$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)}$$

$$p(A \mid B, C) = \frac{p(B \mid A, C)\,p(A \mid C)}{p(B \mid C)}$$

$$p(\theta \mid D, M) = \frac{p(D \mid \theta, M)\,p(\theta \mid M)}{p(D \mid M)}$$

$$\text{Posterior} = \frac{\text{Likelihood} \quad \text{Prior}}{\text{Evidence}}$$

What we design: $p(D \mid \theta, M),\ p(\theta \mid M)$

What we design: $p(D \mid \theta, M),\ p(\theta \mid M)$

What we want: $p(\theta \mid D, M),\ p(D \mid M)$

What we design: $p(D \mid \theta, M),\ p(\theta \mid M)$

What we want: $p(\theta \mid D, M),\ p(D \mid M)$

It is straightforward to compute $p(D \mid \theta, M)p(\theta \mid M)$ anywhere in $\{\theta\}$, but that doesn't give us much insight

$p(\theta \mid D, M)$ is a **probability density function**

$p(\theta\,|\,D, M)$ is a **probability density function**

One way to get insight into a PDF is by computing **moments**

$$\text{e.g., } \langle \theta \rangle = \int \theta \; p(\theta\,|\,D, M)\; d\theta$$

$p(\theta \,|\, D, M)$ is a **probability density function**

One way to get insight into a PDF is by computing **moments**

$$\text{e.g., } \langle \theta \rangle = \int \underline{\theta \; p(\theta \,|\, D, M) \; d\theta}$$

Even if we wanted to look at just one parameter

$$p(\theta^i \,|\, D) = \int p(\theta \,|\, D, M) \; d\theta^j d\theta^k d\theta^l \ldots$$

But if somebody gave you a set of θ values **drawn from** $p(\theta^i \mid D, M)$

$$\{\theta^i\} \rightarrow \{\theta^i_0, \theta^i_1, \theta^i_2, \ldots, \theta^i_N\}$$

But if somebody gave you a set of θ values **drawn from** $p(\theta^i \,|\, D, M)$

$$\{\theta^i\} \rightarrow \{\theta^i_0, \theta^i_1, \theta^i_2, \ldots, \theta^i_N\}$$

we can compute moments etc. the old fashioned way

$$\langle \theta^i \rangle \simeq \frac{\sum_n \theta^i_n}{N}$$

We want a procedure for generating **independent identically distributed** (iid) samples from $p(\theta \mid D, M)$

We want a procedure for generating **independent identically distributed** (iid) samples from $p(\theta | D, M)$

i.e., a random number generator for $p(\theta | D, M)$

`numpy.random.normal()` returns i.i.d. samples from N[0,1]
`numpy.random.uniform()` returns i.i.d. samples from U[0,1]

We want a procedure for generating **independent identically distributed** (iid) samples from $p(\theta | D, M)$

i.e., a random number generator for $p(\theta | D, M)$

`numpy.random.normal()` returns i.i.d. samples from N[0,1]
`numpy.random.uniform()` returns i.i.d. samples from U[0,1]

Sets from RNGs are a subset of a more general class called **Markovian processes**

$$p(\theta_N | \theta_0, \theta_1, \ldots, \theta_{N-1}) = p(\theta_N | \theta_{N-1})$$

Markov Chain Monte Carlo (MCMC):

Markov Chain Monte Carlo (MCMC):

A class of algorithms for generating (approximately) i.i.d. samples (i.e., a *Markov chain*) from some **target distribution**

The *Markov chain* must satisfy the **detailed balance** condition

$$p(y \mid x)p(x) = p(x \mid y)p(y)$$

The *Markov chain* must satisfy the **detailed balance** condition

$$p(y \mid x)p(x) = p(x \mid y)p(y)$$

This is just Bayes' theorem again!

The *Markov chain* must satisfy the **detailed balance** condition

$$\boxed{p(y \mid x)} p(x) = p(x \mid y) p(y)$$

This is just Bayes' theorem again!

Probability of y following x in the Markov Chain. We don't know this.

The *Markov chain* must satisfy the <u>**detailed balance**</u> condition

$$\boxed{p(y \mid x)}\, p(x) = p(x \mid y) p(y)$$

This is just Bayes' theorem again!

Probability of y following x in the Markov Chain. We don't know this.

Replace $p(y \mid x)$ with something we can calculate: $q(y \mid x)$

The *Markov chain* must satisfy the **detailed balance** condition

$$\boxed{p(y|x)}p(x) = p(x|y)p(y)$$

This is just Bayes' theorem again!

Probability of y following x in the Markov Chain. We don't know this.

Replace $p(y|x)$ with something we can calculate: $q(y|x)$

Proposal Distribution

e.g. $N[\mu, \sigma^2]$, $U[x_{\min}, x_{\max}]$

Our detailed balance equation becomes in inequality

$$q(y|x)p(x) > q(x|y)p(y)$$

Our detailed balance equation becomes in inequality

$$q(y \mid x)p(x) \boxed{>} q(x \mid y)p(y)$$

The chain transitions from x to y too often, and y to x too rarely

Our detailed balance equation becomes in inequality

$$q(y \mid x)p(x) \boxed{>} q(x \mid y)p(y)$$

The chain transitions from x to y too often, and y to x too rarely

To recover detailed balance, reweight by **transition probability** $\alpha \in [0,1]$

$$q(y \mid x)p(x)\alpha(x, y) = q(x \mid y)p(y)\alpha(y, x)$$

Our detailed balance equation becomes in inequality

$$q(y \mid x)p(x) \boxed{>} q(x \mid y)p(y)$$

The chain transitions from x to y too often, and y to x too rarely

To recover detailed balance, reweight by **transition probability** $\alpha \in [0,1]$

$$q(y \mid x)p(x)\boxed{\alpha(x,y)} = q(x \mid y)p(y)\boxed{\alpha(y,x)}$$

probability of transitioning from x to y

probability of transitioning from y to x

Our detailed balance equation becomes in inequality

$$q(y\,|\,x)p(x) > q(x\,|\,y)p(y)$$

The chain transitions from x to y too often, and y to x too rarely

To recover detailed balance, reweight by **transition probability** $\alpha \in [0,1]$

1

$$q(y\,|\,x)p(x)\alpha(x,y) = q(x\,|\,y)p(y)\alpha(y,x)$$

Our detailed balance equation becomes in inequality

$$q(y \mid x)p(x) > q(x \mid y)p(y)$$

The chain transitions from x to y too often, and y to x too rarely

To recover detailed balance, reweight by **transition probability** $\alpha \in [0,1]$

$$\alpha(x, y) = \frac{p(y)q(x \mid y)}{p(x)q(y \mid x)}$$

Our detailed balance equation becomes in inequality

$$q(y \mid x)p(x) \boxed{>} q(x \mid y)p(y)$$

The chain transitions from x to y too often, and y to x too rarely

To recover detailed balance, reweight by **transition probability** $\alpha \in [0,1]$

$$\alpha(x, y) = \frac{p(y)q(x \mid y)}{p(x)q(y \mid x)} \equiv H(x, y)$$

Our detailed balance equation becomes in inequality

$$q(y \mid x)p(x) \boxed{>} q(x \mid y)p(y)$$

The chain transitions from x to y too often, and y to x too rarely

To recover detailed balance, reweight by **transition probability** $\alpha \in [0,1]$

$$\alpha(x, y) = \frac{p(y)q(x \mid y)}{p(x)q(y \mid x)} \equiv \boxed{H(x, y)}$$

Hasting's Ratio

Now reverse the inequality

$$H(x, y) > 1 \quad \text{but} \quad \alpha(x, y) \in [0,1]$$

Now reverse the inequality

$$H(x, y) > 1 \quad \text{but} \quad \alpha(x, y) \in [0,1]$$

To generalize:

$$\boxed{\alpha(x, y) = \min\left[1,\ H(x, y)\right]}$$

Markov Chain Monte Carlo (MCMC):

- Start at some state x

Markov Chain Monte Carlo (MCMC):

- Start at some state x
- Propose new state y with $q(y \mid x)$

Markov Chain Monte Carlo (MCMC):

- Start at some state x
- Propose new state y with $q(y \mid x)$
- Accept with probability $\alpha(x, y) = \min\left[1,\ H(x, y)\right]$

Markov Chain Monte Carlo (MCMC):

- Start at some state x
- Propose new state y with $q(y \mid x)$
- Accept with probability $\alpha(x, y) = \min\left[1,\ H(x, y)\right]$
    - if ACCEPT state $\rightarrow y$
    - if REJECT state $\rightarrow x$

Markov Chain Monte Carlo (MCMC):

Repeat

- Start at some state x
- Propose new state y with $q(y|x)$
- Accept with probability $\alpha(x,y) = \min\left[1,\ H(x,y)\right]$
  - if ACCEPT state $\rightarrow y$
  - if REJECT state $\rightarrow x$

Markov Chain Monte Carlo (MCMC):

Now map to our inference problem:

State: Model parameters θ

Target Distribution: Posterior $p(\theta \mid D, M)$

Markov Chain Monte Carlo (MCMC):

Now map to our inference problem:

State: Model parameters θ

Target Distribution: Posterior $p(\theta \,|\, D, M)$

$$\boxed{\alpha(\theta, \theta') = \frac{p(D \,|\, \theta', M)\; p(\theta' \,|\, M)\; q(\theta \,|\, \theta')}{p(D \,|\, \theta, M)\; p(\theta \,|\, M)\; q(\theta' \,|\, \theta)}}$$

To gain insight, always consider limiting cases:

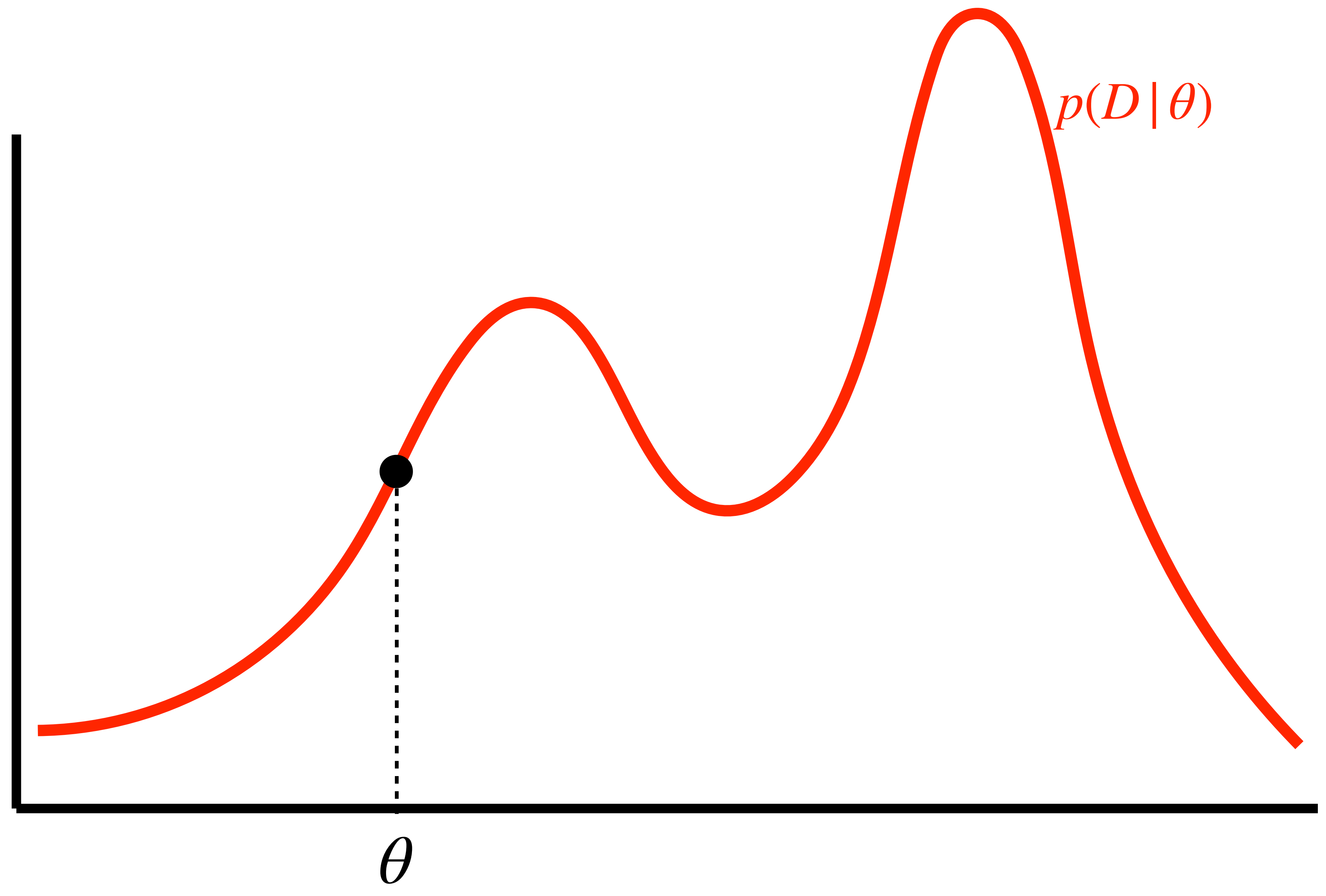
$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#1. Symmetric proposal, uniform priors

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)}{p(D\,|\,\theta, M)}$$



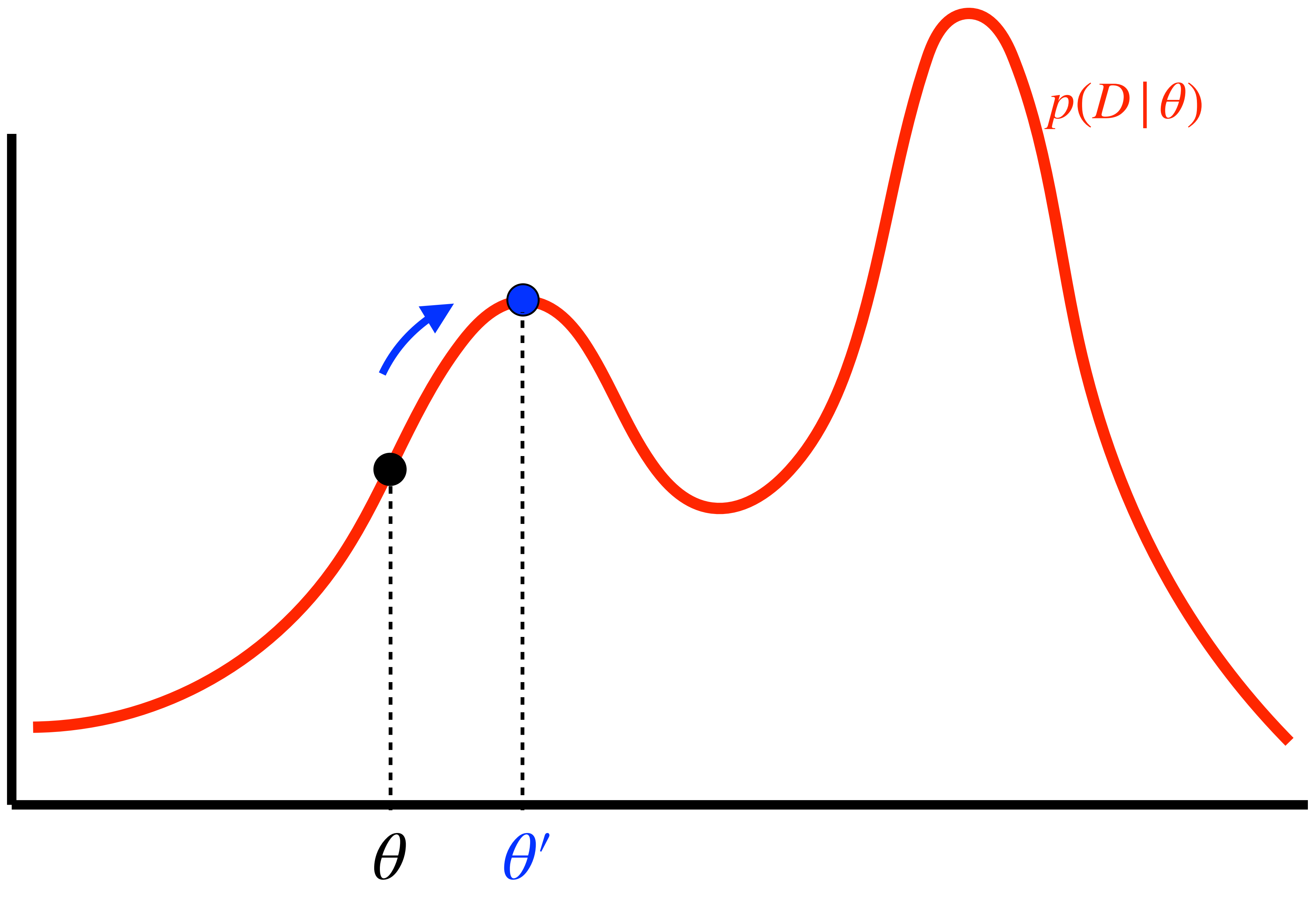
To gain insight, always consider limiting cases:

$$\alpha(\theta,\theta') = \frac{p(D\,|\,\theta',M)\;p(\theta'\,|\,M)\;q(\theta\,|\,\theta')}{p(D\,|\,\theta,M)\;p(\theta\,|\,M)\;q(\theta'\,|\,\theta)}$$

#1. Symmetric proposal, uniform priors

$$\alpha(\theta,\theta') = \frac{p(D\,|\,\theta',M)}{p(D\,|\,\theta,M)}$$

$p(D\,|\,\theta') > p(D\,|\,\theta)$ ACCEPT $\rightarrow$ Always go “uphill”



To gain insight, always consider limiting cases:

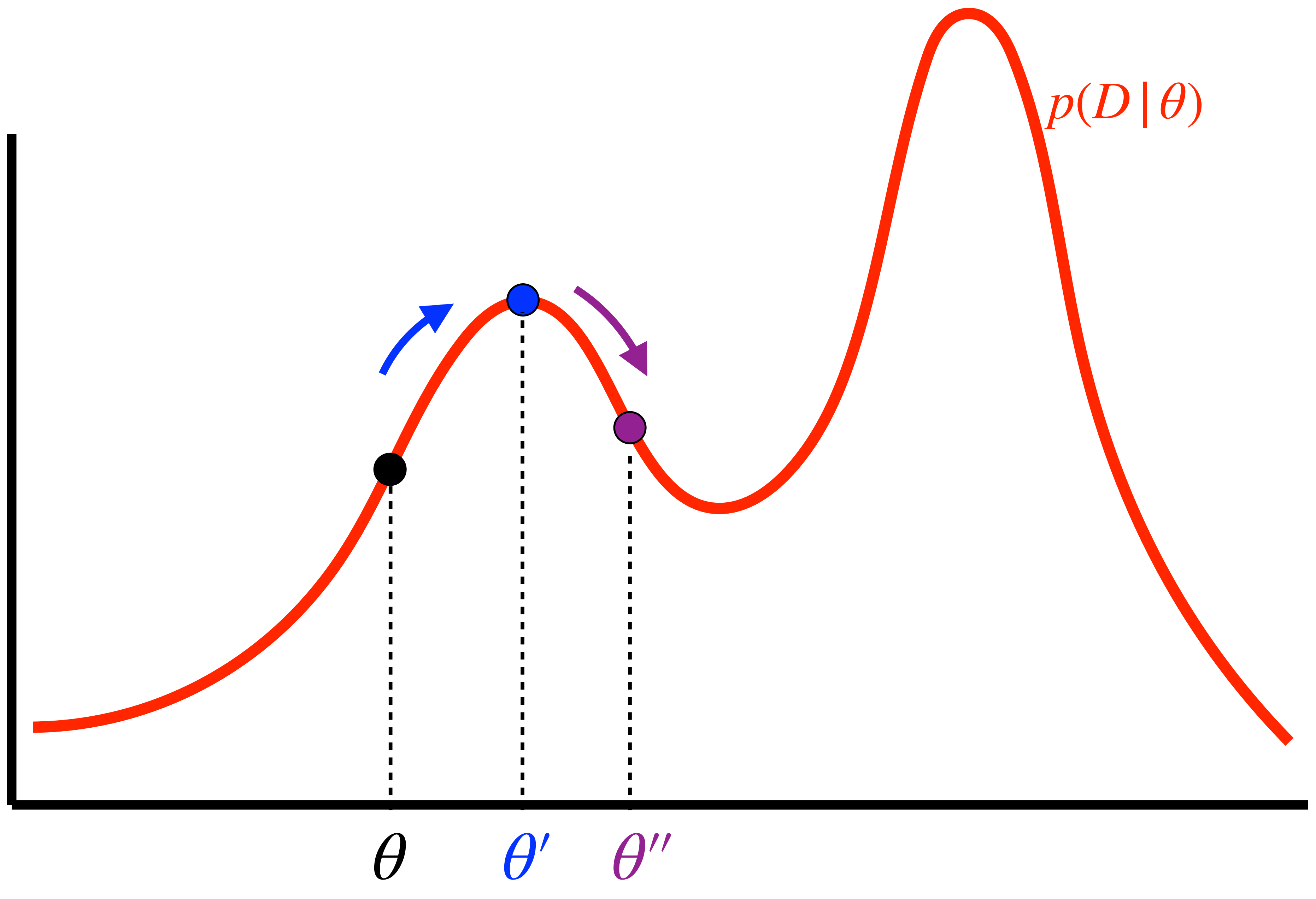
$$\alpha(\theta,\theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#1. Symmetric proposal, uniform priors

$$\alpha(\theta,\theta') = \frac{p(D\,|\,\theta', M)}{p(D\,|\,\theta, M)}$$

$p(D\,|\,\theta') > p(D\,|\,\theta)$ ACCEPT $\rightarrow$ Always go "uphill"

$p(D\,|\,\theta'') < p(D\,|\,\theta')$ ROLL THE DICE $\rightarrow$ Go "downhill" w/ probability $\alpha(\theta',\theta'')$



To gain insight, always consider limiting cases:

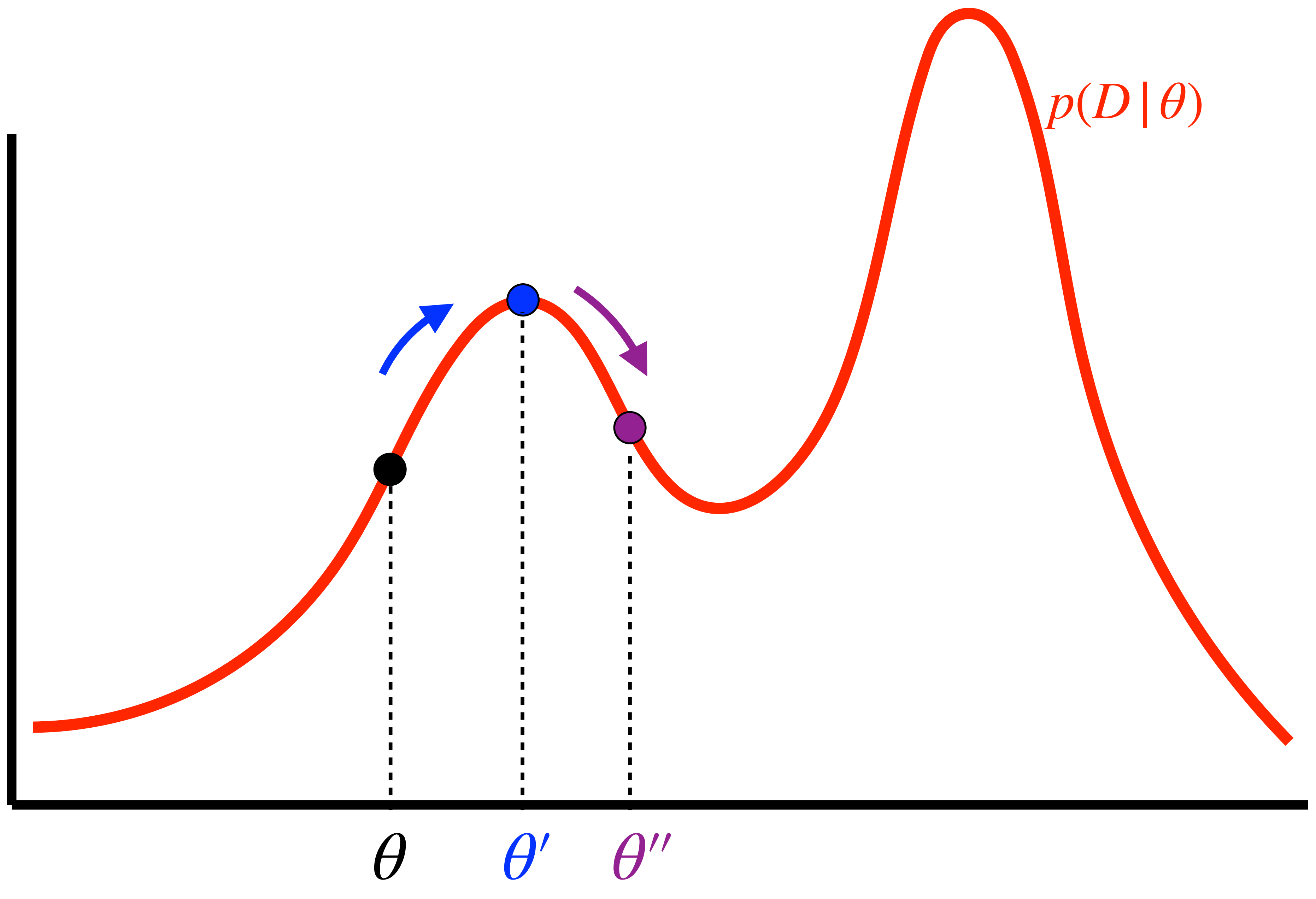
$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#1. Symmetric proposal, uniform priors

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)}{p(D\,|\,\theta, M)}$$

$p(D\,|\,\theta') > p(D\,|\,\theta)$ ACCEPT $\rightarrow$ Always go “uphill”

$p(D\,|\,\theta'') < p(D\,|\,\theta')$ ROLL THE DICE $\rightarrow$ Go “downhill” w/ probability $\alpha(\theta', \theta'')$



This guarantees that MCMCs *eventually* converge

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#2. Proposal distribution matches target distribution

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D \,|\, \theta', M)\ p(\theta' \,|\, M)\ q(\theta \,|\, \theta')}{p(D \,|\, \theta, M)\ p(\theta \,|\, M)\ q(\theta' \,|\, \theta)}$$

#2. Proposal distribution matches target distribution

$$\alpha(\theta, \theta') = \frac{p(D \,|\, \theta', M) q(\theta \,|\, \theta')}{p(D \,|\, \theta, M) q(\theta' \,|\, \theta)},\ q(\theta \,|\, \theta') = p(\theta \,|\, D)$$

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#2. Proposal distribution matches target distribution

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M) q(\theta\,|\,\theta')}{p(D\,|\,\theta, M) q(\theta'\,|\,\theta)}, \; q(\theta\,|\,\theta') = p(\theta\,|\,D)$$

$$\alpha(\theta, \theta') = 1$$

Everything gets accepted!
Maximally efficient

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D \,|\, \theta', M)\; p(\theta' \,|\, M)\; q(\theta \,|\, \theta')}{p(D \,|\, \theta, M)\; p(\theta \,|\, M)\; q(\theta' \,|\, \theta)}$$

#2. Proposal distribution matches target distribution

$$\alpha(\theta, \theta') = \frac{p(D \,|\, \theta', M) q(\theta \,|\, \theta')}{p(D \,|\, \theta, M) q(\theta' \,|\, \theta)}, \quad q(\theta \,|\, \theta') = p(\theta \,|\, D)$$

$$\alpha(\theta, \theta') = 1$$

Everything gets accepted!
Maximally efficient

Designing good proposal distributions is where the work hides

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\ p(\theta'\,|\,M)\ q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\ p(\theta\,|\,M)\ q(\theta'\,|\,\theta)}$$

#3. Hammer at the true value θ^{TRUTH}

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#3. Hammer at the true value θ^{TRUTH}

$$q(\theta'\,|\,\theta) = \delta(\theta' - \theta^{\mathrm{TRUTH}})$$

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#3. Hammer at the true value θ^{TRUTH}

$$q(\theta'\,|\,\theta) = \delta(\theta' - \theta^{\mathrm{TRUTH}})$$

$$\alpha(\theta, \theta^{\mathrm{TRUTH}}) = \frac{p(\theta^{\mathrm{TRUTH}}\,|\,D)}{p(\theta\,|\,D)} \frac{\delta(\theta - \theta^{\mathrm{TRUTH}})}{\delta(0)} = 0?!?!$$

To gain insight, always consider limiting cases:

$$\alpha(\theta, \theta') = \frac{p(D\,|\,\theta', M)\; p(\theta'\,|\,M)\; q(\theta\,|\,\theta')}{p(D\,|\,\theta, M)\; p(\theta\,|\,M)\; q(\theta'\,|\,\theta)}$$

#3. Hammer at the true value θ^{TRUTH}

$$q(\theta'\,|\,\theta) = \delta(\theta' - \theta^{\mathrm{TRUTH}})$$

$$\alpha(\theta, \theta^{\mathrm{TRUTH}}) = \frac{p(\theta^{\mathrm{TRUTH}}\,|\,D)}{p(\theta\,|\,D)} \frac{\delta(\theta - \theta^{\mathrm{TRUTH}})}{\delta(0)} = 0?!?!$$

The proposal is a teenager: The harder you push the harder it pushes back.

$$\alpha(\theta, \theta') = \frac{\boxed{p(D \mid \theta', M)\; p(\theta' \mid M)}\,\boxed{q(\theta \mid \theta')}}{\boxed{p(D \mid \theta, M)\; p(\theta \mid M)}\,\boxed{q(\theta' \mid \theta)}}$$

Always go uphill, but try going down as well

Counterbalance how hard you're pushing

INTERMISSION

What about $p(D \mid M)$?

$$p(D \mid M) = \int d\theta \; p(D \mid \theta, M) p(\theta \mid M)$$

$$p(D \mid M) = \int d\theta \; p(D \mid \theta, M) p(\theta \mid M)$$

Large-dimensional
integral == HARD

Recast as 1D integral

$$Z = \int L \, dX$$

Evidence

Likelihood

$dX = p(\theta | M) d\theta$

Prior volume element

Recast as 1D integral

$$Z = \int L \, dX$$

Evidence

Likelihood

$dX = p(\theta | M) d\theta$

Prior volume element

$$X \equiv X(\lambda) = \int_{L(\theta)>\lambda} p(\theta | M) d\theta$$

Recast as 1D integral

$$Z = \int L \, dX$$

Evidence

Likelihood

$dX = p(\theta \,|\, M) d\theta$

Prior volume element

$$X \equiv X(\lambda) = \int_{L(\theta) > \lambda} p(\theta \,|\, M) d\theta$$

Cumulant prior mass covering all likelihood values $> \lambda$

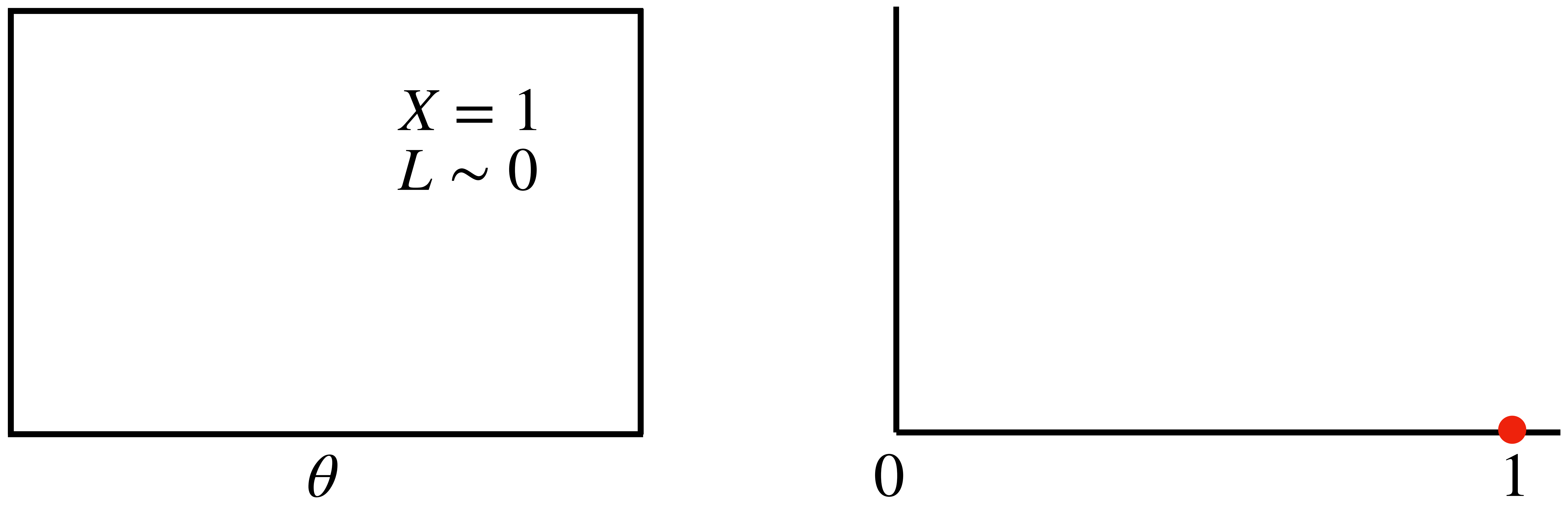
As λ increases, X goes from $1 \rightarrow 0$

λ is small: likelihood is low -> prior

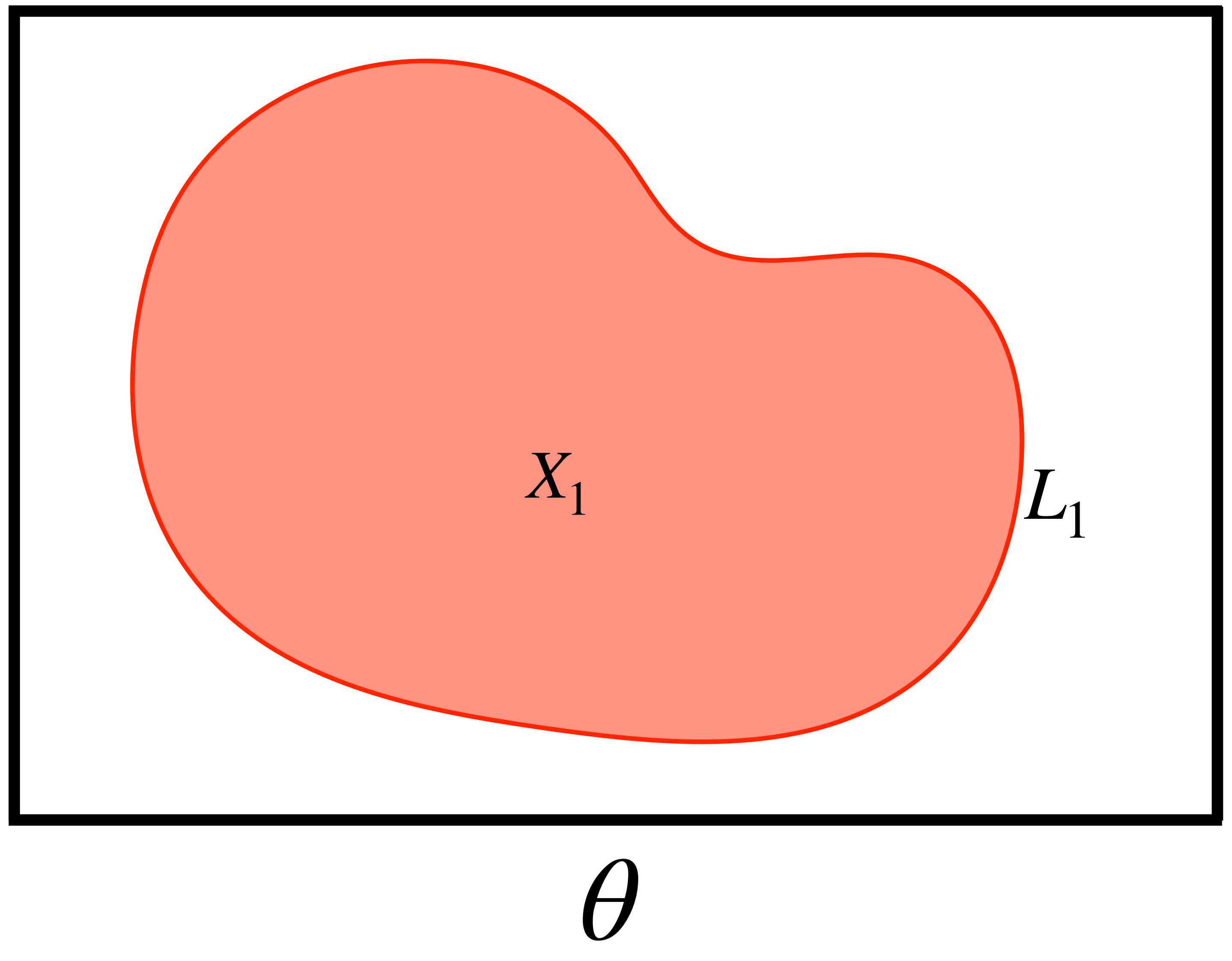
so, $$Z = \int_0^1 L(X)\, dX$$

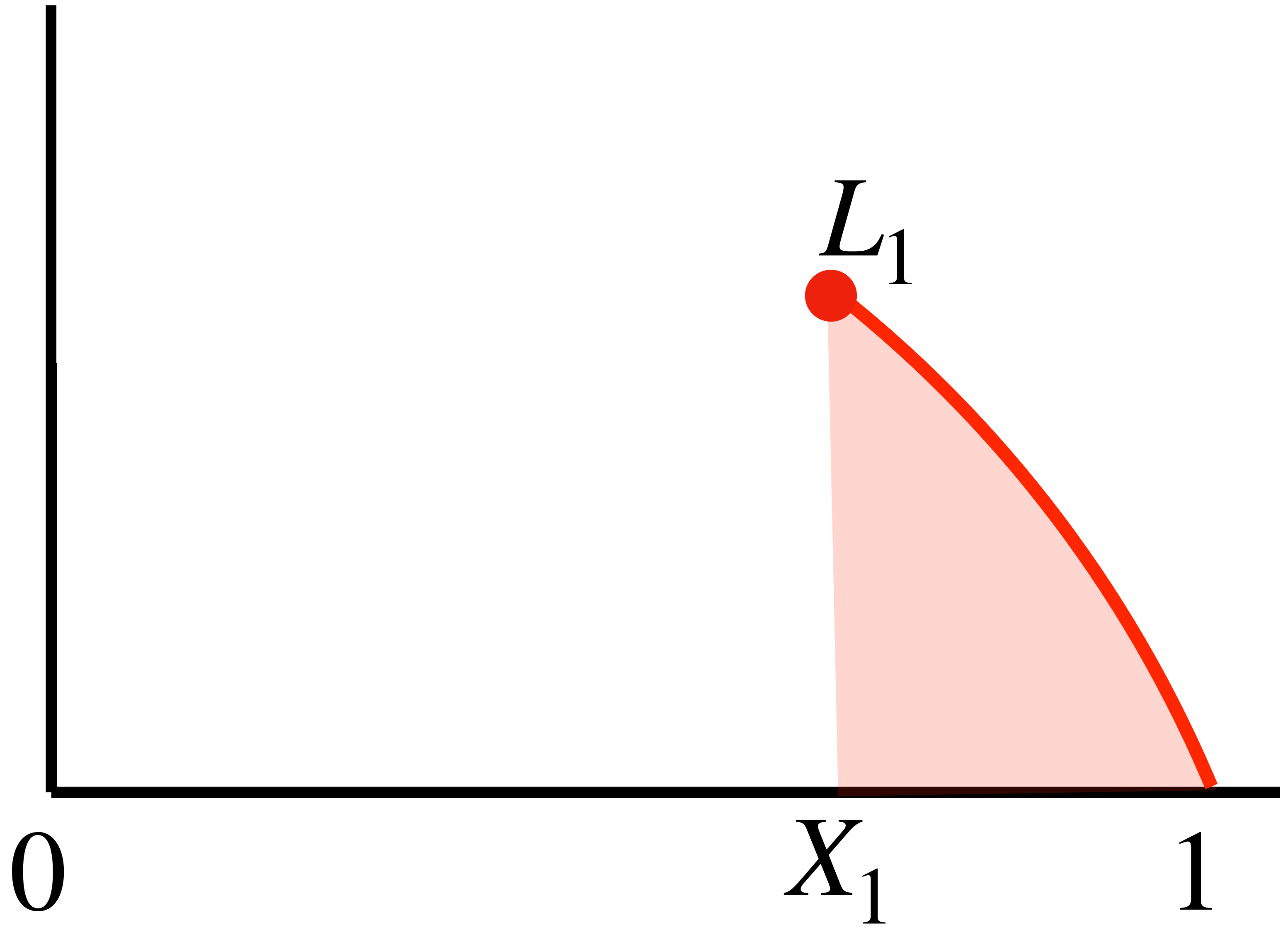
λ is large: likelihood is high

$L(X)$ is positive and monotonically decreasing. We can do this.

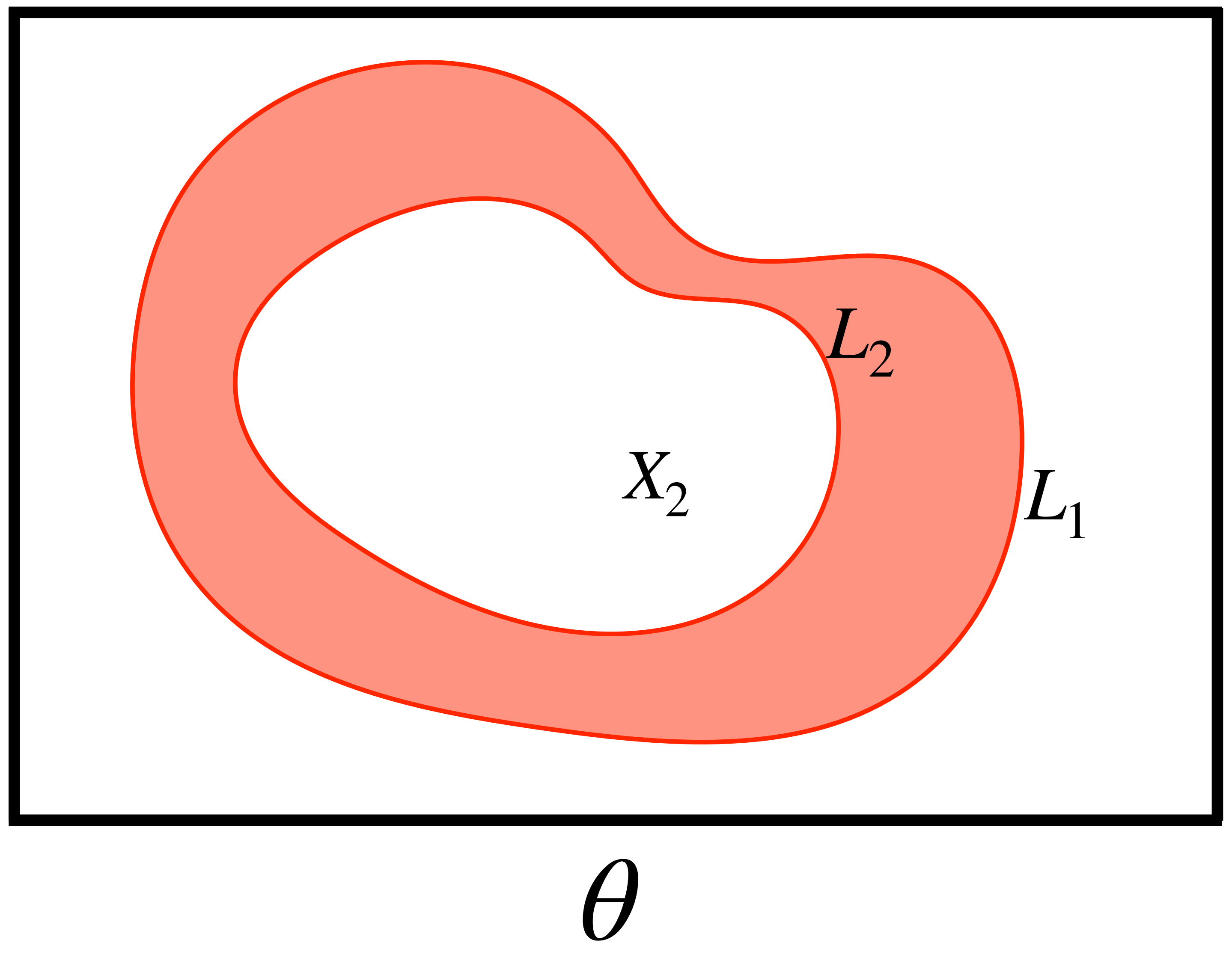


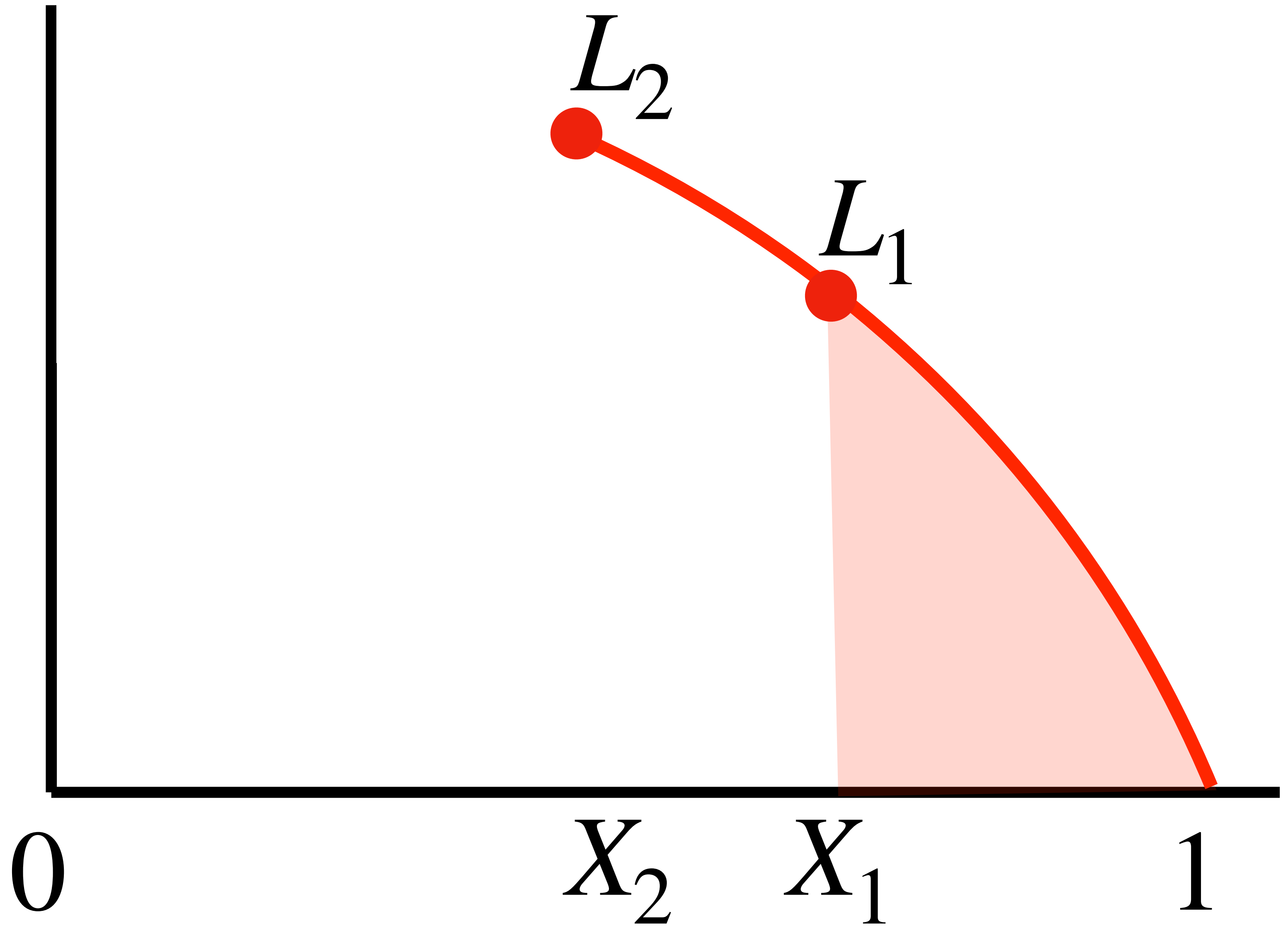
$L(X)$ is positive and monotonically decreasing. We can do this.



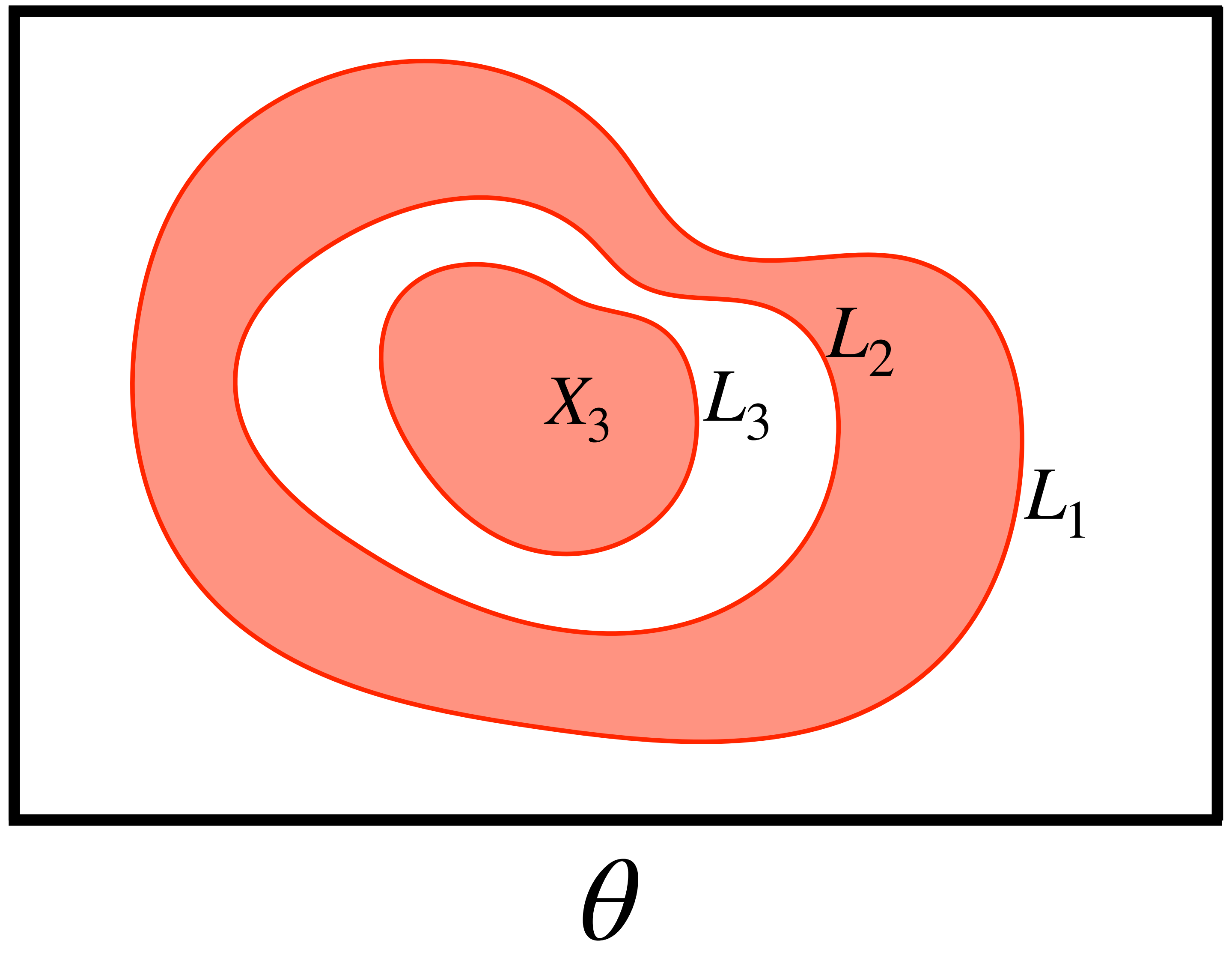


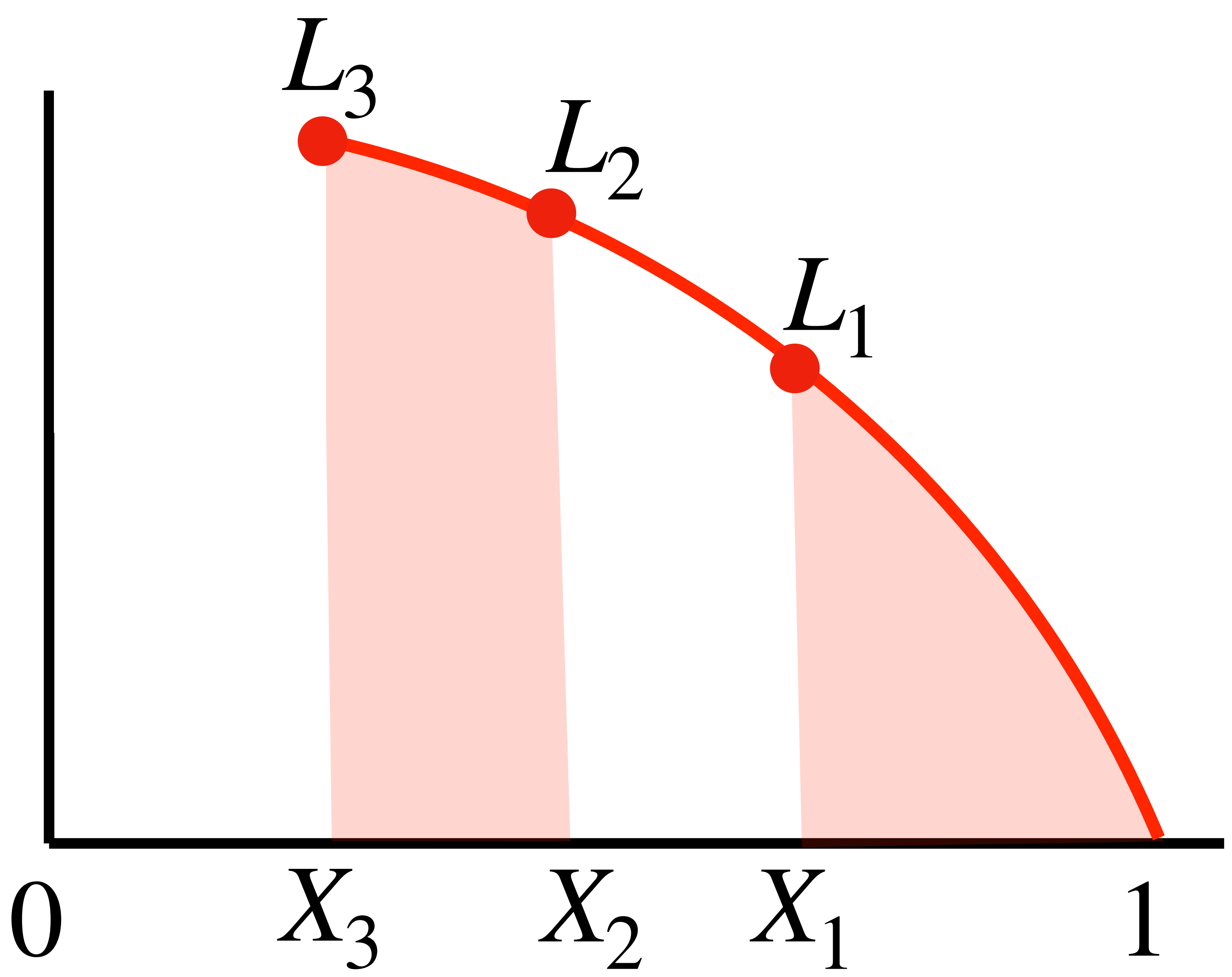
$L(X)$ is positive and monotonically decreasing. We can do this.



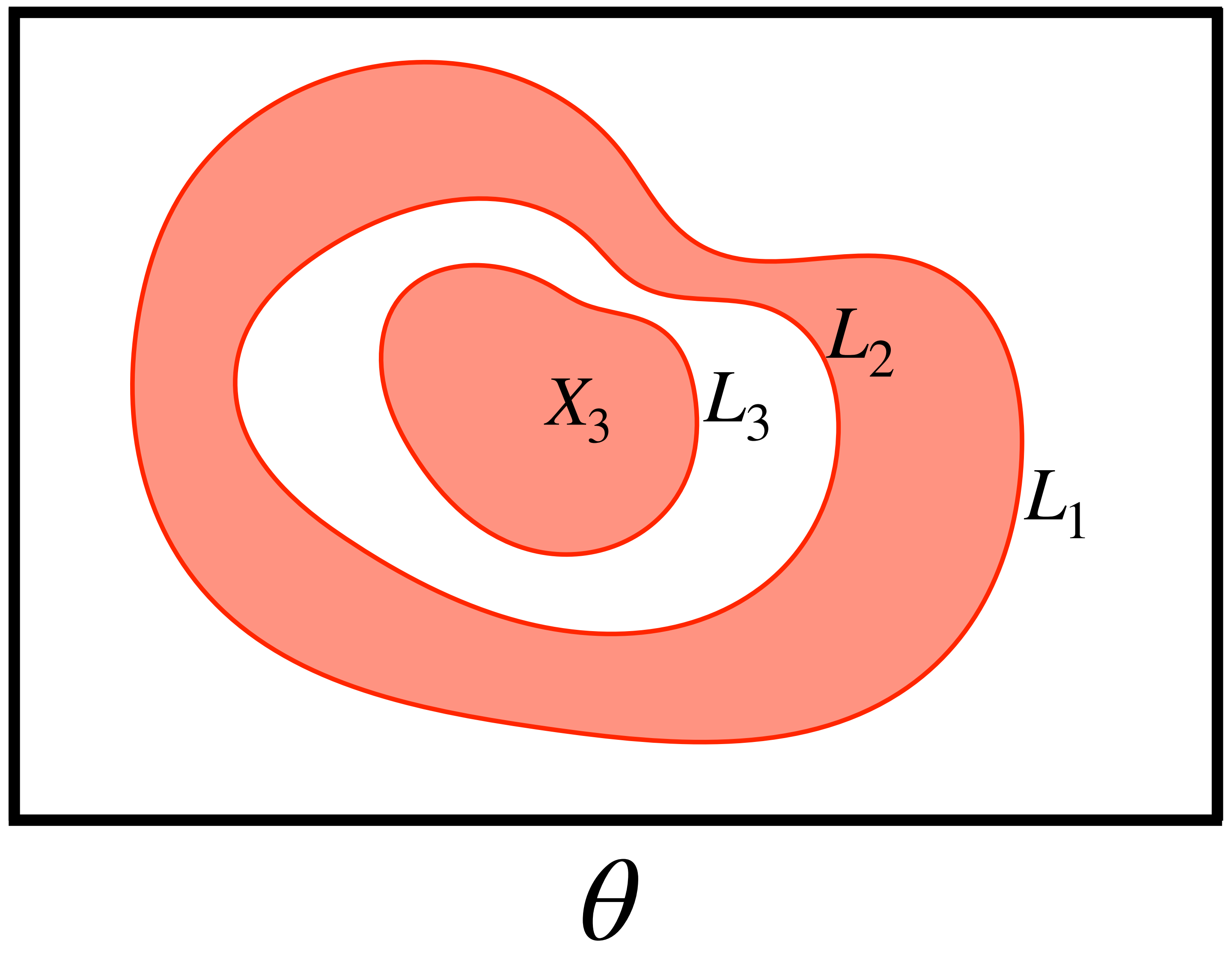


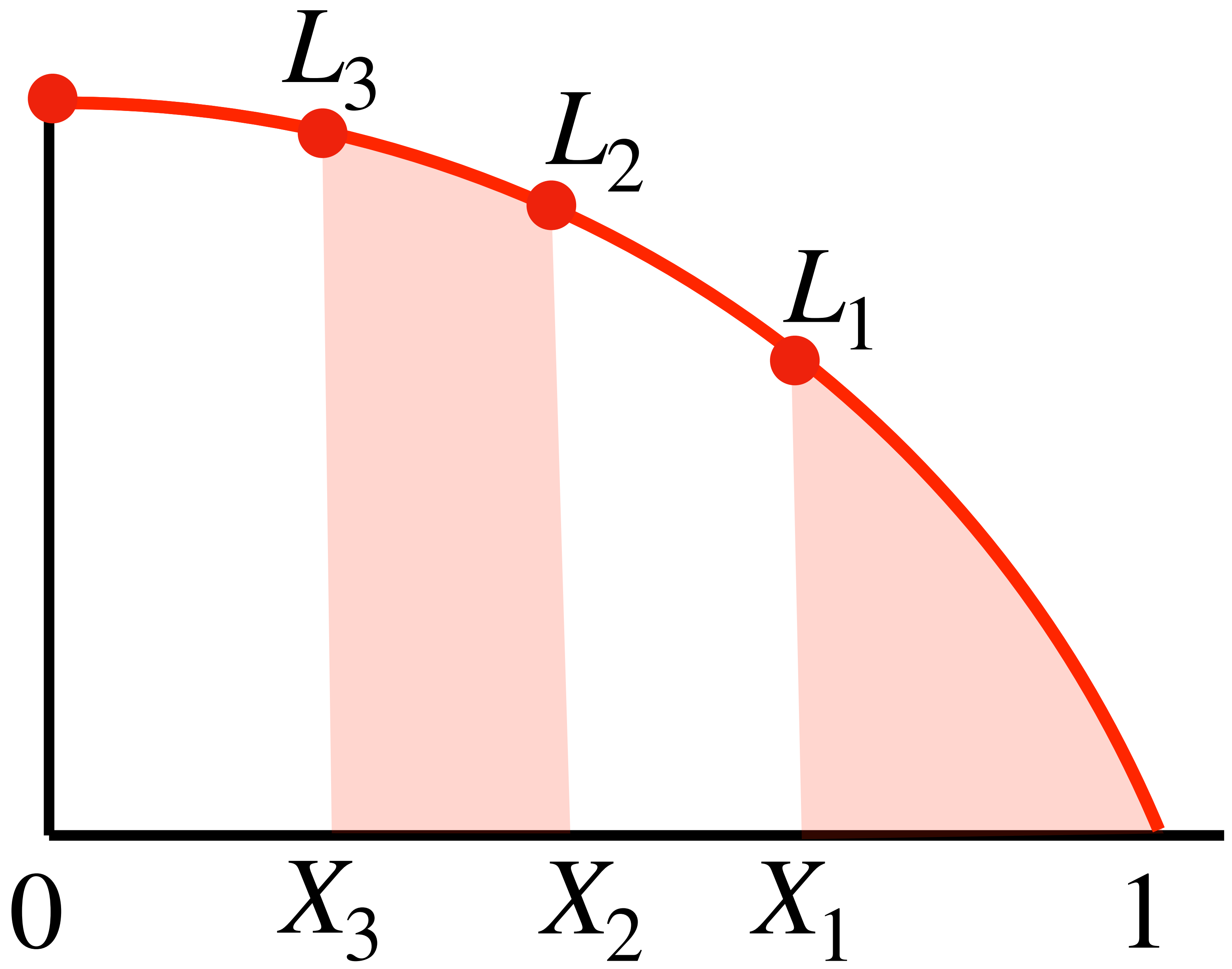
$L(X)$ is positive and monotonically decreasing. We can do this.





$L(X)$ is positive and monotonically decreasing. We can do this.





Nested Sampling:

- Start w/ N points sampling prior $\theta_1, \ldots, \theta_N$

$$Z = 0, \; X_0 = 1, \; i = 1$$

Nested Sampling:

- Start w/ N points sampling prior $\theta_1, \ldots, \theta_N$

$$Z = 0,\ X_0 = 1,\ i = 1$$

- Record the lowest $L(\theta_n)$ as L_i

Nested Sampling:

- Start w/ N points sampling prior $\theta_1, \ldots, \theta_N$

$$Z = 0,\ X_0 = 1,\ i = 1$$

- Record the lowest $L(\theta_n)$ as L_i
- $X_i \sim e^{-i/N},\ Z = Z + L_i(X_{i-1} - X_i)$

Nested Sampling:

- Start w/ N points sampling prior $\theta_1, \ldots, \theta_N$

$$Z = 0,\ X_0 = 1,\ i = 1$$

- Record the lowest $L(\theta_n)$ as L_i
- $X_i \sim e^{-i/N},\ Z = Z + L_i(X_{i-1} - X_i)$
- Draw new θ_N from prior constrained by $L(\theta_n) > L_i$

Nested Sampling:

- Start w/ N points sampling prior $\theta_1, \ldots, \theta_N$

$$Z = 0,\ X_0 = 1,\ i = 1$$

Repeat
$i = i + 1,\ i < i_{\max}$

- Record the lowest $L(\theta_n)$ as L_i
- $X_i \sim e^{-i/N},\ Z = Z + L_i(X_{i-1} - X_i)$
- Draw new θ_N from prior constrained by $L(\theta_n) > L_i$

Nested Sampling:

- Start w/ N points sampling prior $\theta_1, \ldots, \theta_N$

$$Z = 0,\ X_0 = 1,\ i = 1$$

Repeat

$i = i + 1,\ i < i_{\max}$

- Record the lowest $L(\theta_n)$ as L_i
- $X_i \sim e^{-i/N},\ Z = Z + L_i(X_{i-1} - X_i)$
- Draw new θ_N from prior constrained by $L(\theta_n) > L_i$
- $Z = Z + \sum_n L(\theta_n) X_i / N$

Nested Sampling:

Generate posterior samples by reweighting $\{\theta_n\}$

$$p_i = L_i(X_{i-1} - X_i)/Z$$

ENCORE

[Other evidence calculators/estimates]

[Other sampling algorithms]