

Bayesian Repulsive Mixtures for Probabilistic Topic Modelling

Student: Jake Hobson

Supervisor: Riccardo Passeggeri

Imperial College London

September 15, 2023

Presentation Outline

- 1 Latent Dirichlet Allocation
- 2 Repulsive Mixture Models
- 3 A Novel Topic Model
- 4 Conclusion

Table of Contents

1 Latent Dirichlet Allocation

2 Repulsive Mixture Models

3 A Novel Topic Model

4 Conclusion

Problem Overview

Problem Overview

- ▶ **Text-based datasets** are becoming ever-more prevalent.

Problem Overview

- ▶ **Text-based datasets** are becoming ever-more prevalent.
- ▶ Naive approaches analyse words/phrases individually.

Problem Overview

- ▶ **Text-based datasets** are becoming ever-more prevalent.
- ▶ Naive approaches analyse words/phrases individually.
- ▶ Can we instead analyse the **latent themes** in the dataset?

Problem Overview

- ▶ **Text-based datasets** are becoming ever-more prevalent.
- ▶ Naive approaches analyse words/phrases individually.
- ▶ Can we instead analyse the **latent themes** in the dataset?
- ▶ Then we can search for topics rather than words/phrases!

Building Topic Models

Building Topic Models

- ▶ **Topic models** are generative models designed to find latent topics in text-based datasets.

Building Topic Models

- ▶ **Topic models** are generative models designed to find latent topics in text-based datasets.
- ▶ **Document** - a collection of words (observed)

Building Topic Models

- ▶ **Topic models** are generative models designed to find latent topics in text-based datasets.
- ▶ **Document** - a collection of words (observed)
- ▶ **Topic** - a discrete distribution over words (latent).

Building Topic Models

- ▶ **Topic models** are generative models designed to find latent topics in text-based datasets.
- ▶ **Document** - a collection of words (observed)
- ▶ **Topic** - a discrete distribution over words (latent).
 - Example: *statistics*

Building Topic Models

- ▶ **Topic models** are generative models designed to find latent topics in text-based datasets.
- ▶ **Document** - a collection of words (observed)
- ▶ **Topic** - a discrete distribution over words (latent).
 - Example: *statistics*

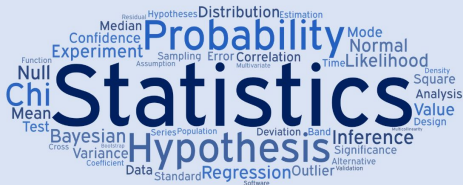


Figure: Potential word distribution of *statistics* topic.

Latent Dirichlet Allocation (LDA)¹

¹Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003)

Latent Dirichlet Allocation (LDA)¹

- ▶ Model assumptions:
 - V words in a known vocabulary.
 - T topics in the dataset.
 - N documents in the dataset.

¹Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003)

Latent Dirichlet Allocation (LDA)¹

- Model assumptions:
- V words in a known vocabulary.
 - T topics in the dataset.
 - N documents in the dataset.

$$\phi_t \sim \text{Dirichlet}_V(\eta), \quad t = 1, \dots, T$$

$$\theta_n \sim \text{Dirichlet}_T(\alpha), \quad n = 1, \dots, N$$

$$z_{nm} | \theta_n \sim \text{Categorical}_T(\theta_n), \quad m = 1, \dots, M_n, \quad n = 1, \dots, N$$

$$w_{nm} | z_{nm}, \phi \sim \text{Categorical}_V(\phi_{z_{nm}}), \quad m = 1, \dots, M_n, \quad n = 1, \dots, N.$$

ϕ_t	t -th topic
θ_n	topic mixture of n -th document
z_{nm}	topic index of m -th word in n -th document
w_{nm}	m -th word in n -th document

¹Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003)

Summary of LDA

Summary of LDA

- **Topics** ϕ_1, \dots, ϕ_T independent from V -dimensional Dirichlet distribution.

Summary of LDA

- ▶ **Topics** ϕ_1, \dots, ϕ_T independent from V -dimensional Dirichlet distribution.
- ▶ **Topic mixtures** $\theta_1, \dots, \theta_N$ independent from T -dimensional Dirichlet distribution.

Summary of LDA

- ▶ **Topics** ϕ_1, \dots, ϕ_T independent from V -dimensional Dirichlet distribution.
- ▶ **Topic mixtures** $\theta_1, \dots, \theta_N$ independent from T -dimensional Dirichlet distribution.
- ▶ For n -th **document**, m -th word w_{nm} generated via:
 - Sample topic for that word according to topic mixture.
 - Sample word according to topic.

Summary of LDA

- ▶ **Topics** ϕ_1, \dots, ϕ_T independent from V -dimensional Dirichlet distribution.
- ▶ **Topic mixtures** $\theta_1, \dots, \theta_N$ independent from T -dimensional Dirichlet distribution.
- ▶ For n -th **document**, m -th word w_{nm} generated via:
 - Sample topic for that word according to topic mixture.
 - Sample word according to topic.
- ▶ Drawbacks: all latent variables are independent and Dirichlet distributed.

Summary of LDA

- ▶ **Topics** ϕ_1, \dots, ϕ_T independent from V -dimensional Dirichlet distribution.
- ▶ **Topic mixtures** $\theta_1, \dots, \theta_N$ independent from T -dimensional Dirichlet distribution.
- ▶ For n -th **document**, m -th word w_{nm} generated via:
 - Sample topic for that word according to topic mixture.
 - Sample word according to topic.
- ▶ Drawbacks: all latent variables are independent and Dirichlet distributed.
- ▶ *How else could we generate the latent variables?*

Table of Contents

1 Latent Dirichlet Allocation

2 Repulsive Mixture Models

3 A Novel Topic Model

4 Conclusion

Bayesian Mixture Models

Bayesian Mixture Models

- ▶ X_1, \dots, X_N are random variables in \mathbb{R}^d , some $d \in \mathbb{N}$.

Bayesian Mixture Models

- ▶ X_1, \dots, X_N are random variables in \mathbb{R}^d , some $d \in \mathbb{N}$.
- ▶ Classic **finite mixture model** of clustered data:

$$X_n | \pi, \mu \sim \sum_{j=1}^J \pi_j f(\cdot | \mu_j), \quad n = 1, \dots, N.$$

where:

- $\mu = (\mu_1, \dots, \mu_J)$ are random variables in $S \subset \mathbb{R}^p$.
- $\pi = (\pi_1, \dots, \pi_J)$ is a random probability vector.

Bayesian Mixture Models

- ▶ X_1, \dots, X_N are random variables in \mathbb{R}^d , some $d \in \mathbb{N}$.
- ▶ Classic **finite mixture model** of clustered data:

$$X_n | \pi, \mu \sim \sum_{j=1}^J \pi_j f(\cdot | \mu_j), \quad n = 1, \dots, N.$$

where:

- $\mu = (\mu_1, \dots, \mu_J)$ are random variables in $S \subset \mathbb{R}^p$.
 - $\pi = (\pi_1, \dots, \pi_J)$ is a random probability vector.
- ▶ Typically $\pi \sim \text{Dirichlet}_J(\gamma)$, some $\gamma > 0$, and $\mu_j \sim_{\text{i.i.d.}} p(\cdot)$.

Bayesian Mixture Models

- ▶ X_1, \dots, X_N are random variables in \mathbb{R}^d , some $d \in \mathbb{N}$.
- ▶ Classic **finite mixture model** of clustered data:

$$X_n | \pi, \mu \sim \sum_{j=1}^J \pi_j f(\cdot | \mu_j), \quad n = 1, \dots, N.$$

where:

- $\mu = (\mu_1, \dots, \mu_J)$ are random variables in $S \subset \mathbb{R}^p$.
- $\pi = (\pi_1, \dots, \pi_J)$ is a random probability vector.
- ▶ Typically $\pi \sim \text{Dirichlet}_J(\gamma)$, some $\gamma > 0$, and $\mu_j \sim_{\text{i.i.d.}} p(\cdot)$.
- ▶ Well-understood & predictable - any more interesting choices?

Normalised Random Measures²

²Beraha, M., Argiento, R., Camerlenghi, F. & Guglielmi, A. (2023).

Normalised Random Measures²

- ▶ View $\{\mu_1, \dots, \mu_J\}$ as a point process - borrow from theory of random measures.

²Beraha, M., Argiento, R., Camerlenghi, F. & Guglielmi, A. (2023).

Normalised Random Measures²

- ▶ View $\{\mu_1, \dots, \mu_J\}$ as a point process - borrow from theory of random measures.
- ▶ Our case of interest: take $\{\mu_1, \dots, \mu_J\} \sim \text{DPP}_J(K)$, where DPP stands for **determinantal point process**.

²Beraha, M., Argiento, R., Camerlenghi, F. & Guglielmi, A. (2023).

Normalised Random Measures²

- ▶ View $\{\mu_1, \dots, \mu_J\}$ as a point process - borrow from theory of random measures.
- ▶ Our case of interest: take $\{\mu_1, \dots, \mu_J\} \sim \text{DPP}_J(K)$, where DPP stands for **determinantal point process**.
- ▶ Consequence: the μ_j are repulsive, i.e. mixture components are well-separated.

²Beraha, M., Argiento, R., Camerlenghi, F. & Guglielmi, A. (2023).

Normalised Random Measures²

- ▶ View $\{\mu_1, \dots, \mu_J\}$ as a point process - borrow from theory of random measures.
- ▶ Our case of interest: take $\{\mu_1, \dots, \mu_J\} \sim \text{DPP}_J(K)$, where DPP stands for **determinantal point process**.
- ▶ Consequence: the μ_j are repulsive, i.e. mixture components are well-separated.
- ▶ Strength of repulsion is controlled by kernel hyper-parameter K .

²Beraha, M., Argiento, R., Camerlenghi, F. & Guglielmi, A. (2023).

Table of Contents

1 Latent Dirichlet Allocation

2 Repulsive Mixture Models

3 A Novel Topic Model

4 Conclusion

Motivation

Motivation

- ▶ Under LDA, all latent variables are independent and Dirichlet distributed.

Motivation

- ▶ Under LDA, all latent variables are independent and Dirichlet distributed.
- ▶ Might not be appropriate in some situations, leading to misspecification.

Motivation

- ▶ Under LDA, all latent variables are independent and Dirichlet distributed.
- ▶ Might not be appropriate in some situations, leading to misspecification.
- ▶ Repulsive mixture model uses very modern framework - never applied to topic modelling.

Motivation

- ▶ Under LDA, all latent variables are independent and Dirichlet distributed.
- ▶ Might not be appropriate in some situations, leading to misspecification.
- ▶ Repulsive mixture model uses very modern framework - never applied to topic modelling.
- ▶ *How does the behaviour of LDA change if we introduce a repulsive mixture component?*

Model Specification

Model Specification

- ▶ Model assumptions:
 - V words; T topics; N docs.
 - J possible topic mixtures.

Model Specification

- ▶ Model assumptions:
 - V words; T topics; N docs.
 - J possible topic mixtures.

$$\{\mu_1, \dots, \mu_J\} \sim \text{DPP}_J(K)$$

$$\pi \sim \text{Dirichlet}_J(\gamma)$$

$$\phi_t \sim \text{Dirichlet}_V(\eta), \quad t = 1, \dots, T$$

$$\theta_n | \mu, \pi \sim \text{Discrete}(\mu; \pi), \quad n = 1, \dots, N$$

$$z_{nm} | \theta_n \sim \text{Categorical}_T(\theta_n), \quad m = 1, \dots, M_n, \quad n = 1, \dots, N$$

$$w_{nm} | z_{nm}, \phi \sim \text{Categorical}_V(\phi_{z_{nm}}), \quad m = 1, \dots, M_n, \quad n = 1, \dots, N.$$

ϕ_t	t -th topic
ρ	mixing distribution (atoms are topic mixtures)
θ_n	topic mixture of n -th document
z_{nm}	topic index of m -th word in n -th document
w_{nm}	m -th word in n -th document

Model Specification

- ▶ Model assumptions:
 - V words; T topics; N docs.
 - J possible topic mixtures.

$$\{\mu_1, \dots, \mu_J\} \sim \text{DPP}_J(K)$$

$$\boldsymbol{\pi} \sim \text{Dirichlet}_J(\boldsymbol{\gamma})$$

$$\phi_t \sim \text{Dirichlet}_V(\boldsymbol{\eta}),$$

$$\theta_n | \boldsymbol{\mu}, \boldsymbol{\pi} \sim \text{Discrete}(\boldsymbol{\mu}; \boldsymbol{\pi}),$$

$$z_{nm} | \theta_n \sim \text{Categorical}_T(\theta_n),$$

$$w_{nm} | z_{nm}, \boldsymbol{\phi} \sim \text{Categorical}_V(\boldsymbol{\phi}_{z_{nm}}),$$

Reminder of LDA:

$$\phi_t \sim \text{Dirichlet}_V(\boldsymbol{\eta})$$

$$\theta_n \sim \text{Dirichlet}_T(\boldsymbol{\alpha})$$

$$z_{nm} | \theta_n \sim \text{Categorical}_T(\theta_n)$$

$$w_{nm} | z_{nm}, \boldsymbol{\phi} \sim \text{Categorical}_V(\boldsymbol{\phi}_{z_{nm}})$$

$$t = 1, \dots, T$$

$$n = 1, \dots, N$$

$$m = 1, \dots, M_n, \quad n = 1, \dots, N$$

$$m = 1, \dots, M_n, \quad n = 1, \dots, N.$$

ϕ_t	t -th topic
ρ	mixing distribution (atoms are topic mixtures)
θ_n	topic mixture of n -th document
z_{nm}	topic index of m -th word in n -th document
w_{nm}	m -th word in n -th document

Theoretical Properties

Theoretical Properties

1. Topic mixtures are (marginally) dependent:

- Topic mixtures coincide with positive probability.
- Distinct topic mixtures are repulsive.
- Documents are clustered based on their topic mixture.

Theoretical Properties

1. Topic mixtures are (marginally) dependent:

- Topic mixtures coincide with positive probability.
- Distinct topic mixtures are repulsive.
- Documents are clustered based on their topic mixture.

2. Topic mixtures have a non-parametric density:

- Density is proportional to $y \mapsto K(y, y)$, where K is DPP kernel.

Theoretical Properties

1. Topic mixtures are (marginally) dependent:

- Topic mixtures coincide with positive probability.
- Distinct topic mixtures are repulsive.
- Documents are clustered based on their topic mixture.

2. Topic mixtures have a non-parametric density:

- Density is proportional to $y \mapsto K(y, y)$, where K is DPP kernel.

3. Informative predictive distribution:

- Predictive distribution of θ_{N+1} given observed $\theta_1, \dots, \theta_N$ has two interesting behaviours.
- θ_{N+1} can coincide with one of the θ_n .
- θ_{N+1} can be distinct from the θ_n - tends to be very different.

Table of Contents

- 1 Latent Dirichlet Allocation
- 2 Repulsive Mixture Models
- 3 A Novel Topic Model
- 4 Conclusion

Conclusion

Conclusion

- ▶ Introduced **novel topic model** based on LDA and a repulsive mixture model.

Conclusion

- ▶ Introduced **novel topic model** based on LDA and a repulsive mixture model.
- ▶ Studied how theoretical properties compared to LDA.

Conclusion

- ▶ Introduced **novel topic model** based on LDA and a repulsive mixture model.
- ▶ Studied how theoretical properties compared to LDA.
- ▶ **Main advantage:** can achieve well-spaced clusters with customisable cluster locations.

Conclusion

- ▶ Introduced **novel topic model** based on LDA and a repulsive mixture model.
- ▶ Studied how theoretical properties compared to LDA.
- ▶ **Main advantage:** can achieve well-spaced clusters with customisable cluster locations.
- ▶ **Main disadvantage:** MCMC for posterior inference. Sampler proposed only for posterior of topics and topic mixtures.

Conclusion

- ▶ Introduced **novel topic model** based on LDA and a repulsive mixture model.
- ▶ Studied how theoretical properties compared to LDA.
- ▶ **Main advantage:** can achieve well-spaced clusters with customisable cluster locations.
- ▶ **Main disadvantage:** MCMC for posterior inference. Sampler proposed only for posterior of topics and topic mixtures.
- ▶ Future work:
 - Constructing kernel K on simplex.
 - Test efficacy of marginal sampler on real-life dataset.

References



Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003)

Latent dirichlet allocation

Journal of Machine Learning Research, 3(4-5), 993–1022



Beraha, M., Argiento, R., Camerlenghi, F. & Guglielmi, A. (2023)

Normalized random measures with interacting atoms for bayesian nonparametric mixtures

arXiv. [Preprint] <https://arxiv.org/abs/2302.09034>