{k} \frac{\exp \theta_{b_j}}{\sum_{u=j}^{k} \exp \theta_{b_u}}. \tag{11}$$

Intuitively, a sample from the Plackett-Luce distribution $b = (b_1, \ldots, b_k)$ is generated as a sequence of samples from categorical distributions. The first component b_1 comes from the categorical distribution with logits θ, then the second components b_2 comes from the categorical distribution with the logits θ without the component θ_{b_1} and so on.

The Plackett-Luce can be used for variational optimization (Staines & Barber, 2012). Indeed, at the lower temperatures $\theta \to \frac{\theta}{T}, T \ll 1$ the distribution converges to a divergent distribution. The mode of the Plackett-Luce distribution is the descending order permutation of the scores $b^0 : \theta_{b^0_1} \geq \cdots \geq \theta_{b^0_k}$, because b^0 permutation maximizes each factor in the product in eq. 11.

Now we will give an alternative definition of the Plackett-Luce distribution.

Lemma 1. *(appears in Grover et al. (2019); Yellott Jr (1977)) Let z be a vector of k independent Gumbel random variables with location parameters specified by score vector θ*

$$z_i = \theta_i - \log(-\log(v_i)),\ v_i \sim \mathrm{uniform}[0,1]. \tag{12}$$

Then for a permutation $b \in S_k$ the probability of event $\{z_{b_1} \geq \cdots \geq z_{b_k}\}$ is

$$p(z_{b_1} \geq \cdots \geq z_{b_k}) = \prod_{j=1}^{k} \frac{\exp \theta_{b_j}}{\sum_{u=j}^{k} \exp \theta_{b_u}}. \tag{13}$$

Similarly to the **Gumbel-max trick**, Lemma 1 shows that an order of a Gumbel-distributed vector is distributed according to the Plackett-Luce distribution. Following the lemma, for Plackett-Luce distributions we define $p(z \mid \theta)$ to be a Gumbel-distributed vector $z_i \sim \mathcal{G}(\theta_i, 1),\ i = 1, \ldots, k$ and $H(z) = \arg\mathrm{sort}(z)$ to be a sorting operation

Our principal discovery is that, similarly to the categorical case, the conditional distribution $p(z|b,\theta)$ factorizes into a sequence of truncated Gumbel distributions. As a consequence, the distribution is reparametrizable and can be used to construct a control variate for a gradient estimator.

Proposition 1. *Let $p(b, z \mid \theta)$ be the joint distribution with $z_i \sim \mathcal{G}(\theta_i, 1)$, $b = \arg\mathrm{sort}(z)$ and normalized parameters $\sum_{j=1}^{k} \exp \theta_j = 1$. Then for uniform i.i.d samples $v_i \sim \mathrm{uniform}[0,1]$ and $\Theta_i = \sum_{j=i}^{k} \exp \theta_{b_j}$ for $i = 1, \ldots, k$ the vector $\tilde{z} = (\tilde{z}_1, \ldots, \tilde{z}_k)$*

$$\tilde{z}_{b_i} = \begin{cases} -\log(-\log v_i) & i = 1 \\ -\log(-\frac{\log v_i}{\Theta_i} + \exp(-\tilde{z}_{b_{i-1}})) & i \geq 2, \end{cases} \tag{14}$$

is a sample from the conditional distriubtion $p(z \mid b, \theta)$.

We prove the proposition in the appendix. The sampling procedure from Proposition 1 has two principal differences from the sampling scheme for the categorical case (see eq. 6). First, the truncation parameter $\tilde{z}_{b_{i-1}}$ now depends on the previous component $i-1$, while for the categorical case the truncation parameter is defined by the maximum component. Second, the location parameter is now a cumulative sum and depends on the previous scores.

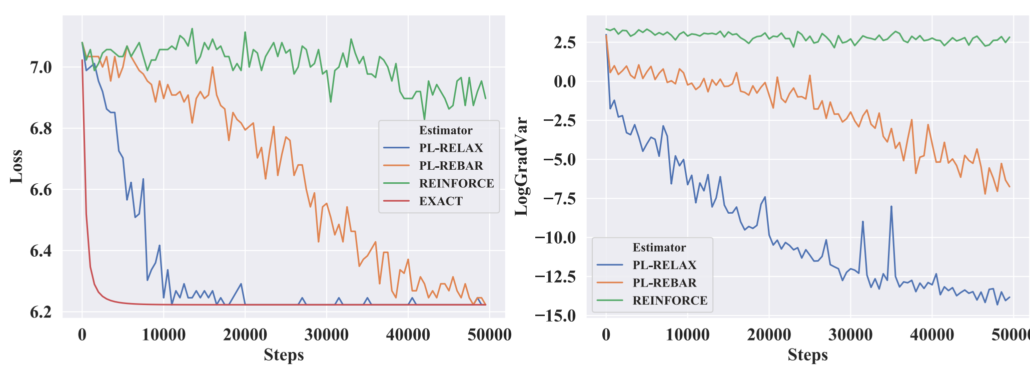

Figure 1: Training curves and log-variance of gradient estimators for different estimators on a toy problem: $\mathbb{E}_{p(b|\theta)}\|P_b - P_{0.05}\|_F^2$

3 Experiment

As a proof of concept we perform an experiment in minimizing $\mathbb{E}_{p(b|\theta)}\|P_b - P_t\|_F^2 = \mathbb{E}_{p(b|\theta)} f(P_b)$ as a function of θ where $p(b|\theta) = \text{Plackett-Luce}(b|\theta)$. P_b is permutation matrix with elements

$p_{i,b_i} = 1$ and P_t is a matrix with $\frac{1}{k} + t$ on the main diagonal and $\frac{1}{k} - \frac{t}{k-1}$ in the remaining positions. This problem can be seen as linear sum assignment problem with specifically constructed doubly stochastic matrix P_t. It is easy to note that taking $k = 2$ and $t = 0.05$ leads to toy problem similar to that of Tucker et al. (2017). We focus on $t = 0.05$ and $k = 8$ to enable computation of exact gradients. For the PL-REBAR estimator we take $c_\phi(z) = \eta f(\sigma(z, \tau))$ where $\sigma(z, \tau)$ is the continuous relaxation of permutations described by Grover et al. (2019). For the PL-RELAX estimator we take $c_\phi(z) = f(\sigma(z, \tau)) + \rho_\phi(z)$ where $\rho_\phi(z)$ is a simple neural network with two linear layers and ReLU activation between them. Figure 1 shows the relative performance and gradient log-variance of REINFORCE, PL-REBAR and PL-RELAX. Although the REINFORCE estimator is unbiased, we can see that the variance of the estimator is too large even for the simple toy task, therefore the method is completely inapplicable for optimization over permutations. On the other hand, the proposed method significantly reduces variance of the gradient and thus converges to optimal. Also, similarly to the toy experiment from Grathwohl et al. (2018) paper, we observe better performance of the PL-RELAX estimator due to free-form control variate parameterized by a neural network. Our PyTorch Paszke et al. (2017) implementation of the gradient estimators is available at https://github.com/agadetsky/pytorch-pl-variance-reduction .

Acknowledgements

This work was partly supported by Sberbank AI Lab and UK EPSRC project EP/P03442X/1. Kirill Struminsky proved proposition 1 and was supported by the Russian Science Foundation grant no. 19-71-30020. The authors thank NRU HSE for providing computational resources, NVIDIA for GPU donations, and Amazon for AWS Cloud Credits.

References

Bello, Irwan, Pham, Hieu, Le, Quoc V, Norouzi, Mohammad, and Bengio, Samy. Neural combinatorial optimization with reinforcement learning. *arXiv preprint arXiv:1611.09940*, 2016.

Friedman, Nir and Koller, Daphne. Being bayesian about network structure. a bayesian approach to structure discovery in bayesian networks. *Machine learning*, 50(1-2):95–125, 2003.

Grathwohl, Will, Choi, Dami, Wu, Yuhuai, Roeder, Geoff, and Duvenaud, David. Backpropagation through the void: Optimizing control variates for black-box gradient estimation. In *International Conference on Learning Representations*, 2018. URL `https://openreview.net/forum?id=SyzKd1bCW`.

Grover, Aditya, Wang, Eric, Zweig, Aaron, and Ermon, Stefano. Stochastic optimization of sorting networks via continuous relaxations. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=H1eSS3CcKX`.

Guiver, John and Snelson, Edward. Bayesian inference for plackett-luce ranking models. In *proceedings of the 26th annual international conference on machine learning*, pp. 377–384. ACM, 2009.

Jang, Eric, Gu, Shixiang, and Poole, Ben. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

Kingma, Diederik P and Welling, Max. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Luce, R Duncan. *Individual Choice Behavior: A Theoretical Analysis*. Courier Corporation, 2005.

Maddison, Chris J, Tarlow, Daniel, and Minka, Tom. A* sampling. In *Advances in Neural Information Processing Systems*, pp. 3086–3094, 2014.

Maddison, Chris J, Mnih, Andriy, and Teh, Yee Whye. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016.

Mnih, Andriy and Gregor, Karol. Neural variational inference and learning in belief networks. *arXiv preprint arXiv:1402.0030*, 2014.

Mohamed, Shakir, Rosca, Mihaela, Figurnov, Michael, and Mnih, Andriy. Monte carlo gradient estimation in machine learning. *arXiv preprint arXiv:1906.10652*, 2019.

Paszke, Adam, Gross, Sam, Chintala, Soumith, Chanan, Gregory, Yang, Edward, DeVito, Zachary, Lin, Zeming, Desmaison, Alban, Antiga, Luca, and Lerer, Adam. Automatic differentiation in PyTorch. In *NIPS Autodiff Workshop*, 2017.

Plackett, Robin L. The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(2):193–202, 1975.

Rezende, Danilo Jimenez, Mohamed, Shakir, and Wierstra, Daan. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

Staines, Joe and Barber, David. Variational optimization. *arXiv preprint arXiv:1212.4507*, 2012.

Tucker, George, Mnih, Andriy, Maddison, Chris J, Lawson, John, and Sohl-Dickstein, Jascha. Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models. In *Advances in Neural Information Processing Systems*, pp. 2627–2636, 2017.

Williams, Ronald J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

Yellott Jr, John I. The relationship between luce's choice axiom, thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2): 109–144, 1977.

Yin, Mingzhang and Zhou, Mingyuan. Arm: Augment-reinforce-merge gradient for discrete latent variable models. *arXiv preprint arXiv:1807.11143*, 2018.

A Conditional Reparametrization for the Plackett-Luce Distribution

We prove Proposition 1 in this section. We first discuss the properties of Gumbel distribution. Then we discuss the generative processes for the densities used for $p(z \mid b, \theta)$ in Eq. 14. Then we show that $p(b \mid z)p(z \mid \theta) = p(b \mid \theta)p(z \mid b, \theta)$ for the unconditional Gumbel density $p(z \mid \theta)$ and the Plackett-Luce distribution $p(b \mid \theta)$.

A.1 Density for the Gumbel distribution and the truncated Gumbel distribution

The density function of the Gumbel distribution with location parameter μ is

$$\phi_\mu(z) = \exp(-z + \mu)\exp(-\exp(-z + \mu)) \tag{15}$$

and the cumulative density function is

$$\Phi_\mu = \exp(-\exp(-z + \mu)). \tag{16}$$

Our derivation of the conditional distribution $p(b \mid z, \theta)$ relies on the additive property of the cumulative density function of the Gumbel distribution

$$\Phi_{\log(\exp\mu + \exp\nu)}(z) = \exp(-\exp(z)(\exp\mu + \exp\nu)) = \Phi_\mu(z)\Phi_\nu(z), \tag{17}$$

which we enfold in the following auxiliary claim.

Lemma 2. *For permutation $b \in S_k$, score vector $\theta \in \mathbb{R}^k$ and $i = 1, \dots, k$ and the argument vector $z \in \mathbb{R}^k$ we have*

$$\phi_{\theta_{b_i}}(z_{b_i})\Phi_{\log(\sum_{j=i+1}^k \exp\theta_{b_j})}(z_{b_i}) = \frac{\exp\theta_{b_i}}{\sum_{j=i}^k \exp\theta_{b_j}}\phi_{\log(\sum_{j=i}^k \exp\theta_{b_j})}(z_{b_i}). \tag{18}$$

Proof. For brevity, we denote $\exp\theta_i$ as p_i. We then rewrite the density $\phi_{\log p_{b_i}}(z_{b_i})$ through the exponent $\exp(-z_{b_i} + \log p_{b_i})$ and c.d.f. $\Phi_{\log p_{b_i}}(z_{b_i})$ and apply the additive property in Eq. 20:

$$\phi_{\log p_{b_i}}(z_{b_i})\Phi_{\log(\sum_{j=i+1}^k p_{b_j})}(z_{b_i}) = p_{b_i}\exp(-z_{b_i})\Phi_{\log p_{b_i}}(z_{b_i})\Phi_{\log(\sum_{j=i+1}^k p_{b_j})}(z_{b_i}) \tag{19}$$

$$= p_{b_i}\exp(-z_{b_i})\Phi_{\log(\sum_{j=i}^k p_{b_j})}(z_{b_i}) = p_{b_i}\frac{\sum_{j=i}^k p_{b_j}}{\sum_{j=i}^k p_{b_j}}\exp(-z_{b_i})\Phi_{\log(\sum_{j=i}^k p_{b_j})}(z_{b_i}) \tag{20}$$

$$= \frac{p_{b_i}}{\sum_{j=1}^k p_{b_j}}\phi_{\log(\sum_{j=i}^k p_{b_j})}(z_{b_i}). \tag{21}$$

The last step collapses the exponent and the c.d.f. into the density function $\phi_{\log(\sum_{j=i}^k p_{b_j})}(z_{b_i})$. $\square$

Finally, to define the density of conditional distribution $p(b \mid z, \theta)$ we define the density of the truncated Gumbel distribution $\phi_\mu^{z_0}(z) \propto \phi_\mu(z)I[z \le z_0]$:

$$\phi_\mu^{z_0}(z) = \frac{\phi_\mu(z)}{\Phi_\mu(z_0)}(z)I[z \le z_0], \tag{22}$$

where the superscript z_0 denotes the truncation parameter.

A.2 Reparametrization for the Gumbel distribution and the truncated Gumbel distribution

The reparametrization trick requires representing a draw from a distribution as a deterministic transformation of a fixed distribution sample and a distribution parameter. For a sample z from the Gumbel distribution $\mathcal{G}(\mu, 1)$ with location parameter μ the representation is

$$z = \mu - \log(-\log v),\ v \sim \mathrm{uniform}[0, 1]. \tag{23}$$

For the Gumbel distribution truncated at z_0 Maddison et al. (2014) proposed an analogous representation

$$z = \mu - \log(-\log v + \exp(-z_0 + \mu)) = -\log\left(-\frac{\log v}{\exp\mu} + \exp(-z_0)\right), \tag{24}$$

for $v \sim \mathrm{uniform}[0, 1]$. In particular, the sampling schemes in Eq. 6 and Eq. 14 generate samples from the truncated Gumbel distribution.

A.3 The derivation of the conditional distribution

We now derive the conditional distribution and the sampling scheme defined in Proposition 1.

The joint distribution of the permutation b and the Gumbel samples z is

$$p(b, z \mid \theta) = p(b \mid z)p(z \mid \theta) = \phi_{\theta_{b_1}}(z_{b_1}) \prod_{i=2}^{k} \Big(\phi_{\theta_{b_i}}(z_{b_i}) I[z_{b_{i-1}} \geq z_{b_i}] \Big) \tag{25}$$

We first multiply and divide the joint density by the c.d.f. $\Phi_{\log(\sum_{i=2}^{k} \exp\theta_{b_i})}(z_{b_1})$ and apply Lemma 2

$$\frac{\Phi_{\log(\sum_{i=2}^{k} \exp\theta_{b_i})}(z_{b_1})}{\Phi_{\log(\sum_{i=2}^{k} \exp\theta_{b_i})}(z_{b_1})} \phi_{\theta_{b_1}}(z_{b_1}) \prod_{i=2}^{k} \cdots = \frac{\exp\theta_{b_1}}{\sum_{i=1}^{k} \exp\theta_{b_i}} \frac{\phi_{\log(\sum_{i=1}^{k} \exp\theta_{b_i})}(z_{b_1})}{\Phi_{\log(\sum_{i=2}^{k} \exp\theta_{b_i})}(z_{b_1})} \prod_{i=2}^{k} \cdots . \tag{26}$$

Next, we apply Lemma 2 to combine the c.d.f. in the denominator $\Phi_{\log(\sum_{j=i}^{k} \exp\theta_{b_j})}(z_{b_{i-1}})$ and the term $\phi_{\theta_{b_i}}(z_{b_i}) I[z_{b_{i-1}} \geq z_{b_i}]$ inside the product

$$\frac{\phi_{\theta_{b_i}}(z_{b_i}) I[z_{b_{i-1}} \geq z_{b_i}]}{\Phi_{\log(\sum_{j=i}^{k} \exp\theta_{b_j})}(z_{b_{i-1}})} = \frac{\phi_{\theta_{b_i}}(z_{b_i}) I[z_{b_{i-1}} \geq z_{b_i}]}{\Phi_{\log(\sum_{j=i}^{k} \exp\theta_{b_j})}(z_{b_{i-1}})} \frac{\Phi_{\log(\sum_{j=i+1}^{k} \exp\theta_{b_j})}(z_{b_i})}{\Phi_{\log(\sum_{j=i+1}^{k} \exp\theta_{b_j})}(z_{b_i})} \tag{27}$$

$$= \frac{\exp\theta_{b_i}}{\sum_{j=i}^{k} \exp\theta_{b_j}} \frac{\phi^{z_{b_{i-1}}}_{\log(\sum_{j=i}^{k} \exp\theta_{b_j})}(z_{b_i})}{\Phi_{\log(\sum_{j=i+1}^{k} \exp\theta_{b_j})}(z_{b_i})} \tag{28}$$

and obtain the truncated distribution $\phi^{z_{b_{i-1}}}_{\log(\sum_{j=i}^{k} \exp\theta_{b_j})}(z_{b_i})$ along with one factor of the Plackett-Luce probability $\frac{\exp\theta_{b_i}}{\sum_{j=i}^{k} \exp\theta_{b_j}}$. Also, after the transformation the summation index in the denominator c.d.f. changes from i to $i+1$. This gives us an induction step that we apply sequentially for $i = 2, \dots, k-1$. For $i = k$ the denominator c.d.f. $\Phi_{\log\exp\theta_k}(z_{b_{k-1}})$ and the product term $\phi_{\log\exp\theta_k}(z_{b_k}) I[z_{k-1} \geq z_k]$ combine into the truncated Gumbel distribution with density $\phi^{z_{b_{k-1}}}_{\log\exp\theta_k}(z_{b_k})$.

As a result, we rearrange $p(b, z \mid \theta)$ into the product of the truncated Gumbel distribution densities $p(z \mid b, \theta)$ and the probability of the Plackett-Luce distribution $p(b \mid \theta)$:

$$\phi_{\log\sum_{j=1}^{k} \exp\theta_{b_j}}(z_{b_1}) \prod_{i=1}^{k} \frac{\exp\theta_{b_i}}{\sum_{j=i}^{k} \exp\theta_{b_j}} \prod_{i=2}^{k} \phi^{z_{b_{i-1}}}_{\log\sum_{j=i}^{k} \exp\theta_j}(z_{b_i}). \tag{29}$$

Finally, to obtain the claim of Proposition 1 we apply the reparametrized sampling scheme defined in Eq. 24.