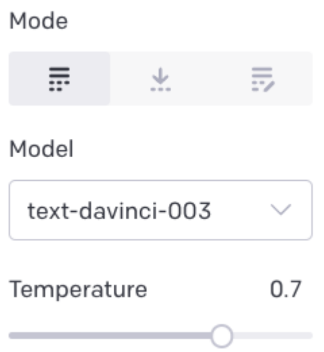
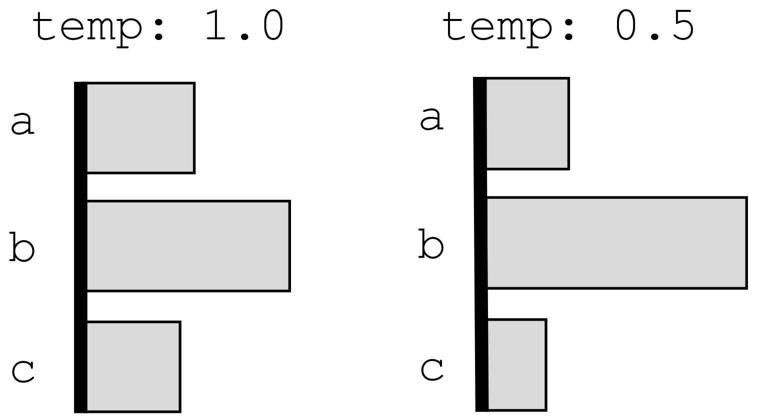
Long Horizon Temperature Scaling

Andy Shih, Dorsa Sadigh, Stefano Ermon

andyshih@cs.stanford.edu



T < 1 biases sampling towards high likelihood regions



$\log p_T(x) = \log p(x)/T - \log Z_{p_T}$

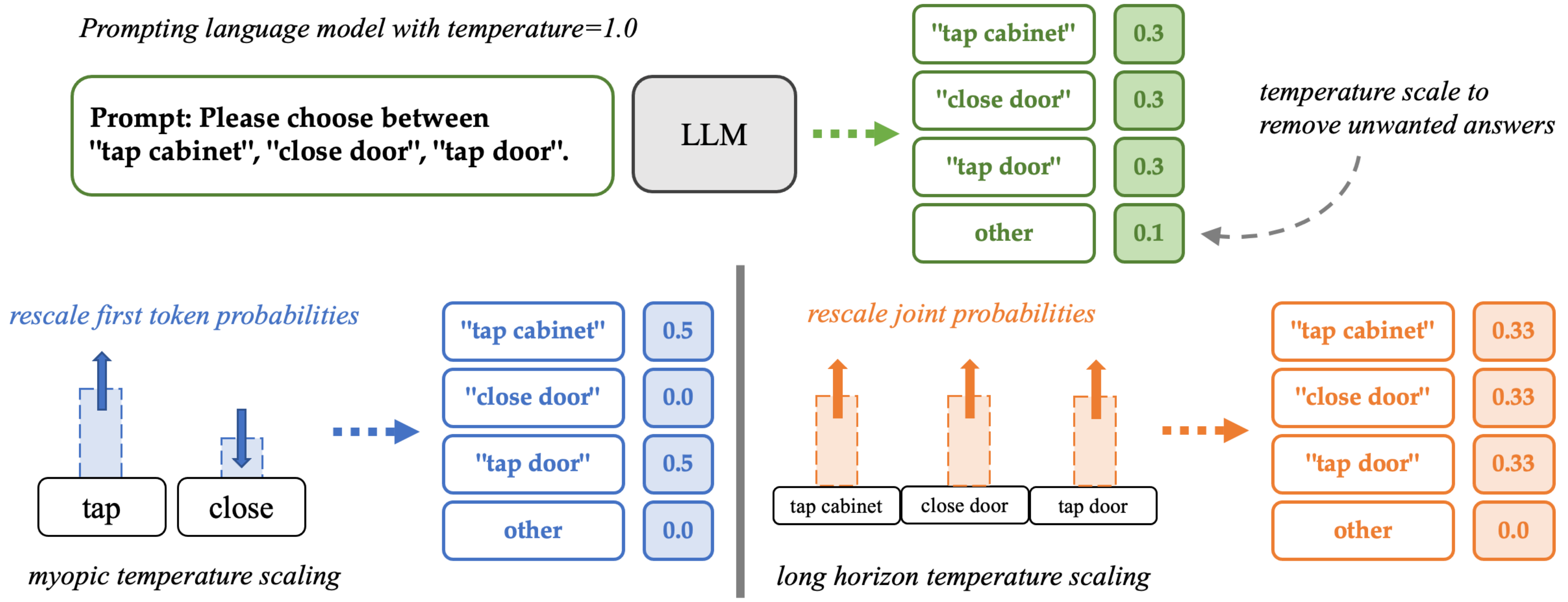
But, greedy decoding has a problem…

$$\log p_T(x) \neq \sum_i \log p_T^{\text{myopic}}(x_i|x_{<i})$$

How are you $\neq$ The thing is

non-myopic argmax — myopic argmax, myopic argmax, myopic argmax

Prompting language model with temperature=1.0



$\hat{p}$	p		q_T
data	model	**distill** →	temperature scaled model

Want: $\log p_T(x) = \log p(x)/T - \log Z_{p_T}$

$$D_{KL}(p_T||q_T) = \mathbb{E}_{x\sim p_T}\left[\frac{\log p(x)}{T} - \log q_T(x)\right] - \log Z_{p_T}$$

$$-\mathbb{E}_{x\sim p_T}[\log q_T(x)]$$

importance sampling

$$-\mathbb{E}_{x\sim p}\frac{e^{\log p(x)/T - \log Z_{p_T}}}{p(x)}[\log q_T(x)]$$

speed up by using data instead of samples

$$-\mathbb{E}_{x\sim\hat{p}}\exp\left(\frac{1-T}{T}\log p(x)\right)[\log q_T(x)]$$

Non-myopic

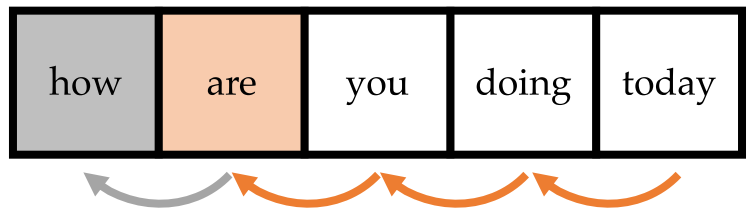
Applicable to all likelihood-based models

Learnable Baseline

multiplicative constant

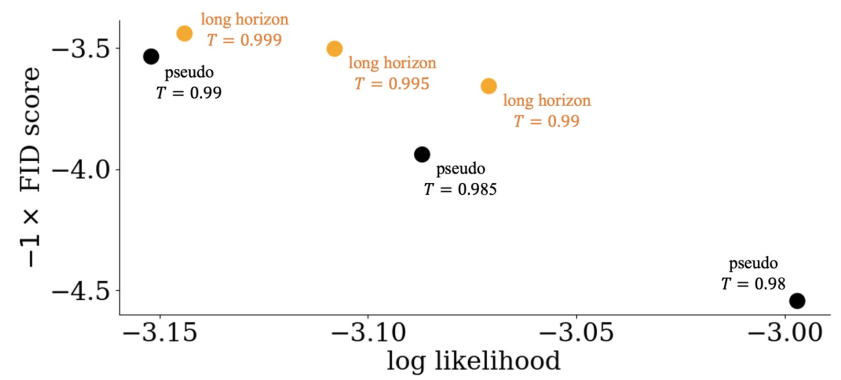
$$-\mathbb{E}_{x\sim p}\frac{e^{\log p(x)/T - b}}{p(x)}[\log q_T(x)] \qquad b = \frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}}\frac{1-T}{T}\log p(x)$$

Suffix likelihood and Index-dependent Baseline (for AR models)

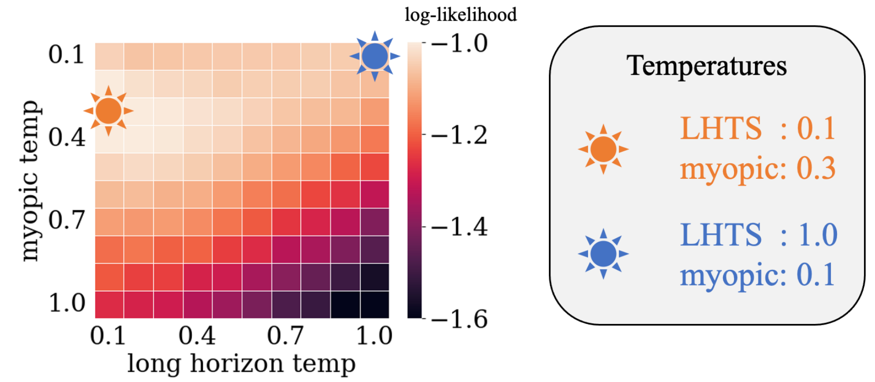


$$b(i) = \frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}}\frac{1-T}{T}\log p(x_{\geq i}|x_{<i})$$

Diffusion Models



Autoregressive Character Models




```
'the proces', ' internati', 'ne nine fi',
'n the latt', ' the const', ' and the m',
'is the fir', 'e three fi', ' of the ma',
```

Likelihood: -0.97
Diversity: Higher

```
' the const', ' the state', ' the state',
' the commu', ' the state', ' the commo',
' the same ', ' the state', ' the south',
```

Likelihood: -1.05
Diversity: Lower

Language Models

model myopic T	gpt2 large 1.0	 0.5	 0.0
LHTS $T = 0.9$	0.249	0.310	**0.317**
pretrained	0.203	0.279	0.290
partition (Quark)	0.213	0.279	0.285