# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$

# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$
- Instead, I have samples from $g$ and the information that $f \approx g$

$$\{Y_k\}_{k=1}^{N}, \qquad Y_k \sim g$$

# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$
- Instead, I have samples from $g$ and the information that $f \approx g$

$$\{Y_k\}_{k=1}^{N}, \qquad Y_k \sim g$$

- These samples from $g$ can be **transformed** (approximately) to samples from $f$ according to :

# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$
- Instead, I have samples from $g$ and the information that $f \approx g$

$$\{Y_k\}_{k=1}^{N}, \qquad Y_k \sim g$$

- These samples from $g$ can be **transformed** (approximately) to samples from $f$ according to :

1. Define: $$w_k = \frac{f(Y_k)}{g(Y_k)} \qquad \hat{w}_k = \frac{w_k}{\sum_{i=1}^{N} w_i}$$

# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$
- Instead, I have samples from $g$ and the information that $f \approx g$

$$\{Y_k\}_{k=1}^{N}, \qquad Y_k \sim g$$

- These samples from $g$ can be **transformed** (approximately) to samples from $f$ according to :

1. Define: $$w_k = \frac{f(Y_k)}{g(Y_k)} \qquad \hat{w}_k = \frac{w_k}{\sum_{i=1}^{N} w_i}$$

2. For $k = 1, \ldots, N$ choose a new sample according to

$$X_k = \begin{cases} Y_1, & \text{w.p. } \hat{w}_1 \\ \vdots & \\ Y_N, & \text{w.p. } \hat{w}_N \end{cases}$$

# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$
- Instead, I have samples from $g$ and the information that $f \approx g$

$$\{Y_k\}_{k=1}^{N}, \qquad Y_k \sim g$$

**Sampling**

- These samples from $g$ can be **transformed** (approximately) to samples from $f$ according to :

1. Define: $$w_k = \frac{f(Y_k)}{g(Y_k)} \qquad \hat{w}_k = \frac{w_k}{\sum_{i=1}^{N} w_i}$$

2. For $k = 1, \ldots, N$ choose a new sample according to

$$X_k = \begin{cases} Y_1, & \text{w.p. } \hat{w}_1 \\ \vdots & \\ Y_N, & \text{w.p. } \hat{w}_N \end{cases}$$

# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$
- Instead, I have samples from $g$ and the information that $f \approx g$

$$\{Y_k\}_{k=1}^N, \qquad Y_k \sim g$$

**Sampling**

- These samples from $g$ can be **transformed** (approximately) to samples from $f$ according to :

1. Define: $$w_k = \frac{f(Y_k)}{g(Y_k)} \qquad \hat{w}_k = \frac{w_k}{\sum_{i=1}^N w_i}$$

**Importance**

2. For $k = 1, \ldots, N$ choose a new sample according to

$$X_k = \begin{cases} Y_1, & \text{w.p. } \hat{w}_1 \\ \vdots & \\ Y_N, & \text{w.p. } \hat{w}_N \end{cases}$$

# Sampling Importance Resampling (SIR)

- I want samples from the distribution $f$
- Instead, I have samples from $g$ and the information that $f \approx g$

$$\{Y_k\}_{k=1}^{N}, \qquad Y_k \sim g$$

**Sampling**

- These samples from $g$ can be **transformed** (approximately) to samples from $f$ according to :

1. Define:

$$w_k = \frac{f(Y_k)}{g(Y_k)} \qquad \hat{w}_k = \frac{w_k}{\sum_{i=1}^{N} w_i}$$

**Importance**

2. For $k = 1, \ldots, N$ choose a new sample according to

$$X_k = \begin{cases} Y_1, & \text{w.p. } \hat{w}_1 \\ \vdots & \\ Y_N, & \text{w.p. } \hat{w}_N \end{cases}$$

**Resampling**

# Transitional Markov Chain Monte Carlo (TMCMC)

- The goal is to sample the distribution:

$$p(\vartheta \,|\, d) \propto p(d \,|\, \vartheta)\pi(\vartheta)$$

# Transitional Markov Chain Monte Carlo (TMCMC)

- The goal is to sample the distribution:

$$p(\vartheta \,|\, d) \propto p(d \,|\, \vartheta)\pi(\vartheta)$$

- Instead, sample iteratively the annealed distribution

$$p_k(\vartheta \,|\, d) \propto p(d \,|\, \vartheta)^{\gamma_k}\pi(\vartheta), \qquad k = 1, \ldots N$$

where

$$\gamma_1 = 0 < \cdots < \gamma_k < \cdots < \gamma_N = 1$$

# Transitional Markov Chain Monte Carlo (TMCMC)

- The goal is to sample the distribution:

$$p(\vartheta \mid d) \propto p(d \mid \vartheta)\pi(\vartheta)$$

- Instead, sample iteratively the annealed distribution

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k}\pi(\vartheta), \quad k = 1, \ldots N$$

where

$$\gamma_1 = 0 < \cdots < \gamma_k < \cdots < \gamma_N = 1$$

- At the last step, we obtain samples from the target distribution.

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

**Sampling Importance Resampling**

**Annealing**

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

## Sampling Importance Resampling

1. If I have samples for $\gamma_1 = 0$ then using **SIR** I have samples approximately for $\gamma_2$

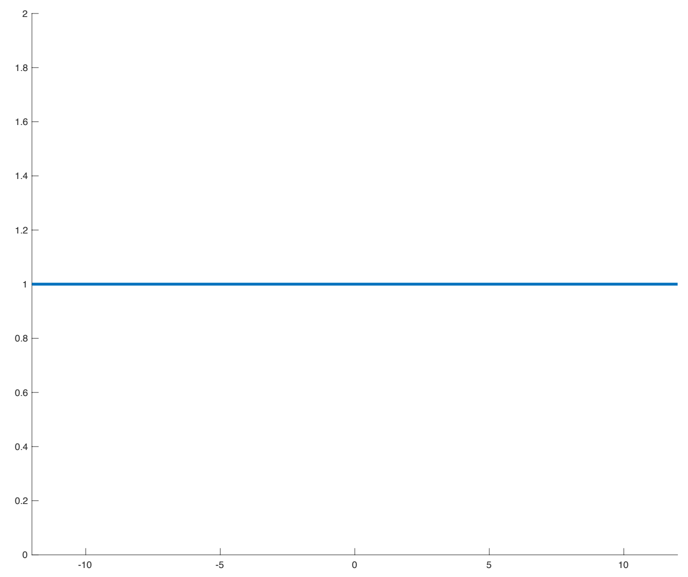
## Annealing

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

## Sampling Importance Resampling

1. If I have samples for $\gamma_1 = 0$ then using **SIR** I have samples approximately for $\gamma_2$

## Annealing



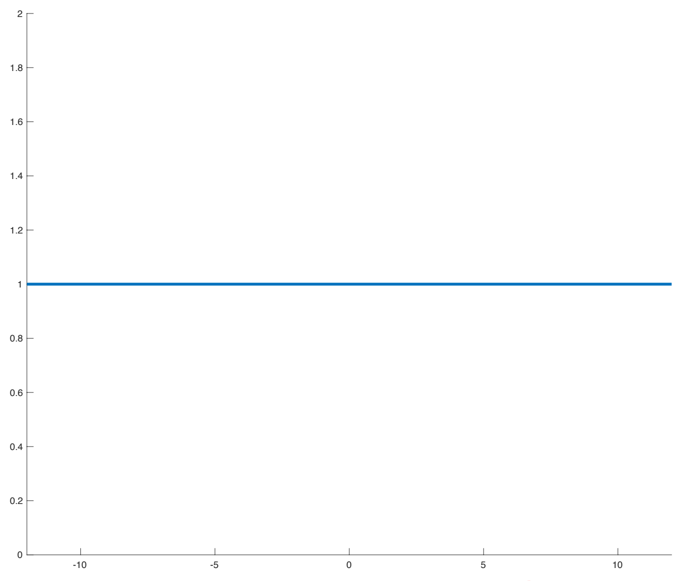
$\gamma_1 = 0$

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

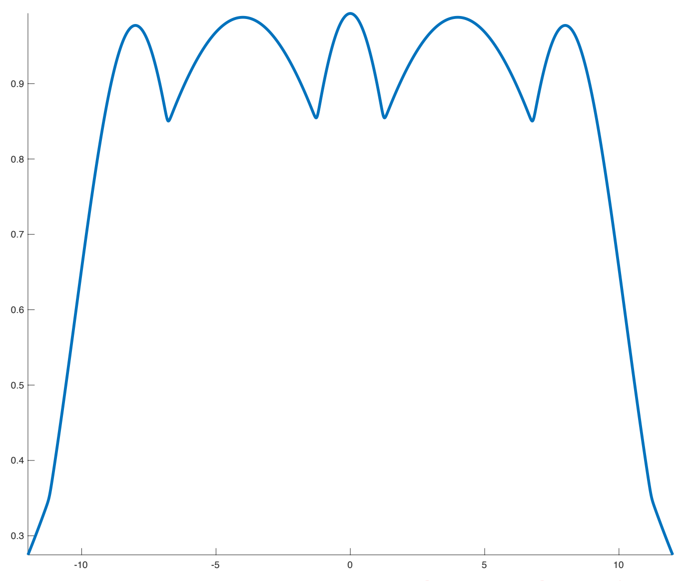
## Sampling Importance Resampling

1. If I have samples for $\gamma_1 = 0$ then using **SIR** I have samples approximately for $\gamma_2$

## Annealing



$\gamma_1 = 0$



$\gamma_2 = 0.01$

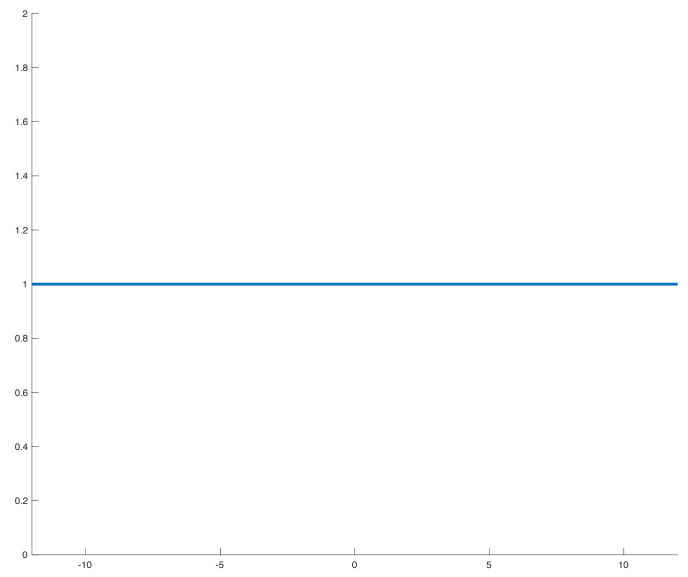
$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

## Sampling Importance Resampling

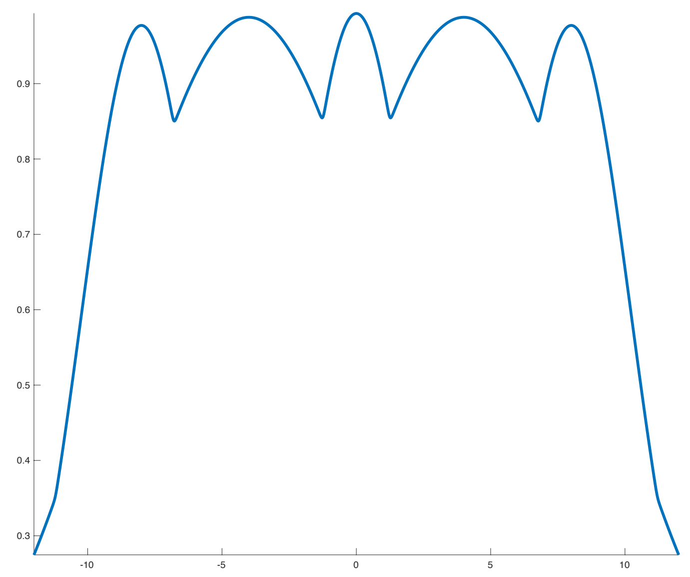
1. If I have samples for $\gamma_1 = 0$ then using **SIR** I have samples approximately for $\gamma_2$

2. If I have samples for $\gamma_2$ then using **SIR** I have samples approximately for $\gamma_3$

## Annealing



$\gamma_1 = 0$



$\gamma_2 = 0.01$

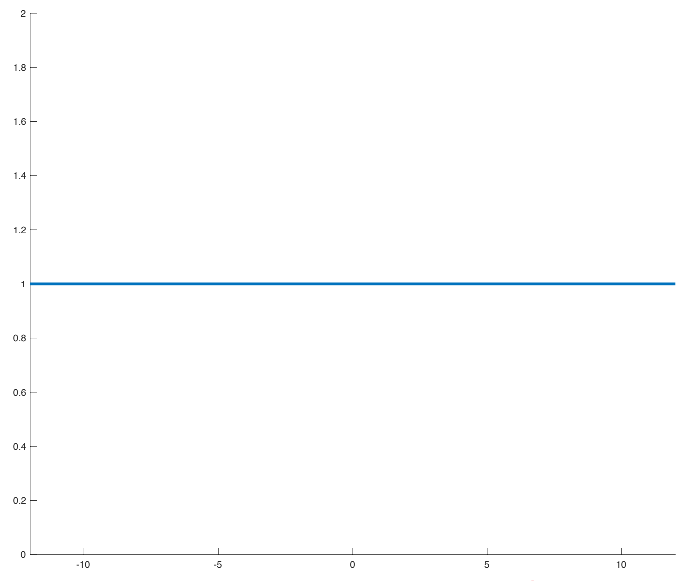
$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

## Sampling Importance Resampling

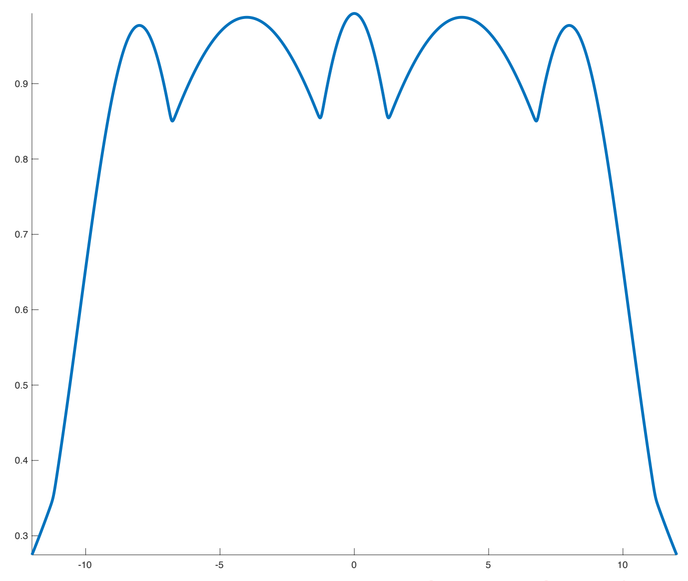
1. If I have samples for $\gamma_1 = 0$ then using **SIR** I have samples approximately for $\gamma_2$

2. If I have samples for $\gamma_2$ then using **SIR** I have samples approximately for $\gamma_3$

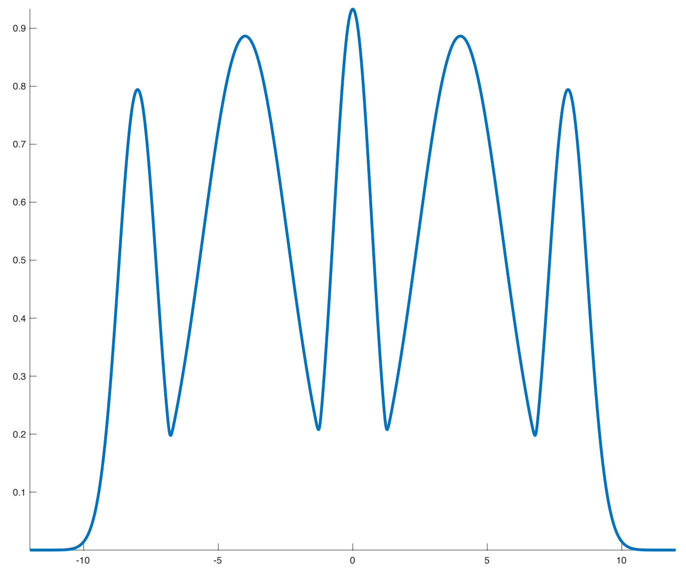
## Annealing



$\gamma_1 = 0$



$\gamma_2 = 0.01$



$\gamma_3 = 0.1$

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

## Sampling Importance Resampling

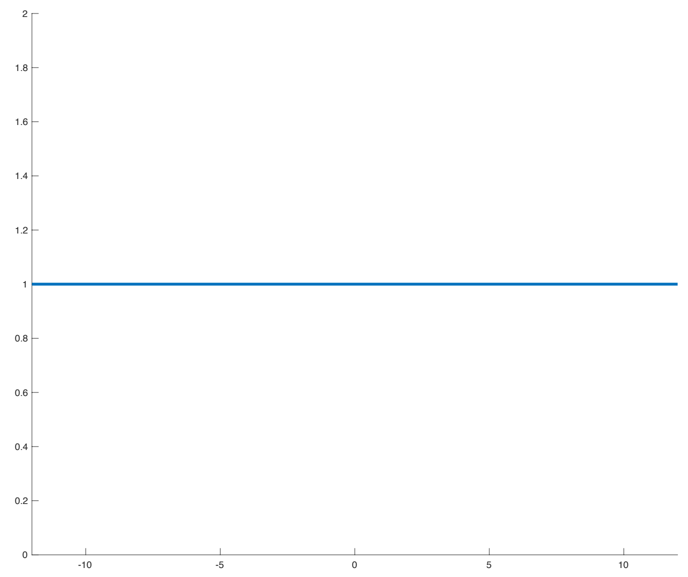
1. If I have samples for $\gamma_1 = 0$ then using **SIR**
   I have samples approximately for $\gamma_2$

2. If I have samples for $\gamma_2$ then using **SIR**
   I have samples approximately for $\gamma_3$

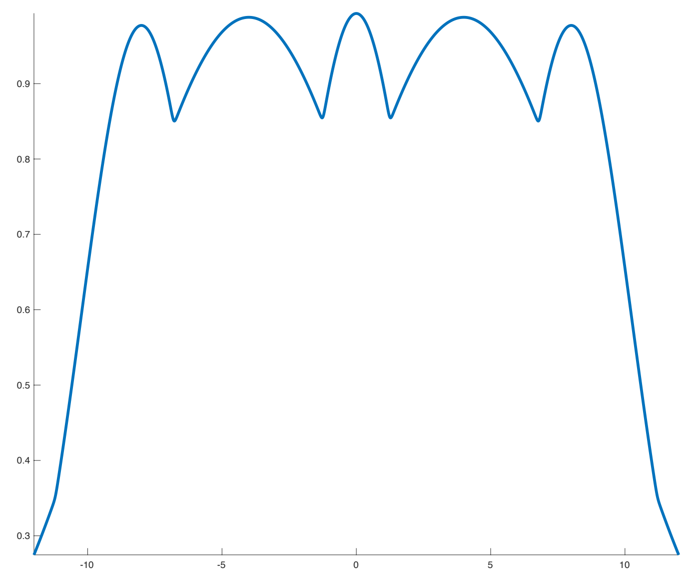
⋮

N. If I have samples for $\gamma_{N-1}$ then using **SIR**
   I have samples approximately for $\gamma_N = 1$

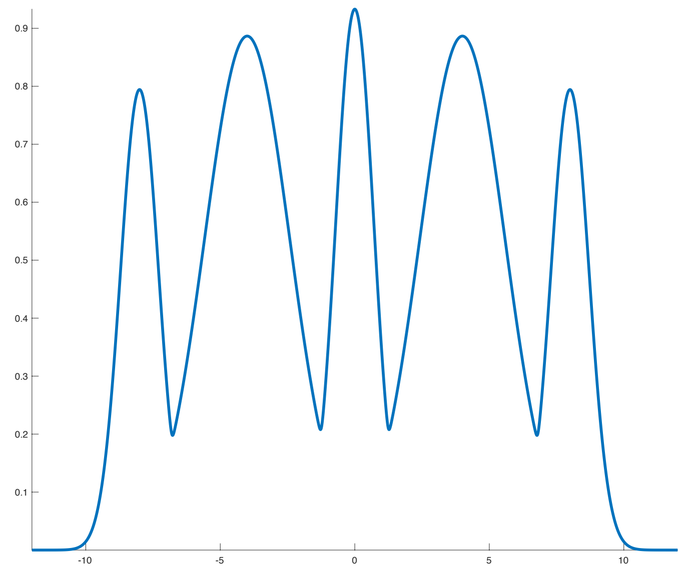
## Annealing



$\gamma_1 = 0$



$\gamma_2 = 0.01$



$\gamma_3 = 0.1$

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

## Sampling Importance Resampling

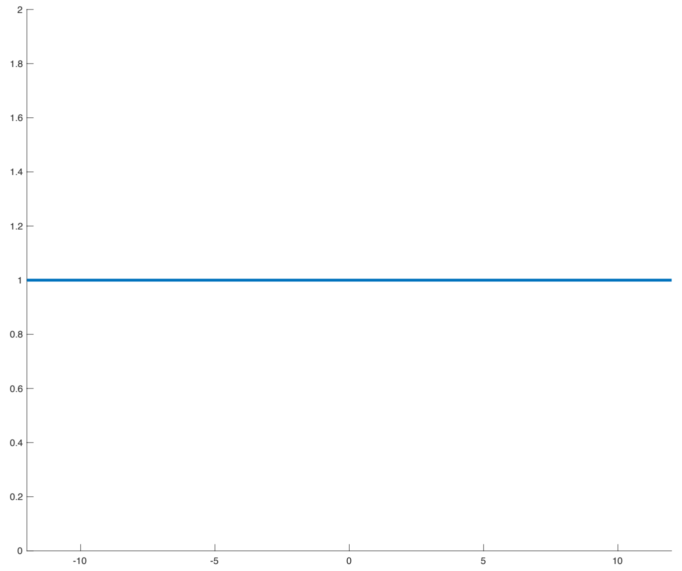
1. If I have samples for $\gamma_1 = 0$ then using **SIR**
   I have samples approximately for $\gamma_2$

2. If I have samples for $\gamma_2$ then using **SIR**
   I have samples approximately for $\gamma_3$

⋮

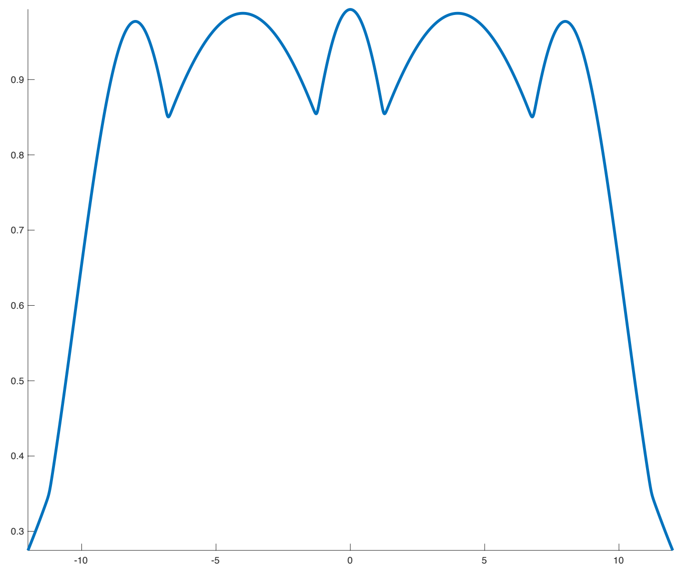
N. If I have samples for $\gamma_{N-1}$ then using **SIR**
   I have samples approximately for $\gamma_N = 1$

## Annealing



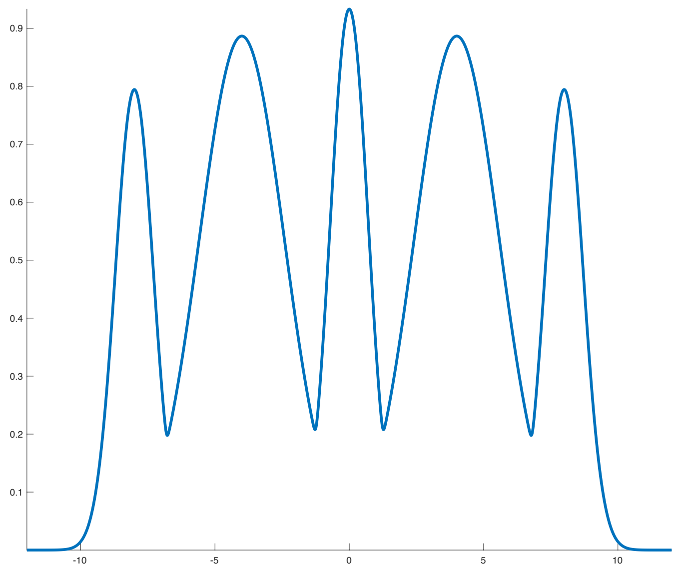
$\gamma_1 = 0$

→



$\gamma_2 = 0.01$

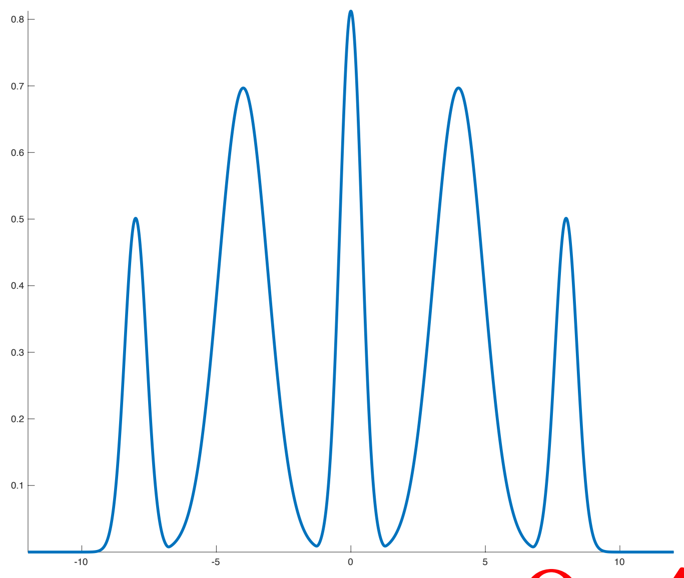
→



$\gamma_3 = 0.1$

→ . . .

→



$\gamma_{N-1} = 0.4$

$$p_k(\vartheta \mid d) \propto p(d \mid \vartheta)^{\gamma_k} \pi(\vartheta)$$

## Sampling Importance Resampling

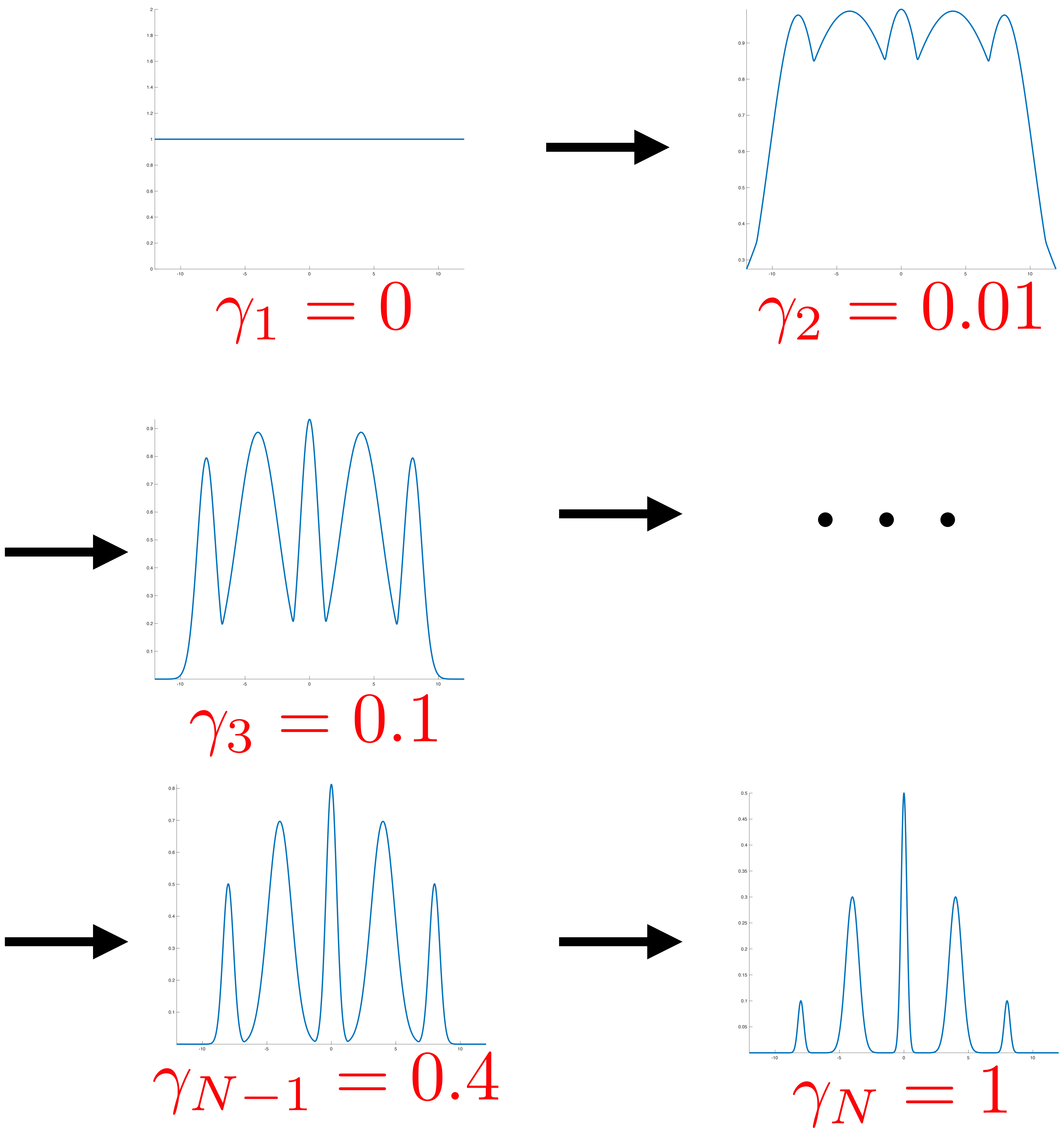
1. If I have samples for $\gamma_1 = 0$ then using **SIR**
   I have samples approximately for $\gamma_2$

2. If I have samples for $\gamma_2$ then using **SIR**
   I have samples approximately for $\gamma_3$

⋮

N. If I have samples for $\gamma_{N-1}$ then using **SIR**
   I have samples approximately for $\gamma_N = 1$

## Annealing



# TMCMC for model selection

- Model selection: an estimator for the denominator is given by,

$$p(d) \approx \prod_{k=1}^{N} S_k$$

# TMCMC for model selection

- Model selection: an estimator for the denominator is given by,

$$p(d) \approx \prod_{k=1}^{N} S_k$$

final number of generation

# TMCMC for model selection

- Model selection: an estimator for the denominator is given by,

$$p(d) \approx \prod_{k=1}^{N} S_k$$

final number of generation

where

$$S_k = \frac{1}{N} \sum_{i=1}^{N} w_i^{(k)}$$

# TMCMC for model selection

- Model selection: an estimator for the denominator is given by,

$$p(d) \approx \prod_{k=1}^{N} S_k$$

final number
of generation

where

$$S_k = \frac{1}{N} \sum_{i=1}^{N} w_i^{(k)}$$

weights at k-th
generation