Royal Institute of Technology

VI INTRODUCTION

LDA

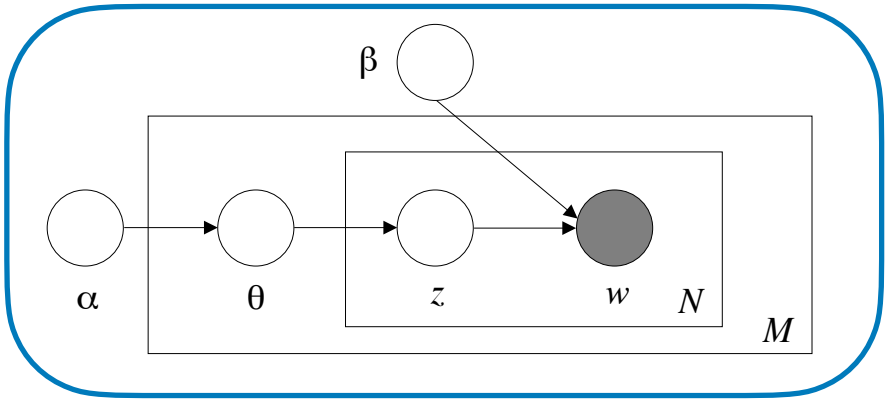
ARTICLE

https://doi.org/10.1038/s42003-020-01247-y OPEN

Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography

Alma Andersson[1✉], Joseph Bergenstråhle[1], Michaela Asp[1], Ludvig Bergenstråhle[1], Aleksandra Jurek[1], José Fernández Navarro[1] & Joakim Lundeberg[1✉]

The field of spatial transcriptomics is rapidly expanding, and with it the repertoire of available technologies. However, several of the transcriptome-wide spatial assays do not operate on a single cell level, but rather produce data comprised of contributions from a – potentially heterogeneous – mixture of cells. Still, these techniques are attractive to use when examining complex tissue specimens with diverse cell populations, where complete expression profiles are required to properly capture their richness. Motivated by an interest to put gene expression into context and delineate the spatial arrangement of cell types within a tissue, we here present a model-based probabilistic method that uses single cell data to deconvolve the cell mixtures in spatial data. To illustrate the capacity of our method, we use data from different experimental platforms and spatially map cell types from the mouse brain and developmental heart, which arrange as expected.



MAXIMUM LIKELIHOOD (ML) AND POSTERIOR PREDICTIVE

- ML
  - Estimate θ_{ML} from training data D and then use

$$y(x, \theta_{ML})$$

- Bayesian
  - Estimate a posterior distribution over θ based on D and then use

$$\int y(x, \theta) p(\theta \mid \mathcal{D}) d\theta$$

VARIATIONAL BAYES - MEAN FIELD

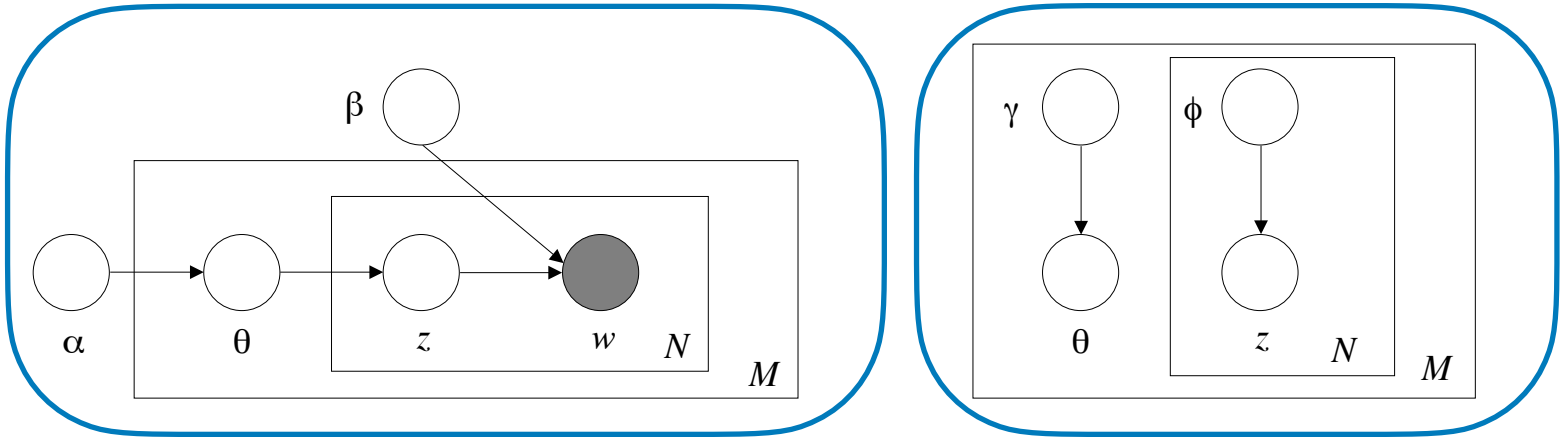


Figure 5: (Left) Graphical model representation of LDA. (Right) Graphical model representation of the variational distribution used to approximate the posterior in LDA.

- Variational Bayes (VB)
  - Technique for approximating a posterior

KL

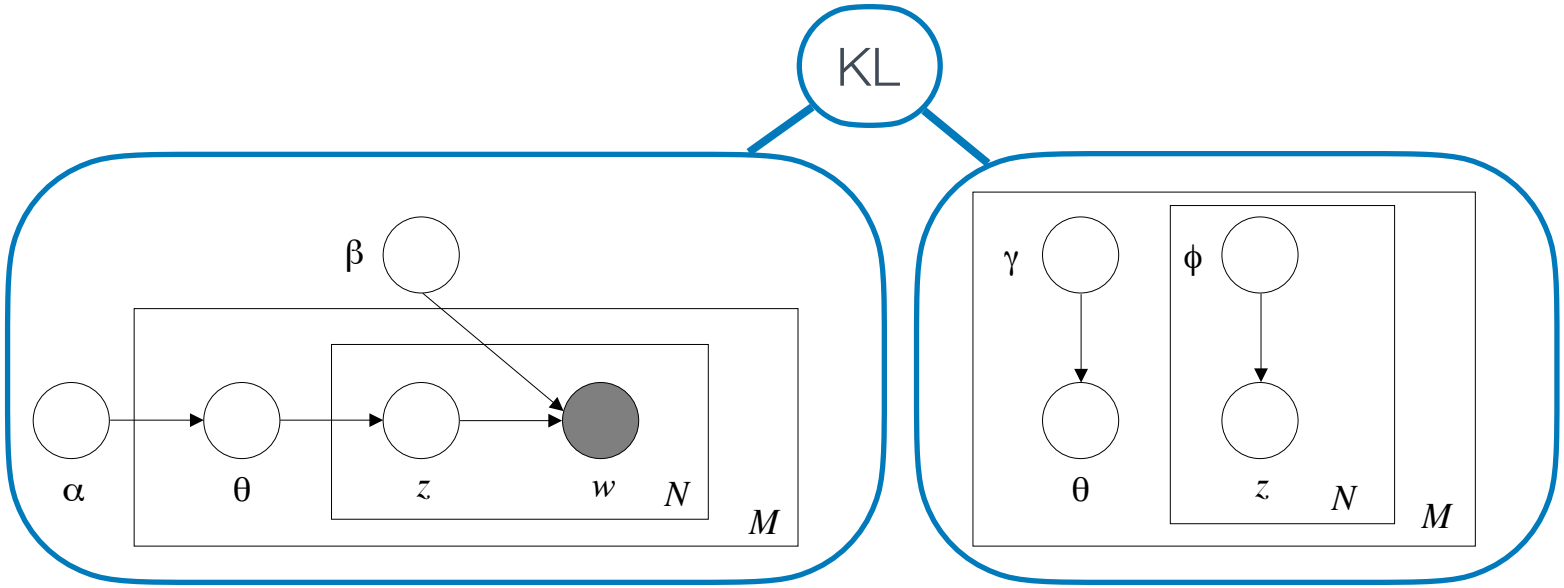


Figure 5: (Left) Graphical model representation of LDA. (Right) Graphical model representation of the variational distribution used to approximate the posterior in LDA.

- Variational Bayes (VB)
  - Technique for approximating a posterior

THE 2 KL BASED APPROACHES

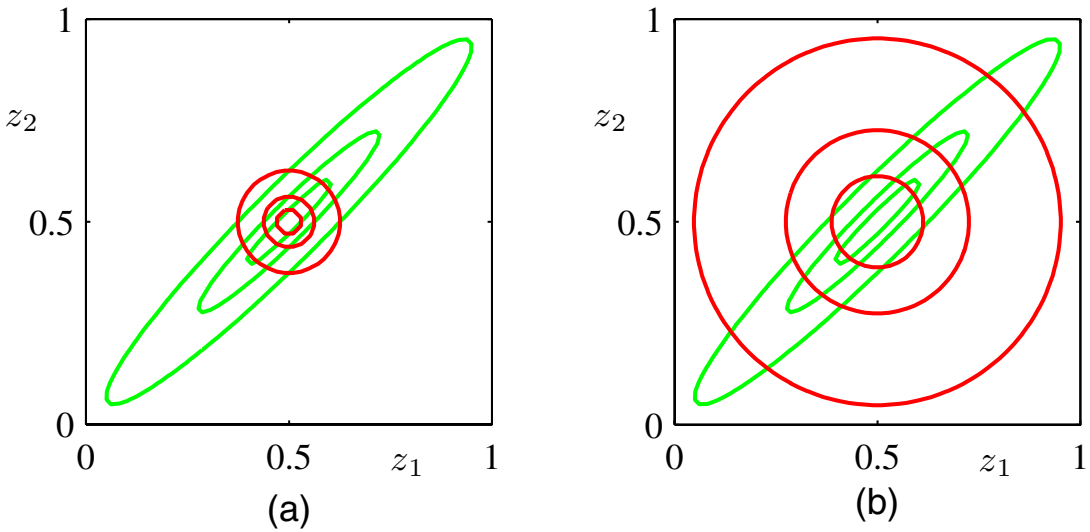
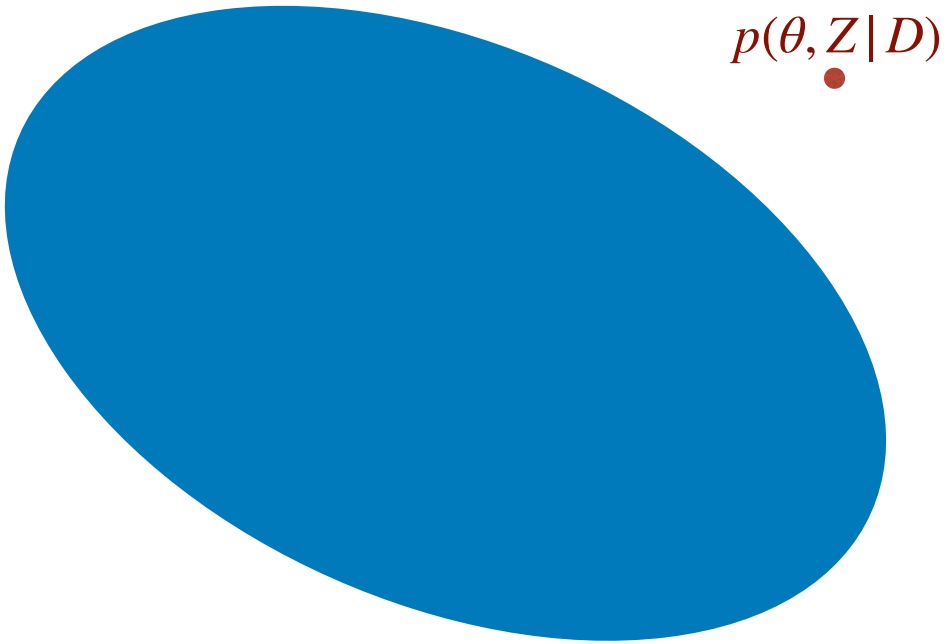


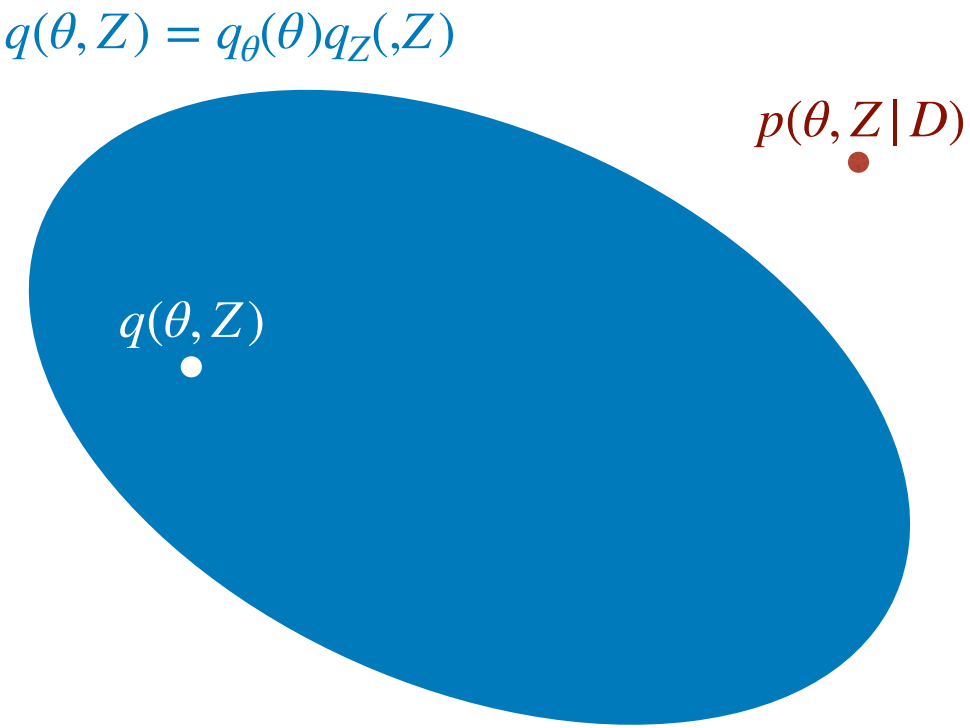
Figure 10.2 Comparison of the two alternative forms for the Kullback-Leibler divergence. The green contours corresponding to 1, 2, and 3 standard deviations for a correlated Gaussian distribution $p(\mathbf{z})$ over two variables z_1 and z_2, and the red contours represent the corresponding levels for an approximating distribution $q(\mathbf{z})$ over the same variables given by the product of two independent univariate Gaussian distributions whose parameters are obtained by minimization of (a) the Kullback-Leibler divergence $\mathrm{KL}(q\|p)$, and (b) the reverse Kullback-Leibler divergence $\mathrm{KL}(p\|q)$.

MINIMIZING KL-MAXIMIZING ELBO

$q(\theta, Z) = q_\theta(\theta) q_Z(,Z)$

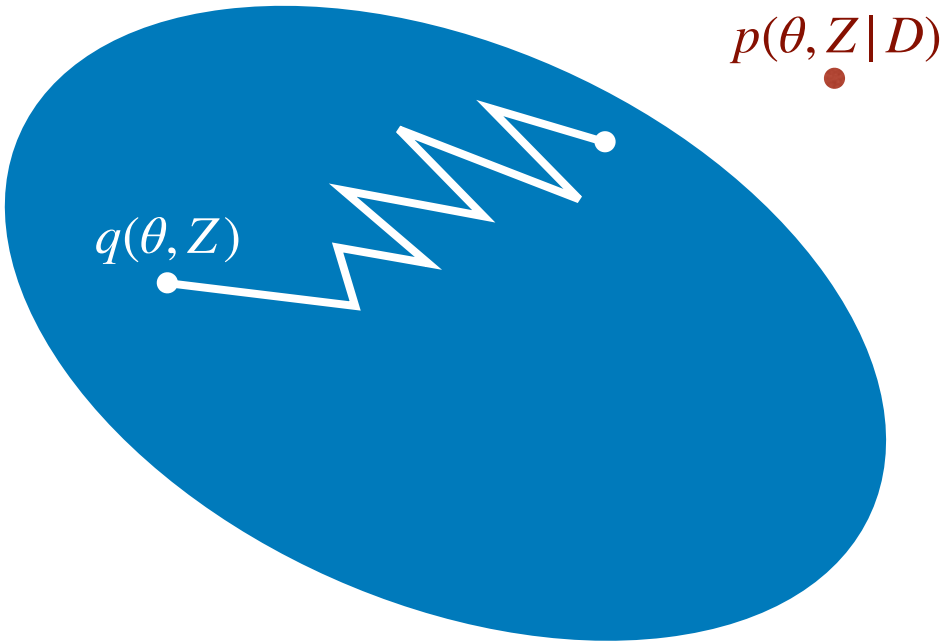


MINIMIZING KL-MAXIMIZING ELBO



MINIMIZING KL-MAXIMIZING ELBO

$q(\theta, Z) = q_\theta(\theta) q_Z(,Z)$



MINIMIZING KL-MAXIMIZING ELBO