Advanced Statistics and Data Analysis

Santiago Barreda
LSA Summer Institute 2019

General Topics

- Coding, running and interpreting Bayesian models.
  - ‘Regular’ model analogues.
  - But also ‘weirder’ stuff that is not possible or very difficult with traditional approaches.
- Writing up and presenting Bayesian analyses.
- Defending Bayesian methods against skeptics.
- Heavily influenced by Gelman, Kruschke, and McElreath.

Homework and Readings

- I will provide readings with background and additional information.
- We will use “Doing Bayesian Data Analysis (second edition)” by John Kruschke as a textbook.
- Homework will involve slight tweaks/extrapolations to the analyses we do in class.
- You will be evaluated on applied knowledge.

What I assume you know about stats

Minimal:

- Main effects and interactions
- Regression and ANOVA
- Fixed and random effects.
- P-values.
- Interpret model coefficients.
- Using LMER (or analogous)

Preferably

- Probability distributions (and likelihood).
- Understanding LMER (or analogous)

What I assume you know about R

Minimal:

- Read in and out of R.
- Know how to get specific columns/rows from matrices and dataframes.
- Basic plotting.
- Basic functions (e.g., c, cbind, seq)
- Understand Booleans.

Preferably

- Slightly-less basic functions (e.g., apply, tapply, aggregate)
- For loops.

Why Bayesian Modeling?

- It works whenever frequentist approaches work, and many times when they don’t.
- You get a lot more information from Bayesian models.
- Frequentist approaches offer off-the shelf, one-size-fits all hypothesis testing machines.
- Bayesian modelling let’s you build the model you want.

What is Bayesian Modeling?

- We will be discussing models that share two important characteristics:

1) Inference is based on the posterior distributions of the parameters in the model.

2) They are fit using Markov-Chain Monte Carlo methods

Basic Probability

$P(A) = \text{Probability that A is true}$

$P(B) = \text{Probability that B is true}$

$P(A \,\&\, B) = P(B|A) * P(A)$

$P(B \,\&\, A) = P(A|B) * P(B)$

Basic Probability

- How likely is a very tall person to play in the NBA?
- Here is some relevant information:
    - 100,994,367 males over 18 in the USA
    - 3,199 men over 6’ 10” in NBA
    - 486 Active NBA players
    - 88 players in the NBA are over 6’ 10”

Basic Probability

Marginal ('overall') probabilities

$P(T)$ = Probability that a man is over 6'10"

$P(B)$ = Probability of playing in the NBA

Joint probabilities

$P(T \;\&\; B) = \text{Probability of being tall AND playing in the NBA}$

$P(B \;\&\; T) = \text{Probability of playing in the NBA AND being tall.}$

Conditional ('if') probabilities

$P(T|B)$ = Probability of being tall <u>given</u> that you play in the NBA

$P(B|T)$ = Probability of playing in the NBA <u>given</u> that you are tall.

Basic Probability

$P(T)$ = 3199 / 100,994,367 = 0.000032

$P(B)$ = 486 / 100,994,367 = 0.0000048

NOT reversible!

$P(B|T)$ = 88 / 3199 = 0.028

$P(T|B)$ = 88 / 486 = 0.18

$P(T \,\&\, B) = P(B|T) * P(T) = 0.028 * 0.000032 = 0.00000087$

$P(B \,\&\, T) = P(T|B) * P(B) = 0.18 * 0.0000048 = 0.00000087$

Basic Probability

$$P(A \,\&\, B) = P(B \,\&\, A)$$

$$P(A|B) * P(B) = P(B|A) * P(A)$$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Bayes Theorem

Posterior Probability

Likelihood

Prior Probability

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Marginal Likelihood

Bayes Theorem

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

data → y; parameters → X

$$P(X|y) = \frac{P(y|X) * P(X)}{P(y)}$$

Bayes Theorem

data parameters

$$P(X|y) = \frac{P(y|X) * P(X)}{P(y)}$$

$$P(X|y) \propto P(y|X) * P(X)$$

Prior Probability

$$P(X|y) \propto P(y|X) * P(X)$$

- The prior probability is the probability that you would observe a parameter value, prior to experimentation.
- We can use informative, mildly informative, or uninformative/diffuse priors.
- Rather than being 'subjective', priors can make models more conservative and less prone to overfitting.

Likelihood

$$P(X|y) \propto P(y|X) * P(X)$$

- The likelihood is the probability that you would observe your data <u>conditional</u> on a given set of parameter values, and the model structure.
- The model structure constitutes a given hypothesis about the world (i.e., the effects represented in X are real).

Posterior Probability

$$P(X|y) \propto P(y|X) * P(X)$$

- The posterior probability of a belief is the ‘updated’ probability of the belief after some observation.
- The prior belief, combined with the likelihood yields the ‘new’, posterior belief.

Is this a Lion?

Is this a Lion?

- We can approach this question in two ways:

  1) Using null-hypothesis significance testing (NHST).

  2) Using the posterior probability of model parameters.

The NHST Approach

- In the NHST approach, we consider only the likelihood function.

$$P(X|y) \propto P(y|X) * P(X)$$

The NHST Approach

- H_1: That thing is a lion.
- H_0: That thing is not a lion. It's a dog.

$$P(X|y) \propto P(y|X) * P(X)$$

The data: the image you are seeing.

The model/parameters: the actual thing that generated the data.

The NHST Approach

- We calculate the probability of observing y under the null hypothesis.

- H_0: That thing is a dog.

$$P(\cancel{X}|y) \propto P(y|X) * P(\cancel{X})$$

P(that an image like that |would be a dog)

The NHST Approach

- When something is a dog, it almost never looks like that. Let's say $p = 0.001$

- We reject the null hypothesis and accept the alternative hypothesis.

- It's a Lion!

- H_1: That thing is a lion.
- ~~H_0: That thing is not a lion. It's a dog.~~

The NHST Approach

- Note that you DID NOT find the probability that it is a lion.

- You found the probability that you would observe that image if it is NOT a lion.

- We also can’t accept the null (that it is a dog), only reject it.

The NHST Approach

- If we had asked of this picture: Is this a lion?

  - H_1: That thing is a lion.
  - H_0: That thing is not a lion. It's a dog.

P(that an image like that |would be a dog)

- We might find that p=0.8.

- We reject H_1, but we cannot conclude that the image is of a dog.

The Bayesian Approach

$$P(X|y) \propto P(y|X) * P(X)$$

- In this approach we combine prior information into our inferences.
- We also directly estimate posterior probabilities and estimates, without convoluted hypothesis testing and rejection.

The Bayesian Approach

P(that it is a lion|given the image I am seeing)

$$P(X|y) \propto P(y|X) * P(X)$$

P(that an image like that |would be a lion)

P(that I would be seeing a lion)

- Investigating the posterior probability of a parameter usually directly gets at our question.

The Bayesian Approach

P(that it is a lion|*given the image I am seeing*)

$$P(X|y) \propto P(y|X) * P(X)$$

P(that an image like that |would be a lion)

P(that I would be seeing a lion)

- We can include information about the model parameters themselves.

The Bayesian Approach

$$P(X|y) \propto P(y|X) * P(X)$$

- Let’s say that the likelihood is 0.8.
- What should the prior be?

- In the quad let’s say its 0.0001.
- On an African safari let’s say its 0.9.

The Bayesian Approach

$$P(X|y) \propto P(y|X) * P(X)$$

- The posterior probability of concluding that this is a lion would be:

- 0.00008 in the quad.
- 0.72 on Safari

The Bayesian Approach

- Two things to note:

- The conclusion may vary based on the prior.
- We are directly investigating the question we want.

$$P(X|y)$$

$$\text{P}(\text{that it is a lion}|\textit{given the image I am seeing})$$

- Our posterior probability is actually the probability of the question we are asking. Not its inverse.

What is Bayesian Modeling?

- We will be discussing models that share two important characteristics:

  1) Inference is based on the posterior distributions of the parameters in the model.

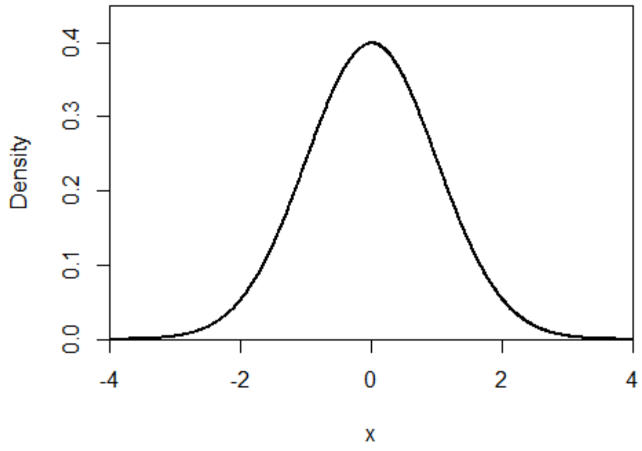
  2) They are fit using Markov-Chain Monte Carlo methods

Estimation via Maximum Likelihood

- Traditional approaches rely on maximum likelihood (ML) estimation.
- This estimates coefficient values and standard errors using point estimates based on the likelihood function.
- P-values may rely on the <u>asymptotical normality</u> of likelihood functions for inference.

Likelihood Functions

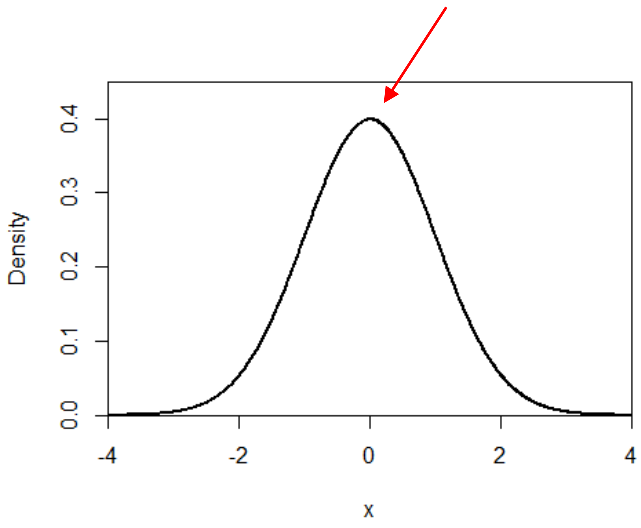
- Probability tells you about the plausibility of an event (i.e., observed data).
- Likelihood tells you about the plausibility of a parameter or hypothesis.



Likelihood Functions

- For fixed data, the likelihood function considers different parameter values.
- It returns a single likelihood for each combination of parameter values.

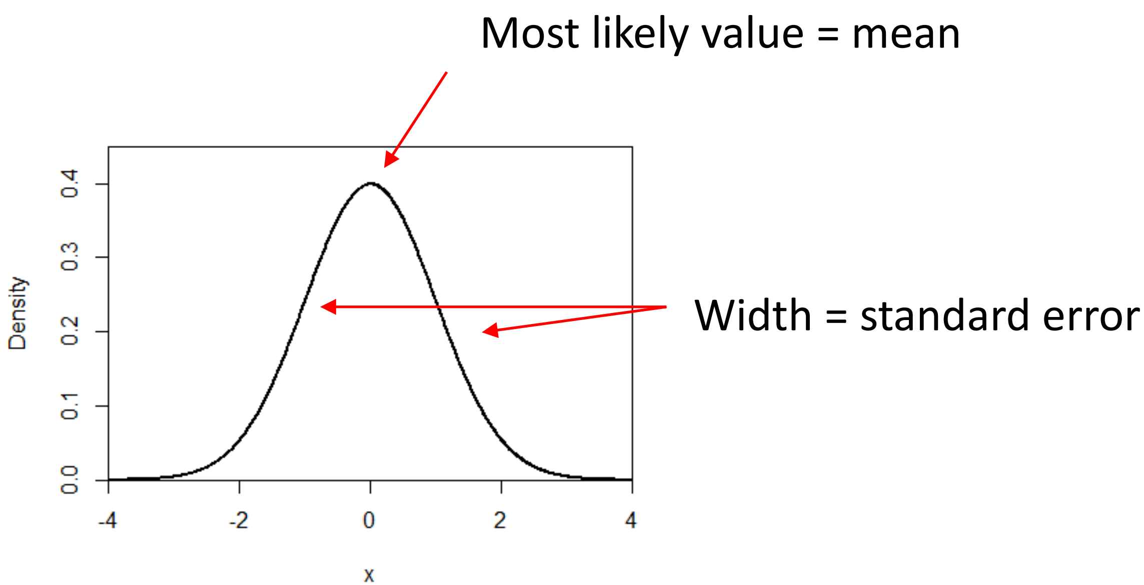
Most likely value



ML Estimation

- We want to estimate two things for each parameter:

  - its mean $\hat{\mu}$
  - its standard error $\hat{\sigma}_{SE}$.

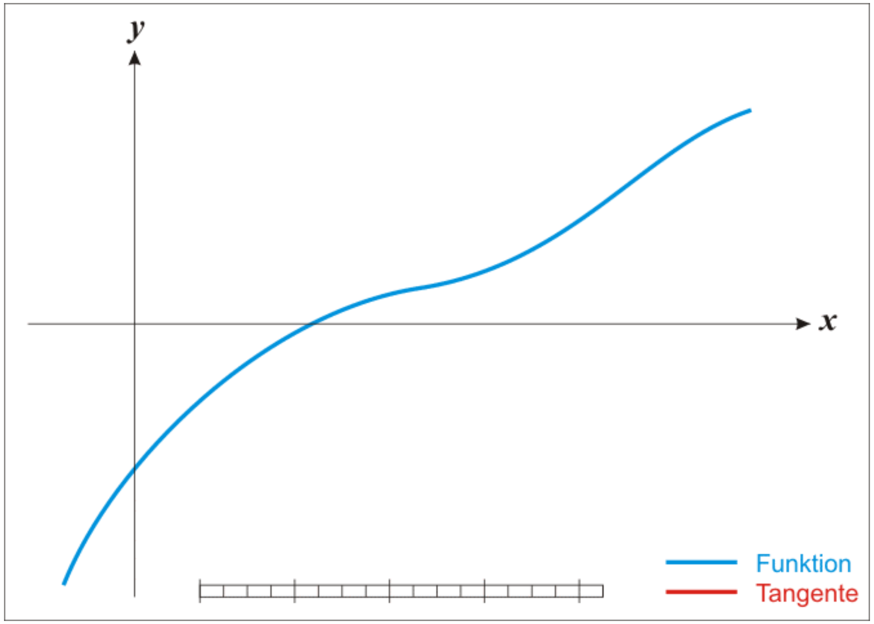


ML Estimation For 'Simple' Cases

- Under certain conditions, ML estimates of $\hat{\mu}$ and $\hat{\sigma}_{SE}$ can be estimated analytically (i.e., exactly).
  - E.g., the sample mean and variance are ML estimates of these parameters given a sample.
  - What does the above mean?

ML Estimation For ‘Complicated’ Cases

- Data with non-normally distributed residuals, non-linear problems, multilevel models.
- These must be estimated ‘numerically’.

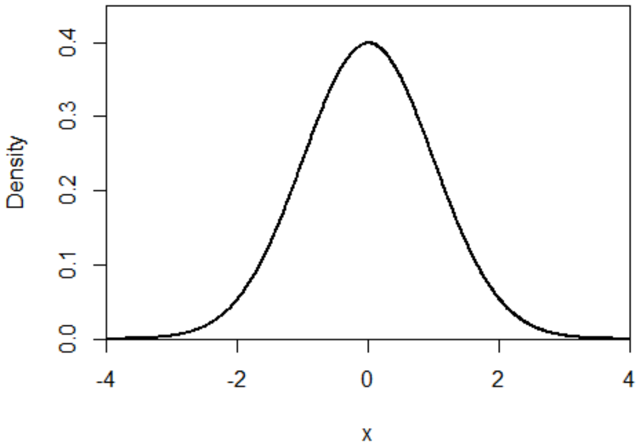


ML Estimation For ‘Complicated’ Cases

- $\hat{\mu}$ and $\hat{\sigma}_{SE}$ can be directly estimated using many different optimizers and the likelihood function.

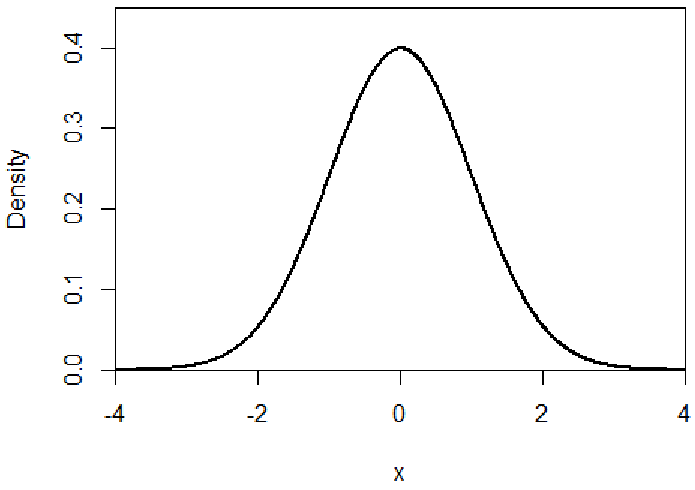
- The shape of the likelihood function approaches a standard normal as n approaches infinity.

- P-values based on Wald-tests are calculated by referring to a standard normal distribution.

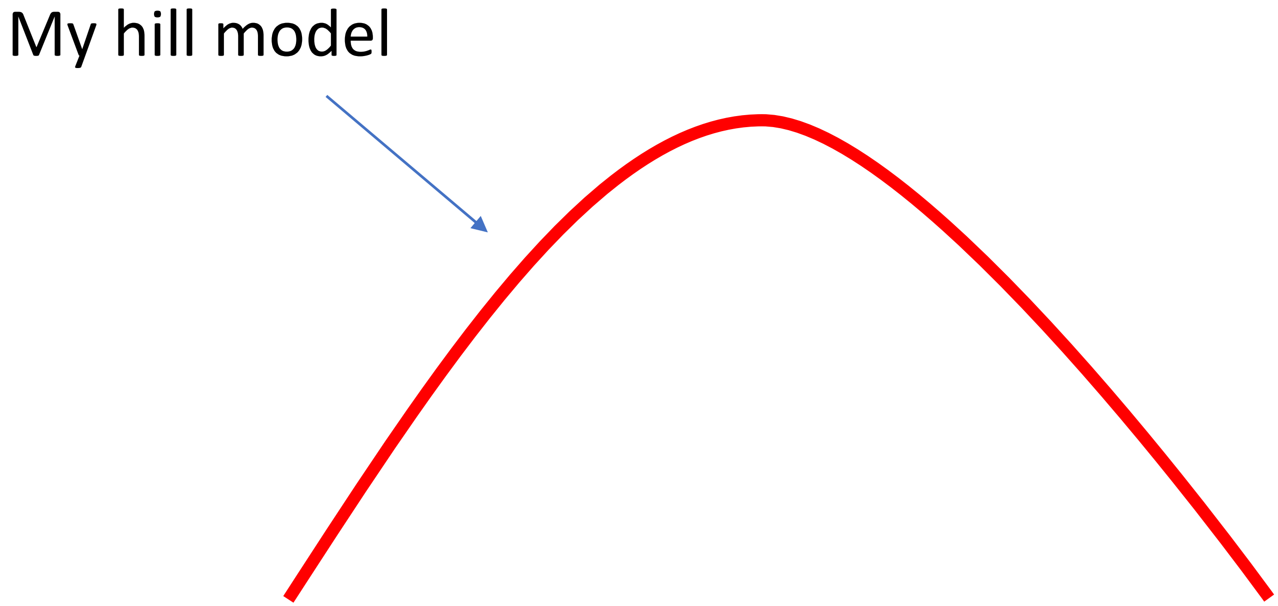


ML Estimation

- NHST based on ML estimation relies on point estimates of parameters.
- Point estimates combine with assumptions about the ‘shape’ of distributions to generate p-values.



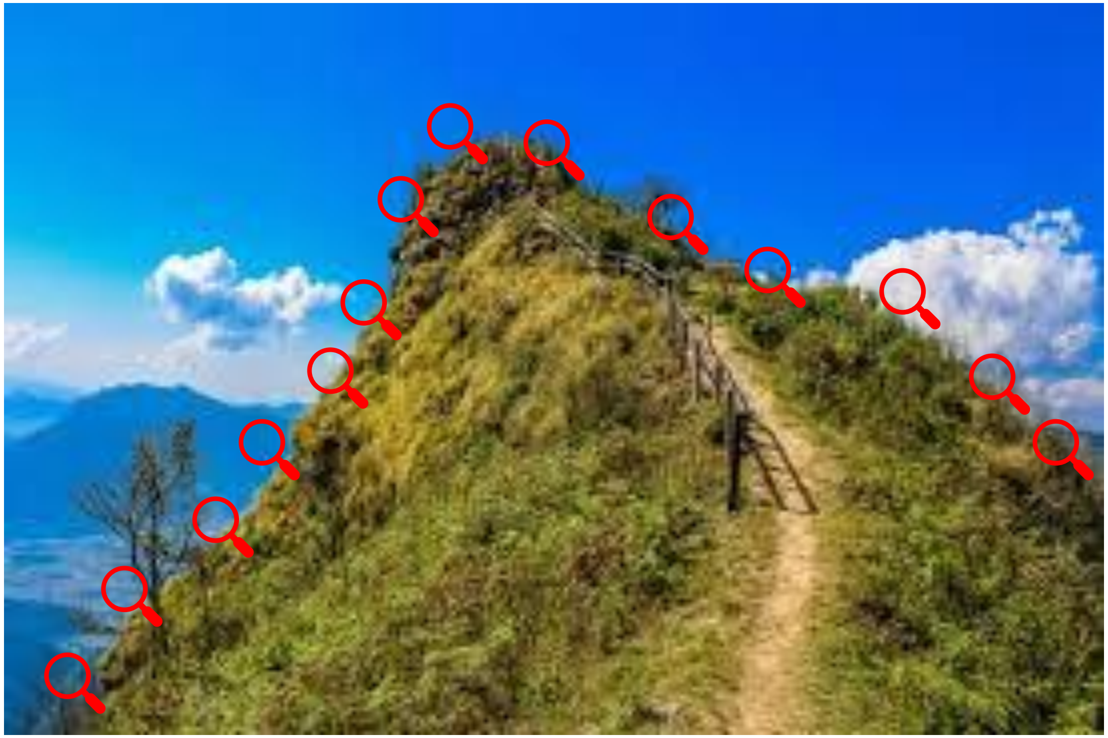
An Analogy



An Analogy

An Analogy

An Analogy



MCMC Estimation

- MCMC samples a given distribution by taking a series of discrete steps, ‘walking’ around the parameter space.
- Values are recorded at each location/step.
- Process continues as long as desired (there are practical time and storage limits).
- Several different ways to approaching the walking, and the acceptance/rejection of samples.

MCMC != Bayesian

- There's nothing particularly 'Bayesian' about MCMC.

- But relying on MCMC makes it possible to fit models that are not possible* with ML methods.

- This includes Bayesian models, but also multilevel models, non-linear regression problems, and models that have no classification.

Samplers

- MCMC lets us directly inspect the posterior distribution of parameters.

- We need a ‘sampler’: software that ‘walks around’ the posterior for us, collecting values.

- Several samplers available. Most popular are:

    - JAGS
    - Stan
    - PyMC

JAGS

- We are going to be using JAGS (Just another Gibbs sampler).
- It’s simple, fast, and works.
- Use of JAGS directly transfers to use of other samplers.