CHAPTER 5

Three Frameworks for Statistical Analysis

In this chapter, we introduce three major frameworks for statistical analysis: **Monte Carlo analysis, parametric analysis, and Bayesian analysis.** In a nutshell, Monte Carlo analysis makes minimal assumptions about the underlying distribution of the data. It uses randomizations of observed data as a basis for inference. Parametric analysis assumes the data were sampled from an underlying distribution of known form, such as those described in Chapter 2, and estimates the parameters of the distribution from the data. Parametric analysis estimates probabilities from observed frequencies of events and uses these probabilities as a basis for inference. Hence, it is a type of frequentist inference. Bayesian analysis also assumes the data were sampled from an underlying distribution of known form. It estimates parameters not only from the data, but also from prior knowledge, and assigns probabilities to these parameters. These probabilities are the basis for Bayesian inference. Most standard statistics texts teach students parametric analysis, but the other two are equally important, and Monte Carlo analysis is actually easier to understand initially. To introduce these methods, we will use each of them to analyze the same sample problem.

Sample Problem

Imagine you are trying to compare the nest density of ground-foraging ants in two habitats—field and forest. In this sample problem, we won't concern ourselves with the scientific hypothesis that you are testing (perhaps you don't even have one at this point); we will simply follow through the process of gathering and analyzing the data to determine whether there are consistent differences in the density of ant nests in the two habitats.

You visit a forest and an adjacent field and estimate the average density of ant nests in each, using replicated sampling. In each habitat, you choose a random location, place a square quadrat of 1-m^2 area, and carefully count all of the ant nests that occur within the quadrat. You repeat this procedure several times in each habitat. The issue of choosing random locations is very important for any type of statistical analysis. Without more complicated methods, such as stratified sampling, randomization is the only safeguard to ensure that we have a representative sample from a population (see Chapter 6).

Each row is an independent observation. The first column identifies the replicate with a unique ID number, the second column indicates the habitat sampled, and the third column gives the number of ant nests recorded in the replicate.

TABLE 5.1 Sample dataset used to illustrate Monte Carlo, parametric, and Bayesian analyses

ID number	Habitat	Number of ant nests per quadrat
1	Forest	9
2	Forest	6
3	Forest	4
4	Forest	6
5	Forest	7
6	Forest	10
7	Field	12
8	Field	9
9	Field	12
10	Field	10

The spatial scale of the sampling determines the scope of inference. Strictly speaking, this sampling design will allow you to discuss differences between forest and field ant densities *only at this particular site*. A better design would be to visit several different forests and fields and sample one quadrat within each of them. Then the conclusions could be more readily generalized.

Table 5.1 illustrates the data in a spreadsheet. The data are arranged in a table, with labeled rows and columns. Each row of the table contains all of the information on a particular observation. Each column of the table indicates a different variable that has been measured or recorded for each observation. In this case, your original intent was to sample 6 field and 6 forest quadrats, but the field quadrats were more time-consuming than you expected and you only managed to collect 4 field samples. Thus the data table has 10 rows (in addition to the first row, which displays labels) because you collected 10 different samples (6 from the forest and 4 from the field). The table has 3 columns. The first column contains the unique ID number assigned to each replicate. The other columns contain the two pieces of information recorded for each replicate: the habitat in which the replicate was sampled (field or forest); and the number of ant nests recorded in the quadrat.[1]

[1] Many statistics texts would show these data as two columns of numbers, one for forest and one for field. However, the layout we have shown here is the one that is recognized most commonly by statistical software for data analysis.

TABLE 5.2 Summary statistics for the sample data in Table 5.1

Habitat	N	Mean	Standard deviation
Forest	6	7.00	2.19
Field	4	10.75	1.50

Following the procedures in Chapter 3, calculate the mean and standard deviation for each sample (Table 5.2). Plot the data, using a conventional bar chart of means and standard deviations (Figure 5.1), or the more informative box plot described in Chapter 3 (Figure 5.2).

Although the numbers collected in the forest and field show some overlap, they appear to form two distinct groups: the forest samples with a mean of 7.0 nests per quadrat, and the field samples with a mean of 10.75 nests per quadrat. On the other hand, our sample size is very small ($n = 6$ forest and $n = 4$ field samples). Perhaps these differences could have arisen by chance or random sampling. We need to conduct a statistical test before deciding whether these differences are significant or not.

Monte Carlo Analysis

Monte Carlo analysis involves a number of methods in which data are randomized or reshuffled so that observations are randomly reassigned to different treatments or groups. This randomization[2] specifies the null hypothesis under consideration: that the pattern in the data is no different from that which we would expect if the observations were assigned randomly to the different groups. There are four steps in Monte Carlo analysis:

1. Specify a test statistic or index to describe the pattern in the data.
2. Create a distribution of the test statistic that would be expected under the null hypothesis.
3. Decide on a one- or two-tailed test.
4. Compare the observed test statistic to a distribution of simulated values and estimate the appropriate P-value as a tail probability (as described in Chapter 3).

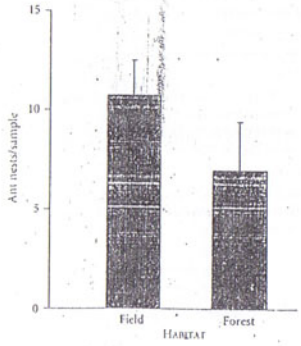
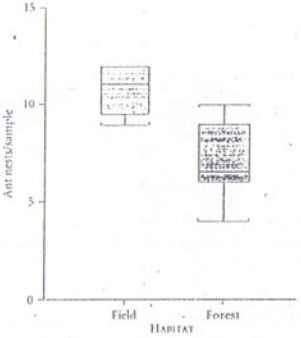


Figure 5.1 Standard bar chart for sample data in Table 5.1. The height of the bar is the mean of the sample and the vertical line indicates one standard deviation above the mean. Monte Carlo, parametric, and Bayesian analyses can all be used to evaluate the difference in means between the groups.

Figure 5.2 Box plot of data in Table 5.1. In a box plot, the central line within the box is the median of the data. The box includes 50% of the data. The top of the box indicates the 75th percentile (upper quartile) of the data, and the bottom of the box indicates the 25th percentile (lower quartile) of the data. The vertical lines extend to the upper and lower deciles (90th and 10th percentiles). For the field sample, there are so few data that the 75th and 90th percentiles do not differ. When data have asymmetric distributions or outliers, box plots may be more informative than standard bar graphs such as Figure 5.1.



[2] Some statisticians distinguish Monte Carlo methods, in which samples are drawn from a known or specified statistical distribution, from randomization tests, in which existing data are reshuffled but no assumptions are made about the underlying distribution. In this book, we use Monte Carlo methods to mean randomization tests. Another set of methods includes bootstrapping, in which statistics are estimated by repeatedly subsampling with replacement from a dataset. Still another set of methods includes jackknifing, in which the variability in the dataset is estimated by systematically deleting each observation and then recalculating the statistics (see Figure 9.8 and the section "Discriminant Analysis" in Chapter 12). Monte Carlo, of course, is the famous gambling resort city on the Riviera, whose citizens do not pay taxes and are forbidden from entering the gaming rooms.

Step 1: Specifying the Test Statistic

For this analysis, we will use as a measure of pattern the absolute difference in the means of the forest and field samples, or DIF:

$$DIF_{obs} = |10.75 - 7.00| = 3.75$$

The subscript "obs" indicates that this DIF value is calculated for the observed data. The null hypothesis is that a DIF_{obs} equal to 3.75 is about what would be expected by random sampling. The alternative hypothesis would be that a DIF_{obs} equal to 3.75 is larger than would be expected by chance.

Step 2: Creating the Null Distribution

Next, estimate what DIF would be if the null hypothesis were true. To do this, use the computer (or a deck of playing cards) to randomly reassign the forest and field labels to the dataset. In the randomized dataset, there will still be 4 field and 6 forest samples, but those labels (Field and Forest) will be randomly reassigned (Table 5.3). Notice that in the randomly reshuffled dataset, many of the observations were placed in the same group as the original data. This will happen by chance fairly often in small datasets. Next, calculate the sample statistics for this randomized dataset (Table 5.4). For this dataset, $DIF_{sim} = |7.75 - 9.00| = 1.25$. The subscript "sim" indicates that this value of DIF is calculated for the randomized, or simulated, data.

In this first simulated dataset, the difference between the means of the two groups ($DIF_{sim} = 1.25$) is smaller than the difference observed in the real data ($DIF_{obs} = 3.75$). This result suggests that means of the forest and field samples may differ more than expected under the null hypothesis of random assignment.

TABLE 5.3 Monte Carlo randomization of the habitat labels in Table 5.1

Habitat	Number of nests per quadrat
Field	9
Field	6
Forest	4
Forest	6
Field	7
Forest	10
Forest	12
Field	9
Forest	12
Forest	10

In the Monte Carlo analysis, the sample labels in Table 5.1 are reshuffled randomly among the different samples. After reshuffling, the difference in the means of the two groups, DIF, is recorded. This procedure is repeated many times, generating a distribution of DIF values (Figure 5.3).

TABLE 5.4 Summary statistics for the randomized data in Table 5.3

Habitat	N	Mean	Standard deviation
Forest	6	9.00	3.286
Field	4	7.75	1.500

In the Monte Carlo analysis, these values represent the mean and standard deviation in each group after a single reshuffling of the labels. The difference between the two means (DIF = |7.75 – 9.00| = 1.25) is the test statistic.

However, there are many different random combinations that can be produced by reshuffling the sample labels.[3] Some of these have relatively large values of DIF_{sim} and some have relatively small values of DIF_{sim}. Repeat this reshuffling exercise many times (usually 1000 or more), then illustrate the distribution of the simulated DIF values with a histogram (Figure 5.3) and summary statistics (Table 5.5).

The mean DIF of the simulated datasets was only 1.46, compared to the observed value of 3.75 for the original data. The standard deviation of 2.07 could be used to construct a confidence interval (see Chapter 3), but it should not be, because the distribution of simulated values (Figure 5.3) has a long right-hand tail and does not follow a normal distribution. An approximate 95% confidence interval can be derived from this Monte Carlo analysis by identifying the upper and lower 2.5 percentiles of the data from the set of simulated DIF values and using these values as the upper and lower bounds of the confidence interval (Efron 1982).

Step 3: Deciding on a One- or Two-Tailed Test

Next, decide whether to use a **one-tailed test** or a **two-tailed test**. The "tail" refers to the extreme left- or right-hand areas under the probability density function (see Figure 3.7). It is these tails of the distribution that are used to determine the cutoff points for a statistical test at $P = 0.05$ (or any other probability level). A one-tailed test uses only one tail of the distribution to estimate the P-value. A two-tailed test uses both tails of the distribution to estimate the P-value. In a

[3] With 10 samples split into a group of 6 and a group of 4, there are

$$\binom{10}{4} = 210$$

combinations that can be created by reshuffling the labels (see Chapter 2). However, the number of unique values of DIF that are possible is somewhat less than this because some of the samples had identical nest counts.

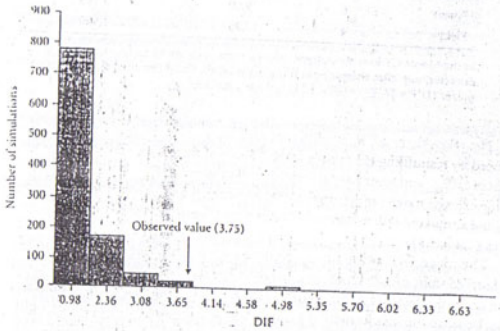


Figure 5.3 Monte Carlo analysis of the data in Table 5.1. For each randomization, the data labels (Forest, Field) were randomly reshuffled among the replicates. Next, the difference (DIF) between the means of the two groups was calculated. This histogram illustrates the distribution of DIF values from 1000 such randomizations. The arrow indicates the single value of DIF observed in the real dataset (3.75). The observed DIF sits well in the right-hand tail of the distribution. The observed DIF of 3.75 was larger than or equal to all but 36 of the simulated DIF values. Therefore, under the null hypothesis of random assignment of samples to groups, the tail probability of finding this observation (or one more extreme) is 36/1000 = 0.036.

one-tailed test, all 5% of the area under the curve is located in one tail of the distribution. In a two-tailed test, each tail of the distribution would encompass 2.5% of the area. Thus, a two-tailed test requires more extreme values to achieve statistical significance than does a one-tailed test.

However, the cutoff value is not the most important issue in deciding whether to use a one- or a two-tailed test. Most important are the nature of the response

TABLE 5.5 Summary statistics for 1000 simulated values of DIF in Figure 5.3

Variable	N	Mean	Standard deviation
DIF_{sim}	1000	1.46	2.07

In the Monte Carlo analysis, each of these values was created by reshuffling the data labels in Table 5.1, and calculating the difference between the means of the two groups (DIF). Calculation of the *P*-value does not require that DIF follow a normal distribution, because the *P*-value is determined directly by the location of the observed DIF statistic in the histogram (see Table 5.6).

variable and the precise hypothesis being tested. In this case, the response variable was DIF, the *absolute* difference in the means of forest and field samples. For the DIF variable, a one-tailed test is most appropriate for *unusually large* values of DIF. Why not use a two-tailed test with DIF? The lower tail of the DIF_{sim} distribution would represent values of DIF that are *unusually small* compared to the null hypothesis. In other words, a two-tailed test would also be testing for the possibility that forest and field means were more similar than expected by chance. This test for extreme similarity is not biologically informative, so attention is restricted to the upper tail of the distribution. The upper tail represents cases in which DIF is unusually large compared to the null distribution.

How could the analysis be modified to use a two-tailed test? Instead of using the absolute value of DIF, use the average difference between forest and field samples (DIF^*). Unlike DIF, DIF^* can take on both positive and negative values. DIF^* will be positive if the field average is greater than the forest average. DIF^* will be negative if the field average is less than the forest average. As before, randomize the data and create a distribution of DIF^*_{sim}. In this case, however, a two-tailed test would detect cases in which the field mean was unusually large compared to the forest mean (DIF^* positive) and cases in which the field mean was unusually small compared to the forest mean (DIF^* negative).

Whether you are using Monte Carlo, parametric, or Bayesian analysis, you should study carefully the response variable you are using. How would you interpret an extremely large or an extremely small value of the response variable relative to the null hypothesis? The answer to this question will help you decide whether a one- or a two-tailed test is most appropriate.

Step 4: Calculating the Tail Probability

The final step is to estimate the probability of obtaining DIF_{obs} or a value more extreme, given that the null hypothesis is true [$P(\text{data}|H_0)$]. To do this, examine the set of simulated DIF values (plotted as the histogram in Figure 5.3), and tally up the number of times that the DIF_{obs} was greater than, equal to, or less than each of the 1000 values of DIF_{sim}.

In 29 of 1000 randomizations, $DIF_{sim} = DIF_{obs}$, so the probability of obtaining DIF_{obs} under the null hypothesis is 29/1000 = 0.029 (Table 5.6). However, when we calculate a statistical test, we usually are not interested in this exact probability as much as we are the tail probability. That is, we want to know the chances of obtaining an observation *as large or larger than* the real data, given that the null hypothesis is true. In 7 of 1000 randomizations, $DIF_{sim} > DIF_{obs}$. Thus, the probability that $DIF_{obs} \geq 3.75$ is (7 + 29)/1000 = 0.036. This tail probability is the frequency of obtaining the observed value (29/1000) plus the frequency of obtaining a more extreme result (7/1000).

TABLE 5.6 Calculation of tail probabilities in Monte Carlo analysis

Inequality	N
$DIF_{sim} > DIF_{obs}$	7
$DIF_{sim} = DIF_{obs}$	29
$DIF_{sim} < DIF_{obs}$	964

Comparisons of DIF_{obs} (absolute difference in the mean of the two groups in the original data) with DIF_{sim} (absolute difference in the mean of the two groups after randomizing the group assignments), N is the number of simulations (out of 1000) for which the inequality was obtained. Because $DIF_{sim} \geq DIF_{obs}$ in $7 + 29 = 36$ trials out of 1000, the tail probability under the null hypothesis of finding DIF_{obs} this extreme is $36/1000 = 0.036$.

Follow the procedures and interpretations of *P*-values that we discussed in Chapter 4. With a tail probability of 0.036, it is unlikely that these data would have occurred given the null hypothesis is true.

Assumptions of the Monte Carlo Method

Monte Carlo methods rest on three assumptions:

1. The data collected represent random, independent samples.
2. The test statistic describes the pattern of interest.
3. The randomization creates an appropriate null distribution for the question.

Assumptions 1 and 2 are common to all statistical analyses. Assumption 1 is the most critical, but it is also the most difficult to confirm, as we will discuss in Chapter 6. Assumption 3 is easy to meet in this case. The sampling structure and null hypothesis being tested are very simple. For more complex questions, however, the appropriate randomization method may not be obvious, and there may be more than one way to construct the null distribution (Gotelli and Graves 1996).

Advantages and Disadvantages of the Monte Carlo Method

The chief conceptual advantage of the Monte Carlo method is that it makes clear and explicit the underlying assumptions and the structure of the null hypothesis. In contrast, conventional parametric analyses often gloss over these features, perhaps because the methods are so familiar. Another advantage of the Monte Carlo method over parametric analysis is that it does not require the assumption that the data are sampled from a specified probability distribution, such as the normal. Finally, Monte Carlo simulations allow you to tailor your statistical test to particular questions and datasets, rather than having to shoehorn them into a conventional test that may not be the most powerful method for your question, or whose assumptions may not match the sampling design of your data.

The chief disadvantage of the Monte Carlo method is that it is computer-intensive and is not included in most traditional statistical packages (but see

Gotelli and Entsminger 2003). As computers get faster and faster, older limitations on computational methods are disappearing, and there is no reason that even very complex statistical analyses cannot be run as a Monte Carlo simulation. However, until such routines become widely available, Monte Carlo methods are available only to those who know a programming language and can write their own programs.[4]

A second disadvantage of Monte Carlo analysis is psychological. Some scientists are uneasy about Monte Carlo because different analyses of the same dataset can yield slightly different answers. For example, we re-ran the analysis on the ant data in Table 5.1 ten times and got *P*-values that ranged from 0.030 to 0.046. Most researchers are more comfortable with parametric analyses, which have more of an air of objectivity because the same *P*-value results each time the analysis is repeated.

A final weakness is that the domain of inference for a Monte Carlo analysis is subtly more restrictive than that for a parametric analysis. A parametric analysis assumes a specified distribution and allows for inferences about the underlying parent population from which the data were sampled. Strictly speaking, inferences from Monte Carlo analyses (at least those based on simple randomization tests) are limited to the specific data that have been collected. However, if a sample is representative of the parent population, the results can be generalized cautiously.

[4] One of the most important things you can do is to take the time to learn a real programming language. Although some individuals are proficient at programming macros in spreadsheets, macros are practical only for the most elementary calculations; for anything more complex, it is actually simpler to write a few lines of computer code than to deal with the convoluted (and error-prone) steps necessary to write macros. There are now many computer languages available for you to choose from. NJG prefers Delphi, which is a high-level version of Pascal with a graphical user interface; AME prefers S-Plus or its open-source version, R, which includes many built-in mathematical and statistical functions. Unfortunately, learning to program is like learning to speak a foreign language—it takes time and practice, and there is no immediate payoff. Sadly, our academic culture doesn't encourage the learning of programming skills (or languages other than English). But if you can overcome the steep learning curve, the scientific payoff is tremendous. The best way to begin is not to take a class, but to purchase the language software, work through the examples in the manual, and then begin with a problem that interests you. Hilborn and Mangel's *The Ecological Detective* (1997) contains an excellent series of ecological exercises to build your programming skills. Not only will programming free you from the chains of canned software packages, it will sharpen your analytical skills and give you new insights into ecological and statistical models. You will have a deeper understanding of a model or a statistic once you have successfully programmed it!

Parametric Analysis

Parametric analysis refers to the large body of statistical tests and theory built on the assumption that the data being analyzed were sampled from a specified distribution. Most statistical tests familiar to ecologists and environmental scientists specify the normal distribution. The parameters of the distribution (e.g., the population mean μ and variance σ^2) are then estimated and used to calculate tail probabilities for a true null hypothesis. A large statistical framework has been built around the simplifying assumption of normality of data. As much as 80% to 90% of what is taught in standard statistics texts falls under this umbrella. Here we use a common parametric method, the analysis of variance (ANOVA), to test for differences in the group means of the sample data.[5] There are three steps in parametric analysis:

1. Specify the test statistic.
2. Specify the null distribution.
3. Calculate the tail probability.

Step 1: Specifying the Test Statistic

Parametric analysis of variance assumes that the data are drawn from a normal, or Gaussian, distribution. The mean and variance of these curves can be estimated from the sample data (Table 5.1) using Equations 3.1 and 3.9. Figure 5.4 shows the distributions used in parametric analysis of variance. The original data are arrayed on the x-axis, and each color represents a different habitat (black circles for the forest samples, green circles for the field samples).

Sir Ronald Fisher

[5] The framework for modern parametric statistical theory was largely developed by the remarkable Sir Ronald Fisher (1890–1962), and the F-ratio is named in his honor (although Fisher himself felt that the ratio needed further study and refinement). Fisher held the Balfour Chair in Genetics at Cambridge University from 1943 to 1957 and made fundamental contributions to the theory of population genetics and evolution. In statistics, he developed the analysis of variance (ANOVA) to analyze crop yields in agricultural systems, in which it may be difficult or impossible to replicate treatments. Many of the same constraints face ecologists in the design of their experiments today, which is why Fisher's methods continue to be so useful. His classic book, *The Design of Experiments* (1935) is still enjoyable and worthwhile reading. It is ironic that Fisher became uneasy about his own methods when dealing with experiments he could design well, whereas today many ecologists apply his methods to ad hoc observations in poorly designed natural experiments (see Chapters 4 and 6). (Photograph courtesy of the Ronald Fisher Memorial Trust.)

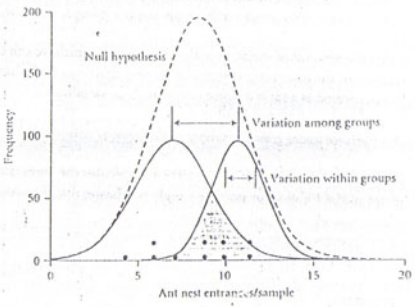


Figure 5.4 Normal distributions based on sample data in Table 5.1 The data in Table 5.1 are shown as symbols (black circles, forest samples; green circles, field samples) that indicate the number of ant nests counted in each quadrat. The null hypothesis is that all the data were drawn from the same population whose normal distribution is indicated by the dashed line. The alternative hypothesis is that each habitat has its own distinct mean (and variance), indicated by the two smaller normal distributions. The smaller the shaded overlap of the two distributions, the less likely it is that the null hypothesis is true. Measures of variation among and within groups are used to calculate the F-ratio and test the null hypothesis.

First consider the null hypothesis: that both sets of data were drawn from a single underlying normal distribution, which is estimated from the mean and variance of all the data (dashed curve; mean = 8.5, standard deviation = 2.54). The alternative hypothesis is that the samples were drawn from two different populations, each of which can be characterized by a different normal distributions, one for the forest and one for the field. Each distribution has its own mean, although we assume the variance is the same (or similar) for each of the two groups. These two curves are also illustrated in Figure 5.4, using the summary statistics calculated in Table 5.2.

How is the null hypothesis tested? The closer the two curves are for the forest and field data, the more likely the data would be collected given the null hypothesis is true, and the single dashed curve best represents the data. Conversely, the more separate the two curves are, the less likely it is that the data represent a single population with a common mean and variance. The area of overlap between these two distributions (shaded in Figure 5.4) should be a measure of how close or how far apart the distributions are.

Fisher's contribution was to quantify that overlap as a ratio of two variables. The first is the amount of variation among the groups, which we can think of as the variance (or standard deviation) of the means of the two groups. The second is the amount of variation within each group, which we can think of as the variance of the observations around their respective means. Fisher's F-ratio can be interpreted as a ratio of these two sources of variation:

$$F = (\text{variance among groups} + \text{variance within groups}) / \text{variance within groups} \quad (5.1)$$

In Chapter 10, we will explain in detail how to calculate the numerator and denominator of the F-ratio. For now, we simply emphasize that the ratio measures the relative size of two sources of variation in the data: variation among groups and within groups. For these data, the F-ratio is calculated as 33.75/ 3.84 = 8.78.

In an ANOVA, the F-ratio is the test statistic that describes (as a single number) the pattern of differences among the means of the different groups being compared.

Step 2: Specifying the Null Distribution

The null hypothesis is that all the data were drawn from the same population, so that any differences between the means of the groups are no larger than would be expected by chance. If this null hypothesis is true, then the variation among groups will be small, and we expect to find an F-ratio of 1.0. The F-ratio will be correspondingly larger than 1.0 if the means of the groups are widely separated (large among-group variation) relative to the variation within groups.[6] In this example, the observed F-ratio of 8.78 is almost 10 times larger than the expected value of 1.0, a result that seems unlikely if the null hypothesis were true.

Step 3: Calculating the Tail Probability

The *P*-value is an estimate of the probability of obtaining an F-ratio = 8.78, given that the null hypothesis is true. Figure 5.5 shows the theoretical distribution of the F-ratio and the observed F-ratio of 8.78, which lies in the extreme right hand tail of the distribution. What is its tail probability (or *P*-value)?

[6] As in Monte Carlo analysis, there is the theoretical possibility of obtaining an F-ratio that is smaller than expected by chance. In such a case, the means of the groups are unusually similar, and differ less than expected if the null hypothesis were true. Unusually small F-ratios are rarely seen in the ecological literature, although Schluter (1990) used them as an index of species-for-species matching of body sizes and community convergence.

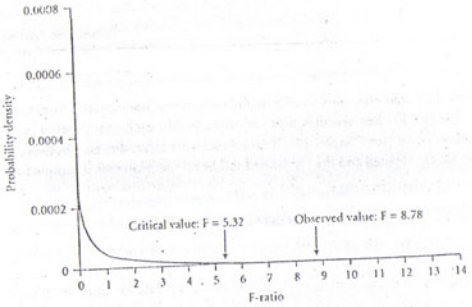


Figure 5.5 Theoretical F-distribution. The larger the observed F-ratio, the more unlikely it would be if the null hypothesis were true. The critical value for this distribution equals 5.32; the area under the curve beyond this point is equal to 5% of the area under the entire curve. The observed F-ratio of 8.78 lies beyond the critical value, so $P(\text{ant data} \mid \text{null hypothesis}) \leq 0.05$. In fact, the actual probability of $P(\text{ant data} \mid \text{null hypothesis}) = 0.018$, because the area under the curve to the right of the observed F-ratio represents 1.8% of the total area under the curve. Compare this result to the *P*-value of 0.036 from the Monte Carlo analysis (Figure 5.3).

The *P*-value of this F-ratio is calculated as the probability mass of the F-ratio distribution (or area under the curve) equal to or greater than the observed F-ratio. For these data, the probability of obtaining an F-ratio as larger or larger than 8.78 (given two groups and a total *N* of 10) equals 0.018. As the *P*-value is smaller than 0.05, we consider it unlikely that such a large F-ratio would occur by chance alone, and we reject the null hypothesis that our data were sampled from a single population.

Assumptions of the Parametric Method

There are two assumptions for all parametric analyses:

1. The data collected represent random, independent samples.
2. The data were sampled from a specified distribution.

As we noted for Monte Carlo analysis, the first assumption of random, independent sampling is always the most important in any analysis. The second assumption is usually satisfied because normal (bell-shaped) distributions are ubiquitous and turn up frequently in the real world.

Specific parametric tests usually include additional assumptions. For example, ANOVA further assumes that variances within each group being compared are equal (see Chapter 10). If sample sizes are large, this assumption can be modestly violated and the results will still be robust. However, if sample sizes are small (as in this example), this assumption is more important.

Advantages and Disadvantages of the Parametric Method

The advantage of the parametric method is that it uses a powerful framework based on known probability distributions. The analysis we presented was very simple, but there are many parametric tests appropriate for complex experimental and sampling designs (see Chapter 7).

Although parametric analysis is intimately associated with the testing of statistical null hypotheses, it may not be as powerful as sophisticated Monte Carlo models that are tailored to particular questions or data. In contrast to Bayesian analysis, parametric analysis rarely incorporates a priori information or results from other experiments. Bayesian analysis will be our next topic—after a brief detour into non-parametric statistics.

Non-Parametric Analysis: A Special Case of Monte Carlo Analysis

Non-parametric statistics are based on the analysis of ranked data. In the ant example, we would rank the observations from largest to smallest and then calculate statistics based on the sums, distributions, or other synthetic measures of those ranks. Non-parametric analyses do not assume a specified parametric distribution (hence the name), but they still require independent, random sampling, (as do all statistical analyses). A non-parametric test is in essence a Monte Carlo analysis of ranked data, and non-parametric statistical tables give *P*-values that would be obtained by a randomization test on ranks. Thus, we have already described the general rationale and procedures for such tests.

Although they are used commonly by some ecologists and environmental scientists, we do not favor non-parametric analyses for three reasons. First, using ranked data wastes information that is present in the original observations. A Monte Carlo analysis of the raw data is much more informative, and often more powerful. One justification that is offered for a non-parametric analysis is that the ranked data may be more robust to measurement error. However, if the original observations are so error-prone that only the ranks are reliable, it is probably a good idea to re-do the experiment using measurement methods that offer greater accuracy. Second, relaxing the assumption of a parametric distribution (e.g., normality) is not such a great advantage, because parametric analyses often are robust to violations of this assumption (thanks to the Central Limit Theorem). Third, non-parametric methods are available only for extremely simple

experimental designs, and cannot easily incorporate covariates or blocking structures. We have found that virtually all needs for statistical analysis of ecological and environmental data can be met by parametric, Monte Carlo, or Bayesian approaches.

Bayesian Analysis

Bayesian analysis is the third major framework for data analysis. Scientists often believe that their methods are "objective" because they treat each experiment as a *tabula rasa* (blank slate): the simple statistical null hypothesis of random variation reflects ignorance about cause and effect. In our example of ant nest densities in forests and fields, our null hypothesis is that the two are equal, or that being in forests and fields has no consistent effect on ant nest density. Although it is possible that no one has ever investigated ant nests in forests and fields before, it is extremely unlikely; our reading of the literature on ant biology prompted us to conduct this particular study. So why not use data that already exist to frame our hypotheses? If our only goal is the hypothetico-deductive one of falsification of a null hypothesis, and if previous data all suggest that forests and fields differ in ant nest densities, it is very likely that we, too, will falsify the null hypothesis. Thus, we needn't waste time or energy doing the study yet again.

Bayesians argue that we could make more progress by specifying the observed difference (e.g., expressed as the DIF or the F-statistic described in the previous sections), and then using our data to extend earlier results of other investigators. Bayesian analysis allows us to do just this, as well as to quantify the probability of the observed difference. This is the most important difference between Bayesian and frequentist methods.

There are six steps in Bayesian inference:

1. Specify the hypothesis.
2. Specify parameters as random variables.
3. Specify the prior probability distribution.
4. Calculate the likelihood.
5. Calculate the posterior probability distribution.
6. Interpret the results.

Step 1: Specifying the Hypothesis

The primary goal of a Bayesian analysis is to determine the probability of the hypothesis *given* the data that have been collected: *P*(H|data). The hypothesis needs to be quite specific, and it needs to be quantitative. In our parametric analysis of ant nest density, the hypothesis of interest (i.e., the alternative hypothesis)

was that the samples were drawn from two populations with different means and equal variances, one for the forest and one for the field. We did not test this hypothesis directly. Instead, we tested the null hypothesis: the observed value of F was no larger than that expected if the samples were drawn from a single population. We found that the observed F-ratio was improbably large ($P = 0.018$), and we rejected the null hypothesis.

We could specify more precisely the null hypothesis and the alternative hypothesis as hypotheses about the value of the F-ratio. Before we can specify these hypotheses, we need to know the critical value for the F-distribution in Figure 5.5. In other words, how large does an F-value have to be in order to have a P-value ≤ 0.05?

For 10 observations of ants in two groups (field and forest), the critical value of the F-distribution (i.e., that value for which the area under the curve equals 5% of the total area) equals 5.32 (see Figure 5.5). Thus, any observed F-ratio greater than or equal to 5.32 would be grounds for rejecting the null hypothesis. Remember that the general hypothetico-deductive statement of the probability of the null hypothesis is $P(\text{data} \mid H_0)$. In the ant nest example, the data result in an F-ratio equal to 8.78. If the null hypothesis is true, the observed F-ratio should be a random sample from the F-distribution shown in Figure 5.5. Therefore we ask what is $P(F_{obs} = 8.78 \mid F_{theoretical})$?

In contrast, Bayesian analysis proceeds by inverting this probability statement: what is the probability of the hypothesis *given* the data we collected [$P(H \mid \text{data})$]? The ant nest data can be expressed as $F = 8.78$. How are the hypotheses expressed in terms of the F-distribution? The null hypothesis is that the ants were sampled from a single population. In this case, the expected value of the F-ratio is small ($F < 5.32$, the critical value). The alternative hypothesis is that the ants were sampled from two populations, in which case the F-ratio would be large ($F = 5.32$). Therefore, the Bayesian analysis of the alternative hypothesis calculates $P(F \geq 5.32 \mid F_{obs} = 8.78)$. By the First Axiom of Probability, $P(F < 5.32 \mid F_{obs}) = 1 - P(F \geq 5.32 \mid F_{obs})$.

A modification of Bayes' Theorem (introduced in Chapter 1) allows us to directly calculate $P(\text{hypothesis} \mid \text{data})$:

$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{hypothesis})P(\text{data} \mid \text{hypothesis})}{P(\text{data})} \quad (5.2)$$

In Equation 5.2, $P(\text{hypothesis} \mid \text{data})$ on the left-hand side of the equation is called the posterior probability distribution (or simply the posterior), and is the quantity of interest. The right-hand side of the equation consists of a fraction. In the numerator, the term $P(\text{hypothesis})$ is referred to as the prior

probability distribution (or simply the prior), and is the probability of the hypothesis of interest *before* you conducted the experiment. The next term in the numerator, $P(\text{data} \mid \text{hypothesis})$, is referred to as the likelihood of the data; it reflects the probability of observing the data given the hypothesis.[7] The denominator, $P(\text{data})$ is a normalizing constant that reflects the probability of the data given all possible hypotheses.[8] Because it is simply a normalizing constant (and so scales our posterior probability to the range [0,1]),

$$P(\text{hypothesis} \mid \text{data}) \propto P(\text{hypothesis})P(\text{data} \mid \text{hypothesis})$$

(where $\propto$ means "is proportional to") and we can focus our attention on the numerator.

Returning to the ants and their F-ratios, we focus on $P(F \geq 5.32 \mid F_{obs} = 8.78)$. We have now specified our hypothesis quantitatively in terms of the relationship between the F-ratio we observe (the data) and the critical value of $F \geq 5.32$ (the hypothesis). This is a more precise hypothesis than the hypothesis dis-

[7] Fisher developed the concept of likelihood as a response to his discomfort with Bayesian methods of inverse probability:

> *What has now appeared, is that the mathematical concept of probability is inadequate to express our mental confidence or diffidence in making such inferences, and that the mathematical quantity which appears to be appropriate for measuring our order of preference among different possible populations does not in fact obey the laws of probability. To distinguish it from probability, I have used the term 'Likelihood' to designate this quantity. (Fisher 1925, p. 10).*

The likelihood is written as $L(\text{hypothesis} \mid \text{data})$ and is directly proportional (but not equal to) the probability of the *observed* data given the hypothesis of interest: $L(\text{hypothesis} \mid \text{data}) = cP(\text{data}_{obs} \mid \text{hypothesis})$. In this way, the likelihood differs from a frequentist P-value, because the P-value expresses the probability of the infinitely many possible samples of the data given the statistical null hypothesis (Edwards 1992). Likelihood is used extensively in information-theoretic approaches to statistical inference (e.g., Hilborn and Mangel 1997; Burnham and Anderson 2002), and it is a central part of Bayesian inference. However, likelihood does not follow the axioms of probability. Because the language of probability is a more consistent way of expressing our confidence in a particular outcome, we feel that statements of the probabilities of different hypotheses (which are scaled between 0 and 1) are more easily interpreted than likelihoods (which are not).

[8] The denominator is calculated as

$$\int P(H_i)P(data \mid H_i)dH$$

cussed in the previous two sections, that there is a difference between the density of ants in fields and in forests.

Step 2: Specifying Parameters as Random Variables

A second fundamental difference between frequentist analysis and Bayesian analysis is that, in a frequentist analysis, parameters (such as the true population means μ_{forest} and μ_{field}, their standard deviations σ^2, or the F-ratio) are *fixed*. In other words, we assume there is a *true value* for the density of ant nests in fields and forests (or at least in the field and forest that we sampled), and we estimate those parameters from our data. In contrast, Bayesian analysis considers these parameters to be *random variables*, with their own associated parameters (e.g., means, variances). Thus, for example, instead of the population mean of ant colonies in the field being a fixed value μ_{field}, the mean could be expressed as a normal random variable with its own mean and variance: $\mu_{field} \sim N(\lambda_{field}, \sigma^2)$. Note that the random variable representing ant colonies does not have to be normal. The type of random variable used for each population parameter should reflect biological reality, not statistical or mathematical convenience. In this example, however, it is reasonable to describe ant colony densities as normal random variables: $\mu_{field} \sim N(\lambda_{field}, \sigma^2)$, $\mu_{forest} \sim N(\lambda_{forest}, \sigma^2)$.

Step 3: Specifying the Prior Probability Distribution

Because our parameters are random variables, they have associated probability distributions. Our unknown population means (the λ_{field} and λ_{forest} terms) themselves have normal distributions, with associated unknown means and variances. To do the calculations required by Bayes' Theorem, we have to specify the prior probability distributions for these parameters—that is, what are probability distributions for these random variables *before* we do the experiment?[9]

[9] The specification of priors is a fundamental division between frequentists and Bayesians. To a frequentist, specifying a prior reflects subjectivity on the part of the investigator, and thus the use of a prior is considered unscientific. Bayesians argue that specifying a prior makes explicit all the hidden assumptions of an investigation, and so it is a more honest and objective approach to doing science. This argument has lasted for centuries (see reviews in Effron 1986 and in Berger and Berry 1988), and was one reason for the marginalization of Bayesians within the statistical community. However, the advent of modern computational techniques allowed Bayesians to work with uninformative priors, such as the ones we use here. It turns out that, with uninformative priors, Bayesian and frequentist results are very similar, although their final interpretations remain different. These findings have led to a recent renaissance in Bayesian statistics, although simple and easy-to-use software is still lacking.

We have two basic choices for specifying the prior. First, we can comb and reanalyze data in the literature, talk to experts, and come up with reasonable estimates for the density of ant nests in fields and forests. Alternatively, we can use an uninformative prior, for which we initially estimate the density of ant nests to be equal to zero and the variances to be very large. (In this example, we set the population variances to be equal to 100,000.) Using an uninformative prior is equivalent to saying that we have no prior information, and that the mean could take on virtually any value with roughly equal probability.[10] Of course, if you have more information, you can be more specific with your prior. Figure 5.6 illustrates the uninformative prior and a (hypothetical) more informative one.

Similarly, the standard deviation σ for μ_{field} and μ_{forest} also has a prior probability distribution. Bayesian inference usually specifies an inverse gamma distribution[11] for the variances; as with the priors on the means, we use an uninformative prior for the variance. We write this symbolically as $\sigma^2 \sim IG(1{,}000, 1{,}000)$ (read "the variance is distributed as an inverse gamma distribution with parameters 1,000 and 1,000). We also calculate the *precision* of our estimate of variance, which we symbolize as τ, where $\tau = 1/\sigma^2$. Here τ is a gamma random variable (the inverse of an inverse gamma is a gamma), and we write this symbolically as $\tau \sim \Gamma(0.001, 0.001)$ (read "tau is distributed as a gamma distribution with parameters 0.001, 0.001"). The form of this distribution is illustrat-

[10] You might ask why we don't use a uniform distribution, in which all values have equal probability. The reason is that the uniform distribution is an improper prior. Because the integral of a uniform distribution is undefined, we cannot use it to calculate a posterior distribution using Bayes' Theorem. The uninformative prior N(0,100,000) is nearly uniform over a huge range, but it can be integrated. See Carlin and Louis (2000) for further discussion of improper and uninformative priors.

[11] The gamma distribution for precision and the inverse gamma distribution for variance are used for two reasons. First, the precision (or variance) needs to take on only positive values. Thus, any probability density function that has only positive values could be used for priors for precision or variance. For continuous variables, such distributions include a uniform distribution that is restricted to positive numbers and the gamma distribution. The gamma distribution is somewhat more flexible than the uniform distribution and allows for better incorporation of prior knowledge.

Second, before the use of high-speed computation, most Bayesian analyses were done using conjugate analysis. In a conjugate analysis, a prior probability distribution is sought that has the same form as the posterior probability distribution. This convenient mathematical property allows for closed-form (analytical) solutions to the complex integration involved in Bayes' Theorem. For data that are normally distributed, the conjugate prior for the parameter that specifies the mean is a normal distribution, and the conjugate prior for the parameter that specifies the precision (= 1/variance) is a gamma distribution. For data that are drawn from a Poisson distribution, the gamma distribu-

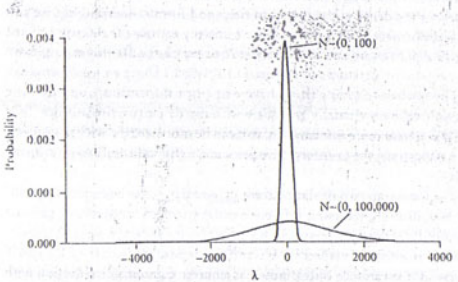


Figure 5.6 Prior probability distributions for Bayesian analysis. Bayesian analysis requires specification of prior probability distributions for the statistical parameters of interest. In this analysis of the data in Table 5.1, the parameter is average ant density, λ. We begin with a simple uninformative prior probability distribution that average ant density is described by a normal distribution with mean 0 and standard deviation of 100,000 (green curve). Because the standard deviation is so large, the distribution is nearly uniform over a large range of values: between −1500 and +1500, the probability is essentially constant (~0.0002), which is appropriate for an uninformative prior. The black curve represents a more precise prior probability distribution. Because the standard deviation is much smaller (100), the probability is no longer constant over a larger range of values, but instead decreases more sharply at extreme values.

tion is the conjugate prior for the parameter that defines the mean (or rate) of the Poisson distribution. For further discussion, see Gelman et al. (1995).

The gamma distribution is a two-parameter distribution, written as $\Gamma(a,b)$, where a is referred to the shape parameter and b is the scale parameter. The probability density function of the gamma distribution is

$$P(X) = \frac{b^a}{\Gamma(a)} X^{(a-1)} e^{-bX}, \text{ for } X > 0$$

where $\Gamma(a)$ is the gamma function $\Gamma(n) = (n-1)!$ for integers $n > 0$. More generally, for real numbers z, the gamma function is defined as

$$\Gamma(z) = \int_0^1 \left[\ln\frac{1}{t}\right]^{z-1} dt$$

The gamma distribution has expected value $E(X) = a/b$ and variance $= a/b^2$. Two distributions used commonly by statisticians are special cases of the gamma distribution. The χ^2 distribution with ν degrees of freedom is equal to $\Gamma(\nu/2, 0.5)$. The exponential distribution with parameter β that was discussed in Chapter 2 is equal to $\Gamma(1,\beta)$.

Finally, if the random variable $1/X \sim \Gamma(a,b)$, then X is said to have an inverse gamma (IG) distribution. To obtain an uninformative prior for the variance of a normal random variable, we take the limit of the IG distribution as a and b both approach 0. This is the reason we use $a = b = 0.001$ as the prior parameters for the gamma distribution describing the precision of the estimate.

ed in Figure 5.7. This use of precision should make some intuitive sense; because our estimate of variability decreases when we have more information, the precision of our estimate increases. Thus, a high value of precision equals a low variance of our estimated parameter.

We now have our prior probability distributions. The unknown population means of ants in fields and forests, μ_{field} and μ_{forest}, are both normal random variables with expected values equal to λ_{field} and λ_{forest} and unknown (but equal) variances, σ^2. These expected means themselves are normal random variables with expected values of 0 and variances of 100,000. The population precision τ is the reciprocal of the population variance σ^2, and is a gamma random variable with parameters (0.001, 0.001):

$$\mu_i \sim N(\lambda_i, \sigma^2)$$
$$\lambda_i \sim N(0, 100{,}000)$$
$$\tau = 1/\sigma^2 \sim \Gamma(0.001, 0.001)$$

These equations completely specify $P(\text{hypothesis})$ in the numerator of Equation 5.2. If we had real prior information, such as the density of ant nests in other fields and forests, we could use those values to more accurately specify the expected means and variances of the λ's.

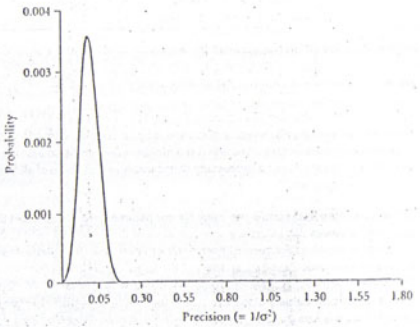


Figure 5.7 Uninformative prior probability distribution for the precision (= 1/variance). Bayesian inference requires not only a specification of a prior distribution for the mean of the variable (Figure 5.6), but also a specification for the precision (= 1/variance). Bayesian inference usually specifies an inverse gamma distribution for the variances. As with the distribution of the means, an uninformative prior is used for the variance. In this case, the variance is distributed as an inverse gamma distribution with parameters 1000 and 1000: $\sigma^2 \sim IG(1{,}000, 1{,}000)$. Because the variance is very large, the precision is small.

Step 4: Calculating the Likelihood

The other quantity in the numerator of Bayes' Theorem (Equation 5.2) is the likelihood, $P(\text{data}_{obs} \mid \text{hypothesis})$. The likelihood is a distribution that is proportional to the probability of the observed data given the hypothesis.[12] Each parameter λ_i and τ of our prior probability distribution has its own likelihood function. In other words, the different values of λ_i have likelihood functions that are normal random variables with means equal to the observed means (here, 7 ant nests per quadrat in the forest and 10.75 ant nests per quadrat in the field; see Table 5.2). The variances are equal to the sample variances (4.79 in the forest and 2.25 in the field). The parameter τ has a likelihood function that is an inverse gamma random variable. Finally, the F-ratio is an F-random variable with expected value (or maximum likelihood estimate)[13] equal to 8.78 (calculated from the data using Equation 5.1).

Step 5: Calculating the Posterior Probability Distribution

To calculate the posterior probability distribution, $P(H \mid \text{data})$, we apply Equation 5.2, multiply the prior by the likelihood, and divide by the normalizing constant (or marginal likelihood). Although this multiplication is straightforward for well-behaved distributions like the normal, computational methods are used to iteratively estimate the posterior distribution for any prior distribution (Carlin and Louis 2000).

In contrast to the results of a parametric or Monte Carlo analysis, the result of a Bayesian analysis is a probability *distribution*, not a single *P*-value. Thus, in this example, we express $P(F \geq 5.32 \mid F_{obs})$ as a *random variable* with expected mean and variance. For the data in Table 5.1, we calculated posterior estimates for all the parameters: λ_{field}, λ_{forest}, $\sigma^2 (= 1/\tau)$ (Table 5.7). Because we used uninformative priors, the parameter estimates for the Bayesian and parametric analyses are similar, though not identical.

The hypothesized F-distribution with expected value equal to 5.32 is shown in Figure 5.8. To compute $P(F \geq 5.32 \mid F_{obs})$, we simulated 20,000 F-ratios using a Monte Carlo algorithm. The average, or expected value, of all of these F-ratios is 9.77; this number is somewhat larger than the frequentist (maximum likelihood) estimate of 8.78 because our sample size is very small ($N = 10$). The spread about the mean is large: SD = 7.495; hence the precision of our estimate is relatively low (0.017).

Step 6: Interpreting the Results

We now return to the motivating question: What is the probability of obtaining an F-ratio ≥ 5.32, given the data on ant nest density in Table 5.1? In other words, how probable is it that the mean ant nest densities in the two habitats really differ? We can answer this question directly by asking what percentage of values in Figure 5.8 are greater than or equal to 5.32. The answer is 67.3%. This doesn't look quite as convincing as the *P*-value of 0.018 (1.8%) obtained in the parametric analysis in the previous section. In fact, the percentage of values in Figure 5.7 that are ≥ 8.78, the observed value for which we found $P = 0.018$ in the parametric analysis section, is 46.5. In other words, the Bayesian analysis (Figure 5.8) indi-

[12] There is a key difference between the likelihood function and a probability distribution. The probability of data given a hypothesis, $P(\text{data} \mid H)$, is the probability of any set of random data given a specific hypothesis, usually the statistical null hypothesis. The associated probability density function (see Chapter 2) conforms to the First Axiom of Probability—that the sum of all probabilities = 1. In contrast, the likelihood is based on only one dataset (the observed sample) and may be calculated for many different hypotheses or parameters. Although it is a function, and results in a distribution of values, the distribution is not a probability distribution, and the sum of all likelihoods does not necessarily sum to 1.

[13] The maximum likelihood is the value for our parameter that maximizes the likelihood function. To obtain this value, take the derivative of the likelihood, set it to 0, and solve for the parameter values. Frequentist parameter estimates are usually equal to maximum-likelihood estimates for the parameters of the specified probability density functions. Fisher claimed, in his system of fiducial inference, that the maximum-likelihood estimate gave a realistic probability of the alternative (or null) hypothesis. Fisher based this claim on a statistical axiom that he defined by saying that given observed data *Y*, the likelihood function $L(H \mid Y)$ contains all the relevant information about the hypothesis H. We do not discuss maximum likelihood estimates any further, as (a) modern computation techniques normally provide them in lieu of asymptotic values for frequentist statistics; and (b) Bayesians use the entire likelihood function in their calculations. For further reading on likelihood methods, see Berger and Wolpert (1984) and Edwards (1992).

TABLE 5.7 Parametric and Bayesian estimators for the means and standard deviations of the data in Table 5.1

Analysis	λ_{Forest}	λ_{Field}	σ_{Forest}	σ_{Field}
Parametric	7.00	10.75	2.19	1.50
Bayesian (uninformed prior)	6.97	10.74	0.91	1.13
Bayesian (informed prior)	7.00	10.74	1.01	1.02

The standard deviation estimators from the Bayesian analysis are slightly smaller because the Bayesian analysis incorporates information from the prior probability distribution. Bayesian analysis may give different results, depending on the shape of the prior distribution and the sample size.

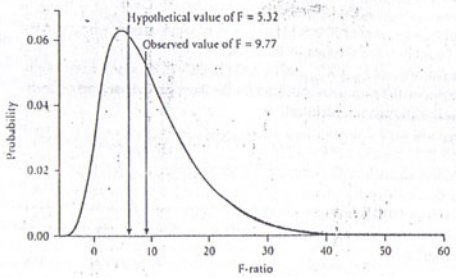


Figure 5.8 Hypothesized F-distribution with an expected value of 5.32. We are interested in determining the probability of $F \geq 5.32$ (the critical value for $P < 0.05$ in a standard F-ratio test), given the data on ant nest densities in Table 5.1. This is the inverse of the traditional null hypothesis, which asks: what is the probability of obtaining the data, given the null hypothesis? In the Bayesian analysis, the posterior probability of $F \geq 5.32$, given the data in Table 5.1, is the proportion of the area under the curve to the right of F = 5.32, which is 0.673. In other words, P(hypothesis that the fields and forests differ in average density of ant nests | observed data in Table 5.1) = 0.673. The most likely posterior value of the F-ratio is 9.77. The proportion of area under the curve to the right of this value is 0.413. The parametric analysis says that the observed data are unlikely given a null distribution specified by the F-ratio [$P(\text{data} \mid H_0) = 0.018$], whereas the Bayesian analysis says that the probability of observing an F-ratio of 9.77 or larger is not unlikely given the data [$P(F \geq 5.32 \mid \text{data}) = 0.673$].

cates $P = 0.67$ that ant nest densities in the two habitats are truly different, given the Bayesian estimate of F = 9.77 [$P(F \geq 5.32 \mid F_{obs}) = 0.67$]. In contrast, the parametric analysis (Figure 5.5) indicates $P = 0.018$ that the parametric estimate of F = 8.78 (or a greater F-ratio) would be found given the null hypothesis that the ant densities in the two habitats are the same [$P(F_{obs} \mid H_0) = 0.018$].

Using Bayesian analysis, a different answer would result if we used a different prior distribution rather than the uninformative prior of means of 0 with large variances. For example, if we used prior means of 15 for the forest and 7 for the field, an among-group variance of 10, and a within-group variance of 0.001, then $P(F \geq 5.32 \mid \text{data}) = 0.57$. Nevertheless, you can see that the posterior probability does depend on the priors that are used in the analysis (Table 5.7).

Finally, we can estimate a 95% Bayesian credibility interval around our estimate of the observed F-ratio. As with the Monte Carlo method, we estimate

the 95% credibility interval as the 2.5 and 97.5 percentiles of the simulated F-ratios. These values are 0.28 and 28.39. Thus, we can say that we are 95% sure that the value of the F-ratio for this experiment lies in the interval [0.28, 28.39]. Note that the spread is large because the precision of our estimate of the F-ratio is low, reflecting the small sample size in our analysis. You should compare this interpretation of a credibility interval with the interpretation of a confidence interval presented in Chapter 3.

Assumptions of Bayesian Analysis

In addition to the standard assumptions of all statistics methods (random, independent observations), the key assumption of Bayesian analysis is that the parameters to be estimated are random variables with known distributions. In our analysis, we also assumed little prior information (uninformative priors), and therefore the likelihood function had more influence on the final calculation of the posterior probability distribution than did the prior. This should make intuitive sense. On the other hand, if we had a lot of prior information, our prior probability distribution (e.g., the black curve in Figure 5.6) would have low variance and the likelihood function would not substantially change the variance of the posterior probability distribution. If we had a lot of prior information, and we were confident in it, we would not have learned much from the experiment. A well-designed experiment should decrease the estimate of the posterior estimate relative to the prior variance.

The relative contributions of prior and likelihood to the posterior estimate of the probability of the mean density of nests in the forest are illustrated in Figure 5.9. In this figure, the prior is flat over the range of the data (i.e., it is an unin-

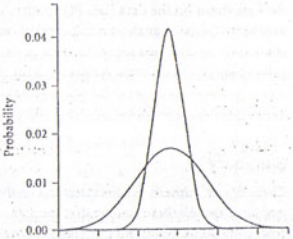


Figure 5.9 Probability densities for the prior, likelihood, and posterior for the mean number of ant nests in the forest plots. In the Bayesian analysis of the data in Table 5.1, we used an uninformative prior distribution with a mean of 0 and a variance of 100,000. This normal distribution generated an essentially uniform range of prior probability values (black) over the range of values for λ_{forest}, the density of forest ants. The likelihood (green) represents the probability based on the observed data (Table 5.1), and the posterior probability (gray) is the product of the two. Notice that the posterior distribution is more precise than the likelihood because it takes into account the (modest) information contained in the prior.

formative prior), the likelihood is the distribution based on the observed data (Table 5.1), and the posterior is the product of the two. Note that the variance of the posterior is smaller than the variance of the likelihood because we had some prior information. However, the expected values of the likelihood (7.0) and the posterior (6.97) are very close, because all values of the prior were approximately equally likely over the range of the data.

After doing this experiment, we have new information on each of the parameters that could be used in analysis of subsequent experiments. For example, if we were to repeat the experiment, we could use the posterior in Figure 5.9 as our prior for the average density of ant nests in other forests. To do this, we would use the values for λ_i and σ_i in Table 5.7 as the estimates of λ_i and σ_i in setting up the new prior probability distributions.

Advantages and Disadvantages of Bayesian Analysis

Bayesian analysis has a number of advantages relative to parametric and Monte Carlo approaches conducted in a frequentist framework. Bayesian analysis allows for the explicit incorporation of prior information, and the results from one experiment (the posterior) can be used to inform (as a prior) subsequent experiments. The results of Bayesian analysis are interpreted in an intuitively straightforward way, and the inferences obtained are conditional on both the observed data and the prior information.

Disadvantages to Bayesian analysis are its computational challenges (even currently available software is difficult to use) and the requirement to condition the hypothesis on the data [i.e., *P*(hypothesis | data)]. The most serious disadvantage of Bayesian analysis is its potential lack of objectivity, because different results will be obtained using different priors. Consequently, different investigators may obtain different results from the same dataset if they start with different preconceptions or prior information. The use of uninformative priors addresses this criticism, but increases the computational complexity.

Summary

Three major frameworks for statistical analysis are Monte Carlo, parametric, and Bayesian. All three assume that the data were sampled randomly and independently. In Monte Carlo analysis, the data are randomized or reshuffled, so that individuals are randomly re-assigned to groups. Test statistics are calculated for these randomized datasets, and the reshuffling is repeated many times to generate a distribution of simulated values. The tail probability of the observed test statistic is then estimated from this distribution. The advantage of Monte Carlo analysis is that it makes no assumptions about the distribution of the data,

it makes the null hypothesis clear and explicit, and it can be tailored to individual datasets and hypotheses. The disadvantage is that it is not a general solution and usually requires computer programming to be implemented.

In parametric analysis, the data are assumed to have been drawn from an underlying known distribution. An observed test statistic is compared to a theoretical distribution based on a null hypothesis of random variation. The advantage of parametric analysis is that it provides a unifying framework for statistical tests of classical null hypotheses. Parametric analysis is also familiar to most ecologists and environmental scientists and is widely implemented in statistical software. The disadvantage of parametric analysis is that the tests do not specify the probability of alternative hypotheses, which often is of greater interest than the null hypothesis.

Bayesian analysis considers parameters to be random variables as opposed to having fixed values. It can take explicit advantage of prior information, although modern Bayesian methods rely on uninformative priors. The results of Bayesian analysis are expressed as probability distributions, and their interpretation conforms to our intuition. However, Bayesian analysis requires complex computation and often requires the investigators to write their own programs.