Path Space Probabilistic Inference

*generated by ChatGPT

Fantastic Path RND and where to find them

- **Fantastic Path RND:** FF-RND, FB-RND

- **Where to find them?**
  - Importance Sampling:
    - Free-energy estimation, density estimation, SMC
  - Parallel Tempering
  - Variational Inference

Fantastic Path RND and where to find them

- **Fantastic Path RND:** FF-RND, FB-RND

- **Where to find them?**
  - Importance Sampling:
    - **Free-energy estimation**, density estimation, SMC
  - **Parallel Tempering**
  - Variational Inference

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Importance sampling / resampling:

e.g., free-energy estimation:

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Importance sampling / resampling:

e.g., free-energy estimation:

$$\Delta F = -\log \frac{Z_p}{Z_q} = -\log \frac{\int \tilde{p}(x) \mathrm{d}x}{Z_q}$$

$$= -\log\left(\frac{1}{Z_q} \int \tilde{q}(x) \frac{\tilde{p}(x)}{\tilde{q}(x)} \mathrm{d}x\right)$$

$$= -\log\left(\int \frac{\tilde{q}(x)}{Z_q} \frac{\tilde{p}(x)}{\tilde{q}(x)} \mathrm{d}x\right)$$

$$= -\log\left(\int q(x) \frac{\tilde{p}(x)}{\tilde{q}(x)} \mathrm{d}x\right)$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Importance sampling / resampling:

e.g., free-energy estimation:

$$\Delta F = -\log \frac{Z_p}{Z_q} = -\log \frac{\int \tilde{p}(x)\mathrm{d}x}{Z_q}$$
$$= -\log(\frac{1}{Z_q}\int \tilde{q}(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x)$$
$$= -\log(\int \frac{\tilde{q}(x)}{Z_q}\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x)$$
$$= -\log(\int q(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x)$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Importance sampling / resampling:

e.g., free-energy estimation:

$$\begin{aligned}
\Delta F &= -\log \frac{Z_p}{Z_q} = -\log \frac{\int \tilde{p}(x)\mathrm{d}x}{Z_q} \\
&= -\log\left(\frac{1}{Z_q}\int \tilde{q}(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right) \\
&= -\log\left(\int \frac{\tilde{q}(x)}{Z_q}\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right) \\
&= -\log\left(\int q(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right)
\end{aligned}$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Importance sampling / resampling:

e.g., free-energy estimation:

$$
\begin{aligned}
\Delta F &= -\log \frac{Z_p}{Z_q} = -\log \frac{\int \tilde{p}(x)\mathrm{d}x}{Z_q} \\
&= -\log\left(\frac{1}{Z_q}\int \tilde{q}(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right) \\
&= -\log\left(\int \frac{\tilde{q}(x)}{Z_q}\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right) \\
&= -\log\left(\int q(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right)
\end{aligned}
$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Importance sampling / resampling:

e.g., free-energy estimation:

$$
\begin{aligned}
\Delta F &= -\log \frac{Z_p}{Z_q} = -\log \frac{\int \tilde{p}(x) \mathrm{d}x}{Z_q} \\
&= -\log\left(\frac{1}{Z_q} \int \tilde{q}(x) \frac{\tilde{p}(x)}{\tilde{q}(x)} \mathrm{d}x\right) \\
&= -\log\left(\int \frac{\tilde{q}(x)}{Z_q} \frac{\tilde{p}(x)}{\tilde{q}(x)} \mathrm{d}x\right) \\
&= -\log\left(\int q(x) w(x)\, \mathrm{d}x\right)
\end{aligned}
$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Free-energy Perturbation

- Importance sampling / resampling:

e.g., free-energy estimation:

$$\Delta F = -\log \frac{Z_p}{Z_q} = -\log \frac{\int \tilde{p}(x)\mathrm{d}x}{Z_q}$$
$$= -\log\left(\frac{1}{Z_q}\int \tilde{q}(x)\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right)$$
$$= -\log\left(\int \frac{\tilde{q}(x)}{Z_q}\frac{\tilde{p}(x)}{\tilde{q}(x)}\mathrm{d}x\right)$$
$$= -\log\left(\int q(x)w(x)\,\mathrm{d}x\right)$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Parallel Tempering:

Parallel tempering

- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

Parallel tempering

- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

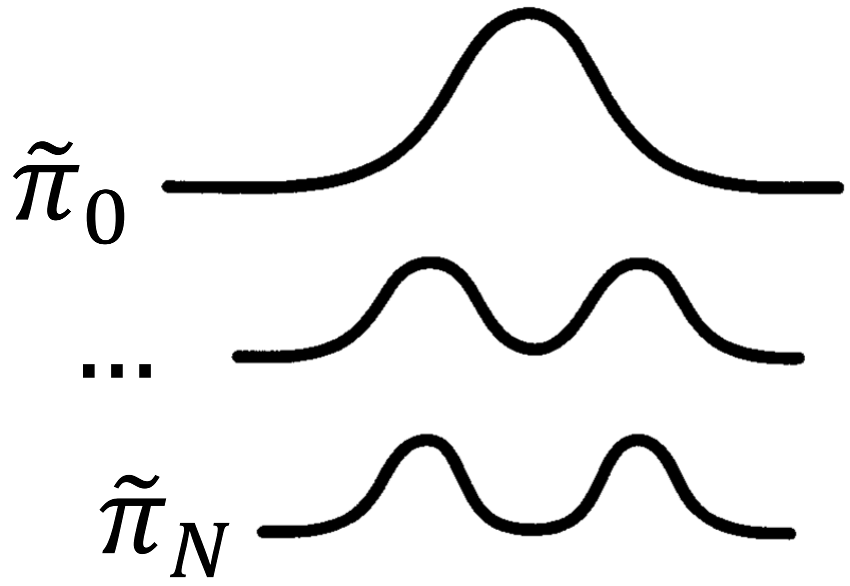
Parallel tempering

- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

Parallel tempering

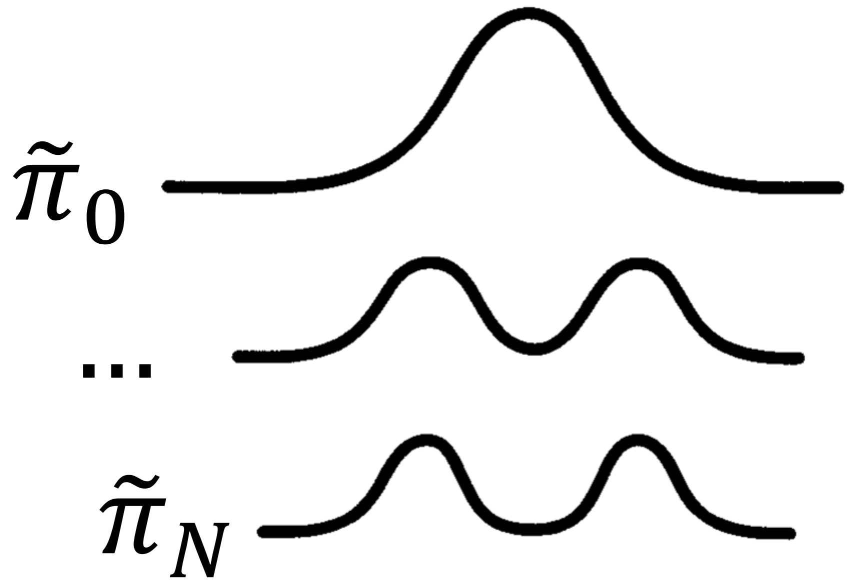
- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

Parallel tempering



- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

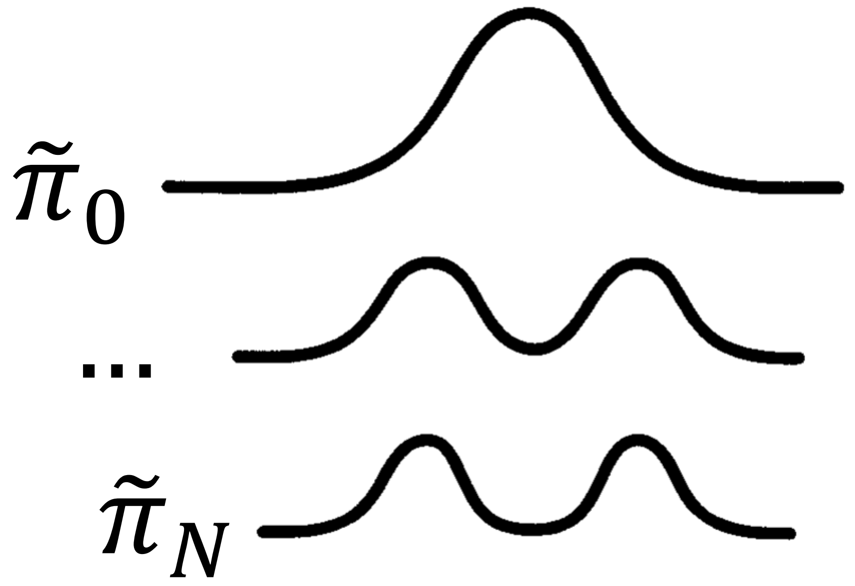
Parallel tempering



- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

1. Local exploration kernel: independent MCMC for each $\tilde{\pi}_n$

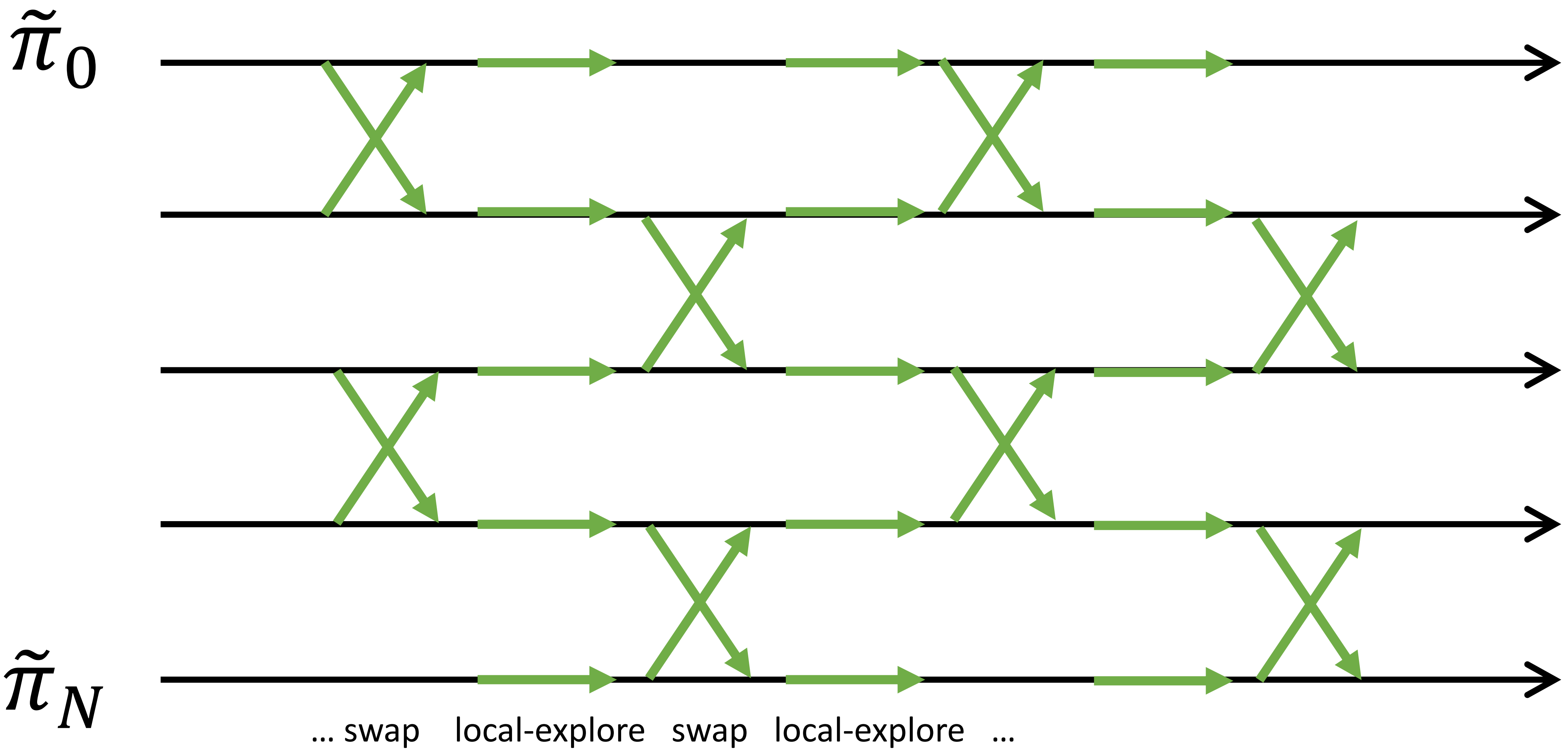
Parallel tempering



- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

1. Local exploration kernel: independent MCMC for each $\tilde{\pi}_n$
2. Communication kernel: swap between all adjacent pairs $(\tilde{\pi}_n, \tilde{\pi}_{n+1})$

Parallel tempering



Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

- Communication in PT:

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density $\tilde{p}(x) \times \tilde{q}(y)$?**

- Communication in PT:

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density $\tilde{p}(x) \times \tilde{q}(y)$?**

- Communication in PT:
  - 💡 current state $(x, y) \sim \tilde{p}(x)\tilde{q}(y)$
  - 💡 Proposal $(x', y') \leftarrow (y, x)$
  - 💡 next state $(x', y') \sim \tilde{p}(x)\tilde{q}(y)$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density $\tilde{p}(x) \times \tilde{q}(y)$?**

- Communication in PT:

💡 current state $(x, y) \sim \tilde{p}(x)\tilde{q}(y)$

💡 Proposal $(x', y') \leftarrow (y, x)$

💡 next state $(x', y') \sim \tilde{p}(x)\tilde{q}(y)$

informally

$$\alpha = \min\{1, \frac{\tilde{p}(x')\tilde{q}(y')\delta_{x,y}(x', y')}{\tilde{p}(x)\tilde{q}(y)\delta_{x',y'}(x, y)}\}$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density $\tilde{p}(x) \times \tilde{q}(y)$?**

- Communication in PT:

💡 current state $(x, y) \sim \tilde{p}(x)\tilde{q}(y)$

💡 Proposal $(x', y') \leftarrow (y, x)$

💡 next state $(x', y') \sim \tilde{p}(x)\tilde{q}(y)$

informally

$$\alpha = \min\{1, \frac{\tilde{p}(x')\tilde{q}(y')\delta_{x,y}(x', y')}{\tilde{p}(x)\tilde{q}(y)\delta_{x',y'}(x, y)}\}$$

Requirement: proposal is involution $f(f(x)) = x$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density** $\tilde{p}(x) \times \tilde{q}(y)$ **?**

- Communication in PT:

💡 current state $(x, y) \sim \tilde{p}(x)\tilde{q}(y)$

💡 Proposal $(x', y') \leftarrow (y, x)$

💡 next state $(x', y') \sim \tilde{p}(x)\tilde{q}(y)$

$$\alpha = \min\{1, \frac{\tilde{p}(x')\tilde{q}(y')}{\tilde{p}(x)\tilde{q}(y)}\}$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density $\tilde{p}(x) \times \tilde{q}(y)$?**

- Communication in PT:

💡 current state $(x, y) \sim \tilde{p}(x)\tilde{q}(y)$

💡 Proposal $(x', y') \leftarrow (y, x)$

💡 next state $(x', y') \sim \tilde{p}(x)\tilde{q}(y)$

$$\alpha = \min\{1, \frac{\tilde{p}(x')\tilde{q}(y')}{\tilde{p}(x)\tilde{q}(y)}\}$$

$$= \min\{1, \frac{\tilde{p}(y)\tilde{q}(x)}{\tilde{p}(x)\tilde{q}(y)}\}$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density $\tilde{p}(x) \times \tilde{q}(y)$?**

- Communication in PT:

  💡 current state $(x, y) \sim \tilde{p}(x)\tilde{q}(y)$

  💡 Proposal $(x', y') \leftarrow (y, x)$

  💡 next state $(x', y') \sim \tilde{p}(x)\tilde{q}(y)$

$$\alpha = \min\{1, \frac{\tilde{p}(x')\tilde{q}(y')}{\tilde{p}(x)\tilde{q}(y)}\}$$

$$= \min\{1, \frac{\tilde{p}(y)\tilde{q}(x)}{\tilde{p}(x)\tilde{q}(y)}\}$$

Density Ratio and where to find them

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Assume replica m has density $\tilde{p}$, replica $m+1$ has density $\tilde{q}$, **how to construct a MCMC "swap" kernel with invariant density $\tilde{p}(x) \times \tilde{q}(y)$?**

- Communication in PT:
  - 💡 current state $(x, y) \sim \tilde{p}(x)\tilde{q}(y)$
  - 💡 Proposal $(x', y') \leftarrow (y, x)$
  - 💡 next state $(x', y') \sim \tilde{p}(x)\tilde{q}(y)$

$$\alpha = \min\{1, \frac{\tilde{p}(x')\tilde{q}(y')}{\tilde{p}(x)\tilde{q}(y)}\}$$

$$= \min\{1, \frac{w(y)}{w(x)}\}$$

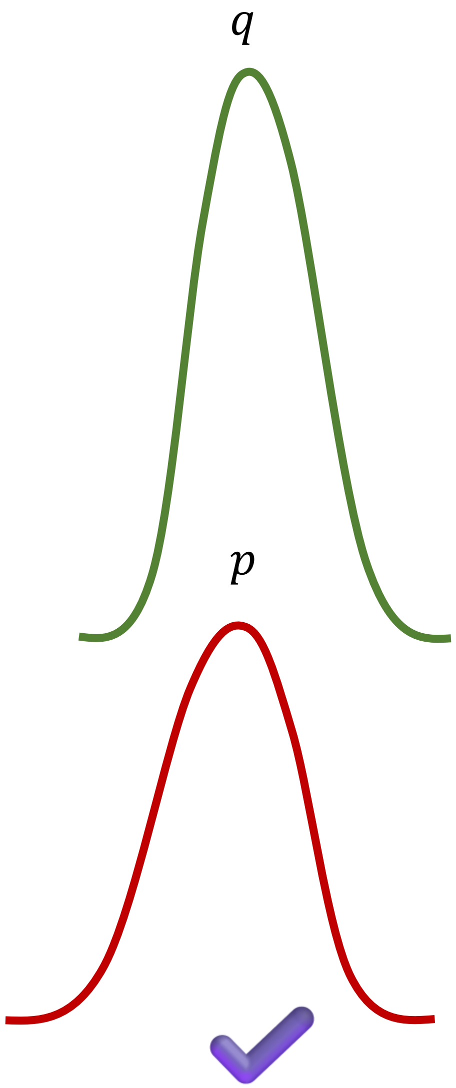
Wrap up

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

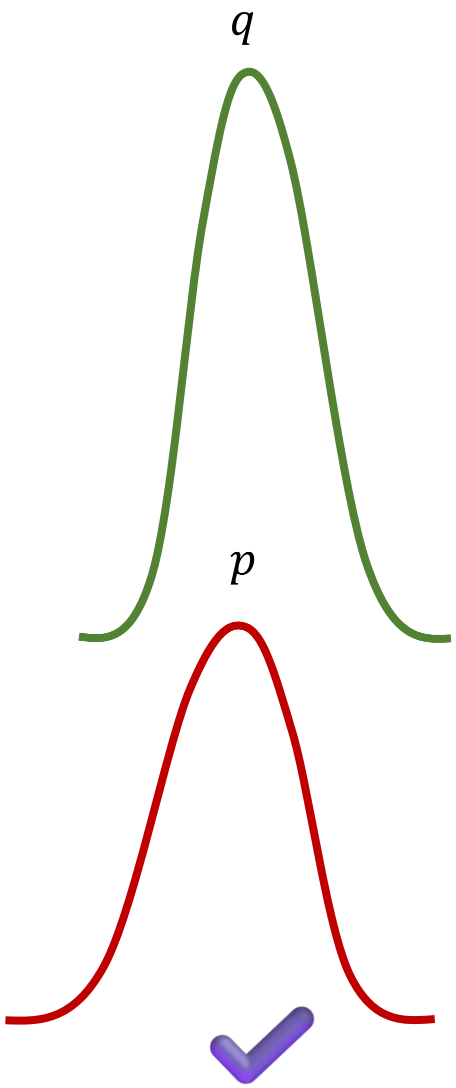
Density ratio: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$

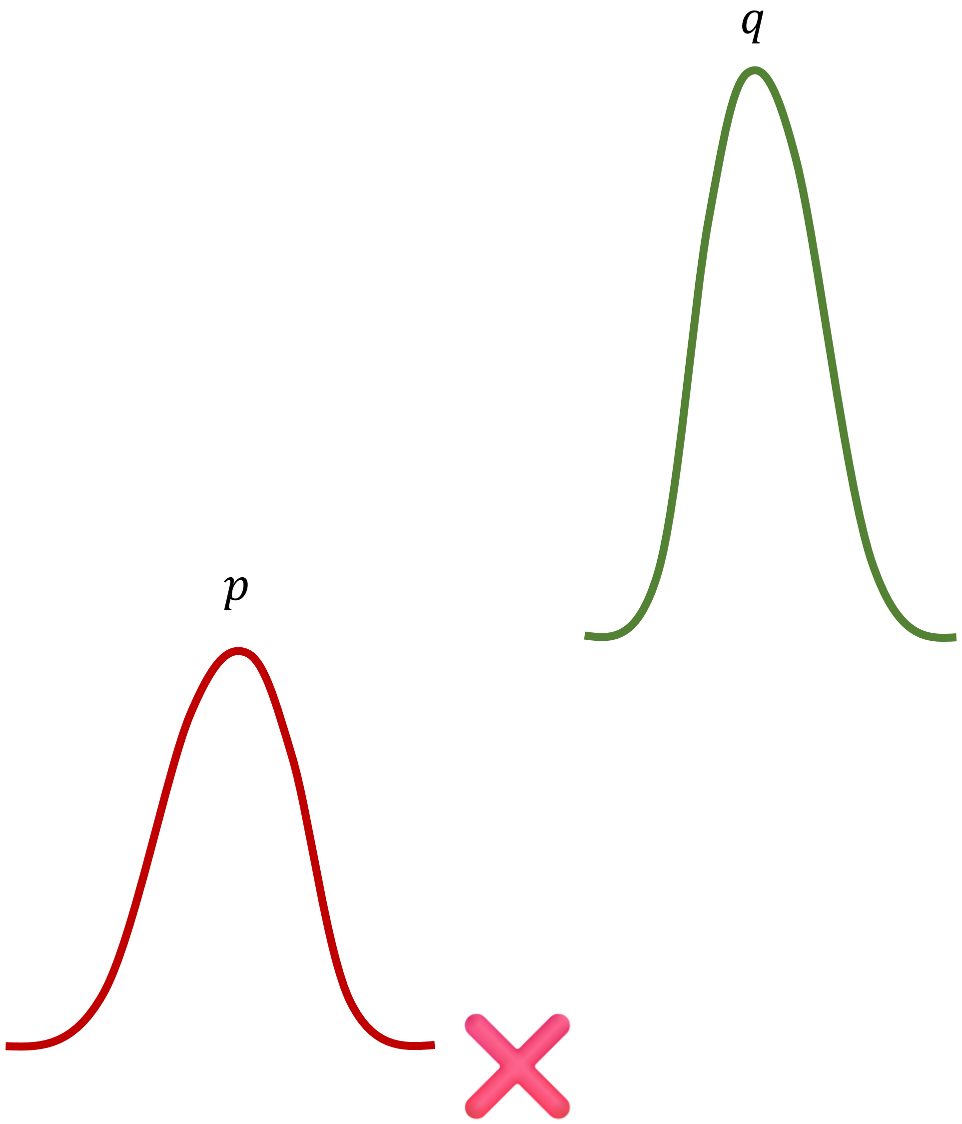
- Importance sampling: $w(x) = \dfrac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\, \mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \dfrac{w(y)}{w(x)}\}$

Limitations

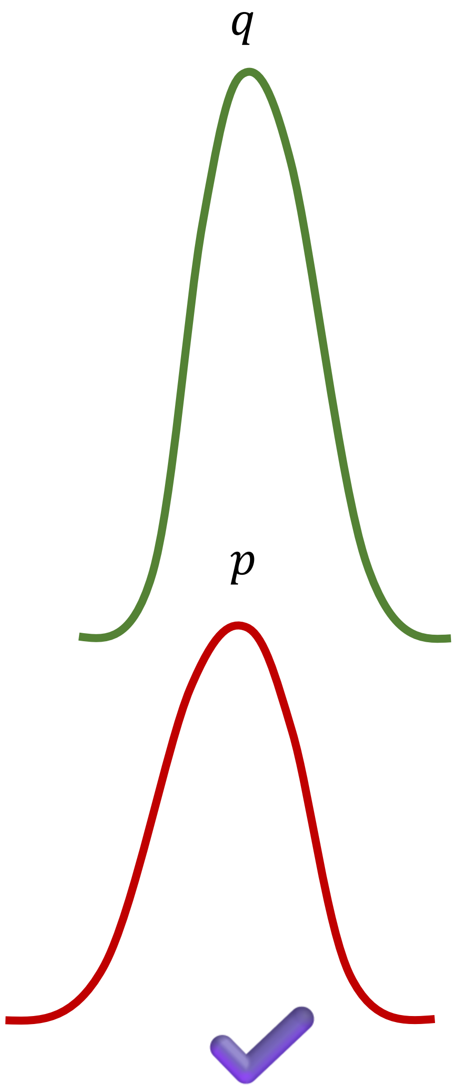


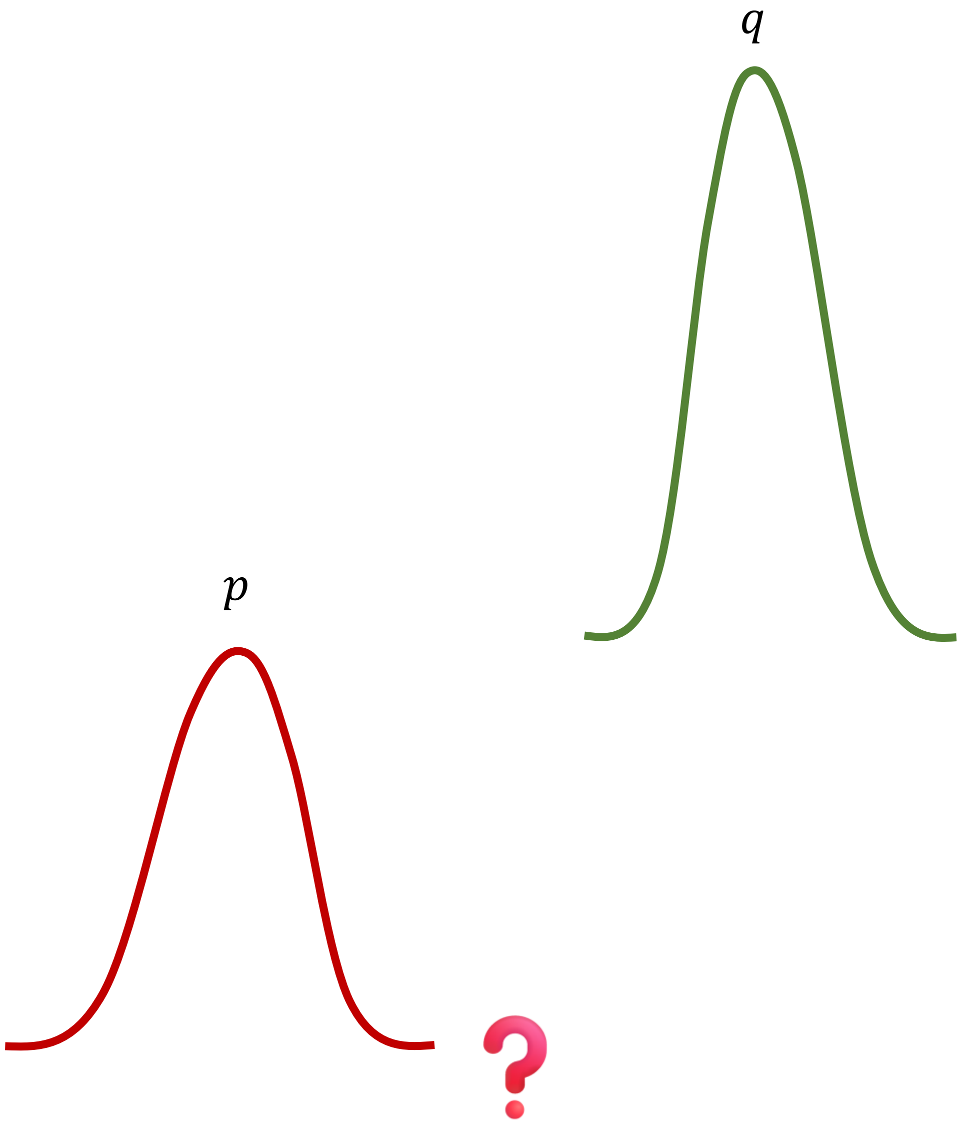
Limitations



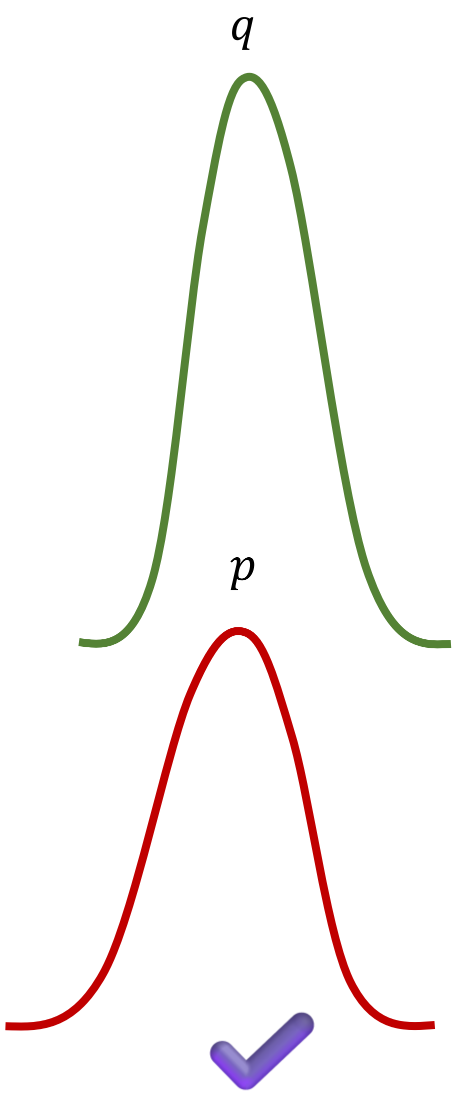


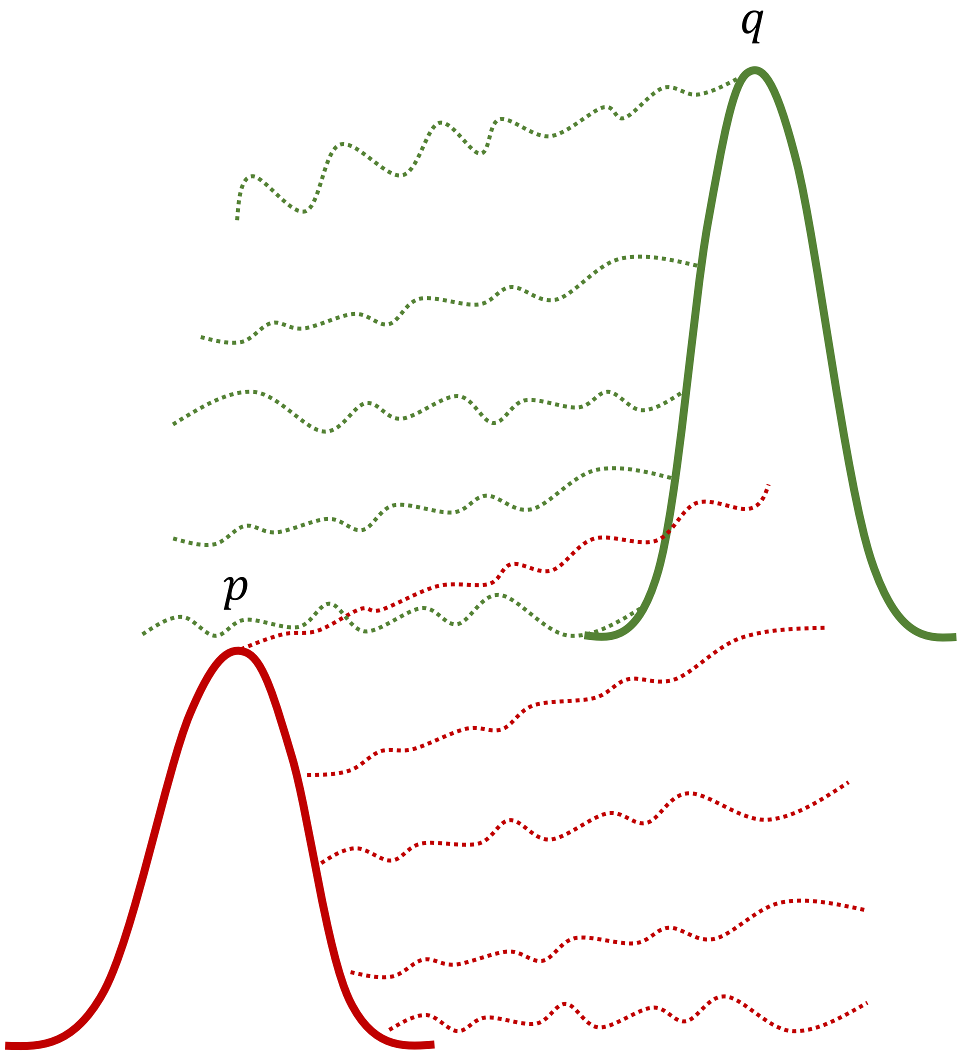
Limitations



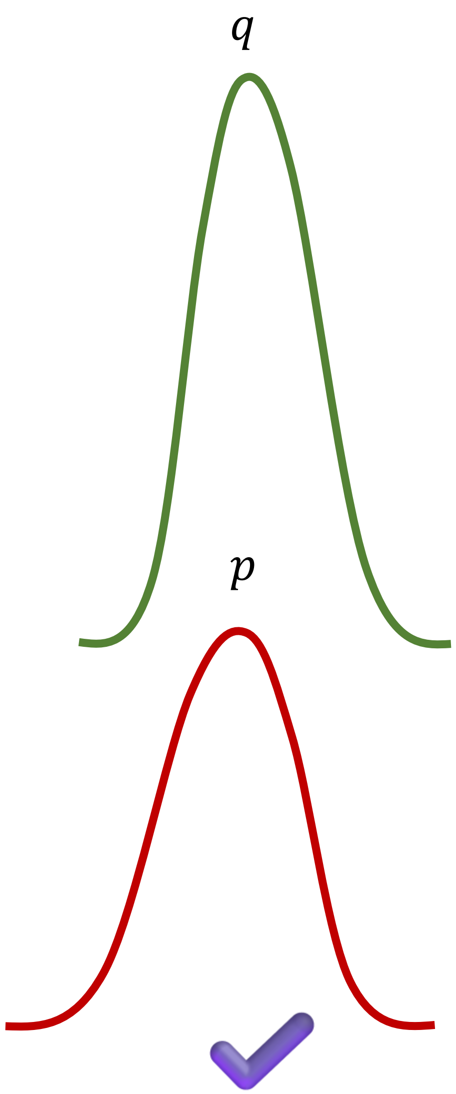


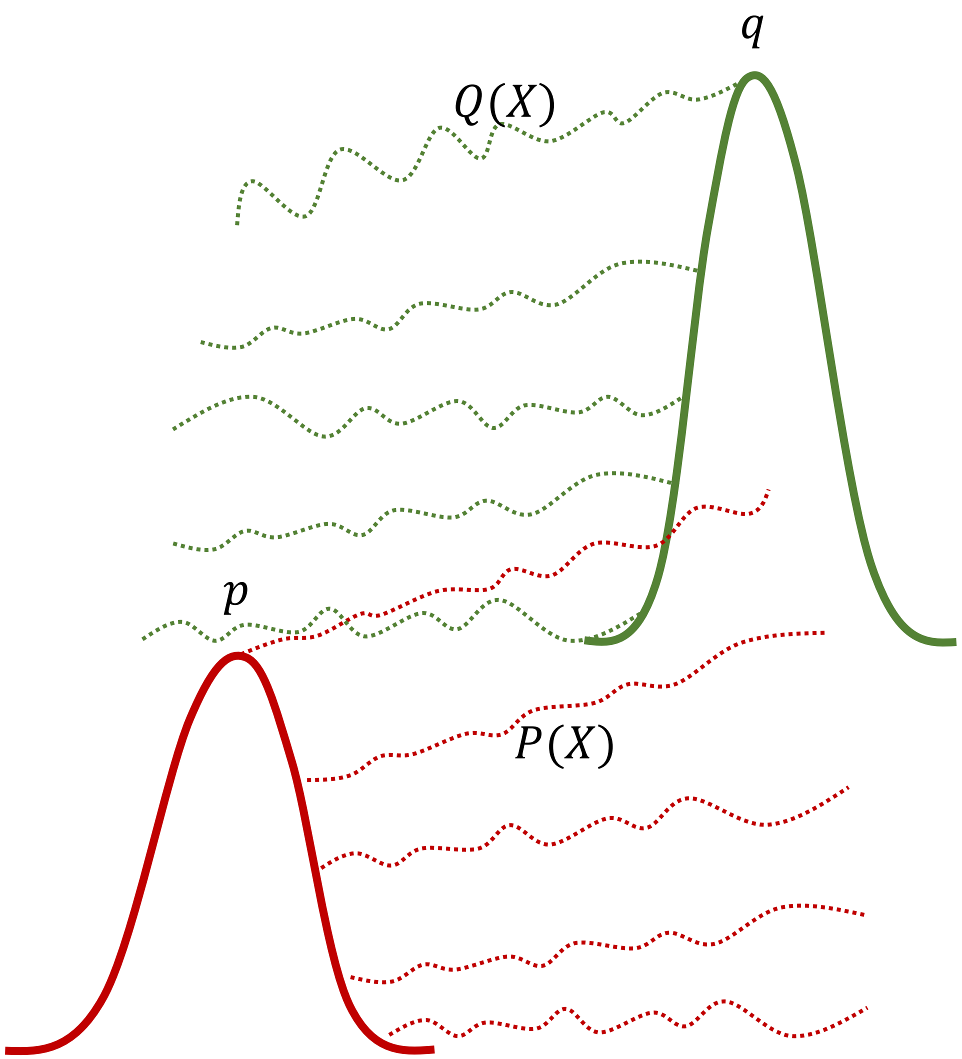
Limitations





Limitations





From Density Ratio to Path RND

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

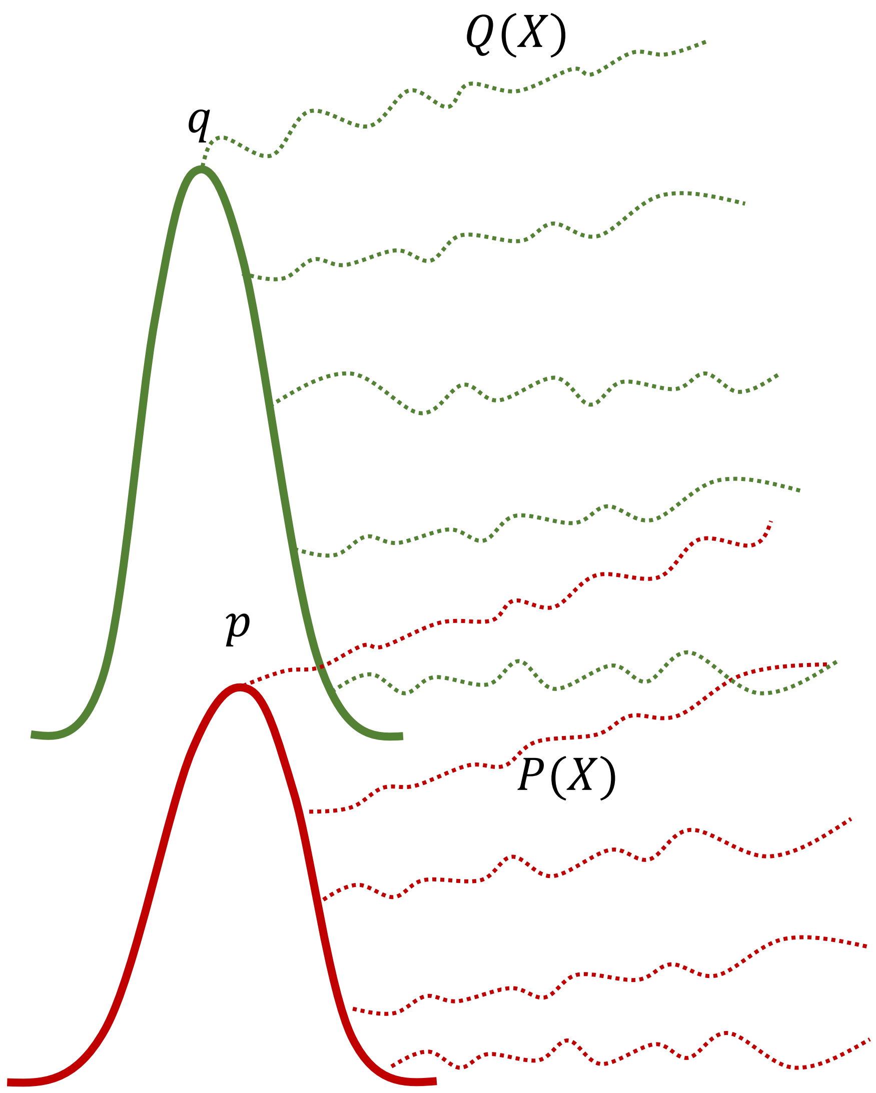
Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

"Density" ratio: $\frac{\mathrm{d}P}{\mathrm{d}Q}(x)$

Forward-forward RND (FF-RND) and Girsanov

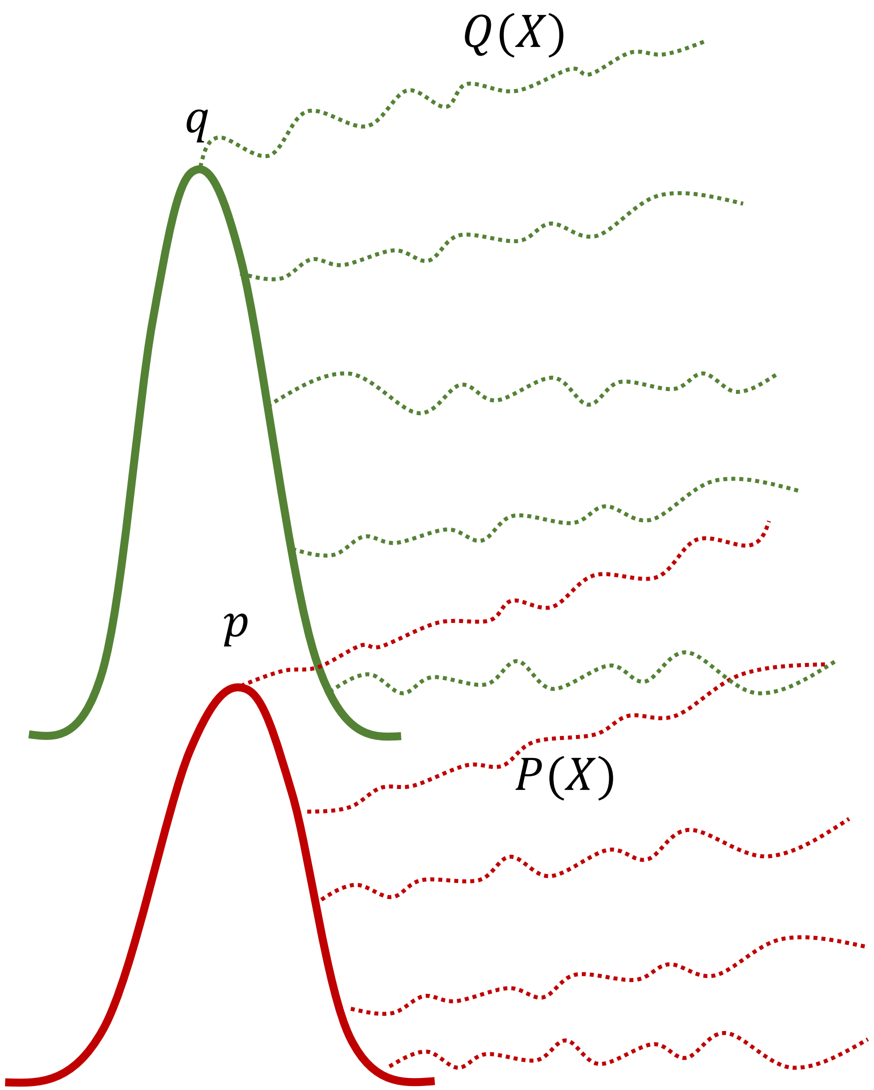
$$P: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0 = p$$
$$Q: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0 = q$$



Forward-forward RND (FF-RND) and Girsanov

$$P: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0 = p$$
$$Q: \ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0 = q$$

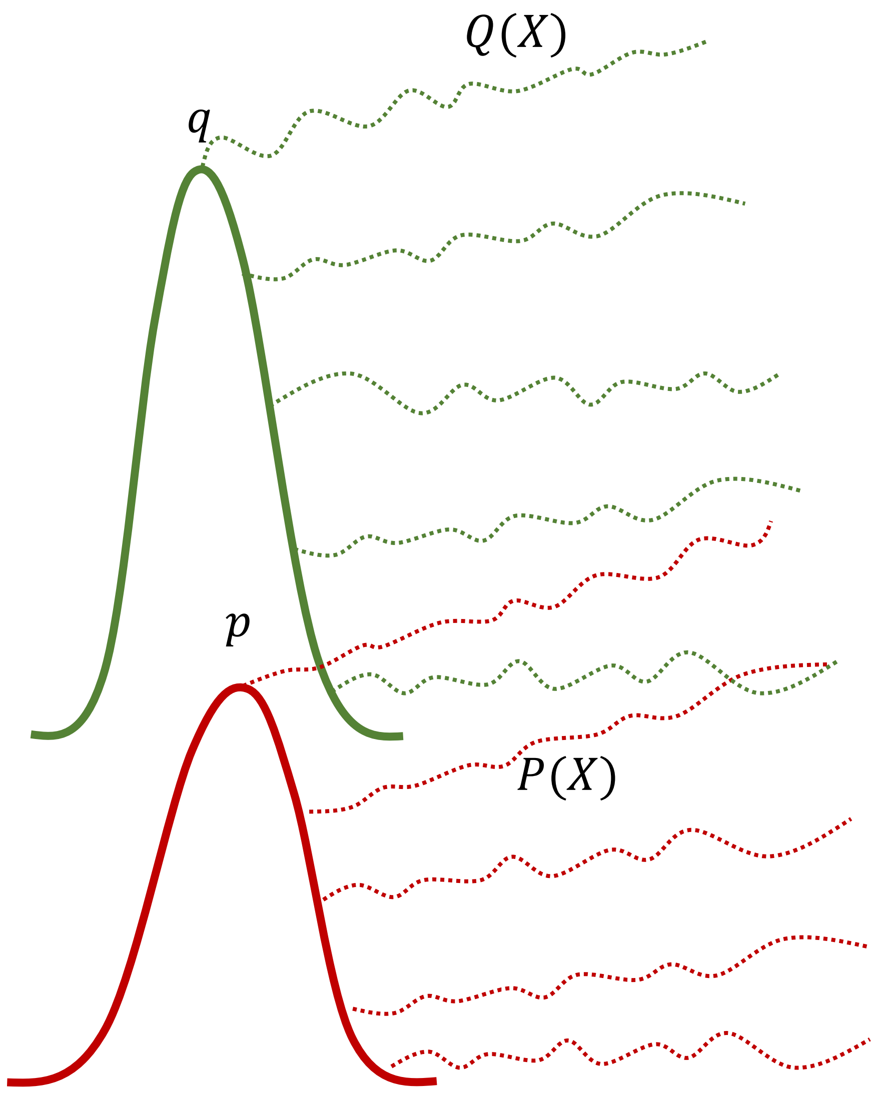
$$\frac{\mathrm{d}P}{\mathrm{d}Q}(X) = \lim \frac{p(X_0)\prod N_1(X_{n+1}|X_n)}{\underbrace{q(X_0)}_{\text{Initial density ratio}}\underbrace{\prod N_2(X_{n+1}|X_n)}_{\text{Transition kernel ratio}}}$$



Forward-forward RND (FF-RND) and Girsanov

$$P:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0 = p$$
$$Q:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0 = q$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(X) = \lim \frac{p(X_0)\prod N_1(X_{n+1}|X_n)}{\underbrace{q(X_0)}_{\text{Initial density ratio}}\underbrace{\prod N_2(X_{n+1}|X_n)}_{\text{Transition kernel ratio}}}$$



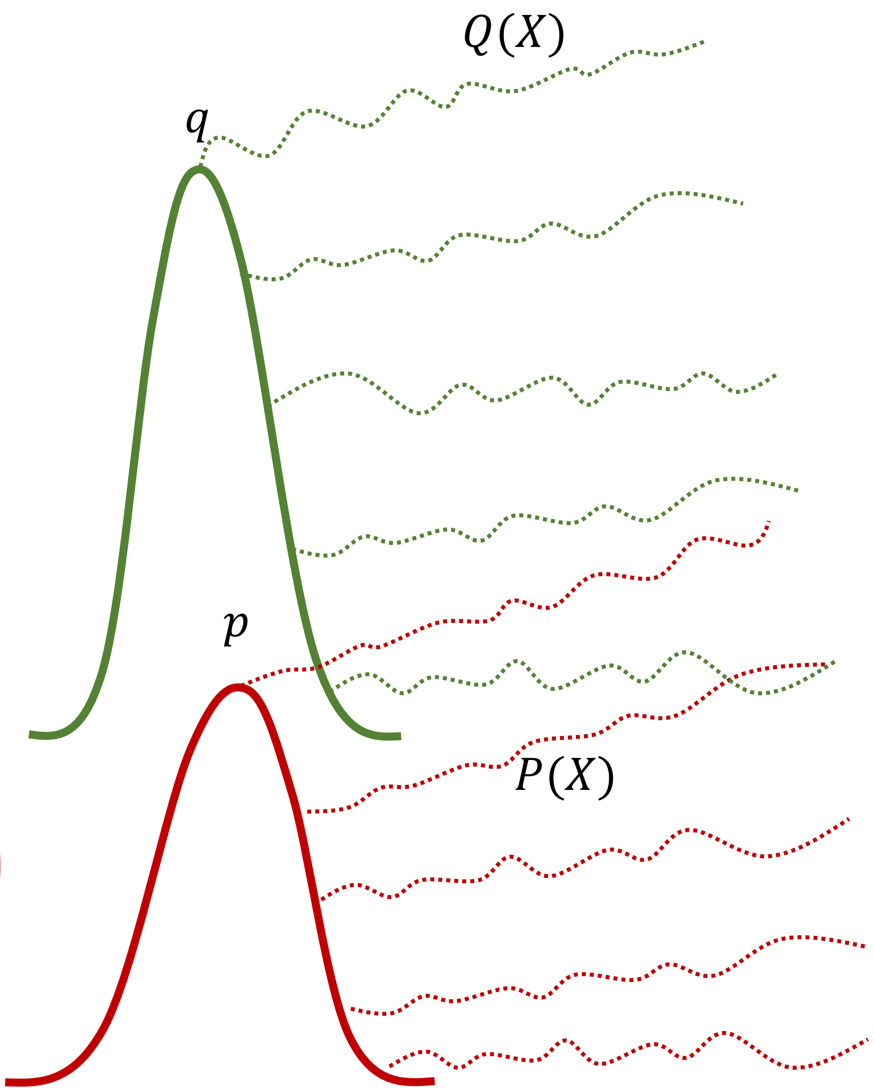
Forward-forward RND (FF-RND) and Girsanov

$$P: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0 = p$$
$$Q: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0 = q$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(X) = \lim \frac{p(X_0)\prod N_1(X_{n+1}|X_n)}{\underbrace{q(X_0)}_{\text{Initial density ratio}}\underbrace{\prod N_2(X_{n+1}|X_n)}_{\text{Transition kernel ratio}}}$$

$$= \frac{p(X_0)}{q(X_0)} \exp\left(\underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t}_{\text{Forward Ito Integral}} - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t \right)$$

Forward Ito Integral $\int a_t(X_t) \cdot \mathrm{d}X_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$



Forward-forward RND (FF-RND) and Girsanov

Unnormalised density?

$$P:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0 = p$$
$$Q:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim q_0 = q$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(X) = \lim \frac{p(X_0)\prod N_1(X_{n+1}|X_n)}{q(X_0)\prod N_2(X_{n+1}|X_n)}$$

Initial density ratio

Transition kernel ratio

$$= \frac{p(X_0)}{q(X_0)}\exp\left(\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

Forward Ito Integral $\int a_t(X_t)\cdot \mathrm{d}X_t = \lim \sum a_n(X_n)\cdot(X_{n+1}-X_n)$

Forward-forward RND (FF-RND) and Girsanov

Unnormalised density?

$$P: \ \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{p}_0 = \tilde{p}$$
$$Q: \ \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{q}_0 = \tilde{q}$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(X) = \lim \frac{\overbrace{p(X_0)}\prod N_1(X_{n+1}|X_n)}{\underbrace{q(X_0)}_{\text{Initial density ratio}}\underbrace{\prod N_2(X_{n+1}|X_n)}_{\text{Transition kernel ratio}}}$$

$$= \frac{p(X_0)}{q(X_0)}\exp\left(\underbrace{\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t}_{\text{Forward Ito Integral}} - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t\right)$$

Forward Ito Integral $\int a_t(X_t)\cdot \mathrm{d}X_t = \lim \sum a_n(X_n)\cdot(X_{n+1} - X_n)$

Forward-forward RND (FF-RND) and Girsanov

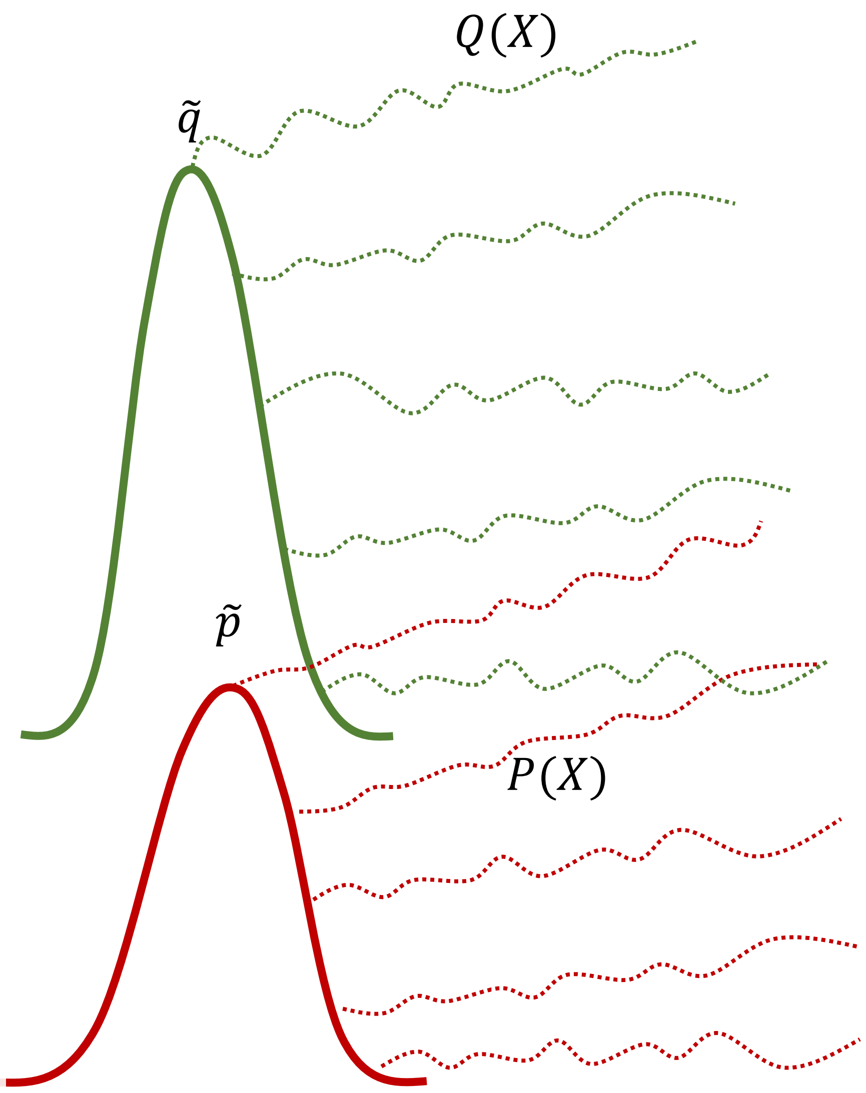
Unnormalised density?

$$P : \mathrm{d}X_t = f(X_t, t)\mathrm{dt} + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{p}_0 = \tilde{p}$$
$$Q : \mathrm{d}X_t = g(X_t, t)\mathrm{dt} + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{q}_0 = \tilde{q}$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(X) = \lim \frac{\underbrace{p(X_0)}_{\text{Initial density ratio}} \underbrace{\prod N_1(X_{n+1}|X_n)}_{\text{Transition kernel ratio}}}{q(X_0)\prod N_2(X_{n+1}|X_n)}$$

$$= \frac{p(X_0)}{q(X_0)} \exp\left(\underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t}_{\text{Forward Ito Integral}} - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t \right)$$

$$\int a_t(X_t) \cdot \mathrm{d}X_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$$



Forward-forward RND (FF-RND) and Girsanov

Unnormalised density?

$$P: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{p}_0 = \tilde{p}$$
$$Q: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{q}_0 = \tilde{q}$$

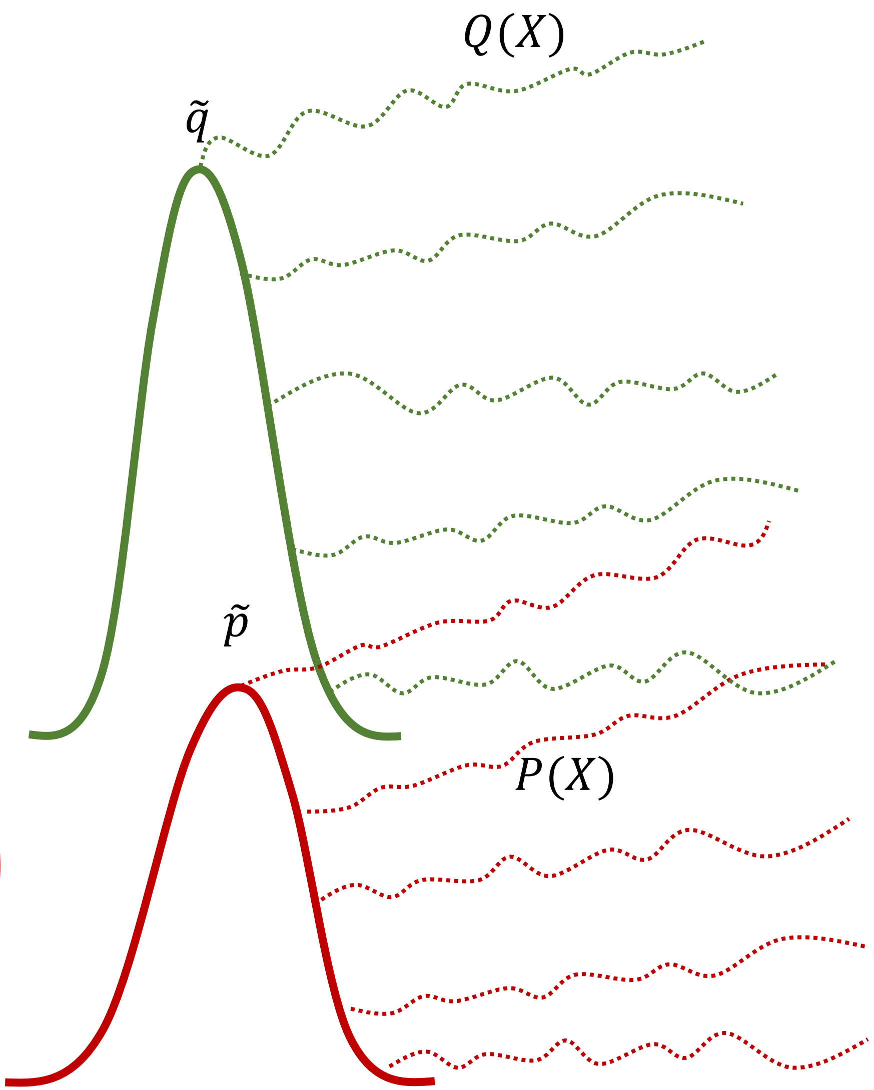
$$w(X) = \frac{Z_p \mathrm{d}P}{Z_q \mathrm{d}Q}(X) = \lim \frac{\tilde{p}(X_0) \prod N_1(X_{n+1}|X_n)}{\tilde{q}(X_0) \prod N_2(X_{n+1}|X_n)}$$

Initial density ratio

Transition kernel ratio

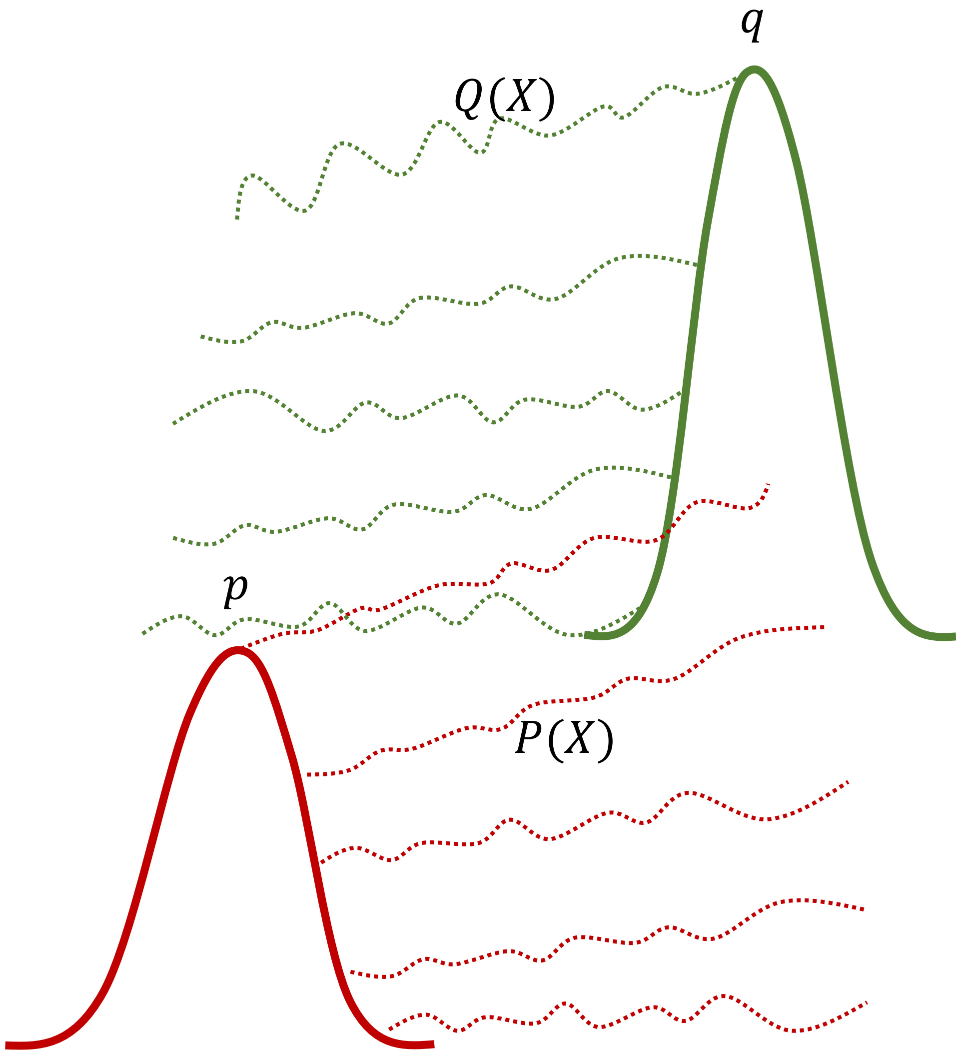
$$= \frac{\tilde{p}(X_0)}{\tilde{q}(X_0)} \exp\left(\underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t}_{\text{Forward Ito Integral}} - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \mathrm{d}X_t + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{d}t \right)$$

Forward Ito Integral $\int a_t(X_t) \cdot \mathrm{d}X_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$



Forward-backward RND (FB-RND)

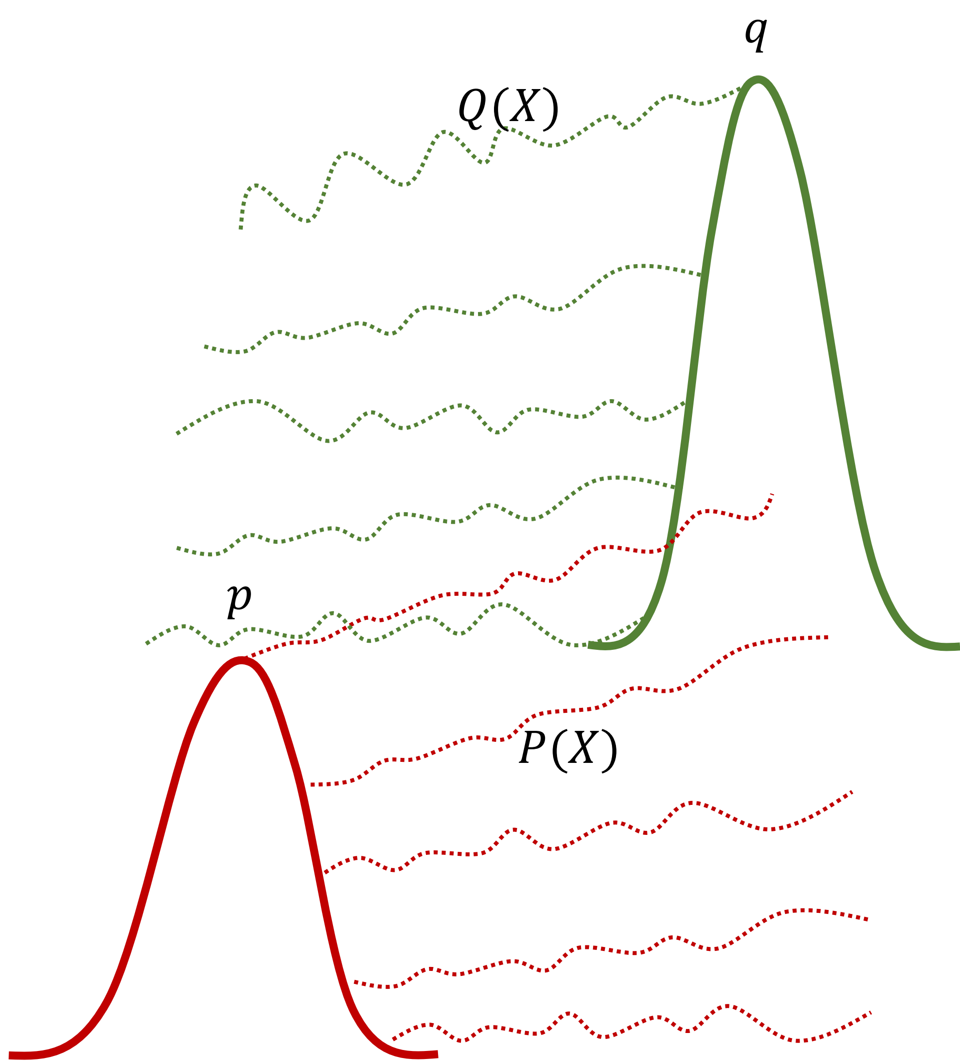
$$P: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{p}_0 = \tilde{p}$$
$$Q: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim \tilde{q}_1 = \tilde{q}$$



Forward-backward RND (FB-RND)

$$P:\ \mathrm{d}X_t = f(X_t,t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{p}_0 = \tilde{p}$$
$$Q:\ \mathrm{d}X_t = g(X_t,t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim \tilde{q}_1 = \tilde{q}$$

$$w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}\overleftarrow{Q}}(X) = \lim \frac{\tilde{p}_0(X_0)\prod N_1(X_{n+1}|X_n)}{\underbrace{\tilde{q}_1(X_1)}_{\text{Initial density ratio}}\underbrace{\prod N_2(X_n|X_{n+1})}_{\text{Transition kernel ratio}}}$$



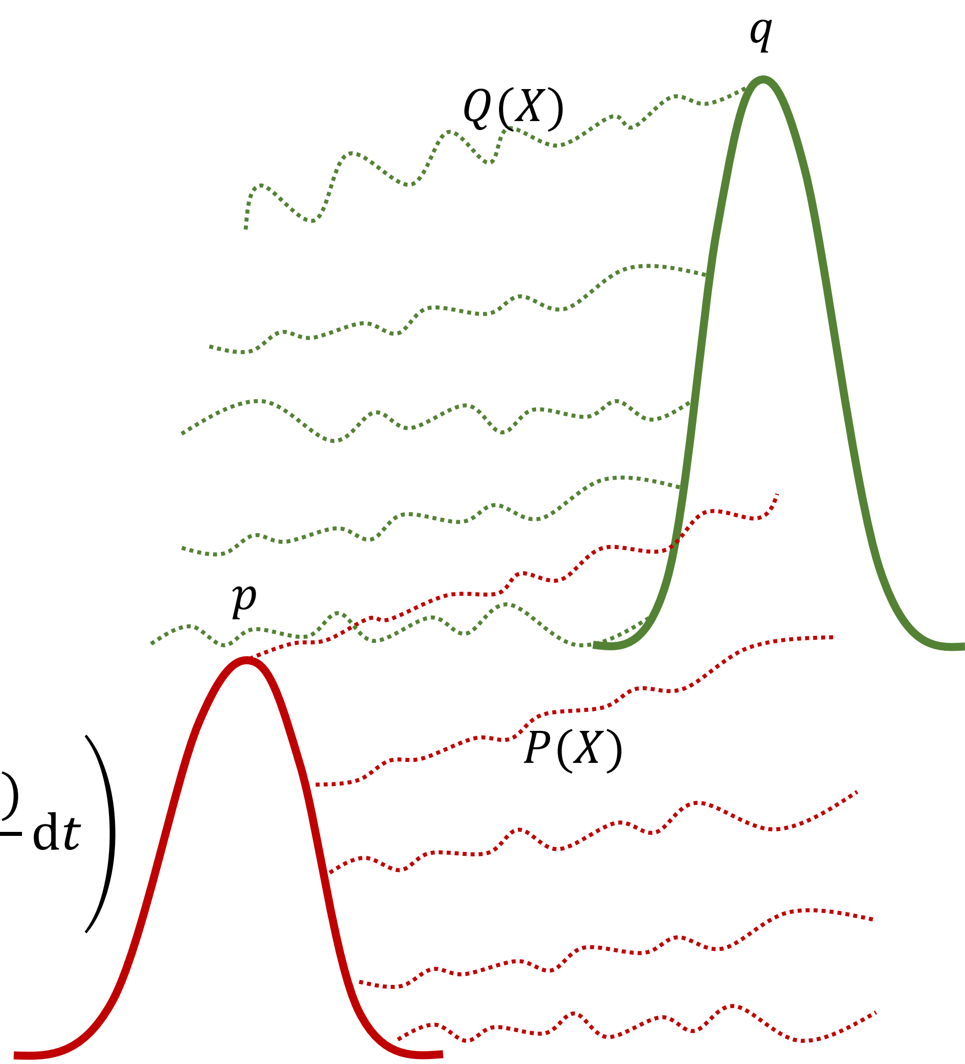
Forward-backward RND (FB-RND)

$$P:\ \mathrm{d}X_t = f(X_t, t)\mathrm{dt} + \sigma_t \mathrm{d}W_t, X_0 \sim \tilde{p}_0 = \tilde{p}$$
$$Q:\ \mathrm{d}X_t = g(X_t, t)\mathrm{dt} + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim \tilde{q}_1 = \tilde{q}$$

$$w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}\overleftarrow{Q}}(X) = \lim \frac{\tilde{p}_0(X_0)\prod N_1(X_{n+1}|X_n)}{\underbrace{\tilde{q}_1(X_1)}_{\text{Initial density ratio}}\underbrace{\prod N_2(X_n|X_{n+1})}_{\text{Transition kernel ratio}}}$$

$$= \frac{\tilde{p}_0\,(X_0)}{\tilde{q}_1(X_1)}\exp\left(\int \frac{f_t(X_t)}{\sigma_t^2}\cdot \mathrm{d}X_t - \frac{f_t^2(X_t)}{2\sigma_t^2}\mathrm{dt} - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2}\cdot \overleftarrow{\mathrm{d}X_t}} + \frac{g_t^2(X_t)}{2\sigma_t^2}\mathrm{dt}\right)$$

$$\int a_t(X_t)\cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1})\cdot (X_{n+1} - X_n)$$ Backward Ito Integral



A Side Note on Stochastic Intergrals

Ito forward integral

$$\int a_t(X_t) \cdot \mathrm{d}X_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$$

Ito backward integral

$$\int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Conversion rule:

$$\int a_t(X_t) \cdot \mathrm{d}X_t - \int a_t(X_t) \cdot \overleftarrow{\mathrm{d}X_t} = -\int \sigma_t^2 \nabla \cdot a_t \mathrm{d}t$$

From Density Ratio to Path RND

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

From Density Ratio to Path RND

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\, \mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

From Density Ratio to Path RND

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$

From Density Ratio to Path RND

Wait…WHY PATH?

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$

From Density Ratio to Path RND

Wait...WHY PATH?

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

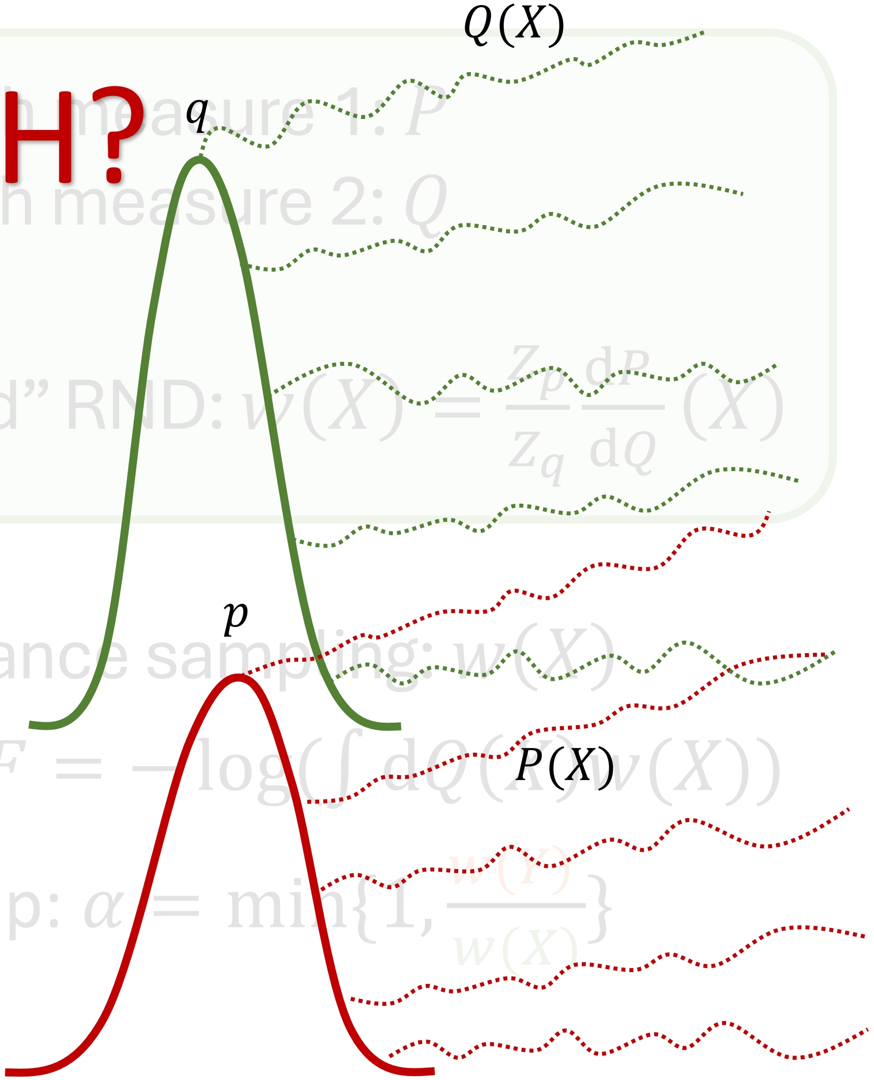
Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\, \mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$



From Density Ratio to Path RND

Unnormalised density 1: $\tilde{p}$

Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P

Path measure 2: Q

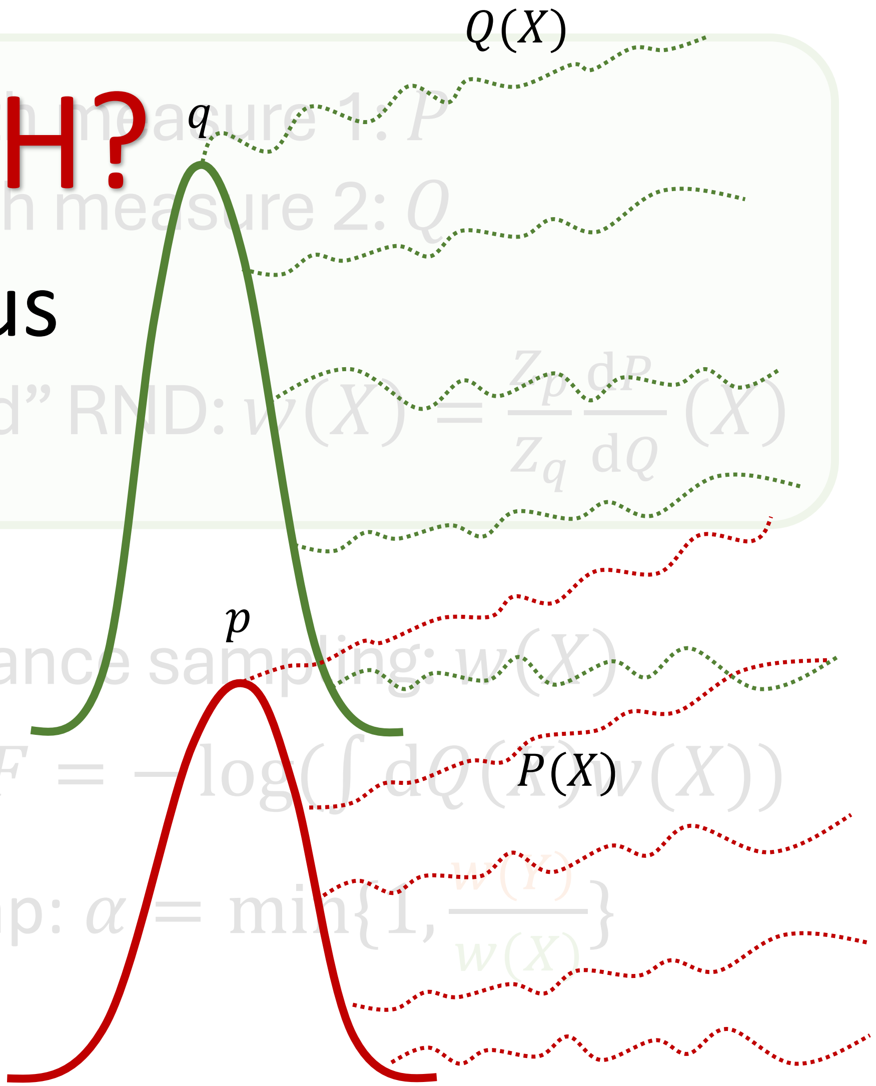
"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x)w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X)w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$

Wait…WHY PATH?

😳 data processing inequality (DPI) told us



From Density Ratio to Path RND

Wait…WHY PATH?

😳 data processing inequality (DPI) told us

😵 $D_f[q||p] \leq D_f[Q||P]$

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

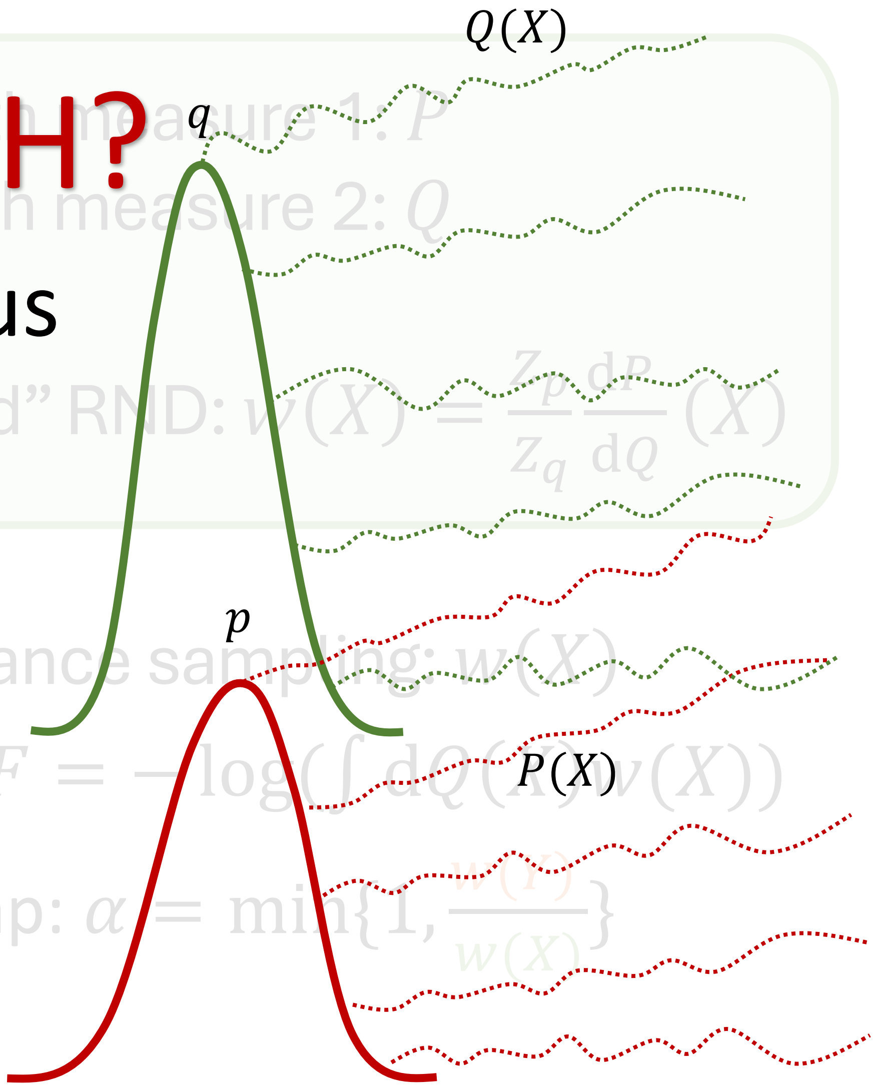
Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

“Unnormalised” RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x)w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X)w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$



From Density Ratio to Path RND

Unnormalised density 1: $\tilde{p}$

Unnormalised density 2: q

Density ratio: $w(x) = \frac{\tilde{p}(x)}{q(x)}$

Path measure 1: P

Path measure 2: Q

“Unnormalised” RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

Wait...WHY PATH?

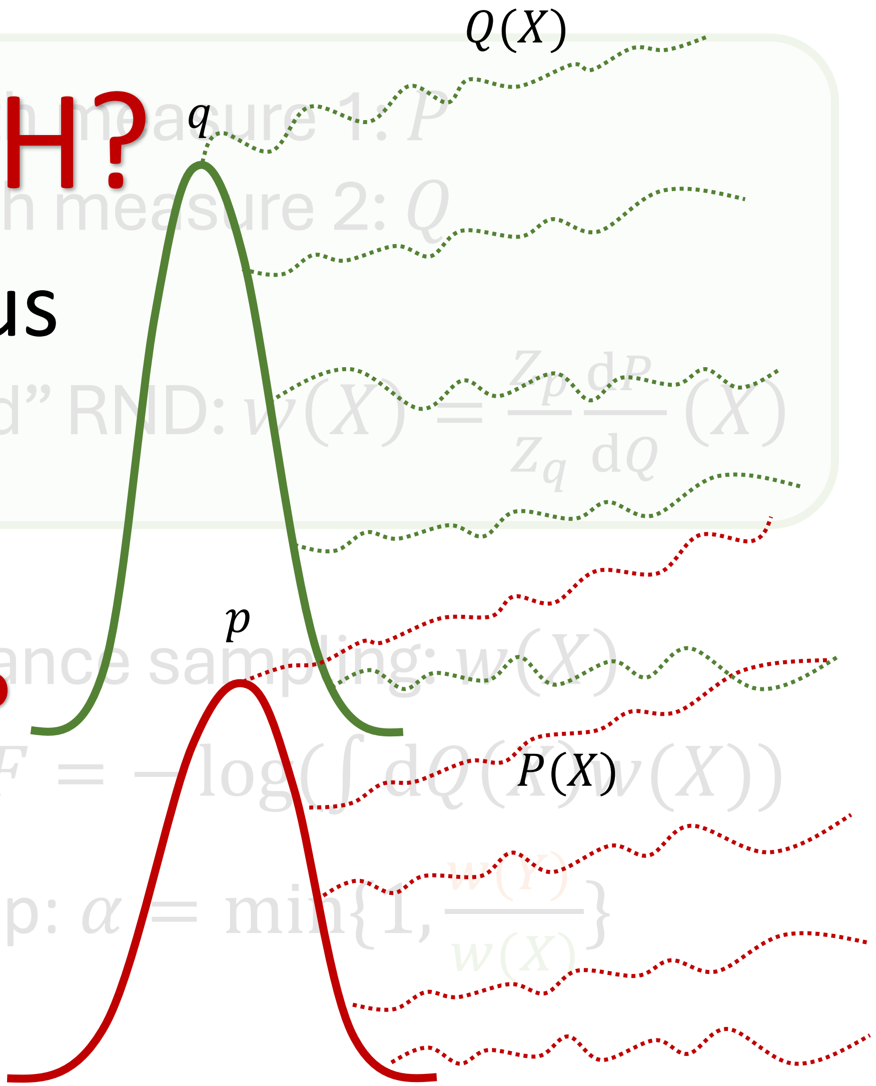
data processing inequality (DPI) told us

$$D_f[q||p] \leq D_f[Q||P]$$

Path weight always has larger variance?

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{q(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$



From Density Ratio to Path RND

Path weight always has **larger variance?**

Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x)w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X)w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$

From Density Ratio to Path RND

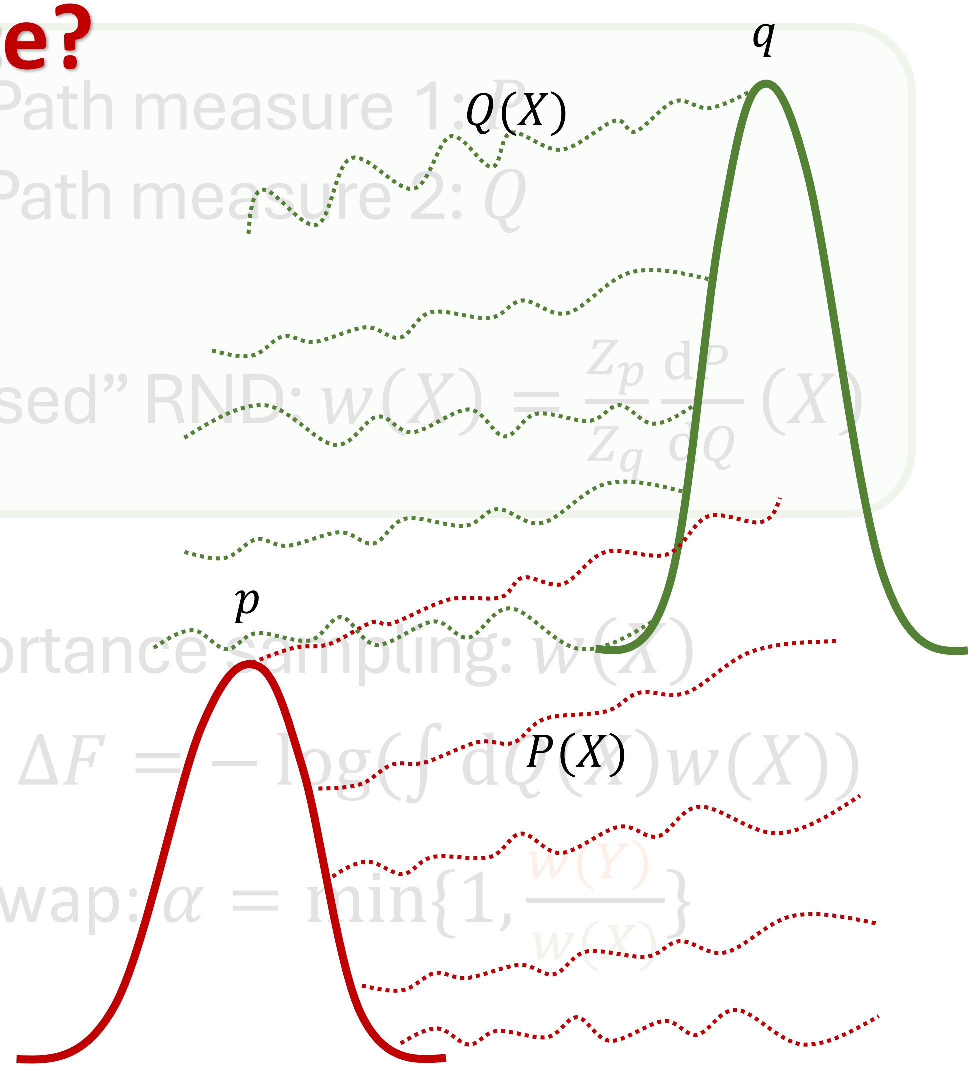
Path weight always has larger variance?

💡 **Not for FB RND!**

Path measure 1: P

Path measure 2: Q

“Unnormalised” RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$



- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance Sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$

From Density Ratio to Path RND

Path weight always has **larger variance?**

💡 **Not for FB RND!**

If $\overleftarrow{Q} = P$ (**time-reversal**)

The path weight will have 0 variance

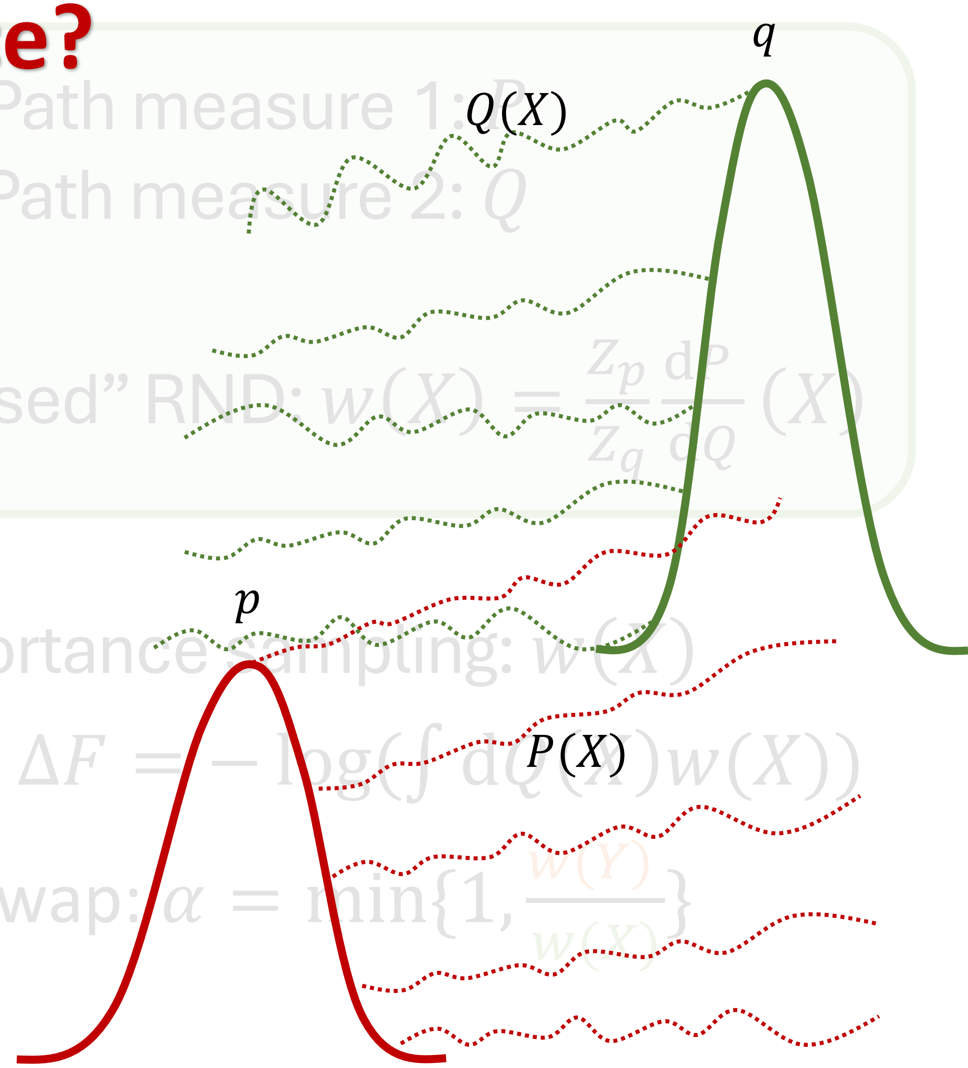
Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x)w(x)\,\mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance Sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X)w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$



Time-reversal and Nelson's relation

$$P: \; \mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, X_0 \sim p_0$$
$$\overleftarrow{Q}: \; \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t}, X_1 \sim p_1$$

"time-reversal" $$\overleftarrow{Q} = P, \mathrm{i.e.}, \frac{\overleftarrow{\mathrm{d}Q}}{\mathrm{d}P} = 1$$

Iff

$$g(\cdot, t) = f(\cdot, t) - \sigma_t^2 \nabla \log p_t(\cdot)$$

From Density Ratio to Path RND

Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$

From Density Ratio to Path RND

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$ → (escorted) Jarzynski/Crooks
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$ → Replica exchange with nonequilibrium switches [1] / Accelerated PT [2]

[1] Ballard, Andrew J., and Christopher Jarzynski. "Replica exchange with nonequilibrium switches." *Proceedings of the National Academy of Sciences* 106.30 (2009): 12224-12229.
[2] Zhang, Leo, et al. "Accelerated Parallel Tempering via Neural Transports." *arXiv preprint arXiv:2502.10328* (2025).

From Density Ratio to Path RND

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

Equilibrium method and nonequilibrium ones are not too different:

One use *Marginal space RND*

One use *Path space RND*

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$ → (escorted) Jarzynski/Crooks
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$ → Replica exchange with nonequilibrium switches [1] / Accelerated PT [2]

[1] Ballard, Andrew J., and Christopher Jarzynski. "Replica exchange with nonequilibrium switches." *Proceedings of the National Academy of Sciences* 106.30 (2009): 12224-12229.
[2] Zhang, Leo, et al. "Accelerated Parallel Tempering via Neural Transports." *arXiv preprint arXiv:2502.10328* (2025).

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t \right)$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{p(X_1)}{q(X_0)}\exp\left(\int \frac{\nabla U_t}{2}\cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4}|\nabla U_t|^2\mathrm{d}t + \int \frac{\nabla U_t}{2}\cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4}|\nabla U_t|^2\mathrm{d}t\right)$$

$$= \frac{p(X_1)}{q(X_0)}\exp\left(\int \frac{\nabla U_t}{2}\cdot \mathrm{d}X_t + \int \frac{\nabla U_t}{2}\cdot \overleftarrow{\mathrm{d}X_t}\right)$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} \right)$$

👉 conversion rule

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} - \frac{\sigma_t^2}{4} |\nabla U_t|^2 \mathrm{d}t \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \frac{\nabla U_t}{2} \cdot \mathrm{d}X_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{\mathrm{d}X_t} \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right)$$

conversion rule

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{dt} \right)$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right)$$

Ito's lemma

$$\mathrm{d}f_t(X_t) = (\partial_t f(X_t) + \sigma_t^2 \Delta f)\mathrm{d}t + \nabla U_t \cdot \mathrm{d}X_t$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t\right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t\right)$$

👉 Ito's lemma

$$\mathrm{d}f_t(X_t) = (\partial_t f(X_t) + \sigma_t^2 \Delta f)\mathrm{d}t + \nabla U_t \cdot \mathrm{d}X_t$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\begin{aligned}
\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} &= \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right) \\
&= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t \right) \\
&= \frac{p(X_1)}{q(X_0)} \exp\left(U_1(X_1) - U_0(X_0) + \int -\partial_t U_t(X_t)\mathrm{d}t \right)
\end{aligned}$$

👉 Ito's lemma

$$\mathrm{d}f_t(X_t) = (\partial_t f(X_t) + \sigma_t^2 \Delta f)\mathrm{d}t + \nabla U_t \cdot \mathrm{d}X_t$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\begin{aligned}
\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} &= \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t \right) \\
&= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t \right) \\
&= \frac{Z_0 \exp(-U_1(X_1))}{Z_1 \exp(-U_0(X_0))} \exp\left(U_1(X_1) - U_0(X_0) + \int -\partial_t U_t(X_t)\mathrm{d}t \right)
\end{aligned}$$

Ito's lemma

$$\mathrm{d}f_t(X_t) = (\partial_t f(X_t) + \sigma_t^2 \Delta f)\mathrm{d}t + \nabla U_t \cdot \mathrm{d}X_t$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\begin{aligned}
\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} &= \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{d}t\right) \\
&= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t\right) \\
&= \frac{Z_0 \cancel{\exp(-U_1(X_1))}}{Z_1 \cancel{\exp(-U_0(X_0))}} \exp\left(\cancel{U_1(X_1) - U_0(X_0)} + \int -\partial_t U_t(X_t)\mathrm{d}t\right)
\end{aligned}$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\begin{aligned}
\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} &= \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot \mathrm{d}X_t + \int \sigma_t^2 \Delta U_t \mathrm{dt} \right) \\
&= \frac{p(X_1)}{q(X_0)} \exp\left(\int \mathrm{d}U_t(X_t) - \partial_t U_t(X_t)\mathrm{d}t \right) \\
&= \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t \right)
\end{aligned}$$

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)$$

💡 Crooks Fluctuation Theorem

Example: Path RND to Jarzynski Equality

$$X_0 \sim q \qquad \mathrm{d}X_t = -\sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = \sigma^2 \nabla U_t(X_t)\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)$$

💡 Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\frac{\overleftarrow{\mathrm{dP}}}{\mathrm{dQ}}\right] = \mathbf{E}_{\mathbf{Q}}\left[\frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t\right)\right] = 1$$

💡 Jarzynski Equation

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)$$

💡 Controlled Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

💡 Escorted Jarzynski Equation

From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \qquad \mathrm{d}X_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overrightarrow{\mathrm{d}W_t},$$

$$X_1 \sim p \qquad \mathrm{d}X_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma\sqrt{2}\,\overleftarrow{\mathrm{d}W_t},$$

$$\frac{\overleftarrow{\mathrm{d}\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \frac{Z_0}{Z_1}\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)$$

💡 Controlled Crooks Fluctuation Theorem

$$\mathbf{E}_{\mathbf{Q}}\left[\exp\left(\int -\partial_t U_t(X_t)\mathrm{d}t - \nabla U_t \cdot u_t \mathrm{d}t + \nabla \cdot u_t \mathrm{d}t\right)\right] = \frac{Z_1}{Z_0}$$

💡 Escorted Jarzynski Equation

Can also be derived via PDEs [1] / Feynman-Kac formula [2]:

[1] Albergo, M. S., & Vanden-Eijnden, E (2025). NETS: A Non-equilibrium Transport Sampler. *ICML 2025.*

[2] Skreta, M., Akhound-Sadegh, T., Ohanesian, V., Bondesan, R., Aspuru-Guzik, A., Doucet, A., ... & Neklyudov, K. (2025). Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *ICML 2025*.

From Density Ratio to Path RND

Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

Equilibrium method and nonequilibrium ones are not too different:

One use *Marginal space RND*

One use *Path space RND*

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$ → (escorted) Jarzynski/Crooks
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$ → Replica exchange with nonequilibrium switches [1] / Accelerated PT [2]

[1] Ballard, Andrew J., and Christopher Jarzynski. "Replica exchange with nonequilibrium switches." *Proceedings of the National Academy of Sciences* 106.30 (2009): 12224-12229.

[2] Zhang, Leo, et al. "Accelerated Parallel Tempering via Neural Transports." *arXiv preprint arXiv:2502.10328* (2025).

From Density Ratio to Path RND

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

Equilibrium method and nonequilibrium ones are not too different:

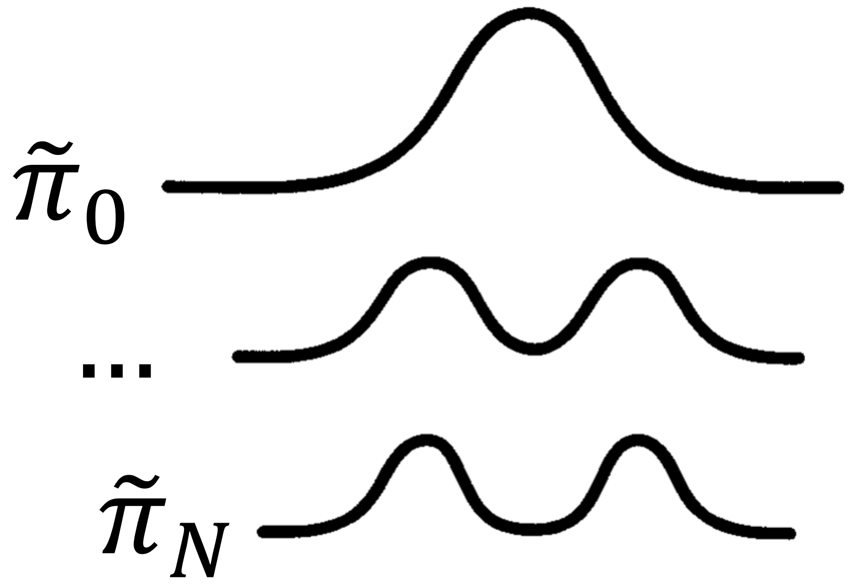
One use *Marginal space RND*

One use *Path space RND*

- ✓ Path Importance sampling: $w(X)$
- ✓ Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$ → (escorted) Jarzynski/Crooks
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$ → Replica exchange with nonequilibrium switches [1] / Accelerated PT [2]

[1] Ballard, Andrew J., and Christopher Jarzynski. "Replica exchange with nonequilibrium switches." *Proceedings of the National Academy of Sciences* 106.30 (2009): 12224-12229.
[2] Zhang, Leo, et al. "Accelerated Parallel Tempering via Neural Transports." *arXiv preprint arXiv:2502.10328* (2025).

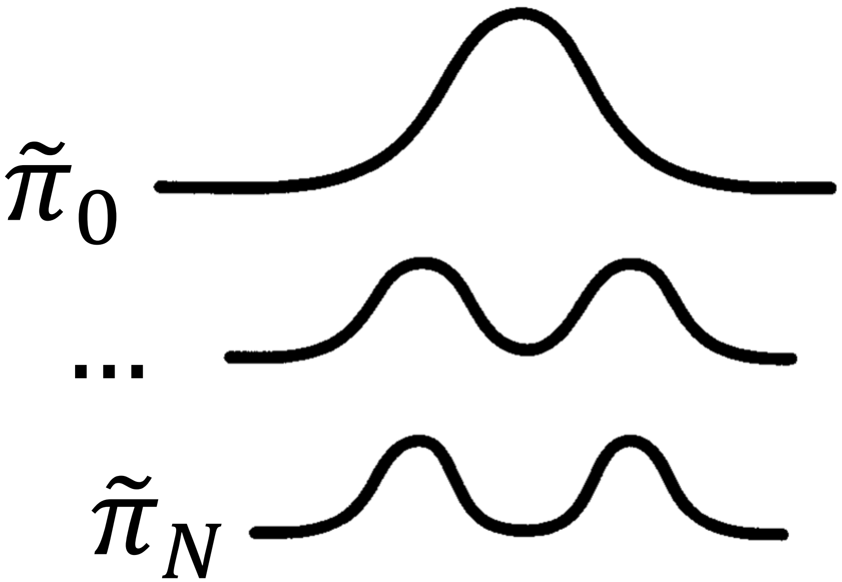
Parallel tempering



- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

1. Local exploration kernel: independent MCMC for each $\tilde{\pi}_n$
2. Communication kernel: swap between all adjacent pairs $(\tilde{\pi}_n, \tilde{\pi}_{n+1})$

Parallel tempering



- **An MCMC algorithm for target density $\tilde{\pi}_N$**
- Workflow:
  - Choose an easy-to-sample reference $\tilde{\pi}_0$
  - Design multiple intermediate targets $\tilde{\pi}_n$
  - Design two MCMC kernels with invariant measure as $\tilde{\pi}_0 \times \tilde{\pi}_1 \times \cdots \times \tilde{\pi}_N$

Unchanged!

1. Local exploration kernel: independent MCMC for each $\tilde{\pi}_n$
2. Communication kernel: swap between all adjacent pairs $(\tilde{\pi}_n, \tilde{\pi}_{n+1})$

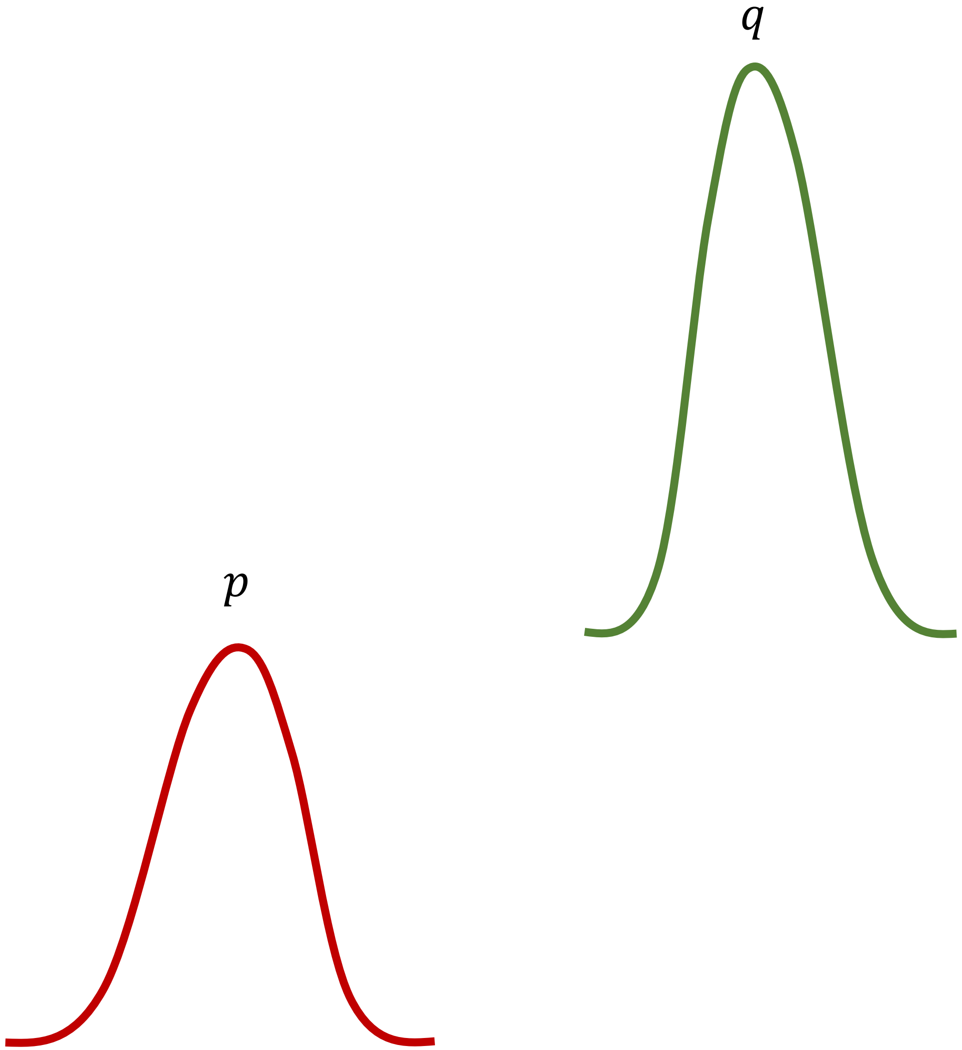
Extend to path!

Parallel tempering Swap in Path Space

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(1) Current state $(x, y) \sim p(x) \times q(y)$



Parallel tempering Swap in Path Space

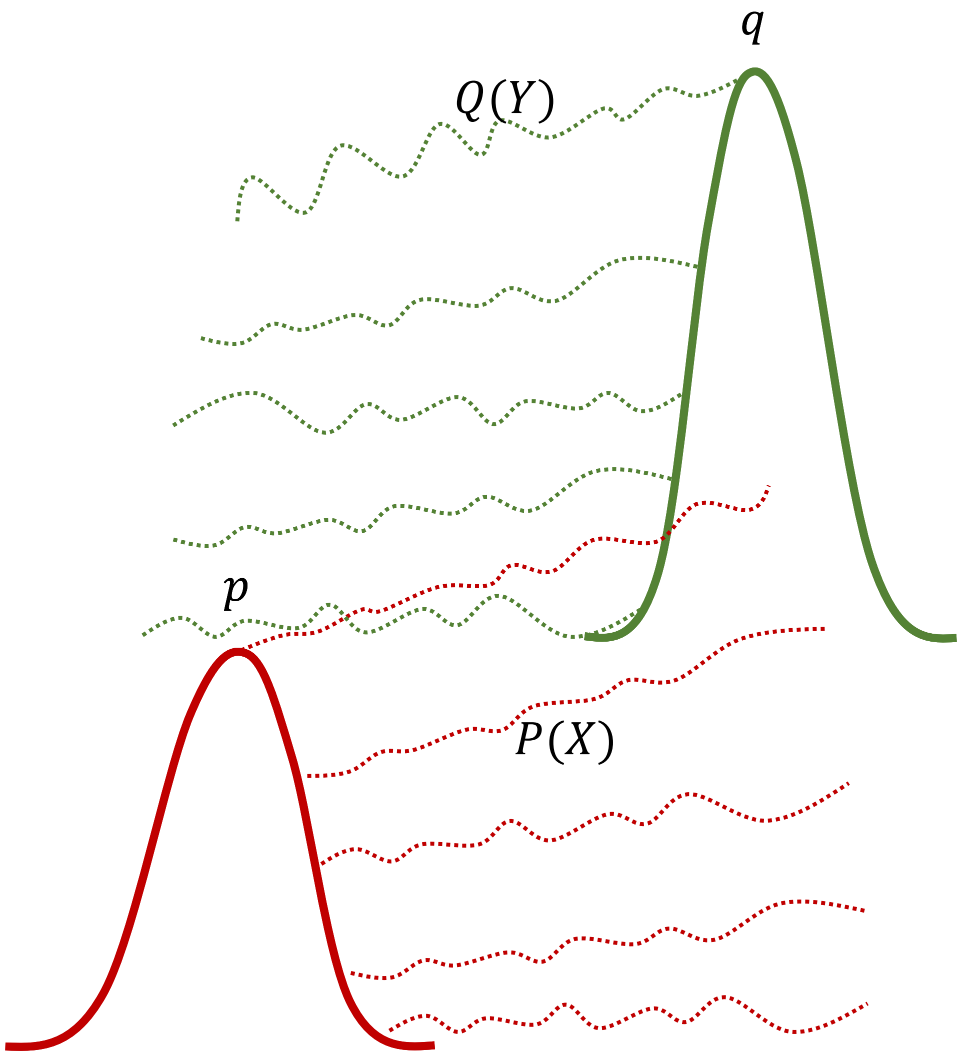
Path measure 1: P

Path measure 2: Q

“Unnormalised” RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(2) Extend current states with path

$$(X, Y) \sim P(X) \times Q(Y)$$

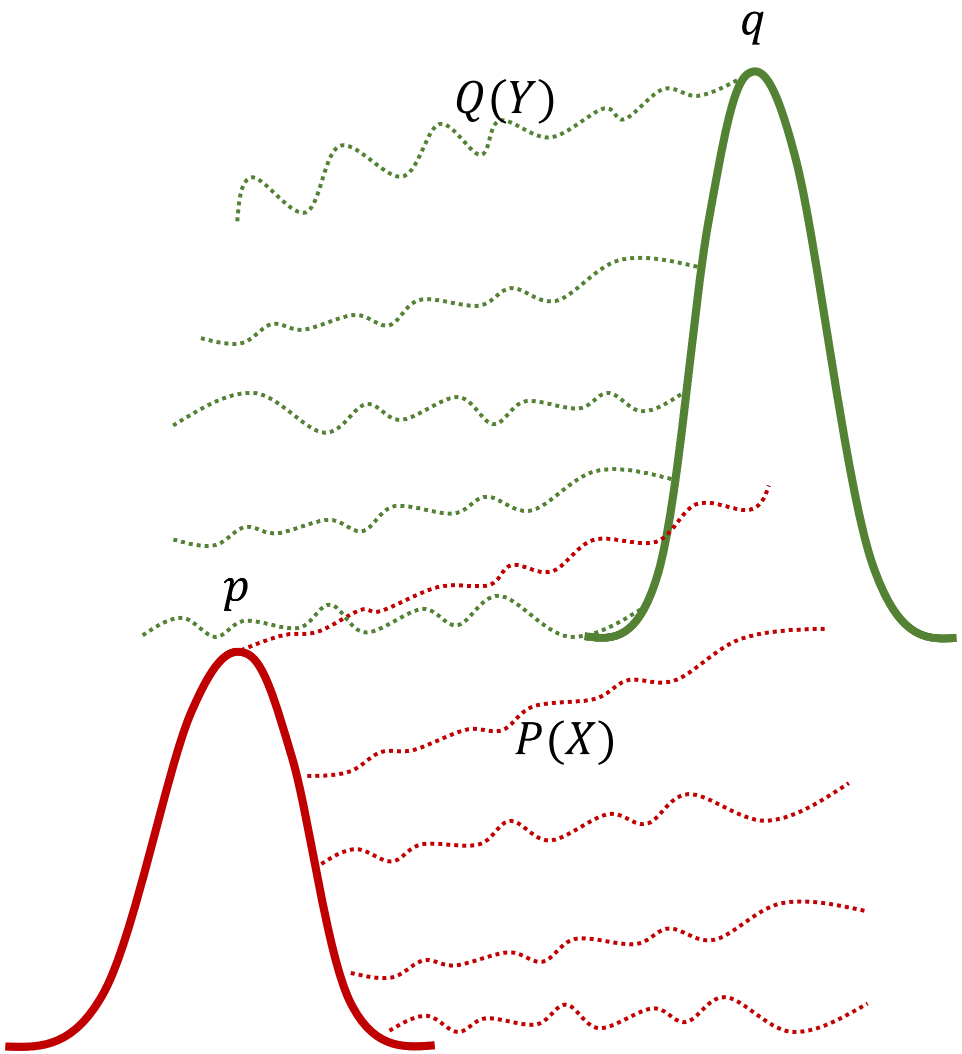


Parallel tempering Swap in Path Space

Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$



(3) Swap the Paths

$$(X', Y') \leftarrow (Y, X)$$

*Note that this proposal function is still involution

Parallel tempering Swap in Path Space

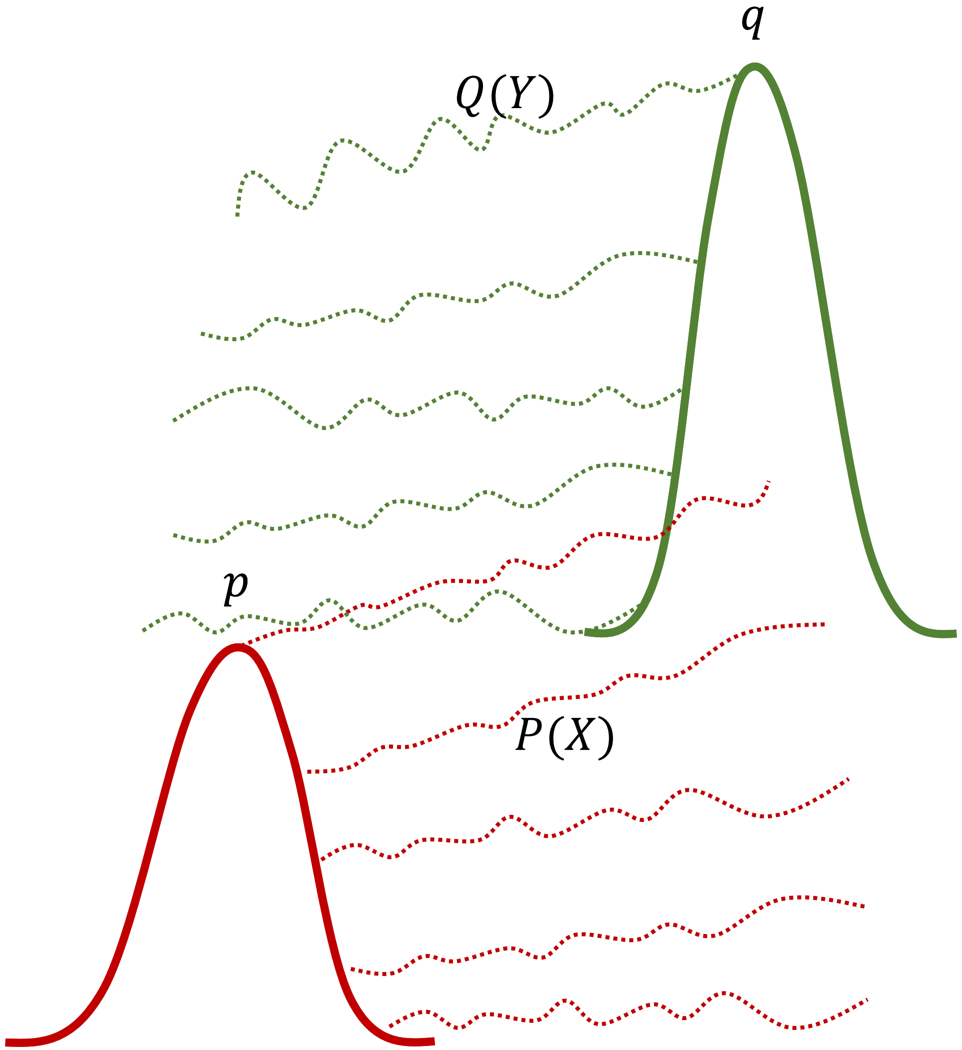
Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(4) Calculate MH correction

$$\alpha = \min\{1, \frac{\mathrm{d}\, P(X') \times Q(Y')}{\mathrm{d}\;\, P(X) \times Q(Y)}\}$$



Parallel tempering Swap in Path Space

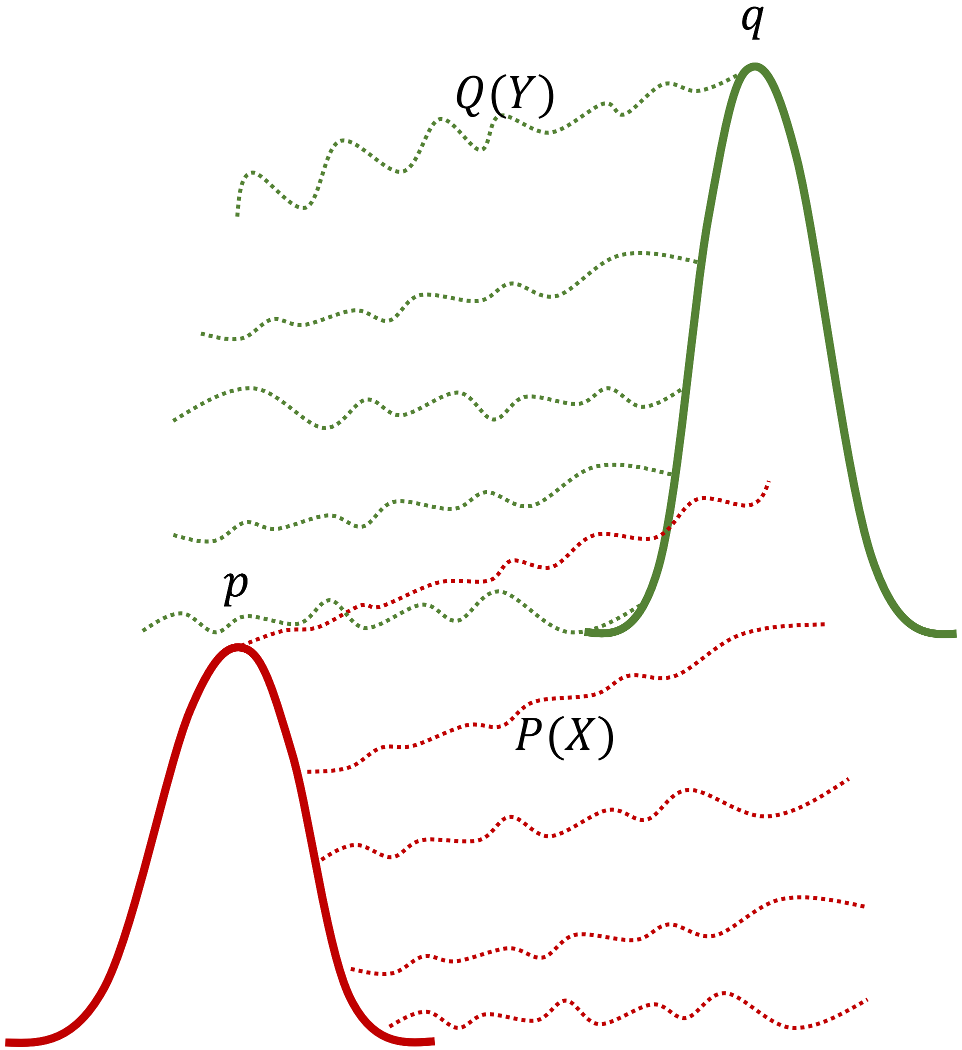
Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(4) Calculate MH correction

$$\alpha = \min\{1, \frac{\mathrm{d}\, P(Y) \times Q(X)}{\mathrm{d}\ P(X) \times Q(Y)}\}$$



Parallel tempering Swap in Path Space

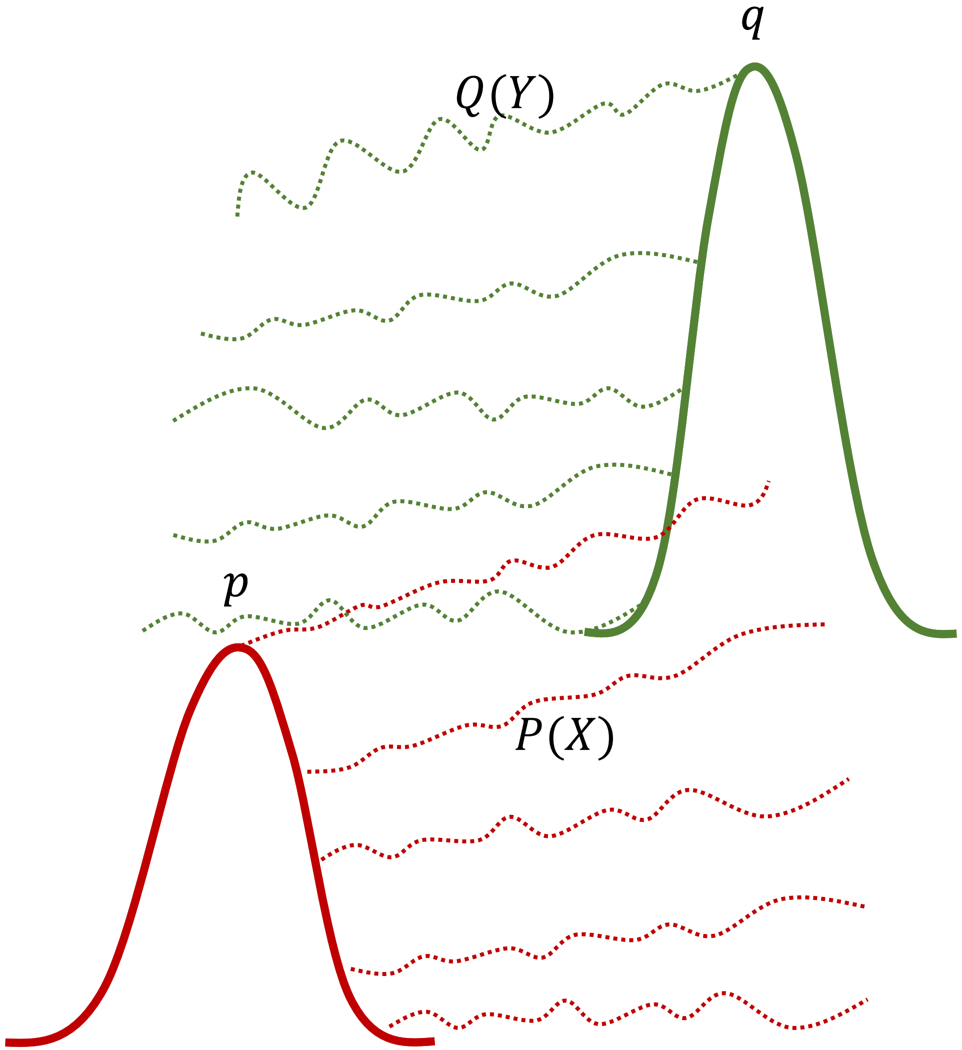
Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(4) Calculate MH correction

$$\alpha = \min\{1, \frac{\mathrm{d}\ P(Y) \times Q(X)}{\mathrm{d}\ \ P(X) \times Q(Y)}\}$$



Parallel tempering Swap in Path Space

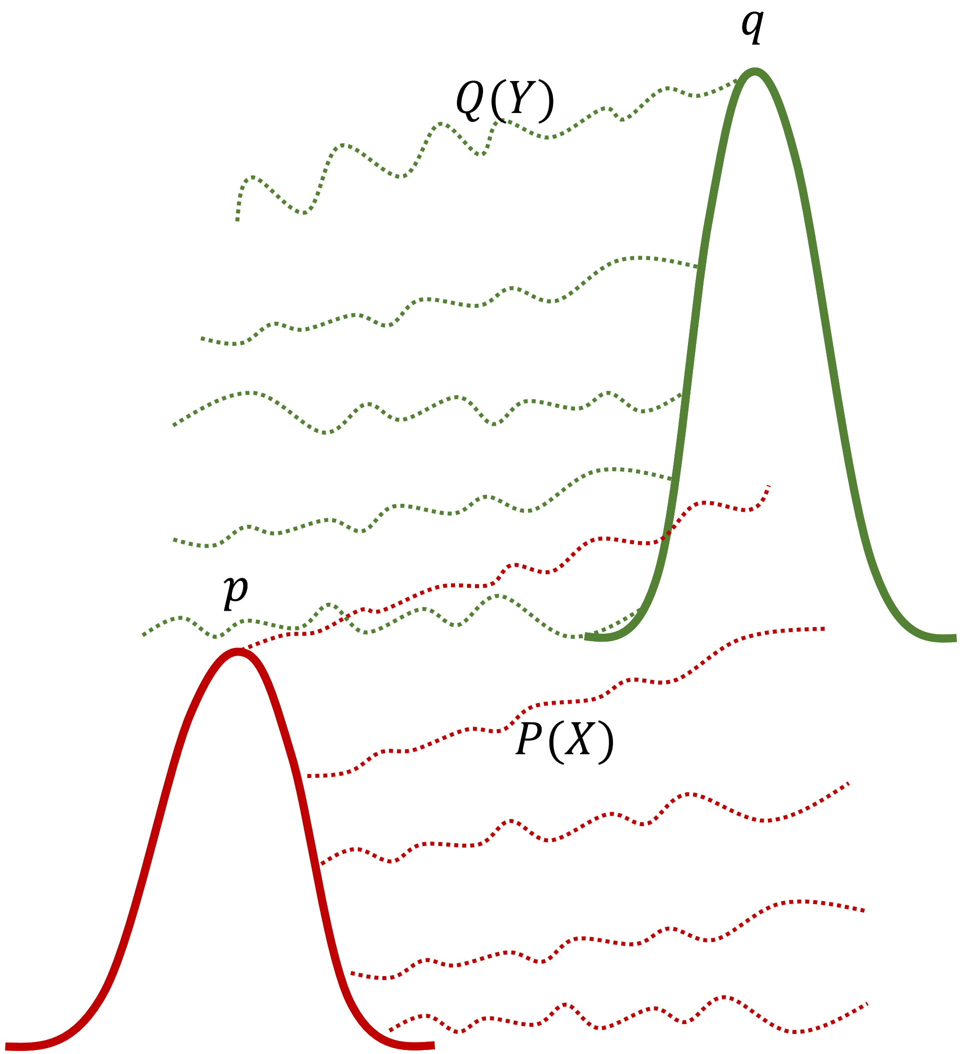
Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(4) Calculate MH correction

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$



Parallel tempering Swap in Path Space

Path measure 1: P

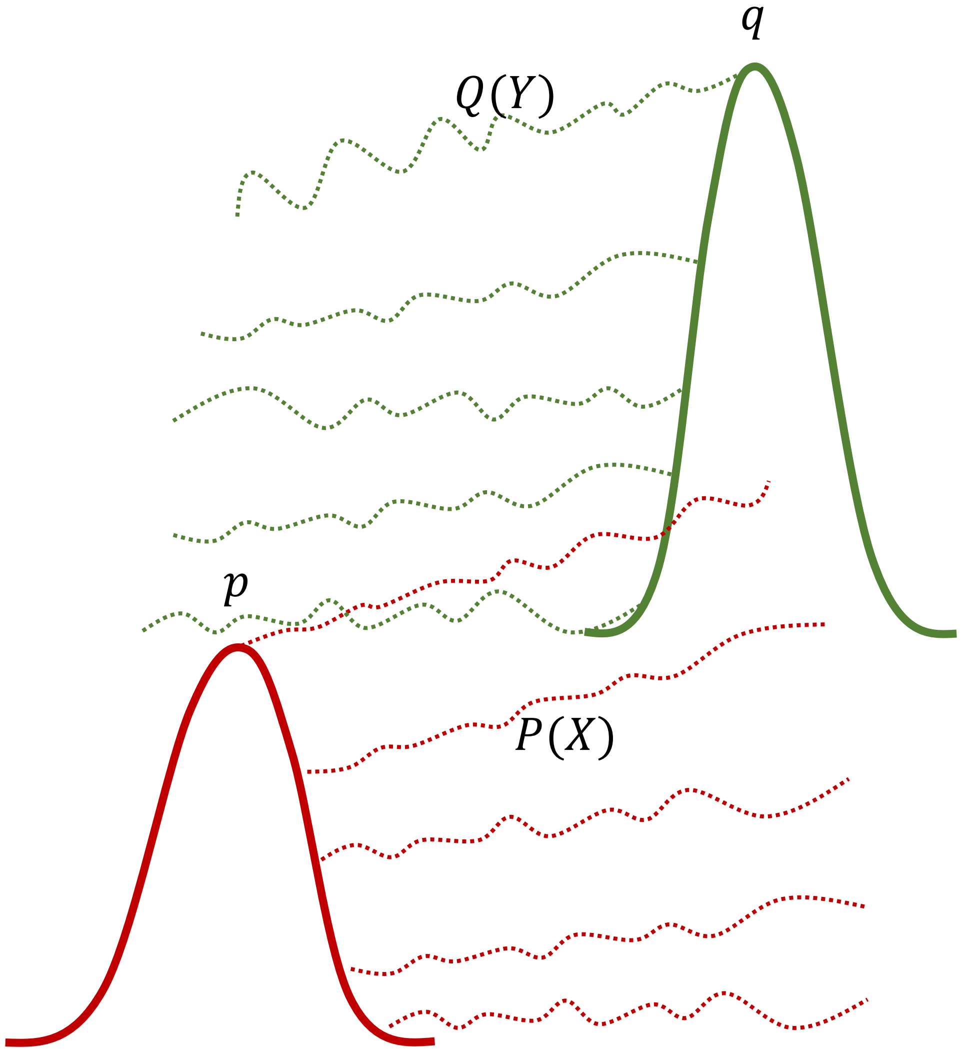
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(4) Calculate MH correction

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

if $P \approx Q, \alpha \approx 1$ 🥳



Parallel tempering Swap in Path Space

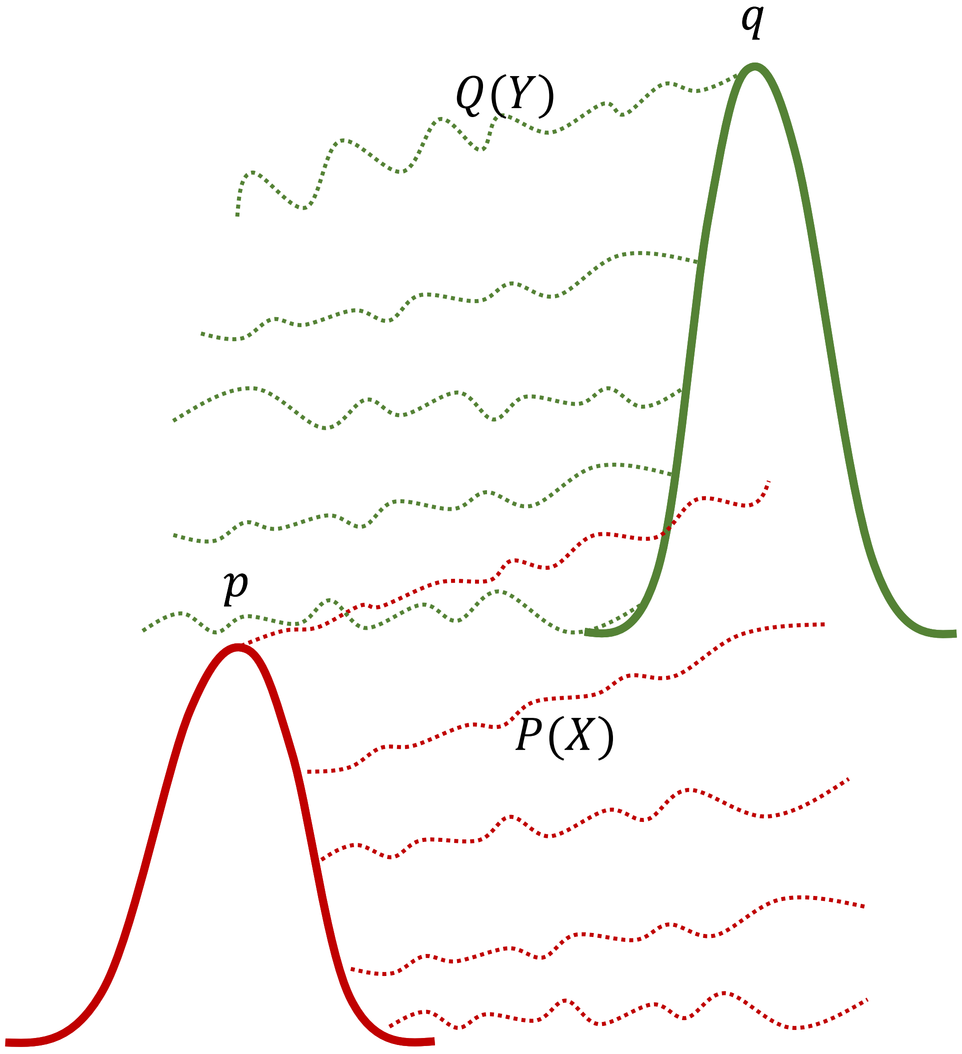
Path measure 1: P

Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

(4) Calculate MH correction

$$\alpha = \min\{1, \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{Z_q}{Z_p}\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

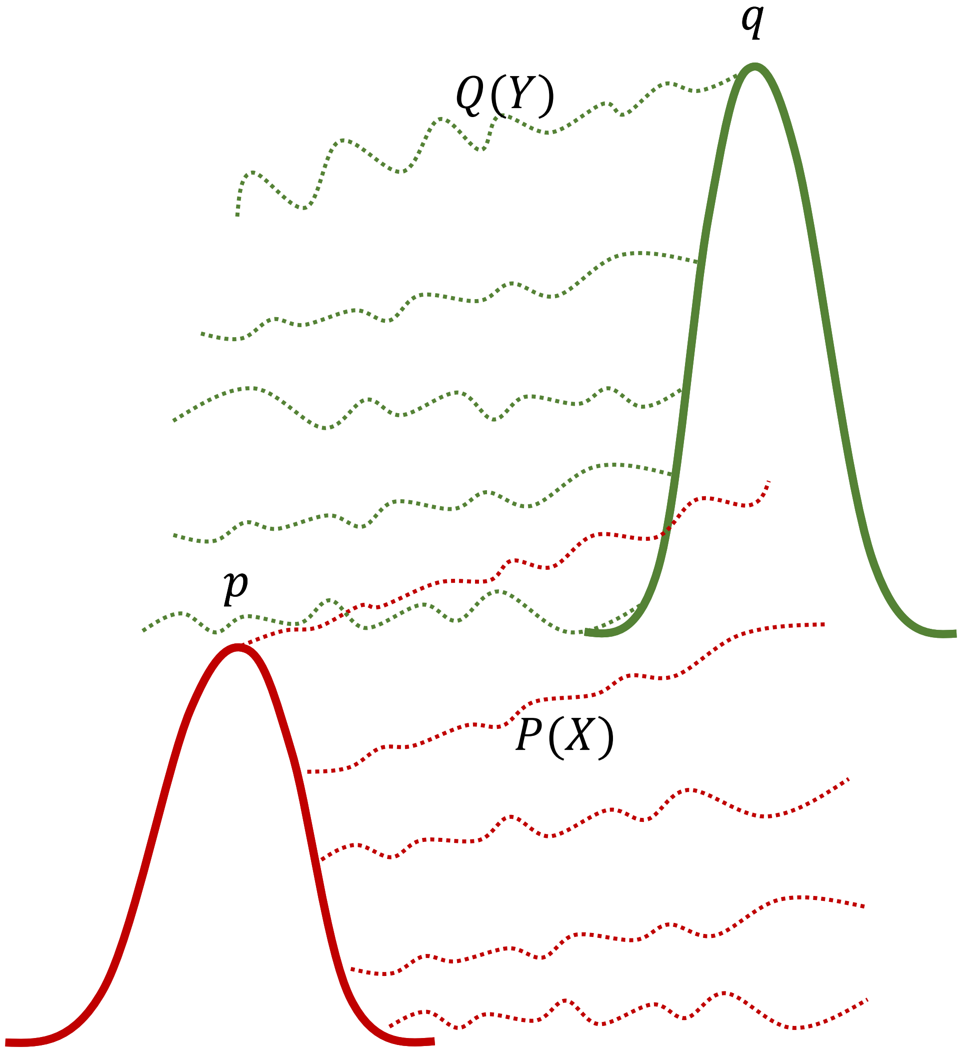


Parallel tempering Swap in Path Space

Path measure 1: P

Path measure 2: Q

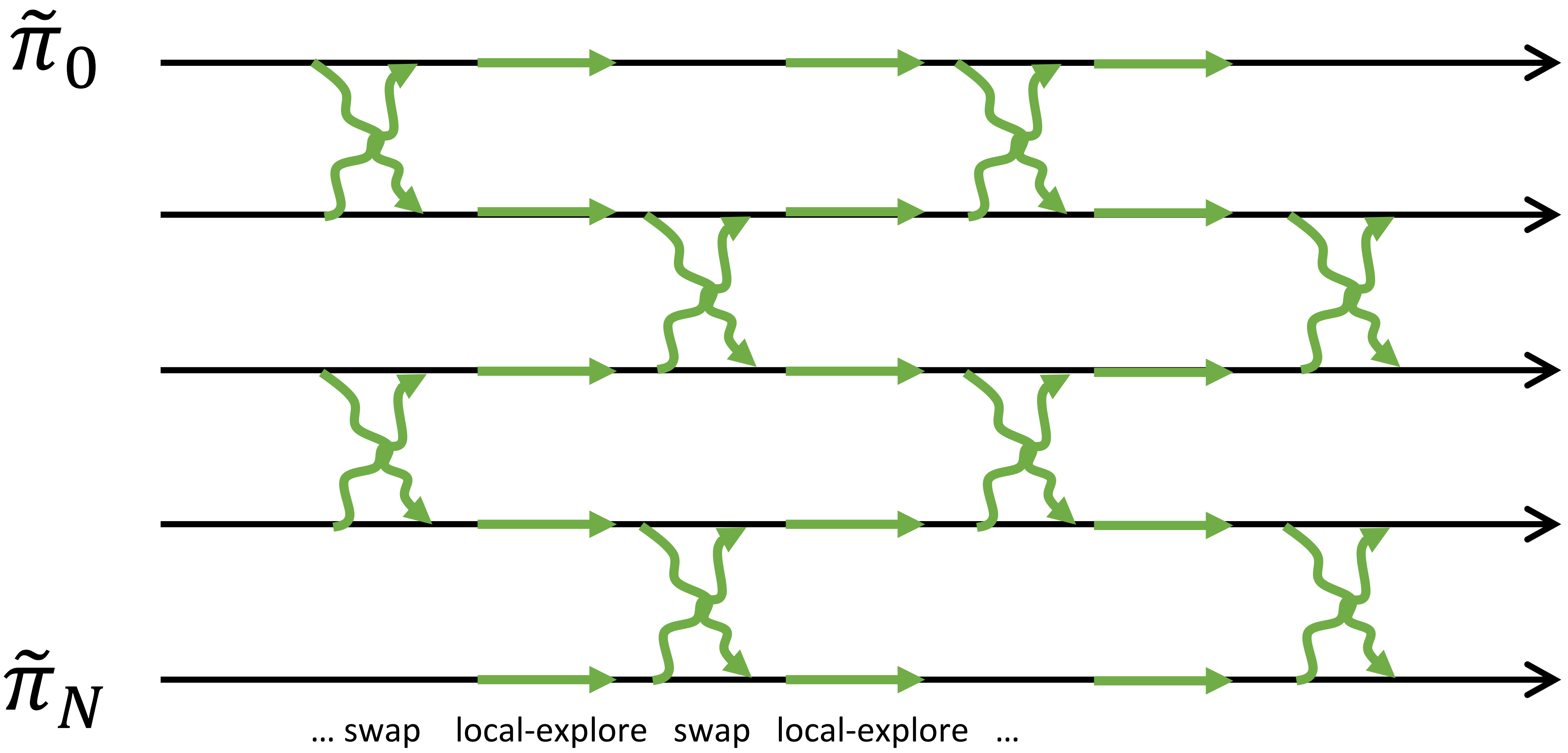
"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q} \frac{\mathrm{d}P}{\mathrm{d}Q}(X)$



How to realise the path?

CMCD Path / Diffusion Path / etc...

Accelerated Parallel tempering in Path Space



Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control

- Our setup so far:
  - Given unnormalised density, generated samples from it
- Diffusion test-time control:
  - Given a pretrained diffusion, steer distribution of generated samples

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control

- Our setup so far:
  - Given unnormalised density, generated samples from it

- Diffusion test-time control:
  - Given a pretrained diffusion, steer distribution of generated samples

tempering:	$\pi_0(x) \propto p_0^j(x)^\beta$ with inverse-temperature $\beta > 0$;
reward-tilting/posterior sampling:	$\pi_0(x) \propto p_0^j(x) \exp(r_0(x))$ with reward/likelihood $r_0(x)$;
model composition:	$\pi_0(x) \propto \prod_j p_0^j(x)$ composing J diffusions $p_0^j, j = 1, \cdots, J$.

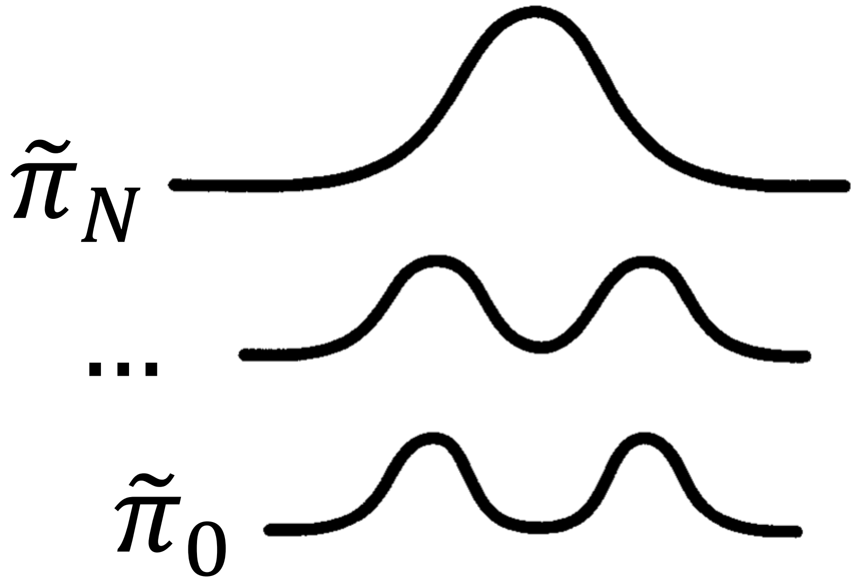
Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control

tempering: $\pi_0(x) \propto p_0^j(x)^\beta$ with inverse-temperature $\beta > 0$;

reward-tilting/posterior sampling: $\pi_0(x) \propto p_0^j(x)\exp(r_0(x))$ with reward/likelihood $r_0(x)$;

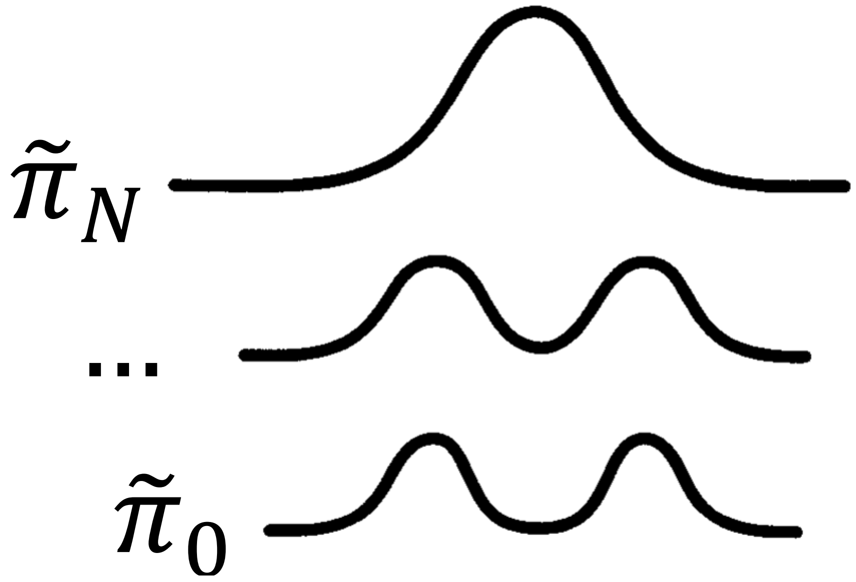
model composition: $\pi_0(x) \propto \prod_j p_0^j(x)$ composing J diffusions $p_0^j, j = 1, \cdots, J$.



Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control

tempering:	$\pi_0(x) \propto p_0^j(x)^\beta$ with inverse-temperature $\beta > 0$;
reward-tilting/posterior sampling:	$\pi_0(x) \propto p_0^j(x)\exp(r_0(x))$ with reward/likelihood $r_0(x)$;
model composition:	$\pi_0(x) \propto \prod_j p_0^j(x)$ composing J diffusions $p_0^j, j = 1, \cdots, J$.

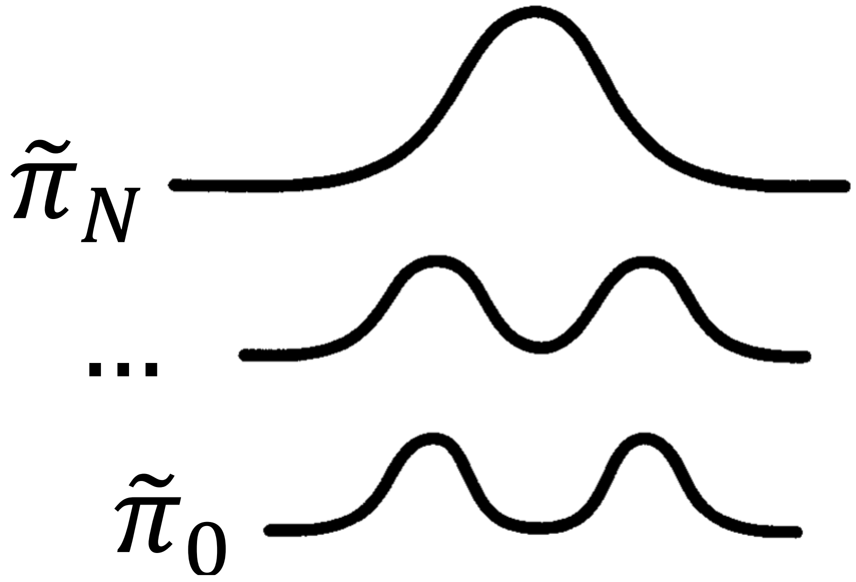


tempering:	$\pi_t(x) \propto p_t^j(x)^\beta$ with inverse-temperature $\beta > 0$;
reward-tilting/posterior sampling:	$\pi_t(x) \propto p_t^j(x)\exp(r_t(x))$ with reward/likelihood $r_t(x)$;
model composition:	$\pi_t(x) \propto \prod_j p_t^j(x)$ composing J diffusions $p_t^j, j = 1, \cdots, J$.

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control

tempering:	$\pi_0(x) \propto p_0^j(x)^\beta$ with inverse-temperature $\beta > 0$;
reward-tilting/posterior sampling:	$\pi_0(x) \propto p_0^j(x)\exp(r_0(x))$ with reward/likelihood $r_0(x)$;
model composition:	$\pi_0(x) \propto \prod_j p_0^j(x)$ composing J diffusions $p_0^j, j = 1, \cdots, J$.



tempering: $\pi_t(x) \propto p_t^j(x)^\beta$ with inverse-temperature $\beta > 0$;

model composition: $\pi_t(x) \propto \prod_j p_t^j(x)$ composing J diffusions $p_t^j, j = 1, \cdots, J$.

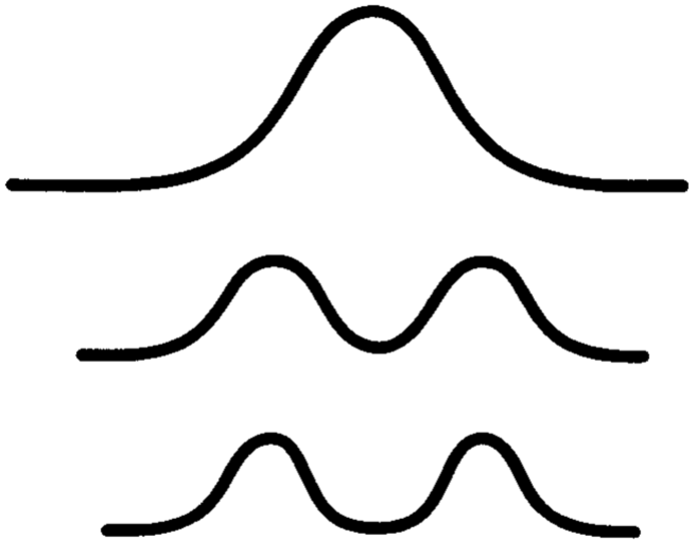
In short, control the marginal of each denoising step using APT

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)



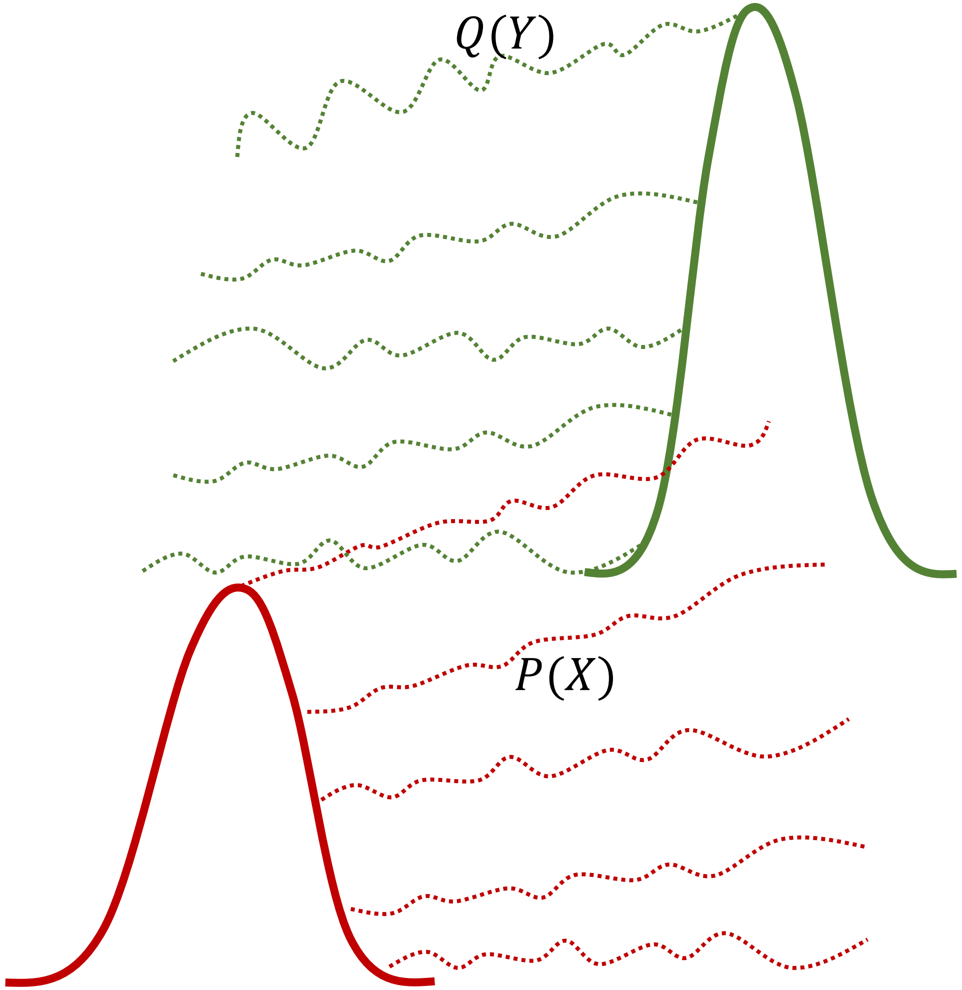
$$\tilde{\pi}_N \propto p_N \exp(r_N)$$

...

$$\tilde{\pi}_0 \propto p_0 \exp(r)$$

Accelerated Parallel tempering in Path Space

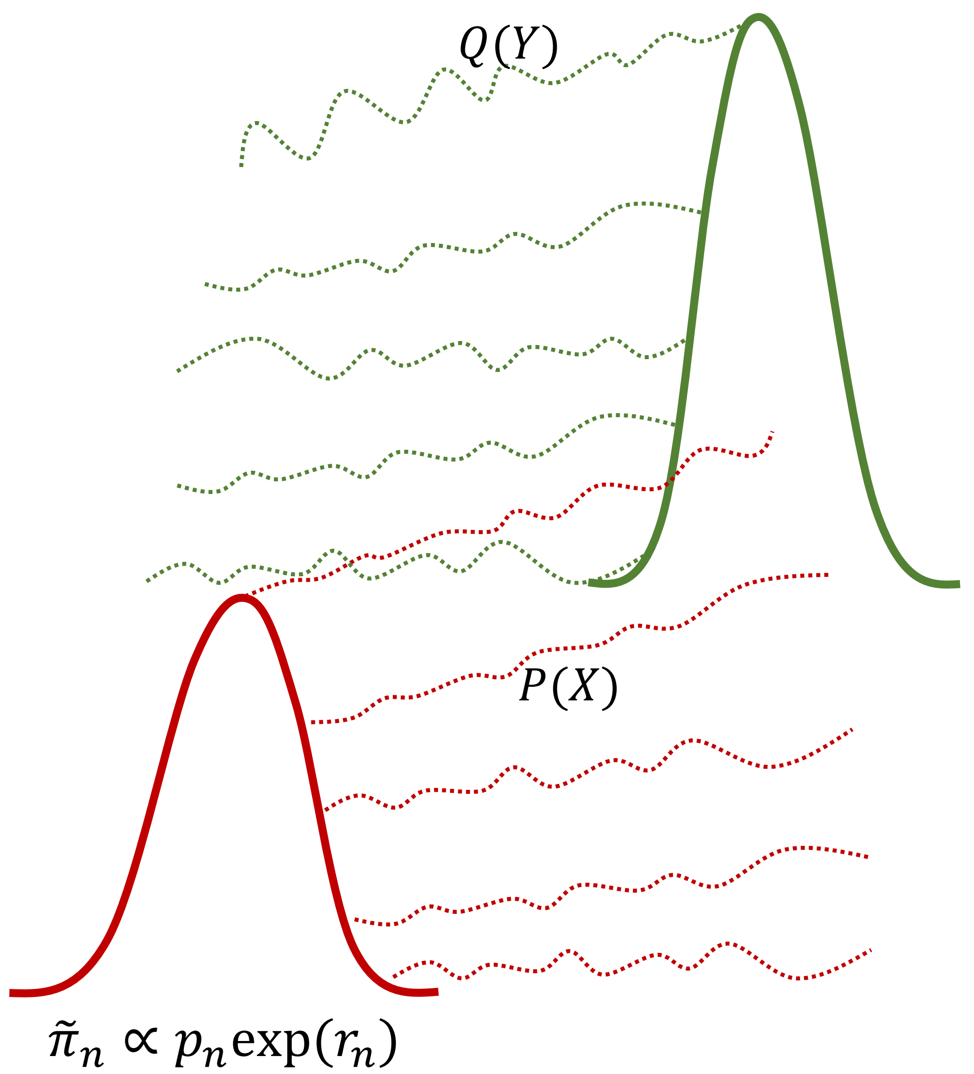
For Diffusion Test-time Control (reward-tilting as example)



Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

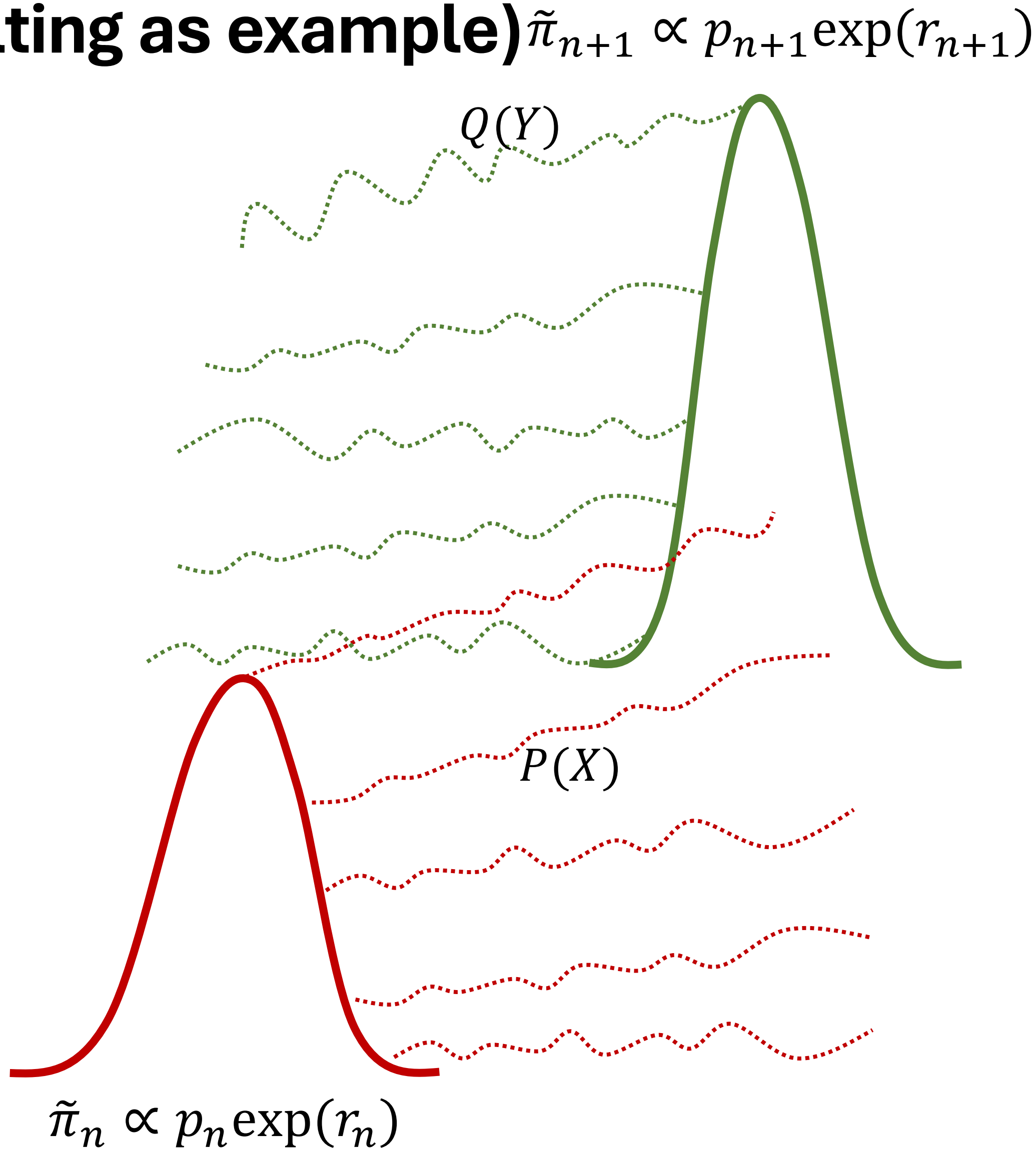


Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q} \propto \frac{\tilde{\pi}_n(X_0)}{\tilde{\pi}_{n+1}(X_1)} \lim \frac{\prod N_1(X_{k+1}|X_k)}{\prod N_2(X_k|X_{k+1})}$$

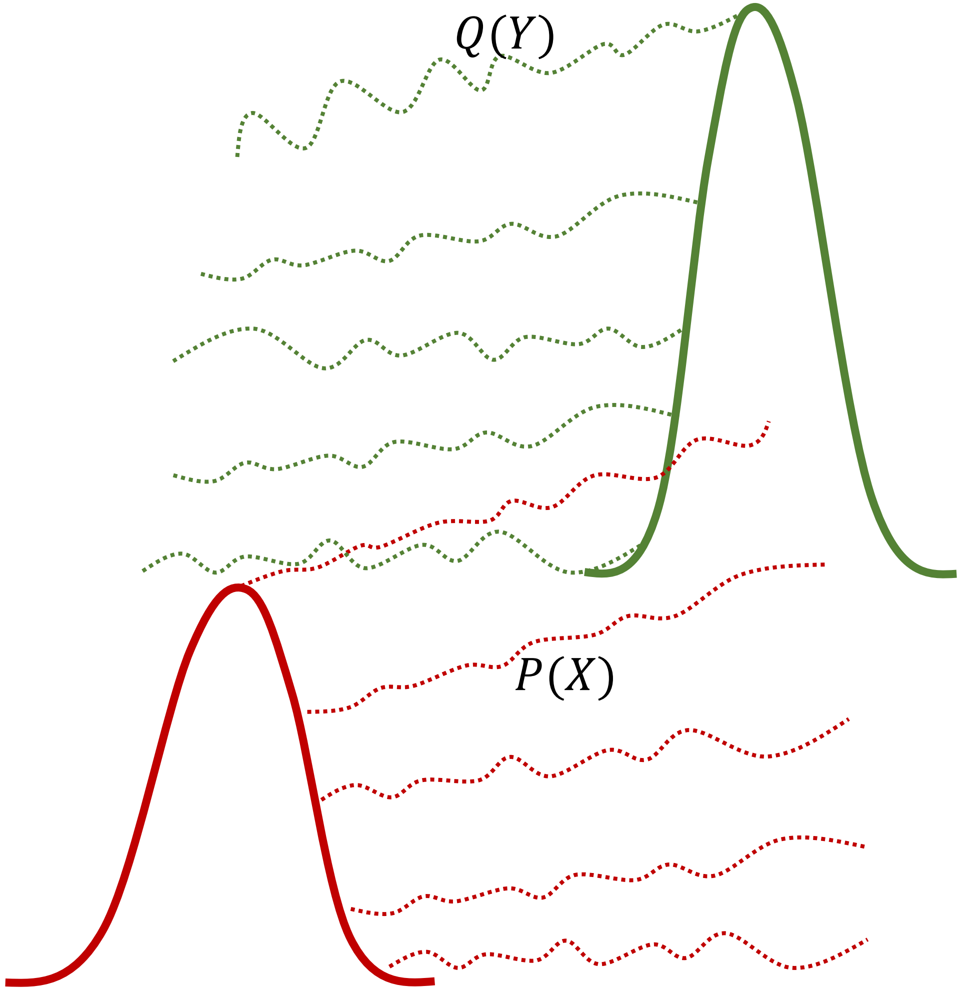


Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example) $\tilde{\pi}_{n+1} \propto p_{n+1}\exp(r_{n+1})$

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

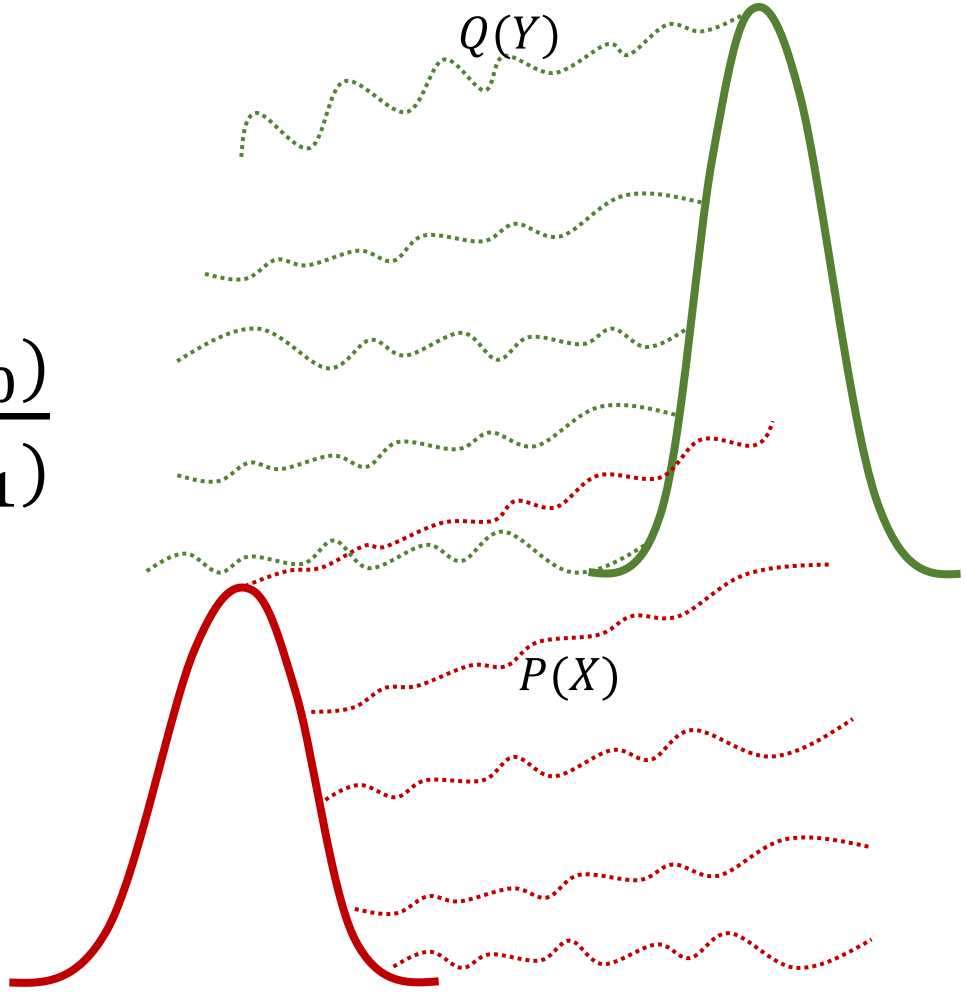
$$\frac{\mathrm{d}P}{\mathrm{d}Q} \propto \frac{\tilde{\pi}_n(X_0)}{\tilde{\pi}_{n+1}(X_1)}\frac{N_1(X_1|X_0)}{N_2(X_0|X_1)}$$



$\tilde{\pi}_n \propto p_n\exp(r_n)$

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)



$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q} \propto \frac{p_n(X_0)}{p_{n+1}(X_1)}\frac{\exp(r_n(X_0))}{\exp(r_{n+1}(X_0))}\frac{N_1(X_1|X_0)}{N_2(X_0|X_1)}$$

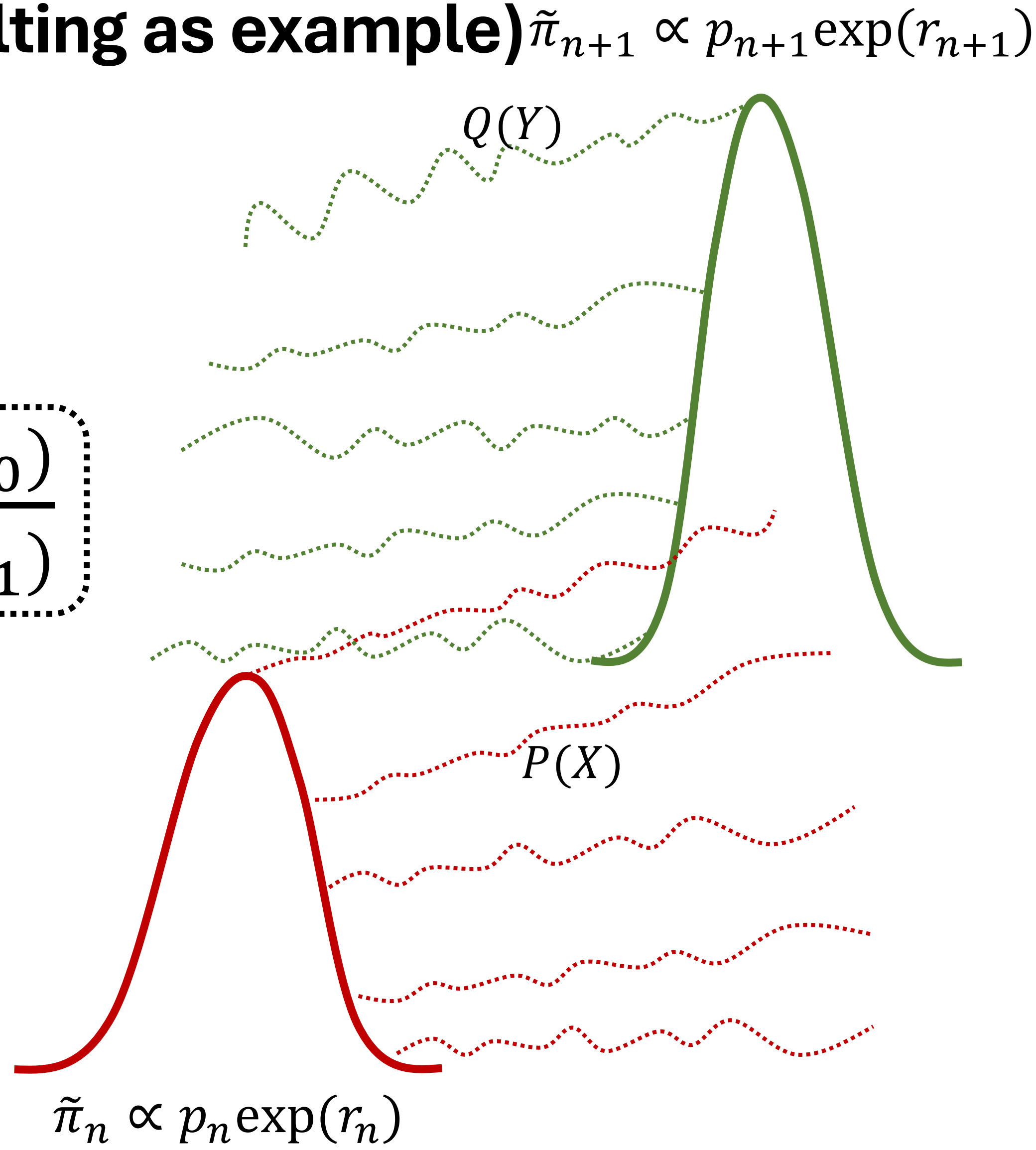
Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q} \propto \frac{p_n(X_0)}{p_{n+1}(X_1)} \boxed{\frac{\exp(r_n(X_0))}{\exp(r_{n+1}(X_0))} \frac{N_1(X_1|X_0)}{N_2(X_0|X_1)}}$$

known

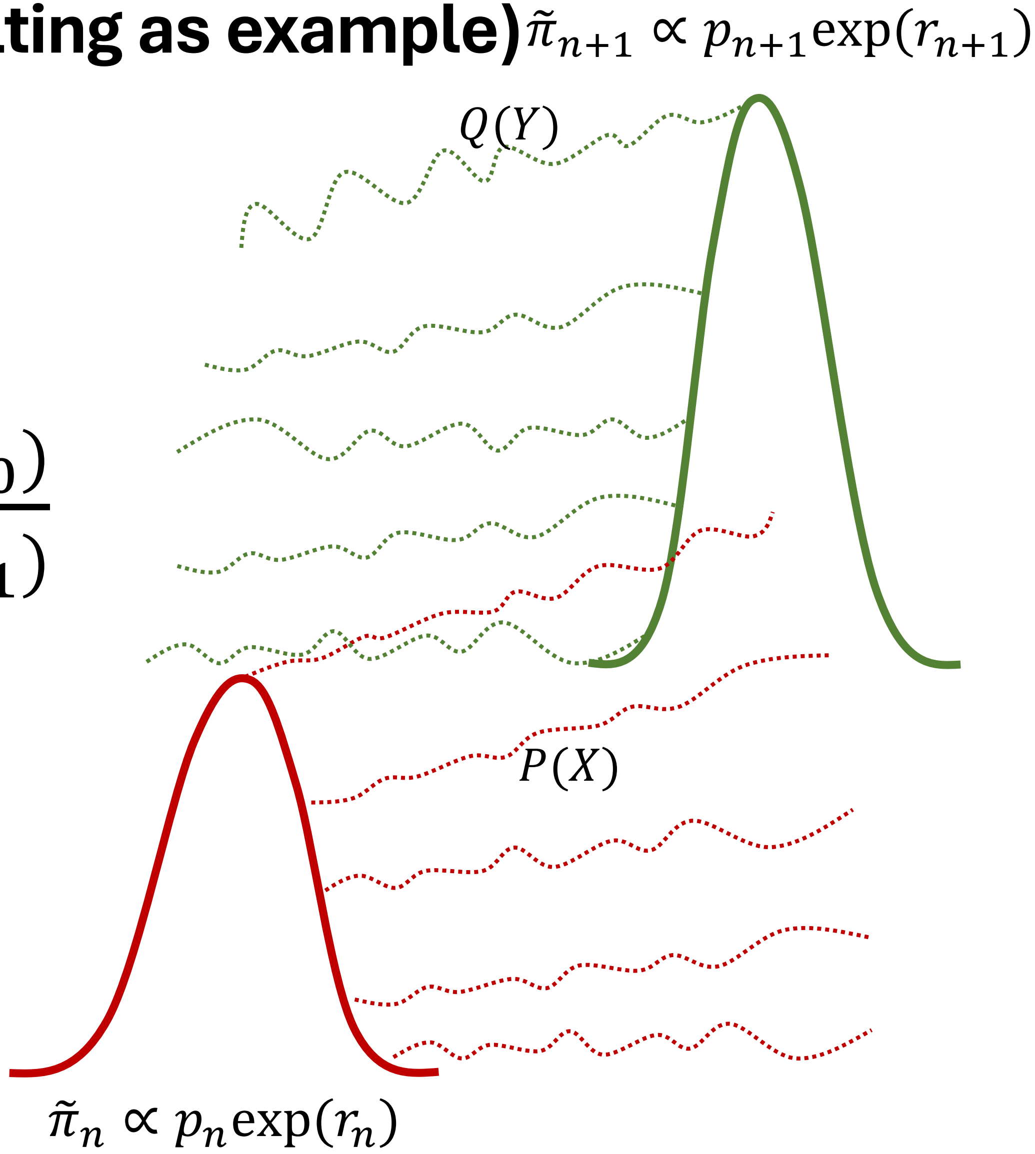


Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q} \propto \underbrace{\boxed{\frac{p_n(X_0)}{p_{n+1}(X_1)}}}_{\text{unknown}} \frac{\exp(r_n(X_0))}{\exp(r_{n+1}(X_0))} \frac{N_1(X_1|X_0)}{N_2(X_0|X_1)}$$



Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\overleftarrow{P}:\ \mathrm{d}X_t = \text{diffusion denoising drift}\,\mathrm{d}t + \sigma_t \mathrm{d}\overleftarrow{W_t} \qquad X_1 \sim p_{n+1}$$

$$P:\ \mathrm{d}X_t = \text{diffusion noising drift}\,\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_n$$

$$\frac{p_n\,(X_0)}{p_{n+1}(X_1)} = \mathbf{?}$$

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\overleftarrow{P}\text{: } \mathrm{d}X_t = \text{diffusion denoising drift}\,\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t} \qquad X_1 \sim p_{n+1}$$

$$P\text{: } \mathrm{d}X_t = \text{diffusion noising drift}\,\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_n$$

$$\Rightarrow \quad \frac{\overleftarrow{dP}}{dP} = 1$$

$$\frac{p_n(X_0)}{p_{n+1}(X_1)} = \textbf{?}$$

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\overleftarrow{P}\colon\ \mathrm{d}X_t = \text{diffusion denoising drift}\,\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t} \qquad X_1 \sim p_{n+1}$$

$$P\colon\ \mathrm{d}X_t = \text{diffusion noising drift}\,\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_n$$

$$\Rightarrow \quad \frac{\overleftarrow{dP}}{dP} = 1$$

$$\frac{p_n\,(X_0)}{p_{n+1}(X_1)}\,\frac{N_{\text{noise}}(X_1|X_0)}{N_{\text{denoise}}(X_0|X_1)} \approx 1$$

Accelerated Parallel tempering in Path Space

For Diffusion Test-time Control (reward-tilting as example)

$$\overleftarrow{P}: \ \mathrm{d}X_t = \text{diffusion denoising drift}\,\mathrm{d}t + \sigma_t \overleftarrow{\mathrm{d}W_t} \qquad X_1 \sim p_{n+1}$$

$$P: \ \mathrm{d}X_t = \text{diffusion noising drift}\,\mathrm{d}t + \sigma_t \mathrm{d}W_t \qquad X_0 \sim p_n$$

$$\Rightarrow \quad \frac{\overleftarrow{dP}}{dP} = 1$$

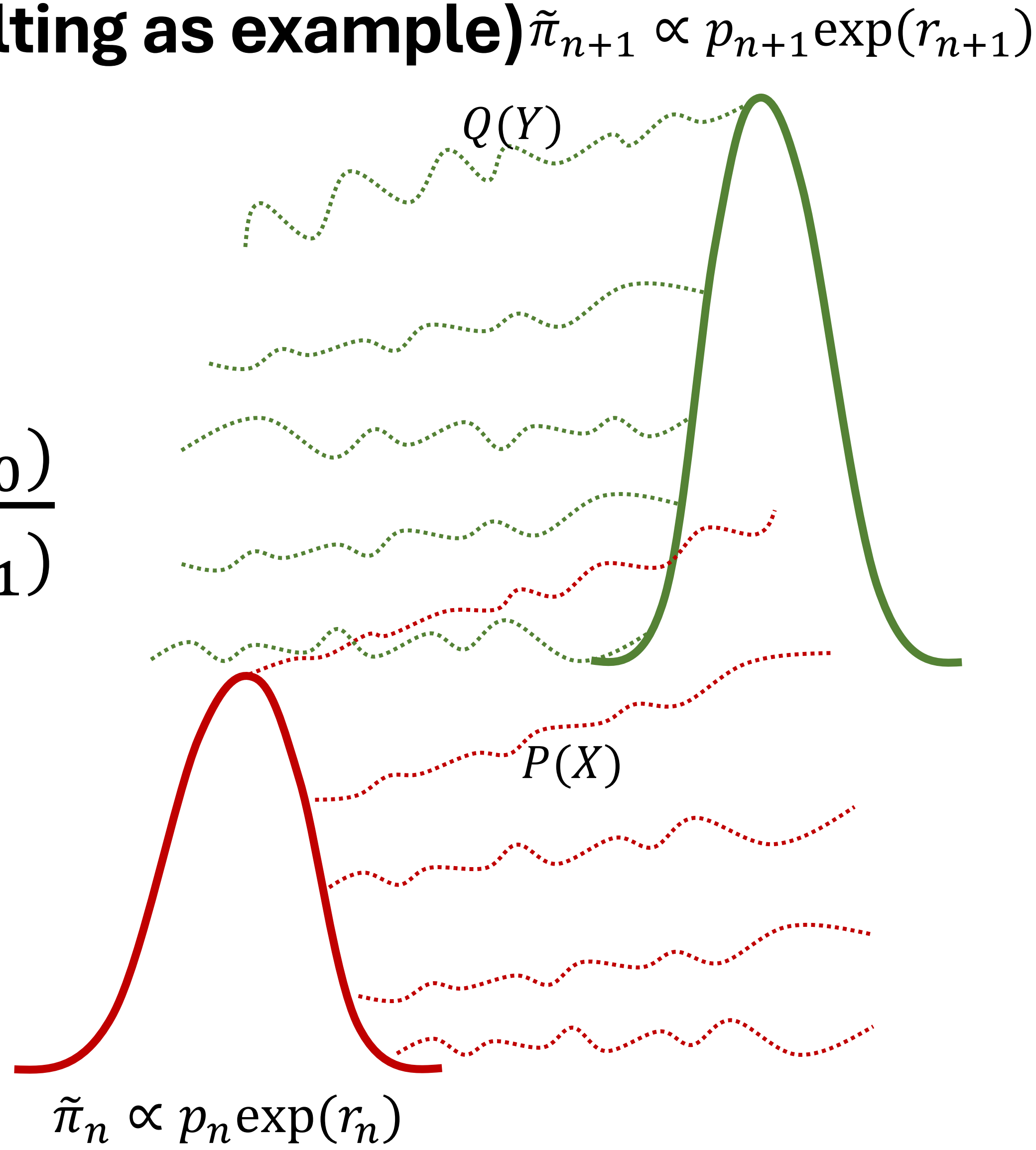
$$\frac{p_n(X_0)}{p_{n+1}(X_1)} \approx \frac{N_{\mathrm{denoise}}(X_0|X_1)}{N_{\mathrm{noise}}(X_1|X_0)}$$

Accelerated Parallel tempering in Path Space

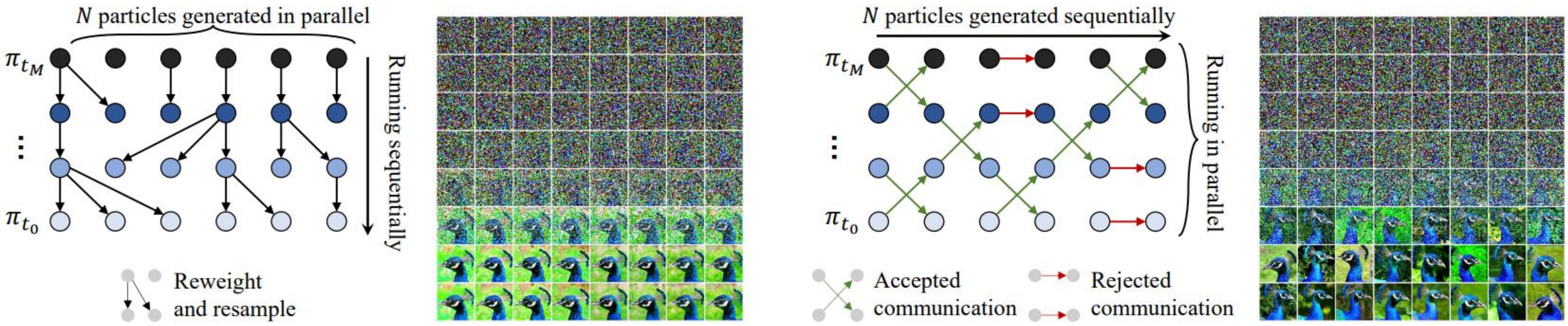
For Diffusion Test-time Control (reward-tilting as example)

$$\alpha = \min\{1, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\,\frac{\mathrm{d}Q}{\mathrm{d}P}(X)\}$$

$$\frac{\mathrm{d}P}{\mathrm{d}Q} \propto \underbrace{\boxed{\frac{p_n(X_0)}{p_{n+1}(X_1)}}}_{\text{unknown}}\,\frac{\exp(r_n(X_0))}{\exp(r_{n+1}(X_0))}\,\frac{N_1(X_1|X_0)}{N_2(X_0|X_1)}$$



CREPE:
Controlling Diffusion with Replica Exchange



CREPE:
Controlling Diffusion with Replica Exchange

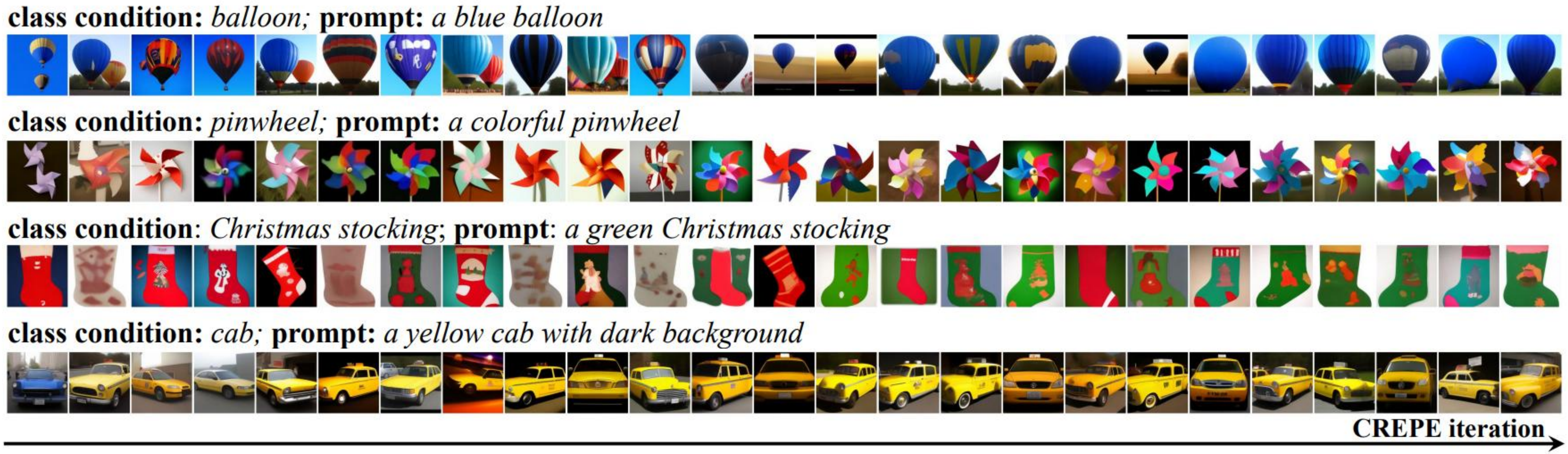
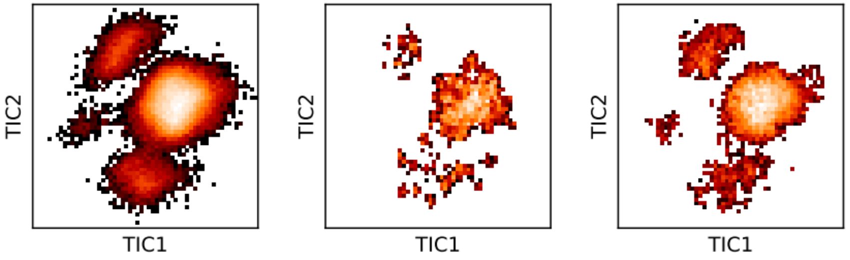


Figure 1: Trajectory of images generated using CREPE for prompted reward-tilting on ImageNet-512, thinned every 8 iterations. After burn-in, the samples align closely with the prompt.

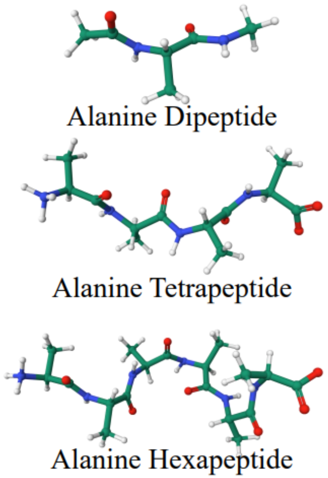
CREPE:
Controlling Diffusion with Replica Exchange



(a) MD (b) SMC (c) CREPE

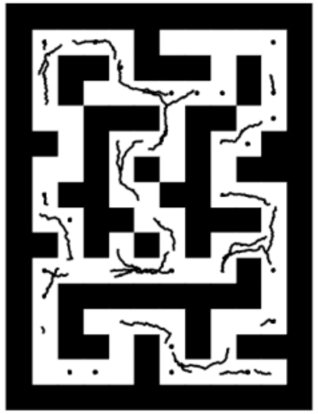
Figure 3: TICA of Alanine Hexapeptide annealed to 600K by SMC and CREPE. CREPE maintains more diversity.

Table 1: Inference-time tempering performance for Alanine Dipeptide, Tetrapeptide and Hexapeptide.

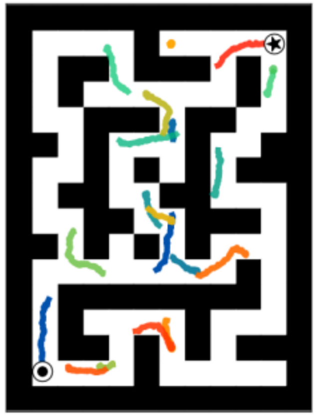


		FKC		**RNE**	**CREPE** (Ours)
		Anneal Score	Anneal Noise		
ALA Dipeptide (800K → 300K)	Energy TVD	0.345 $\pm$ 0.010	0.894 $\pm$ 0.002	0.391 $\pm$ 0.006	**0.224** $\pm$ 0.005
	Distance TVD	0.023 $\pm$ 0.001	0.036 $\pm$ 0.001	0.024 $\pm$ 0.001	**0.019** $\pm$ 0.000
	Sample W2	0.293 $\pm$ 0.001	0.282 $\pm$ 0.001	0.282 $\pm$ 0.001	**0.264** $\pm$ 0.001
	TICA MMD	0.116 $\pm$ 0.003	0.108 $\pm$ 0.004	0.168 $\pm$ 0.007	**0.096** $\pm$ 0.014
ALA Tetrapeptide (800K → 500K)	Energy TVD	**0.122** $\pm$ 0.012	0.436 $\pm$ 0.007	0.154 $\pm$ 0.006	**0.122** $\pm$ 0.004
	Distance TVD	**0.014** $\pm$ 0.000	0.015 $\pm$ 0.000	**0.013** $\pm$ 0.001	**0.013** $\pm$ 0.001
	Sample W2	0.923 $\pm$ 0.008	0.892 $\pm$ 0.001	0.893 $\pm$ 0.005	**0.856** $\pm$ 0.004
	TICA MMD	0.183 $\pm$ 0.020	0.138 $\pm$ 0.017	0.155 $\pm$ 0.009	**0.035** $\pm$ 0.002
ALA Hexapeptide (800K → 600K)	Energy TVD	**0.091** $\pm$ 0.006	0.206 $\pm$ 0.005	**0.087** $\pm$ 0.003	0.398 $\pm$ 0.001
	Distance TVD	0.018 $\pm$ 0.000	0.020 $\pm$ 0.001	**0.010** $\pm$ 0.001	**0.009** $\pm$ 0.001
	Sample W2	1.585 $\pm$ 0.001	1.652 $\pm$ 0.012	1.618 $\pm$ 0.001	**1.299** $\pm$ 0.004
	TICA MMD	0.088 $\pm$ 0.004	0.068 $\pm$ 0.010	0.042 $\pm$ 0.004	**0.009** $\pm$ 0.001

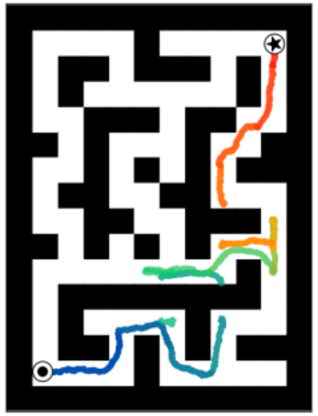
CREPE:
Controlling Diffusion with Replica Exchange



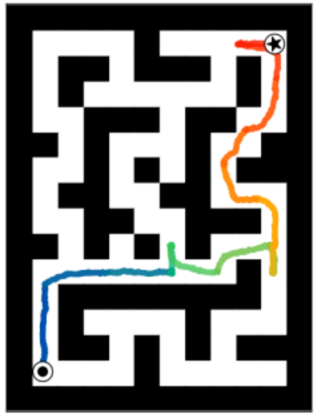
Example of training trajectories.



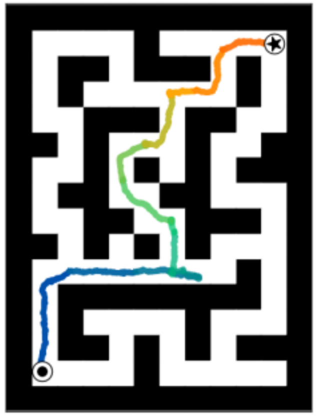
Trajectory after 1 PT iteration.



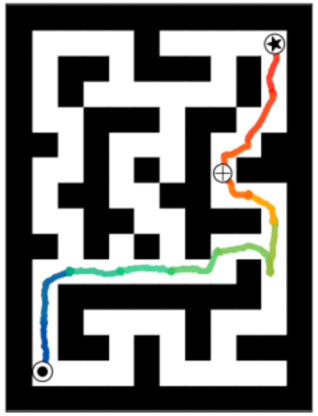
Trajectory after 10k PT iterations.



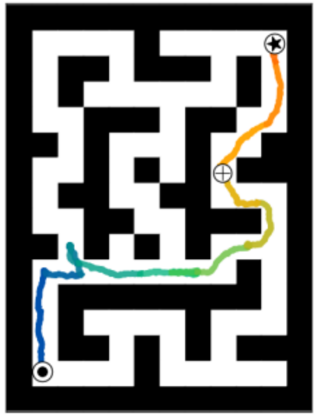
Trajectory after 50k PT iterations.



Trajectory after 100k PT iterations.



Trajectory after 101k PT iteration.



Trajectory after 150k PT iterations.

CREPE:
Controlling Diffusion with Replica Exchange

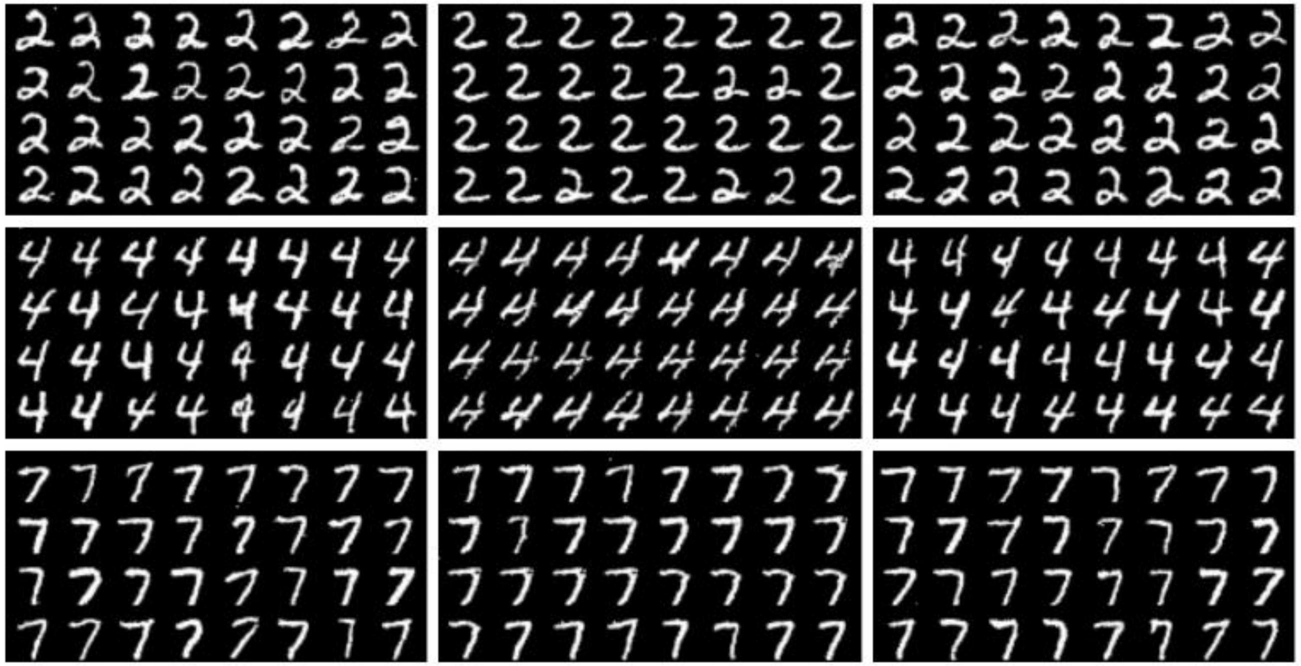
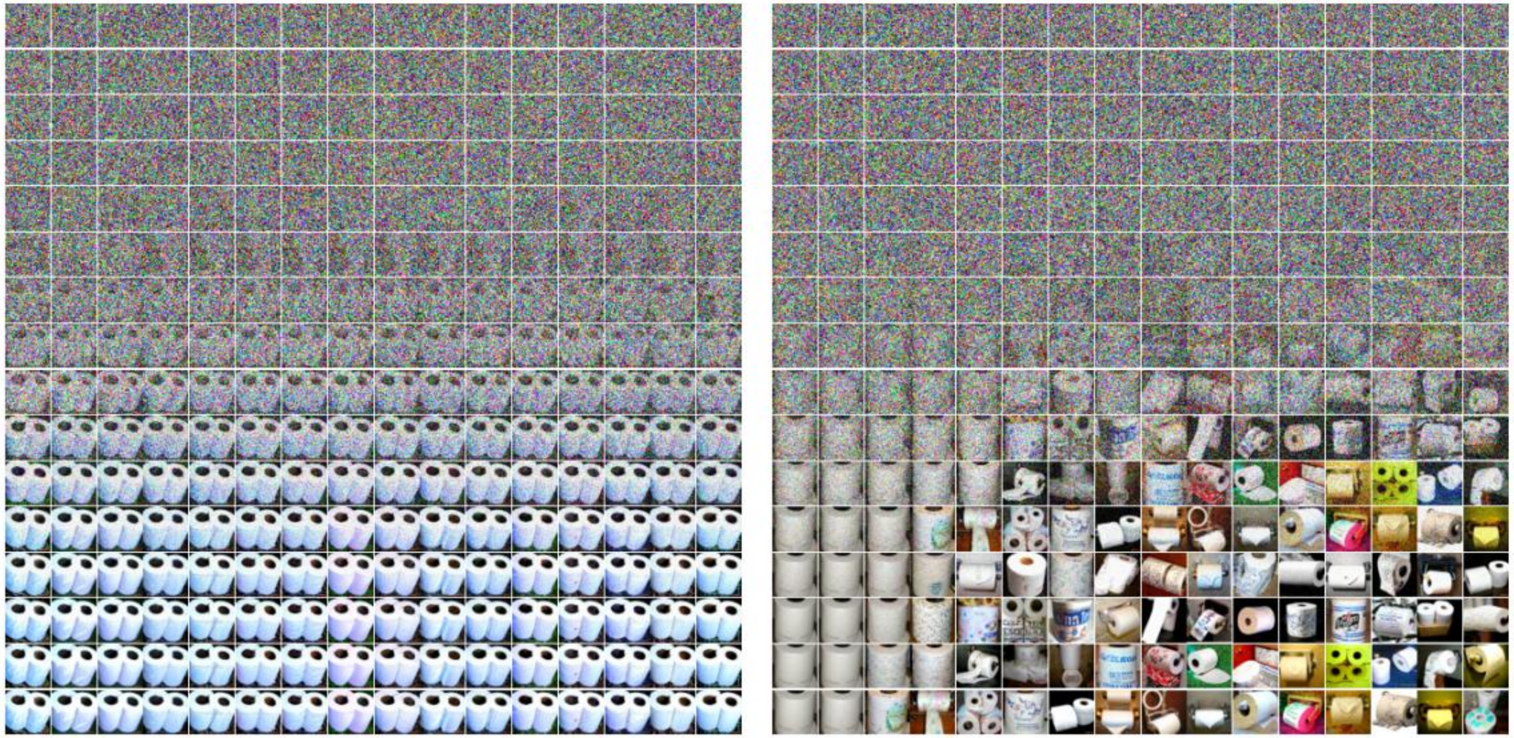


Figure 7: MNIST samples generated by CFG, and debiased by SMC and CREPE.

CREPE:
Controlling Diffusion with Replica Exchange



(a) FKC (b) CREPE

Figure 11: CFG Debiasing with FKC and CREPE for class “toilet tissue” (idx 999).

From Density Ratio to Path RND

Unnormalised density 1: $\tilde{p}$
Unnormalised density 2: $\tilde{q}$

Density ratio: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$

Path measure 1: P
Path measure 2: Q

"Unnormalised" RND: $w(X) = \frac{Z_p}{Z_q}\frac{\mathrm{d}P}{\mathrm{d}Q}(X)$

- Importance sampling: $w(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$
- FEP: $\Delta F = -\log(\int q(x) w(x)\, \mathrm{d}x)$
- PT Swap: $\alpha = \min\{1, \frac{w(y)}{w(x)}\}$

- Path Importance sampling: $w(X)$
- Path FEP: $\Delta F = -\log(\int \mathrm{d}Q(X) w(X))$
- Path PT Swap: $\alpha = \min\{1, \frac{w(Y)}{w(X)}\}$

Collaborators (random order):
Free-energy estimator with adaptive transport






Collaborators (random order):
Accelerated parallel tempering







Collaborators (random order):
Controlling diffusion with Replica Exchange

References

(1) Free energy estimation on Path:

He*, J., Du*, Y., Vargas, F., Wang, Y., Gomes, C. P., Hernández-Lobato, J. M., & Vanden-Eijnden, E. (2025). FEAT: Free energy Estimators with Adaptive Transport. *NeurIPS 2025*.

(2) Parallel Tempering on Path:

Zhang*, L., Potaptchik*, P., He*, J., Du, Y., Doucet, A., Vargas, F., ... & Syed, S. (2025). Accelerated Parallel Tempering via Neural Transports. *ICLR 2026.*

(3) Parallel Tempering for Diffusion Control:

He, J., Jeha, P., Potaptchik, P., Zhang, L., Hernández-Lobato, J. M., Du, Y., ... & Vargas, F. (2025). CREPE: Controlling Diffusion with Replica Exchange. *ICLR 2026*.