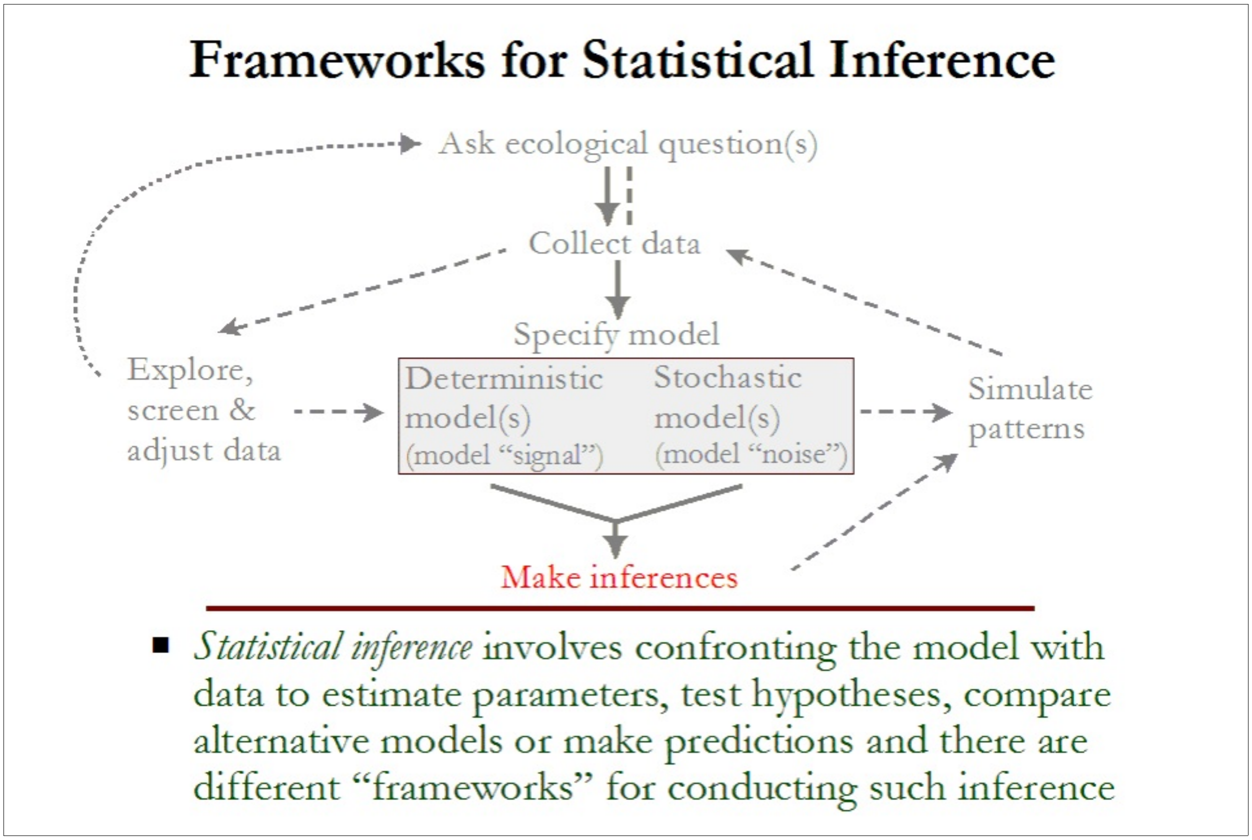
Analysis of Environmental Data

Chapter 6a. Conceptual Foundations:

Frameworks for Statistical Inference



1. Frameworks for statistical inference

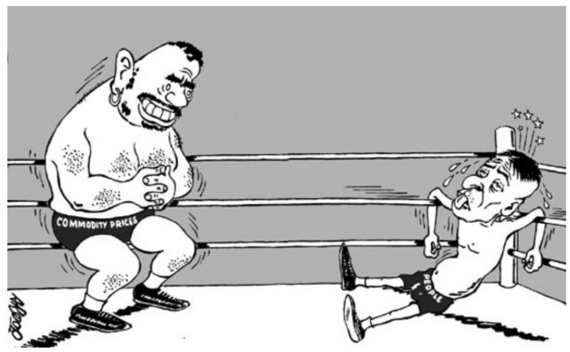
Recall that the general purpose of statistical inference is to make statements about processes extrinsic to the data – the underlying environmental system. More specifically, some attempt is made to describe the underlying population(s) from which the sample was drawn; i.e., to estimate parameters of an approximating model of the underlying environmental process, to test specific hypotheses about the underlying environmental process, to chose among alternative explanations of the underlying environmental process, or to make predictions for samples not yet collected. Not surprisingly, there are different frameworks or approaches for conducting such inference.

Statistical Inference Frameworks...

Parametric versus nonparametric inference

- *Parametric inference...* inference based on statistical models that assume a stochastic component based on a known probability distribution
- *Nonparametric inference...* inference based on statistical models that do not assume any particular error distribution

Strong inferences Vs. Weak inferences



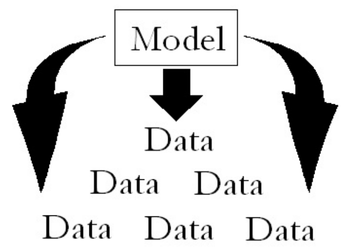
Parametric inference.–In modern statistics there are two major conceptual frameworks (or paradigms) for conducting parametric statistical inference – classical *frequentist* and *Bayesian*. Both of these frameworks require statistical models to assume a known probability distribution for the stochastic component, which makes them "parametric". *Maximum likelihood* is sometimes described as a third framework, but it is really just a particular approach within the general frequentist framework. The distinction between these frameworks is sometimes blurred, since maximum likelihood and Bayesian methods are both based on the *likelihood* – the probability of the observed data given a particular model or choice of parameters. Thus, likelihood serves as a bridge between the frequentist and Bayesian frameworks.

Nonparametric inference.–Sometimes it is not possible to assume any particular error distribution for the model and "nonparametric" statistical methods must be used. Inferences derived from nonparametric methods are generally much weaker than those from parametric methods, because without a probability distribution for the error it is difficult to conceive of the statistical model as a data-generating mechanism for the underlying environmental system. For example, there is no easy way to simulate data from a model without specifying a probability distribution for the error. Moreover, even if we can estimate the parameters of the model, without a probability distribution it is impossible to say whether these are the most likely or probable values for the underlying population. In fact, nonparametric inference really isn't a conceptual framework or paradigm for conducting statistical inference at all, it's more like the lack of an inference framework. Nevertheless, we include it here as a "framework" because there are many occasions when nonparametric methods are useful or required.

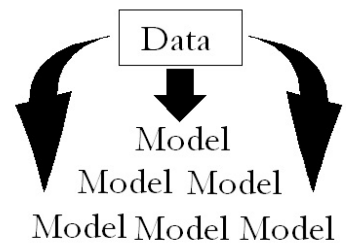
Statistical Inference Frameworks...

Philosophical distinction between parametric frameworks

- *Frequentist...* what is the probability of observing the sample data given the fixed parameters, and finds the estimates that would make the data the most likely outcome under repeated sampling



- *Bayesian...*what is the probability of the model (parameters) given the observed data (and prior expectations), and finds the population parameters that are most probable

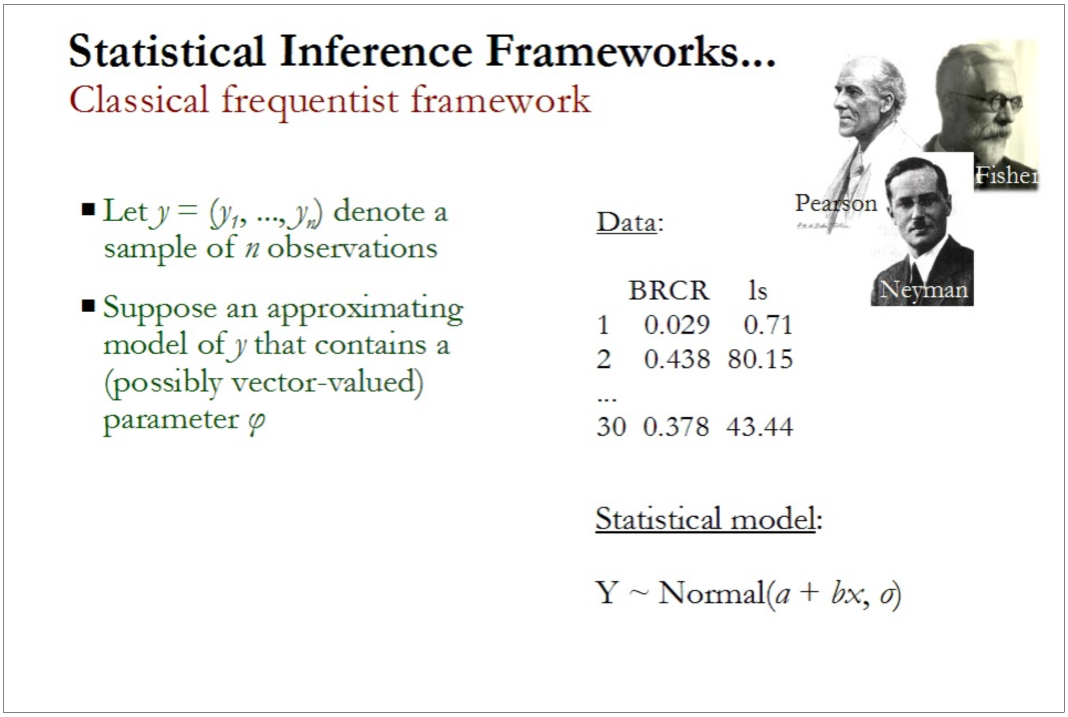


2. Philosophical distinction between parametric frameworks

The frequentist and Bayesian inference frameworks can be thought of as more than mere methodological frameworks, but as philosophies as well.

Frequentist.–The frequentist generally believes that there is a true underlying model (and parameters) that defines the way the environmental system works, which we cannot perfectly observe? Consequently, the parameters are assumed to be fixed but unknown, while the sample data are viewed as a random outcome of the underlying model; i.e., an imperfect representation of the underlying truth. The frequentist asks: what is the probability of observing the sample data given the fixed parameters, and finds the values (estimates) of those parameters that would make the data the most likely (frequently occurring) outcome under repeated sampling of the system (if one were able to repeatedly sample the system).

Bayesian.–The Bayesian, on the other hand, generally believes that the only knowable truth is the sample data itself and therefore does not worry about whether there are true fixed parameters or not. Consequently, the sample data is assumed to be true (since it is observed), while the model parameters are viewed as random variables. Accordingly, the Bayesian asks: what is the probability of the model (parameters) given the observed data (and prior expectations), and finds the population parameters that are most probable.



3. Classical frequentist inference framework

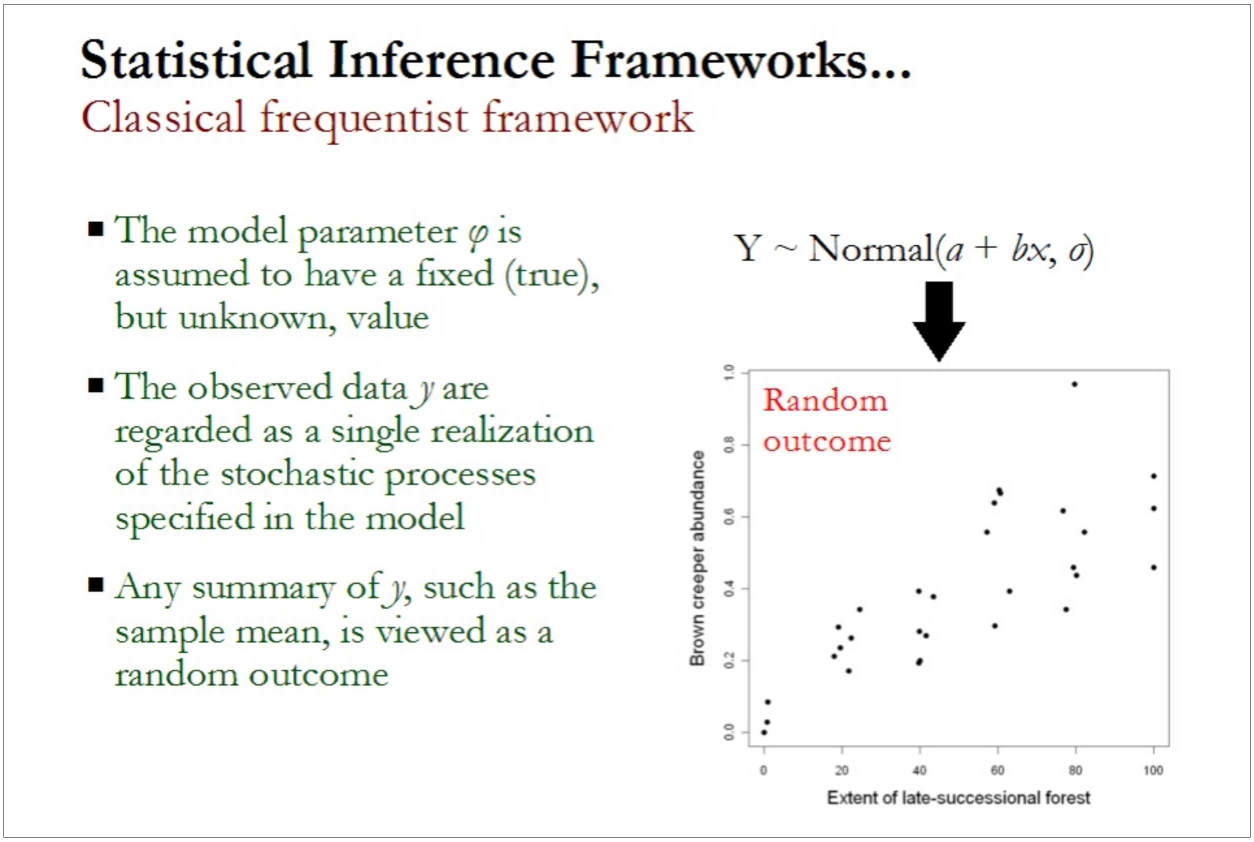
Classical frequentist inference (due to Fisher, Neyman, and Pearson) is simply one of the ways in which we can make statistical inferences and is the one that is typically presented in introductory statistics classes. The essence of the approach is as follows:

- Let $y = (y_1, ..., y_n)$ denote a sample of n observations.
- Suppose an approximating model of y that contains a (possibly vector-valued) parameter φ.

Example: We illustrate the approach using a simple ecological model based on the now familiar Oregon birds data set. For this example, let's examine the relationship between brown creeper abundance and the extent of late-successional forest across 30 subbasins in the central Oregon Coast Range based on the real data collected. For now, we will ignore the real data and simply simulate some data based on a particular statistical model. Let's assume a linear model (deterministic component) with normal errors (stochastic component), which we can write as:

$$\mathrm{Y} \sim \mathrm{Normal}(a + bx, \sigma)$$

which specifies that Y (brown creeper abundance) is a random variable drawn from a normal distribution with a mean $a + bx$ and standard deviation σ. In this notation, a and b are parameters of the deterministic linear model (intercept and slope, respectively), x is data (%late-successional forest), and σ is a parameter of the stochastic component of the model (the standard deviation of the normally distributed errors). This means that the i^{th} value of Y, y_i, is equal to $a + bx_i$ plus a normally distributed error with mean equal to the linear model and variance σ^2.



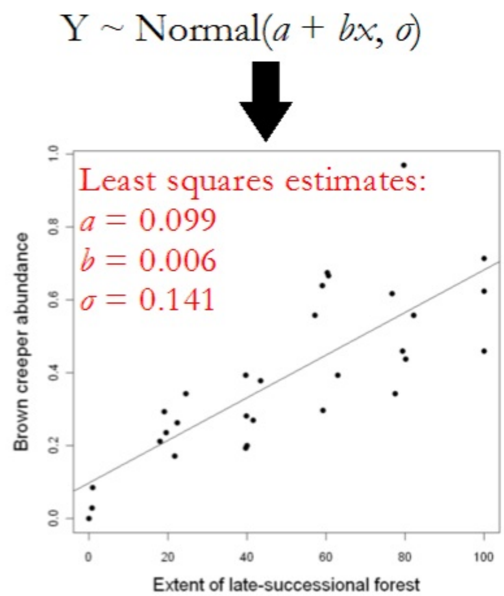
- The model parameter φ is assumed to have a fixed, but unknown, value.
- The observed data y are regarded as a single realization of the stochastic processes specified in the model.
- Any summary of y, such as the sample mean $\bar{y}$, is viewed as a random outcome.

<u>Example</u>: The parameters of our statistical model, a, b and σ, are assumed to have true fixed, but unknown, values, which we cannot perfectly observe. Our sample data are considered a single random outcome of that underlying model. Summary descriptive statistics such as the mean of brown creeper abundance or the Pearson's correlation coefficient between brown creeper abundance and %late-successional forest are random outcomes as well.

Statistical Inference Frameworks...

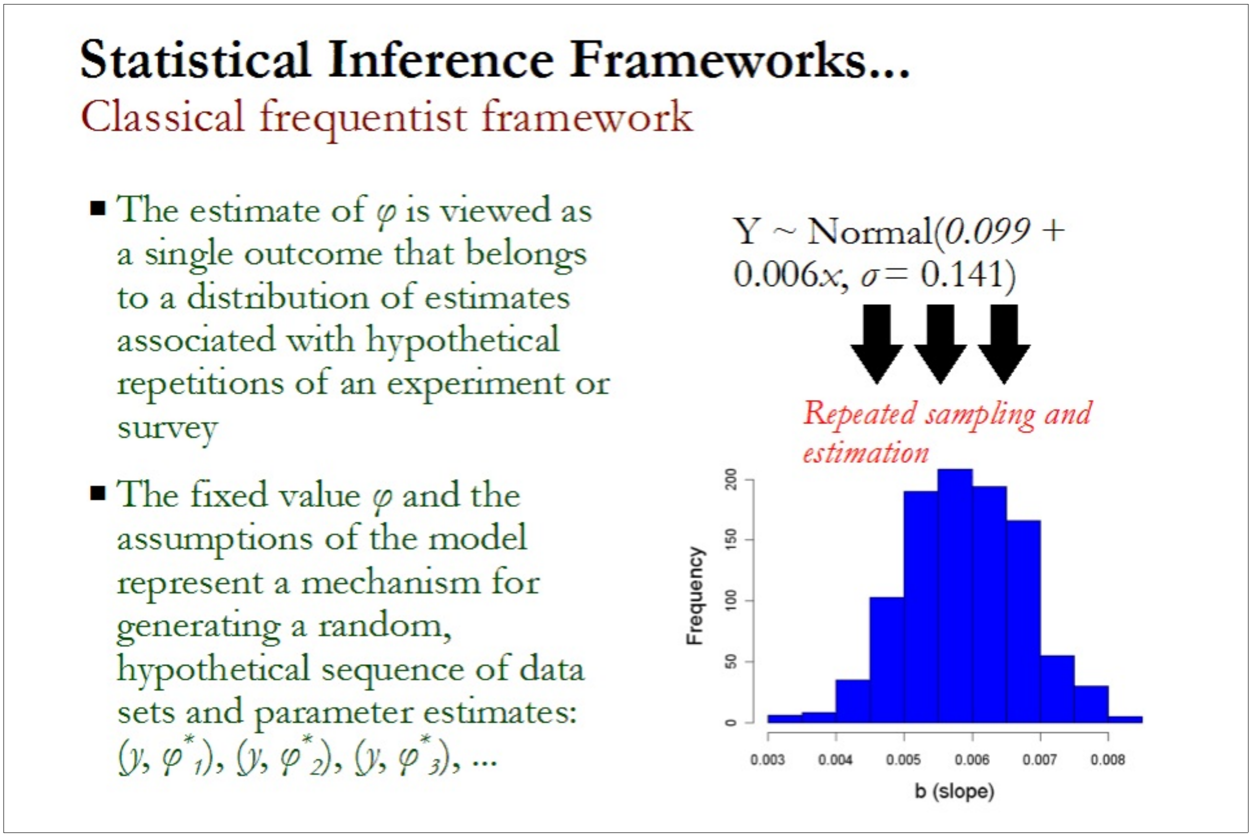
Classical frequentist framework

- A procedure for estimating the value of φ is called an *estimator* and the result of its application to a particular data set yields an *estimate* of the fixed parameter φ
- The *estimate* is viewed as a random outcome because it is a function of y, which is also regarded as a random outcome



- A procedure for estimating the value of φ is called an *estimator* and the result of its application to a particular data set yields an *estimate* of the fixed parameter φ.
- The *estimate* is viewed as a random outcome because it is a function of y, which is also regarded as a random outcome.

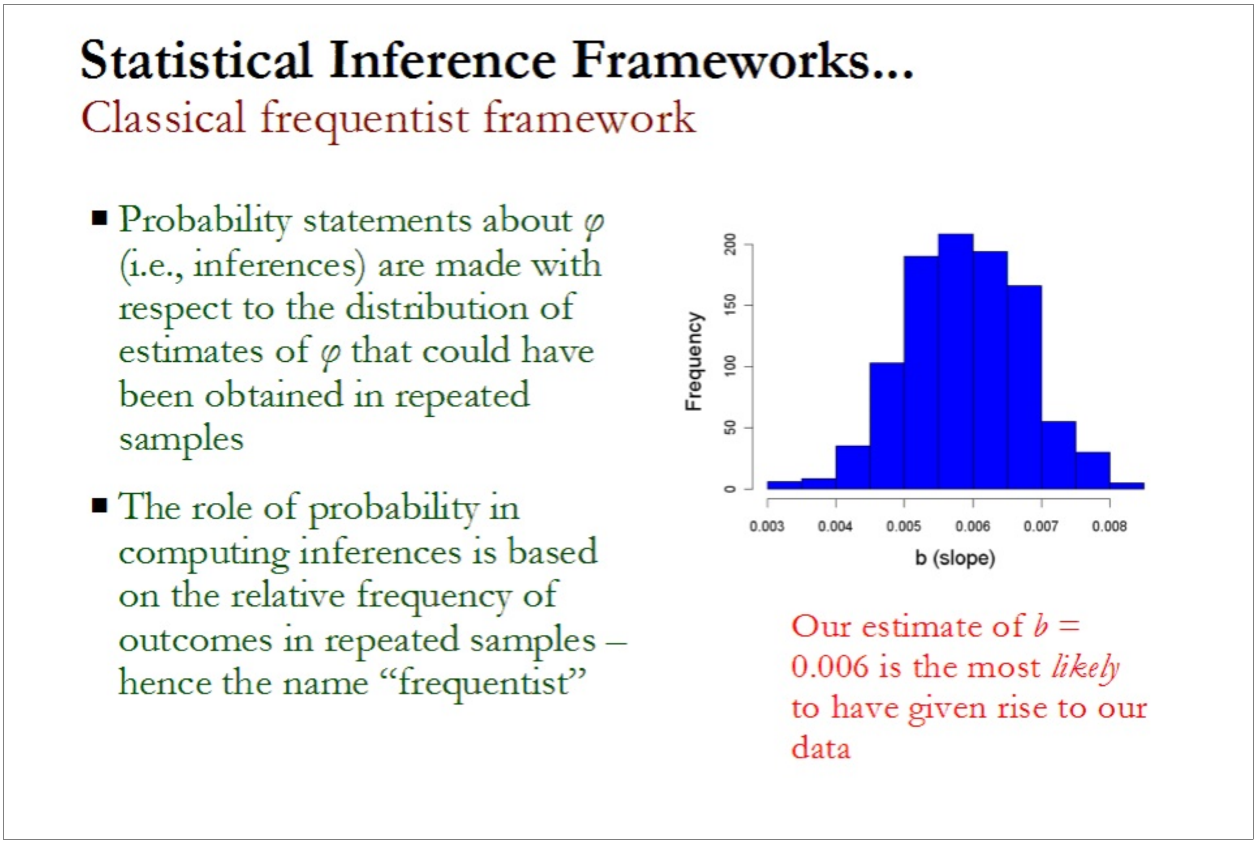
<u>Example</u>: Here we used the method of ordinary least squares (OLS) to estimate the values of the parameters a, b and σ, but we could have just as easily used the method of maximum likelihood estimation (MLE). Under the assumptions of this model, namely that the errors are normally distribution, these two methods produce equivalent point estimates of the parameters. For now, we need not worry about the details of the particular estimation method (more on this in the next chapters). Suffice it to say that either method produces parameter estimates that make our observed data the most likely outcome under repeated sampling.



- To make inferences about φ, classical statistics appeals to the idea of hypothetical outcomes under *repeated sampling*; i.e., the estimate of φ is viewed as a single outcome that belongs to a distribution of estimates associated with hypothetical repetitions of an experiment or survey.
- Under this view, the fixed value φ and the assumptions of the model represent a mechanism for generating a random, hypothetical sequence of data sets and parameter estimates:

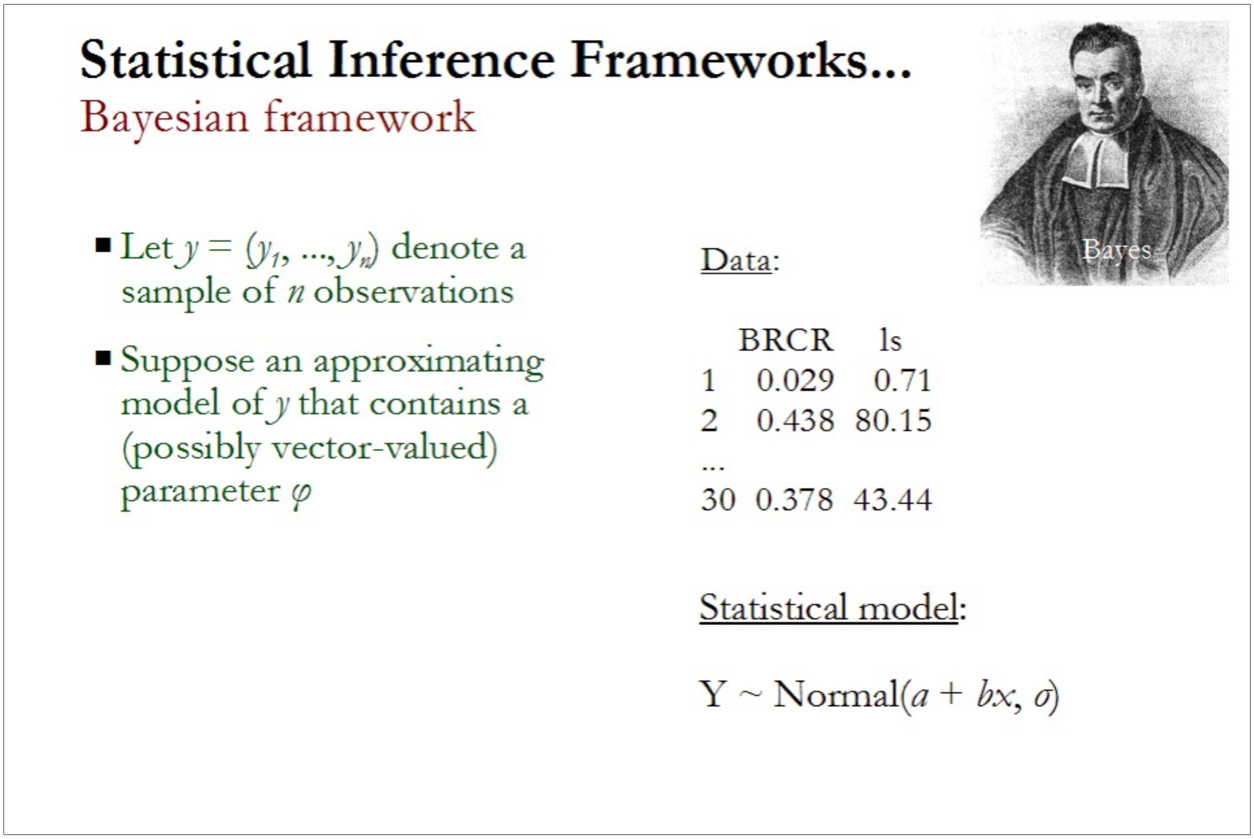
$$(y, \hat{\varphi}_1), (y, \hat{\varphi}_2), (y, \hat{\varphi}_3), \ldots$$

<u>Example</u>: Let's assume that our parameter estimates are the true, fixed values of the population parameters; a=0.099 b=0.006 and σ=0.141 . If we were to repeatedly draw samples from this population and each time estimate the parameters from the sample, we would generate a distribution of estimates in which the true values are the most likely or frequently occurring. Note, this is a hypothetical distribution, as repeated sampling of the population is almost never practical in the real world.



- Therefore, probability statements about φ (i.e., inferences) are made with respect to the distribution of estimates of φ that could have been obtained in repeated samples.
- In classical statistics, the role of probability in computing inferences is based on the relative frequency of outcomes in repeated samples – hence the name “frequentist”.
- Note, frequentists never use probability directly as an expression of degrees of belief in the magnitude of φ. Probability statements are based entirely on the hypothetical distribution of $\hat{\varphi}$.

Example: In this example, using a frequentist approach we could not legitimately claim that our estimates are the most probable, only that they are the most likely to have given rise to our data – if we had been able to repeatedly sample the population. In practice, however, likelihood is often interpreted as a probability even though it is not in the strictest sense – a point that a Bayesian is keen to point out.



4. Bayesian inference framework

Bayesian inference represents an alternative approach to model-based inference. Surprisingly, the Bayesian framework is actually much older than the frequentist framework, dating back to 1763 with a paper written by Thomas Bayes, but it fell out of favor with the advent of the frequentist approach in the early 1900s and only recently, since the 1990s, has regained popularity. It is now the hottest area of modern statistics. The essence of the approach is as follows:

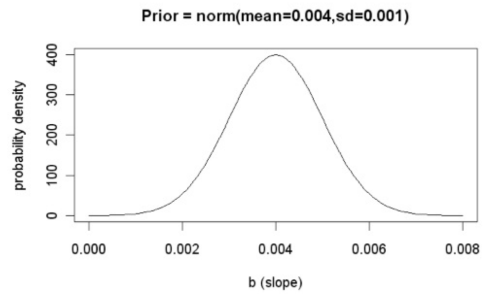
- Let $y = (y_1, ..., y_n)$ denote a sample of n observations.
- Suppose an approximating model of y that contains a (possibly vector-valued) parameter φ.

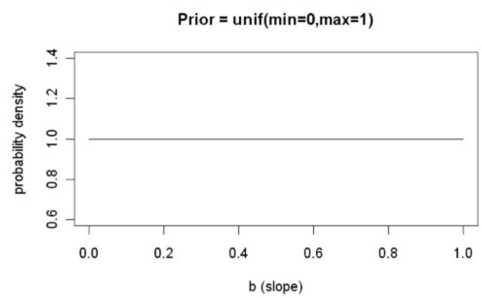
<u>Example</u>: We use the same example to illustrate the differences between frameworks. Note, the sample data is the same, as is the approximating model.

Statistical Inference Frameworks...

Bayesian framework

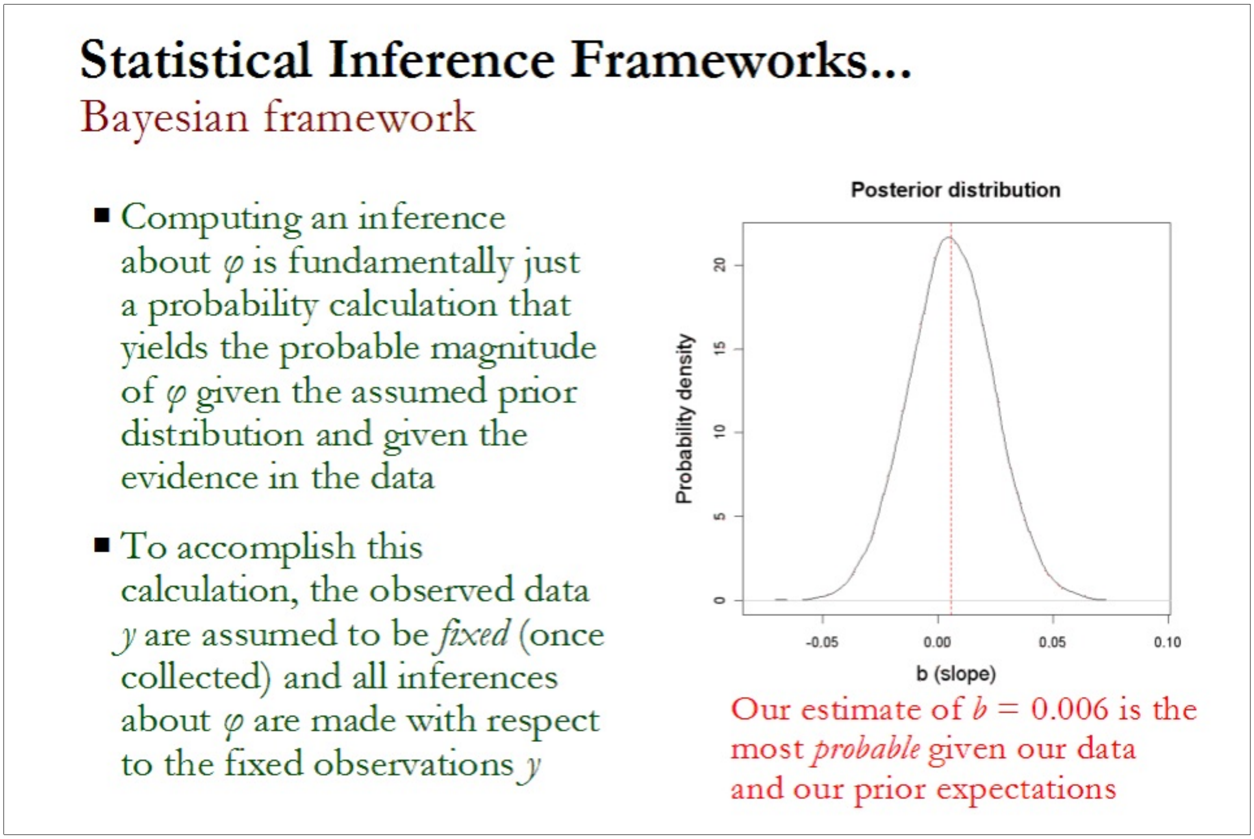
- The model parameter φ is treated as a random variable (not fixed) and the approximating model is elaborated to include a probability distribution for φ that specifies one's beliefs about the magnitude of φ prior to having observed the data – this elaboration is called the *prior distribution*





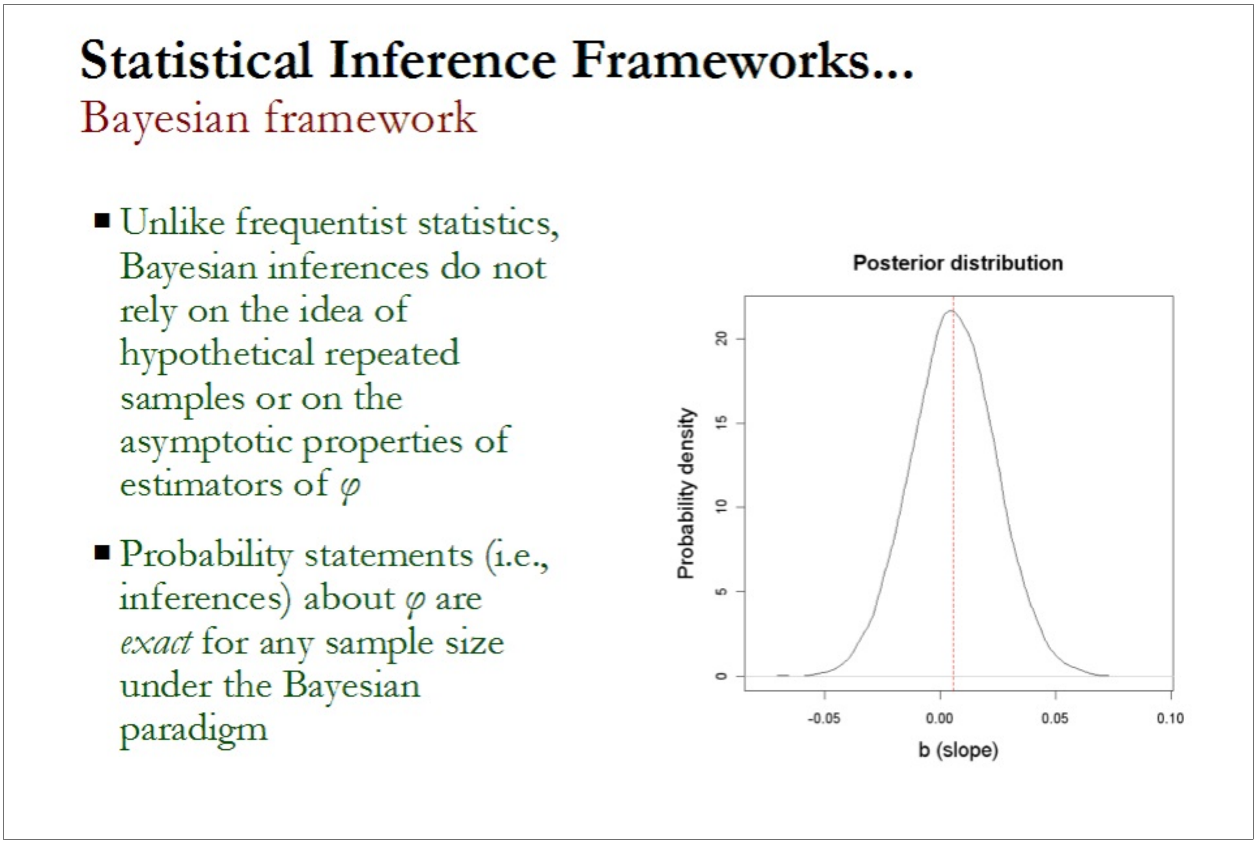
- In frequentist inference, the model parameter φ is assumed to have a fixed, but unknown, value. In Bayesian inference, the model parameter θ is treated as a random variable and the approximating model is elaborated to include a probability distribution for φ that specifies one's beliefs about the magnitude of φ prior to having observed the data – this elaboration is called the *prior distribution*. The prior distribution is necessary in order to make the Bayesian approach work, but we will not worry about that here.

<u>Example</u>: The parameters of our statistical model, a, b and σ, are treated as random variables, not fixed as in the frequentist approach. In fact, in the Bayesian approach, it is mute whether the parameters are believed to be truly fixed or not since we can never confirm them as such. The best we can do is make probability statements about their values. Importantly, in the Bayesian approach we have to specify our prior belief about the values of the parameters. For example, we might specify our prior belief that the slope parameter (b) in the linear model is normally distributed with a mean of 0.004 and a standard deviation of 0.001 and then see whether the sample data conforms to this expectation or differs.



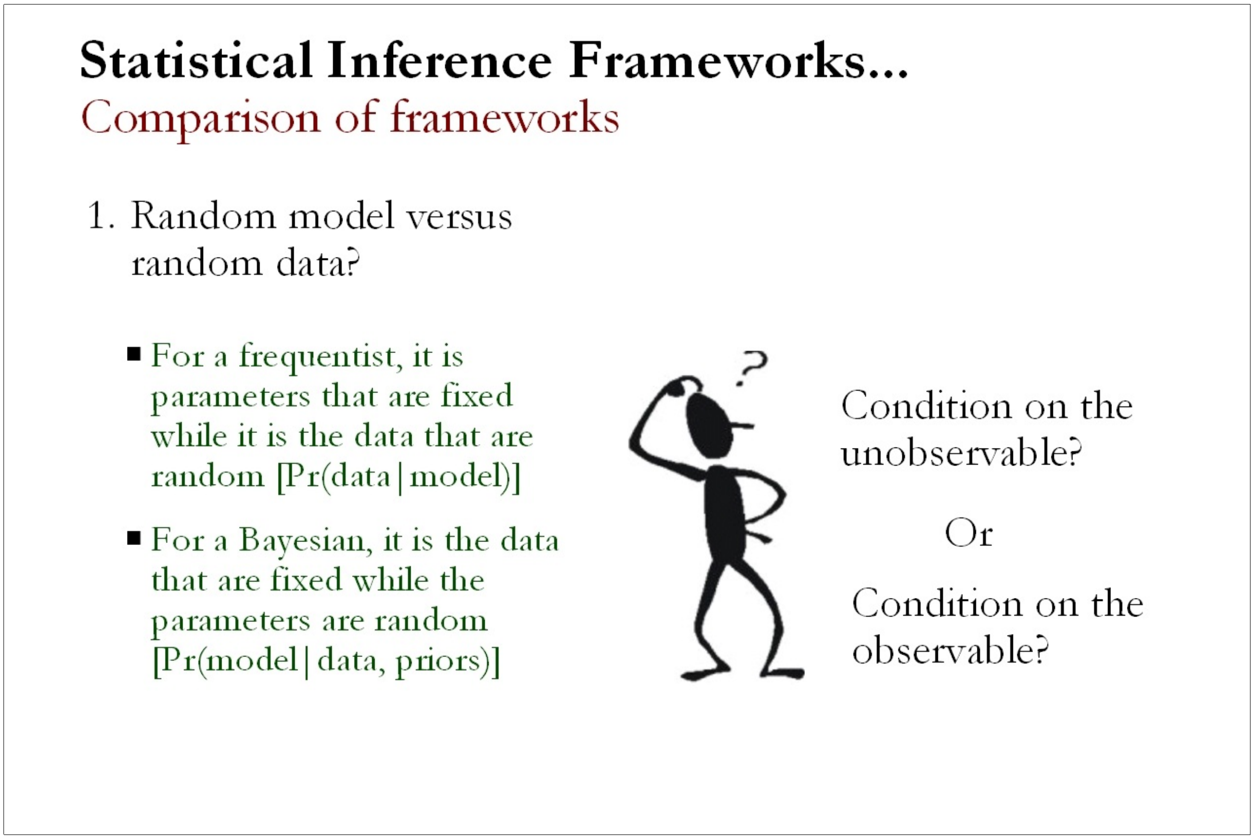
- In the Bayesian view, computing an inference about φ is fundamentally just a probability calculation that yields the probable magnitude of φ given the assumed prior distribution and given the evidence in the data.
- To accomplish this calculation, the observed data y are assumed to be *fixed* (once the sample has been obtained), not a random outcome as in the frequentist view, and all inferences about φ are made with respect to the fixed observations y.

Example: Here we used a Bayesian method to estimate the most probable values of the parameters a, b and σ. For now, we need not worry about the details of this particular estimation method (more on this in the next section). Suffice it to say that this method produces parameter estimates that are the most probable given our observed data and prior beliefs.



- Unlike frequentist statistics, Bayesian inferences do not rely on the idea of hypothetical repeated samples or on the asymptotic properties of estimators of φ.
- Probability statements (i.e., inferences) about φ are *exact* for any sample size under the Bayesian paradigm.

Example: Recall that in the frequentist approach we estimated parameters that, if true, would make the data the most likely outcome under hypothetical repeated sampling. In the Bayesian approach, however, we make no reliance on hypothetical repeated sampling. Instead, we estimate the *posterior probability distribution* of each of parameter – our best estimate of its probability distribution given the observed data and some prior knowledge of its distribution. For now, let's not worry about how we estimate a posterior probability distribution, since it involves a couple of tricks. What matters is that with our estimate of the posterior distribution, we can ask questions like what is the most probable value of that distribution?



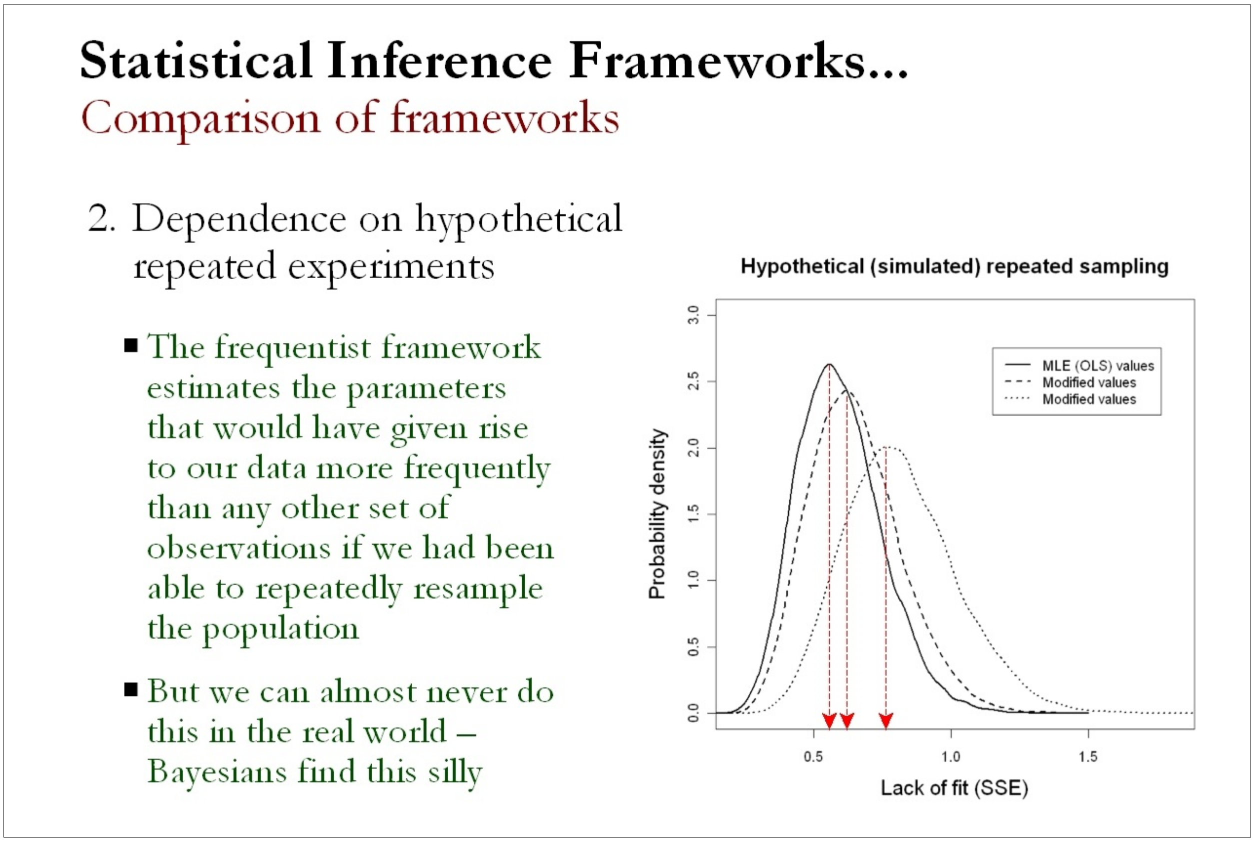
5. Comparison of Inference frameworks

The classical frequentist and Bayesian inference frameworks differ in a number of important and some not so important ways. Each framework has proponents that are quick to point out the drawbacks of the other. Let's review a few of the more notable differences and opposing points of view here; we will address additional points of comparison later.

1. *Random model versus random data.*–Implicit in the frequentist approach to estimation is that there is a fixed quantity in nature, the parameter(s), that we wish to measure. In short, for a frequentist it is parameters that are fixed while it is the data that are random. This is diametrically opposed to the Bayesian point of view. For a Bayesian it is the data that are fixed while the parameters are random. Since we can never know whether any model is true, Bayesians argue that it does not make sense to condition on something we can never observe. Bayesians argue that we should instead determine the probability of the model given the data (i.e., condition on the data), since the data is the only thing we know for certain.

The notion that the parameter is a fixed quantity in nature causes problems in interpretation for the frequentist. When estimating a parameter we usually derive both point and interval estimates. The point estimate is our best guess at the parameter's true value; the interval estimate is our best guess at the likely interval (range of values) that contains the parameter's true value. When we construct the interval estimate (called a confidence interval) for a parameter, we like to treat the interval as a probability statement for the parameter—a set of likely values. But in truth, if the parameter is fixed, then it is either in the interval we've constructed or it's not. There's no probability associated with it. The probability instead derives from the sampling distribution. For example, we call it a 95% confidence interval because we're guaranteed that 95% of the intervals we might have constructed if we had obtained all possible samples from the population do in fact contain the true parameter value. All we can do is hope that this is one of the lucky ones. The bottom line is that when the parameter is treated as real and fixed, then it's only our methods that can have probability associated with them.

As far as the existence of a true value of the parameter in nature, Bayesians are of an open mind. The parameter may or may not be real, but in the Bayesian perspective it doesn't matter. All we know about the parameter is what we believe about it. As knowledge accumulates, our beliefs about the parameter become more focused. Since the value of the parameter is a matter of belief, and probability is a matter of belief, for all intents and purposes the parameter can be viewed as a random quantity. Consequently, to a Bayesian, parameters are random variables, not fixed constants. As a result, confidence intervals pose no philosophical dilemma for a Bayesian. Since parameters are random, we can make probability statements about their values. Thus, a confidence interval for a parameter is a probability statement about the likely values of the parameter. To avoid confusion with frequentist confidence intervals, Bayesians often call their intervals "credibility intervals".



2. *Reliance on hypothetical repeated sampling*.–For the frequentist, everything hinges on the notion of repeated sampling to generate the sampling distribution of the statistic. Here, the definition of probability depends on a series of hypothetical repeated experiments that are often impossible in any practical sense. Importantly, in the frequentist framework, we assume that the specified model is true and that the data represent a random outcome. We try to estimate the parameters of the model that if true would give rise to our observed data as the most frequently occurring outcome. For example, to say that the probability of heads is one half when a fair coin is tossed once means that if we were to flip a fair coin repeatedly the long run relative frequency of heads is one half. But since we only observe our data once (typically), we have to estimate the parameters that would have given rise to our data more frequently than any other set of observations if we had been able to repeatedly resample the population. Because we can almost never actually do this in the real world, Bayesians find this framework silly.

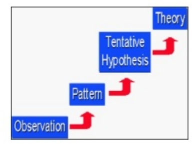
For the Bayesian, there is no reliance on hypothetical repeated sampling. The probabilistic interpretation of parameters instead stems from the fact that parameters are treated as random variables, not fixed constants, which allows us to make probability statements about their values directly. However, the catch is that to do this we have to specify prior probability distributions for each parameter – our prior belief in the value of each parameter. Given the data in hand and these prior distributions, we can derive parameter estimates that are the most probable values – without hypothetical repeated sampling. Frequentists object to the use of priors.

Statistical Inference Frameworks…

Comparison of frameworks

3. Role of priors – subjective versus objective approaches

- For the Bayesian, science is subjective and is a constant dynamic between prior and posterior probabilities
- For the Frequentist, science is inherently objective and has no use for subjective quantities such as prior probabilities

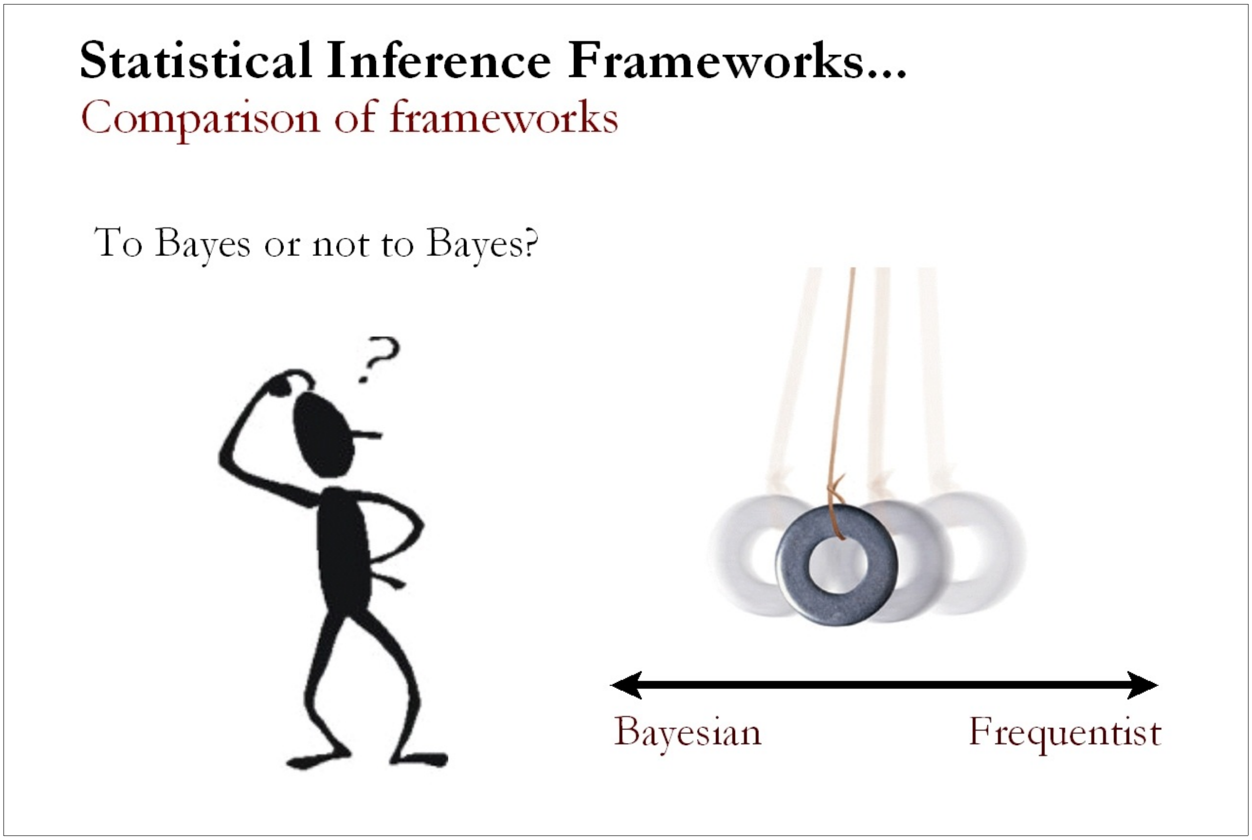


3. *Role of priors - subjective versus objective approaches.*--A Bayesian takes pride in the fact that the Bayesian method of estimation, which finds the most probable parameters given the data at hand and prior knowledge, essentially encapsulates inductive science in a single formula. In science we develop theories about nature. Observation and experiment then cause us to modify those theories about nature. Further data collection causes further modification. Thus, science is a constant dynamic between prior and posterior probabilities, with prior probabilities becoming modified into posterior probabilities through experimentation at which point they become the prior probabilities of future experiments. Thus, the Bayesian perspective accounts for the cumulative nature of science.

The frequentist retort is that this is a mischaracterization of science. Science is inherently objective and has no use for subjective quantities such as prior probabilities. Science should be based on data alone and not the prejudices of the researcher.

The Bayesian rejoinder is first that science does have a subjective component. The "opinions" of scientists dictate the kinds of research questions they pursue. In any case, if there is concern that a prior probability unfairly skews the results, the analysis can be rerun with other priors or with uninformative priors that do not skew the results in any direction. In fact, in Bayesian analysis it is fairly typical to carry out analyses with a range of priors to demonstrate that results are robust to the choice of prior. In any event, with large sample sizes, say >30, the priors may have little influence on the result anyways.

In fact both schools of thought are correct here. While Bayesians may have described the ideal scientific method, in truth consensus in science is informal at best. Perhaps there should be a current prior probability in vogue for everything in science, but typically there isn't. Without this consensus the inherent subjectivity of priors does seem to be a problem, at least in some cases.



To Bayes or Not to Bayes? This is the question most often asked today. Unfortunately, there is no simple answer since it depends on so many things. Clearly, the current momentum in ecology is towards the Bayesian framework, but is this because it is fashionable or because it is inherently superior – this remains to be seen. The pragmatic modeler will become familiar with both frameworks, learn the advantages and disadvantages of each and seek to understand the conditions under which the two frameworks give the same or very similar answers – in which case it doesn't matter – and when they differ and why.