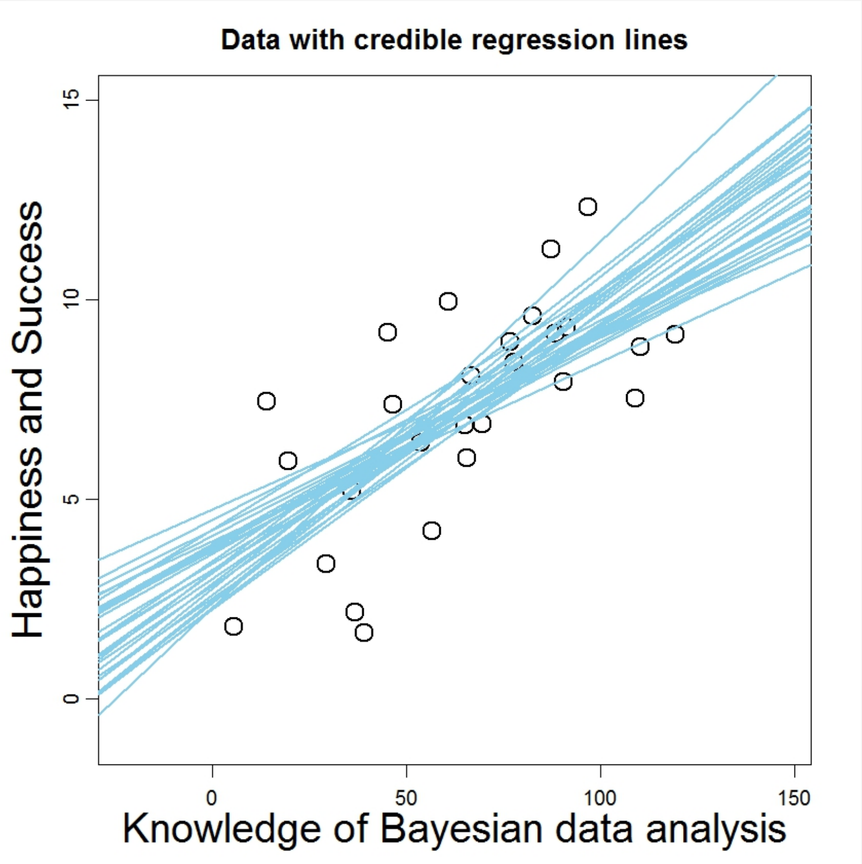
STATS 331



Credit: John Kruschke

Introduction to Bayesian Statistics
Semester 2, 2016

Bayesian parameter estimation

- Today we'll do two examples of Bayesian parameter estimation, and one with a prediction
- For both, we'll get the likelihood by defining the *sampling distribution* first.
- Will look at a different "non-informative" prior distribution that isn't uniform

Bayes' Rule for Parameter Estimation

$$
\begin{aligned}
p(\theta|x) &= \frac{p(\theta)p(x|\theta)}{p(x)} \\
p(\theta|x) &\propto p(\theta)p(x|\theta) \\
\texttt{posterior} &\propto \texttt{prior} \times \texttt{likelihood}.
\end{aligned}
$$

This works for discrete and continuous distributions

Taxi Problem

- You fall asleep in a foreign city
- When you wake up in the morning, you see a taxi drive by, that says

"This is taxi number 42"

How many taxies are in the city?

Image credit: wikipedia

Let's calculate the posterior in R

Auckland Volcano Example

Image is in Public domain

The Problem (based on STATS 210)

- In the last 20,000 years (call this 1 time unit), there have been 20 volcanic eruptions.
- What's the probability of an eruption in the next 50 years (0.0025 time units)?

Volcanoes and Poisson Processes

- We don't need to know much detail about the Poisson process

```
   *              *  *                   *              * *    *

-----------------------------------------------------------------> time
```

Except

- If we ask the question “how many events” we need the Poisson distribution

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Poisson Distribution

- This is a probability distribution for the data (in this problem).

- Can predict the count *x*
- Could even predict the count in the future

- *But useless if we don't know the value of the parameter λ!!!!*

The Data

$x = 20$

(and we will need to remember this occurred over a period of 20,000 years)

Let's Estimate λ

- Set of possible values? Units are volcanoes per 20,000 years. Let's go from 1 to 100 in steps of 1.

- This discreteness is an **approximation**

Bayes Box: Possible Values

Possible Values λ	Prior $p(\lambda)$	Likelihood $p(x\|\lambda)$	Prior x Likelihood $p(\lambda)p(x\|\lambda)$	Posterior $p(\lambda\|x)$
1	0.01			
2	0.01			
3	0.01			
4	0.01			
...	...			
97	0.01			
98	0.01			
99	0.01			
100	0.01			
Totals	1		$p(x)$	1

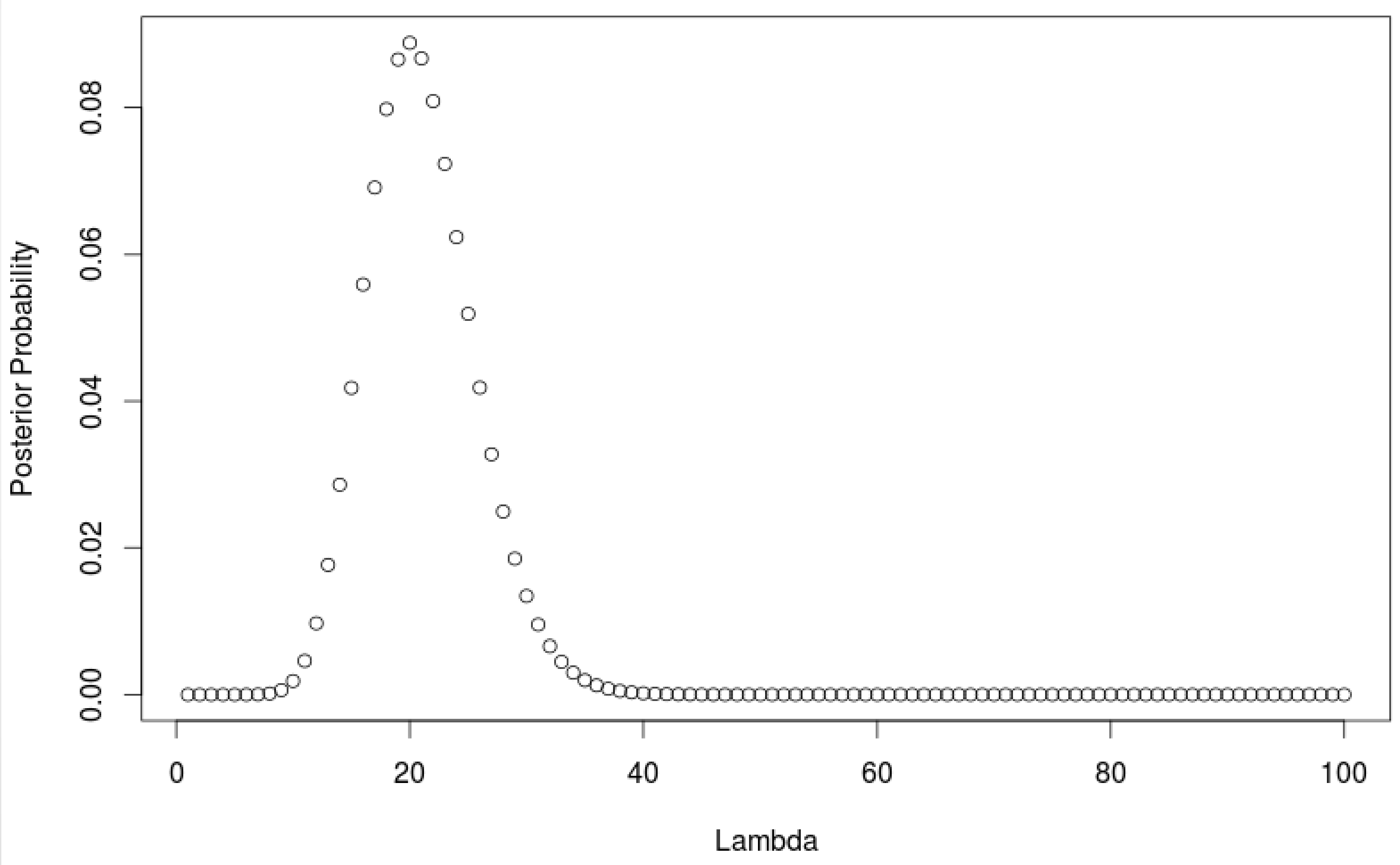
Likelihood

- Compare with Lab 1, Question 2

- In parameter estimation problems, the likelihood is obtained from the sampling distribution but with the observed values of the data plugged in

Let's calculate the posterior in R

Posterior Distribution for λ



Another Prior: Log-Uniform

Q: How long is a piece of string?

Another Prior: Log-Uniform

A: Twice the distance from the middle to one end.

The Log-Uniform Prior

- Useful for a parameter θ:

  a) known to be positive

  b) uncertain by orders of magnitude

$p(\theta)$ proportional to $1/\theta$.

Probability vs. Probability Density

```
lambda = seq(1, 100,
by=1)

prior = 1/lambda

prior = prior/sum(prior)

...

Z = sum(prior*likelihood)
```

```
step = 0.01

lambda = seq(1, 100,
by=step)

prior = 1/lambda

prior = prior/
(step*sum(prior))

...

Z =
step*sum(prior*likelihood)
```

- Let’s redo the problem using the log-uniform prior, and using probability densities (prior and posterior for λ) instead of probabilities.

Prediction

- Our original question was not “what is λ” but rather “what is the probability of an eruption in the next 50 years”

Prediction: Classical Approach (210)

- Step 1: use an estimator (single number guess) to estimate the value of the parameter

- Step 2: use the sampling distribution to make the prediction, assuming your parameter estimate was perfectly correct.

:-/

Prediction: Bayesian Approach (331)

- Step 1: get the posterior distribution for the parameter(s)
- Step 2: For all possible values of the parameter, calculate the probability you're interested in *if that value were true* (i.e. conditional on the parameter)
- Step 3: Calculate the *posterior expected value* of the results from Step 2.

:-)

This procedure is not *invented* - it is derivable from the sum rule.

The two results

- Classical

```
> 1 - dpois(0, lambda=0.0025*20)
[1] 0.04877058
```

Bayesian

```
> # Conditional probabilities
> prob = 1 - dpois(0, 0.0025*lambda)
> # Marginal probability
> sum(step*posterior*prob)
[1] 0.04871122
```

In this case, the results were very similar.
But that's not always the case.

Food for thought:

When would you expect the two methods to give different results?