Tools for Parameter Estimation and Propagation of Uncertainty

Brian Borchers
Department of Mathematics
New Mexico Tech
Socorro, NM 87801
borchers@nmt.edu

Outline

- Models, parameters, parameter estimation, and uncertainty.
- A simplified example, the van Genuchten model.
- Methods for propagation of uncertainty.
- Nonlinear regression as an approach to parameter estimation.
- A Bayesian approach to the parameter estimation problem.
- How does this apply to CRONUS?

Models and Parameters

- In this talk we will consider models of the form

  $$y = f(x, p)$$

  where x is an independent variable, y is a dependent variable, p is some collection of parameters that enter into the model.

- The model might be physically based, or completely empirical, or somewhere in between.

- In practice there will always be uncertainties in x, y, and p.

- Furthermore, there might be uncertainty in f itself.

Three Problems

- The **parameter estimation** problem is to find p given a suitable collection of x and y values and/or information from other experiments.
- Once we have p, we can consider the problem of **predicting** y from x.
- We will also be interested in the **backwards prediction** problem of estimating x from y.
- In CRONUS, x would be the exposure time for a sample, y would be the TCN production, and p would include factors such as the cosmic ray flux, reaction cross sections, etc.

The van Genuchten Model

- In hydrology, the van Genuchten model is often used to relate the volumetric water content of an unsaturated soil to the head.
- The model is
$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{(1 + (-\alpha h)^n)^{(1-1/n)}}.$$
- Here θ_s is the volumetric water content at saturation, and θ_r is the residual volumetric water content at a very large negative head. We will assume that these are known.
- The two parameters α and n can be fit to laboratory measurements.

Propagation of Errors by Linearization

- Let δx be the uncertainty in x, δp_i be the uncertainty in the ith parameter p_i, and let δy be the uncertainty in y.
- For independent random errors,

$$\delta y = \sqrt{\left(\frac{\partial f}{\partial x}\delta x\right)^2 + \left(\frac{\partial f}{\partial p_1}\delta p_1\right)^2 + \ldots + \left(\frac{\partial f}{\partial p_m}\delta p_m\right)^2}.$$

- For errors that may not be independent,

$$\delta y \leq \left|\frac{\partial f}{\partial x}\delta x\right| + \left|\frac{\partial f}{\partial p_1}\delta p_1\right| + \ldots + \left|\frac{\partial f}{\partial p_m}\delta p_m\right|.$$

Example

- Suppose that $\alpha = 0.013 \pm 0.001$, and $n = 1.86 \pm 0.04$. Also, $\theta_s = 0.44$ and $\theta_r = 0.09$. Suppose that $h = -100 \pm 1$. What is $\theta(h)$?
- Simply substituting the nominal values into the formula gives $\theta(-100) = 0.31$.
- Using the propagation of errors formula, we obtain an uncertainty of $\delta\theta = 0.01$.
- If we allow for correlation between α, n, and h, we obtain $\delta\theta \leq 0.014$.

Example

Note that computing $\delta\theta$ required the computation of derivatives with respect to α, n, and h.

$$\frac{d\theta}{d\alpha} = -\frac{(\theta_s - \theta_r)\,(1 - 1/n)\,(-\alpha\,h)^n\,n}{\left((1 + (-\alpha\,h)^n)^{1-1/n}\right)\alpha\,(1 + (-\alpha\,h)^n)}.$$

$$\frac{d\theta}{dh} = -\frac{(\theta_s - \theta_r)\,(1 - 1/n)\,(-\alpha\,h)^n\,n}{\left((1 + (-\alpha\,h)^n)^{1-1/n}\right)h\,(1 + (-\alpha\,h)^n)}.$$

$$\frac{d\theta}{dn} = -\frac{(\theta_s - \theta_r)\left(\frac{\ln(1+(-\alpha\,h)^n)}{n^2} + \frac{(1-1/n)(-\alpha\,h)^n\,\ln(-\alpha\,h)}{(1+(-\alpha\,h)^n)}\right)}{\left((1 + (-\alpha\,h)^n)^{1-1/n}\right)}.$$

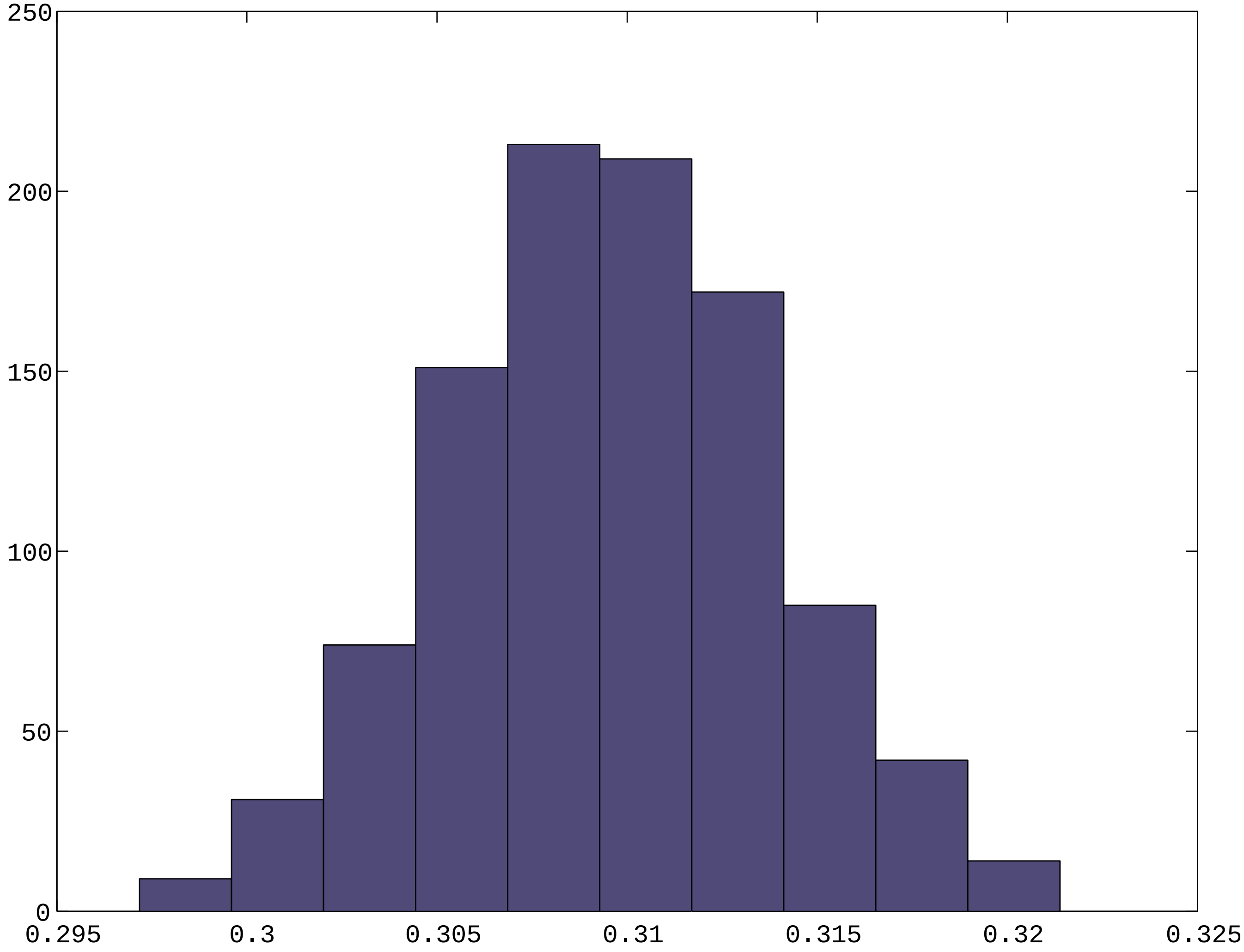
Monte Carlo Propagation of Errors

- We are given nominal values for x and the parameters p, together with their associated uncertainties.
- Generate many (say $n = 1000$) random values of x and p using the nominal values and uncertainties.
- For each realization of x and p compute y.
- Produce a histogram, average, and standard deviation of the y values.

Example

- We randomly generated one thousand values of α, n, and h. The values of α were normally distributed with mean 0.013 and standard deviation 0.001. The values of n were normally distributed with mean 1.86 and standard deviation 0.04. The values of h were normally distributed with mean -100 and standard deviation 1.
- For these 1,000 realizations, the mean value of $\theta(h)$ was 0.3095 with standard deviation 0.0043.
- In this case the linearized propagation of errors formula gave a somewhat larger uncertainty in θ than the uncertainty revealed by the Monte Carlo technique.

Example



Advantages of the Monte Carlo Approach

- The Monte Carlo approach avoids any errors associated with the linearization of the model. These can be substantial, especially when the model is highly nonlinear in its parameters or when the parameter uncertainties are quite large.
- The Monte Carlo approach produces a distribution for the uncertain output as well as the mean and standard deviation.
- The propagation of errors formula requires the computation of derivatives, which can be quite complicated for larger models. The Monte Carlo approach needs only a black box routine for computing $f(x, p)$.
- The Monte Carlo approach can handle correlated parameters as well as independent parameters.

Nonlinear Regression

- Nonlinear regression is the classical approach to estimating the parameters p from a collection of x, y measurements.
- We assume that the x values x_1, x_2, ..., x_n are known exactly, but the y values y_1, y_2, ..., y_n have independent and normally distributed errors with standard deviations σ_1, σ_2, ..., σ_n.
- The basic idea is to solve the nonlinear least squares problem

$$\min \chi^2(p) = \sum_{i=1}^{n} \left(\frac{y_i - f(x_i, p)}{\sigma_i} \right)^2 .$$

- Call the optimal parameter vector p^*.

Nonlinear Regression

- The value of $\chi^2_{\text{obs}} = \chi^2(p^*)$ provides an important statistical measure of the goodness of fit.
- In theory, the value of χ^2_{obs} should follow a χ^2 distribution with $n - m$ degrees of freedom.
- An extremely large value of χ^2_{obs} would be extremely unlikely, indicating that parameters were not fit correctly or that there was some other problem with the model.
- The p-value is the probability of obtaining a χ^2 value that is larger than the one that was actually obtained.

Covariance of the Fitted Parameters

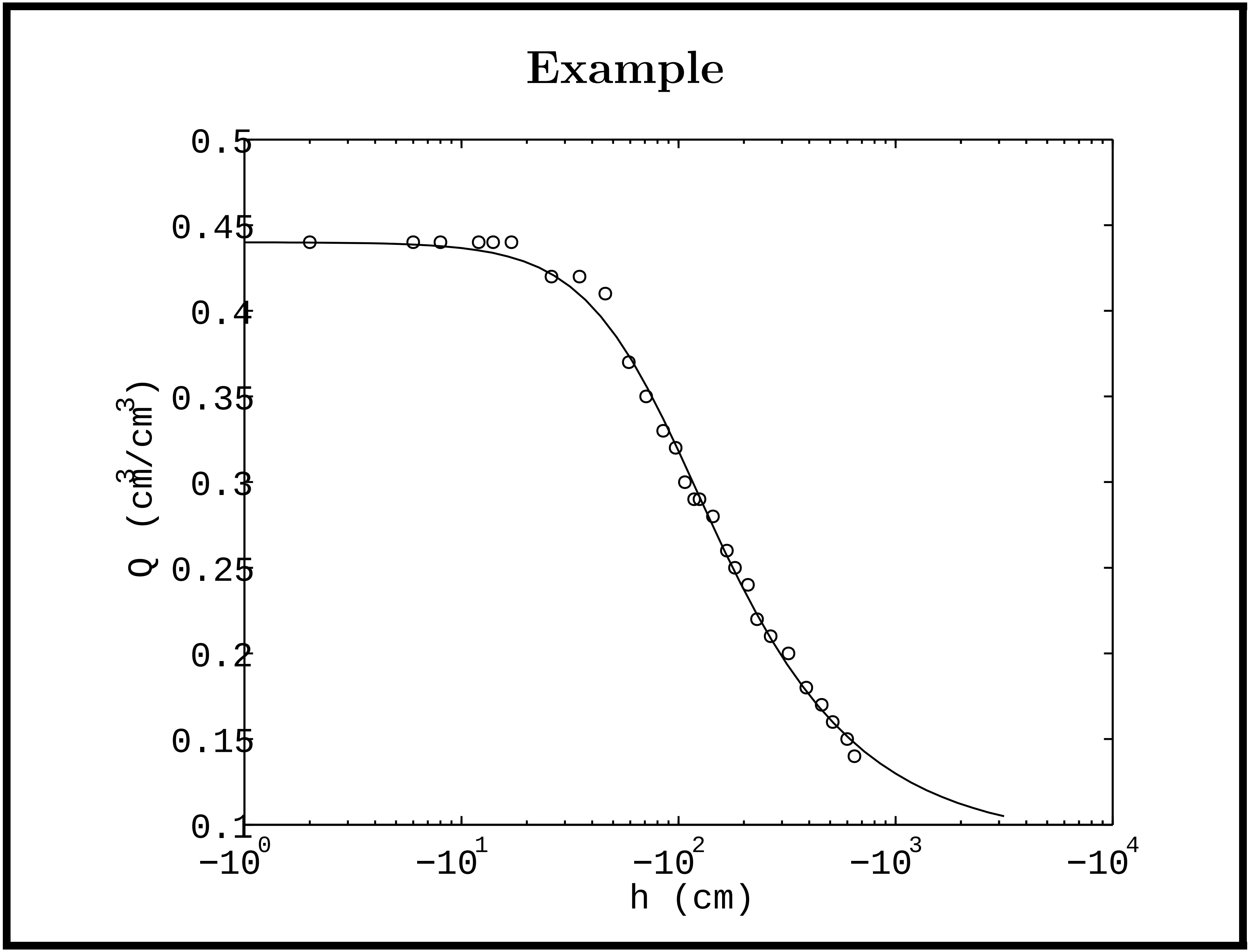
- Let
$$f_i(p) = \frac{y_i - f(x_i, p)}{\sigma_i}.$$
- Let
$$J(p) = \begin{bmatrix} \frac{\partial f_1(p)}{\partial p_1} & \cdots & \frac{\partial f_1(p)}{\partial p_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(p)}{\partial p_1} & \cdots & \frac{\partial f_n(p)}{\partial p_m} \end{bmatrix}.$$
- The covariance matrix for the fitted parameters is then
$$C = \left(J(p^*)^T J(p^*) \right)^{-1}.$$
- From C, we can construct approximate 95% confidence intervals and confidence ellipsoids for the fitted parameters.

Example

- Returning to our earlier example, suppose that we have the following data.

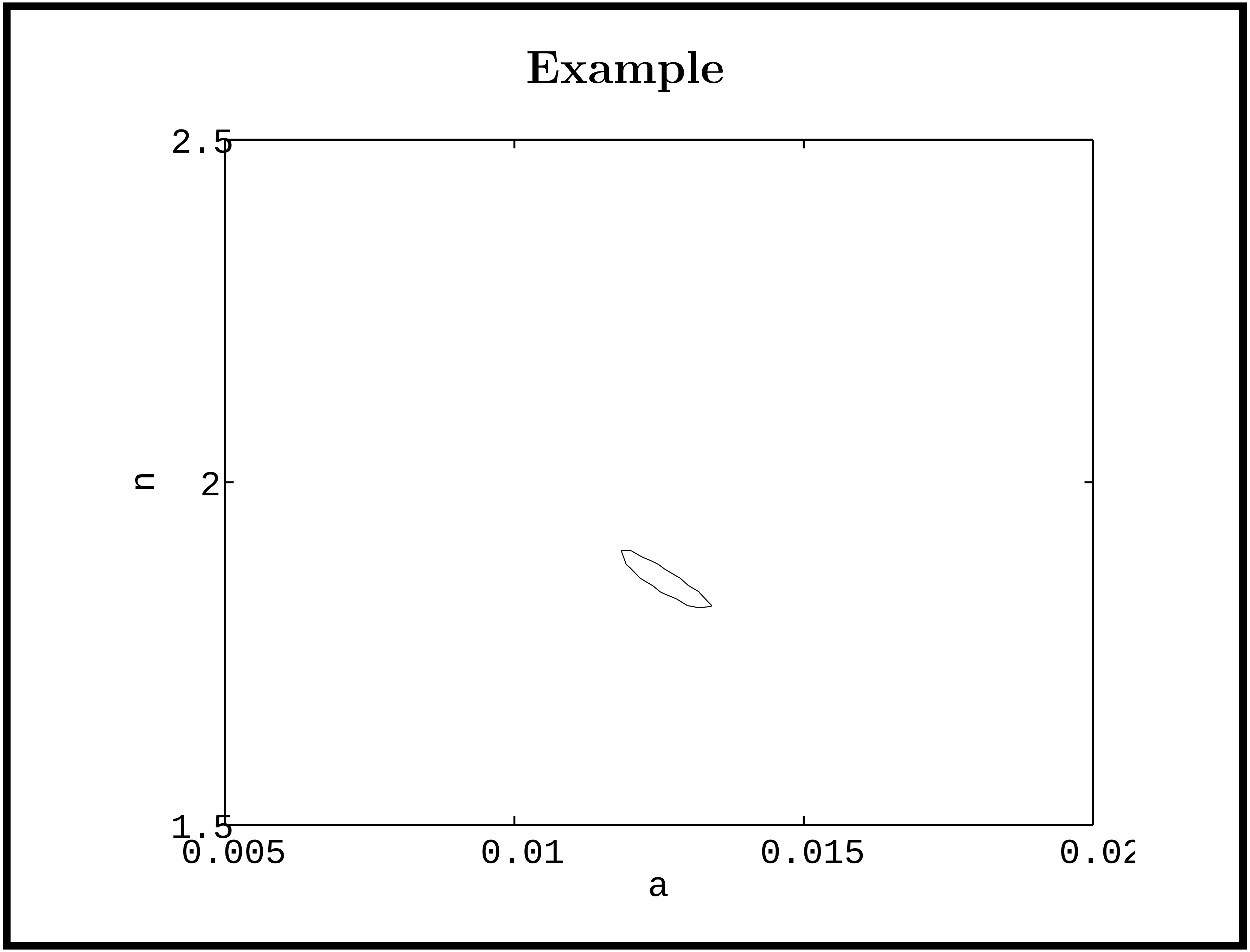
h	-2	-6	-8	-12	-14	-17	-26	-35	-46	-59
θ	0.44	0.44	0.44	0.44	0.44	0.44	0.42	0.42	0.41	0.37
h	-71	-85	-97	-107	-118	-125	-144	-167	-182	-209
θ	0.35	0.33	0.32	0.30	0.29	0.29	0.28	0.26	0.25	0.24
h	-230	-266	-321	-388	-457	-514	-599	-647		
θ	0.22	0.21	0.20	0.18	0.17	0.16	0.15	0.14		

- The measurements of θ are assumed to be accurate to 2%.
- From other measurements, we know that $\theta_r = 0.09$ and $\theta_s = 0.44$.


Example
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
Q (cm3/cm3)
−10^0
−10^1
−10^2
−10^3
−10^4
h (cm)

Example

- $\chi^2_{\text{obs}} = 31.3$. Since there are 28 data points and two parameters to estimate, we have 26 degrees of freedom, and this χ^2 value is not large. The p–value is 22%.
- 95% confidence intervals for the individual parameters are $\alpha = 0.01257 \pm 0.00066$ and $n = 1.85664 \pm 0.03667$.
- However, the two parameters are strongly correlated, with $\rho(\alpha, n) = -0.90$.
- The following plot shows a 95% confidence ellipsoid for the parameters.


Example
2.5
2
1.5
n
0.005
0.01
0.015
0.02
a

The Bayesian Approach

- In the Bayesian approach, we use probability distributions to quantify the uncertainty in the model parameters.
- We begin with a **prior distribution** for the parameters, $\text{prob}(p|I)$, that depends on any initial information that we might have.
- As data becomes available, we update the distribution using Bayes' theorem
$$\text{prob}(p|d,I) = \frac{\text{prob}(p|I)\text{prob}(d|p,I)}{\int \text{prob}(p|I)\text{prob}(d|p,I)dp}.$$
- $\text{prob}(d|p,I)$ is the **likelihood** of observing data d given parameters p and prior information I.
- $\text{prob}(p|d,I)$ is the **posterior distribution** for the parameters.

The Bayesian Approach

- In practice, computing the integral in the denominator of Bayes' theorem can be computationally difficult. This integral is a normalizing constant for the posterior distribution.
- For many purposes, a simplified version of the formula is sufficient:

$$\text{prob}(p|d, I) \propto \text{prob}(p|I)\text{prob}(d|p, I).$$

- For example, we can find the vector of parameters p^* that has the maximum posterior probability without evaluating the integral. This vector p^* is called the maximum a posteriori (MAP) solution.
- Also, we can generate random values for the parameters according to the posterior distribution without computing the integral.

The Bayesian Approach

- An important practical advantage of the Bayesian approach is that we can repeatedly apply Bayes' theorem to incorporate additional data.
- If we receive data d_1, d_2, …, d_n, then

$$\mathrm{prob}(p|d_1, \ldots, d_n, I) \propto \mathrm{prob}(p|I)\mathrm{prob}(d_1|p, I) \cdots \mathrm{prob}(d_n|p, I).$$

Example

- Returning to our earlier example, suppose that we know from prior experience that α lies between 0.005 and 0.02 and that n lies between 1.0 and 3.0.
- Since we do not know anything else, we simply assign a flat prior distribution to α and n.
- This distribution needs to be normalized so that it integrates out to one. The result is

$$\text{prob}(p, I) = 33.3333$$

Example

- Now, we obtain a data point: at a head of $h = -2$, $\theta = 0.44$. Since the measurement of θ is assumed to have a normally distributed measurement error with standard deviation $\sigma = 0.008$, the likelihood is normal.

$$\text{prob}(d_1|\alpha, n, I) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\left(\theta_r + \frac{\theta_s - \theta_r}{(1 + (-\alpha(-2.0))^n)^{(1-1/n)}} - 0.44\right)^2 / (2\sigma^2)\right)$$

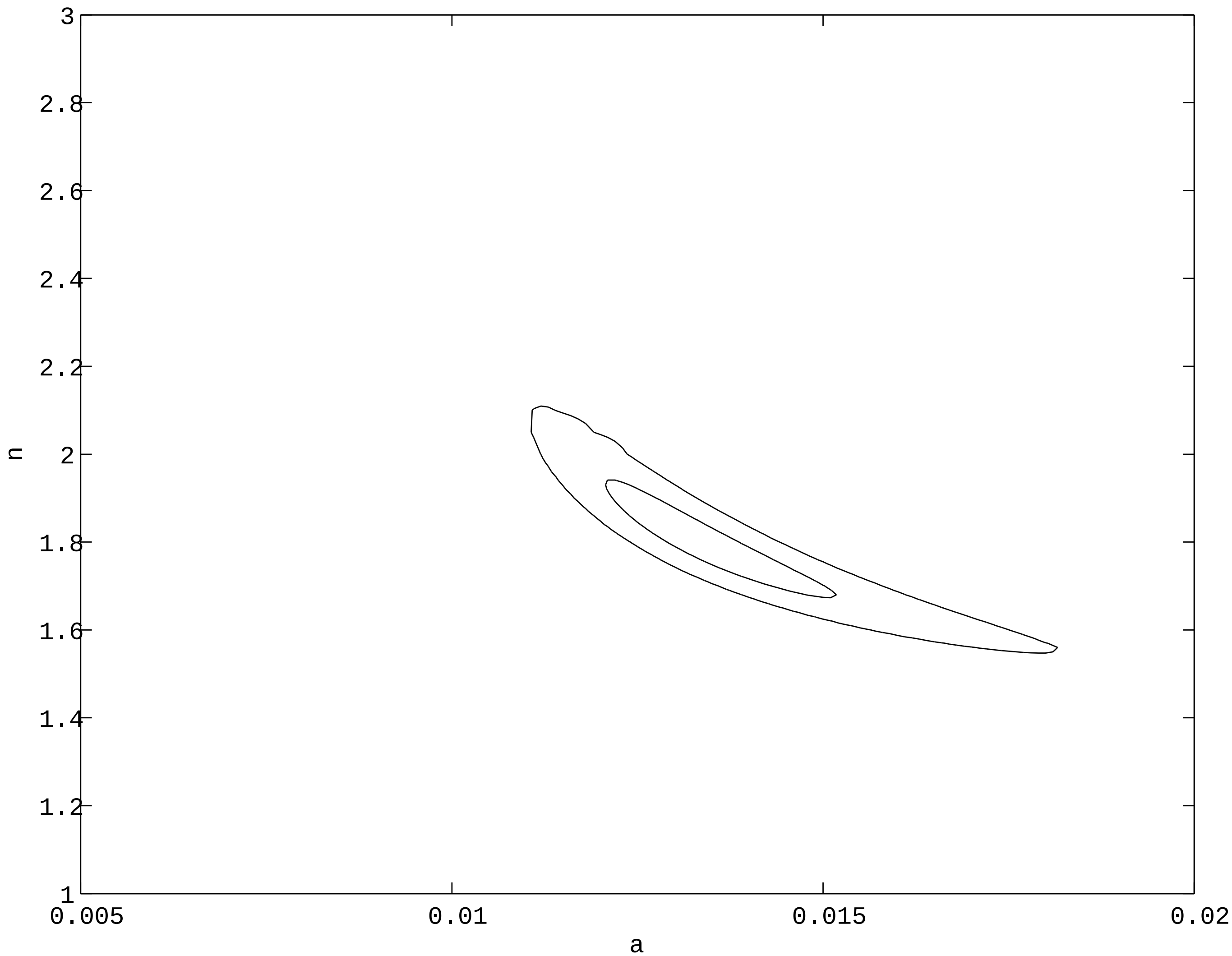
- Now we can compute the posterior distribution

$$\text{prob}(p|d_1, I) \propto \text{prob}(p|I)\text{prob}(d_1|p, I).$$

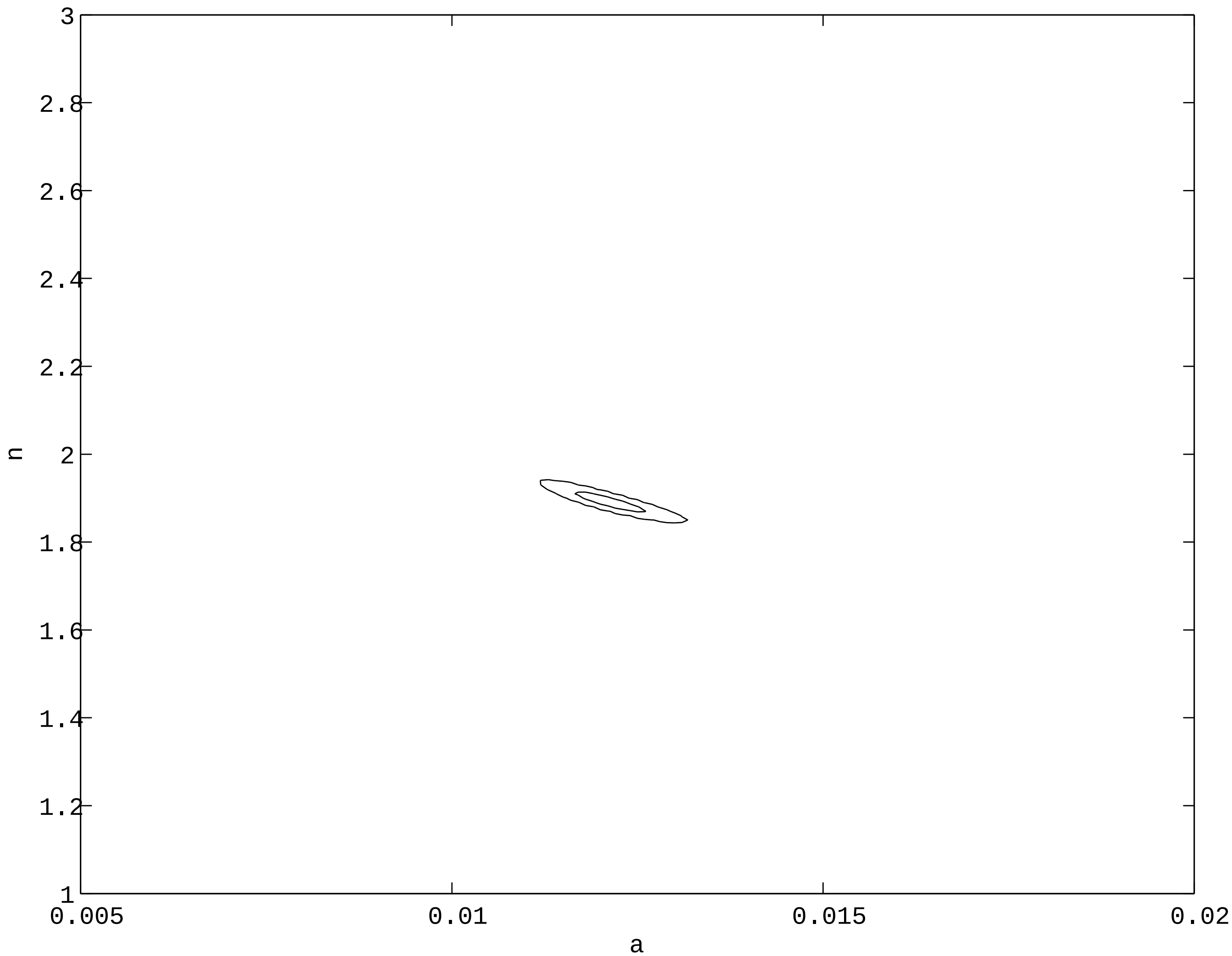
- In the same way, we can incorporate data points $d_2 = (-59, 0.37)$, $d_3 = (-125, 0.29)$, ...

- Here are contour plots of the posterior distribution after 3, 10, and 28 data points.

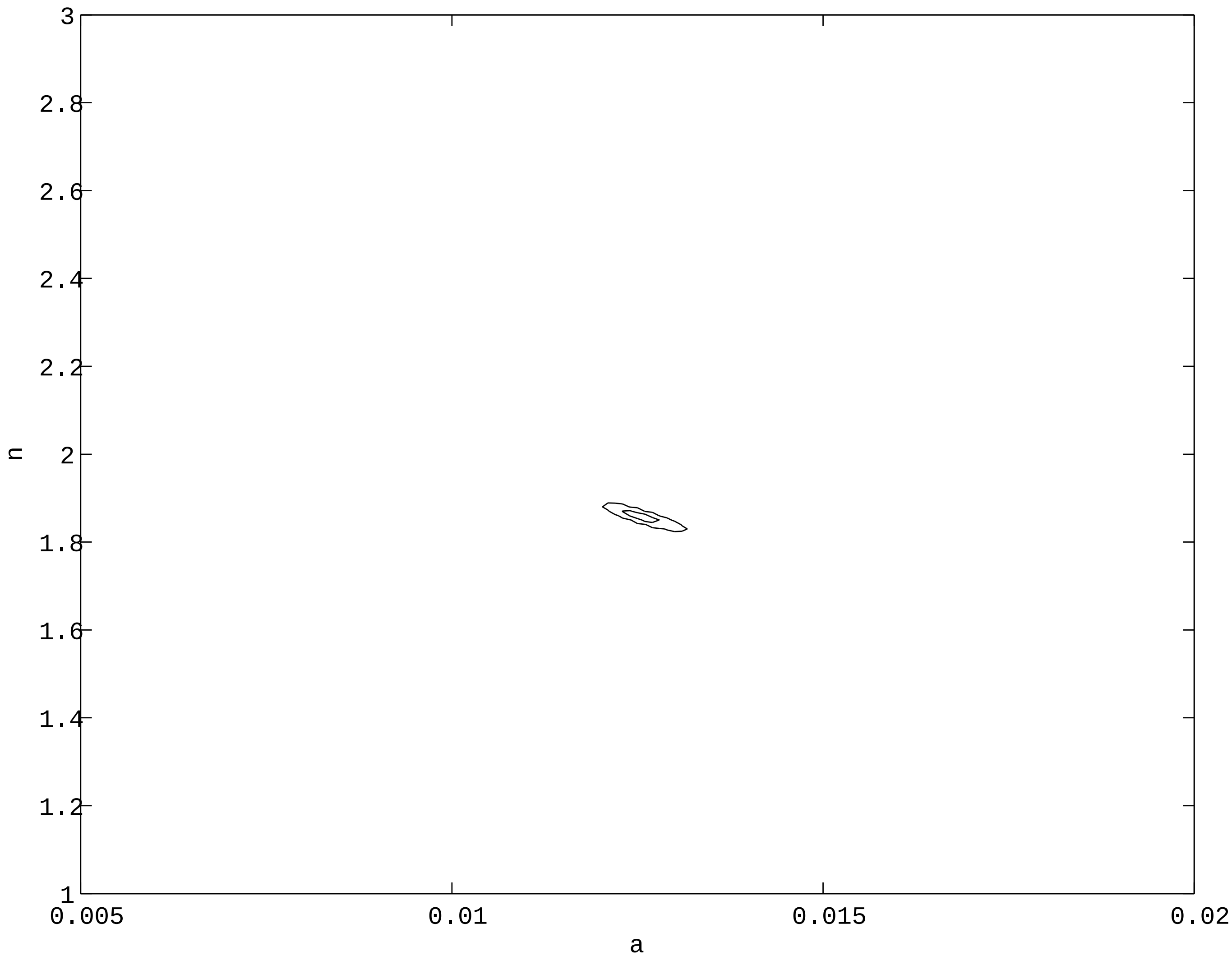
Example



Example



Example



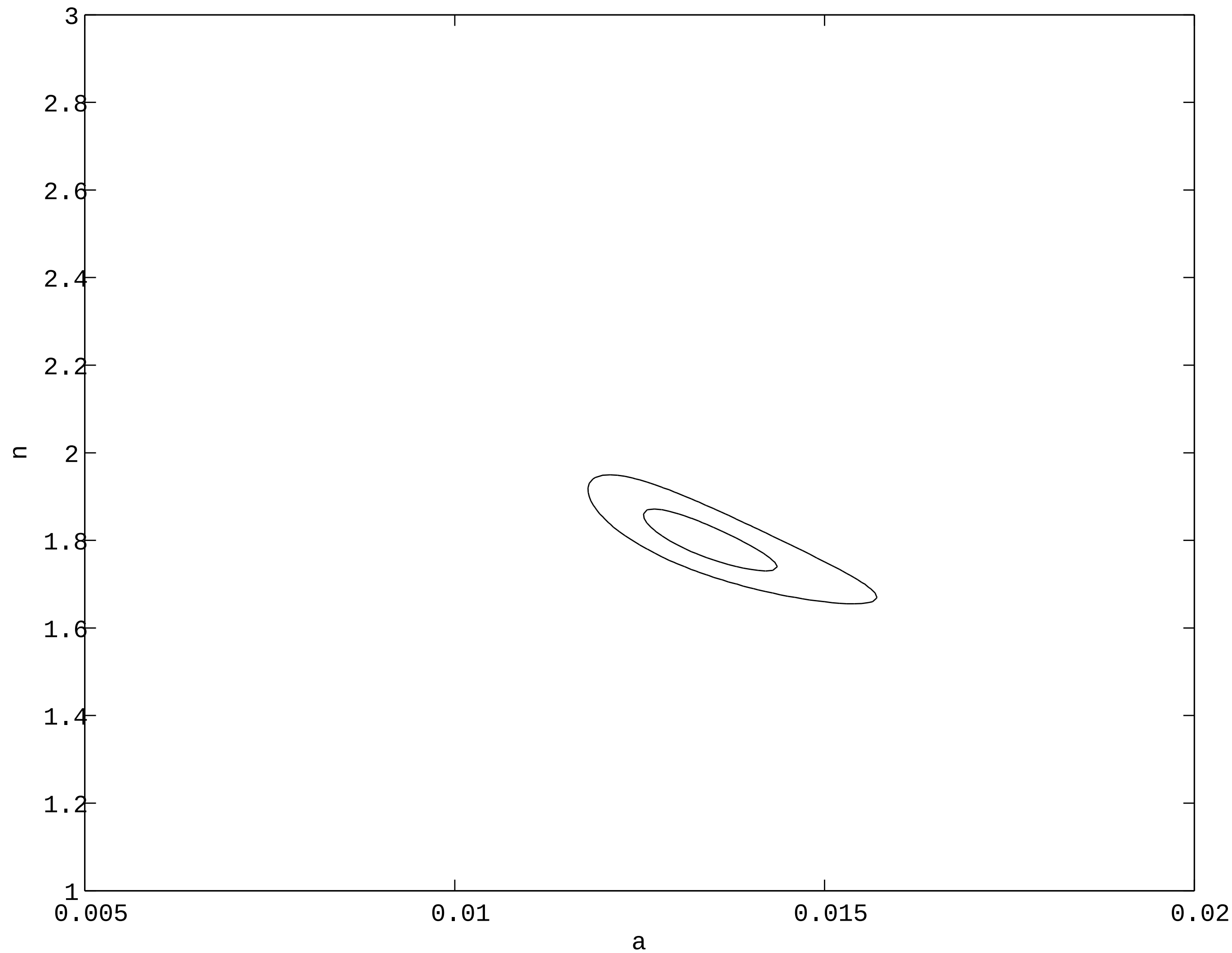
Example

- Now, suppose that we knew from some other measurements of the soil that n was near 1.8.
- We could assign a prior distribution of

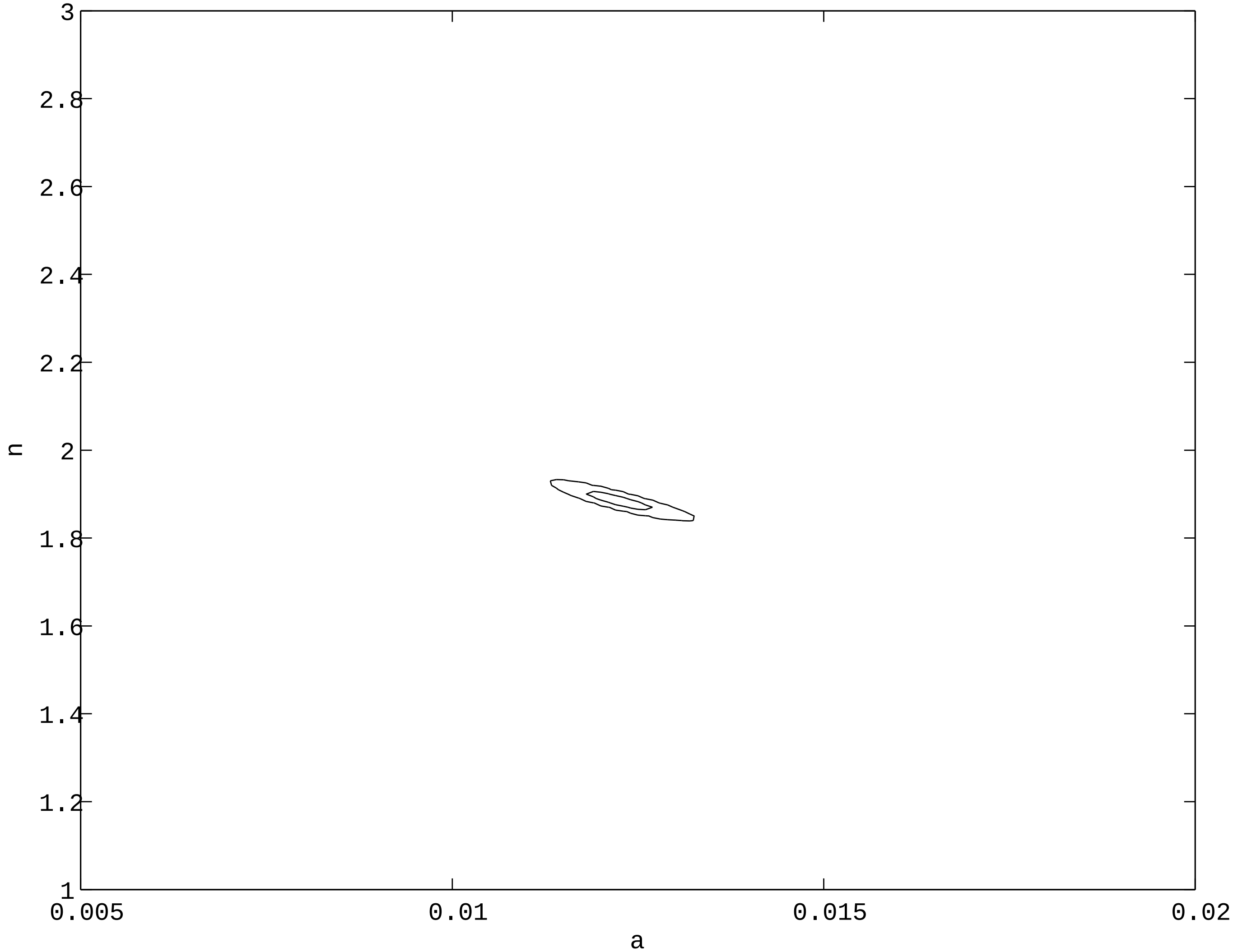
$$\text{prob}(\alpha, n|I) \propto \exp\left(-\left(\frac{n-1.8}{0.1}\right)^2\right).$$

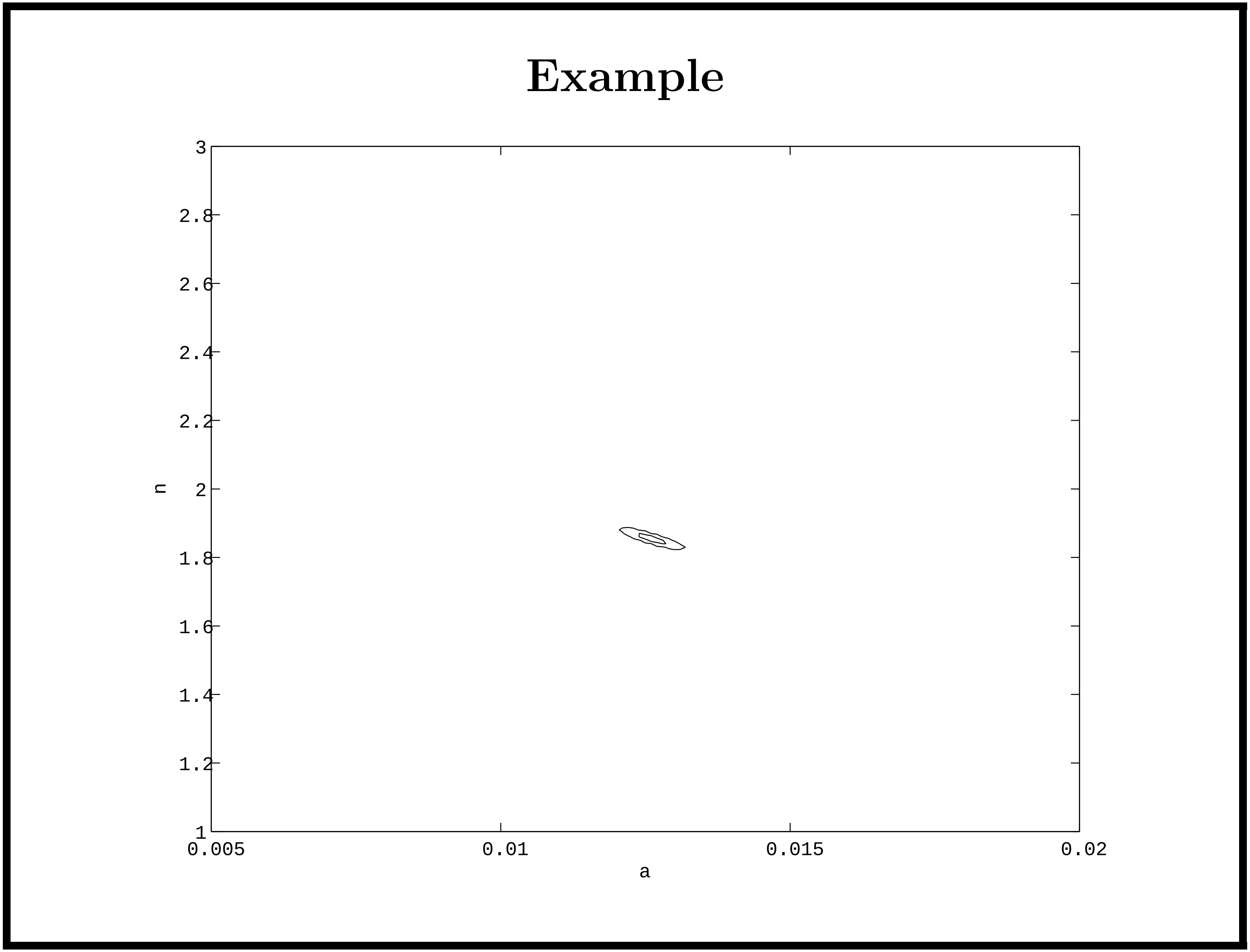
- Here are contour plots of the posterior distribution after 3, 10, and 28 data points.

Example



Example

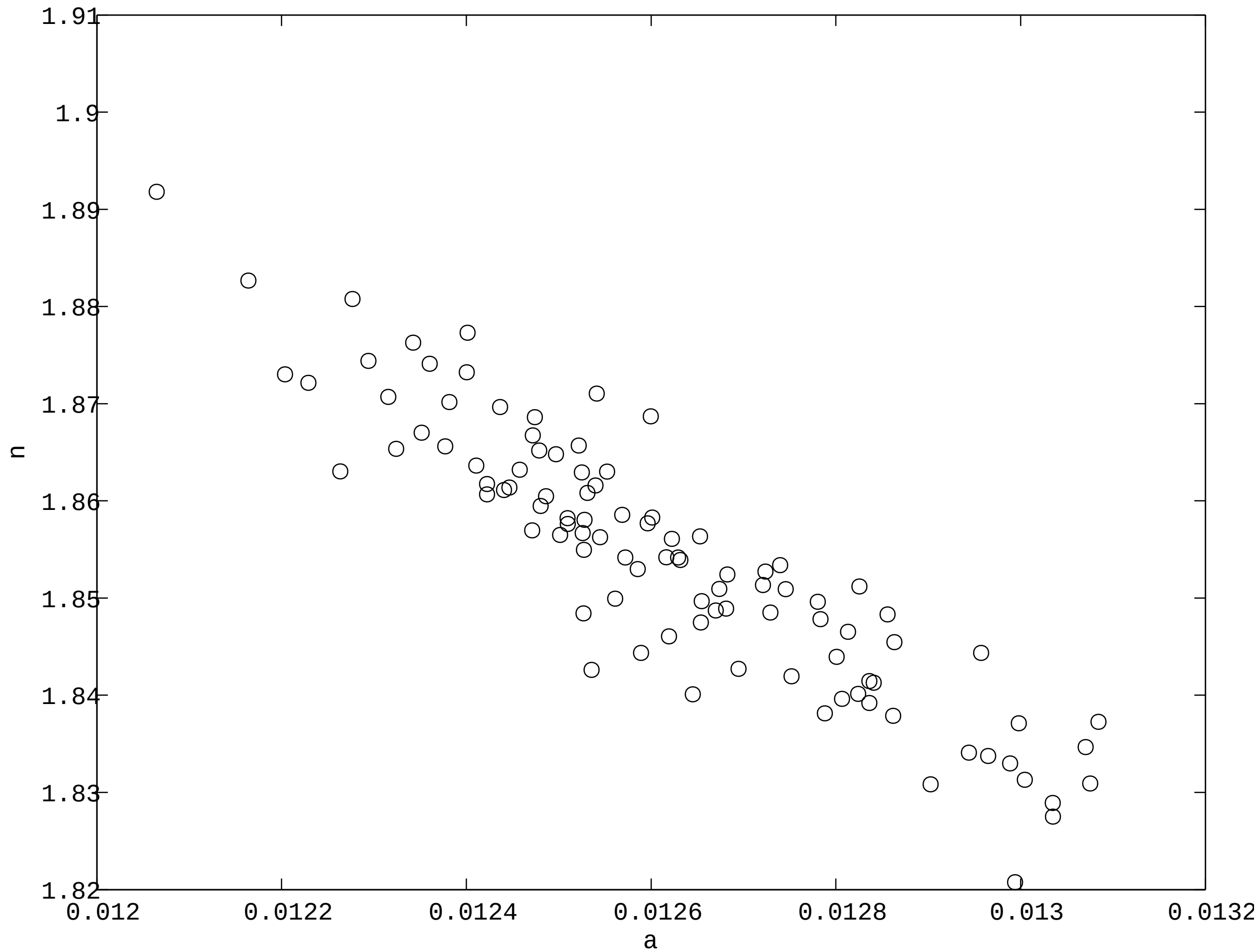



Example
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
n
0.005
0.01
0.015
0.02
a

Example

- Once we have a posterior distribution, we can easily generate random sets of parameters according to the posterior distribution.
- This can be used as input to a Monte Carlo procedure for propagation of uncertainty.
- The following plot shows 100 random sets of parameters for our example.

Example


1.91
1.9
1.89
1.88
1.87
1.86
1.85
1.84
1.83
1.82
n
0.012
0.0122
0.0124
0.0126
0.0128
0.013
0.0132
a

Advantages of the Bayesian Approach

- The Bayesian approach allows us to incorporate data from many different kinds of experiments, as well as prior information that may not be in the form of experimental data.
- The Bayesian approach is incremental, not "all-at-once" as in nonlinear regression. We can simply incorporate new data as it becomes available.
- The Bayesian approach with sampling from the posterior distribution does not require us to linearize the nonlinear model to obtain uncertainties for the parameters.
- In the Bayesian approach, the posterior distribution provides a natural input to a procedure for propagating uncertainty in x and p to y.

Some Questions

- How do we deal with multiple data sets that bear on the same parameter?
- This is easy- just construct appropriate likelihood functions for each type of data set.

Some Questions

- How do we deal with anomalous data?
- Sometimes anomalous data is just wrong, while other times it points to fundamental modeling problems. By examining the likelihood of each new data point, we can identify data that don't fit with the existing model. Then the challenge is to determine whether the data are bad or whether the data are pointing to a modeling error. Unfortunately, this is not something that can be done statistically.

Some Questions

- How do we intercompare disparate types of data?
- If the data are all valid, then they should be incorporated into the posterior distribution. If there is some doubt as to the validity of a data set, then the specifics of that experiment need to be investigated.

What Next?

- The first step in the parameter estimation process is to settle on the parameters that need to be incorporated into the model and the exact form of the function that relates the parameters to TCN production.
- If there are conflicting models, a heirarchical Bayesian approach may be needed to distinguish between them.
- Likelihood functions need to be developed for the different types of experiments that have been and will be conducted.
- As data sets become available, they need to be incorporated into the analysis. At each stage there will be a current working distribution for the parameters.

What Next?

- Once enough data has been accumulated, the working distribution can be published along with the software for performing TCN production predicition and TCN dating. This will be an iterative process, with versions 1.0, 2.0, ...
- Once the statistical model is well established, we can perform additional "what if" analysis to determine how the results of different types of experiments might reduce the uncertainty in the parameters and the resulting uncertainty in TCN dates.

How Can You Help?

- For this effort to worthwhile to the community of researchers that use TCN methods, there must be widespread understanding and acceptance of the statistical model produced by the CRONUS project.
- If the CRONUS community can agree on reasonable prior distributions for the parameters this would help greatly. To the extent that there is disagreement, the prior distributions need to reflect the different theories.
- There's an obvious need to integrate the results of the parameter estimation with the development of software for TCN applications.
- Experimenters providing data for the analysis will have to provide information about the experiments and help with the development of Likelihood functions.

References

[1] R. C. Aster, B. Borchers, and C. T. Thurber. *Parameter Estimation and Inverse Problems*. Elsevier Academic Press, San Diego, 2004.

[2] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis,* 2nd ed. Chapman & Hall/CRC, Boca Raton, FL, 2003.

[3] W.R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

[4] D. S. Sivia. *Data Analysis: A Bayesian Tutorial For Scientists and Engineers*. Oxford University Press, New York, 1996.