An Introduction to Bayesian Inference

Michael Betancourt

April 8, 2009

8.882

An Example

$$\mathbf{x} \sim \mathcal{N}(\mu, \sigma)$$

$$\{x_i\} = \{x_1, \ldots, x_m\}$$

$$\mu? \;\; \sigma?$$

The General Problem

$$\mathbf{x} \sim p\left(x_1, \ldots, x_m | \alpha_1, \ldots, \alpha_n\right)$$

$$\{x_i\} = \{x_1, \ldots, x_m\}$$

$$\alpha_1, \ldots, \alpha_n?$$

Estimators

$$\tilde{\alpha}_j = f\left(x_1, \ldots, x_m\right)$$

$$\lim_{m \to \infty} \tilde{\alpha}_j = \alpha_j$$

Estimators: An Example

$$\mathbf{x} \sim \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Estimators: An Example

$$\mathbf{x} \sim \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$\mu = \langle \mathbf{x} \rangle \qquad \sigma^2 = \langle \mathbf{x}^2 \rangle - \langle \mathbf{x} \rangle^2$$

Estimators: An Example

$$\mathbf{x} \sim \mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$\mu = \langle \mathbf{x} \rangle \qquad \sigma^2 = \langle \mathbf{x}^2 \rangle - \langle \mathbf{x} \rangle^2$$

$$\tilde{\mu} = \frac{\sum_i x_i}{m} \equiv \langle x_i \rangle \qquad \tilde{\sigma}^2 = \frac{\sum_i (x_i - \langle x \rangle)^2}{m-1}$$

Estimators: The Danger

$$\lim_{m \to \infty} \tilde{\alpha}_j = \alpha_j ?$$

Estimators: The Danger

$$\lim_{m \to \infty} \tilde{\alpha}_j = \alpha_j?$$

$$m \neq \infty!$$

Estimators: The Danger

$$\lim_{m\to\infty} \tilde{\alpha}_j = \alpha_j ?$$

$$m \neq \infty !$$

$$p\left(\tilde{\alpha}_j | \alpha_j, m\right)$$

Estimators: The Limitation

$$\mathbf{x} \sim \alpha_1 \exp\left[-\alpha_2 (x - \alpha_3)^3\right], x \in (0, \infty)$$

Maximum Likelihood

$$P\left(x_1,\ldots,x_m\right) = \prod_i p\left(x_i|\alpha_1,\ldots,\alpha_n\right) \equiv L$$

$$\alpha_i{}^{ML} = \arg\max L$$

Maximum Log Likelihood

$$logL = \sum_i \log p\left(x_i | \alpha_1, \ldots, \alpha_n\right)$$

$$\{\alpha_i\}^{ML} = \arg\max L = \arg\max logL$$

χ^2 Minimization

$$L \propto \prod_i \exp\left[-\left(\frac{x_i - f(x)}{\sigma_i}\right)^2\right] = \exp\left[-\sum_i \left(\frac{x_i - f(x)}{\sigma_i}\right)^2\right] = \exp\left[-\chi^2\right]$$

$$\{\alpha_j\}^{ML} = \arg\max L = \arg\min \chi^2 = \{\alpha_j\}^{\chi^2}$$

Maximum Likelihood: An Example

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

Maximum Likelihood: An Example

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

$$\frac{\partial L}{\partial \mu} = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right] \frac{\sum_i (x_i-\mu)}{\sigma^2}$$

Maximum Likelihood: An Example

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

$$\frac{\partial L}{\partial \mu} = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right] \frac{\sum_i (x_i-\mu)}{\sigma^2}$$

$$0 = \frac{\sum_i (x_i-\mu_{ML})}{\sigma_{ML}^2} = \frac{\sum_i x_i - m\mu_{ML}}{\sigma_{ML}^2}$$

Maximum Likelihood: An Example

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

$$\frac{\partial L}{\partial \mu} = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right] \frac{\sum_i (x_i-\mu)}{\sigma^2}$$

$$0 = \frac{\sum_i (x_i-\mu_{ML})}{\sigma_{ML}^2} = \frac{\sum_i x_i - m\mu_{ML}}{\sigma_{ML}^2}$$

$$\mu_{ML} = \frac{\sum_i x_i}{m}$$

Maximum Likelihood Biases

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

Maximum Likelihood Biases

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

$$\frac{\partial L}{\partial \sigma} = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right] \frac{1}{\sigma}\left(-m + \frac{\sum_i (x_i-\mu)^2}{\sigma^2}\right)$$

Maximum Likelihood Biases

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

$$\frac{\partial L}{\partial \sigma} = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right] \frac{1}{\sigma} \left(-m + \frac{\sum_i (x_i-\mu)^2}{\sigma^2}\right)$$

$$0 = -m + \frac{\sum_i (x_i-\mu_{ML})^2}{\sigma_{ML}^2} = -m + \frac{\sum_i (x_i-\langle x\rangle)^2}{\sigma_{ML}^2}$$

Maximum Likelihood Biases

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right]$$

$$\frac{\partial L}{\partial \sigma} = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left[-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\right] \frac{1}{\sigma}\left(-m + \frac{\sum_i (x_i-\mu)^2}{\sigma^2}\right)$$

$$0 = -m + \frac{\sum_i (x_i-\mu_{ML})^2}{\sigma^2_{ML}} = -m + \frac{\sum_i (x_i-\langle x\rangle)^2}{\sigma^2_{ML}}$$

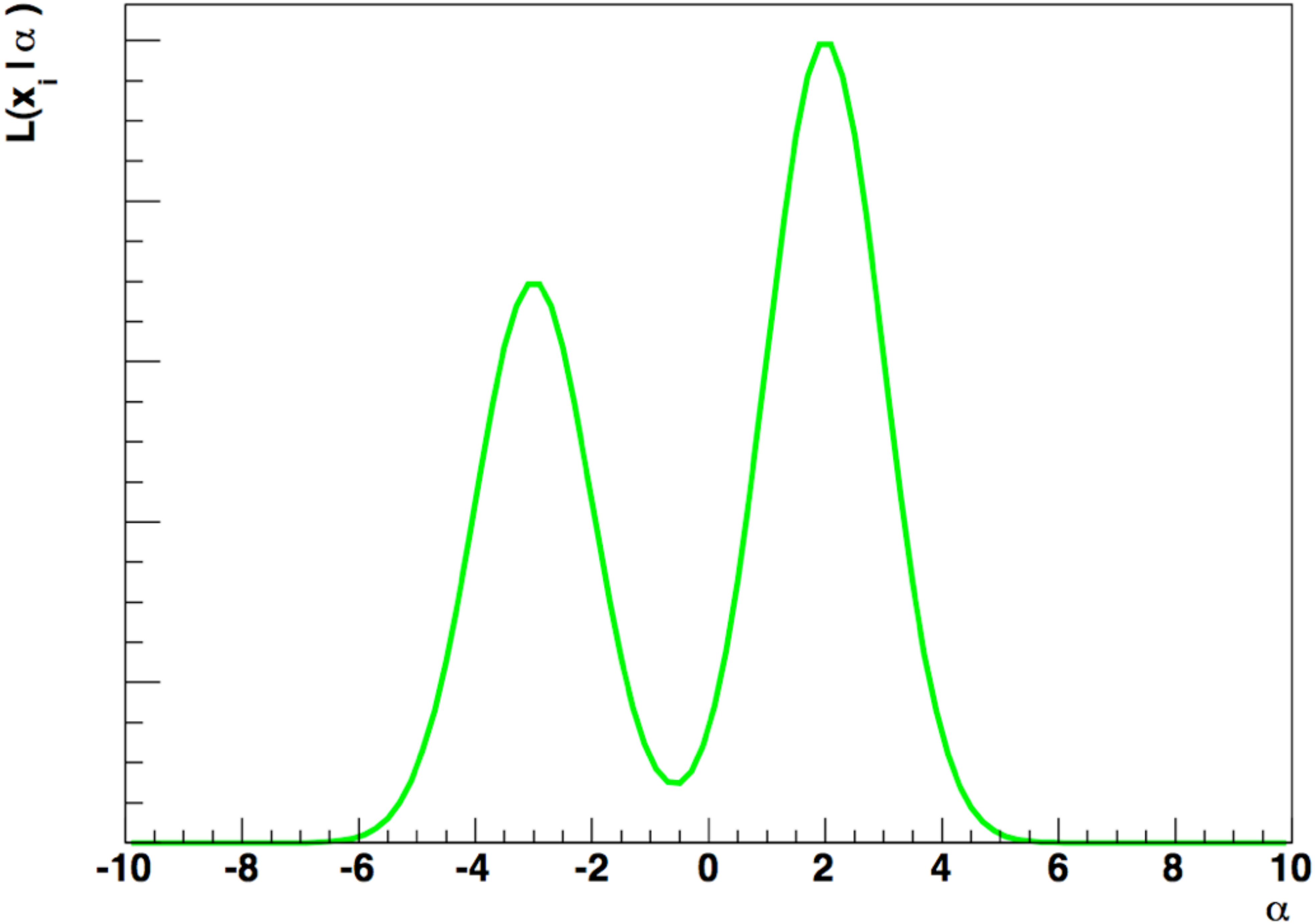
$$\sigma^2_{ML} = \frac{\sum_i (x_i-\langle x\rangle)^2}{m}$$

Maximum Likelihood Subtleties

$$L(\beta_1, \ldots, \beta_n) = L(\alpha_1, \ldots, \alpha_n) \left| \frac{\partial(\alpha_1, \ldots, \alpha_n)}{\partial(\beta_1, \ldots, \beta_n)} \right|$$

$$\{\beta_i\}^{ML} \neq \{f(\alpha_i)\}^{ML}$$

Bimodal Distributions



Back to the Drawing Board

$$p\left(\mathbf{b}\middle|\mathbf{a}\right)?$$

Back to the Drawing Board

$$p\left(\mathbf{b}|\mathbf{a}\right)?$$

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a},\mathbf{b}\right)}{p\left(\mathbf{a}\right)} \qquad p\left(\mathbf{a}|\mathbf{b}\right) = \frac{p\left(\mathbf{a},\mathbf{b}\right)}{p\left(\mathbf{b}\right)}$$

Back to the Drawing Board

$$p\left(\mathbf{b}|\mathbf{a}\right)?$$

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a},\mathbf{b}\right)}{p\left(\mathbf{a}\right)} \qquad p\left(\mathbf{a}|\mathbf{b}\right) = \frac{p\left(\mathbf{a},\mathbf{b}\right)}{p\left(\mathbf{b}\right)}$$

$$p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right) = p\left(\mathbf{a},\mathbf{b}\right) = p\left(\mathbf{b}|\mathbf{a}\right) p\left(\mathbf{a}\right)$$

Back to the Drawing Board

$$p\left(\mathbf{b}|\mathbf{a}\right)?$$

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a},\mathbf{b}\right)}{p\left(\mathbf{a}\right)} \qquad p\left(\mathbf{a}|\mathbf{b}\right) = \frac{p\left(\mathbf{a},\mathbf{b}\right)}{p\left(\mathbf{b}\right)}$$

$$p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right) = p\left(\mathbf{a},\mathbf{b}\right) = p\left(\mathbf{b}|\mathbf{a}\right) p\left(\mathbf{a}\right)$$

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right)}{p\left(\mathbf{a}\right)} = \frac{p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right)}{\int \mathrm{d}\mathbf{b}\, p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right)}$$

Bayes Theorem

$$p\left(\mathbf{b}\middle|\mathbf{a}\right) = \frac{p\left(\mathbf{a}\middle|\mathbf{b}\right) p\left(\mathbf{b}\right)}{\int \mathrm{d}\mathbf{b}\, p\left(\mathbf{a}\middle|\mathbf{b}\right) p\left(\mathbf{b}\right)}$$

Bayes Theorem

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a}|\mathbf{b}\right)p\left(\mathbf{b}\right)}{\int \mathrm{d}\mathbf{b}\, p\left(\mathbf{a}|\mathbf{b}\right)p\left(\mathbf{b}\right)}$$

Bayes Theorem

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right)}{\int \mathrm{d}\mathbf{b}\, p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right)}$$

Bayes Theorem

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right)}{\int \mathrm{d}\mathbf{b}\, p\left(\mathbf{a}|\mathbf{b}\right) p\left(\mathbf{b}\right)}$$

Bayes Theorem

$$p\left(\mathbf{b}|\mathbf{a}\right) = \frac{p\left(\mathbf{a}|\mathbf{b}\right)p\left(\mathbf{b}\right)}{\int \mathrm{d}\mathbf{b}\, p\left(\mathbf{a}|\mathbf{b}\right)p\left(\mathbf{b}\right)}$$

A New Approach?

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right) = \frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

A New Approach?

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right) = \frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

The $\{\alpha_j\}$ are not random variables!

Frequentist Statistics

$$P\left(x_i\right) = \frac{N\left(x_i\right)}{N} \rightarrow \frac{\omega\left(x_i\right)}{\Omega}$$

Uncertainty of Belief

$$P(\text{Rain Tomorrow})$$

Uncertainty of Belief

$$P\left(\text{Rain Tomorrow}\right)$$

$$P\left(\text{Tevatron Finds The Higgs}\right)$$

Uncertainty of Belief

$$P(\text{Rain Tomorrow})$$

$$P(\text{Tevatron Finds The Higgs})$$

$$P(\text{Graduate})$$

The Cox Axioms

- Axiom One

$$P\left(\bar{A}|B\right) = f\left(P\left(A|B\right)\right)$$

- Axiom Two

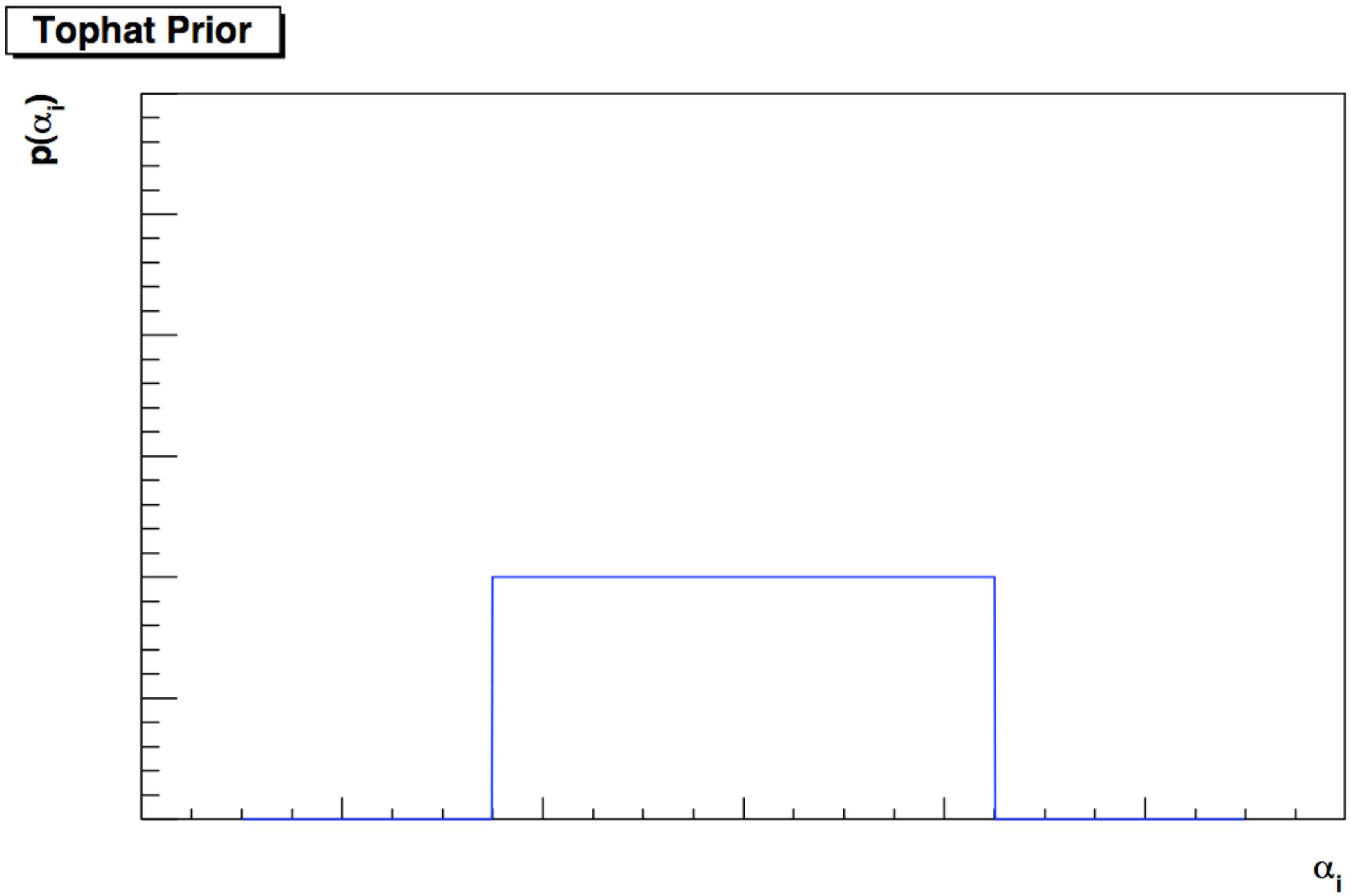
$$P\left(A_1 A_2|B\right) = g\left(P\left(A_1|B\right), P\left(A_2|A_1 B\right)\right)$$

A New Approach!

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)=\frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

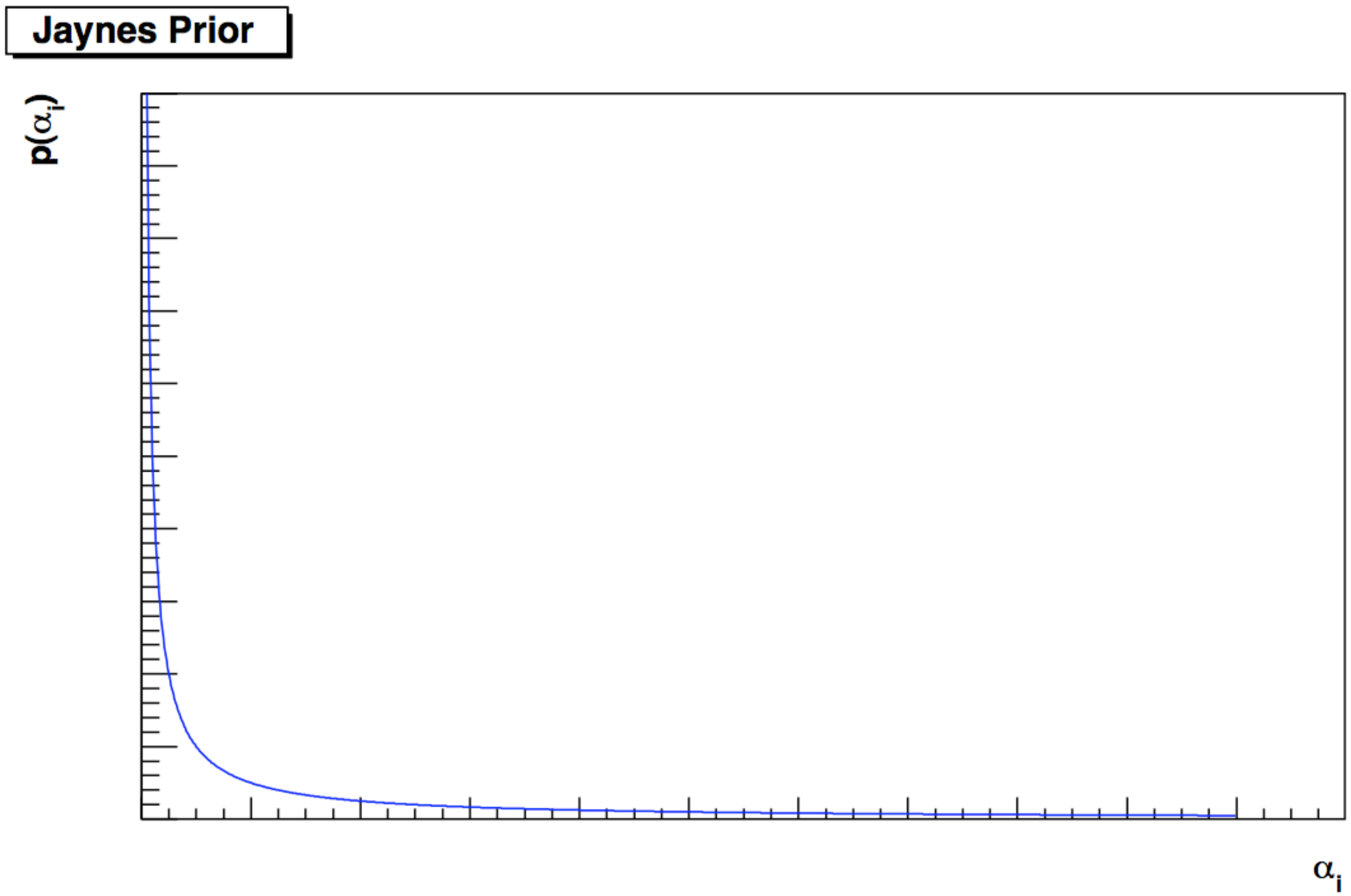
Selecting Priors

$$\alpha_i \in (-\infty, \infty), \;\; p(\alpha_i) = \frac{1}{\sigma_{\alpha_i}}$$



Selecting Priors

$$\alpha_i \in (0,\infty),\ \ p(\alpha_i) = \frac{1}{\alpha_i}$$



An Example

$$p(x|\alpha) = \alpha \exp[-\alpha x]$$

An Example

$$p(x|\alpha) = \alpha \exp[-\alpha x]$$

$$p(\alpha|x_1,\ldots,x_m) = \frac{p(x_1,\ldots,x_m|\alpha)\,p(\alpha)}{p(x_1,\ldots,x_m)} = \frac{\alpha^m \exp[-\alpha \sum_i x_i]\,p(\alpha)}{p(x_1,\ldots,x_m)}$$

An Example

$$p(x|\alpha) = \alpha \exp[-\alpha x]$$

$$p(\alpha|x_1,\ldots,x_m) = \frac{p(x_1,\ldots,x_m|\alpha)\,p(\alpha)}{p(x_1,\ldots,x_m)} = \frac{\alpha^m \exp[-\alpha \sum_i x_i]\,p(\alpha)}{p(x_1,\ldots,x_m)}$$

$$p(\alpha|x_1,\ldots,x_m) = \frac{\alpha^m \exp[-\alpha \sum_i x_i]\,1/\alpha}{p(x_1,\ldots,x_m)}$$

An Example

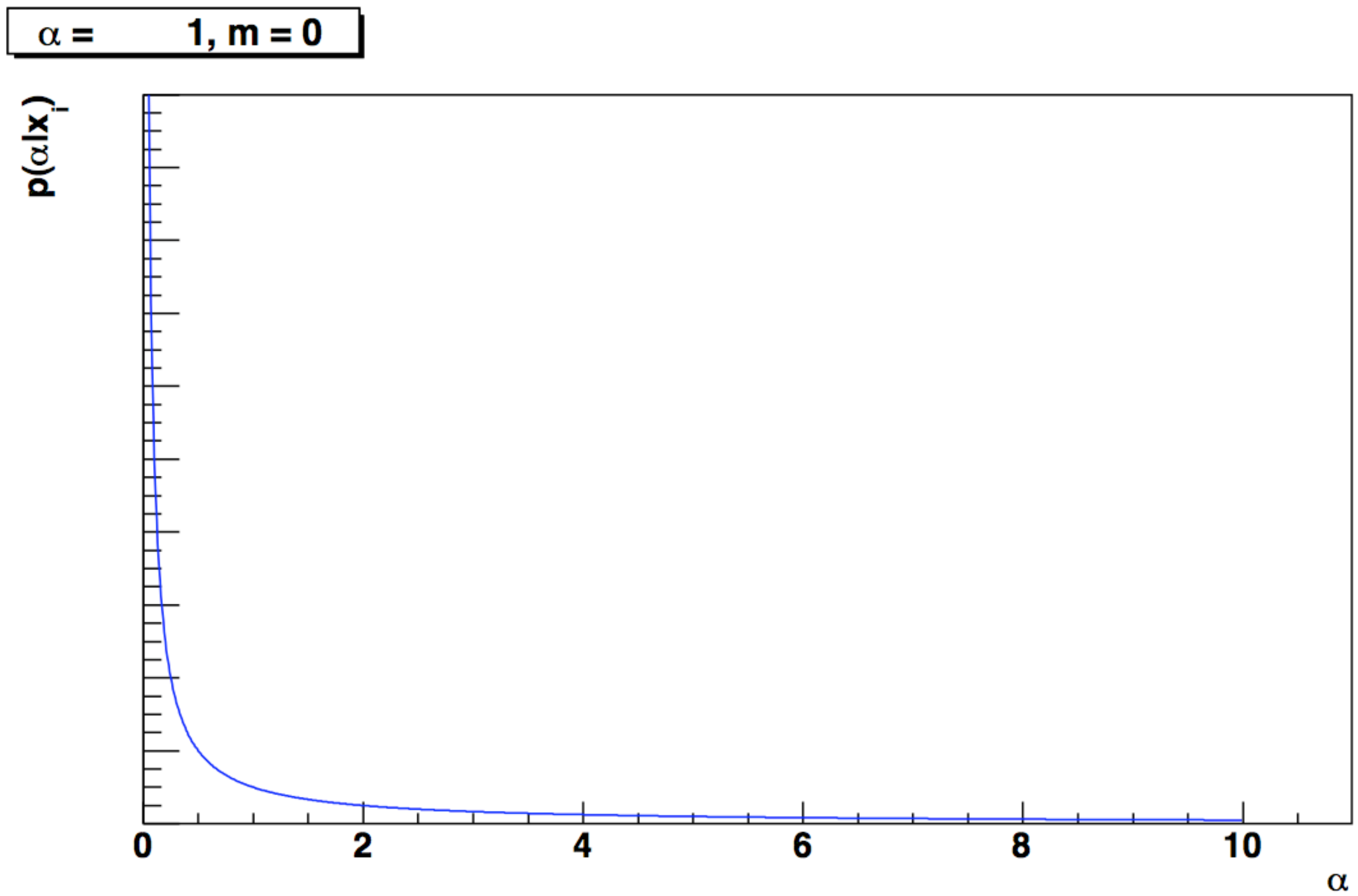
$$p(x|\alpha) = \alpha \exp[-\alpha x]$$

$$p(\alpha|x_1,\ldots,x_m) = \frac{p(x_1,\ldots,x_m|\alpha)\, p(\alpha)}{p(x_1,\ldots,x_m)} = \frac{\alpha^m \exp[-\alpha \sum_i x_i]\, p(\alpha)}{p(x_1,\ldots,x_m)}$$

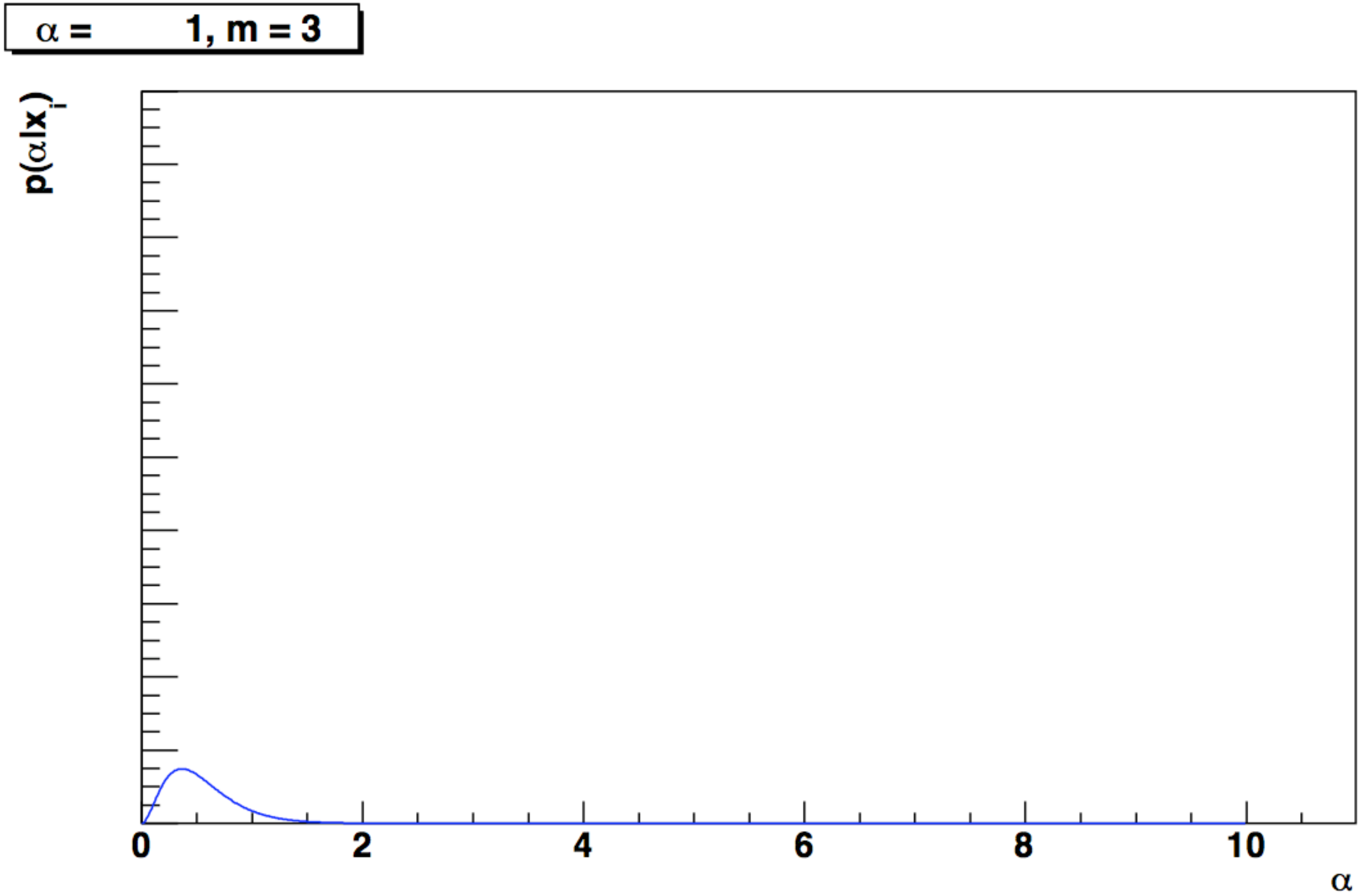
$$p(\alpha|x_1,\ldots,x_m) = \frac{\alpha^m \exp[-\alpha \sum_i x_i]\, 1/\alpha}{p(x_1,\ldots,x_m)}$$

$$p(\alpha|x_1,\ldots,x_m) = \frac{(\sum_i x_i)^m}{(m-1)!}\alpha^{m-1} \exp\left[-\alpha \sum_i x_i\right]$$

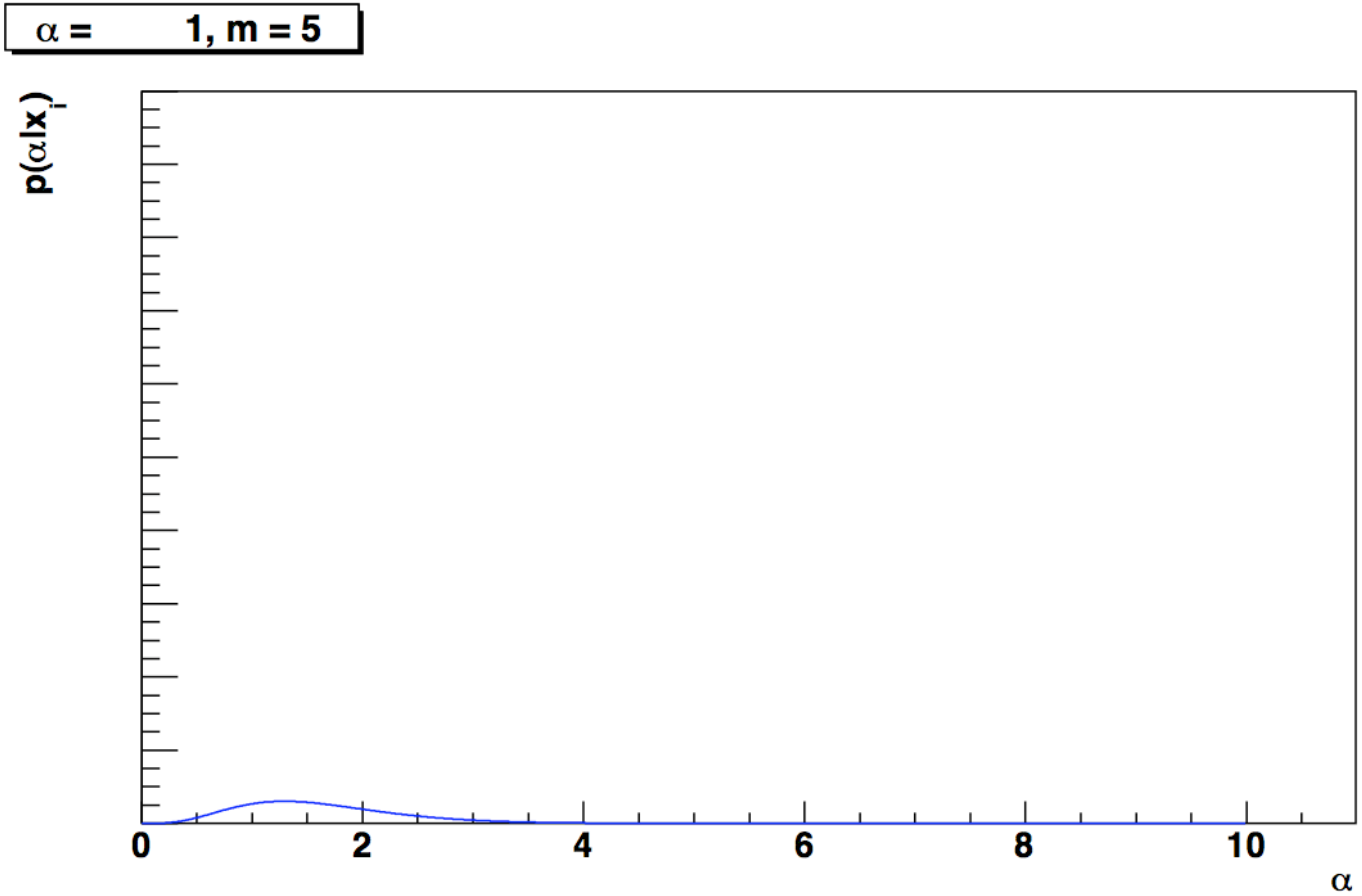
An Example



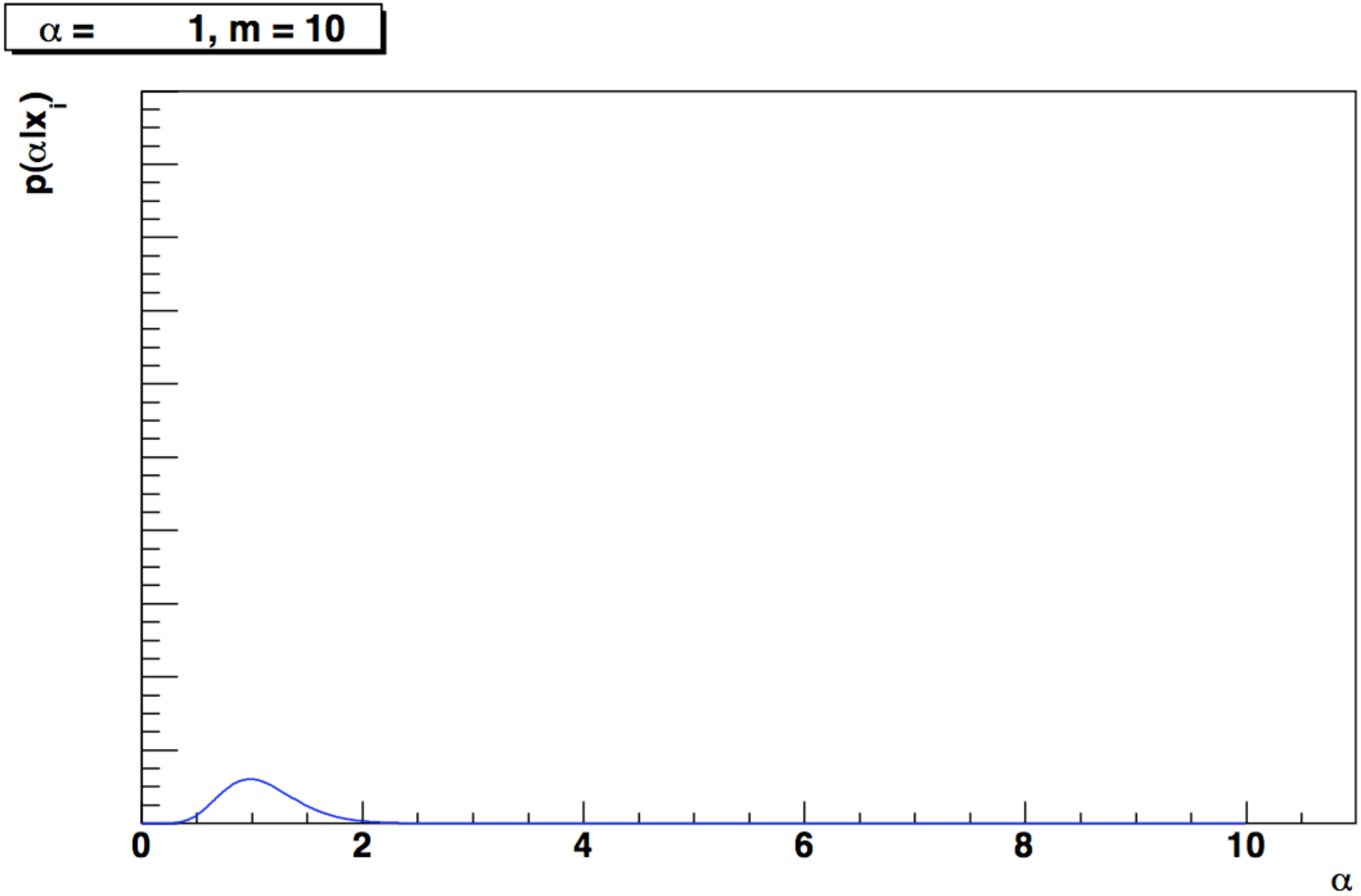
An Example



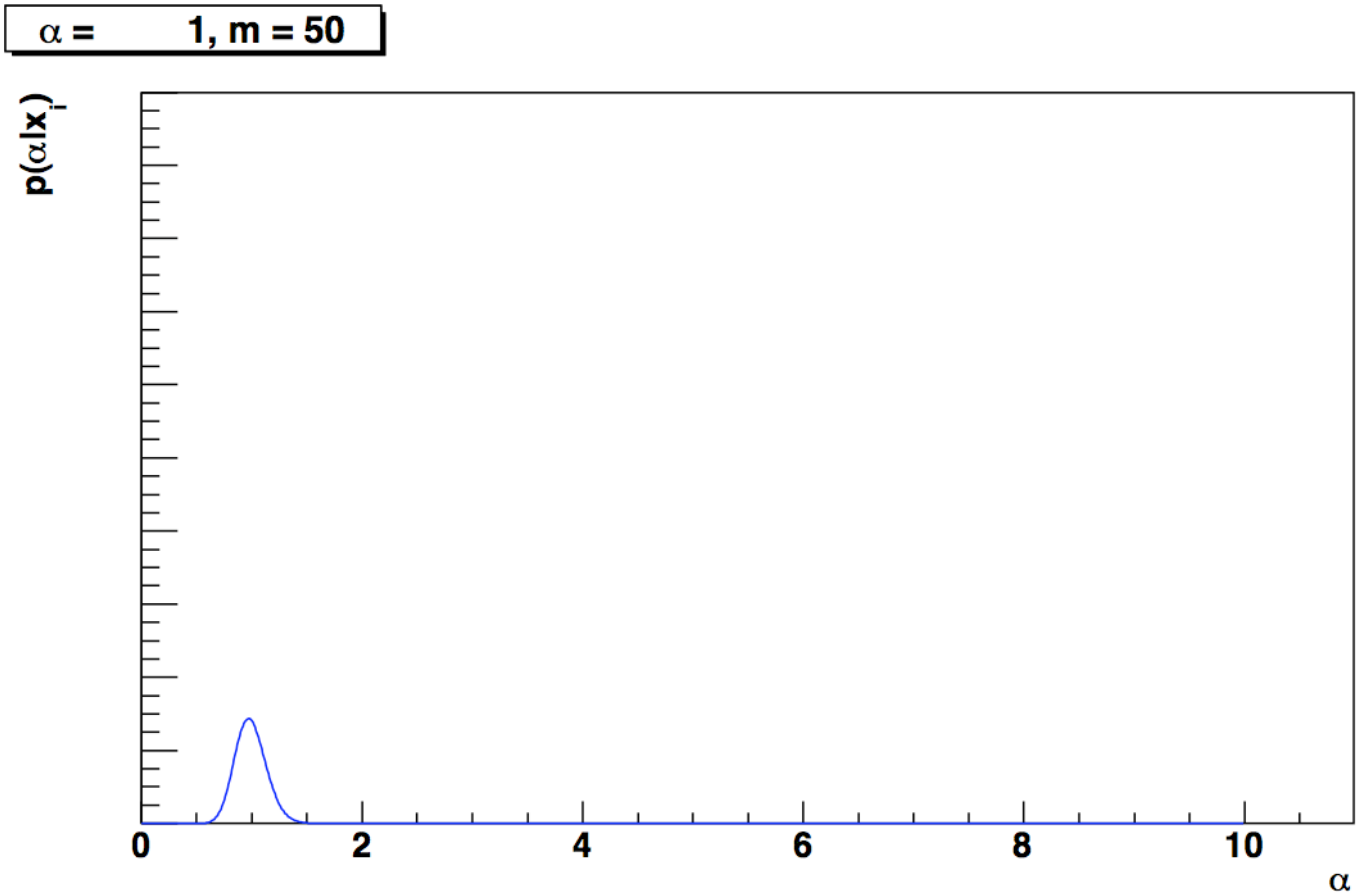
An Example



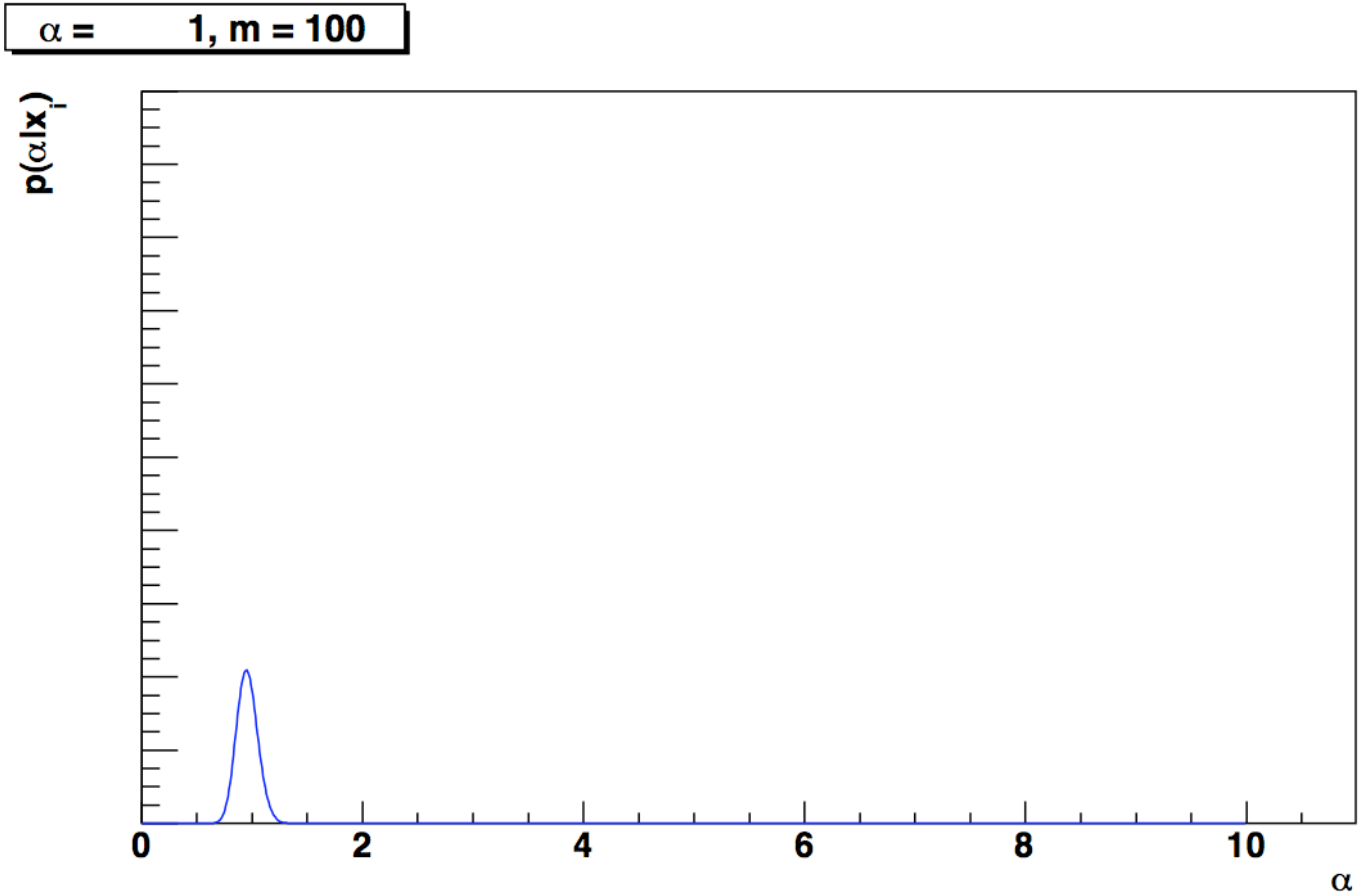
An Example



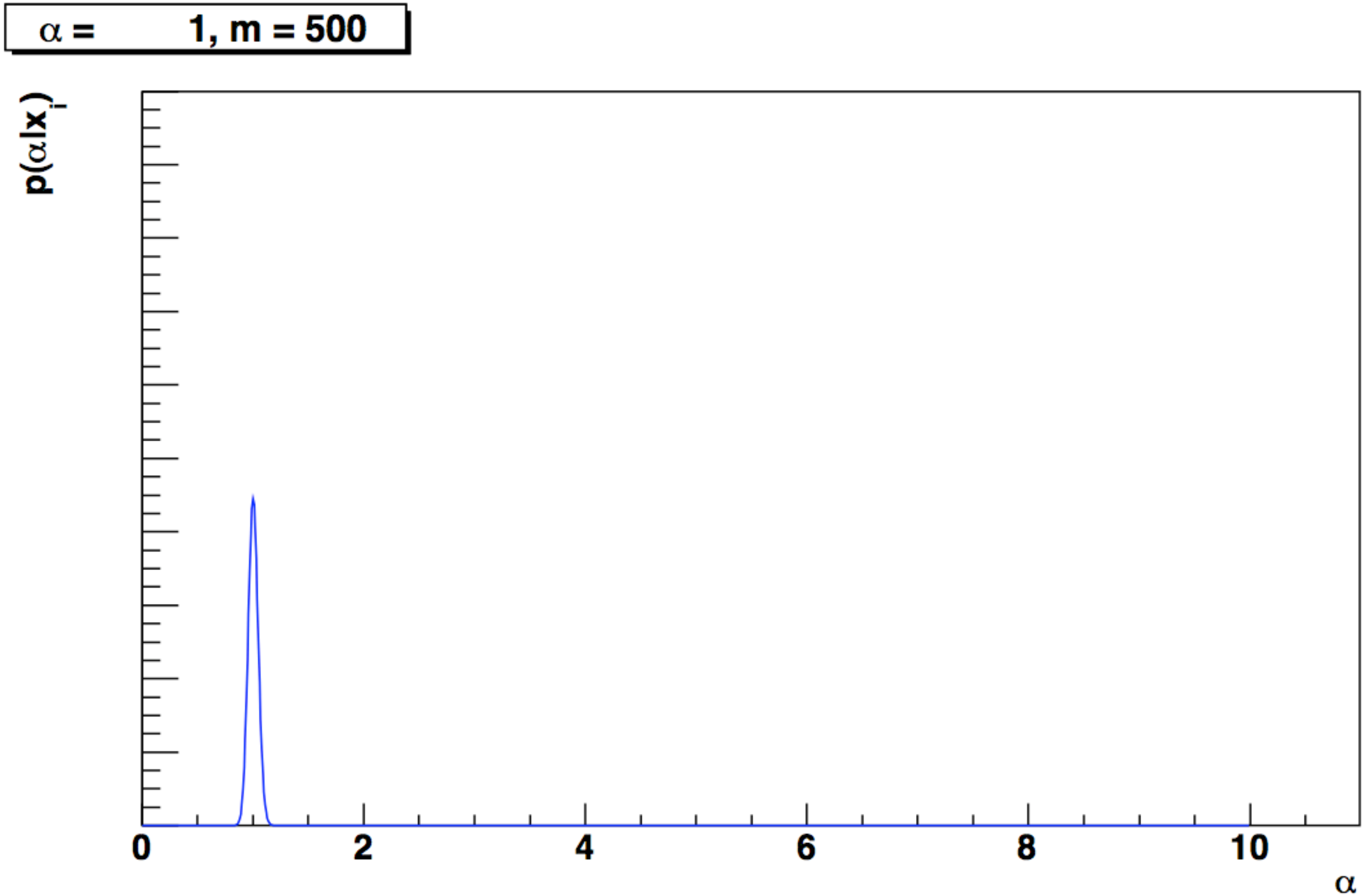
An Example



An Example



An Example



An Example Revisted

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)=\frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

An Example Revisted

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)=\frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

$$p\left(\mu,\sigma|x_1,\ldots,x_m\right)=\frac{L\left(x_1,\ldots,x_m|\mu,\sigma\right)p\left(\mu\right)p\left(\sigma\right)}{p\left(x_1,\ldots,x_m\right)}$$

An Example Revisted

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)=\frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

$$p\left(\mu,\sigma|x_1,\ldots,x_m\right)=\frac{L\left(x_1,\ldots,x_m|\mu,\sigma\right)p\left(\mu\right)p\left(\sigma\right)}{p\left(x_1,\ldots,x_m\right)}$$

$$p\left(\mu,\sigma|x_1,\ldots,x_m\right)\propto\left(2\pi\sigma^2\right)^{-\frac{m}{2}}\exp\left[-\frac{\sum_i\left(x_i-\mu\right)^2}{2\sigma^2}\right]\frac{1}{\sigma_\mu}\frac{1}{\sigma}$$

Marginalization

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)=\frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

$$p\left(\alpha_i\right)=\int \mathrm{d}^{n-1}\alpha\, p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)$$

Marginalization: An Example

$$p(\sigma|x_1,\ldots,x_m) = \int \mathrm{d}\mu\, p(\mu,\sigma|x_1,\ldots,x_m)$$

Marginalization: An Example

$$p\left(\sigma|x_1,\ldots,x_m\right) = \int \mathrm{d}\mu p\left(\mu,\sigma|x_1,\ldots,x_m\right)$$

$$p\left(\sigma|x_1,\ldots,x_m\right) = \int \mathrm{d}\mu \frac{L\left(x_1,\ldots,x_m|\mu,\sigma\right) p\left(\mu\right) p\left(\sigma\right)}{p\left(x_1,\ldots,x_m\right)}$$

Marginalization: An Example

$$p(\sigma|x_1,\ldots,x_m) = \int \mathrm{d}\mu\, p(\mu,\sigma|x_1,\ldots,x_m)$$

$$p(\sigma|x_1,\ldots,x_m) = \int \mathrm{d}\mu \frac{L(x_1,\ldots,x_m|\mu,\sigma)\, p(\mu)\, p(\sigma)}{p(x_1,\ldots,x_m)}$$

$$p(\sigma|x_1,\ldots,x_m) = \frac{p(\sigma)}{p(x_1,\ldots,x_m)} \int \mathrm{d}\mu\, L(x_1,\ldots,x_m|\mu,\sigma)\, p(\mu)$$

Marginalization: An Example

$$p\left(\sigma|x_1,\ldots,x_m\right)=\int \mathrm{d}\mu p\left(\mu,\sigma|x_1,\ldots,x_m\right)$$

$$p\left(\sigma|x_1,\ldots,x_m\right)=\int \mathrm{d}\mu \frac{L\left(x_1,\ldots,x_m|\mu,\sigma\right)p\left(\mu\right)p\left(\sigma\right)}{p\left(x_1,\ldots,x_m\right)}$$

$$p\left(\sigma|x_1,\ldots,x_m\right)=\frac{p\left(\sigma\right)}{p\left(x_1,\ldots,x_m\right)}\int \mathrm{d}\mu L\left(x_1,\ldots,x_m|\mu,\sigma\right)p\left(\mu\right)$$

$$p\left(\sigma|x_1,\ldots,x_m\right)=\frac{p\left(\sigma\right)}{p\left(x_1,\ldots,x_m\right)}\left(2\pi\sigma^2\right)^{-\frac{m}{2}}\frac{1}{\sigma_\mu}\sqrt{\frac{2\pi\sigma^2}{m}}\exp\left[-\frac{\sum_i\left(x_i-\langle x\rangle\right)^2}{2\sigma^2}\right]$$

Marginalization: An Example

$$p(\sigma|x_1,\ldots,x_m) = \frac{p(\sigma)}{p(x_1,\ldots,x_m)}\left(2\pi\sigma^2\right)^{-\frac{m}{2}}\frac{1}{\sigma_\mu}\sqrt{\frac{2\pi\sigma^2}{m}}\exp\left[-\frac{\sum_i\left(x_i-\langle x\rangle\right)^2}{2\sigma^2}\right]$$

Marginalization: An Example

$$p(\sigma|x_1,\ldots,x_m) = \frac{p(\sigma)}{p(x_1,\ldots,x_m)}\left(2\pi\sigma^2\right)^{-\frac{m}{2}}\frac{1}{\sigma_\mu}\sqrt{\frac{2\pi\sigma^2}{m}}\exp\left[-\frac{\sum_i\left(x_i-\langle x\rangle\right)^2}{2\sigma^2}\right]$$

$$p(\sigma|x_1,\ldots,x_m) = \frac{1/\sigma}{p(x_1,\ldots,x_m)}\left(2\pi\sigma^2\right)^{-\frac{m}{2}}\frac{1}{\sigma_\mu}\sqrt{\frac{2\pi\sigma^2}{m}}\exp\left[-\frac{\sum_i\left(x_i-\langle x\rangle\right)^2}{2\sigma^2}\right]$$

Marginalization: An Example

$$p\left(\sigma|x_1,\ldots,x_m\right)=\frac{p\left(\sigma\right)}{p\left(x_1,\ldots,x_m\right)}\left(2\pi\sigma^2\right)^{-\frac{m}{2}}\frac{1}{\sigma_\mu}\sqrt{\frac{2\pi\sigma^2}{m}}\exp\left[-\frac{\sum_i\left(x_i-\langle x\rangle\right)^2}{2\sigma^2}\right]$$

$$p\left(\sigma|x_1,\ldots,x_m\right)=\frac{1/\sigma}{p\left(x_1,\ldots,x_m\right)}\left(2\pi\sigma^2\right)^{-\frac{m}{2}}\frac{1}{\sigma_\mu}\sqrt{\frac{2\pi\sigma^2}{m}}\exp\left[-\frac{\sum_i\left(x_i-\langle x\rangle\right)^2}{2\sigma^2}\right]$$

$$\arg\max p\left(\sigma|x_1,\ldots,x_m\right)=\sqrt{\frac{\sum_i\left(x_i-\langle x\rangle\right)^2}{m-1}}$$

Frequentist vs. Bayesian

$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right) = \frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

Frequentist vs. Bayesian

$$p(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m) = \frac{\overbrace{p(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n)}^{L!}\, p(\alpha_1,\ldots,\alpha_n)}{p(x_1,\ldots,x_m)}$$

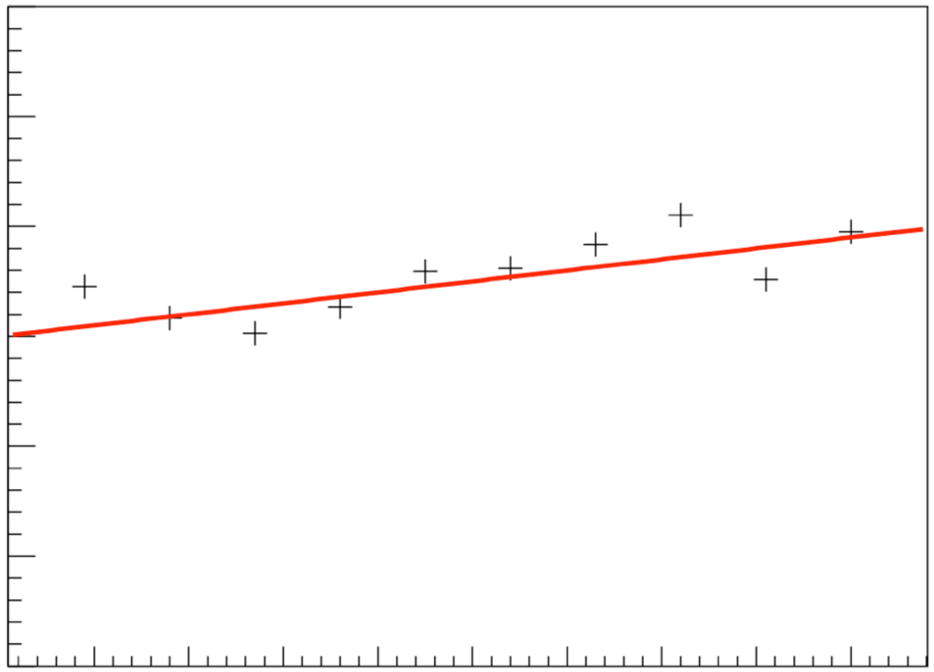
Bayesian Inference in Practice

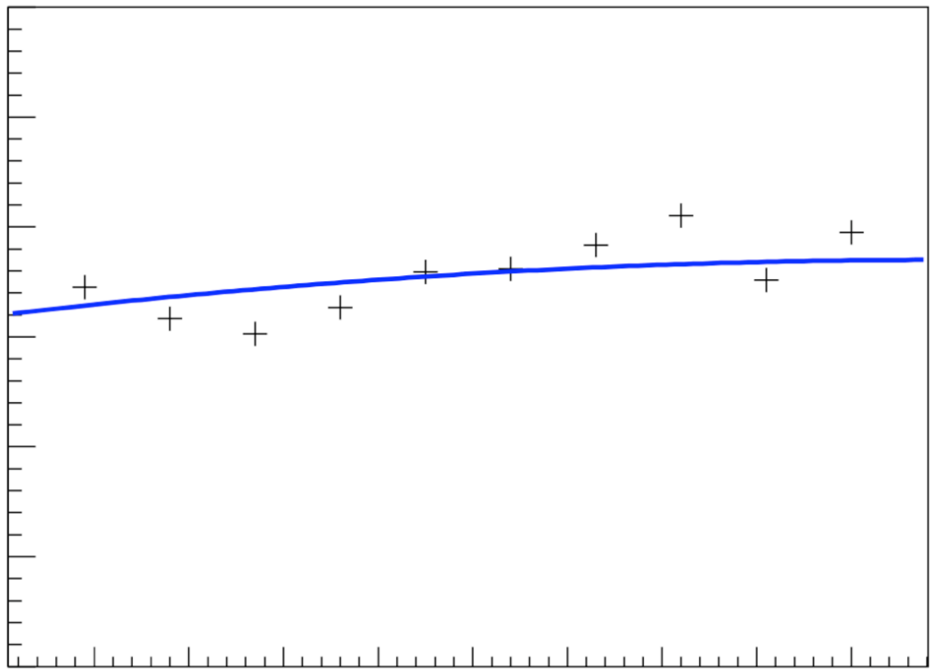
Frequentist	➲	Bayesian
Optimization	➲	Expectation
(Differentiation)	➲	(Integration)

The Opposition

- “Priors are Subjective, Science Isn’t”

Application: Model Selection



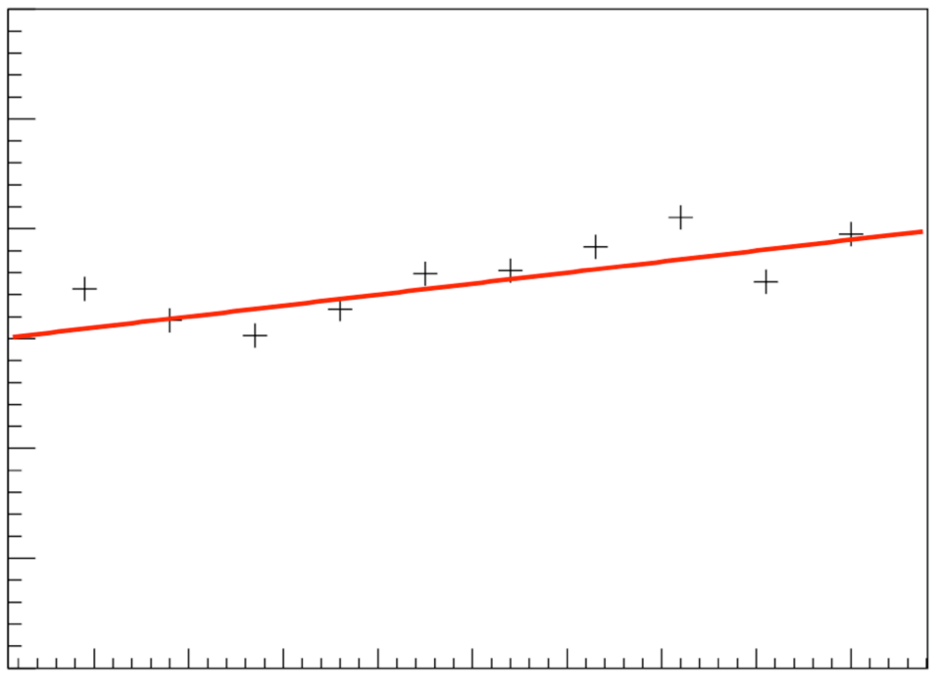


Application: Model Selection

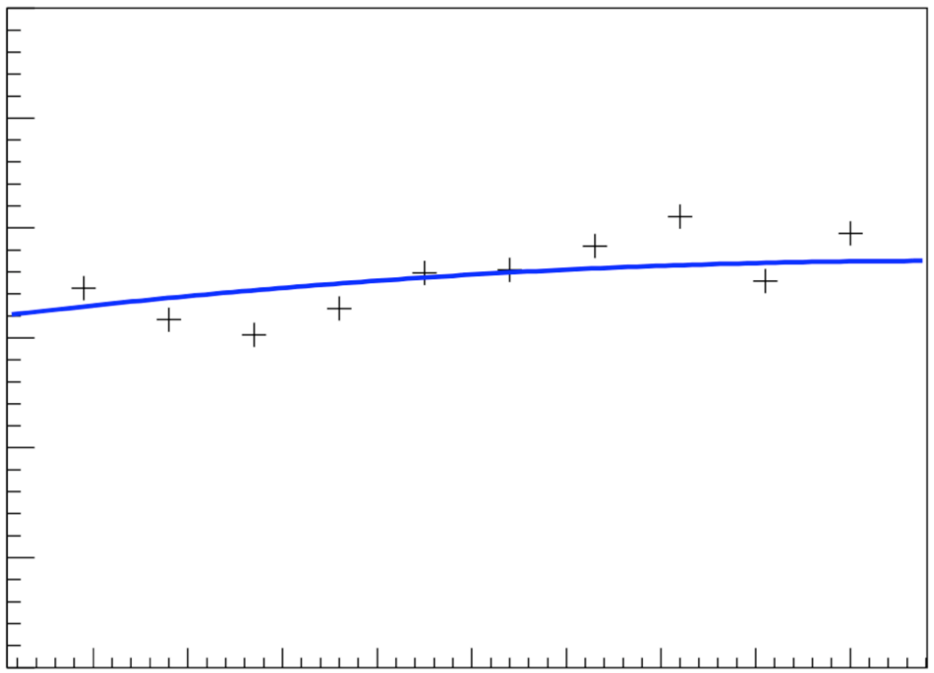
$$p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)=\frac{p\left(x_1,\ldots,x_m|\alpha_1,\ldots,\alpha_n\right)p\left(\alpha_1,\ldots,\alpha_n\right)}{p\left(x_1,\ldots,x_m\right)}$$

$$E=p\left(x_1,\ldots,x_m\right)=\int \mathrm{d}^n\alpha\, p\left(\alpha_1,\ldots,\alpha_n|x_1,\ldots,x_m\right)p\left(\alpha_1,\ldots,\alpha_n\right)$$

Application: Model Selection

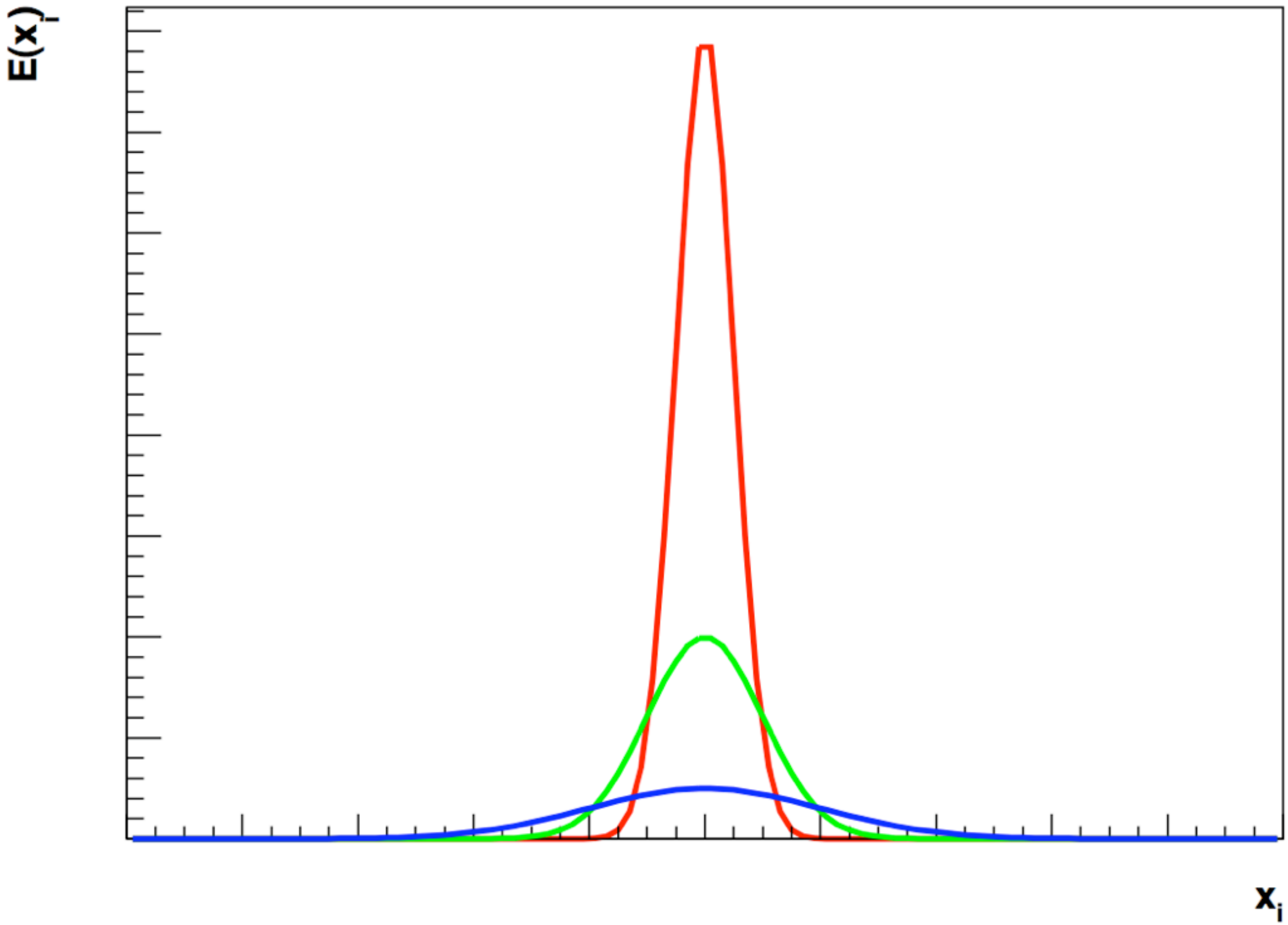


$E\,(\text{Line})$



$E\,(\text{Quadratic})$

The Bayesian Razor



Application: Systematics

Energy Scale	5%
Background Subtraction	1%
Trigger Bias	2%
Luminosity	4%
...	n%
Total	7%

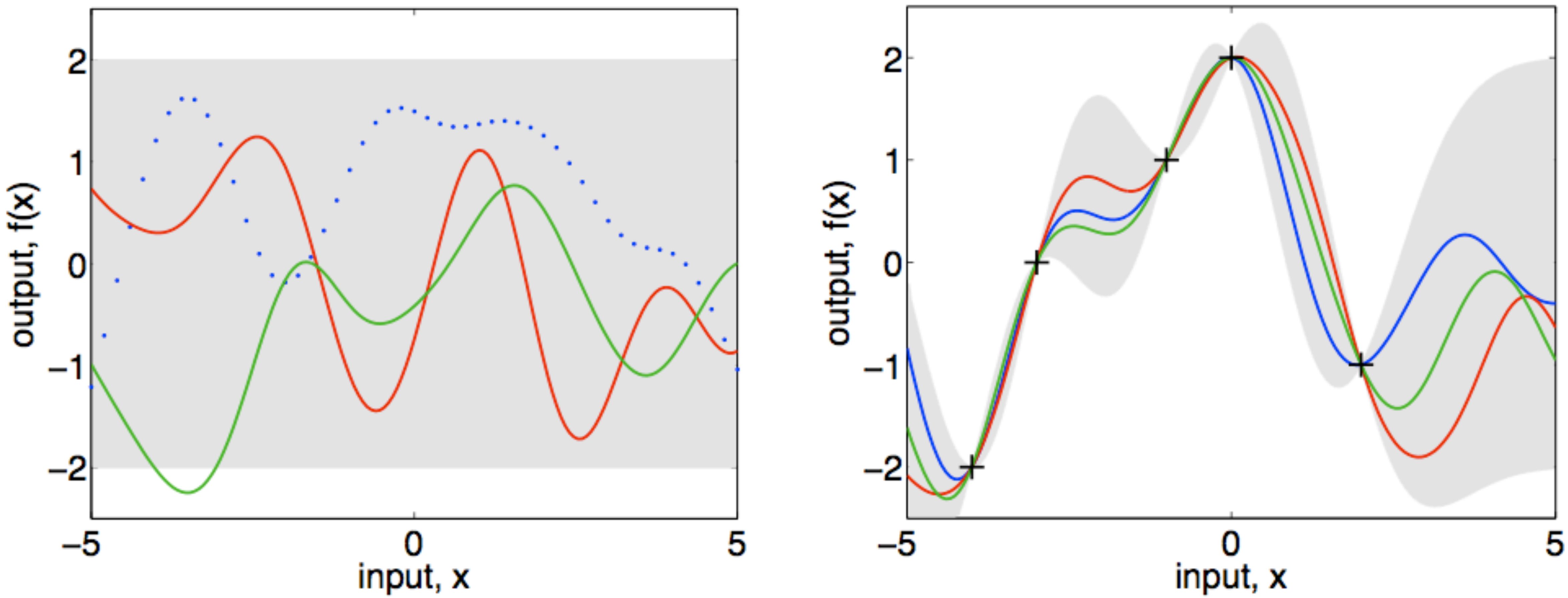
$$\Delta N = \frac{1}{2}\left[N\left(x_1, \ldots, x_m \middle| \alpha_1, \ldots, \alpha_i + 0.7, \ldots, \alpha_m\right) - N\left(x_1, \ldots, x_m \middle| \alpha_1, \ldots, \alpha_i - 0.7, \ldots, \alpha_m\right)\right]$$

Application: Systematics

$$p(\sigma)$$

$$p(N) = \int \mathrm{d}\sigma\, p(N|\sigma)\, p(\sigma)$$

Application: Nonparametric Estimation



Further Reading

- *Information Theory, Inference, and Learning Algorithms*,
  MacKay, http://www.inference.phy.cam.ac.uk/mackay/itila/
- *Data Analysis*,
  Sivia with Skilling, Oxford 2006
- *Lectures on Probability, Entropy, and Statistical Physics,*
  Caticha, http://arxiv.org/abs/0808.0012v1