An Introduction to Bayesian Statistics

Robert Weiss

Department of Biostatistics
UCLA Fielding School of Public Health
robweiss@ucla.edu

Sept 2015

What Exactly is Bayesian Statistics?

- A philosophy of statistics.
- A generalization of classical statistics.
- An approach to statistics that explicitly incorporates expert knowledge in modeling data.
- A language to talk about statistical models and expert knowledge.
- An approach that allows complex models and explains why they are helpful.
- A different method for fitting models.
- An easier approach to statistics.

Prior to Posterior

Bayes Theorem

- Prior – which parameter values you think are likely and unlikely.
- Collect data.
- Data gives us *Likelihood* – which parameter values the data consider likely
- Update prior to
- Posterior – what values you think are likely and unlikely given prior info and data.
- Prior and posterior are both probability distributions
- Bayes Theorem:

$$\text{Posterior} = c * \text{Prior} * \text{Data Likelihood} \tag{1}$$

Prior Distributions

- Prior: Probability distribution summarize beliefs about the unknown parameters *prior* to seeing the data.
- In practice: a combination of expert knowledge and practical choices.
- Priors can be specified by
  - A point estimate (My systolic blood pressure is around 123), and
  - A measure of uncertainty (I might be in error by 3 points).
  - (Is that a maximal error, an SD, or perhaps 2$*$SD?) Need to decide (1 SD). And,
  - A practical choice about the density: normal distribution.
  - $N(123, 3^2)$.

Priors and Bayesian methods keep getting reinvented with other names

- penalized likelihood; penalized regression
- Random Effects Models (REMs)
- Generalized Linear Mixed Models (GLMMs)
- Hierarchical models
- multi-level models (MLM)
- Pitman estimator
- Shrinkage estimator
- Ridge regression
- Stein estimation
- James Stein estimators
- Regularization
- Best Linear Unbiased Prediction (BLUP)
- Restricted Maximum Likelihood (REML)
- Lasso

- Data is random before you see it, both classical and Bayesian statistics.
- After you see data, it is fixed in Bayesian statistics (not classical).
- Uncertainty in parameter values is quantified by probability distributions.
- Parameters have distributions before you see the data [Prior distribution].
- Parameters have updated distributions after you see the data [Posterior distribution].
- The conclusion of an analysis is a posterior distribution.
- Point estimates, sds, 95% CIs: summaries of the posterior distribution.

- 123 is the value I find most likely.
- 3 is my assessment of the standard deviation.
- 123 ± 3 is $(120, 126)$ is about a 68% interval.
- $(117, 129)$ is I expect my true blood pressure to fall in this interval 95% of the time.
- That 68% or 95% would be over all prior intervals that I specify for various quantities.

Collecting My Data

From CVS and from Kaiser

- I went into Rite-Aid and took 5 measurements in a row.
- Measures were: 125, 126, 112, 120, 116.

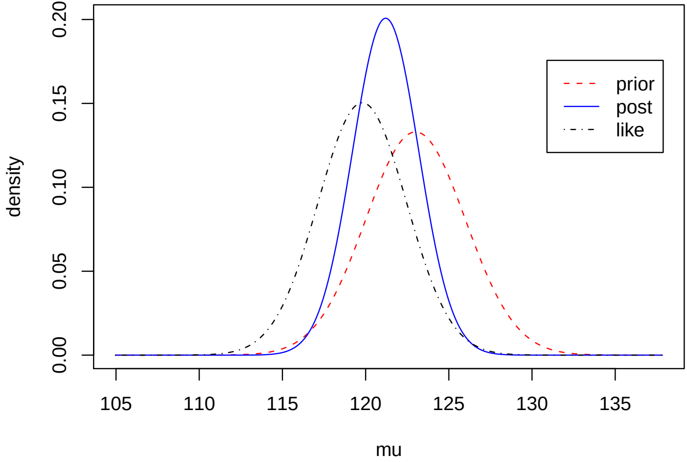
Collecting My Data

From CVS and from Kaiser

- I went into Rite-Aid and took 5 measurements in a row.
- Measures were: 125, 126, 112, 120, 116.
- Compare to: A Kaiser visit, a measurement of 115. I was surprised.
- At Kaiser: 115 was common 3 years ago.
- Last 5 Kaiser data measures were 127 (2010), 115 (2010), 117 (2007), 115, 116.

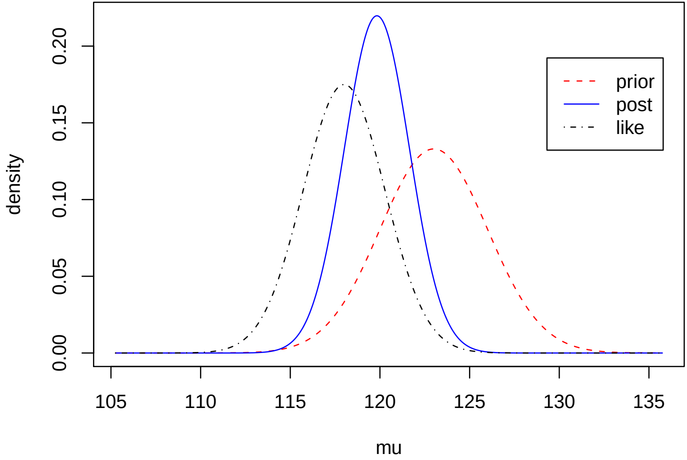
Graphical Results for Rob's Blood Pressure

Prior, Data Likelihood, Posterior, CVS



Graphical Results for Rob's Blood Pressure 2

Prior, Data Likelihood, Posterior from Kaiser



Interpreting the Figure

Rob's Blood Pressure

- Red curve is prior - what I believed before seeing the data.
- *X* axis values where red curve is highest are most plausible values *a priori*.
- Where red curve is low, values are less plausible.
- Black curve is *likelihood* –
- Where black curve is high are values supported by data.
- Blue curve is posterior – combination of prior and likelihood.
- Where blue curve is high are most plausible values given data and my prior beliefs.

Inferences for Rob, CVS data

Inference	Mean	SD	9% CI
Posterior	121.2	2.0	(117.2, 125.2)
Likelihood	119.8	2.7	(114.5, 125.1)
Prior	123	3	(117.0, 129.0)

Table: Likelihood gives the classical inference. Posterior is the Bayesian solution. Prior is what you guessed prior to seeing the data. Posterior sd is smaller than classical answer and posterior CI is narrower, ie more precise.

$$\bar{\mu} = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \bar{y} + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \mu_0$$

$\bar{\mu}$ – Posterior mean
μ_0 – Prior mean
τ – Prior sd
σ^2 – Sampling (data) sd
n – Sample size
$\bar{y}$ – Data mean

Posterior mean is a weighted average of the prior mean and the data mean.

Other Sources of Prior Information

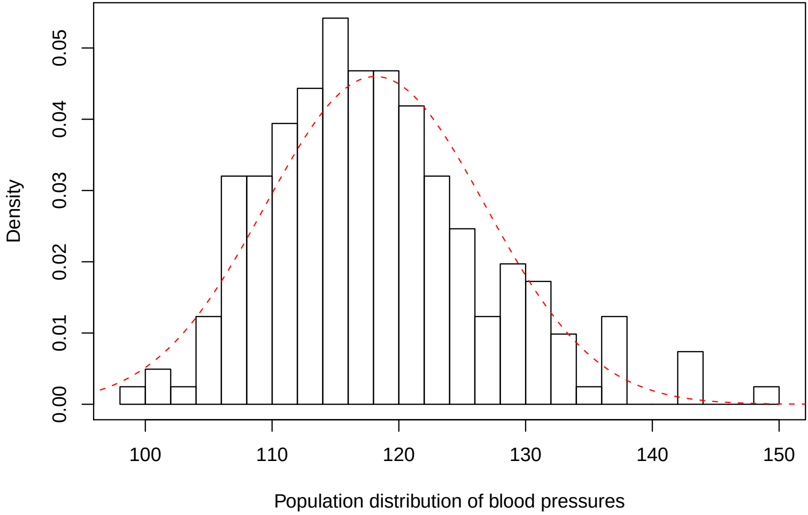
- In your analyses, *prior information* came from you, the experts on your own blood pressure.
- There are other places to get prior information.
- Knowledge of usual blood pressure values can be used as prior knowledge.
- For example, nurses' average SBP is 118.
- Nurses' population sd is 8.7.
- Sampling sd of SBP is 13.

UCLA Nurse's Blood Pressure Study

- An average of 47 measurements of Ambulatory Blood Pressure on
- 203 Nurses.
- Calculate average SBP for each nurse.
- Pretend this is their “true" average SBP.
- Average SBP is 118, sd of SBP is 8.7.
- SD of SBP measures around nurse's average is 13.
- Suppose we record a single SBP measurement of 140.
- Does the nurse likely have 140 SBP, or is it a somewhat high reading from someone with 125 SBP, or a somewhat low reading from someone with SBP=155?

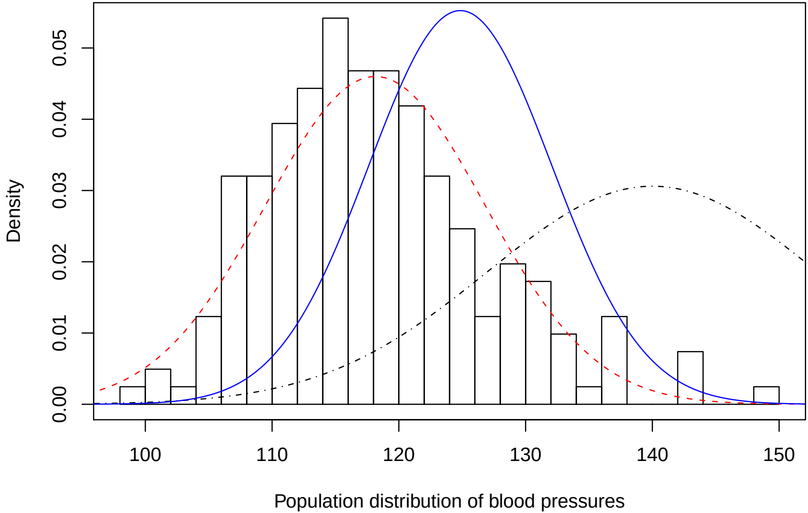
Nurse's Blood Pressure Study

The Population Distribution of Blood Pressures



Nurse's Blood Pressure Study

We observe SBP=140



Compare Bayesian Analysis to Classical Analysis

- Sample a single observation y_1 from one nurse in the data set.
- Model the true mean μ a priori as normal with a prior mean of 118 and standard deviation of 8.7.
- The sampling sd is 13.
- The classical estimate is just the value of the single measurement.
- The Bayes measurement is calculated as illustrated for the measurement of 140 and as you did in the homework.
- Repeat for all subjects.
- Which result is closer to the true value? Calculate root mean squared error, and count which statistical approach comes closer to true value more often.

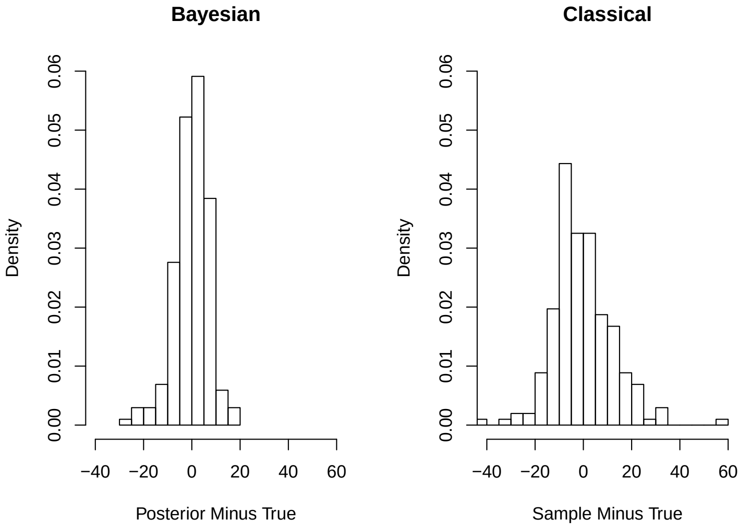
Bayesian Analysis Beats Classical Analysis

Badly

- Mean squared error
- Bayesian method – 51 units. (7.1)
- Classical method – 153 units. (12.4)
- Bayes method is 3 times more precise.
- Bayesian prior worth roughly 3 observations.
- Bayesian estimate closer 140 times.
- Classical estimate closer 63 times.

Nurse's Blood Pressure Study

Histogram of Bayesian And Classical Errors



- Bayesian errors are packed more closely in towards zero.
- Almost never larger than 20 units.
- Mostly between -10 and 10 units.
- Some classical errors are even as large as -40 and +60.
- Generally between -20 and 20, but roughly 10% outside that range.
- Bayesian analysis tends to eliminate/reduce ridiculous estimates and make those results more reasonable.

- Model all data from the Nurse's Blood Pressure Data set.
- Model is n=203 copies of our simpler blood pressure data model.
- Plus a model of the population of true blood pressures.
- SBP population mean($\approx$ 118) and sd$\approx$ 8.7 are unknown parms.
- Sampling sd (sigma$\approx$ 13) is estimated.
- Estimate all 203 individual nurse's average SBPs.
- A *random effects model*.
- The random effects are the individual nurses' (unknown) true SBP values.

- Classical statistics confidence interval
  - 95% confidence intervals constructed from repeated data sets will contain the true parameter value 95% of the time.
  - Throwing horseshoes: true parameter fixed, the data is random. Data attempts to throw a ringer capturing true parameter.
- Bayesian confidence (*posterior* or *credible*) interval
  - Probability *this interval* contains true parameter is 95%.
  - Data is fixed, parameter is unknown. Throwing a dart.
- Both: The parameter is fixed but unknown.

Confidence is a statement about the statistician. Bayesian posterior probability is a statement about the particular scientific problem being studied.

- Classical hypothesis test
  - *p*-value is the probability of observing a test statistic as extreme or more extreme assuming the null hypothesis is true.
- Bayesian hypothesis test.
  - The probability that the null (alternative) hypothesis is true.

The classical statement requires one more leap of faith: if the p-value is small, either something unusual occurred or the null hypothesis must be false. The Bayesian statement is a direct statement about the probability of the null hypothesis.

- Classical hypothesis test
  - H_0 is treated differently from H_A.
  - Only two hypotheses
  - H_0 must be nested in H_A.
- Bayesian hypothesis test.
  - May have 2 or 3 or more hypotheses.
  - Hypotheses are treated symmetrically.
  - Hypotheses need not be nested.

Example: Three hypotheses $H_1 : \beta_1 < \beta_2$, $H_2 : \beta_1 = \beta_2$, $H_3 : \beta_1 > \beta_2$ can be handled in a Bayesian framework. The result is three probabilities. The probability that each hypothesis is true given the data.

Orthodontic fixtures (braces) gradually reposition teeth by sliding the teeth along an arch wire attached to the teeth by brackets. Each bracket was attached to the arch wire at one of four angles: 0°, 2°, 4°, or 6°. Outcome is frictional resistance which is expected to increase linearly as the angle increases. Four different types of fixtures are used in this study.

- 4 groups.
- Interest in intercepts and slopes (on angle) in each group.
- Are intercepts same or different?
- Which slopes are equal?
- For any two slopes, which is higher? Lower?

ANOVA table

Treatments, Angle, Treatments by Angle

	F value	Degrees Freedom	p-value
Treatment	.25	3	.86
Angle	185	1	$<.0001$
Treatment$\times$Angle	7.43	3	$<.0001$

Table: Results from the analysis of covariance with interaction. The *p*-value for treatment is for differences between treatment intercepts, and the *p*-value for treatment$\times$angle is for differences between treatment-specific slopes.

Pairwise *p*-values

Are they significant? What about multiple comparisons?

		Treatment		
		2	3	4
Treatment	1	.03	.04	.04
	2		$<$.0001	.80
	3			.0003

Table: P-values for differences between treatment-specific slopes from the previous ANOVA.

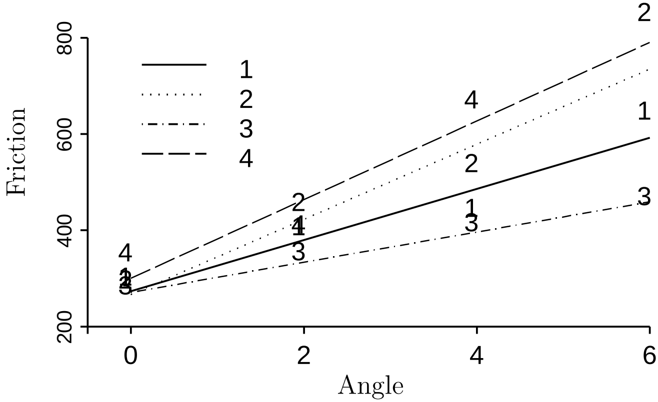
Orthodontic Example

Plot of Data

- Lines are fitted values from usual least squares regression.
- Points are average responses for each group 1 through 4 and each angle.
- Four intercepts and four slopes in plot.
- Which intercepts appear to be possibly equal, which slopes appear to possibly be equal?

Orthodontic Example

Data and fitted lines

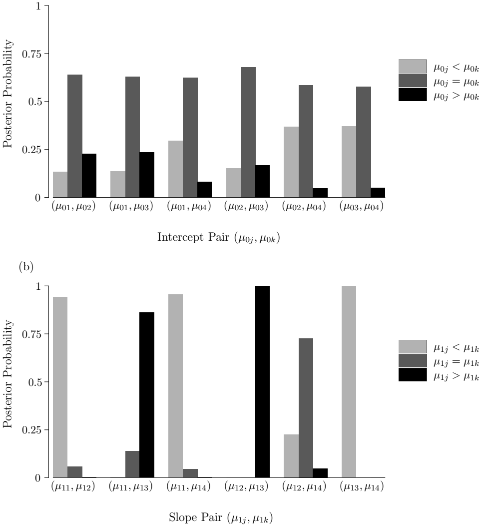


Alber and Weiss (2009). A model selection approach to analysis of variance and covariance. Statistics in Medicine, 28, 1821-1840.

- Data y_{ij} is the jth observation in group i and has angle x_{ij}.
- $y_{ij} = \alpha_i + \beta_i x_{ij} +$ error.
- Simple linear regression within group
- Different intercept α_i and different slope β_i in each group.
- Prior on intercepts is that any two groups may or may not be equal.
- Same for slopes: Any two slopes may or may not be equal.
- Results on next page.
- For each pair of intercepts or slopes, the figure plots the probability that the first slope (intercept) is less than the next slope (intercept), the probability that they are equal, the probability that the first is greater than the second.

Orthodontic Example

Results



Discussion of Orthodontic Results

- All pairs of intercepts have probability about .6 of being equal.
- Intercepts, groups 1, 2, 3: Roughly equal probabilities of one being larger or smaller than the other if not equal for groups.
- Modest probability that group 4 is larger than others, if not equal.
- Slopes: 1 less than 2, 4.
- Slope 3 less than 1, 2, and 4.
- Only slopes 2, 4 may be equal.