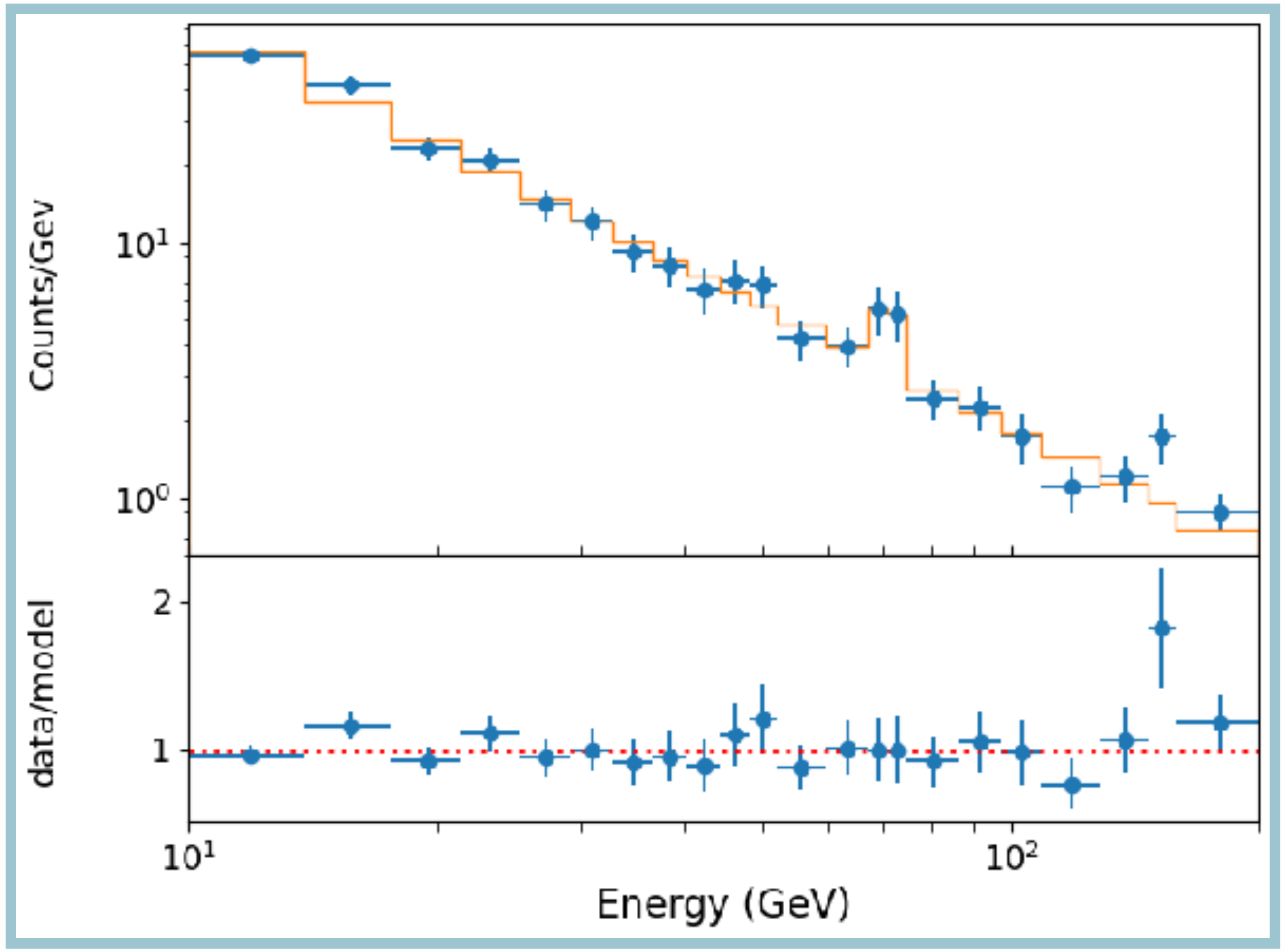
Sampling + Model Comparison, Part II

Universiteit van Amsterdam

d.huppenkothen@uva.nl

What about physics?



Data: X-ray spectrum

Model: Power law + Gaussian line

Likelihood: Poisson

Parameters: $A_{\mathrm{PL}}, \Gamma, A_{\mathrm{Line}}, E_{\mathrm{Line}}, \sigma_{\mathrm{Line}} A_{\mathrm{Line}}$

Priors: ???

What priors should we choose?

$$p(A_{\mathrm{PL}}, \Gamma, A_{\mathrm{Line}}, E_{\mathrm{Line}}, \sigma_{\mathrm{Line}} \,|\, D) = \frac{p(D \,|\, A_{\mathrm{PL}}, \Gamma, A_{\mathrm{Line}}, E_{\mathrm{Line}}, \sigma_{\mathrm{Line}}) p(A_{\mathrm{PL}}, \Gamma, A_{\mathrm{Line}}, E_{\mathrm{Line}}, \sigma_{\mathrm{Line}})}{p(D)}$$

$$p(A_{\mathrm{PL}}, \Gamma, A_{\mathrm{Line}}, E_{\mathrm{Line}}, \sigma_{\mathrm{Line}} \,|\, D) = \frac{p(D \,|\, A_{\mathrm{PL}}, \Gamma, A_{\mathrm{Line}}, E_{\mathrm{Line}}, \sigma_{\mathrm{Line}}) p(A_{\mathrm{PL}}) p(\Gamma) p(A_{\mathrm{Line}}) p(E_{\mathrm{Line}}) p(\sigma_{\mathrm{Line}})}{p(D)}$$

What assumption have we made here?

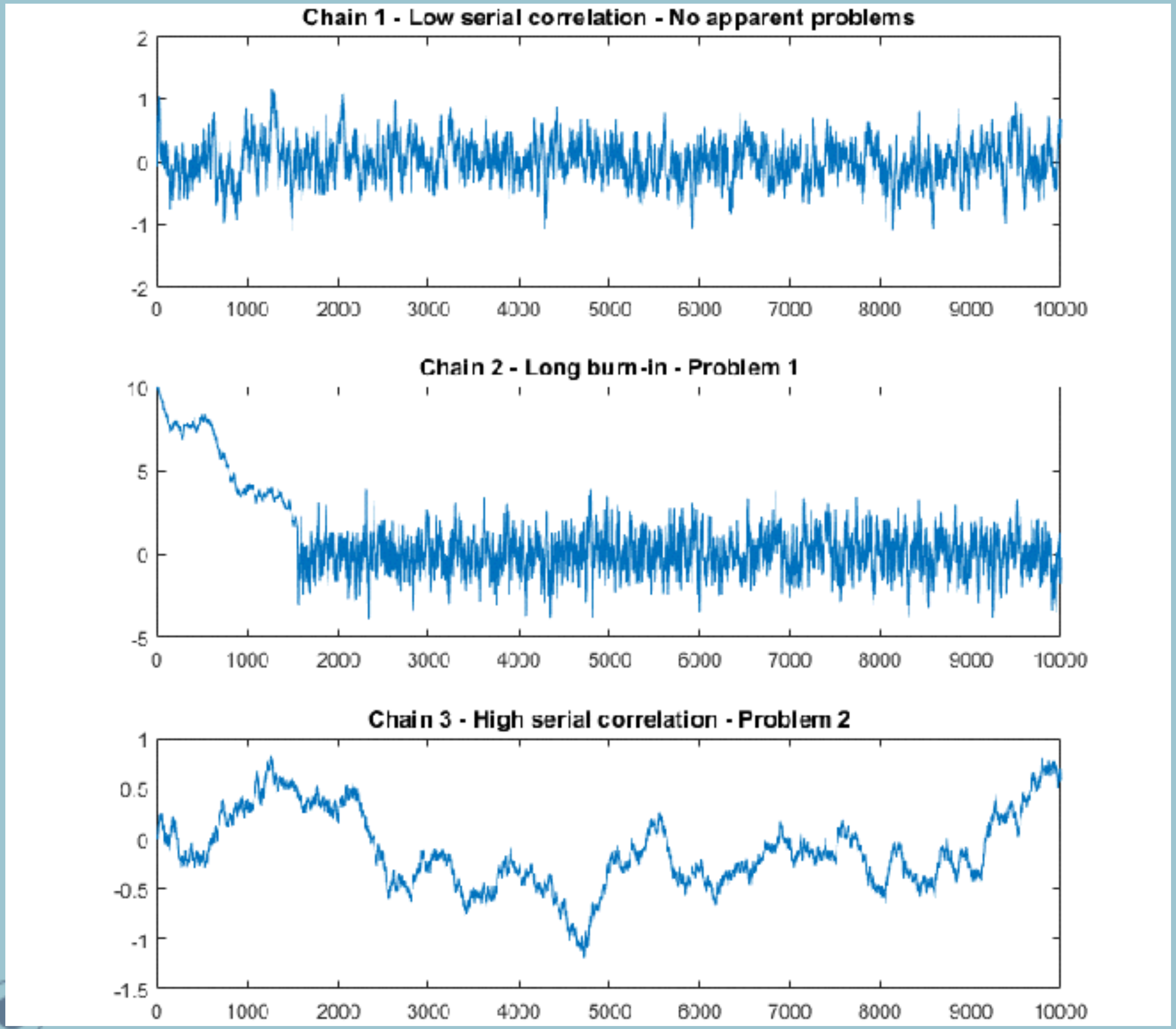
Independence of priors!

Markov Chain Monte Carlo: Assessing Convergence

Assessing Convergence:

trace plots

How do I know whether my Markov chain is sampling from the posterior?

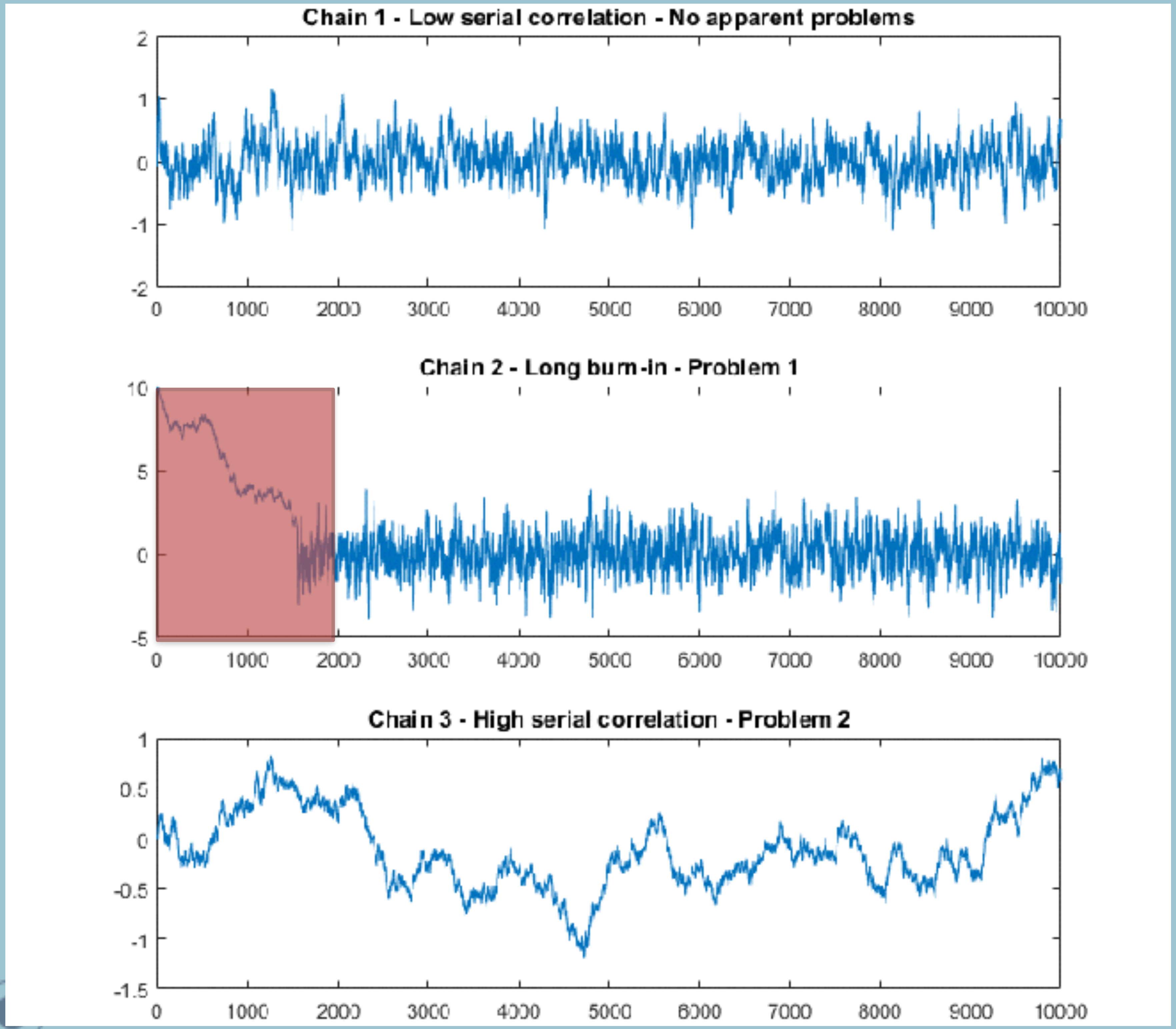


https://www.statlect.com/fundamentals-of-statistics/Markov-Chain-Monte-Carlo-diagnostics

Assessing Convergence:

trace plots

How do I know whether my Markov chain is sampling from the posterior?



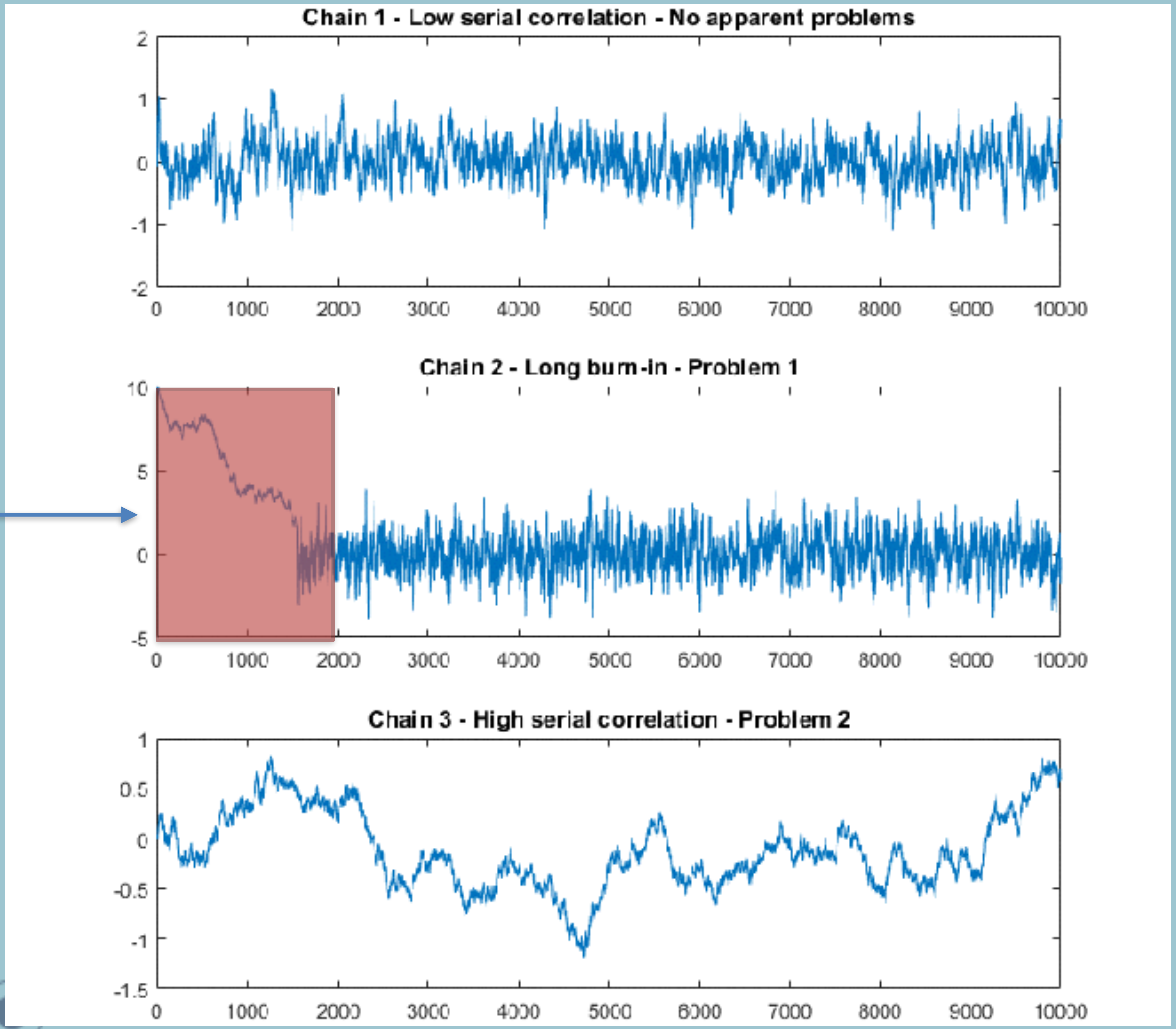
https://www.statlect.com/fundamentals-of-statistics/Markov-Chain-Monte-Carlo-diagnostics

Assessing Convergence:

trace plots

How do I know whether my Markov chain is sampling from the posterior?

Cut out this part

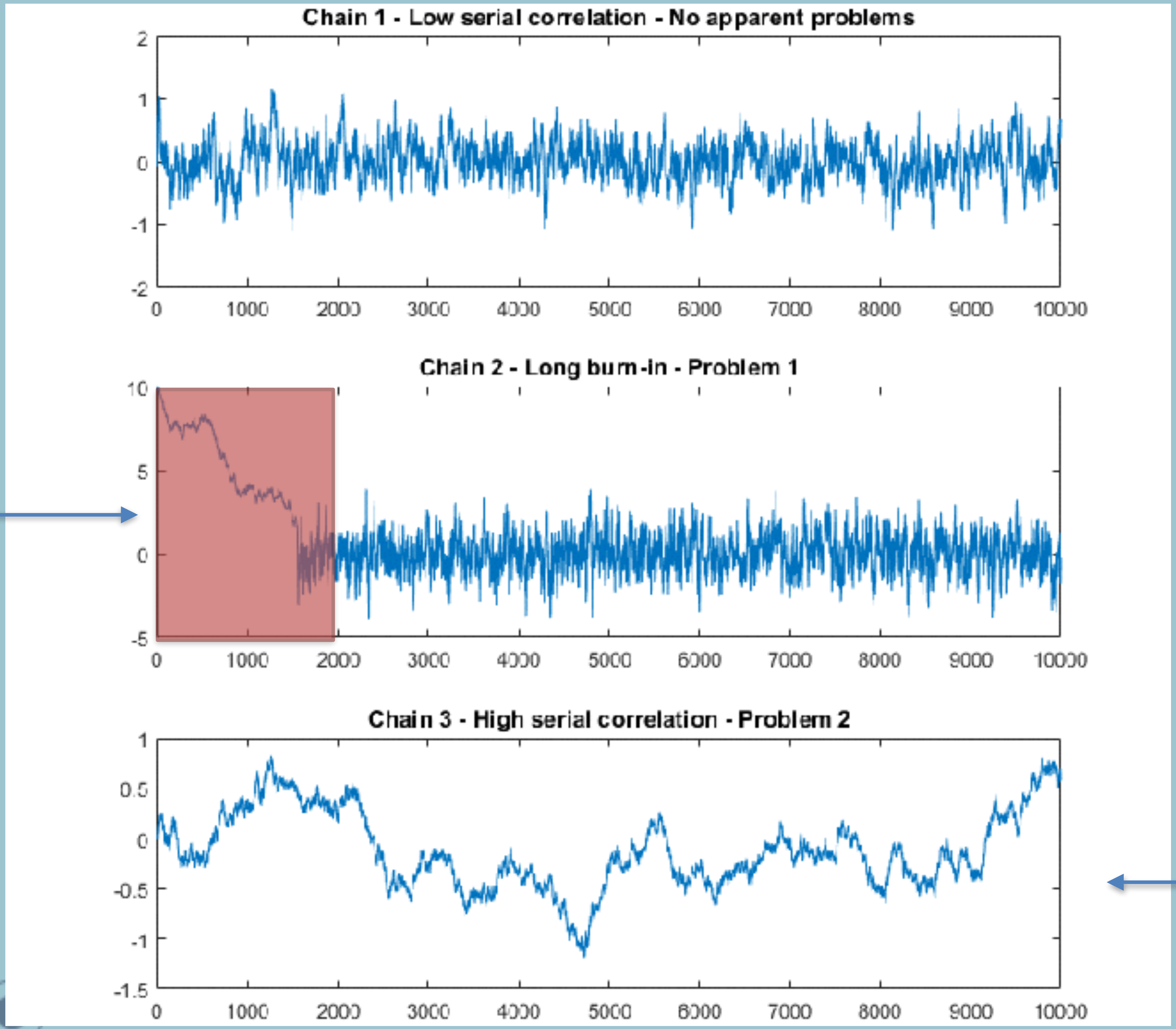


https://www.statlect.com/fundamentals-of-statistics/Markov-Chain-Monte-Carlo-diagnostics

Assessing Convergence:

trace plots

How do I know whether my Markov chain is sampling from the posterior?



Cut out this part

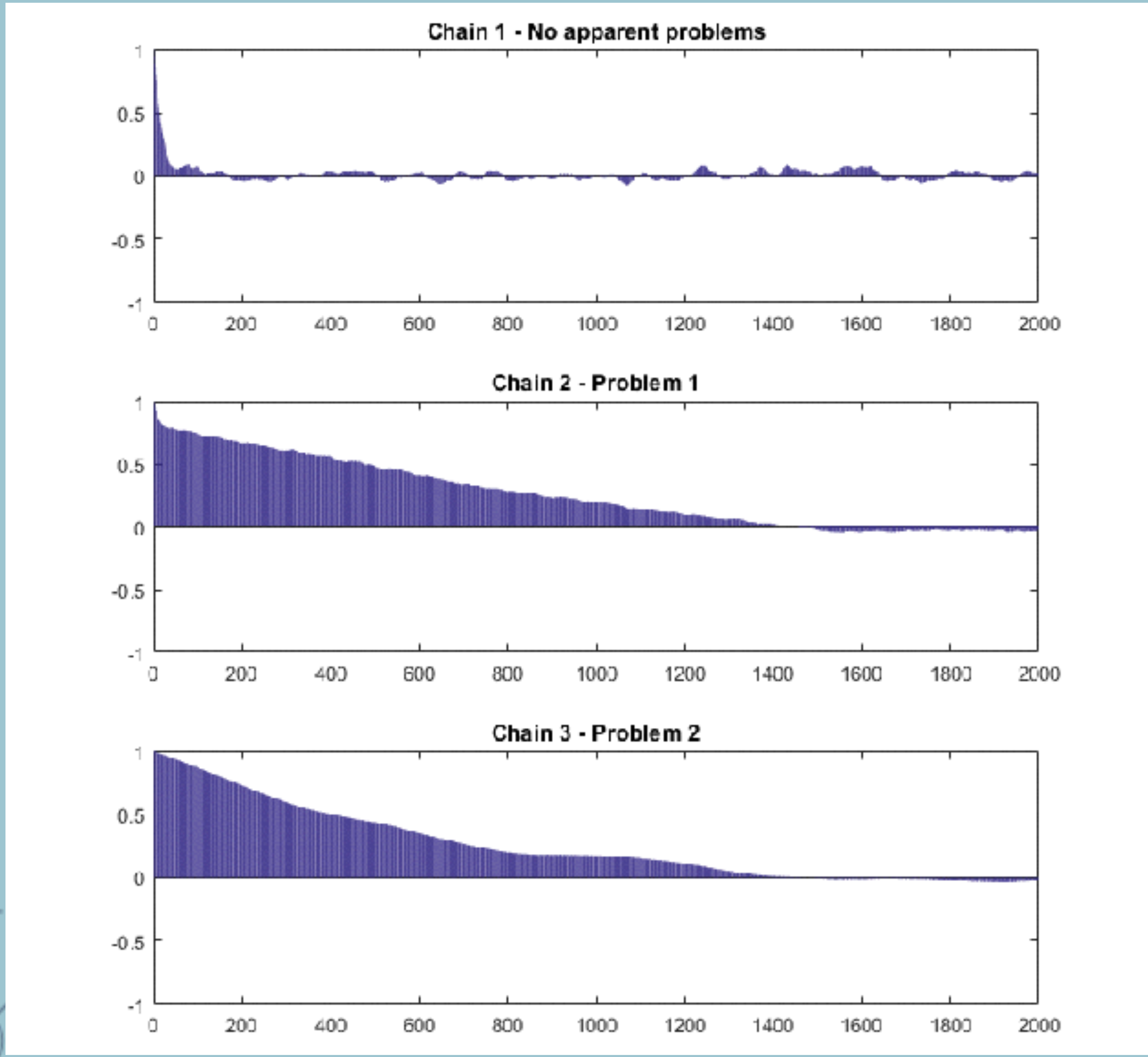
Sample for much longer, change $q(\theta \mid \theta_0)$

https://www.statlect.com/fundamentals-of-statistics/Markov-Chain-Monte-Carlo-diagnostics

Assessing Convergence:

autocorrelation

"How long do I have to sample until I get to an independent sample?"



https://www.statlect.com/fundamentals-of-statistics/Markov-Chain-Monte-Carlo-diagnostics

Assessing Convergence:

Gelman-Rubin Statistic

Compare the variance <u>within chains</u> to the variance <u>between chains</u>

$$\bar{x}_j = \frac{1}{L}\sum_{i=1}^{L} x_i^{(j)}$$ Mean value of chain j

Assessing Convergence:

Gelman-Rubin Statistic

Compare the variance <u>within chains</u> to the variance <u>between chains</u>

$$\bar{x}_j = \frac{1}{L}\sum_{i=1}^{L} x_i^{(j)}$$ Mean value of chain j

$$\bar{x}_* = \frac{1}{J}\sum_{j=1}^{J} \bar{x}_j$$ Mean of the means of all chains

Assessing Convergence:

Gelman-Rubin Statistic

Compare the variance <u>within chains</u> to the variance <u>between chains</u>

$$\overline{x}_j = \frac{1}{L}\sum_{i=1}^{L} x_i^{(j)}$$ Mean value of chain j

$$\overline{x}_* = \frac{1}{J}\sum_{j=1}^{J} \overline{x}_j$$ Mean of the means of all chains

$$B = \frac{L}{J-1}\sum_{j=1}^{J}(\overline{x}_j - \overline{x}_*)^2$$ Variance of the means of the chains

Assessing Convergence:

Gelman-Rubin Statistic

Compare the variance <u>within chains</u> to the variance <u>between chains</u>

$$\bar{x}_j = \frac{1}{L}\sum_{i=1}^{L} x_i^{(j)}$$ Mean value of chain j

$$\bar{x}_* = \frac{1}{J}\sum_{j=1}^{J} \bar{x}_j$$ Mean of the means of all chains

$$B = \frac{L}{J-1}\sum_{j=1}^{J}(\bar{x}_j - \bar{x}_*)^2$$ Variance of the means of the chains

$$W = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{L-1}\sum_{i=1}^{L}(x_i^{(j)} - \bar{x}_j)^2\right)$$ Averaged variances of the individual chains across all chains

Assessing Convergence:

Gelman-Rubin Statistic

Compare the variance within chains to the variance between chains

$$\bar{x}_j = \frac{1}{L}\sum_{i=1}^{L} x_i^{(j)}$$ Mean value of chain j

$$\bar{x}_* = \frac{1}{J}\sum_{j=1}^{J} \bar{x}_j$$ Mean of the means of all chains

$$B = \frac{L}{J-1}\sum_{j=1}^{J} (\bar{x}_j - \bar{x}_*)^2$$ Variance of the means of the chains

$$W = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{L-1}\sum_{i=1}^{L}(x_i^{(j)} - \bar{x}_j)^2\right)$$ Averaged variances of the individual chains across all chains

An estimate of the Gelman-Rubin statistic R then results as[1]

$$R = \frac{\frac{L-1}{L}W + \frac{1}{L}B}{W}.$$

When L tends to infinity and B tends to zero, R tends to 1.

Summarising Posterior Probability Distributions

You have sampled a posterior PDF, which has produced samples (vectors of parameters). What do you do now?

Summarising Posterior Probability Distributions:

Marginalised Distributions + Credible Intervals

Reminder: Marginalisation

Summarising Posterior Probability Distributions:

Marginalised Distributions + Credible Intervals

Reminder: Marginalisation

$$p(x) = \int_{-\infty}^{\infty} p(x, y)dy = \int_{-\infty}^{\infty} p(x \mid y)p(y)dy$$

Summarising Posterior Probability Distributions:

Marginalised Distributions + Credible Intervals

Reminder: Marginalisation

$$p(x) = \int_{-\infty}^{\infty} p(x, y)dy = \int_{-\infty}^{\infty} p(x \,|\, y)p(y)dy$$

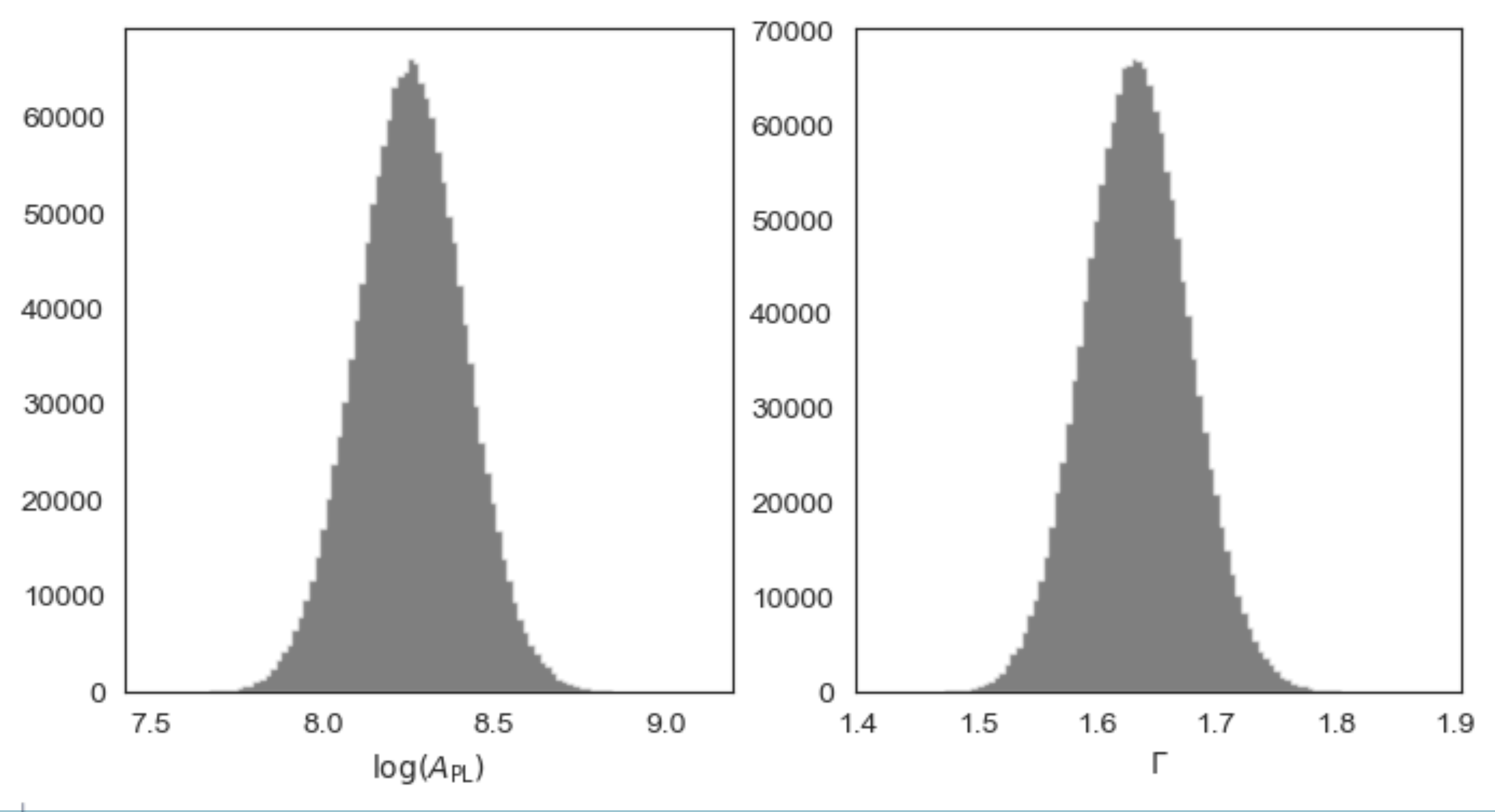
$$p(\theta_1 \,|\, D) = \int_{-\infty}^{\infty} p(\theta_1, \theta_2 \,|\, D)p(\theta_1)p(\theta_2)d\theta_2$$

Summarising Posterior Probability Distributions:

Marginalised Distributions + Credible Intervals

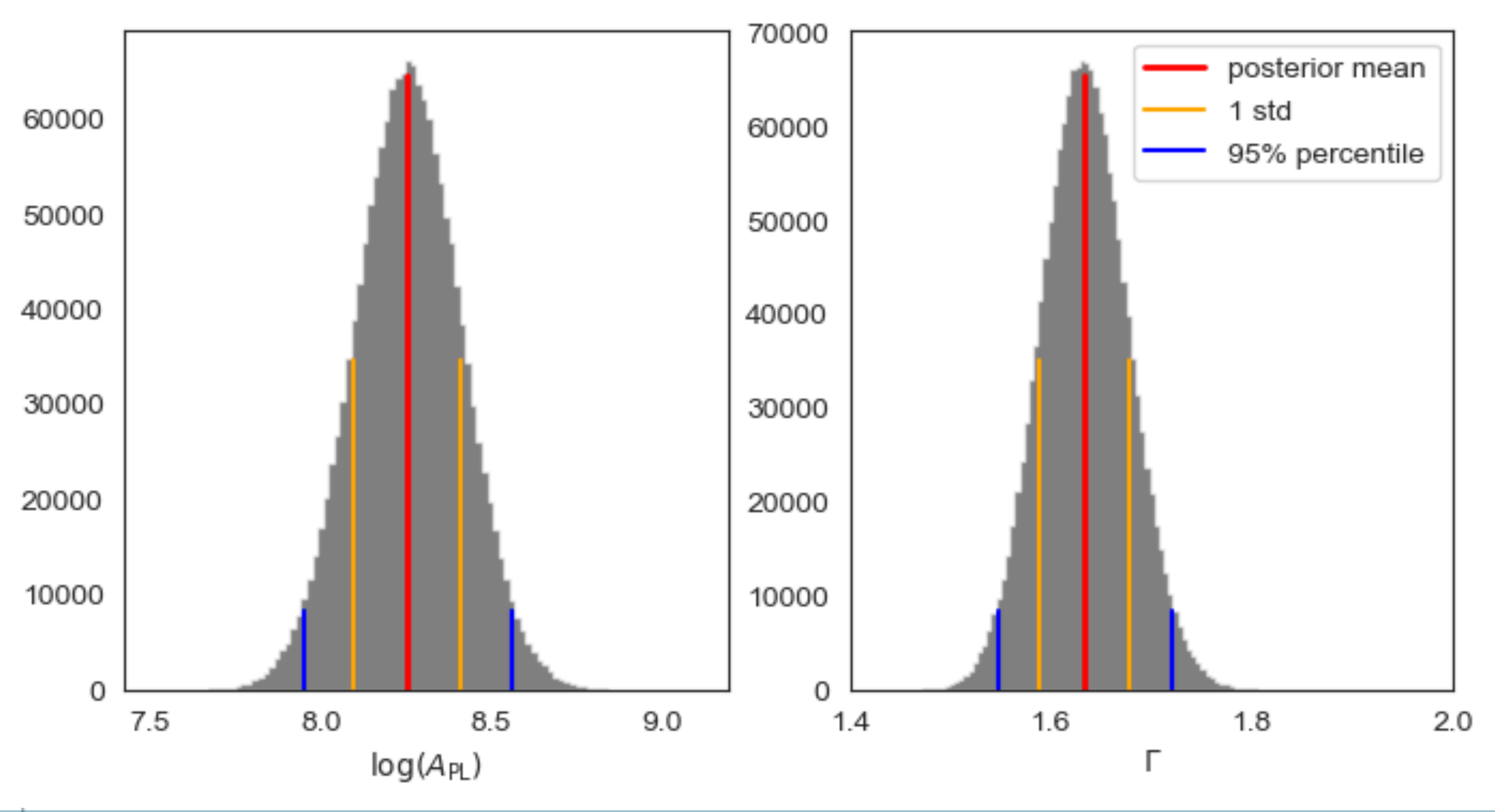
Summarising Posterior Probability Distributions:

Marginalised Distributions + Credible Intervals



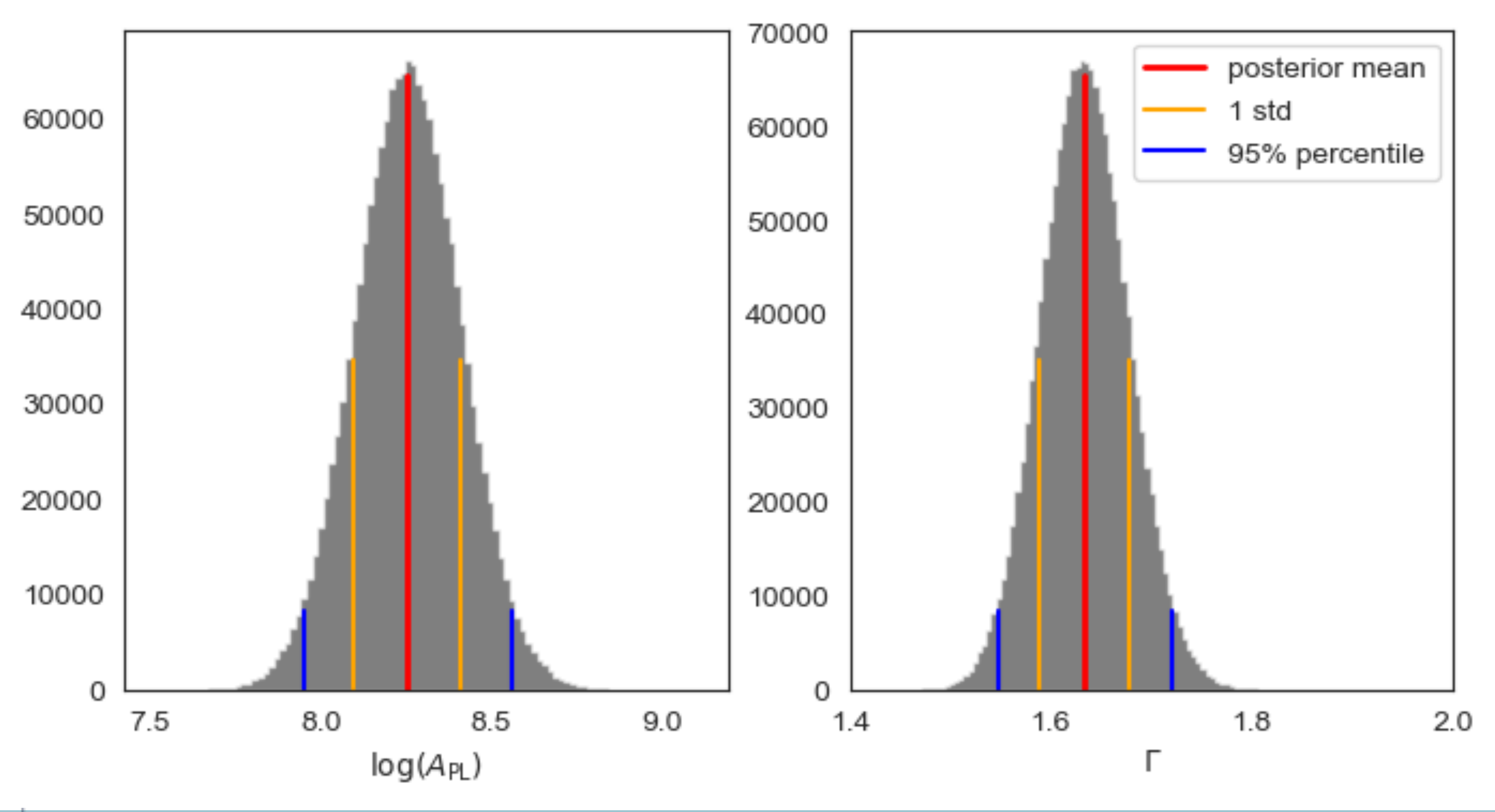
Summarising Posterior Probability Distributions:

Marginalised Distributions + Credible Intervals



Summarising Posterior Probability Distributions:

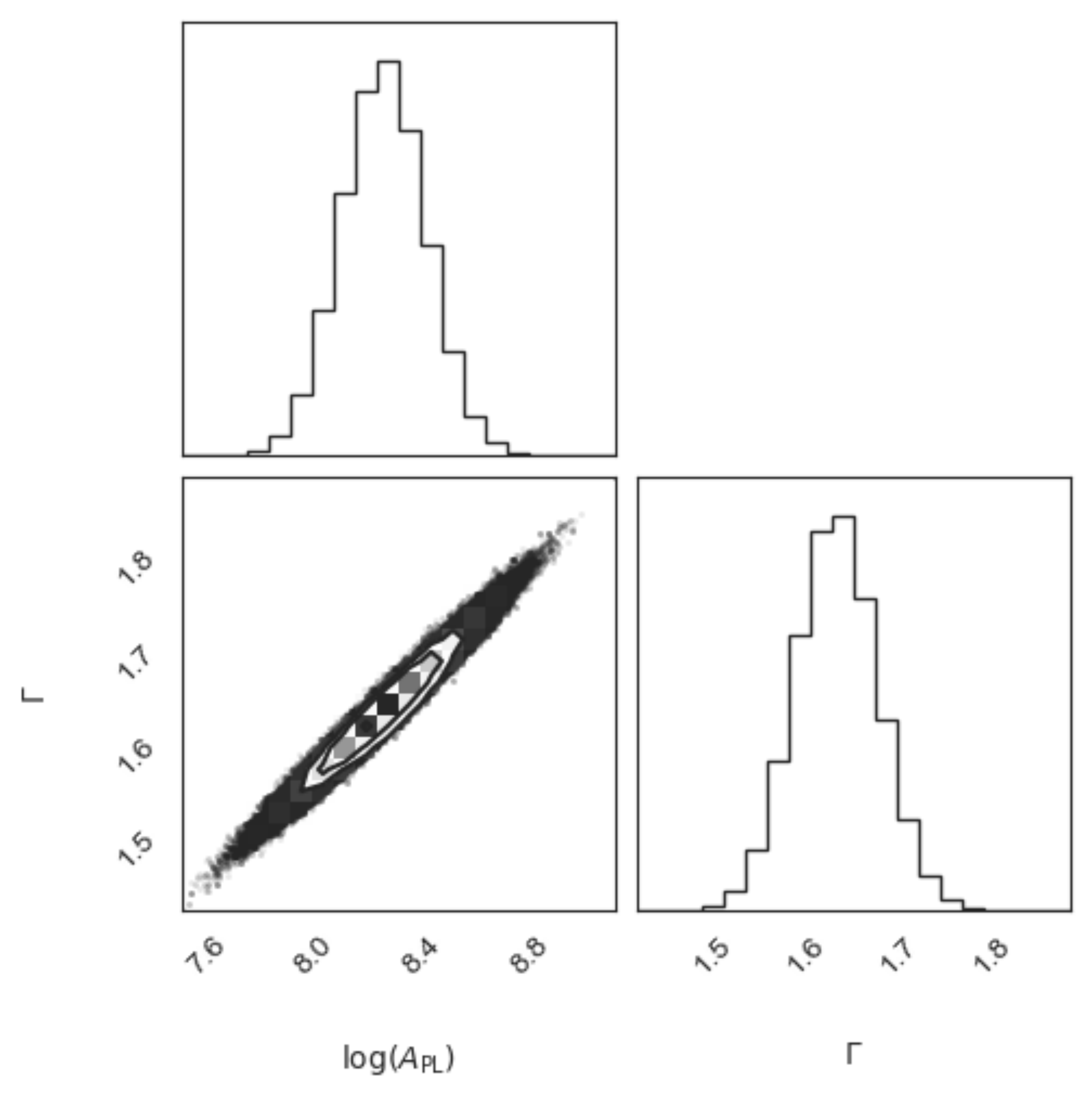
Marginalised Distributions + Credible Intervals



Considering just one dimension of your MCMC sample automatically marginalises over the other dimensions! Magic!

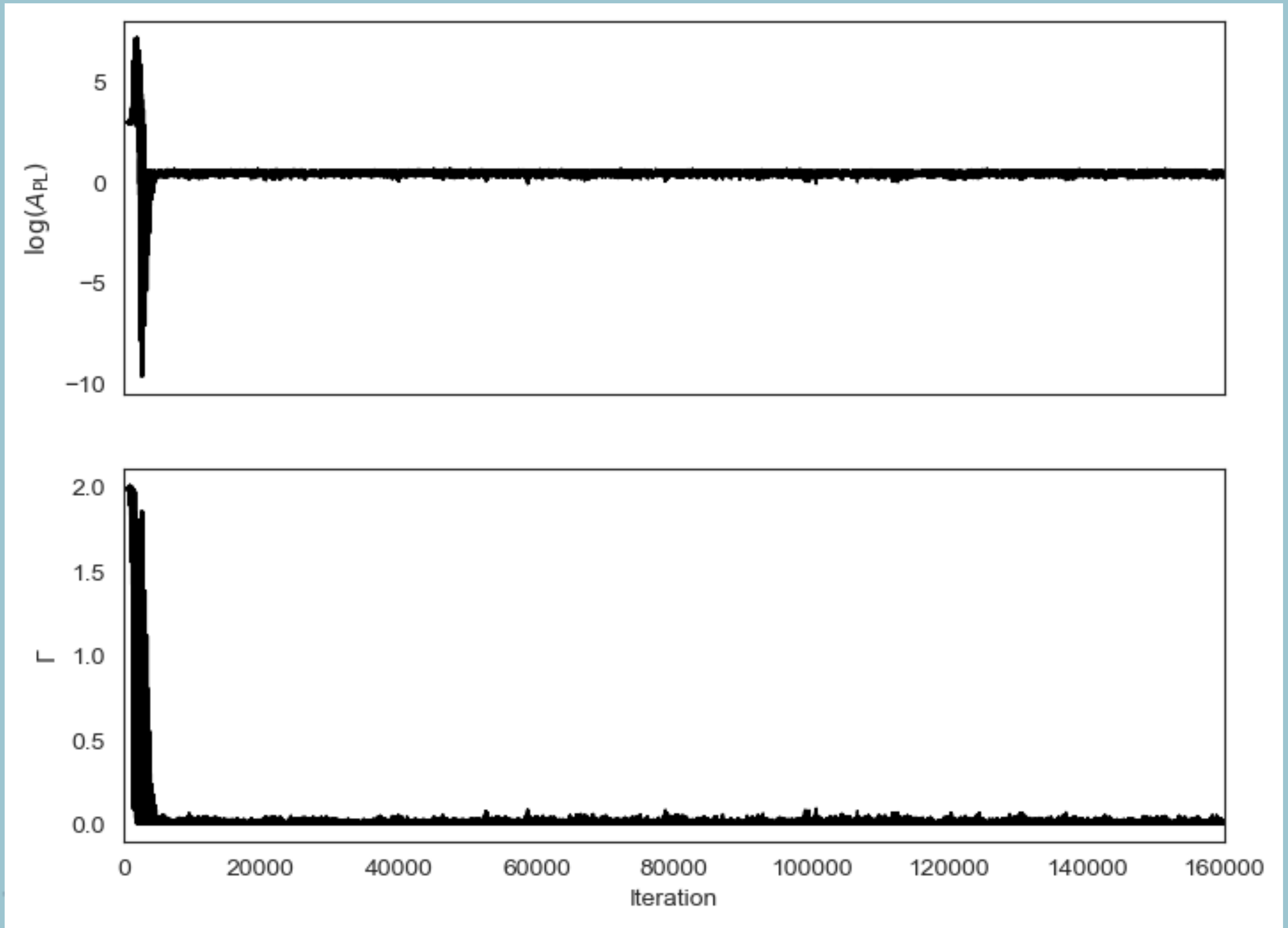
Summarising Posterior Probability Distributions:

Corner Plots or Pair Plots



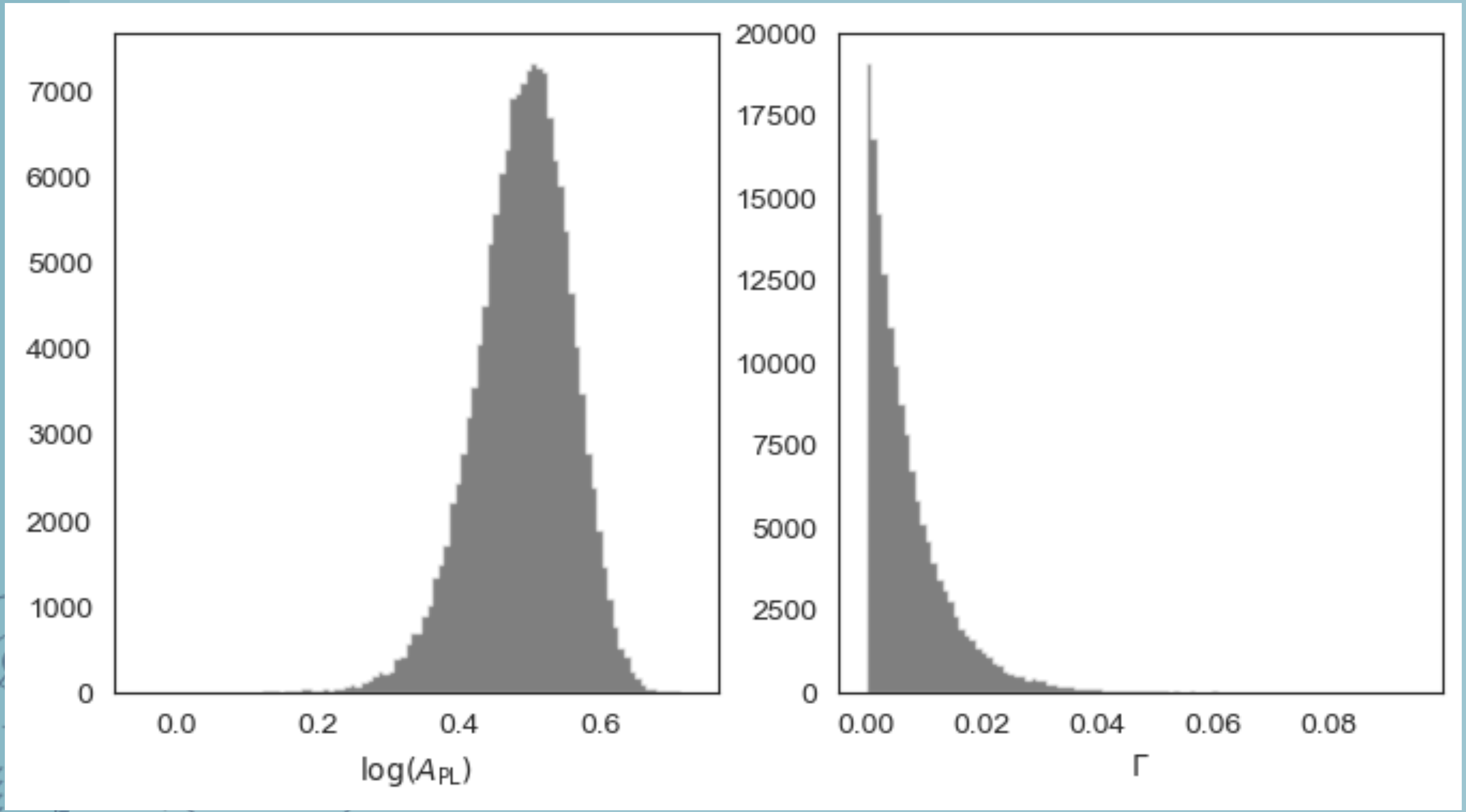
How do you know if/ when things go wrong?

Earlier this week ...



How do you know if/ when things go wrong?

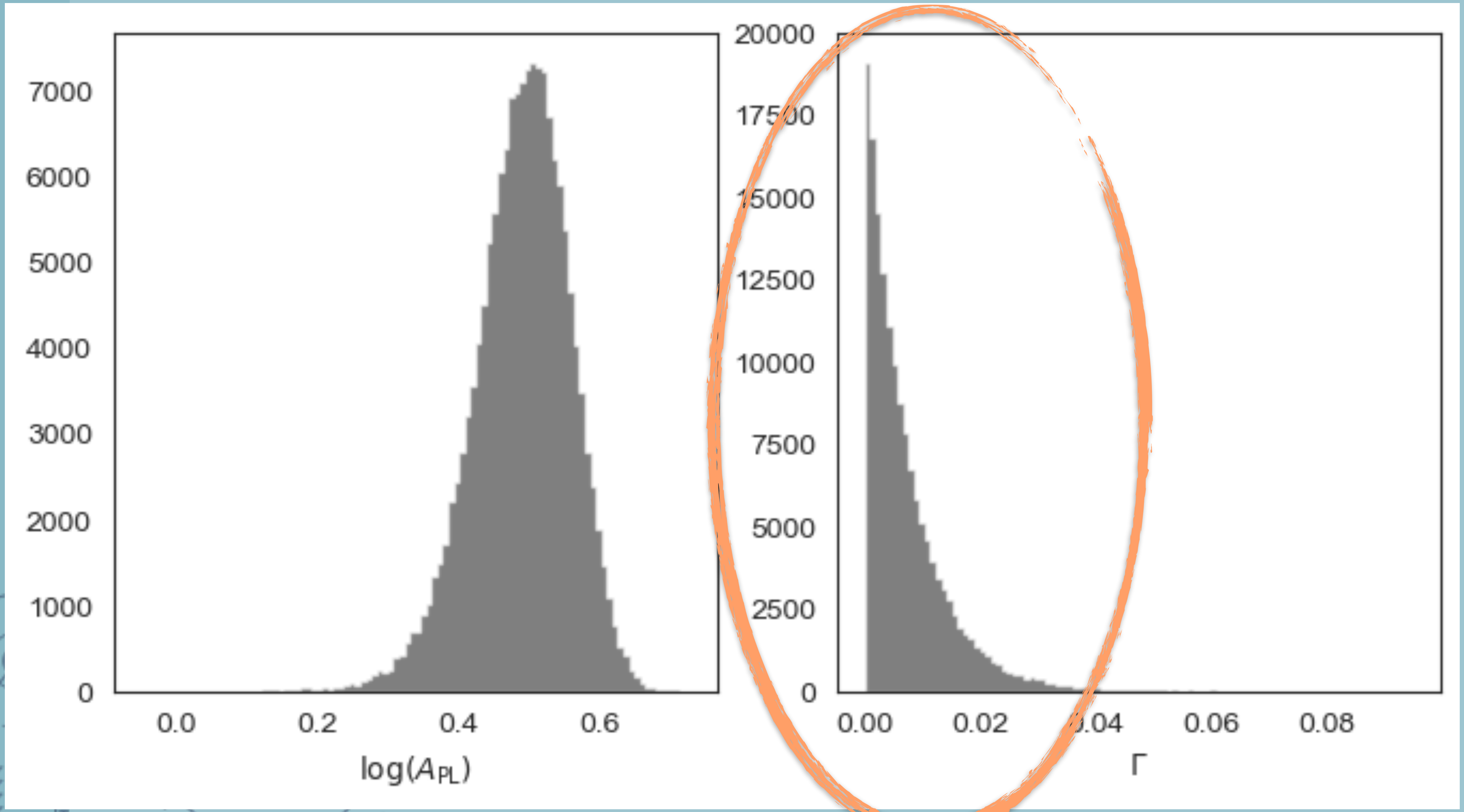
Earlier this week ...



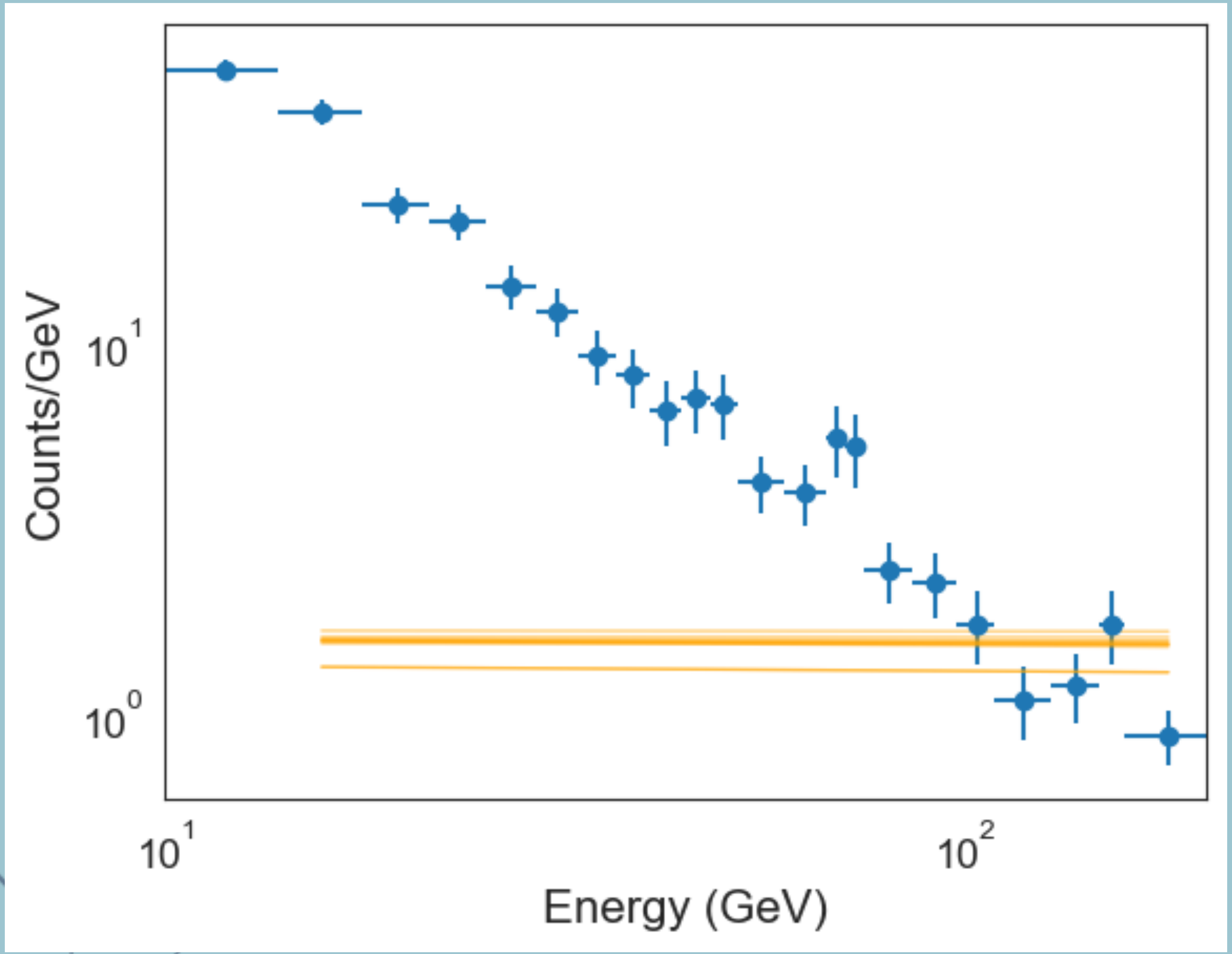
How do you know if/ when things go wrong?

Earlier this week ...

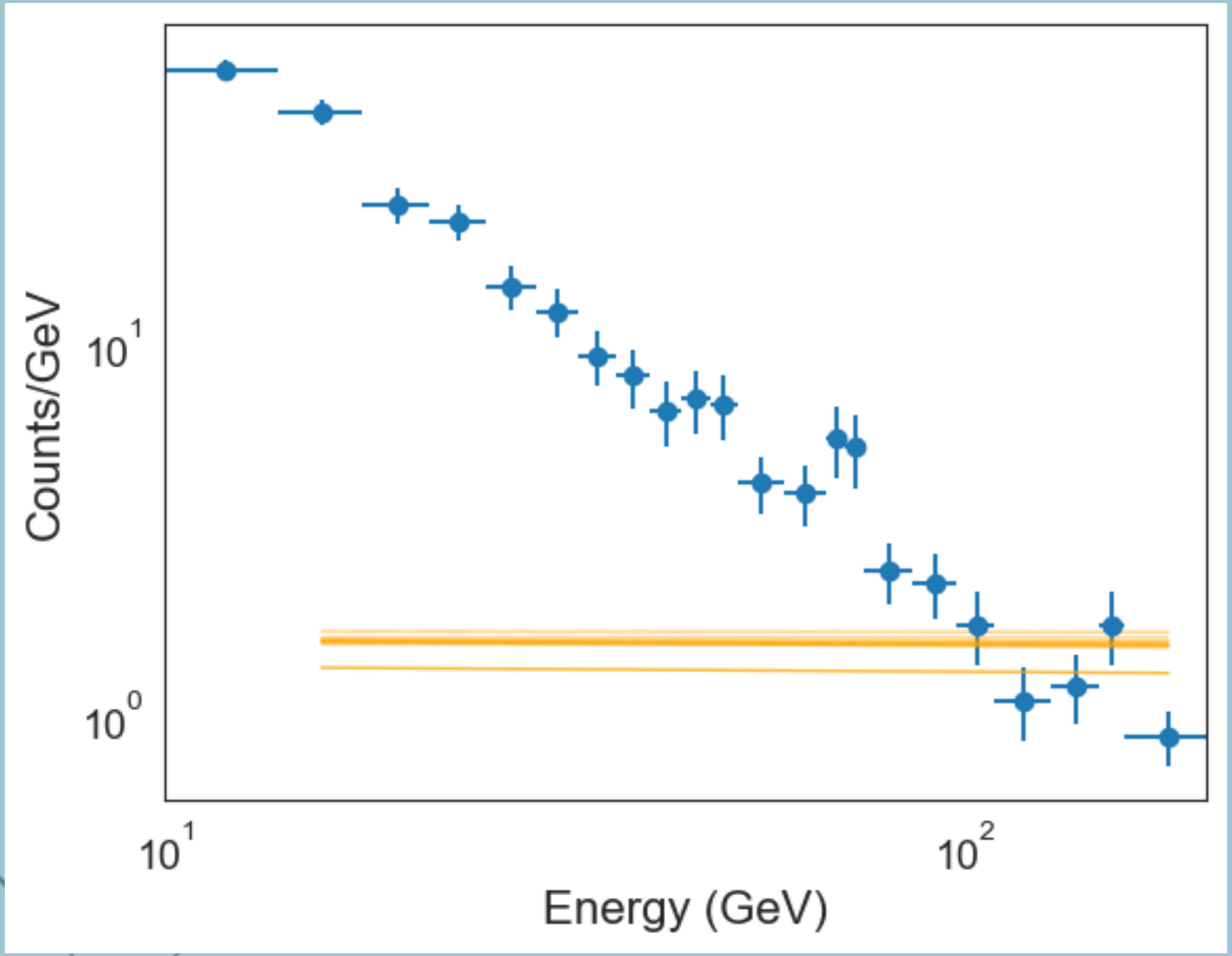
???



Plot posterior draws:

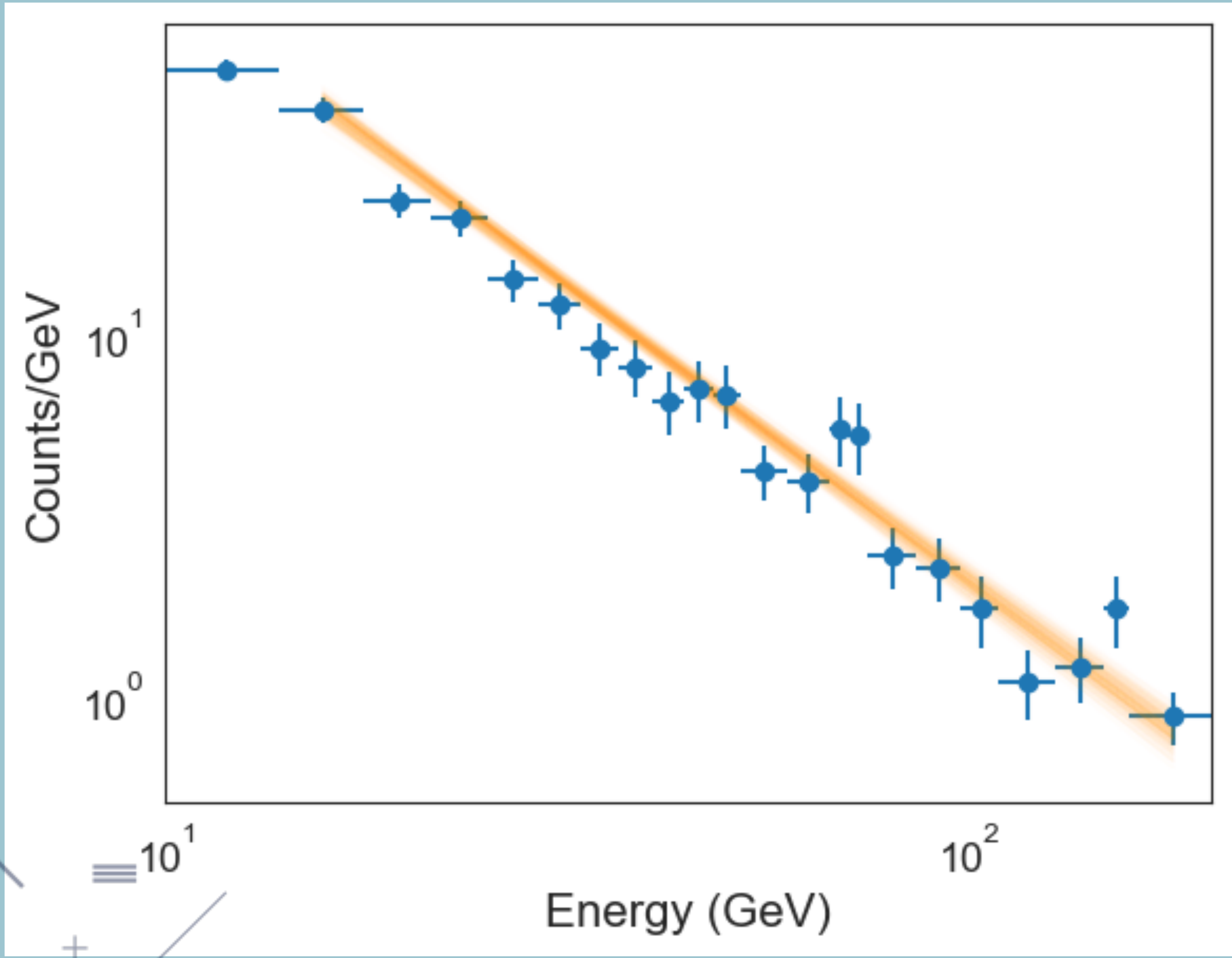


Plot posterior draws:

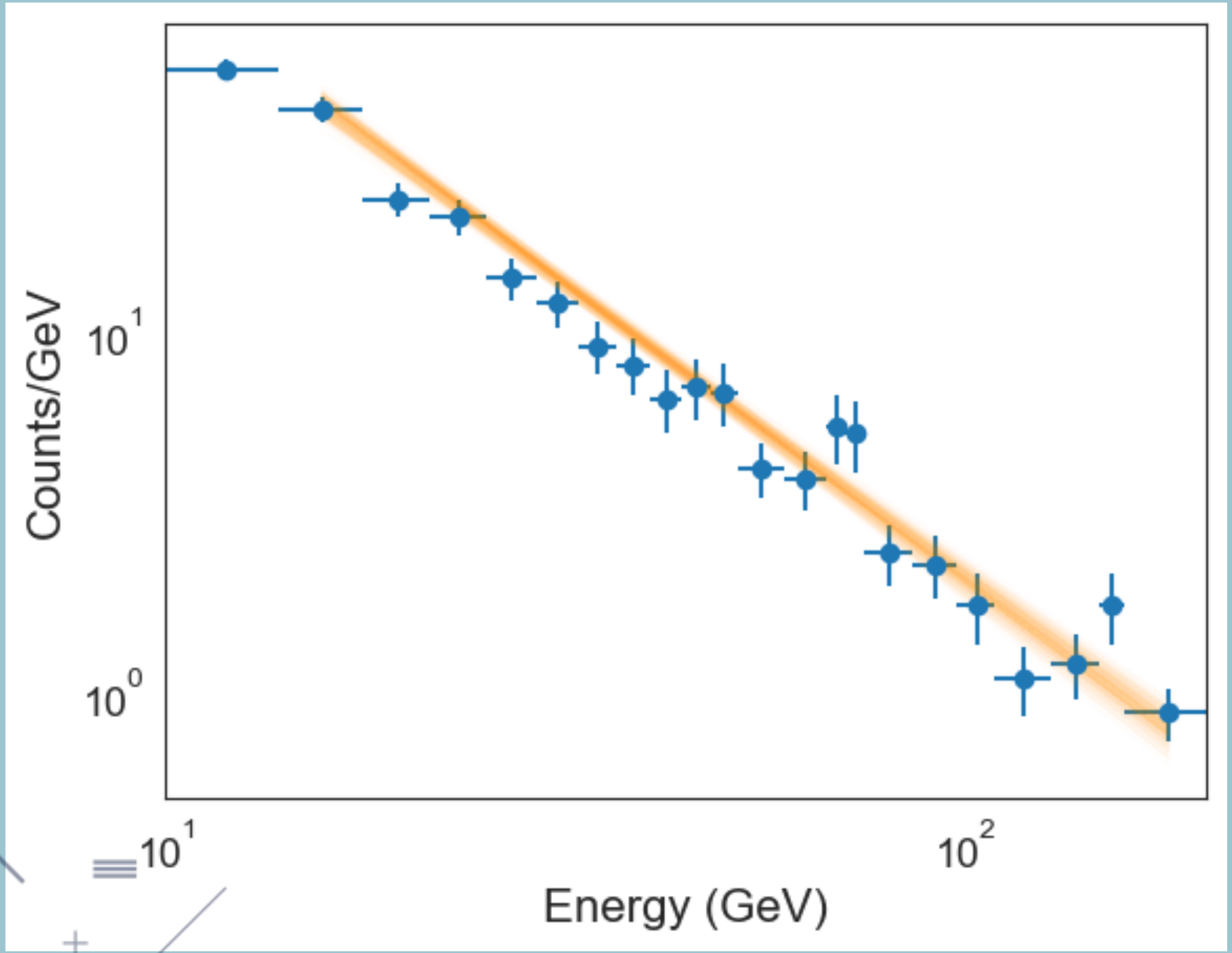


Γ was defined with the wrong sign!

Fixing Γ:



Fixing Γ:



Plotting posterior draws allows you to see the range of models the posterior produces

MCMC Notes and Pitfalls

Further reading:
https://arxiv.org/pdf/1710.06068

MCMC Notes and Pitfalls

Further reading: https://arxiv.org/pdf/1710.06068

You can only do MCMC on the posterior, not on the likelihood!

MCMC Notes and Pitfalls

Further reading:
https://arxiv.org/pdf/1710.06068

You can only do MCMC on the posterior, not on the likelihood!

You don’t want to use MCMC to find the maximum of the distribution (but in many scientific applications, the maximum is not what you’re interested in anyway)

MCMC Notes and Pitfalls

Further reading:
https://arxiv.org/pdf/1710.06068

You can only do MCMC on the posterior, not on the likelihood!

You don't want to use MCMC to find the maximum of the distribution (but in many scientific applications, the maximum is not what you're interested in anyway)

MCMC does not do well on multi-modal distributions (i.e. distributions with multiple peaks), especially when those peaks are well-separated (try it out!)

MCMC Notes and Pitfalls

Further reading: https://arxiv.org/pdf/1710.06068

You can only do MCMC on the posterior, not on the likelihood!

You don't want to use MCMC to find the maximum of the distribution (but in many scientific applications, the maximum is not what you're interested in anyway)

MCMC does not do well on multi-modal distributions (i.e. distributions with multiple peaks), especially when those peaks are well-separated (try it out!)

The proposal distribution is an important tuning factor in Metropolis-Hastings MCMC (try it out!)

MCMC Notes and Pitfalls

Further reading: https://arxiv.org/pdf/1710.06068

You can only do MCMC on the posterior, not on the likelihood!

You don't want to use MCMC to find the maximum of the distribution (but in many scientific applications, the maximum is not what you're interested in anyway)

MCMC does not do well on multi-modal distributions (i.e. distributions with multiple peaks), especially when those peaks are well-separated (try it out!)

The proposal distribution is an important tuning factor in Metropolis-Hastings MCMC (try it out!)

There are a algorithms that try to circumvent the last two issues (e.g. Affine Invariant MCMC Ensemble Samplers, https://emcee.readthedocs.io)

MCMC Notes and Pitfalls

Further reading: https://arxiv.org/pdf/1710.06068

You can only do MCMC on the posterior, not on the likelihood!

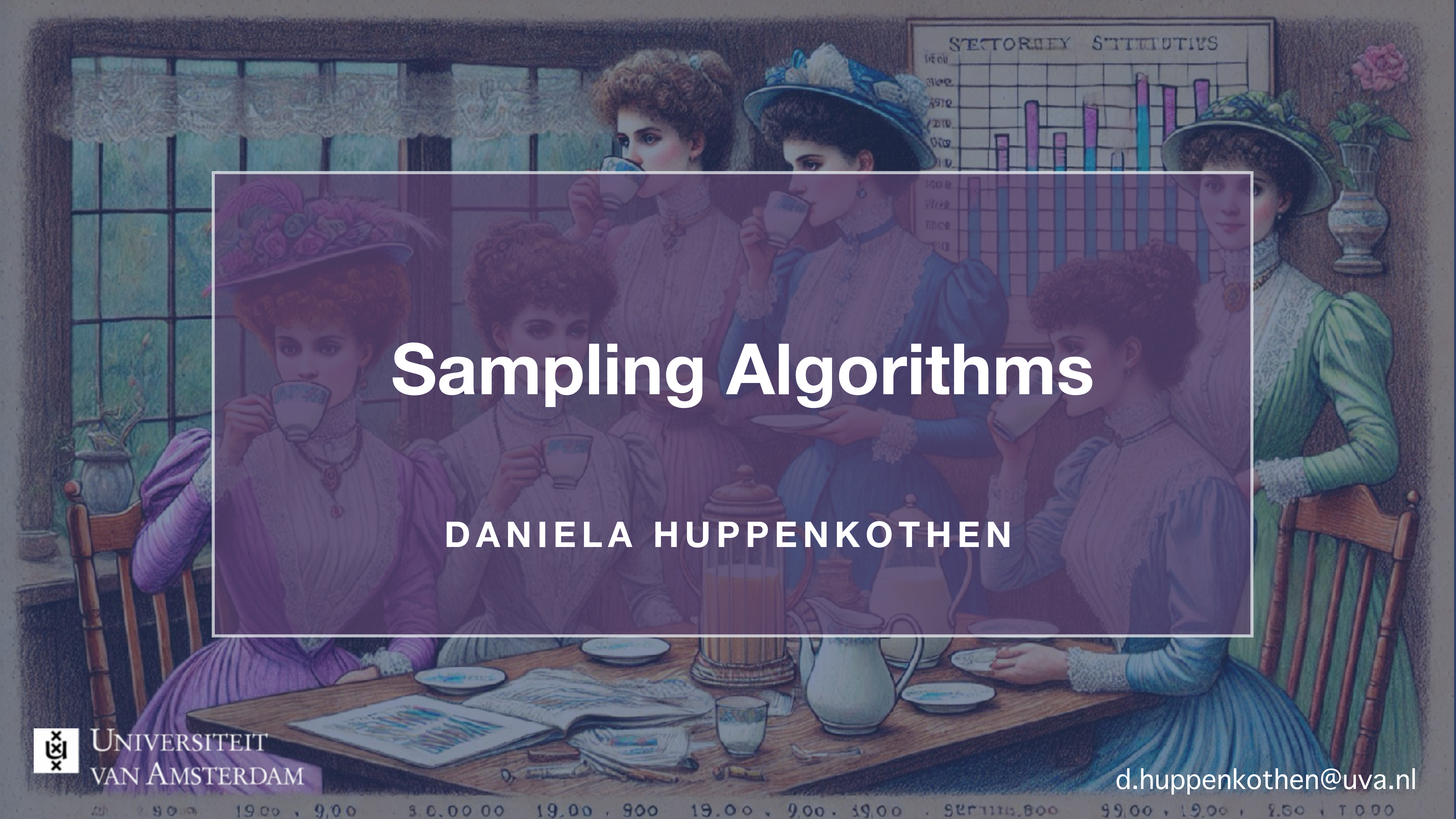
You don't want to use MCMC to find the maximum of the distribution (but in many scientific applications, the maximum is not what you're interested in anyway)

MCMC does not do well on multi-modal distributions (i.e. distributions with multiple peaks), especially when those peaks are well-separated (try it out!)

The proposal distribution is an important tuning factor in Metropolis-Hastings MCMC (try it out!)

There are a algorithms that try to circumvent the last two issues (e.g. Affine Invariant MCMC Ensemble Samplers, https://emcee.readthedocs.io)

Sampling is not a one-size-fits-all solution! Different problems may do better with different sampling algorithms! And there are many (see also: Hamiltonian Monte Carlo, Importance Sampling, Slice Sampling, Gibbs Sampling, Nested Sampling, AI-supported sampling algorithms, ...)



Sampling Algorithms

DANIELA HUPPENKOTHEN

Universiteit van Amsterdam

d.huppenkothen@uva.nl

Gibbs Sampling

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*
- At each iteration t, each θ_j^t is sampled from the *conditional distribution given all the other components* $p(\theta_j \mid \theta_{-j}^{t-1}, y)$

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*
- At each iteration t, each θ_j^t is sampled from the *conditional distribution given all the other components* $p(\theta_j \mid \theta_{-j}^{t-1}, y)$
- The Metropolis algorithms and Gibbs sampler can be used in combination

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*
- At each iteration t, each θ_j^t is sampled from the *conditional distribution given all the other components* $p(\theta_j \mid \theta_{-j}^{t-1}, y)$
- The Metropolis algorithms and Gibbs sampler can be used in combination
  - Useful when some parameters can be sampled directly while others cannot

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*
- At each iteration t, each θ_j^t is sampled from the *conditional distribution given all the other components* $p(\theta_j \mid \theta_{-j}^{t-1}, y)$
- The Metropolis algorithms and Gibbs sampler can be used in combination
  - Useful when some parameters can be sampled directly while others cannot
  - Update parameters in *blocks*

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*
- At each iteration t, each θ_j^t is sampled from the *conditional distribution given all the other components* $p(\theta_j \mid \theta_{-j}^{t-1}, y)$
- The Metropolis algorithms and Gibbs sampler can be used in combination
  - Useful when some parameters can be sampled directly while others cannot
  - Update parameters in *blocks*
- Relies on knowing the conditional posterior distributions

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*
- At each iteration t, each θ_j^t is sampled from the *conditional distribution given all the other components* $p(\theta_j \mid \theta_{-j}^{t-1}, y)$
- The Metropolis algorithms and Gibbs sampler can be used in combination
  - Useful when some parameters can be sampled directly while others cannot
  - Update parameters in *blocks*
- Relies on knowing the conditional posterior distributions
- Can become computationally inefficient for large numbers of parameters

Gibbs Sampling

- Useful in multidimensional problems where you want to know the target distribution $p(\theta \mid y)$
  - i.e., you have d parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$
- Instead of sampling all parameters simultaneously at every step in the Markov chain, sample one (or more) at a time based on the *conditional* distribution of all the other current parameter values
- also known as *alternating conditional sampling*
- At each iteration t, each θ_j^t is sampled from the *conditional distribution given all the other components* $p(\theta_j \mid \theta_{-j}^{t-1}, y)$
- The Metropolis algorithms and Gibbs sampler can be used in combination
  - Useful when some parameters can be sampled directly while others cannot
  - Update parameters in *blocks*
- Relies on knowing the conditional posterior distributions
- Can become computationally inefficient for large numbers of parameters
- Tend to get stuck in regions of parameter space for a long time, especially if high correlation in posterior and many dimensions

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.
5. Repeat from step 2 until you have as many draws as you want.

Slice Sampling

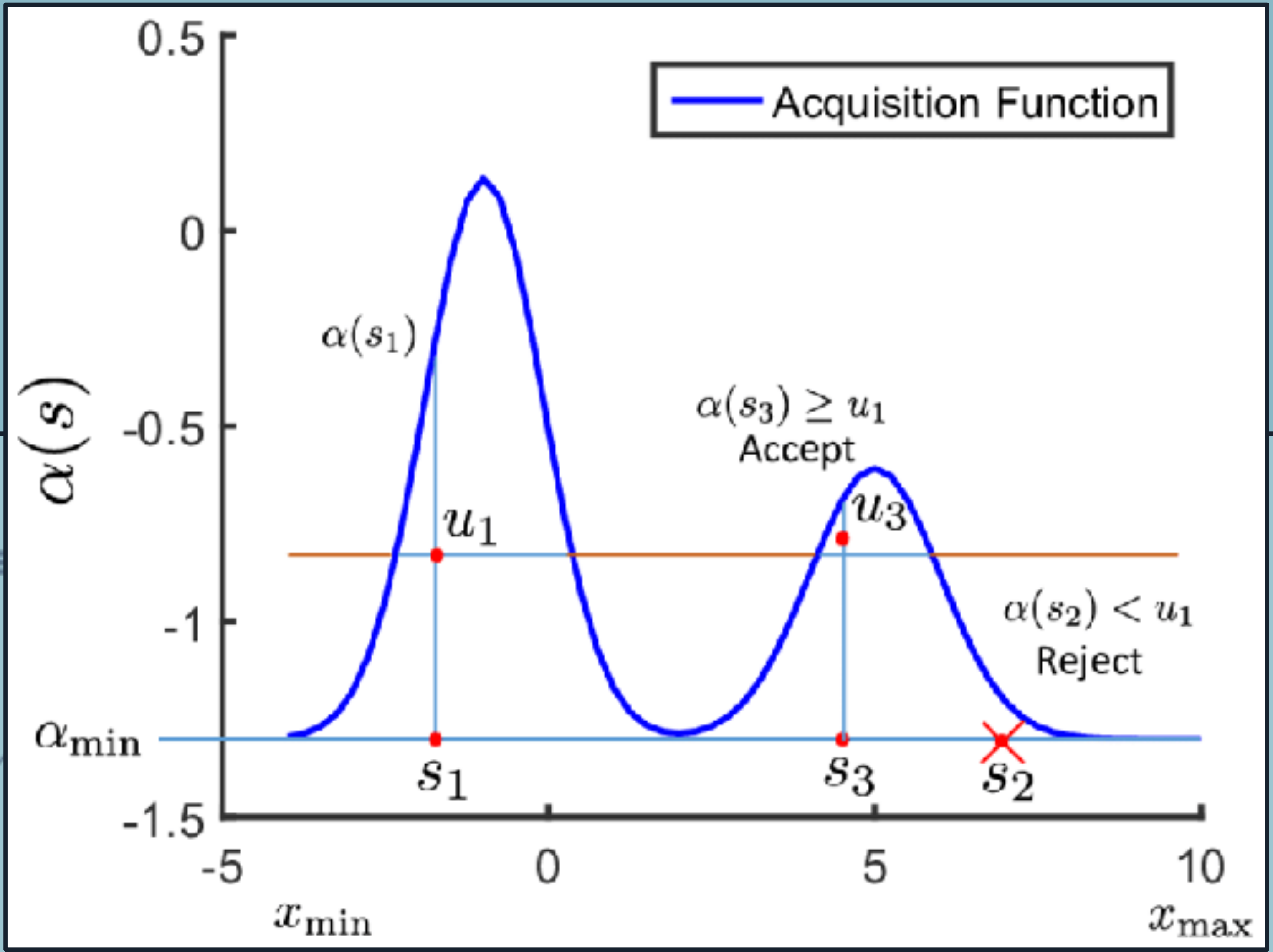
Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.
5. Repeat from step 2 until you have as many draws as you want.



Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.
5. Repeat from step 2 until you have as many draws as you want.

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.
5. Repeat from step 2 until you have as many draws as you want.

- In 1D: easier to implement than Metropolis Hastings

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.
5. Repeat from step 2 until you have as many draws as you want.

- In 1D: easier to implement than Metropolis Hastings
- Slice-sampling auto-adjusts the step size

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.
5. Repeat from step 2 until you have as many draws as you want.

- In 1D: easier to implement than Metropolis Hastings
- Slice-sampling auto-adjusts the step size
- Often slower than Metropolis Hastings.

Slice Sampling

Further reading:

- https://www.cs.toronto.edu/pub/radford/slice-aos.pdf
- arXiv:1001.0175

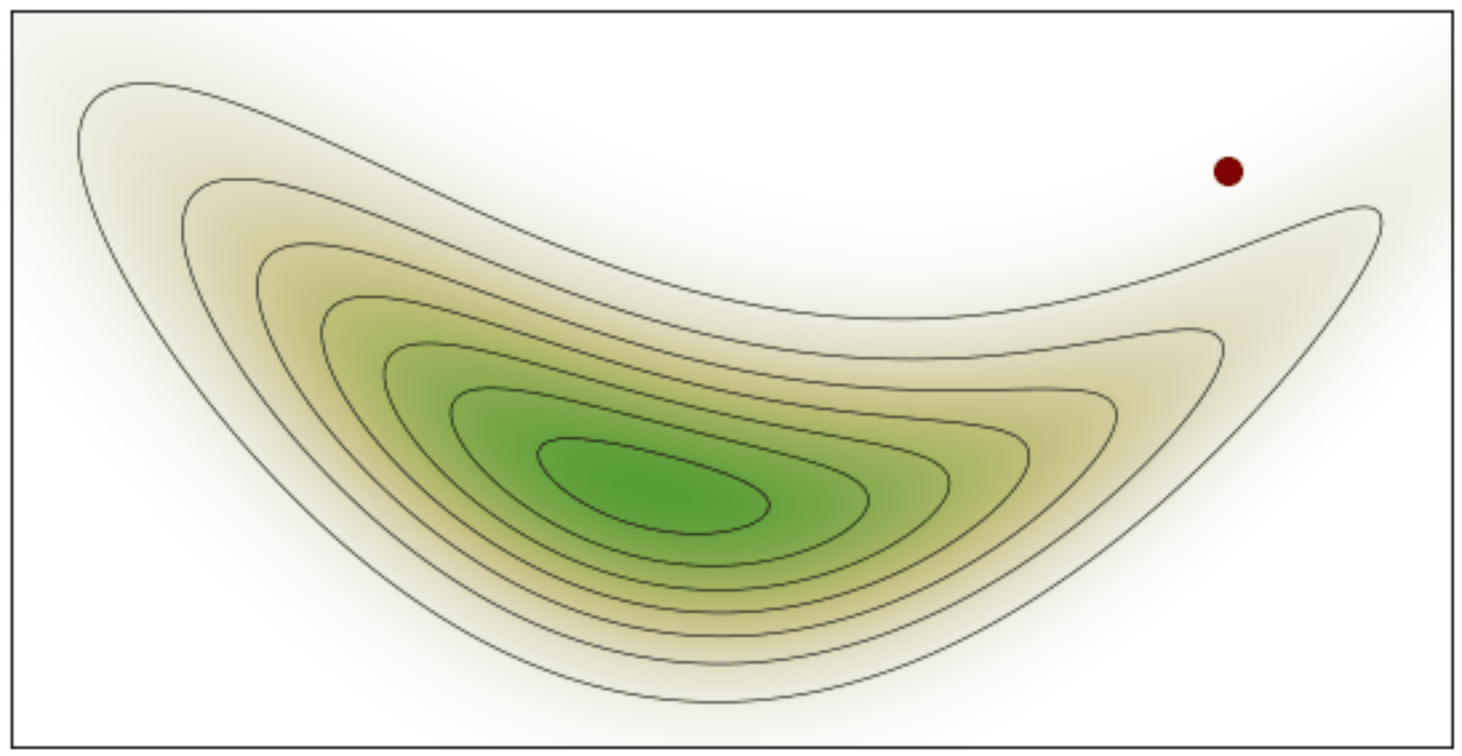
Issue with MH and Gibbs: you have to know the step size!

Slice sampling:

1. Choose a random starting value of x, let's call it x_1
2. Uniformly sample on interval [0, f(x1)], let's call it a
3. Imagine a horizontal line at $y = a$: figure out all the line segments under the curve.
4. From all the line segments, draw a value of x uniformly.
5. Repeat from step 2 until you have as many draws as you want.

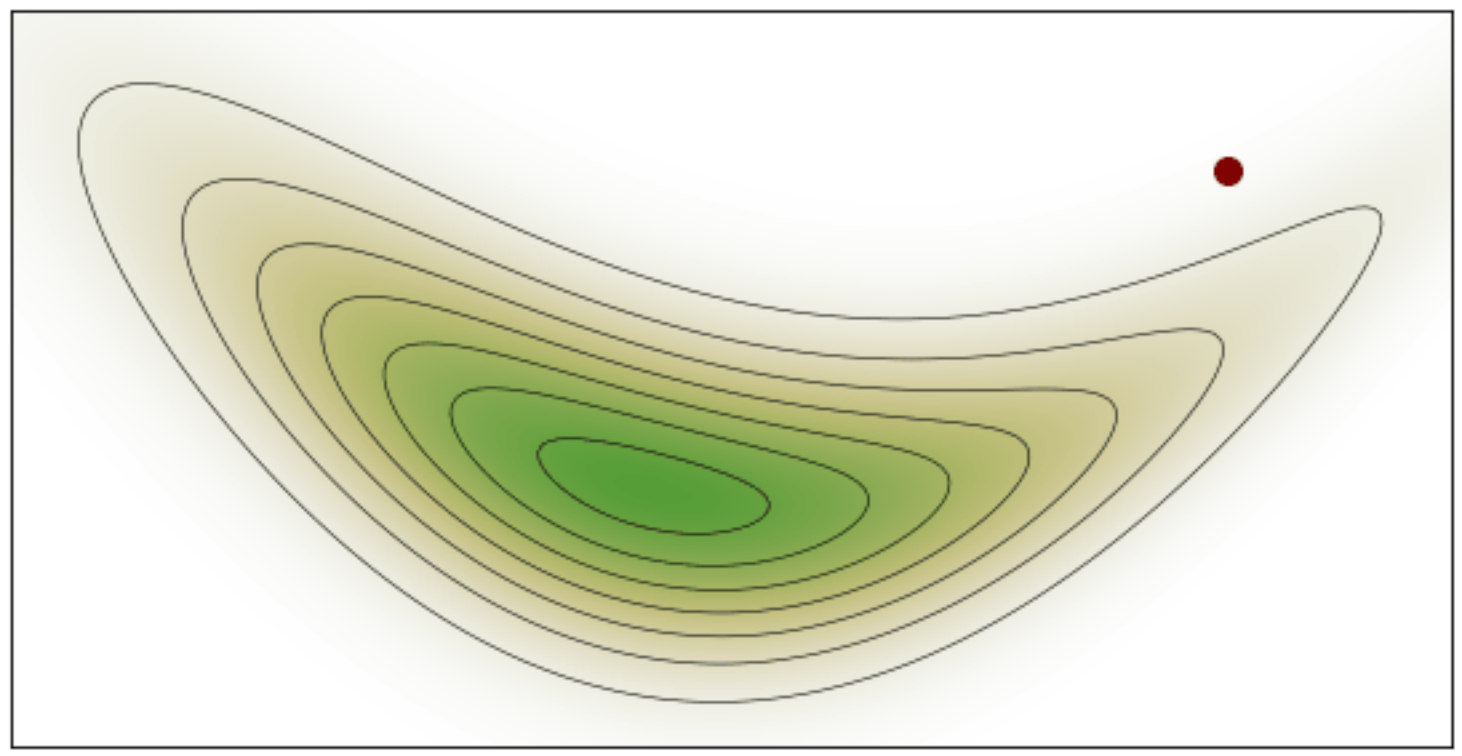
- In 1D: easier to implement than Metropolis Hastings
- Slice-sampling auto-adjusts the step size
- Often slower than Metropolis Hastings.
- Finding the roots of the intersection between horizontal line and distribution is pretty tricky.

Hamiltonian Monte Carlo



Hamiltonian Monte Carlo

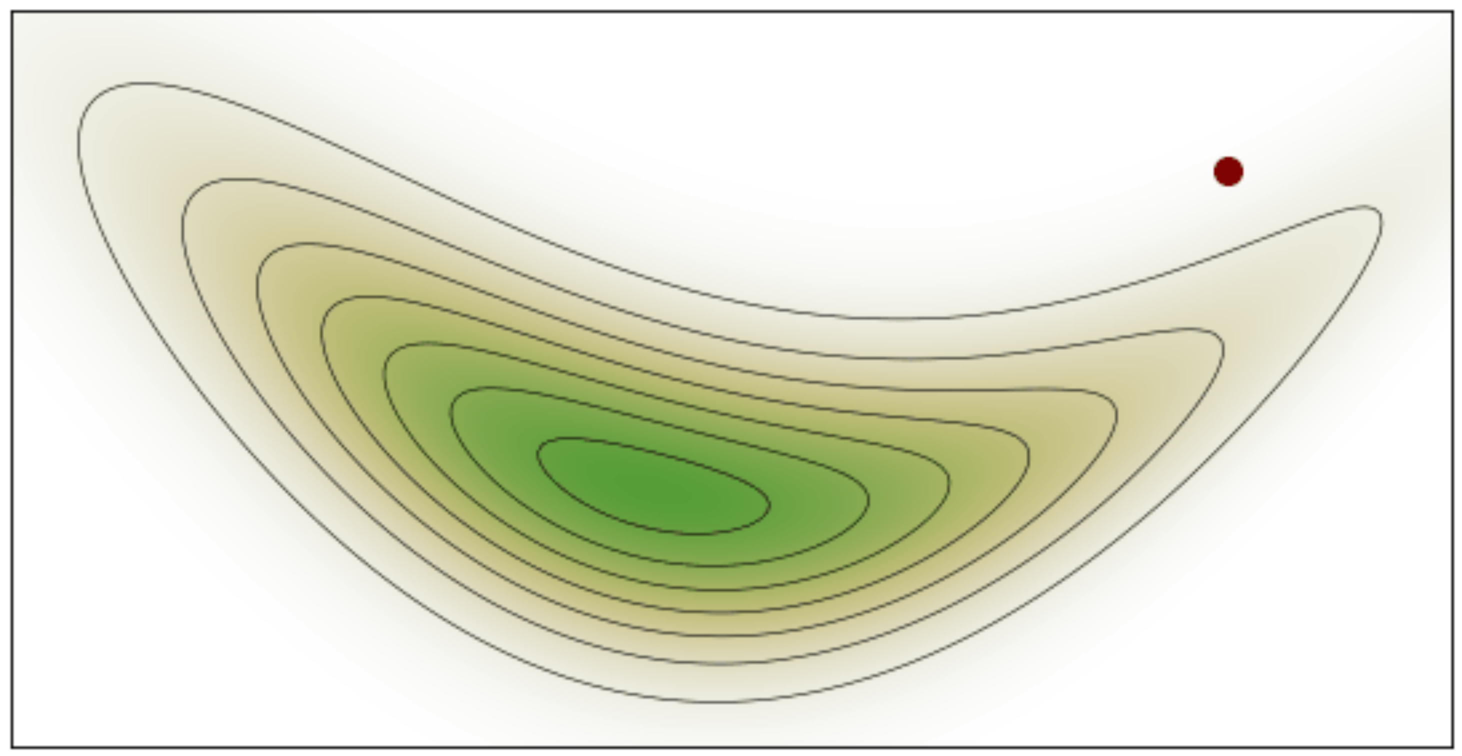
Use Hamiltonian dynamics to propose a move to a new point in state space



Hamiltonian Monte Carlo

Use Hamiltonian dynamics to propose a move to a new point in state space

Reduces correlations between successive samples by proposing moves to distant states which maintain a high probability of acceptance

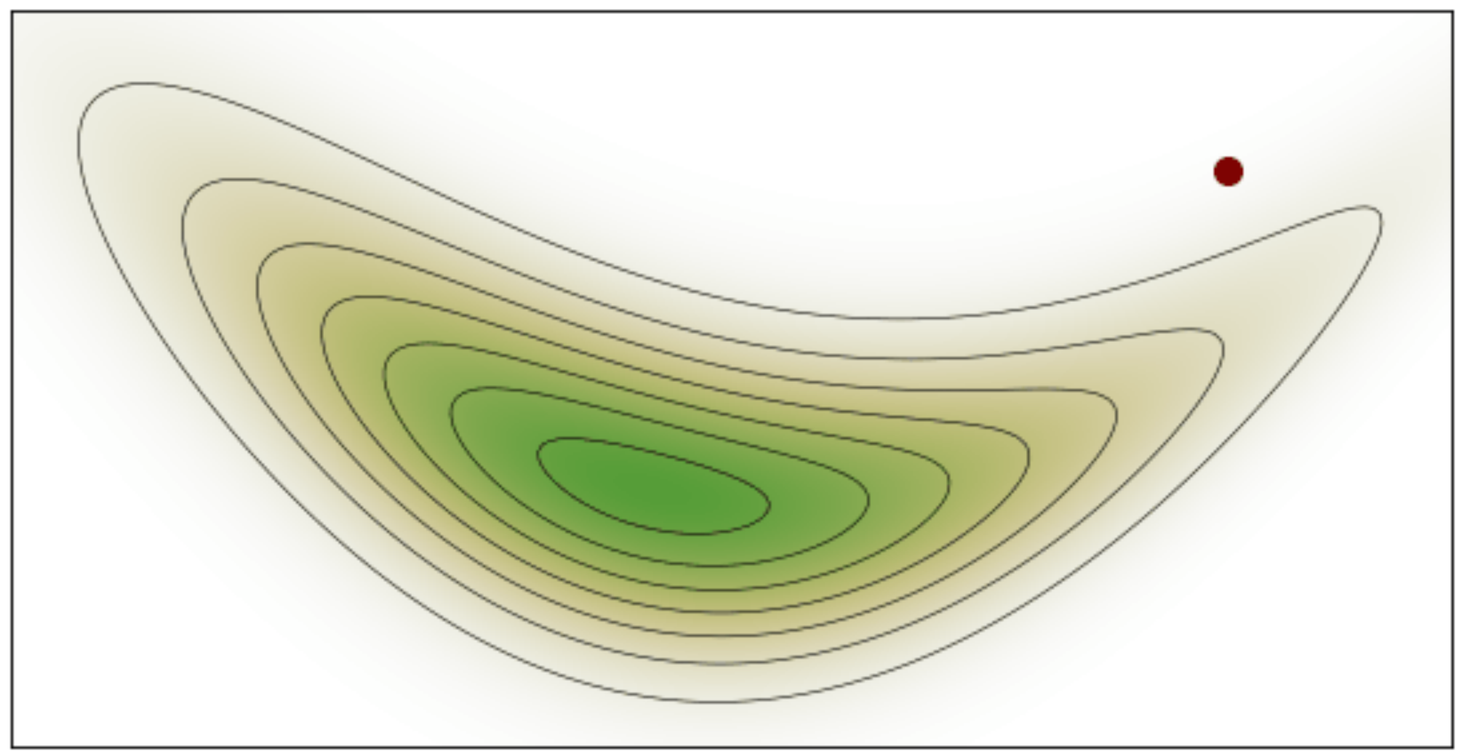


Hamiltonian Monte Carlo

Use Hamiltonian dynamics to propose a move to a new point in state space

Reduces correlations between successive samples by proposing moves to distant states which maintain a high probability of acceptance

Needs fewer Markov chain samples (shorter chains)



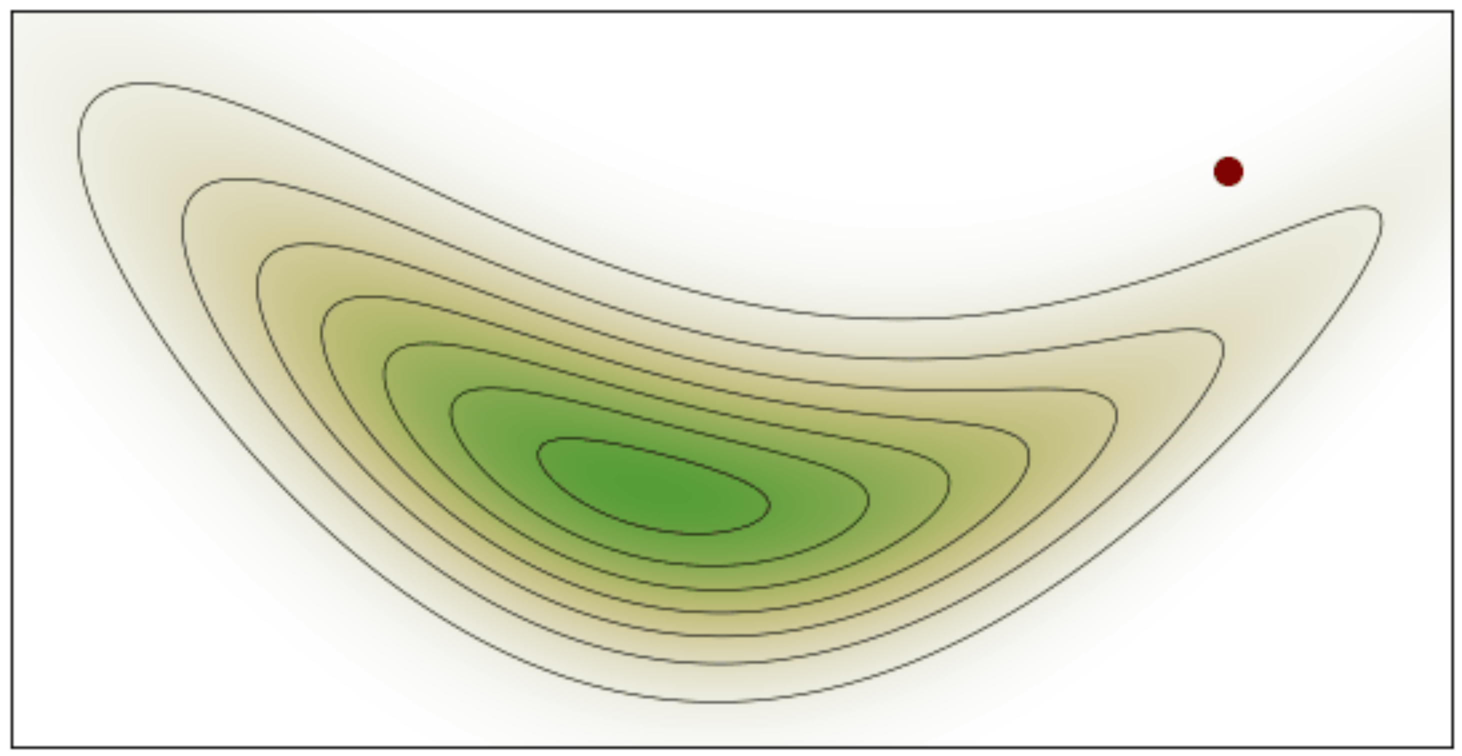
Hamiltonian Monte Carlo

Use Hamiltonian dynamics to propose a move to a new point in state space

Reduces correlations between successive samples by proposing moves to distant states which maintain a high probability of acceptance

Needs fewer Markov chain samples (shorter chains)

Requires access to (or the ability to approximate) the Hamiltonian (i.e. gradients) of the posterior



Hamiltonian Monte Carlo

Use Hamiltonian dynamics to propose a move to a new point in state space

Reduces correlations between successive samples by proposing moves to distant states which maintain a high probability of acceptance

Needs fewer Markov chain samples (shorter chains)

Requires access to (or the ability to approximate) the Hamiltonian (i.e. gradients) of the posterior