Theoretical Underpinnings

A Gentle Introduction to Bayesian Analysis with Applications to QuantCrit
ASHE Workshop

Alberto Guzman-Alvarez
Taylor Burtch
Benjamin Skinner

18 November 2023

Bayesian statistics is all about probabilities. Before jumping into our hands-on coding activity, we'll spend just a few minutes discussing the basic probability relationships underlying Bayesian statistical analysis.

A little bit of probability

For Bayesian statistics, there are three basic probability types that are useful to know, both as they are defined and as they relate to one another.

Probability types

- Marginal probability: $P(A)$
- Joint probability: $P(A, B)$
- Conditional probability: $P(A \mid B)$

When a **joint probability** is comprised of independent variables — like coin flips — we can simply decompose it into the product of the **marginal probabilities**:

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

However, if one set of variables depend on the other set, then it's not quite as straightforward. An example of this might be: what are your odds of correctly guessing the correct number between 1 and 10 after you've already heard X wrong guesses (your odds change depending on how many incorrect guesses you get to hear).

When there is *conditional dependence*, then the **joint probability** is the **conditional probability** of the first variable, A, times the **marginal probability** of the second variable, B:

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

Bayes Theorem

Knowing these relationships, we can quickly derive **Bayes' Theorem**, which is:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Derived:

$$
\begin{aligned}
P(A,B) &= P(A,B) && \text{identity} \\
P(A \mid B)P(B) &= P(B \mid A)P(A) && \text{condition for both A and B} \\
P(A \mid B) &= \frac{P(B \mid A)P(A)}{P(B)} && \text{divide by P(B)}
\end{aligned}
$$

Okay great! But what do we do with this? How exactly is Bayes' Theorem useful?

Priors, likelihoods, posteriors

Bayes' Theorem represents a way to incorporate prior information into current probability calculations. We don't have to pretend we're brand new here — we know things! Bayes gives us a formal way to use this knowledge.

To make this interpretation of Bayes a little clearer, let's change the notation just a bit. Instead of A and B, which are vague, we'll use X and θ:

- X: knowns (e.g., data)
- θ: unknowns (e.g., probabilities/parameters)

which gives us,

$$
P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)}
$$

In most applied work, we can drop $P(X)$, which leaves us,

$$
\underbrace{P(\theta \mid X)}_{posterior} \propto \underbrace{P(X \mid \theta)}_{likelihood} \cdot \underbrace{P(\theta)}_{prior}
$$

which is read as, *the **posterior** is proportional to the **likelihood** times the **prior**.*

- **Prior**: $P(\theta)$
- **Likelihood**: $P(X \mid \theta)$
- **Posterior**: $P(\theta \mid X)$

In plain language, we have existing beliefs about $P(\theta)$ that we modify with data, X, to produce new beliefs, $P(\theta \mid X)$. How much our existing beliefs change depends on a combination of how strongly we hold them. If we have strong prior beliefs, no data will really change them — our beliefs won't be very different. Conversely, if we have weak prior beliefs, our updated beliefs will be mostly a function of what we observe in our data.

Comparison to frequentist statistics

Most quantitative work in education (and in social sciences more generally) falls under the frequentist paradigm. There are historical reasons for this, both philosophical and technological. Philosophically, Bayesian approaches have been accused of being too subjective (the Bayesian retort is that frequentist approaches contain subjective elements as well — they just aren't formally incorporated into the analysis). Technologically, Bayesian posteriors can be difficult to directly compute except for very simple (read: boring) problems. It's only been with the rise of modern computing power that Bayesian approaches for interesting applied problems have been possible.

Briefly, for those trained in frequestist (likely econometric) paradigm, here are a few differences between frequentist and Bayesian statistical approaches to applied work:

	Frequentist	Bayesian
X (Data)	Random	Fixed
θ (Parameters)	Fixed	Random
$\hat{\theta}$ (Output)	Single value	Distribution of values
Error for $\hat{\theta}$	Computed using asymptotic formula	Computed directly from distribution
Interpretation	Values that make data most likely	Most likely values given data
Statistical significance	Binary decision rules (e.g., p values)	Direct probabilistic decision

Applicability to QuantCrit

Key benefits of Bayesian approach for applied QuantCrit analyses:

1. Clear incorporation / acknowledgment of prior (subject) beliefs
2. Ability to provide estimates using small data sets
3. Ability to provide estimates for small groups that otherwise might be dropped
4. Provide estimates that are more easily interpreted by stakeholders, data or aggregated owners, and participants with the purpose of supporting actionable, antiracist, social justice-oriented policy

To make a point that we will return to in the next module: performing Bayesian analyses does not, in and of itself, mean one is working within the QuantCrit paradigm or any critical paradigm. Deep engagement with critical theories, frameworks, and positionalities is also required. We see Bayesian analyses as a specific set of tools that lend themselves well to a critical approach. Because they are less often taught in quantitatively-focused education research methods courses, we hope this workshop provides a short introduction. It remains up to the researcher, however, to interrogate these tools and the results they provide with same level of rigor demanded by critical frameworks of all quantitative approaches.