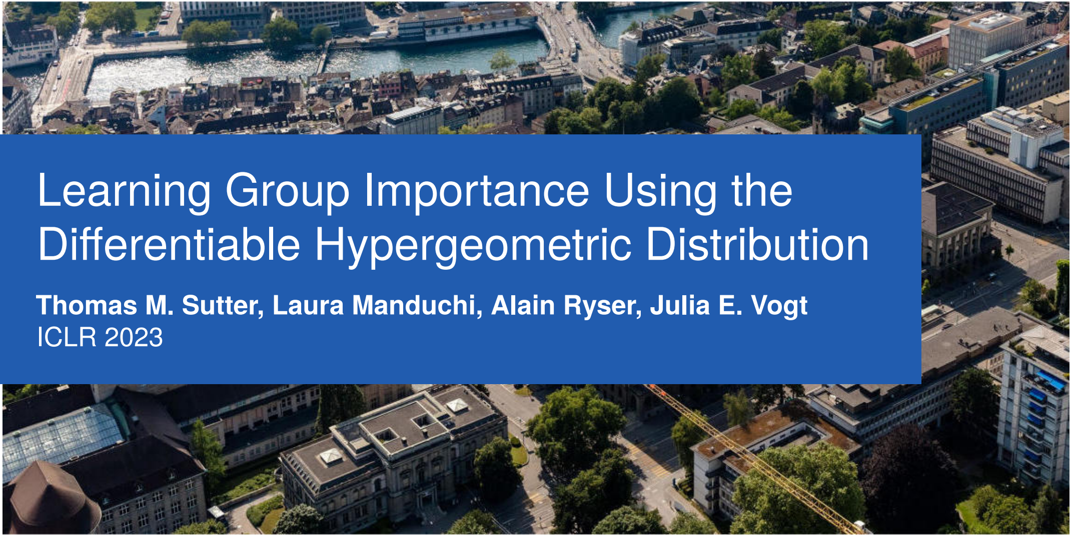
ETH zürich

Medical Data Science

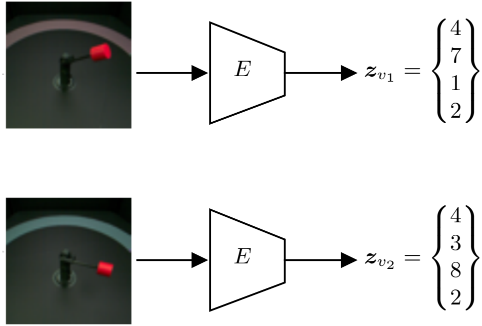


Learning Group Importance Using the Differentiable Hypergeometric Distribution

Thomas M. Sutter, Laura Manduchi, Alain Ryser, Julia E. Vogt

ICLR 2023

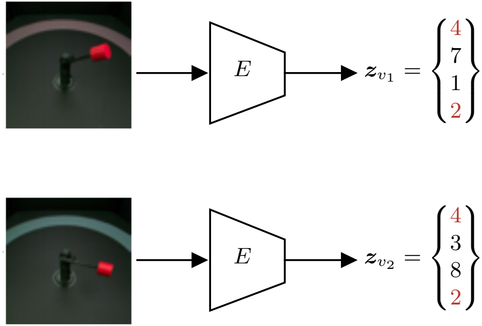
Motivation



- We are interested in learning from pairs of frames $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2]$
- Weak-supervision: **a subset of all generative factors is shared** between the frames
- **Neither true number of generative nor independent/shared factors is known in general**

Robot arm images taken from Locatello et al. [2020]

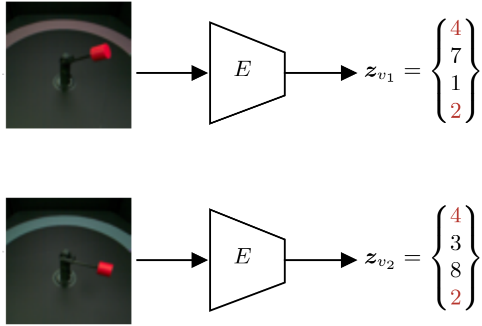
Motivation



- We are interested in learning from pairs of frames $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2]$
- Weak-supervision: **a subset of all generative factors is shared** between the frames
- **Neither true number of generative nor independent/shared factors is known in general**

Robot arm images taken from Locatello et al. [2020]

Motivation



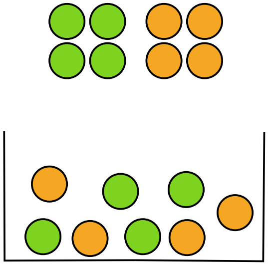
Can we model the number of shared and independent factors?

Robot arm images taken from Locatello et al. [2020]

Motivation

$m_s = 4$ $m_i = 4$

$N = 8$

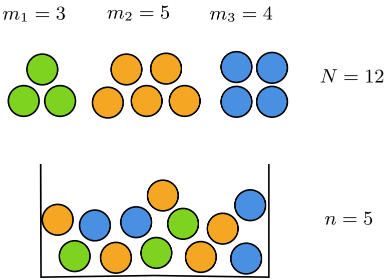


$n = 4$

Can we model the number of shared and independent factors?

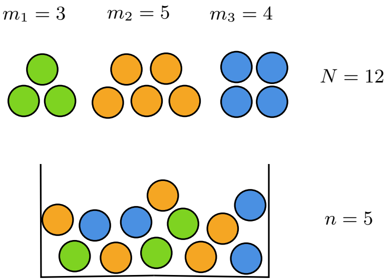
Robot arm images taken from Locatello et al. [2020]

Multivariate Hypergeometric Distribution



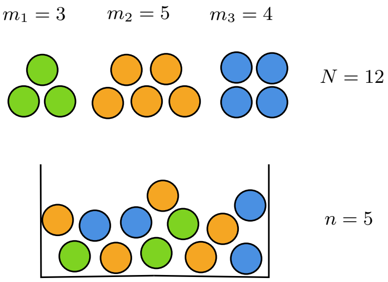
$$P(\boldsymbol{X} = \boldsymbol{x}) = p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\prod_{i=1}^{c} \binom{m_i}{x_i}}{\binom{N}{n}} \tag{1}$$

Multivariate Hypergeometric Distribution



$$p_{\boldsymbol{X}}(\boldsymbol{x}) \propto \prod_{i=1}^{c} \binom{m_i}{x_i} \quad (1)$$

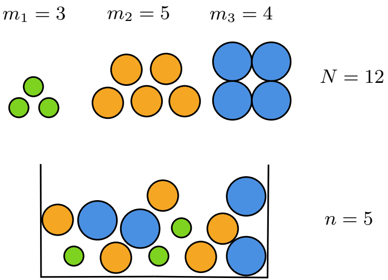
Hypergeometric Distribution
Noncentral [Fisher, 1935]



$$p_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\omega}) \propto \prod_{i=1}^{c} \binom{m_i}{x_i} \omega_i^{x_i} \quad (2)$$

ω_i: group importance parameter of group i

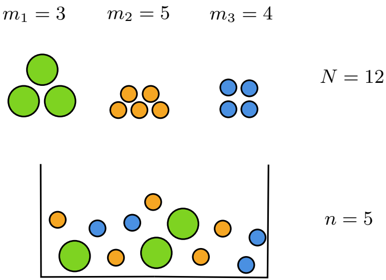
Hypergeometric Distribution
Noncentral [Fisher, 1935]



$$p_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\omega}) \propto \prod_{i=1}^{c} \binom{m_i}{x_i} \omega_i^{x_i} \tag{2}$$

ω_i: group importance parameter of group i

Hypergeometric Distribution
Noncentral [Fisher, 1935]

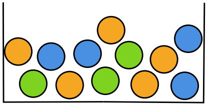


$$p_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\omega}) \propto \prod_{i=1}^{c} \binom{m_i}{x_i} \omega_i^{x_i} \quad (2)$$

ω_i: group importance parameter of group i

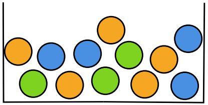
Method

1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.



Method

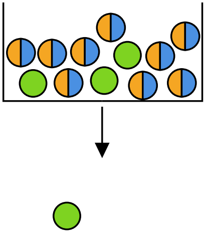
1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.



$$
\begin{aligned}
p_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\omega}) &= p_{\boldsymbol{X}}(x_1, x_2, x_3; \boldsymbol{\omega}) \\
&= p_{X_1}(x_1; \boldsymbol{\omega}) p_{X_2}(x_2 \mid x_1; \boldsymbol{\omega}) p_{X_3}(x_3 \mid x_1, x_2; \boldsymbol{\omega})
\end{aligned}
$$

Method

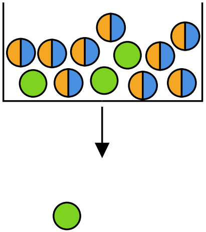
1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.



$$N = 12$$
$$n = 5$$
$$m_L = m_1 \quad \text{and} \quad m_R = m_2 + m_3$$
$$\omega_L = \omega_1 \quad \text{and} \quad \omega_R = \frac{\omega_2 m_2 + \omega_3 m_3}{m_R}$$
$$X_1 \sim p_{X_L}(n, m_L, m_R, \omega_L, \omega_R)$$

Method

1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.



$$N = 12$$

$$n = 5$$

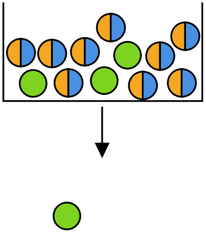
$$m_L = m_1 \quad \text{and} \quad m_R = m_2 + m_3$$

$$\omega_L = \omega_1 \quad \text{and} \quad \omega_R = \frac{\omega_2 m_2 + \omega_3 m_3}{m_R}$$

$$X_1 \sim p_{X_L}(n, m_L, m_R, \omega_L, \omega_R)$$

Method

1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.



$$N = 12$$

$$n = 5$$

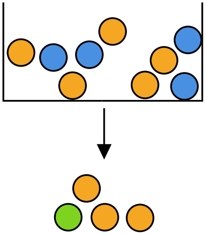
$$m_L = m_1 \quad \text{and} \quad m_R = m_2 + m_3$$

$$\omega_L = \omega_1 \quad \text{and} \quad \omega_R = \frac{\omega_2 m_2 + \omega_3 m_3}{m_R}$$

$$X_1 \sim p_{X_L}(n, m_L, m_R, \omega_L, \omega_R)$$

Method

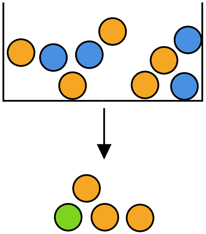
1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.



$$N = 9$$
$$n = 4$$
$$m_L = m_2 \quad \text{and} \quad m_R = m_3$$
$$\omega_L = \omega_2 \quad \text{and} \quad \omega_R = \omega_3$$
$$X_2 \sim p_{X_L}(n, m_L, m_R, \omega_L, \omega_R)$$

Method

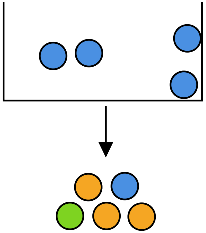
1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.



$$N = 9$$
$$n = 4$$
$$m_L = m_2 \quad \text{and} \quad m_R = m_3$$
$$\omega_L = \omega_2 \quad \text{and} \quad \omega_R = \omega_3$$
$$X_2 \sim p_{X_L}(n, m_L, m_R, \omega_L, \omega_R)$$

Method

1. Reformulate the multivariate distribution as a sequence of interdependent and conditional univariate hypergeometric distributions.
2. Calculate the probability mass function of the respective univariate distributions.
3. Sample from the conditional distributions utilizing the Gumbel-Softmax trick.

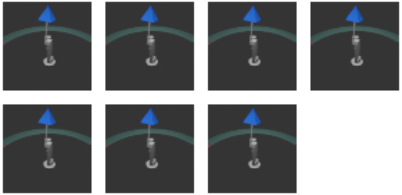


$$N = 4$$
$$n = 1$$
$$m_L = m_3 \quad \text{and} \quad m_R = 0$$
$$\omega_L = \omega_3 \quad \text{and} \quad \omega_R = 0$$
$$X_3 \sim p_{X_L}(n, m_L, 0, \omega_L, 0)$$

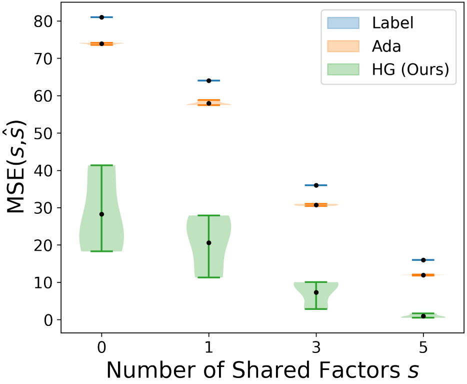
WSL: Dataset

`mpi3 toy`

- synthetic dataset with 7 generative factors
  - color
  - shape
  - size
  - camera height
  - background color
  - horizontal axis
  - vertical axis
- Dataset originally introduced as part of the Disentanglement challenge at Neurips 2019 [Gondal et al., 2019]
- We use `disentanglement_lib` for the experiments [Locatello et al., 2020]



WSL: Experiments & Results
Estimation of Number of Shared Factors



LabelVAE: [Bouchacourt et al., 2018, Hosoya, 2018]
AdaVAE: [Locatello et al., 2020]

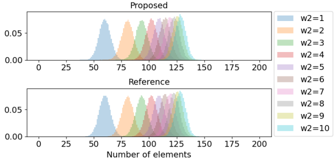
WSL: Experiments & Results
Downstream Tasks

	$s=0$	$s=1$		$s=3$		$s=5$	
	I	S	I	S	I	S	I
LABEL	0.14 ± 0.01	0.19 ± 0.03	0.16 ± 0.01	$\mathbf{0.10} \pm 0.00$	0.23 ± 0.01	$\mathbf{0.34} \pm 0.00$	0.00 ± 0.00
ADA	0.12 ± 0.01	0.19 ± 0.01	0.15 ± 0.01	$\mathbf{0.10} \pm 0.03$	0.22 ± 0.02	0.33 ± 0.03	0.00 ± 0.00
HG	$\mathbf{0.18} \pm 0.01$	$\mathbf{0.22} \pm 0.05$	$\mathbf{0.19} \pm 0.01$	0.08 ± 0.02	$\mathbf{0.28} \pm 0.01$	0.28 ± 0.01	0.01 ± 0.00

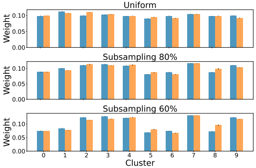
LabelVAE: [Bouchacourt et al., 2018, Hosoya, 2018]
AdaVAE: [Locatello et al., 2020]

In the paper, we have

- Detailed derivation of method
- Additional experiments, incl.
  - Kolmogorov-Smirnov test to compare proposed differentiable sampling to reference implementation [Kolmogorov, 1933, Smirnov, 1939]
  - MVHG as prior distribution in a clustering experiment



Histograms over Random Samples



Learned Cluster Weights

References I

D. Bouchacourt, R. Tomioka, and S. Nowozin. Multi-level variational autoencoder: Learning disentangled representations from grouped observations. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

R. A. Fisher. The logic of inductive inference. *Journal of the royal statistical society*, 98(1):39–82, 1935.

M. W. Gondal, M. Wuthrich, D. Miladinovic, F. Locatello, M. Breidt, V. Volchkov, J. Akpo, O. Bachem, B. Schölkopf, and S. Bauer. On the Transfer of Inductive Bias from Simulation to the Real World: a New Disentanglement Dataset. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

H. Hosoya. A simple probabilistic deep generative model for learning generalizable disentangled representations from grouped data. *CoRR*, abs/1809.0, 2018.

A. Kolmogorov. Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, 4: 83–91, 1933.

F. Locatello, B. Poole, G. Rätsch, B. Schölkopf, O. Bachem, and M. Tschannen. Weakly-supervised disentanglement without compromises. In *International Conference on Machine Learning*, pages 6348–6359. PMLR, 2020.

N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2):3–14, 1939.

ETH zürich

Visit us at our poster #58 on
Mon 1 May 11:30 a.m. CEST — 1:30 p.m. CEST

Link to Paper

Link to Github Repository