Bayesian Inference for Comparative Research

POSC 3410 – Quantitative Methods in Political Science

Steven V. Miller

Department of Political Science

Goal for Today

Introduce students to basic Bayesian inference.

"Frequentist" Inference and Research Design

You should be familiar with our discussion of research design and quantitative analysis to this point.

- Concepts, measures, variables, et cetera.
- Research design and the logic of control.
- Random sampling of the population (i.e. inferential statistics).
- Regression (linear or logistic) as estimating cause and effect.

“Frequentist” Inference and Research Design

We summarize inference as follows.

- If our regression coefficient is at least $\pm$ 1.96 standard errors from zero, we reject the null hypothesis.
- The regression coefficient is “statistically significant” in support of our hypothesis.

We know this because central limit theorem tell us this is true.

"Statistically Significant" Frequentist Inference

The simplicity of "statistically significant" is powerful and deceptive.

- When $z = 1.96$, we would observe a coefficient that far from zero five times in 100 random samples, on average.

Notice more carefully what's happening.

- We assume a fixed parameter (here: the null).
- We make statements of relative frequencies of extreme results under it.

"Statistically Significant" Frequentist Inference

Does that really make sense?

- Central limit theorem says it's true.

However, it depends on two things we routinely don't have.

1. Known population parameters
2. Repeated sampling

Probability and Frequentist Inference

Objectivist probability is the foundation for classical statistics.

Objectivist Probability

For example, the probability of a tossed coin landing heads up is a characteristic of the coin itself.

- By tossing it infinitely and recording the results, we can estimate the probability of a head.

Formally:

$$Pr(A) = \lim_{n \to \infty} \frac{m}{n}$$

. . . where:

- n: number of trials
- m: number of times we observe event A
- A: outcome in question (here: a coin landing heads up).

Objectivist Probability and Frequentist Inference

We can understand why classical statistics is **frequentist** and **objectivist**.

- Frequentist: probability is a long-run relative *frequency* of an event.
- Objectivist: probability is a characteristic of the object itself.
    - e.g. cards, dice, coins, roulette wheels.

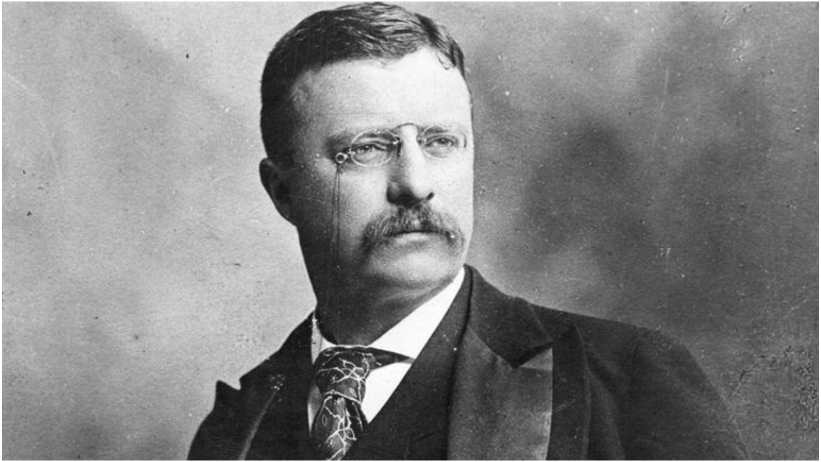
Bayesian Probability

Bayesian probability statements are states of mind about the states of the world and not states of the world, per se.

- It is a *belief* of some event occurring.
- It is characterized as *subjective* probability accordingly.

There are constraints, but nonetheless a substantial amount of variation allowed on probabilistic statements.

Bayesian Probability: An Unintuitive Application



What is the probability that Teddy Roosevelt is the 25th U.S. President?

Bayesian Probability

A Bayesian approach:

- What is my degree of belief that statement is true?

A frequentist approach:

- Well, was he or wasn't he?

Since there is only one experiment for this phenomenon, the frequentist probability is either 0 or 1?

- The phenomena is neither standardized nor repeatable.

Bayesian Probability

Even greater difficulties arise for future events. For example:

- What is the probability that Trump wins the White House in 2016?
- What is the probability of a terrorist attack in the U.S. in the next five years?
- What is the probability of a war between the U.S. and Russia?
- What is the probability ISIL takes over Damascus?

Bayesian Inference

These are all perfectly legitimate and interesting questions.

- However, frequentist inference offers no helpful answer.

Bayesian inference does offer a helpful route in **Bayes' theorem**.

Bayesian Inference

The probability of event A given B for a continuous space:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

With only two possible outcomes: A and ∼A

$$p(A|B) = \frac{p(B|A)p(A)}{p(B|A)p(A) + p(B| \sim A)p(\sim A)}$$

Bayesian Inference: An Illustration with Pregnancy Tests

Suppose a woman wants to know if she's pregnant.

- She acquires a name-brand test that purports to be 90% reliable.
    - i.e. if you're pregnant, you'll test positive 90% of the time.
- It gives false positives 50% of the time.
    - i.e. if you're not pregnant, you'll test positive 50% of the time.
- Suppose the probability of getting pregnant after a sexual encounter is $p = .15$
    - *Note*: this is just one number I found. I'm not that kind of doctor.

Bayesian Inference: An Illustration with Pregnancy Tests

Suppose the woman tested positive.

- She knows her test purports 90% accuracy in testing positive, given she is pregnant.
- *She wants to know if she's pregnant, given she tested positive*.

Bayesian Inference: An Illustration with Pregnancy Tests

We are interested in $p(\text{preg} \mid \text{test} +)$. We know the following:

- $p(\text{test} + \mid \text{preg}) = .90$
- $p(\text{preg}) = .15$ (conversely: $p(\sim\text{preg}) = .85$).
- $p(\text{test} + \mid \sim\text{preg}) = .50$.

We have this derivation of Bayes' theorem.

$$p(\text{preg}|\text{test } +) = \frac{p(\text{test } +|\text{preg})p(\text{preg})}{p(\text{test } +|\text{preg})p(\text{preg}) + p(\text{test } +| \sim \text{preg})p(\sim \text{preg})}$$

Bayesian Inference: An Illustration with Pregnancy Tests

We can now answer $p(\text{preg} \mid \text{test} +)$.

$$p(\mathrm{preg}|\mathrm{test}\ +) = \frac{(.90)(.15)}{(.90)(.15) + (.50)(.85)} = \frac{.135}{.135 + .425} = .241$$

This is far from the belief you'd get from "90% accuracy" and a single positive test.

Posterior Probability

However, this quantity is important for Bayesians in its own right: a **posterior probability**.

- It's an updated probability of event A (being pregnant) after observing the data B (the positive test).
- She has a prior belief of being pregnant ($p = .15$), which is now updated to $p = .241$.

Does this mean the woman is really not pregnant?

Posterior Probability

She should take the updated posterior probability as "prior information" (i.e. $p(\text{preg}) = .241$, and $p(\sim\text{preg}) = .759$) and take another test.

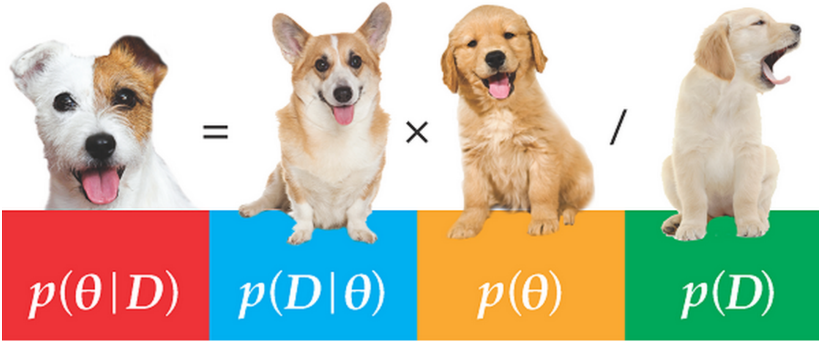
- Assume, again, she tested positive.

$$p(\text{preg}|\text{test} +) = \frac{(.90)(.241)}{(.90)(.241) + (.50)(.759)} = \frac{.216}{.216 + .379} = .363$$

Posterior Probability

In other words, keep repeating tests until you're convinced, but don't begin agnostic each time.

Bayesian Inference



Bayesian inference uses this uncontroversial imputation of conditional probability as a foundation for statistical inference.

Bayesian Inference

We say the posterior distribution (i.e. likelihood of the unknown parameter given the data) is *proportional to* the likelihood of the data multiplied by our prior expectations of it.

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

…where $\propto$ means "is proportional to" in symbol form.

Western and Jackman (1994)

Nonstochastic and Weak Data

Two properties of comparative research violate foundations for frequentist inference.

1. Nonstochastic data (i.e. non-random DGP)
2. Weak data

Nonstochastic Data

Frequentist inference assumes data are generated by a repeated mechanism like a coin flip (hence: RDGP).

- A sample statistic is just one possible result from a draw of a probability distribution of the population.

Nonstochastic Data

However, political scientists can define the sample on the population. Examples:

- OECD countries
- Militarized interstate disputes
- Supreme Court decisions

You know what this is. We called this a **census**.

Nonstochastic Data

Frequentist inference is inapplicable to the nonstochastic setting.

- If we took another random draw, we'd get the exact same data.
- "Updating" the data doesn't generate a new random sample.
- Appeals to a "superpopulation" don't help either.

Weak Data

This takes on two forms in political science research.

1. Small n
2. Collinearity

If the population of interest is "advanced industrial societies", our n is limited to 15 to 21.

- We run out of degrees of freedom quickly when adding controls.

Weak Data

The issue of **multicollineary** also arises in weak data with small n.

- This is when two predictors are so highly correlated that their estimated partial effects are uninformative.

This is relevant to a debate Western and Jackman address: what accounts for the percentage of the work force that is unionized?

- Wallerstein: size of civilian labor force (-).
- Stephens: industrial concentraion (+).

Weak Data

Problem: both are highly collinear ($r =$ -.92).

- In normal regression, one has to be dropped.
- We can still estimate this in Bayesian regression.

Using Prior Information

TABLE 1

Prior Means (Standard Deviations) and Their Substantive Interpretations for Bayesian Regression Analysis of Union Density

VARIABLE	PRIOR	SUBSTANTIVE INTERPRETATION
Wallerstein's priors		
Left government	.3 (.15)	One year of left-wing government increases union density by about 1 percentage point. A year of left-wing government may increase union density by as much as 2 percentage points of union density, but its effect is almost certainly not negative.
Logged labor-force size	−5 (2.5)	Doubling the size of the labor force would reduce union density about $\ln(2) \times 5 \approx 3.5$ percentage points. This increase in labor-force size may generate a union decline as big as 7 percentage points, but a growing labor force is unlikely to increase union density.
Economic concentration	0 (10^6)	The diffuse prior indicates that the researcher has no strong prior beliefs about the sign or magnitude of an effect. When the explanatory variables are uncorrelated, the diffuse prior yields posteriors that are approximately given by the sample data.
Stephens's priors		
Left government	.3 (.15)	Like Wallerstein's prior, one year of left-wing government increases union density by 1 percentage point.
Logged labor-force size	0 (10^6)	Diffuse prior.
Economic concentration	10 (5)	If economic concentration were to increase by 100% in relation to the United States, union density would increase by 10 percentage points. This increase in the concentration ratio may generate a density increase as large as 20 percentage points, but any increase in concentration is unlikely to decrease union density.

Note: Left government is measured by Wilensky's (1981) cumulative index of left-wing government; logged labor-force size is the natural log of the size (in thousands) of the dependent labor force in the year that union density is measured; and economic concentration is measured by the four-firm concentration ratio, in proportion to the United States.

Using Prior Information

Notice what's happening with our prior information.

- Wallerstein and Stephens agree on the effect of left governments.
- They disagree on the two other variables.

Using Uninformative Priors

Table 2 provides posterior distributions with uninformative priors.

- When we do this, we allow the data from the sample to have a larger effect over the ensuing posterior distribution.

Using Uninformative Priors

TABLE 2

Posterior Distributions with Noninformative Prior Information in a Regression Analysis of Union Density

INDEPENDENT VARIABLES	MEAN (S.D.)	5TH PERCENTILE	95TH PERCENTILE
Intercept	97.59 (57.48)	3.04	192.14
Left government	.27 (.08)	.15	.39
Size	−6.46 (3.79)	−12.70	−.22
Concentration	.35 (19.25)	−31.32	32.02

Note: These results are equivalent to the ordinary least squares estimates ($N = 20$)

Using Uninformative Priors

We see that the effects of left governments and logged labor force size are significant.

- Prima facie, Wallerstein is right.
- The industrial concentration variable is insignificant.

Using Informative Priors

In the interest of brevity, let's focus on just Table 3.

- We are looking at the regression results using both sets of prior information.

Using Informative Priors

TABLE 3

Posterior Distributions with Stephens's and Wallerstein's Informative Priors in a Regression Analysis of Union Density ($N = 20$)

INDEPENDENT VARIABLES	MEAN (S.D.)	5TH PERCENTILE	95TH PERCENTILE
Wallerstein's prior			
Intercept	82.43 (32.83)	28.42	136.43
Left government	.28 (.07)	.17	.39
Logged labor-force size	−5.44 (2.09)	−8.87	−2.00
Economic concentration	4.87 (12.41)	−15.54	25.28
Stephens's prior			
Intercept	70.82 (19.87)	38.13	103.51
Left government	.27 (.07)	.16	.38
Logged labor-force size	−4.79 (1.77)	−7.70	−1.88
Economic concentration	9.38 (4.84)	1.42	17.34

Interpreting Table 3

Using Wallerstein's priors:

- Posterior estimates for left-wing governments remain precise.
    - Actually gain a little precision too.
- Prior information makes confidence interval for labor-force size much less diffuse.
- No effect of industrial concentration.

Interpreting Table 3

Using Stephens' priors:

- Same posterior estimates for left-wing governments.
- Labor-force size estimate still significant, though magnitude decreases.
- Significant effect of industrial concentration.
  - But notice: we had prior beliefs about that effect!

The data we ultimately observed don't discount the effect of industrial concentration if you build in the prior belief.

Conclusion

Bayesians highlight how many liberties we can take with our research design if we're not careful.

- A census (a non-random DGP) does not permit conventional statistical inference.
- Collinearity magnifies problems of weak data.

Importantly, why start agnostic of the population parameter if we do not have to do this?

- If you have prior information, use it.

Table of Contents