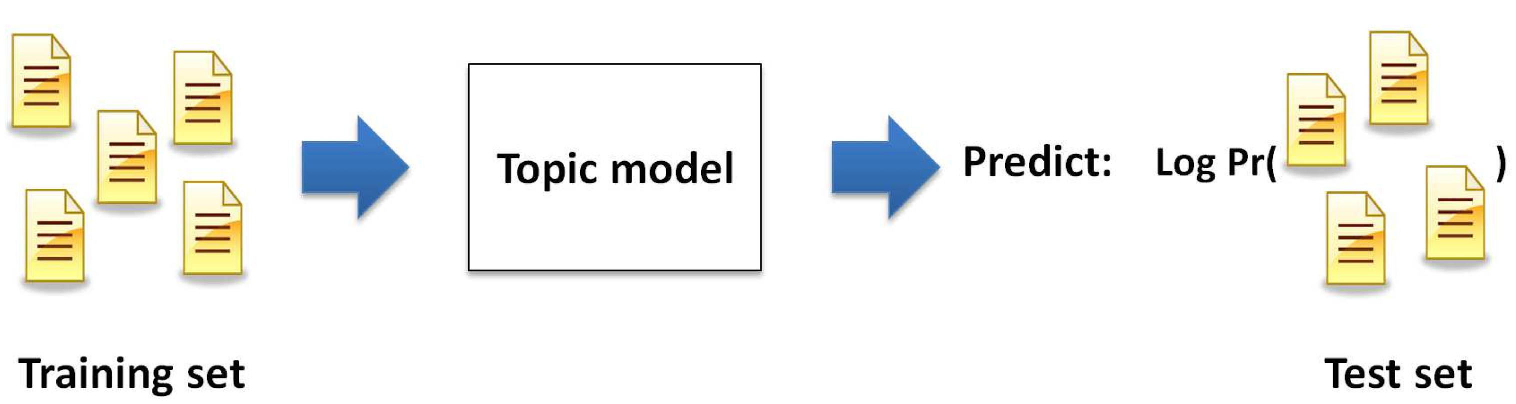
Annealing Paths for the Evaluation of Topic Models

James Foulds and Padhraic Smyth
University of California, Irvine

Abstract

- **Evaluating the predictive performance of topic models** is expensive and unreliable, despite recent advances in learning and inference algorithms.
- **Annealed importance sampling** (AIS), a Monte Carlo technique which operates by annealing between two distributions, has previously been successfully used for topic model evaluation.
- We introduce **new AIS annealing paths** which **anneal from one topic model to another**, thereby estimating the **relative predictive performance** of the models, and the **progress of topic model training algorithms**, efficiently and reliably.

Evaluating Topic Models

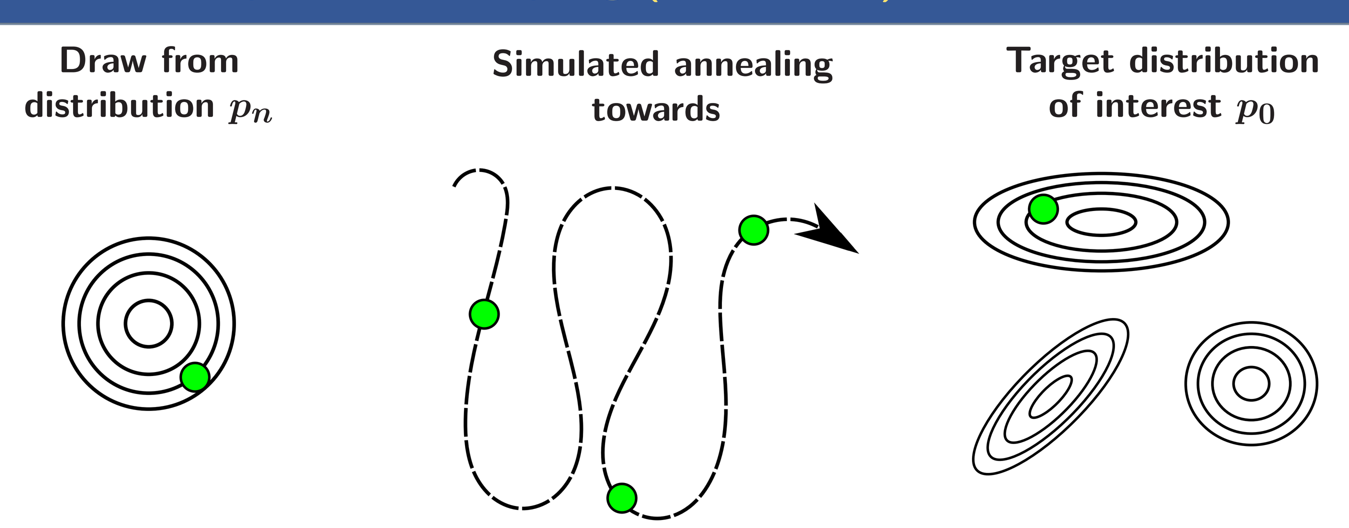


- For every held-out document d, computing the **likelihood** involves an **intractable sum** or an **intractable integral**,

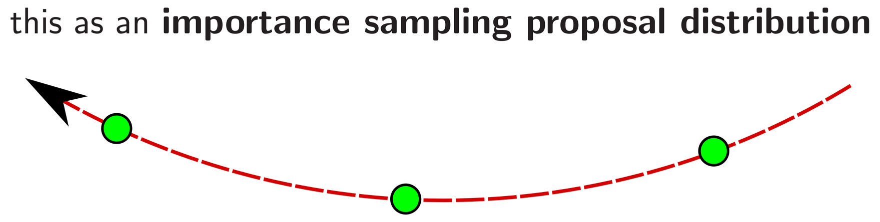
$$Pr(w^{(d)}|\Phi,\alpha) = \sum_{z^{(d)}} Pr(w^{(d)}, z^{(d)}|\Phi,\alpha)$$
$$= \int_{\theta} Pr(w^{(d)}, \theta^{(d)}|\Phi,\alpha) \,.$$

- We need to approximate this for each of possibly **tens of thousands** of documents in the test set!

Annealed Importance Sampling (Neal, 2001)



Use this as an **importance sampling proposal distribution** for:



Annealing in the reverse direction, from the **target** to the **source**.

- The importance samples can be used to estimate the ratio of normalizing constants of $f_0 \propto p_0$ and $f_n \propto p_n$.

$$\frac{\sum w^{(i)}}{N} \Rightarrow \frac{\int f_0(x)dx}{\int f_n(x)dx}$$

- Wallach *et al.* (2009) show how to employ **AIS in the context of topic models** to estimate $Pr(w^{(d)}|\Phi, \alpha^{(d)})$:
  - Perform AIS on the topic assignments $z^{(d)}$, collapsing out $\theta^{(d)}$.
  - Anneal from the **prior** to the **posterior**.
    - Draw initial state from the prior over z, $f_n = Pr(z^{(d)}|\alpha^{(d)})$.
    - Anneal towards a distribution proportional to the posterior, $f_0 = Pr(w^{(d)}, z^{(d)}|\phi, \alpha^{(d)})$.
- Estimate the likelihood by taking the **average of the importance weights**:

$$\frac{\sum w^{(i)}}{N} \Rightarrow \frac{\sum_{z^{(d)}} Pr(w^{(d)}, z^{(d)}|\phi,\alpha^{(d)})}{\sum_{z^{(d)}} Pr(z^{(d)}|\alpha^{(d)})}$$
$$= \frac{Pr(w^{(d)}|\phi,\alpha^{(d)})}{1} = Pr(w^{(d)}|\phi,\alpha^{(d)}) \,.$$

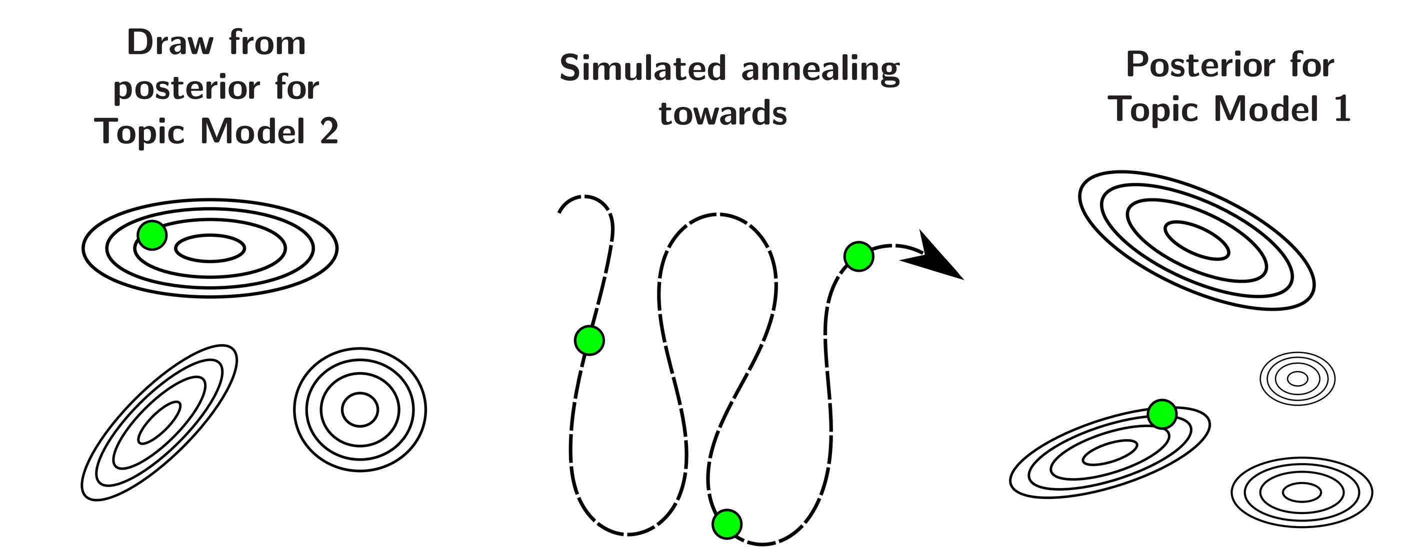
Ratio-AIS for Comparing Topic Models

- Typically for evaluation we are interested in the **relative** performance of topic model 1 (e.g. a new model) and topic model 2 (e.g. vanilla LDA):

$$\log Pr(w^{(d)}|\phi^{(1)},\alpha^{(d,1)}) - \log Pr(w^{(d)}|\phi^{(2)},\alpha^{(d,2)})$$
$$= \log \frac{Pr(w^{(d)}|\phi^{(1)},\alpha^{(d,1)})}{Pr(w^{(d)}|\phi^{(2)},\alpha^{(d,2)})} \,.$$

- This could be estimated by running AIS **once for each model**. However, AIS is already designed to compute a ratio. We can **estimate this ratio directly**.

Ratio-AIS



$$f_n = Pr(w^{(d)}, z^{(d)}|\phi^{(2)},\alpha^{(d,2)}) \qquad f_0 = Pr(w^{(d)}, z^{(d)}|\phi^{(1)},\alpha^{(d,1)})$$

$$\sum \frac{w^{(i)}}{N} \Rightarrow \frac{Pr(w^{(d)}|\phi^{(1)},\alpha^{(d,1)})}{Pr(w^{(d)}|\phi^{(2)},\alpha^{(d,2)})}$$

- It remains to choose the sequence of intermediate distributions. We consider two alternatives:

Geometric Averages Path

$$f_j(z^{(d)}) = f_0(z^{(d)})^{\beta_j} f_n(z^{(d)})^{1-\beta_j}$$

Convex Combinations Path

$$f_j(z^{(d)}) = Pr(w^{(d)}, z^{(d)}|\Phi_j = \beta_j\Phi^{(1)} + (1-\beta_j)\Phi^{(2)}$$
$$\alpha_j = \beta_j\alpha^{(1)} + (1-\beta_j)\alpha^{(2)}) \,.$$

- This approach **avoids several sources of Monte Carlo error** incurred by running AIS for each model separately. Specifically, the standard method
  - estimates the denominator of a ratio even though it is a constant (=1),
  - uses different z's for both models,
  - and is run twice, introducing Monte Carlo noise each time.
- An easy **convergence check**: anneal in the reverse direction to compute the reciprocal.

Iteration-AIS for Evaluating Topic Model Learning Algorithms

- We evaluate learning algorithms by using the **learned model at each iteration as an AIS intermediate distribution**, and using ratio-AIS to **anneal between each successive model**.
- This provides a **warm-start** with successively more effective temperatures, potentially leading to better estimates as the learning algorithm proceeds.

Iteration-AIS

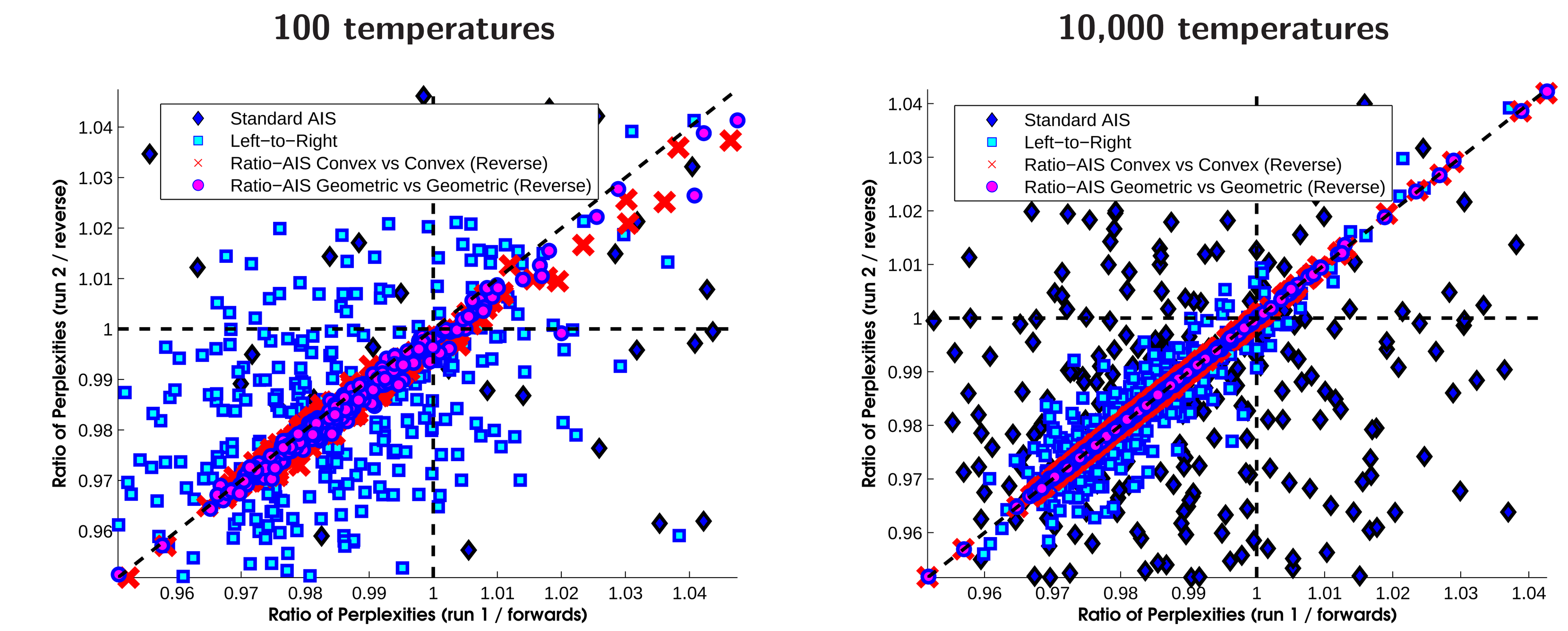
Prior	$\Rightarrow$	Model at Iteration 1	$\Rightarrow$	Model at Iteration 2	$\Rightarrow \ldots \Rightarrow$	Model at Iteration I
	Wallach et al.		*Ratio-AIS*		*Ratio-AIS*	

$$\log w^{(i,t)} = \log w^{(i,t-1)} + \sum_{j=0}^{n-1} \log \frac{f_{t,j}(z_{t,j})}{f_{t,j+1}(z_{t,j})}$$

Experimental Analysis on the NIPS and ACL Corpora

Comparing Learned Topics with Perturbed Topics

- **Dots below 1**: Unperturbed topics are better (likely **correct**)
- **Dots on the diagonal**: Two repeated runs of the method produce the same perplexity ratio

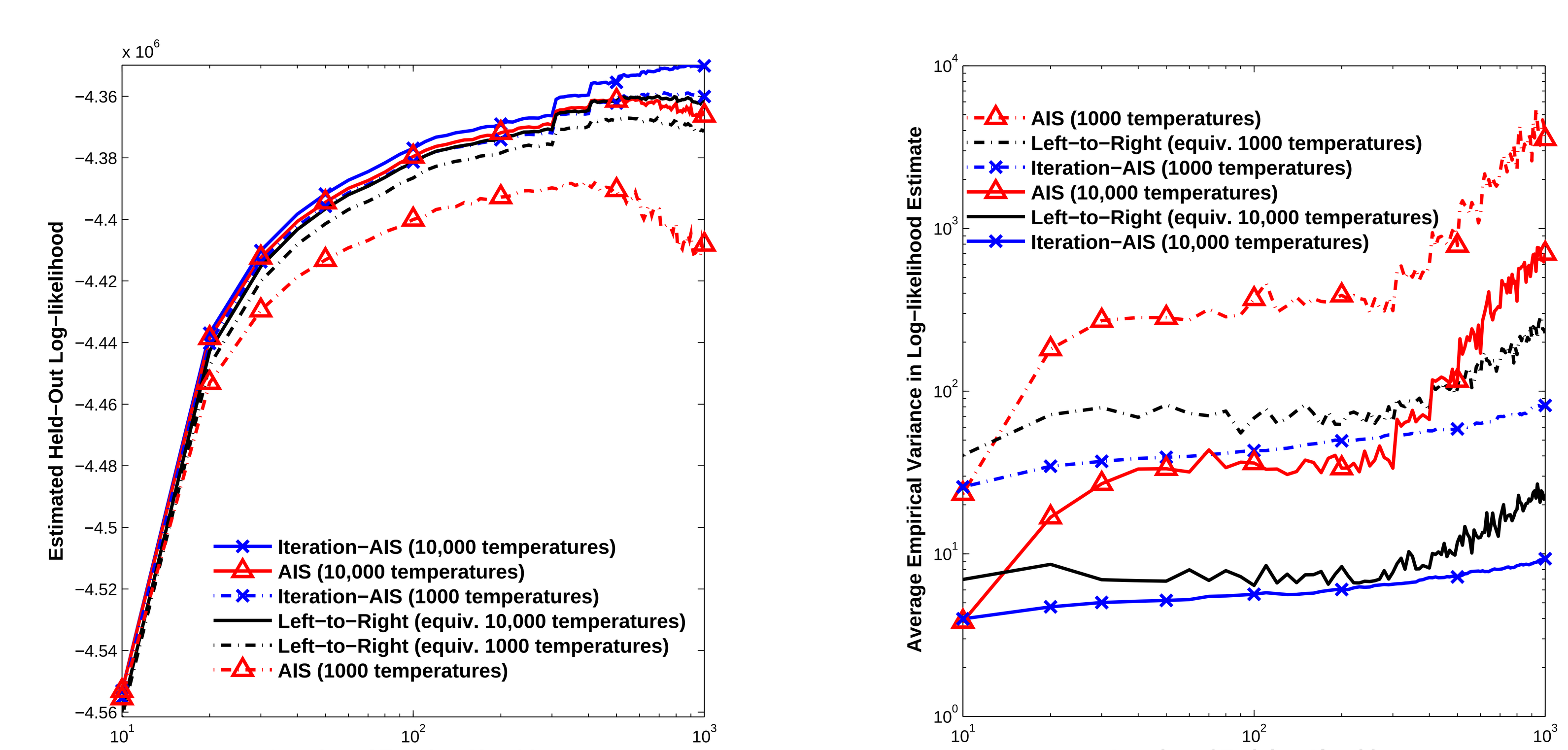


% Correct	Left to Right	Standard AIS	Ratio-AIS Geometric	Ratio-AIS Geom. (reverse)	Ratio-AIS Convex	Ratio-AIS Convex (reverse)
NIPS (cheap)	63.8	48.8	83.8	**89.2**	84.6	87.7
NIPS (expensive)	84.6	62.3	86.9	**87.7**	**87.7**	**87.7**
ACL (cheap)	80.2	50.8	88.3	92.0	88.3	**92.3**
ACL (expensive)	**90.7**	75.2	90.3	90.3	90.3	90.3

Comparing Asymmetric α and Symmetric α Topic Models

Correlation with Long LR Run	Left to Right	Standard AIS	Ratio-AIS Geometric	Ratio-AIS Geom. (reverse)	Ratio-AIS Convex	Ratio-AIS Convex (reverse)
NIPS (cheap)	0.947	0.619	0.973	0.975	0.976	**0.981**
NIPS (expensive)	**0.993**	0.852	0.981	0.982	0.981	0.982
ACL (cheap)	0.965	0.578	0.984	0.983	**0.987**	0.986
ACL (expensive)	**0.995**	0.892	0.989	0.989	0.990	0.989
Variance of Perplexity Ratio	Left to Right	Standard AIS	Ratio-AIS Geometric	Ratio-AIS Geom. (reverse)	Ratio-AIS Convex	Ratio-AIS Convex (reverse)
NIPS (cheap)	2.6×10^{-4}	2.6×10^{-3}	2.0×10^{-5}	1.5×10^{-5}	$\mathbf{8.2 \times 10^{-6}}$	9.8×10^{-6}
NIPS (expensive)	1.7×10^{-5}	6.0×10^{-4}	1.4×10^{-6}	1.2×10^{-6}	6.9×10^{-7}	$\mathbf{5.8 \times 10^{-7}}$
ACL (cheap)	1.7×10^{-4}	3.6×10^{-3}	1.6×10^{-5}	1.3×10^{-5}	7.7×10^{-6}	$\mathbf{6.6 \times 10^{-6}}$
ACL (expensive)	1.4×10^{-5}	5.6×10^{-4}	1.1×10^{-6}	9.4×10^{-7}	7.4×10^{-7}	$\mathbf{5.1 \times 10^{-7}}$
Corpus-Level Perplexity Ratio	Left to Right	Standard AIS	Ratio-AIS Geometric	Ratio-AIS Geom. (reverse)	Ratio-AIS Convex	Ratio-AIS Convex (reverse)
NIPS (cheap)	0.984	**0.975**	1.01	0.992	1.01	0.994
NIPS (expensive)	**0.989**	0.990	1.00	0.999	1.00	0.998
ACL (cheap)	0.984	**0.980**	1.00	0.985	1.00	0.988
ACL (expensive)	**0.987**	0.989	0.994	0.992	0.996	0.992

Evaluating Iteration-AIS



References

Neal, R.M. 2001. Annealed importance sampling. *Statistics and Computing*, **11**(2), 125–139.
Wallach, H.M., Murray, I., Salakhutdinov, R., & Mimno, D. 2009. Evaluation methods for topic models. *ICML*.

{jfoulds, smyth}@ics.uci.edu