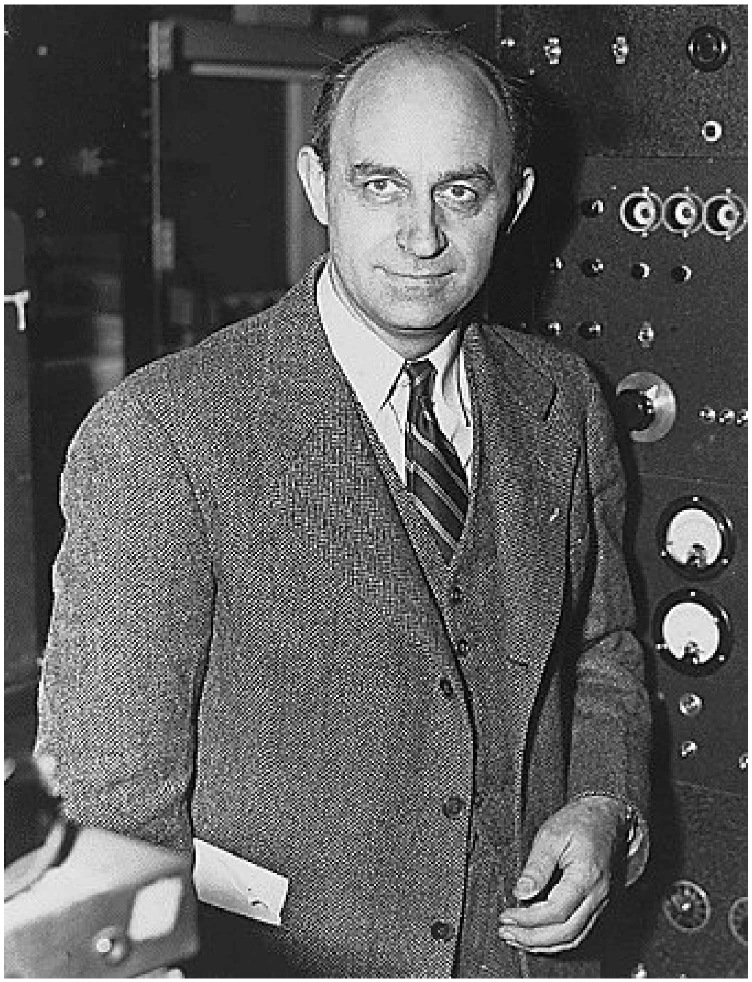
Monte Carlo

— Monte Carlo basics and Motivation
— Rejection sampling
— Importance sampling
— Next time: Markov chain Monte Carlo

Iain Murray
`http://iainmurray.net/`

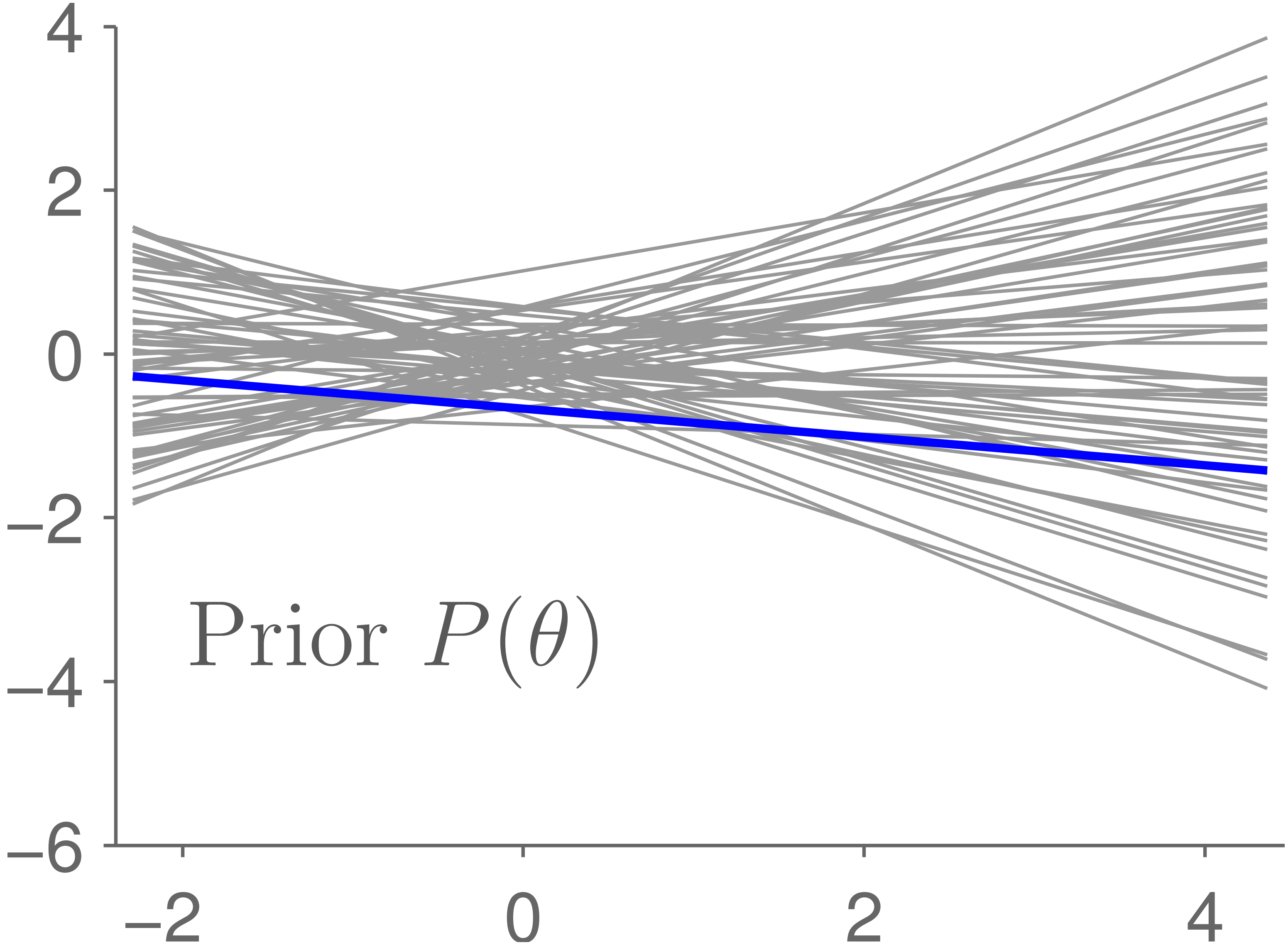
Monte Carlo and Insomnia



Enrico Fermi (1901–1954) took great delight in astonishing his colleagues with his remakably accurate predictions of experimental results. . . he revealed that his "guesses" were really derived from the statistical sampling techniques that he used to calculate with whenever insomnia struck in the wee morning hours!

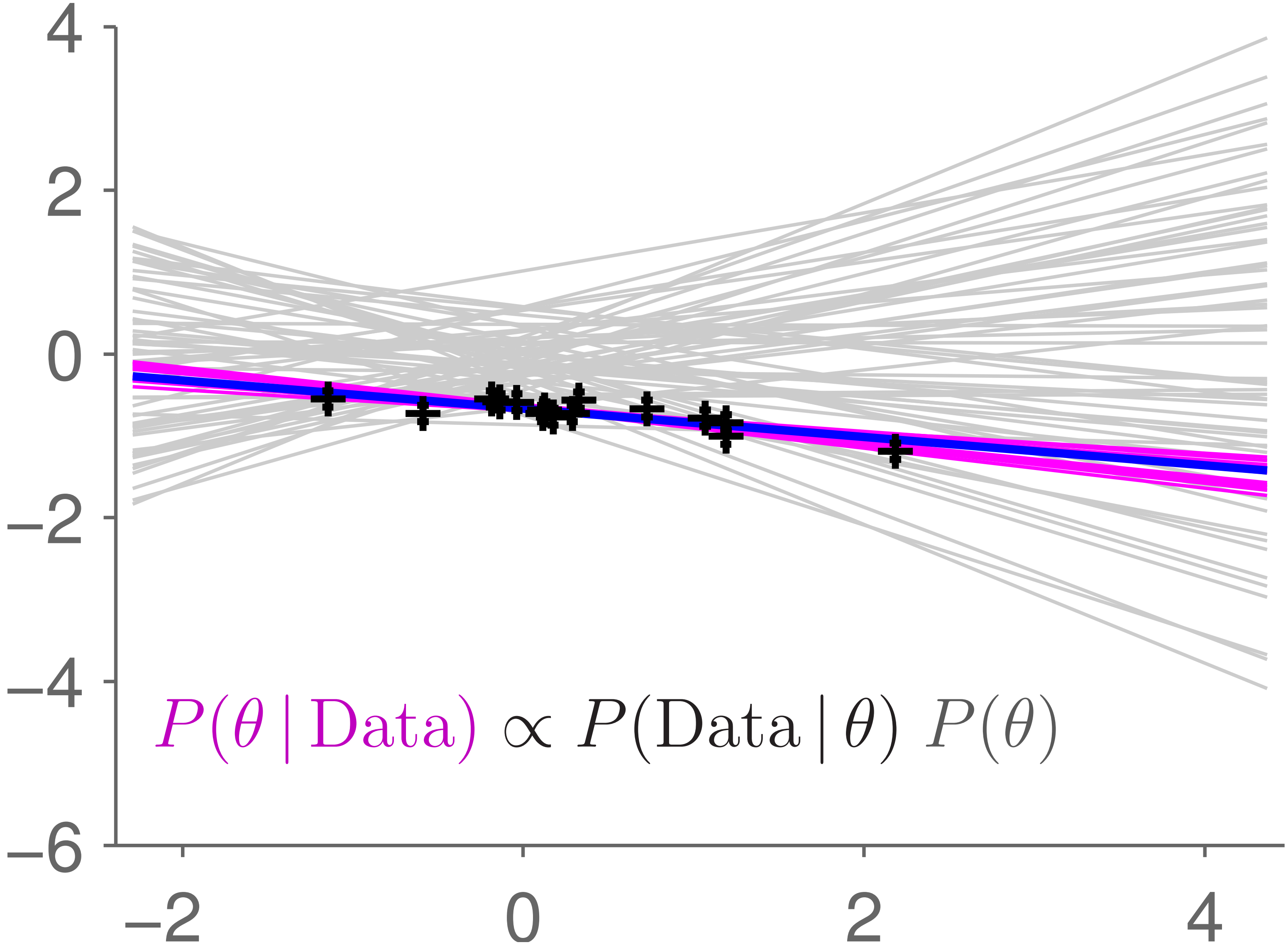
—*The beginning of the Monte Carlo method*, N. Metropolis

Linear Regression: Prior



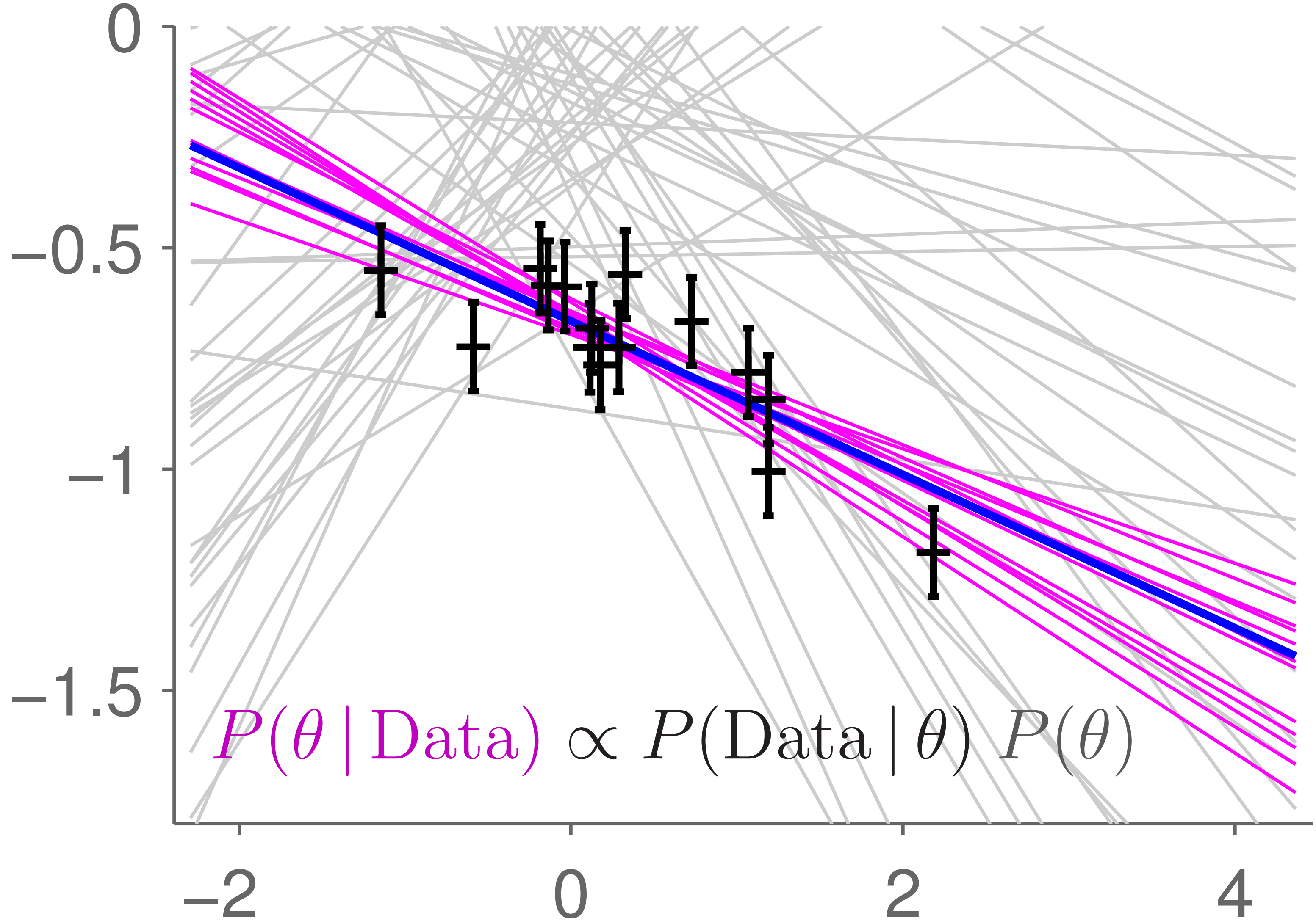
Input $\rightarrow$ output mappings considered plausible before seeing data.

Linear Regression: Posterior



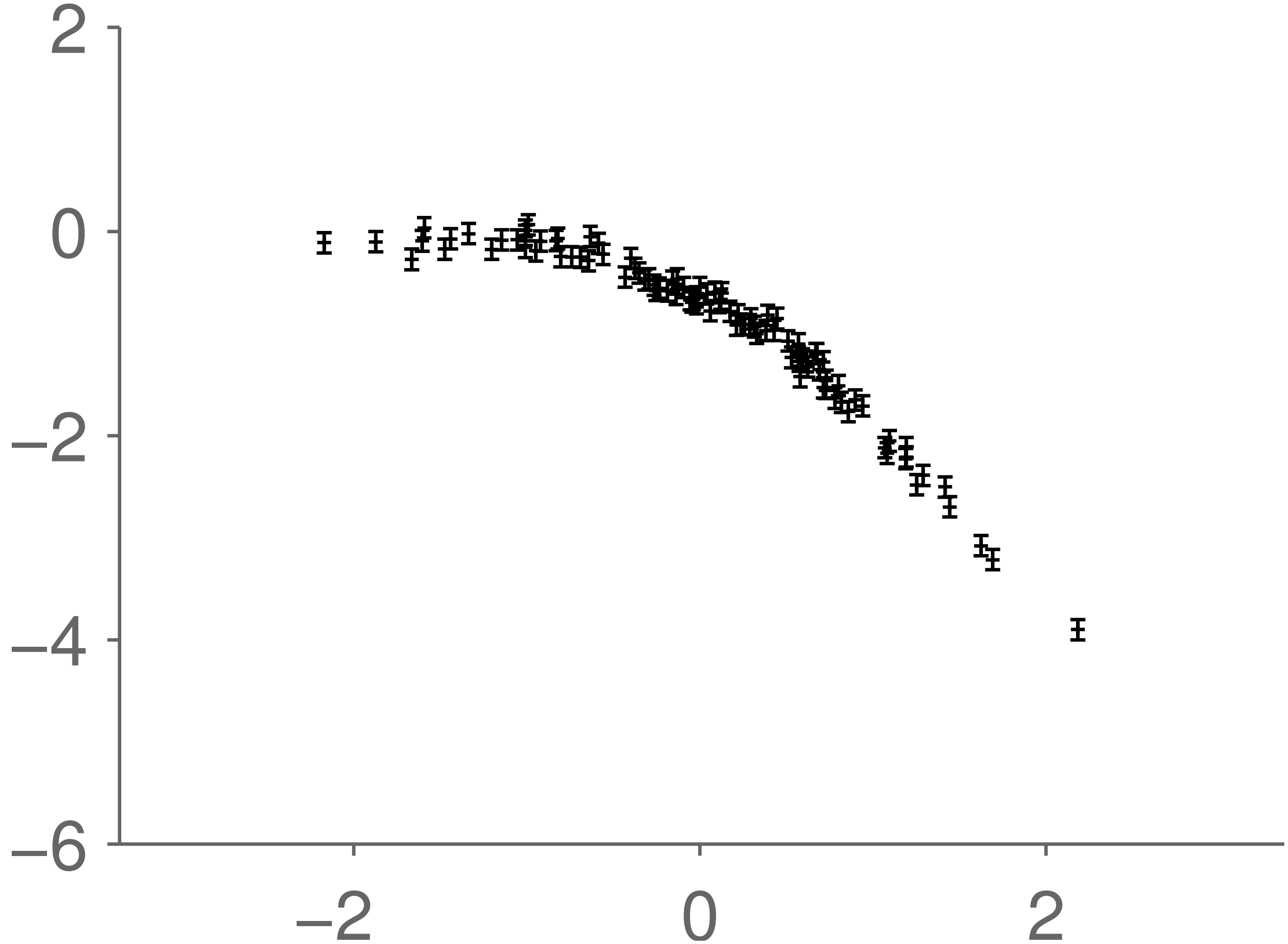
Posterior much more compact than prior.

Linear Regression: Posterior



Draws from posterior. Non-linear error envelope. Possible explanations linear.

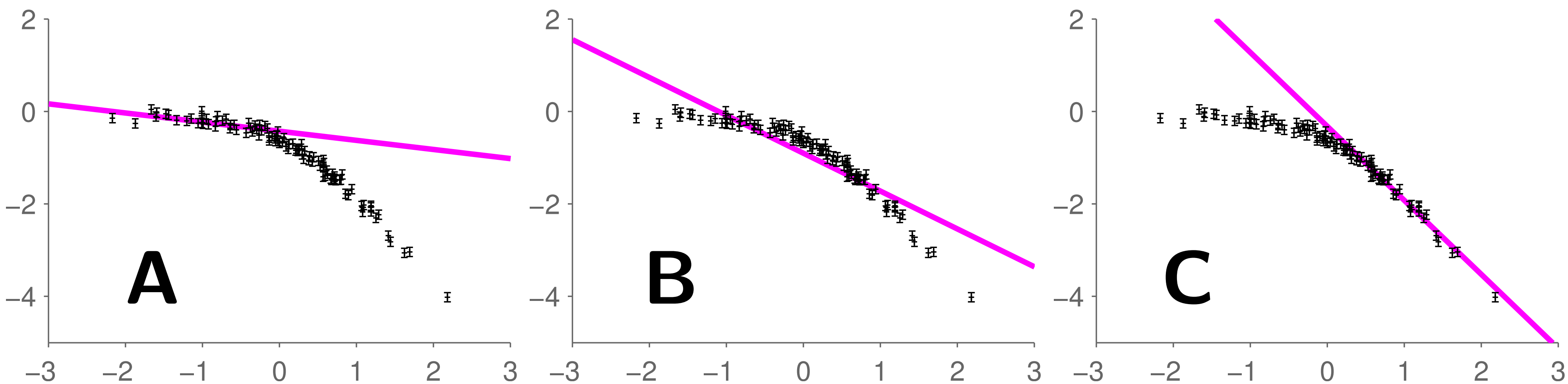
Model mismatch



What will Bayesian linear regression do?

Quiz

Given a (wrong) linear assumption, which explanations are typical of the posterior distribution?

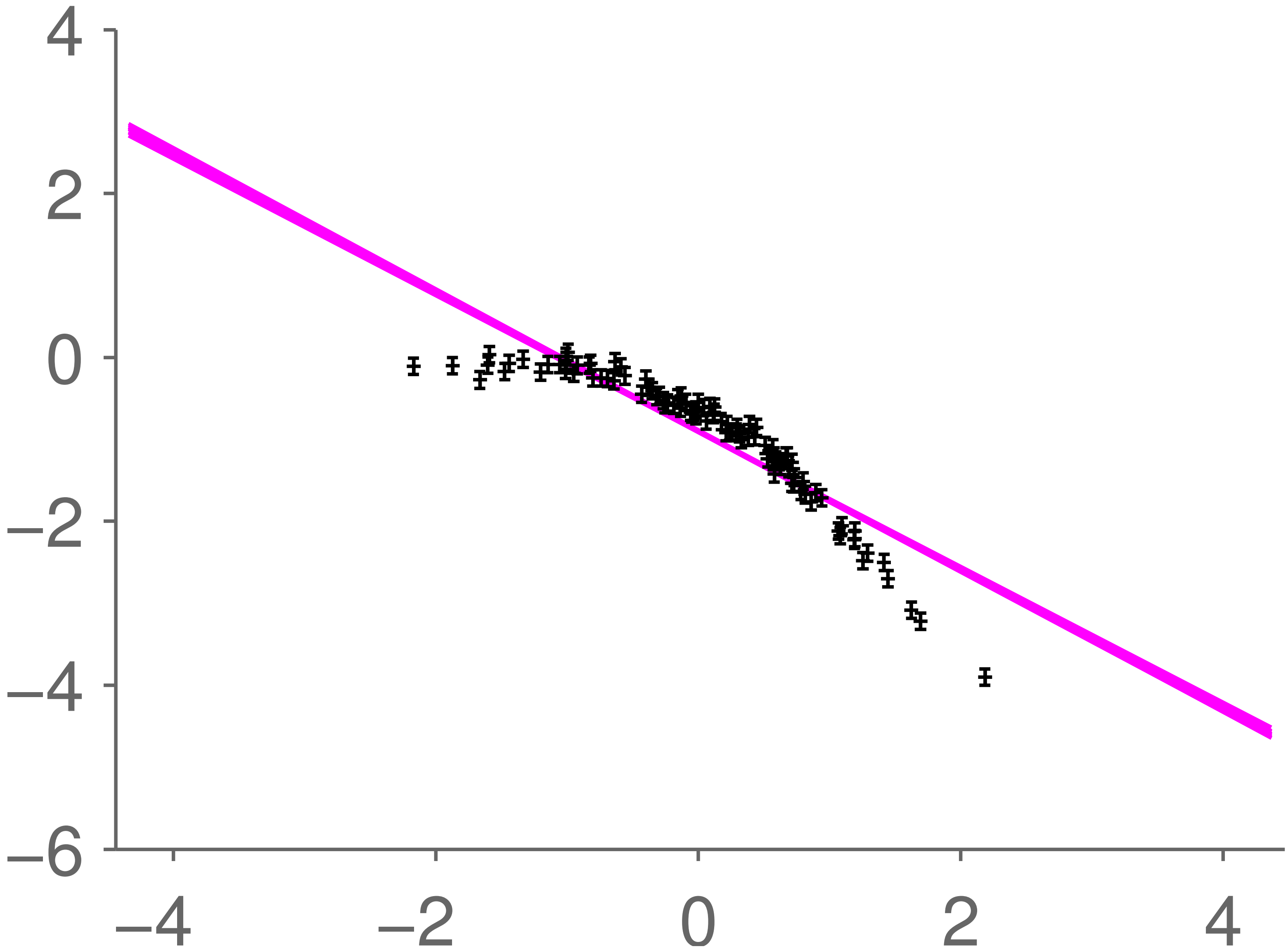


D All of the above

E None of the above

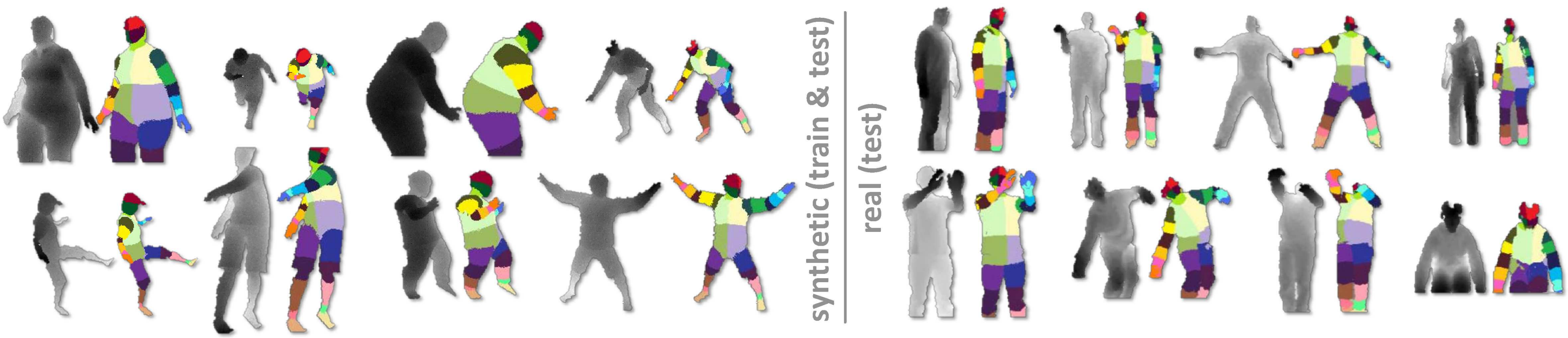
Z Not sure

‘Underfitting’



Posterior *very* certain despite blatant misfit. Prior ruled out truth.

Microsoft Kinect (Shotton et al., 2011)



Eyeball modelling assumptions

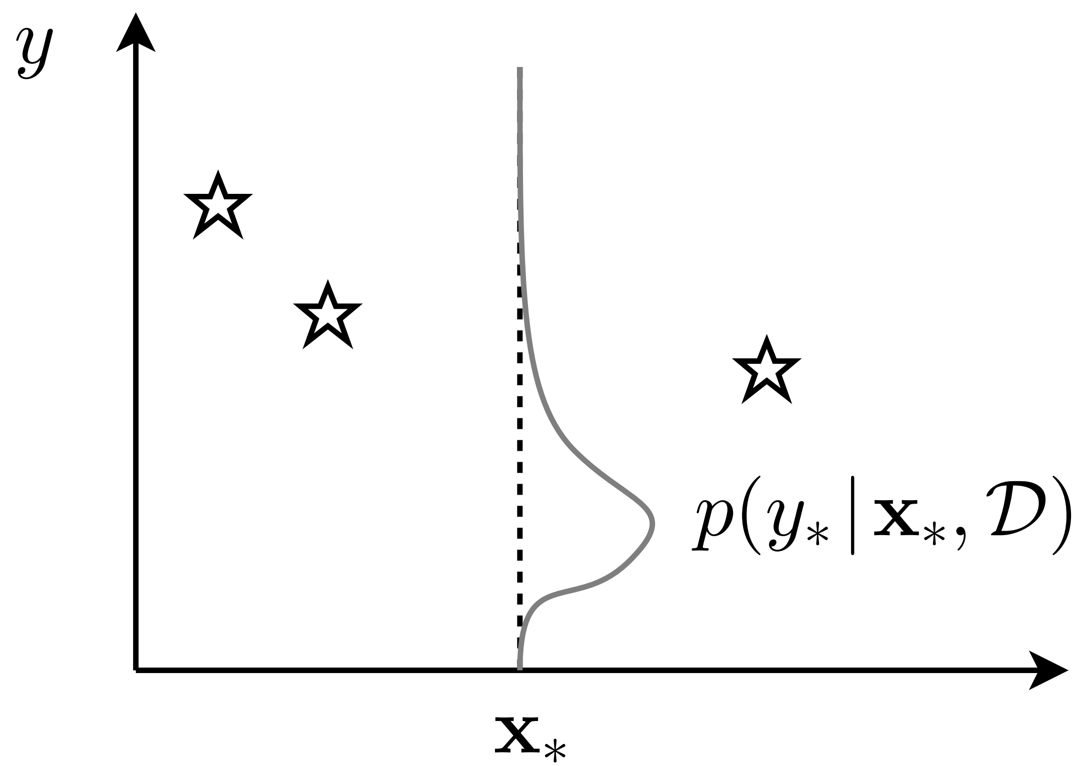
Generate training data

Random forest applied to fantasies

The need for integrals

$$p(y_* \,|\, \mathbf{x}_*, \mathcal{D}) = \int \mathrm{d}\theta \; p(y_*, \theta \,|\, \mathbf{x}_*, \mathcal{D})$$

$$= \int \mathrm{d}\theta \; p(y_* \,|\, \theta, \cancel{\mathcal{D}}) \, p(\theta \,|\, \cancel{\mathbf{x}_*}, \mathcal{D})$$



A statistical problem

What is the average height of the people in this room?
Method: measure our heights, add them up and divide by N.

What is the average height f of people p in Edinburgh $\mathcal{E}$?

$$E_{p\in\mathcal{E}}[f(p)] \equiv \frac{1}{|\mathcal{E}|}\sum_{p\in\mathcal{E}} f(p), \quad \text{“intractable”?}$$

$$\approx \frac{1}{S}\sum_{s=1}^{S} f\big(p^{(s)}\big), \quad \text{for random survey of } S \text{ people } \{p^{(s)}\} \in \mathcal{E}$$

Surveying works for large and notionally infinite populations.

Simple Monte Carlo

In general:

$$\int f(x) P(x) \, \mathrm{d}x \;\approx\; \frac{1}{S} \sum_{s=1}^{S} f(x^{(s)}), \quad x^{(s)} \sim P(x)$$

Example: making predictions

$$\begin{aligned} P(x \,|\, \mathcal{D}) \;&=\; \int P(x \,|\, \theta) \, p(\theta \,|\, \mathcal{D}) \, \mathrm{d}\theta \\ &\approx\; \frac{1}{S} \sum_{s=1}^{S} P(x \,|\, \theta^{(s)}), \quad \theta^{(s)} \sim p(\theta \,|\, \mathcal{D}) \end{aligned}$$

Many other integrals appear throughout statistical machine learning

Properties of Monte Carlo

Estimator: $\int f(x)\, P(x)\, \mathrm{d}x \;\approx\; \hat{f} \;\equiv\; \frac{1}{S}\sum_{s=1}^{S} f(x^{(s)}), \quad x^{(s)} \sim P(x)$

Estimator is unbiased:

$$\mathbb{E}_{P(\{x^{(s)}\})}\left[\hat{f}\right] \;=\; \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}_{P(x)}[f(x)] \;=\; \mathbb{E}_{P(x)}[f(x)]$$

Variance shrinks $\propto 1/S$:

$$\mathrm{var}_{P(\{x^{(s)}\})}\left[\hat{f}\right] \;=\; \frac{1}{S^2}\sum_{s=1}^{S}\mathrm{var}_{P(x)}[f(x)] \;=\; \mathrm{var}_{P(x)}[f(x)]\,/S$$

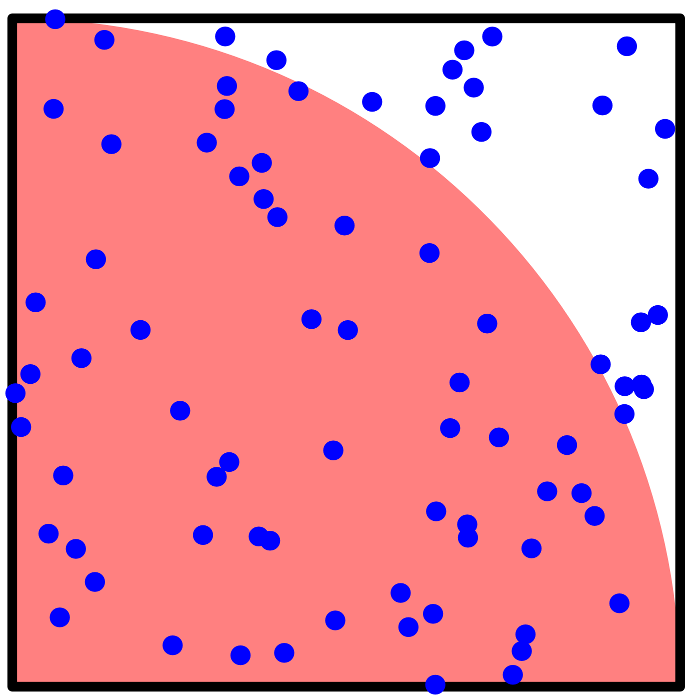
"Error bars" shrink like $\sqrt{S}$

Aside: don't always sample!

"Monte Carlo is an extremely bad method; it should be used only when all alternative methods are worse."

— Alan Sokal, 1996

A dumb approximation of π



$$P(x,y) = \begin{cases} 1 & 0 < x < 1 \ \text{and} \ 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\pi = 4 \iint \mathbb{I}\left((x^2 + y^2) < 1\right) P(x,y) \, \mathrm{d}x \, \mathrm{d}y$$

```
octave:1> S=12; a=rand(S,2); 4*mean(sum(a.*a,2)<1)
ans = 3.3333
octave:2> S=1e7; a=rand(S,2); 4*mean(sum(a.*a,2)<1)
ans = 3.1418
```

Alternatives to Monte Carlo

There are other methods of numerical integration!

Example: (nice) 1D integrals are easy:

```
octave:1> 4 * quadl(@(x) sqrt(1-x.^2), 0, 1, tolerance)
```

Gives π to 6 dp's in 108 evaluations, machine precision in 2598.

(NB Matlab's `quadl` fails at `tolerance=0`, but Octave works.)

In higher dimensions sometimes determinstic approximations work: Variational Bayes, Laplace, . . . (covered later)

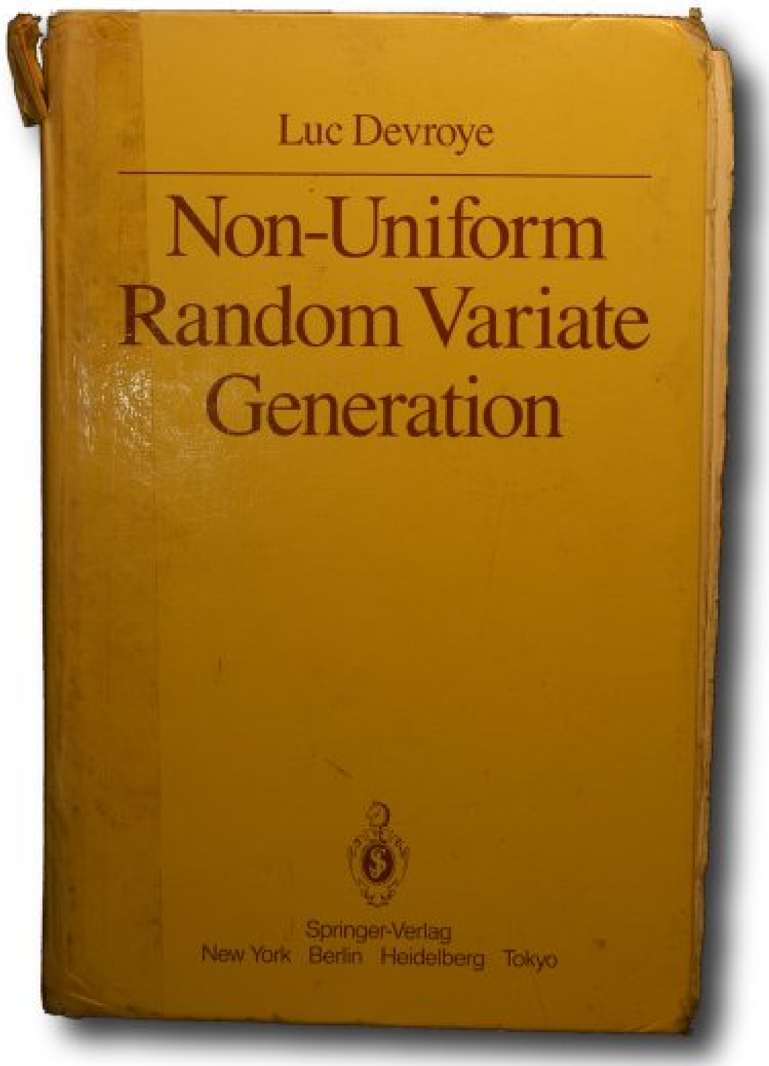
Reminder

Want to sample to approximate expectations:

$$\int f(x)P(x)\,\mathrm{d}x \;\approx\; \frac{1}{S}\sum_{s=1}^{S} f(x^{(s)}), \quad x^{(s)} \sim P(x)$$

How do we get the samples?

Sampling simple distributions



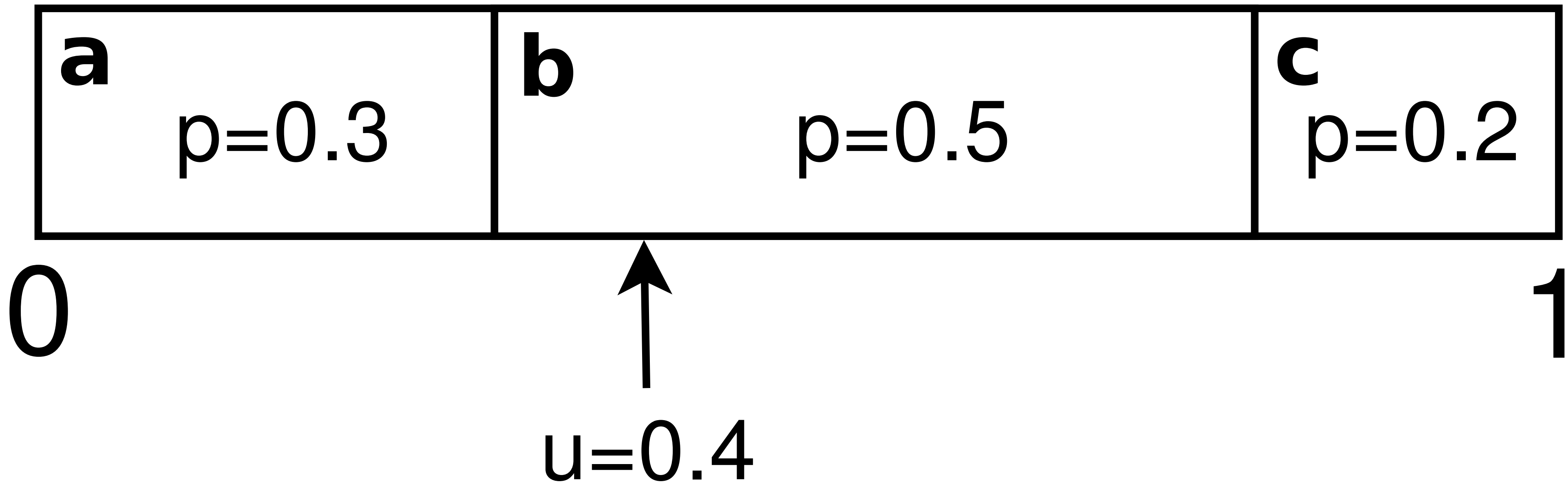
Use library routines for univariate distributions

(and some other special cases)

This book (free online) explains how some of them work

`http://cg.scs.carleton.ca/~luc/rnbookindex.html`

Sampling discrete values



$$u \sim \text{Uniform}[0,1]$$

$$u = 0.4 \quad \Rightarrow \quad x = \mathsf{b}$$

There are more efficient ways for large numbers of values and samples. See Devroye book.

Sampling from densities

How to convert samples from a Uniform[0,1] generator:

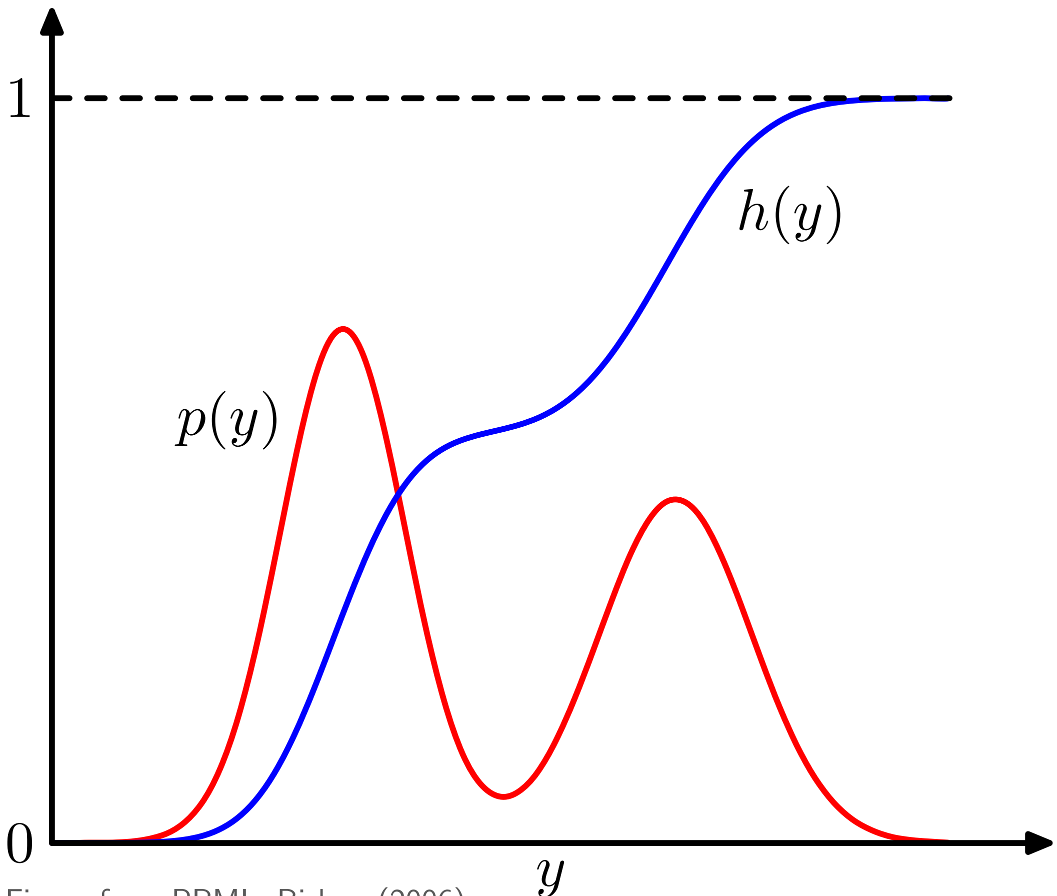


Figure from PRML, Bishop (2006)

$$h(y) = \int_{-\infty}^{y} p(y')\,\mathrm{d}y'$$

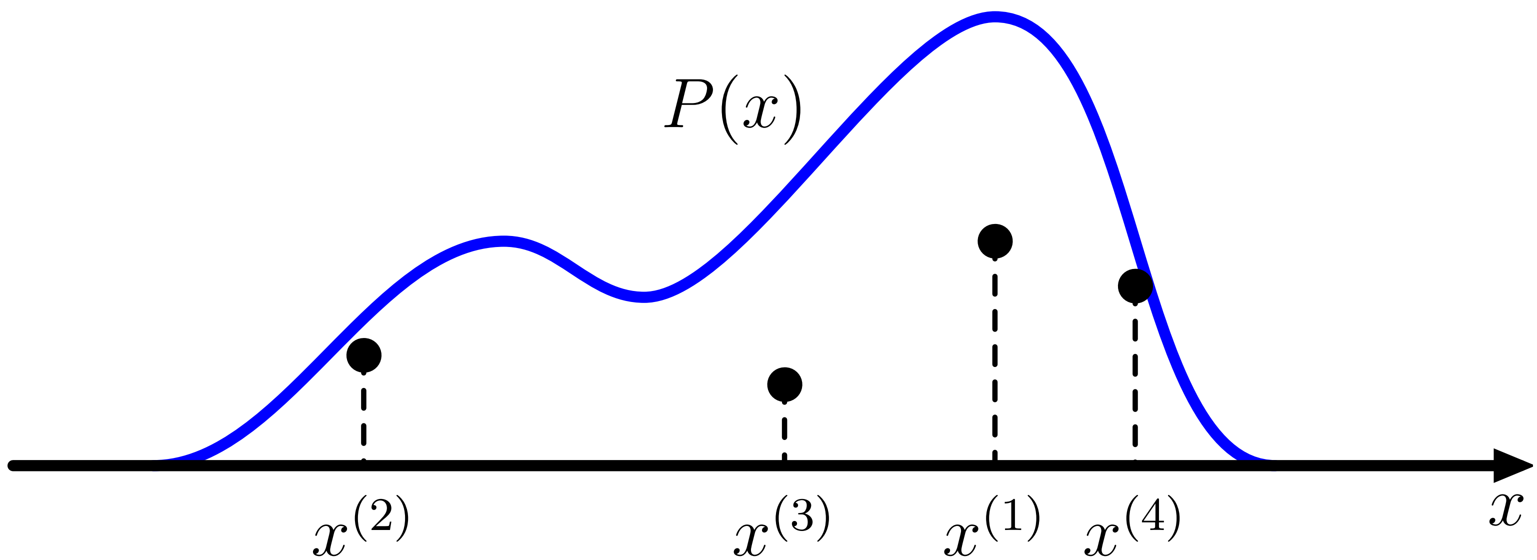
$u \sim$ Uniform[0,1]

Sample, $y(u) = h^{-1}(u)$

Although we can't always compute and invert $h(y)$

Sampling from densities

Draw points uniformly under the curve:



Probability mass to left of point $\sim$ Uniform[0,1]

Rejection sampling

Sampling from $\pi(x)$ using tractable $q(x)$:

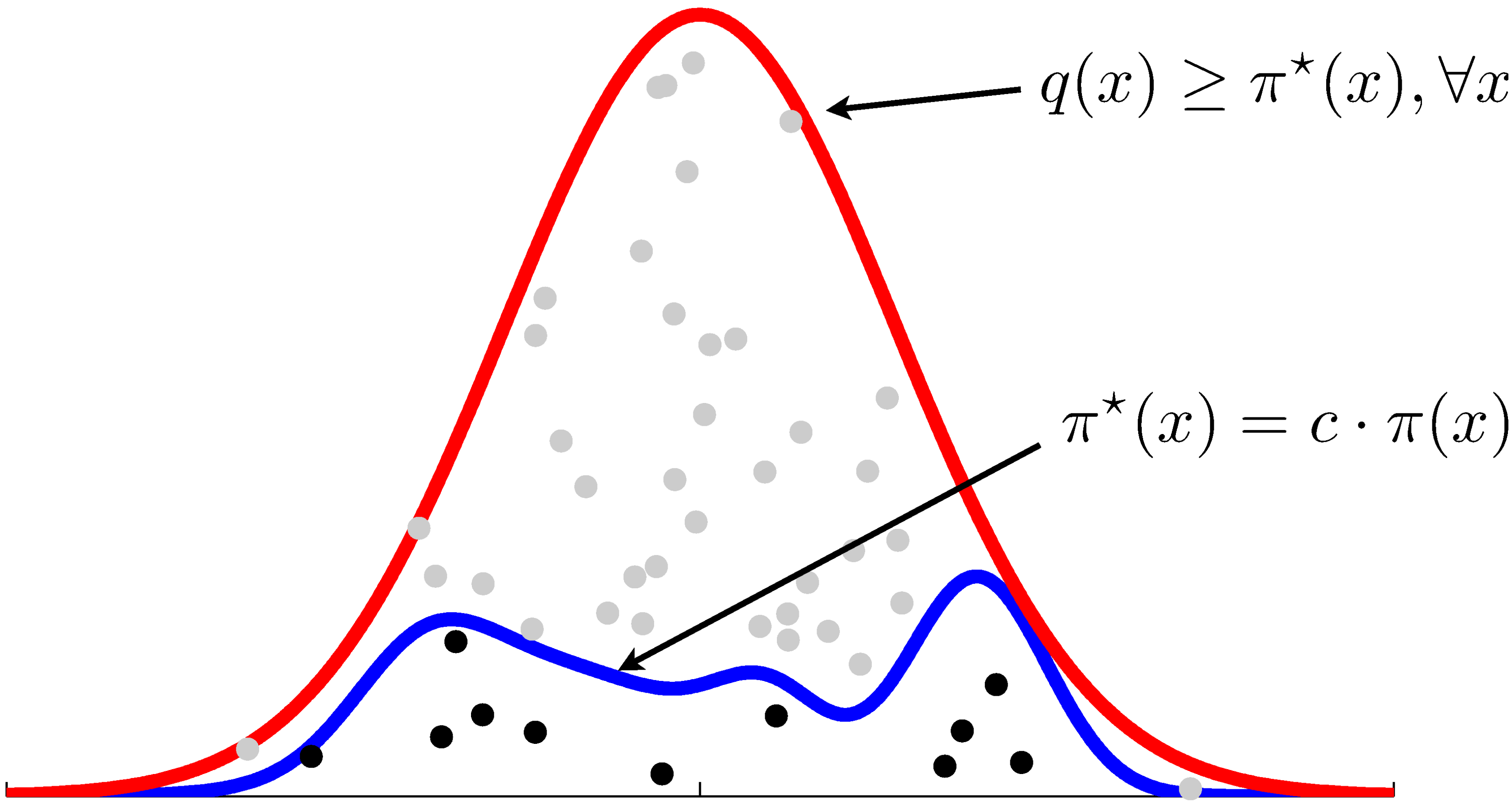


Figure credit: Ryan P. Adams

Importance sampling

Rewrite integral: expectation under simple distribution Q:

$$\int f(x)\, P(x)\, \mathrm{d}x \;=\; \int f(x)\, \frac{P(x)}{Q(x)}\, Q(x)\, \mathrm{d}x,$$

$$\approx\; \frac{1}{S} \sum_{s=1}^{S} f(x^{(s)})\, \frac{P(x^{(s)})}{Q(x^{(s)})}, \qquad x^{(s)} \sim Q(x)$$

Simple Monte Carlo applied to any integral.

Unbiased and independent of dimension?

Importance sampling (2)

If only know $P(x) = P^*(x)/\mathcal{Z}_P$ up to constant:

$$\int f(x)\, P(x)\, \mathrm{d}x \;\approx\; \frac{\mathcal{Z}_Q}{\mathcal{Z}_P}\, \frac{1}{S} \sum_{s=1}^{S} f(x^{(s)})\, \underbrace{\frac{P^*(x^{(s)})}{Q^*(x^{(s)})}}_{w^{*(s)}}, \quad x^{(s)} \sim Q(x)$$

$$\approx\; \cancel{\frac{1}{S}} \sum_{s=1}^{S} f(x^{(s)})\, \frac{w^{*(s)}}{\cancel{\frac{1}{S}} \sum_{s'} w^{*(s')}}$$

This estimator is **consistent** but **biased**

Exercise: Prove that $\mathcal{Z}_P/\mathcal{Z}_Q \approx \frac{1}{S}\sum_s w^{*(s)}$

Application to large problems

Approximations scale badly with dimensionality

Example: $P(x) = \mathcal{N}(0,\, \mathbb{I}), \quad Q(x) = \mathcal{N}(0,\, \sigma^2\mathbb{I})$

Rejection sampling:

Requires $\sigma \geq 1$. Fraction of proposals accepted $= \sigma^{-D}$

Importance sampling:

Var$[P(x)/Q(x)] \;=\; \left(\frac{\sigma^2}{2-1/\sigma^2}\right)^{D/2} - 1$

Infinite / undefined variance if $\sigma \leq 1/\sqrt{2}$

Summary so far

- **Monte Carlo**
  approximate expectations with a sample average
- **Rejection sampling**
  draw samples from complex distributions
- **Importance sampling**
  apply Monte Carlo to ‘any’ sum/integral

Next: High dimensional problems: MCMC