CBE—Life Sciences Education
Vol. 9, 298–310, Fall 2010

Article

On a Calculus-based Statistics Course for Life Science Students

Joseph C. Watkins

Department of Mathematics, University of Arizona, Tucson, AZ 85721-0089

Submitted March 17, 2010; Revised June 7, 2010; Accepted June 21, 2010
Monitoring Editor: John Jungck

The choice of pedagogy in statistics should take advantage of the quantitative capabilities and scientific background of the students. In this article, we propose a model for a statistics course that assumes student competency in calculus and a broadening knowledge in biology. We illustrate our methods and practices through examples from the curriculum.

INTRODUCTION

When considering quantitative training for the next generation of scholars, the most persistently requested advanced skills expressed by the research-active life scientists at the University of Arizona are the use of statistics and comfort with calculus. Consequently, one centerpiece for the curricular activities at the University of Arizona is the development of three core mathematics courses–a two-semester-long sequence that integrates calculus and differential equations and, based upon that knowledge, a one-semester-long course in statistics.

Benefiting from changes in approach in the calculus and differential equations course, students enrolled in the statistics course are acquainted and have some facility with open-ended questions. In addition, because these students are typically juniors and seniors, they bring a much broader knowledge base in the life sciences than they had when they entered the calculus classroom for the first time.

Most life science students, presumed to be proficient in college algebra, are taught a variety of procedures and standard tests under a well-developed pedagogy. This approach is sufficiently refined so that students have a good intuitive understanding of the underlying principles presented in the course. However, if the needs presented by the science fall outside the battery of methods presented in the standard curriculum, then students are typically at a loss to adjust the procedures to accommodate the additional demand.

On the other hand, mathematics students frequently have a course in the theory of statistics as a part of their undergraduate program of study. In this case, the standard curriculum repeatedly finds itself close to the very practically minded subject that statistics is. However, the demands of the syllabus provide very little time to explore these applications with any sustained attention.

Our goal for life science students at the University of Arizona is to find a middle ground. In their overview "Undergraduate Statistics Education and the National Science Foundation," Hall and Rowell (2008) note that statistics education reformers have until recently overlooked the issues associated with an introductory postcalculus statistics curriculum. Their notable exceptions are the data-oriented active-learning approach of Rossman and Chance (2004), Virtual Laboratories in Statistics by Siegrist (2004), and case studies–based approaches to teach mathematical statistics (Nolan and Speed, 1999, 2000).

Despite the fact that calculus is a routine tool in the development of statistics, the benefits to students who have learned calculus are very rarely used in the statistics curriculum for undergraduate biology students. Our objective is to meet this need with a one-semester course in statistics that moves forward in recognition of the coherent body of knowledge provided by statistical theory having an eye consistently on the application of the subject. Even though such a course may not be able to achieve the same degree of completeness now presented by the two more standard courses described above, the question is whether it leaves the student capable of understanding what statistical thinking is, understanding how to integrate this with scientific procedures and quantitative modeling, how to ask statistics experts productive questions, and how to implement their ideas using statistical software and

DOI: 10.1187/cbe.10–03–0035
Address correspondence to: Joseph C. Watkins (jwatkins@math.arizona.edu).

other computational tools. For a thoughtful essay on these issues, see Nolan and Lang (2009).

The efforts described above, despite having similar goals, arrive at very different approaches. In this article, we shall introduce the course at the University of Arizona with an annotated syllabus and through classroom examples, take home assignments, and end-of-the-semester projects.

AN ANNOTATED SYLLABUS

The four parts of the course—organizing and collecting data, an introduction to probability, estimation procedures, and hypothesis testing—are the standard building blocks of many statistics courses. This section highlights some of the features of a calculus-based course.

Organizing and Collecting Data

Much of this is standard and essential—organizing categorical and quantitative data, appropriately displayed as bar charts, histograms, boxplots, time plots, and scatterplots, and summarized using medians, quartiles, means, weighted means, trimmed means, standard deviations, and regression lines. We use this as an opportunity to introduce students to the statistical software package R (R Development Core Team, 2009) and to add additional summaries like the empirical cumulative distribution function and the empirical survival function. One example integrating the use of this is the comparison of the lifetimes of wild-type and transgenic mosquitoes and a discussion of the best strategy to display and summarize data if the goal is to examine the differences in these two genotypes of mosquitoes in their ability to carry and spread malaria. A bit later, we will do an integration by parts exercise to show that the mean lifetime of the mosquitoes is the area under the survival function.

Collecting data under a good design is introduced early in the course, and discussion of the underlying principles of experimental design is an abiding issue throughout this course. With each new mathematical or statistical concept comes an enhanced understanding of what an experiment might uncover through a more sophisticated design than what was previously thought possible. The students are given readings on design of experiment and examples using R to create simple and stratified random samples. The recommended lecture is to tell a story that illustrates how these issues appear in personal research activities.

Introduction to Probability

Probability theory is the analysis of random phenomena. It is built on the axioms of probability and is explored, for example, through the introduction of random variables. The goal of probability theory is to uncover properties arising from the phenomena under study. Statistics is devoted to the analysis of data. The goal of statistical theory is to articulate as well as possible what model of random phenomena underlies the production of the data. The focus of this section of the course is to develop those probabilistic ideas that relate most directly to the needs of statistics.

Thus, we must study the axioms of probability to the extent that the students understand conditional probability and independence. Conditional probability is necessary to develop Bayes formula, which we will later use to give a taste of the Bayesian approach to statistics. Independence will be needed to describe the likelihood function in the case of an experimental design that is based on independent observations. Densities for continuous random variables and mass function for discrete random variables are necessary to write these likelihood functions explicitly. Expectation will be used to standardize a sample sum or sample mean and to perform method of moments estimates.

Random variables are developed for a variety of reasons. Some, like the Poisson random variable or the gamma random variable, arise from considerations based on Bernoulli trials or exponential waiting. The hypergeometric random variable helps us understand the difference between sampling with and without replacement. The F, t, and χ^2 random variables will later become test statistics. Uniform random variables are the ones simulated by random number generators. Because of the central limit theorem, the normal family is the most important among the list of parametric families of random variables.

The flavor of the course returns to becoming more authentically statistical with the law of large numbers and the central limit theorem. These are largely developed using simulation explorations and first applied to simple Monte Carlo techniques and importance sampling to evaluate integrals. One cautionary tale is an example of the failure of these simulation techniques when applied without careful analysis. If one uses, for example, Cauchy random variables in the evaluation of some quantity, then the simulated sample means can appear to be converging only to experience an abrupt and unpredictable jump. The lack of convergence of an improper integral reveals the difficulty.

The central object of study is, of course, the central limit theorem. It is developed both in terms of sample sums and sample means and used in relatively standard ways to estimate probabilities. However, in this course, we can introduce the delta method, which adds ideas associated to the central limit theorem to the context of propagation of error.

Estimation Procedures

In the simplest possible terms, the goal of estimation theory is to answer the question: *What is that number?* An estimator is a statistic (i.e., a function of the data). We look to two types of estimation techniques—method of moments and maximum likelihood and several criteria for an estimator (e.g., bias and variance). Several examples including capture-recapture and the distribution of fitness effects are developed for both types of estimators. The variance is estimated using the delta method for method of moments estimators and using Fisher information for maximum likelihood estimators. An analysis of bias is based on quadratic Taylor series approximations and the properties of expectations. R is routinely used in simulations to gain insight into the quality of estimators.

The point estimation techniques are followed by interval estimation and, notably, by confidence intervals. This brings us to the familiar one- and two-sample t intervals for population means and one- and two-sample z intervals for population proportions. In addition, we can return to the delta method and the observed Fisher information to construct confidence intervals associated, respectively, with method of moment estimators and maximum likelihood estimators.

Hypothesis Testing

For hypothesis testing, we begin with the central issues—null and alternative hypotheses, type I and type II errors, test statistics and critical regions, significance, and power. We begin with the ideas of likelihood ratio tests as best tests for a simple hypothesis. This is motivated by a game. Extensions of this result, known as the Neyman Pearson lemma, form the basis for the *t* test for means, the χ^2 test for goodness of fit, and the *F* test for analysis of variance. These results follow from the application of optimization techniques from calculus, including Langrange multiplier techniques to develop goodness of fit tests.

The desire of a powerful test is articulated in a variety of ways. In engineering terms, power is called sensitivity. We illustrate this with a radon detector. An insensitive instrument is a risky purchase. This can be either because the instrument is substandard in the detection of fluctuations or poor in the statistical test that results in an algorithm to announce a change in radon level. An insensitive detector has the undesirable property of not sounding its alarm when the radon level has indeed risen.

The course ends by looking at the logic of hypotheses testing and the results of different likelihood ratio analyses applied to a variety of experimental designs. The delta method allows us to extend the resulting test statistics to multivariate nonlinear transformations of the data.

EXAMPLES FROM THE CURRICULUM

In this section, we describe three examples from the University of Arizona course. These examples are presented to highlight how a calculus-based course differs from an algebra-based course. These abbreviated descriptions cannot bring the same sense of background preparation or the breadth of issues under consideration that a student experiences in the classroom. Consequently, more detailed notes are available from the author upon request.

For the first two examples, we begin with the presentation seen in a typical algebra-based statistics course and then provide extensions of these ideas that are possible for those students who have good calculus skills. The third example describes a strategy to introduce the concept of likelihood. In subsequent classes, the students will use this as motivation for the rationale for the optimization problems in the development of likelihood ratio tests.

Extensions on Regression

Given observations $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$, ordinary linear regression has as its goal to find the best linear fit $g(x|\alpha, \beta) = \alpha + \beta x$ to the data. The least squares criterion refers to the minimization of the sum of squares

$$SS(\alpha, \beta) = \sum_{i=1}^{n} (y_i - g(x_i \mid \alpha, \beta))^2$$

over all real parameter values α and β. Calling $\hat{\alpha}$ and $\hat{\beta}$ the values that achieve this minimum, we obtain the regression line

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

fit to the response γ_i. Ordinary linear regression is a staple of any introductory statistics course [see, e.g., Moore *et al.* (2007) or Agresti and Franklin (2008)]. Students compute regression lines using software or a formula and learn to check the appropriateness of the regression line by examining the residuals (i.e., the differences $y_i - \hat{y}_i$ between the data and the fit). Informally, a good fit would show no structure in the residual plot. Two departures from this are common. Either

- the size of the residuals depends on x, or
- the sign of the residuals depends on x.

The goal here is to have the ability to extend the use of regression beyond the formulas and qualitative reasoning. To address the first case above, we can modify the least squares criterion and solve a weighted least squares regression with weight function w,

$$SS_w(\alpha, \beta) = \sum_{i=1}^{n} w(x_i)(y_i - g(x_i \mid \alpha, \beta))^2.$$

Later in the course, after the students have learned about maximum likelihood estimation, they learn that the weights should often be chosen to be inversely proportional to the variance of the residual.

Here, we focus on the second case in which the scatterplots display a curved relationship. With the tools of college algebra, this relationship can be transformed to be suitable for linear regression by guessing the relationship and applying the inverse transformation to the response variable. This is seen most frequently in scatterplots of a quantity over time exhibiting exponential growth or decay. Thus, we take the logarithm of the response variable and proceed.

This strategy is sometimes too superficial to be successful. For students who have been acquainted with differential equations, we can consider a common chemical reaction,

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P$$

Here E is an enzyme, S is the substrate, ES is the substrate-bound enzyme, and P is the product. The production rate, $V = d[P]/dt$, of the product is measured for five different concentrations $[S]$ of the substrate to obtain the following data: (See Figure 1.)

$[S]$ (mM)	1	2	5	10	20
V (nmol/sec)	1.0	1.5	2.2	2.5	2.9

Naive attempts to give a transformation that yields a linear relationship are unlikely to succeed. Despite intense speculation, students have never been able to guess correctly. A model-based approach is needed. With this in mind, we begin with the law of mass action to obtain differential equations for the concentrations

$$\frac{d[ES]}{dt} = k_1[E][S] - (k_{-1} + k_2)[ES] \text{ and } \frac{d[P]}{dt} = k_2[ES].$$

The strategy used by Michaelis and Menten applies to situations in which the concentration of the substrate-bound

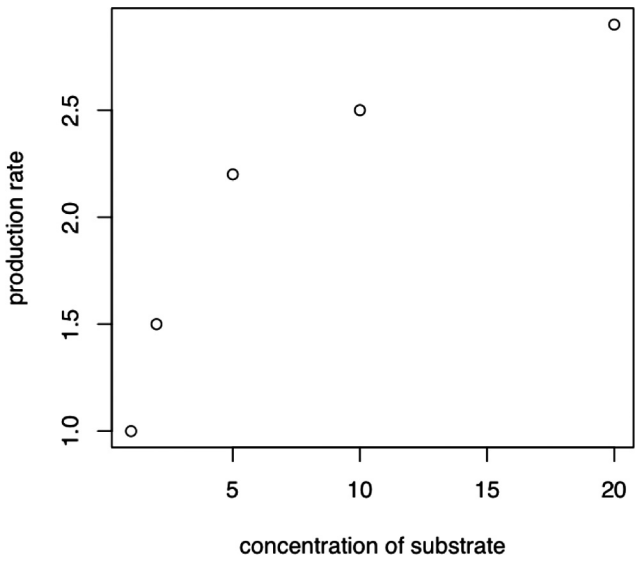

Figure 1. Michaelis–Menten kinetics. Measurements of production rate $V = d[P]/dt$ versus $[S]$, substrate concentration.

enzyme (and hence also the unbound enzyme) change much more slowly than those of the product and substrate. If, by measuring the change of $[ES]$ directly over time or by arguing that if the concentration of enzyme $[E]$ is small compared with the concentration of substrate $[S]$, we find that

$$0 \approx \frac{d[ES]}{dt} \text{ and, thus, } [E][S] \approx K_m[ES] \text{ where } K_m = \frac{k_{-1} + k_2}{k_1},$$

the Michaelis constant. Equipped with these ideas, we find after some algebraic manipulation, the well-known Lineweaver-Burk double reciprocal plot

$$\frac{1}{V} = \frac{K_m + [S]}{V_{\max}[S]} = \frac{K_m}{V_{\max}}\frac{1}{[S]} + \frac{1}{V_{\max}}.$$

In this way, we have found that the desired linear relationship is between $1/V$ and $1/[S]$. We can now perform ordinary least squares on the transformed data and estimate $V_{\max}$, the maximum rate of production, and K_{m}. [See, for example, Nelson and Cox (2005), pp. 204–210.]

However, this classical approach to estimation is no longer considered satisfactory (Piegorsch and Bailer, 2005). Examination of residual plots reveals the drawback—the residuals measured as the reciprocal of the concentration become magnified for very small concentrations. Consequently, the preferred method is to use a nonlinear least squares criterion. In this case, given data $(V_1, [S]_1)$, $(V_2, [S]_2)$, … $(V_n, [S]_n)$, find the values of $V_{\max}$ and $[S]$ that minimize

$$SS(V_{\max}, K_m) = \sum_{j=1}^{n} (V_j - g([S]_j \mid V_{\max}, K_m))^2$$

where

$$g([S] \mid V_{\max}, K_m) = V_{\max}\frac{[S]}{K_m + [S]}.$$

This minimization problem does not yield explicit equations for the estimators $\hat{V}_{max}$ and $\hat{K}_m$. Rather than embarking on a minicourse on numerical techniques for optimization, we use this opportunity to discuss the intricacies of the nonlinear regression analysis and strategies to engage an expert in statistical science on the science, the experimental protocol, and the data.

One of the aspects of any new course is the understandable uncertainty by the students that the approaches presented in class have applicability to questions of their own concern. Thus, in the problem sets, we ask the students to use the ideas presented in lecture and make substantial use of them in another biological context. In this case, the ideas on nonlinear transformations are further explored by the students in the following question addressed by Wiehe and Stephan (1993):

Due to selection, neutral sites near genes are hitchhiked along with the gene to higher levels in the population until a recombination event separates the neutral site from the selected gene. This reduces the diversity of the genome near genes. The nucleotide diversity π versus recombination rate ρ is given for 17 gene regions in *Drosophila melanoganster*. After a short amount of reasoning based on the nature of recombination and selection, the students learn the relationship

$$\pi = \frac{\alpha\rho}{\beta + \rho}.$$

This expression is also amenable to a Lineweaver-Burk double reciprocal plot. Thus, we rewrite this expression as

$$\frac{1}{\pi} = \frac{\beta}{\alpha}\frac{1}{\rho} + \frac{1}{\alpha},$$

a linear expression in the variables $1/\pi$ and $1/\rho$. As the scatterplot of these reciprocal variables in Figure 2 shows, we can see that the correlation is not very high. In this case, strength of selection is a good candidate for a hidden or lurking variable. In this data set, we are likely looking at genes that range from small selection that will not reduce diversity much to large selection that will. In the same way that estimated value of $V_{\max}$ gives the reaction rate at high concentration of the substrate, the estimated value of α gives the nucleotide diversity at recombination distances ρ far away from genes. This value gives a sense of the intrinsic diversity of the *Drosophila* genome due solely to the effects of neutral mutation and demographic history.

The Central Limit Theorem, Propagation of Error, and the Delta Method

George Polya (1920) used the phrase the central limit theorem because of its centrality to the theory of probability. In simple terms, the central limit theorem states that, irrespective of the distribution of the observations from a simple random sample, its sample mean $\bar{X}$ has, for sufficiently many observations, approximately a normal distribution. Many of the commonly used hypotheses rely on a test statistic that is based on the normal distribution. Thus, the use of the t test, the χ^2 test, and the analysis of variance F test depends on the appropriateness of the normal distribution

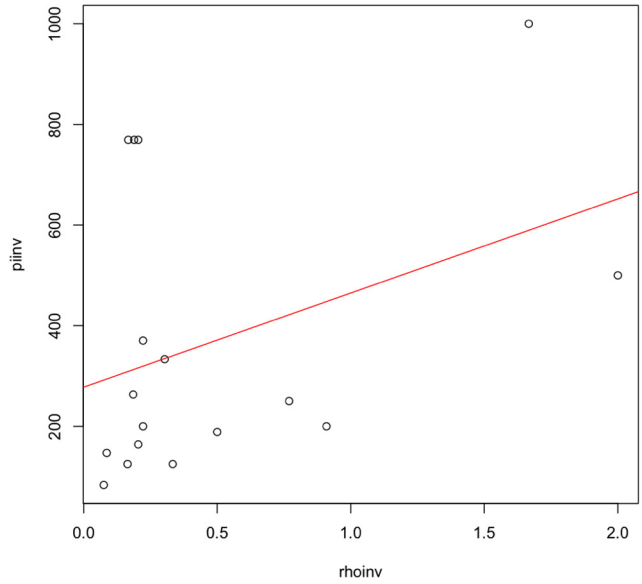

Figure 2. Double reciprocal plot for genetic hitchhiking. $1/\pi$ versus $1/\rho$. The regression line y intercept is 277.8, giving an estimate $\hat{\alpha} = 1/277.8 = 0.0036$.

as an approximation for the distribution of sample means computed from the data. So, understanding the use of these statistics depends in part on grasping the logic behind the central limit theorem.

Unfortunately, the best proofs of the central limit theorem rely on indirect and sophisticated methods. Moreover, these proofs yield very little insight into the emergence, with an increasing number of observations, of the bell curve for the distribution of $\bar{X}$. Consequently, rather than proving the central limit theorem in an introductory statistics course at this level, the typical pedagogical choice is to use graphical and other empirical methods to convey the key ideas (see Ross (2009) or Pitman (1999) for a proof).

Most elementary courses dedicate a week to the understanding of the central limit theorem. This development might culminate with a graph like that seen in Figure 3. This shows the standardization of random variables and relates it to the test statistics that the students will later encounter in the study of hypothesis testing. This can then be used to estimate probabilities for $\bar{X}$ by evaluating z scores.

The extension we can have in a calculus-based course is motivated by interest in nonlinear transformations of the measured quantities. Some physics, chemistry, and engineering students have seen a bit of this in the study of propagation of error (see, e.g., Meyer [1975], Bevington and Robinson [2002]). Having investigated the properties of variance and covariance in the context of the independence of random variables, the goal is to combine the benefits of the central limit theorem and the propagation of error analyses.

After investigating some one-dimensional examples like those seen in Figure 4, the students are prepared to examine a more sophisticated example. For this, we consider a suggestion by Powell (2007) to questions in avian biology. To introduce one of his examples, define the fecundity B as the mean number of female fledglings per year. Then, B is a product of three quantities,

$$B = F \cdot p \cdot N,$$

where F equals the mean number of female fledglings per successful nest, p equals nest survival probability, and N equals the mean number of nests built per female per year. Let's collect measurements on n_1 nests to count female fledglings in a successful nest, check n_2 nests for survival probability, and follow n_3 females to count the number of successful nests per year. To begin, assume that the experimental design is structured so that measurements are independent.

A critical value for B is 1. If B is greater than 1, then the population grows over time. If it is less than 1, then the population is headed toward extinction. So, for example, if our estimate $\hat{B} = 1.03$, then we may or may not be confident that the actual value of B is greater than 1 depending on the way the estimator $\hat{B}$ distributes its values.

The central limit theorem can directly determine appropriate normal distributions to approximate $\bar{F}$, the sample mean of the number of female fledglings per successful nest, $\hat{p}$, the sample proportion of surviving nests, and $\bar{N}$, the sample mean number of nests built per female per year. If our estimate $\hat{B}$ of the fecundity is the product of these three numbers, then what is a good approximation for the distribution of this statistic?

Propagation of error analysis suggests that we find a linear approximation to B. Because the measurements for F,

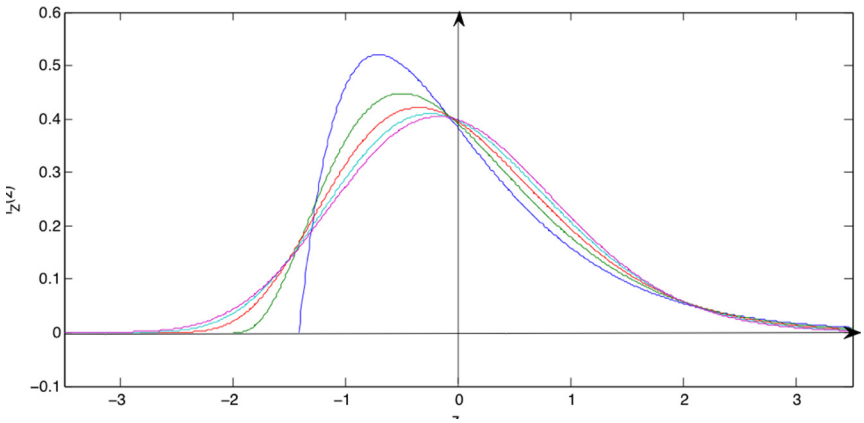

Figure 3. Displaying the central limit theorem graphically. Density of the standardized version of the sum of n independent exponential random variables for $n = 2$ (dark blue), 4 (green), 8 (red), 16 (light blue), and 32 (magenta). Note how the skewness of the exponential distribution slowly gives way to the bell curve shape of the normal distribution.

p, and N are independent, then $\sigma_{\hat{B}}^2$, the variance of $\hat{B}$ is easy to approximate. This solution give us the size of the error, but we do not yet know its distribution.

The delta method extends the propagation of error analysis by noting that, by the central limit theorem, the random quantity given by the linear approximation can be approximated by a normal distribution. Returning to the question of estimating fecundity, the delta method tells us that $\hat{B}$ can be approximated by a normal distribution. The variance can be determined from σ_F^2, the variance in the fecundity measurement and σ_N^2, the variance in the number of nests built per adult female per year. By combining this information, we derive in the class an expression that highlights in the contribution to the variance in the estimate of fecundity originating from each of the three measurements:

$$\frac{\sigma_{\hat{B}}^2}{B^2} \approx \frac{1}{n_1}\frac{\sigma_F^2}{F^2} + \frac{1}{n_2}\frac{1-p}{p} + \frac{1}{n_3}\frac{\sigma_N^2}{N^2}.$$

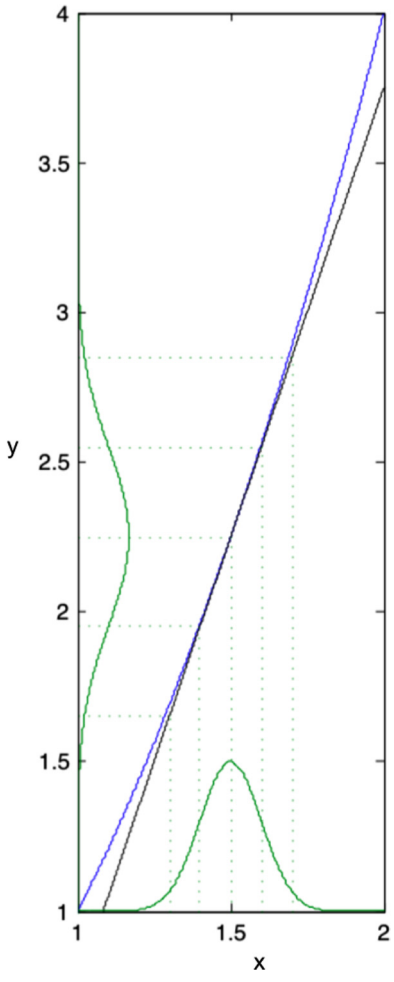

Figure 4. Illustrating the delta method. Here the mean $\mu = EX = 1.5$ and the blue curve $h(x) = x^2$. Thus, $\overline{X}$ is approximately normal with mean close to 2.25 and $\sigma_{h(\bar{X})} \approx 30\sigma_{\bar{X}}$. The bell curve on the y axis is the reflection of the bell curve on the x axis about the (black) tangent line $y = h(\mu) + h'(\mu)(x - \mu)$.

This formula now plays an essential role in settling on a data collecting protocol. With a goal to bring the standard deviation σ_B down as much as possible given available resources, the choices of n_1, n_2, and n_3 are under control of the biologists. In addition, if independence of observations between the estimation of F, p, and N is an issue, we can extend the propagation of error and with it the delta method by including the covariances between the pairs of the sets of observations.

This application of the calculus greatly extends the applicability of the central limit theorem in the practical application of statistics. This idea is explored by the students who apply the delta method to describe the estimator for the focal length f of a convex lens based on repeated measurement of the distance, s_1, to an object and the distance, s_2, to its image using the thin lens formula,

$$\frac{1}{f} = \frac{1}{s_1} + \frac{1}{s_2}.$$

(See Figure 5.)

The appearances of the delta method do not end here. We shall use this technique to determine the variance for a method of moments estimator. Later, we can use these ideas to construct test statistics for hypothesis tests beyond those generally encountered in more elementary courses.

Likelihood Ratios

The study of statistical hypotheses begins with a shot of jargon that will take a student some time to absorb. Understanding the terminology is important because it codifies the

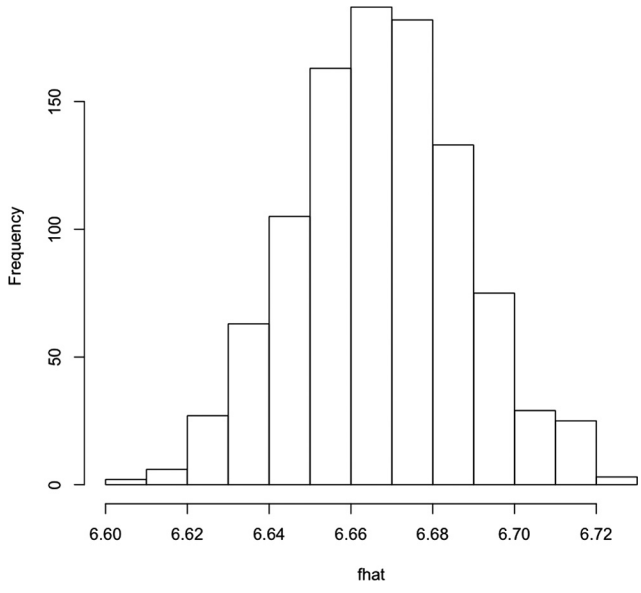

Figure 5. Simulating a sampling distribution for the estimate of focal length. In this example, the distance from a convex lens to an object $s_1 = 12$ cm and to its image $s_2 = 15$ cm. The standard deviation of the measurement is 0.1 cm for s_1 and 0.5 cm for s_2. Based on 25 measurements each for s_1 and s_2, the delta method gives 0.0207 for the standard deviation of the estimate for f. This agrees to three decimal places with the value based on the 1000 simulations summarized in the histogram above. The bell curve shape shows the quality of an approximation to a normal random variable.

relationship between hypothesis testing and scientific advancement. Under a simple hypothesis, we write the test as:

H_0 : data are from distribution 0

versus H_1 : data are from distribution 1.

H_0 is called the null hypothesis. H_1 is called the alternative hypothesis. The possible actions are:

- Reject the hypothesis. Rejecting the hypothesis when it is true is called a type I error or a false positive. Its probability α is called the size of the test or the significance level.
- Fail to reject the hypothesis. Failing to reject the hypothesis when it is false is called a type II error or a false negative. If the false negative probability is β, then the power of the test is $1 - \beta$.

The rejection of the hypothesis is based on whether or not the data X land in a critical region C. Thus,

reject H_0 if and only if $X \in C$.

Given a choice α for the size of the test, a critical region C is called best or most powerful if it has the lowest probability of a type II error among all regions that have size α (Table 1).

	H_0 is true	H_1 is true
reject H_0	type I error	OK
fail to reject H_0	OK	type II error

Table 1: **The language of hypothesis testing.**

$$\beta = P\{X \notin C \mid H_1 \text{ is true}\}$$

To see whether the students can discover on their own the best possible test, we play a game. The goal here is to establish a foundation for the fundamental role of likelihood ratios. This will form the rationale behind the optimization problems that must be solved in order to derive likelihood ratio tests. Both the null hypothesis (our side) and the alternative hypothesis (our opponent) are given 100 points among the values for the data X that run from -11 to 11. These are our likelihood functions. These can be created and displayed quickly in R using the commands:

```
> x<-c(-11:11)
> L0<-c(0,0:10,9:0,0)
> L1<-sample(L0,length(L0))
> data.frame(x,L0,L1)
```

Thus L_1 is a random rearrangement of the values in L_0. Here is the output from one simulation.

The goal of this game is to pick values x so that your accumulated points increase as quickly as possible from your likelihood L_0, keeping your opponent's points from L_1 as low as possible. The natural start is to pick values of x so that $L_1(x) = 0$. Then, the points you collect begin to add up without your opponent gaining anything.

x	-2	3	5	7
L_0 total	8	15	20	23
L_1 total	0	0	0	0

Being ahead by a score of 23–0 can be translated into a best critical region in the following way. If we take as our critical region C all the values for x except -2, 3, 5, and 7, then, the size of the test $\alpha = 0.77$ and the power of the test $1 - \beta = 1.00$ because there is no chance of type II error with this critical region.

Understanding the next choice is crucial. Candidates are

$x = 4$, with $L_0(4) = 6$ against $L_1(4) = 1$ and

$x = 1$, with $L_0(1) = 9$ against $L_1(1) = 2$.

After some discussion, the class will come to the conclusion that having a high ratio is more valuable than having a high difference. The choice 6 against 1 is better than 9 against 2 because choosing 6 against 1 twice will put us in a better place than the single choice of 9 against 2. Now we can pick the next few candidates, keeping track of the size and the power of the test with the choice of critical region being the values of x not yet chosen (see Table 2).

From this exercise we see how the likelihood ratio test is the choice for a most powerful test. This is more carefully described in the Neyman-Pearson lemma [See Hoel *et al.* (1972) and Hogg and Tanis (2009)].

Neyman Pearson Lemma

Let L_0 denote the likelihood function for the random variable X corresponding to H_0 and L_1 denote the likelihood function for the random variable X corresponding to H_1. If there exists a critical region C of size α and a nonnegative constant k such that

$$\frac{L_1(x)}{L_0(x)} \geq k \text{ for } x \in C \text{ and } \frac{L_1(x)}{L_0(x)} \leq k \text{ for } x \notin C,$$

then C is the most powerful critical region of size α.

Using R, we can complete the table for L_0 total and L_1 total.

```
> o<-order(L1/L0)
> sumL0<-cumsum(L0[o])
> sumL1<-cumsum(L1[o])
> alpha<-1-sumL0/100
> beta<-sumL1/100
> data.frame(x[o],L0[o],L1[o],
+ sumL0,sumL1,alpha,1-beta)
```

x	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
$L_0(x)$	0	0	1	2	3	4	5	6	7	8	9	10	9	8	7	6	5	4	3	2	1	0	0
$L_1(x)$	3	8	7	5	7	1	3	10	6	0	6	4	2	5	0	1	0	4	0	8	2	9	9

x	-2	3	5	7	4	1	-6	0	-5
L_0 total	8	15	20	23	29	38	42	52	57
L_1 total	0	0	0	0	1	3	4	8	11
α	0.92	0.85	0.80	**0.77**	0.71	0.62	0.58	0.48	0.43
$1-\beta$	1.00	1.00	1.00	**1.00**	0.99	0.97	0.96	0.92	0.89

Table 2: Results for Neyman-Peason game.

Completing the curve, known as the receiver operator characteristic (ROC), is shown in Figure 6. The ROC can be used to discuss the inevitable trade-offs between type I and type II errors. For example, by the mere fact that the graph is increasing, we can see that by setting a more rigorous test achieved by lowering the level of significance (decreasing the value on the horizontal axis) necessarily reduces the power (decreasing the value on the vertical axis) [see, e.g., Zweig and Campbell (1993)].

Now the students are prepared to understand the reasoning behind the use of critical values for the t statistic, the χ^2 statistic, or the F statistic as related to the critical value k in extensions of the ideas of likelihood ratios and can extend the likelihood ratio tests to more novel situations.

These examples have been presented here in such a way as to emphasize transitions from the use of algebra to the use of calculus in the learning of concepts in statistics. However, students do not notice such a shift in the sense that they do not see calculus as a special tool. Limits need to be evaluated, rates change, areas under curves need to be determined, functions need to be maximized, and their concavity needs to be assessed. Moreover, calculus is more than a powerful computational tool: it provides a way of thinking that enhances the students' view of what is possible. I asked a student how much is calculus used in the course. The answer she gave, "So much we don't think about it," is a goal for this course.

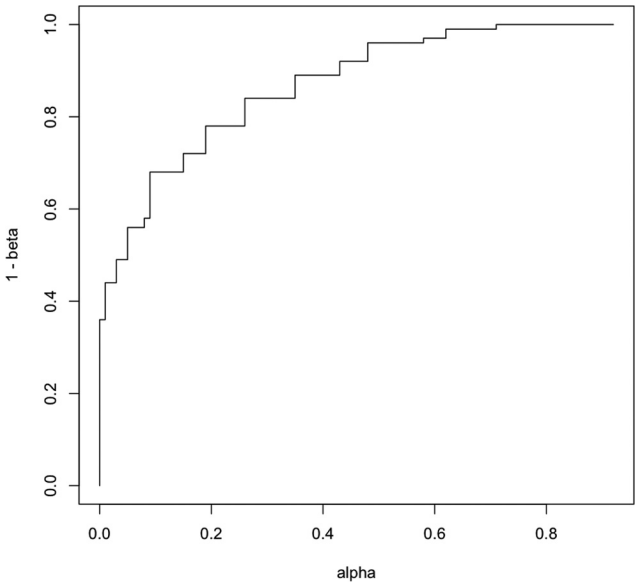

Figure 6. Receiver operator characteristic. The graph of $\alpha = P\{X \in C \mid H_0 \text{ is true}\}$ (significance) versus $1-\beta = P\{X \in C \mid H_1 \text{ is true}\}$ (power) in the example. The horizontal axis α is also called the false positive fraction (FPF). The vertical axis $1-\beta$ is also called the true positive fraction (TPF).

For many years, quantitative subjects as diverse as physics and economics have developed distinct pedagogies for an algebra-based and for a calculus-based course. For example, Newton's second law applied to projectile motion or to the motion of springs gives, for the calculus student, simple differential equations from which the basic algebraic or trigonometric relationships for the variables position, velocity, acceleration, and time can be derived. For the algebra student, these relationships have to be taken on faith. In a similar way, students can receive a more comprehensive and elegant understanding of the fundamental principles of statistical science if they are equipped with a working knowledge of calculus.

PROJECT DESCRIPTIONS

One important feature of the course is the end-of-the-semester project. Typically for this assignment, students work in pairs. All of the suggested projects are based on research activities that have taken place at the University of Arizona, and all of the projects require statistical analyses beyond the methods presented in the course. This gives the students the opportunity to work as a team on a project and to work with one of the scientists (typically a graduate student or postdoctoral fellow) who was involved in the original research. This way, the students obtain a first-hand account of the fundamental questions that are meant to be addressed by the research and learn how to use the ideas from their statistics course and apply them to new situations. They certainly see the nature of open-ended questions that are a part of a research scientist's daily life.

We describe two projects to give a sense of the breadth of the choices from the point of view of both biology and statistics.

Language and Genes in Sumba

Human populations and the languages that they speak change over time. The movement of people and the innovations they make in their languages are difficult to observe and quantify over short periods and impossible to witness over long periods. Consequently, researchers have been forced to undertake indirect approaches to infer associations between human languages and human genes. Many well-known studies focus their questions on the movement of people on continental scales. These stud-

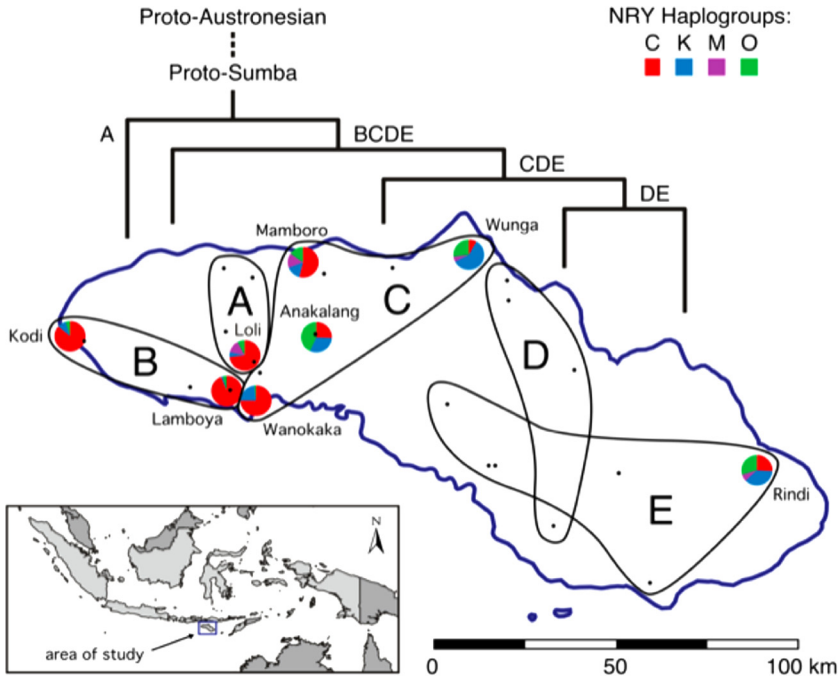

Figure 7. Phylogenetic and geographic distribution of languages and Y chromosome haplogroups.

ies led Diamond and Bellwood (2003) to suggest that many of the correlations that we presently see in languages and genes result from the movement of prehistoric farmers from the places in which these agricultural techniques first arose.

Using the Indonesian island of Sumba as a test case, Lansing *et al.* (2007) look to see the degree in which these correlations can be seen at much smaller scales. Both genetic and archeological evidence place the first migration of anatomically modern humans in Southeast Asia and Oceana between 40,000 and 45,000 years ago. Further archeological evidence places the transition on Sumba from the original hunter-gatherer technology to the neolithic technology between 3500 and 4000 years ago. At that time, a small number of farmers speaking an Austronesian language likely came into contact with resident foragers speaking presumably a Papuan language.

To investigate the paternal histories of the Sumbanese, Lansing *et al.* (2007) obtained genetic samples from 352 men inhabiting eight villages. Their genetic information is derived from the Y chromosome. Based on the Y Chromosome Consortium worldwide genealogical tree of paternal ancestry, haplogroup O appears to be associated with the expansion of Austronesian societies from southeast Asia to Indonesia and Oceana.

The linguistic data consist of 29 200-word Swadesh lists from sites well distributed throughout the island. Swadesh lists are built from words ascribed to meanings that are basic to everyday life. Using the methodology of comparative linguistics, some words from different language lists can be traced to a common ancestral word. These techniques have been used previously to construct a Proto-Austronesian (PAn) language. In addition, a phylogenetic tree built from these 29 word lists branches to form five major language subgroups and leads us to the conclusion that the present day Sumbanese all speak a language derived from a single common ancestral Proto-Sumbanese. From this we can count the number of words on the Swadesh word lists that are derived from the PAn language. This information is summarized in Figure 7 (Lansing *et al.*, 2007).

Statistical Procedures

The Mantel test (Sokal and Rohlf, 1994) computes the significance of the correlation between two positive symmetric $n \times n$ matrices. The ij entry in a matrix is meant to give a distance between sites i and j. Distance matrices M and N are square and the calculations for the test are carried out on the entries above the diagonal. The computation yields a statistic:

$$Z = \sum_{i=1}^{n} \sum_{j=i+1}^{n} M_{ij} N_{ij}$$

similar to the correlation statistic well known to the students. The null hypothesis is that the observed relationship between the two distance matrices could have been obtained by any random arrangement of the observations.

In this example, we have $n = 8$ villages and three measures of distance—geographic, linguistic, and genetic. In previous correlation computations, changing one observation was not at all prohibited by the structure of the problem. However, distance matrices are highly constrained. For example, if we change one entry in the matrix by altering distance from one village to the next, then we must also change other entries in the matrix to compensate. These constraints force us to look for a refinement in the standard hypothesis testing procedure for correlation.

To see whether the value of Z is bigger than what we might find by chance, we perform a permutation on one of the matrices. For example, if we were to keep the labels on the villages geographic distances in M and rearrange the distances in the N matrix with a permutation π, then we have a new statistic:

$$Z^{\pi} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} M_{ij} N_{\pi(i)\pi(j)}.$$

If the alternative hypothesis holds, then Z is likely to be bigger than most of the Z^{π}. Indeed, the P value of the test is a fraction of the Z^{π} that are larger than Z. The results of these tests are summarized in Table 3.

	r	P-value
Genetics/geography	0.011	0.518
Genetics/language	0.358	0.023
Geography/language	0.673	< 0.001

Table 3: Mantel test results for correlation.

We next test the null assumption of no correlation between the fraction of O haplogroup men and retained Swadesh list PAn cognates among the eight villages. The alternative is that these two quantities are positively correlated. A naive strategy to analyze the correlation in Figure 8 is to perform simple linear regression using software and report the P value for a test of zero slope. This method wastes much of the effort in collecting the substantial amount of genetic and linguistic information necessary to plot each of the points in Figure 8. We can tell that the method is not most powerful in that if we were to double the amount of data and obtain the same regression line, then the P value remains the same.

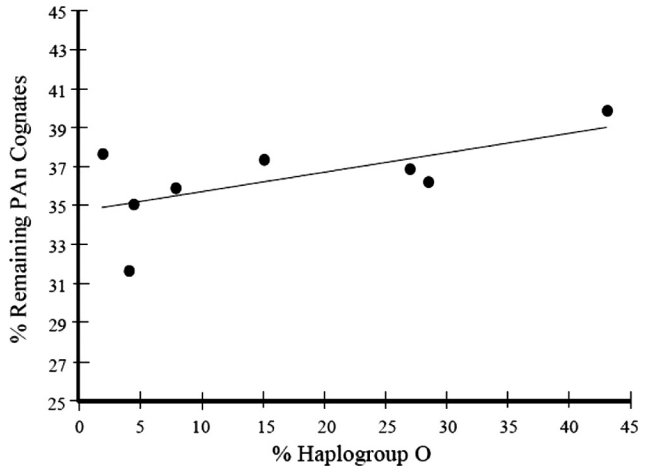

Figure 8. Scatterplot of PAn cognates versus the percentage of sample from haplogroup O. The correlation is 0.627 (Lansing *et al.*, 2007).

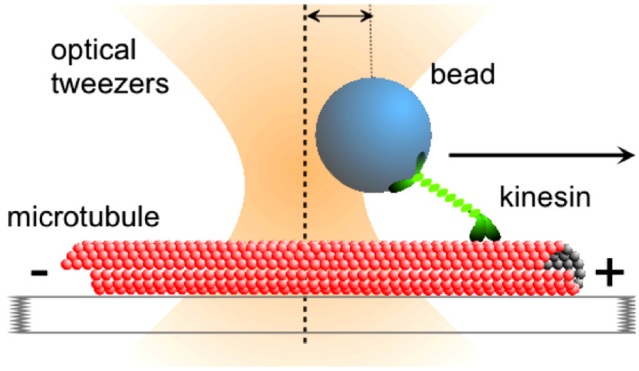

Figure 9. Single molecule experiments. Set-up for kinesin attached to a bead, translocating along a microtubule subject to optical tweezers pulling on bead. (Walton, 2002).

An exact or even an approximate computation of the distribution of correlation under the null hypothesis is difficult. Consequently a bootstrap analysis (Efron and Tibshirani, 1994) is used. Let p be the fraction of men throughout all of Sumba typed as O haplogroup and let q be the fraction of neutral words throughout all of Sumba that are PAn cognates. Under the null hypothesis, the data from each of the villages are independent Bernoulli trials based on these parameters. The students know that, under these conditions, the distribution of O haplogroup men and PAn cognate words from each village will have binomial distributions and can easily simulate them using R. What is unknown are the actual values of p and q. The bootstrap suggests that these two parameters should be estimated from the actual data, estimations that the students have seen in simpler contexts. Next, pseudodata should be simulated repeatedly under the null hypothesis.

For each of the resulting bootstrap samples, the students can compute the correlation. Repeat this procedure many times to create the bootstrap distribution of correlation under the null hypothesis. The bootstrapped P value, 0.047, for the test of no correlation is simply the fraction of bootstrap correlations greater than 0.627, the observed value of correlation.

Force-dependent Kinetic Models for Single Molecules

One of the most widely used relations in chemistry is the Arrhenius relation,

$$k \, \alpha \exp - \frac{\Delta G^{\ddagger}}{k_B T}.$$

Here k is a reaction rate, k_B is Boltzmann's constant, T is the absolute temperature, and $\Delta G^{\ddagger}$ is the free energy (see, Nelson and Cox [2005], pp. 489–506.) Svante Arrhenius could not have imagined the development of micromanipulation techniques that allow measurement and control of piconewton size forces and nanometer size displacements with time resolution as short as a millisecond (see Figure 9). Such modern techniques have made it possible to probe, under

mechanical load, the kinetics of biomolecular processes at the single-molecule level. The goal of this research is to examine the Arrhenius relationship at this level.

Taking displacement to be our reaction coordinate, the free energy $\Delta G^{\ddagger}$ may be split into the sum of the unloaded free energy, $\Delta G_0^{\ddagger}$, and the work done against or assisted by the external force, F,

$$\Delta G^{\ddagger} = \Delta G_0^{\ddagger} + F\delta$$

where δ is the displacement to the transition state. The sign of δ depends on whether the applied force helps or hinders the reaction. Setting k_0 to be the zero-force rate, we have

$$k = k_0 e^{F\delta/k_BT}$$

for the form of the force-dependent rates in the system.

At the level of a single interaction, then for a simple reaction,

$$\text{bound} \rightarrow \text{unbound},$$

describing an irreversible detachment of two molecules with no substeps, the physics leads us to two assumptions:

- The dwell time is based on thermal fluctuations and thus possesses the memorylessness property. In other words, the dwell times are random and follow an exponential distribution (see Ross [2009]). Thus, the density of dwell times takes the form

$$f(t \mid \lambda) = \lambda \exp(-\lambda t)$$

for some rate $\lambda > 0$. The mean of this random variable is $1/\lambda$.

- The Arrenhius relation holds at the level of a single molecule, i.e., the mean dwell time equals

$$\tau_0 e^{-F\delta/k_BT}, \text{ where } \tau_0 = \frac{1}{k_0}.$$

Consequently, $\lambda = e^{F\delta/k_BT}/\tau_0$ and the density of dwell times is

$$f(t \mid \tau_0, \delta, F) = \frac{1}{\tau_0} e^{F\delta/k_BT} \exp\left(-\frac{t}{\tau_0} e^{F\delta/k_BT}\right).$$

The force F is under the control of the experimentalist. The parameters τ_0 and δ need to be estimated from data. Moreover, because we have no way of estimating either τ_0 or δ separately, they need to be estimated simultaneously.

Previously published methods used seat of the pants bootstrapping techniques. The graduate student who brought this problem to my attention performed some numerical experimentation, and it appears that the estimators do not even converge as the number of observations increases. However, students can apply the concepts in the course to design most powerful tests and with it asymptotically narrowest possible confidence intervals.

Statistical Procedures

If our data are from an experiment with independently measured forces $\mathbf{F} = (F_1, F_2, \ldots, F_n)$ and corresponding independent dwell times $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, we have an explicit expression for the likelihood function $L(\tau_0, \delta | \mathbf{t}, \mathbf{F})$. The equations for the maximum likelihood estimators,

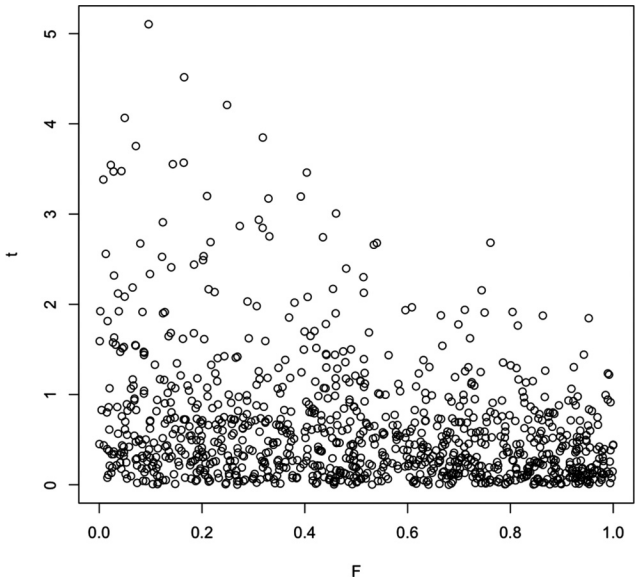

Figure 10. Force versus dwell time for the simulated data. The Arrhenius relationship is not easy to visualize in the scatterplot of 1000 observations.

$$\frac{\partial}{\partial \delta} L(\hat{\tau}_0, \hat{\delta} \mid \mathbf{t}, \mathbf{F}) = 0 \text{ and } \frac{\partial}{\partial \tau_0} L(\hat{\tau}_0, \hat{\delta} \mid \mathbf{t}, \mathbf{F}) = 0,$$

do not separate to produce closed forms for the maximum likelihood estimators $\hat{\tau}_0$ and $\hat{\delta}$. However, they can be rearranged to obtain two algebraic relationships.

$$\hat{\tau}_0 = \frac{1}{n} \sum_{i=1}^{n} t_i e^{F_i \hat{\delta}/k_BT} \text{ and } \hat{\tau}_0 = \frac{1}{n\bar{F}} \sum_{i=1}^{n} t_i F_i e^{F_i \hat{\delta}/k_BT}.$$

This facilitates a simple graphical interpretation for $\hat{\tau}_0$ and $\hat{\delta}$ as the intersection point of the curves determined by these two relationships.

This experiment has not yet been performed (Kalafut, Liang, Watkins, and Visscher, unpublished results), so we ask the students to simulate data and find the maximum likelihood estimates. This gives us the opportunity to discuss the value of generating data via simulation and test the inference methods on these data before moving to the actual data. To keep the situation simple, we take $\delta = 1$ and $\tau_0 = 1$. We also take the forces uniformly distributed between 0 and k_bT. As can be seen from the simulated data, any relationship in Figure 10 between the dwell times and the force would be difficult to discern by inspection.

To simulate the data with $n = 1000$:

```
> F<-runif(1000); t<-rep(0,1000)
> for(k in 1:1000){t[k]<-rexp(1,exp(F[k]))}
> plot(F,t)
```

If we solve numerically for the estimates in this simulation, we obtain $\hat{\tau}_0$ and $\hat{\delta}$ for the intersection of the two curves defined above. This is displayed graphically in Figure 11.

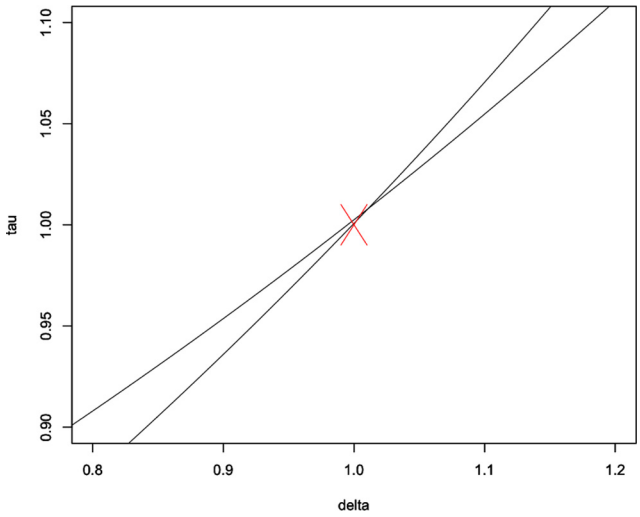

Figure 11. Maximum likelihood estimation. The two curves are the two expressions for $\hat{\tau}_0$ as a function of $\hat{\delta}$ for the simulated data. Here $\hat{\tau}_0 =$ 1.008 and $\hat{\delta} =$ 1.011. The actual values, $\tau_0 = 1$ and $\delta = 1$, are indicated by the red $\times$.

The covariance matrix for $\hat{\tau}_0 =$ 1.008 and $\hat{\delta} =$ 1.011 can be approximated by the inverse of the Fisher information matrix. In this case, we can compute this matrix explicitly and take its inverse.

$$I(\tau_0, \delta)^{-1} = \frac{1}{n\mathrm{var}(F)} \begin{pmatrix} \tau_0^2 \bar{F}^2 & \tau_0 k_B T \bar{F} \\ \tau_0 k_B T \bar{F} & (k_B T)^2 \end{pmatrix}.$$

This information matrix is the fundamental object that directs the design of an experiment. For example, we can reduce the variance of our estimators $\hat{\tau}_0$ and $\hat{\delta}$ and hence the length of confidence intervals by making the variance of F as large as possible. This is accomplished by having the force be as small as the laser trap allows and as large as possible maintaining the stability for the single molecule. On the other hand the variance depends on the product $n\mathrm{var}(F)$. Thus, the ability to perform many measurements may be an easier experiment to perform to achieve a desired length for confidence intervals.

The remaining three projects are drawn from questions on bacterial growth and division, on otoacoustic emissions in the human ear, and on flour beetle population dynamics. The expectation is that if the student has a genuine interest in the life sciences, then at least one of these projects will be enticing.

ASSESSMENT AND IMPACT

At the University of Arizona, we are seeing a transformation in the view that mathematics plays in the education of life scientist students. To see this evolution reflected in the statistics class, we saw six students choose to add a mathematics minor during the first time the course was given. For the second and third time, the students entered the class as mathematics minors. The course, offered for a fourth time this fall, is full and the capacity is being increased to accommodate demand.

The head of our math center is now writing to students who have completed the first two years of mathematics course work and is encouraging them to take the calculus-based statistics course. In addition, more than a fourth of the students (26 of 98) in the Undergraduate Biology Research Program are either mathematics majors or minors. Several researchers who have had statistics students in their lab are recommending it to other students.

From the course evaluations, the students especially liked "the applicability of statistics to multiple areas of interest," that "the homework sets were challenging and engaging," "relating course work to software used in research," having "taken stats classes before I feel like I actually learned some of the basics of stats in this one," "learn(ing) about the real world," "forc(ing) me to learn a subject I really didn't enjoy (I like stats now)," that "the study examples were particularly effective as motivation to learn the material," "using R and relating everything to the real world," and that "the homework was nicely challenging." Thus, we can see that students are favorably impressed by the applicability of the course and the ability of theoretical considerations to lead to strategies for addressing practical problems. After some gentle reminding, they recognize that the material in some of the more mathy courses in their past were essential in their ability to address life science questions with their present level of understanding. Students are showing more facility with software as they continue to use it and having a general sense that mathematical and statistical issues are aspects of the challenges that come with exciting research.

The students, especially those who are working in a laboratory, were invited to select their own topic of interest. About a third of the students picked this option. Yeast genetics, earthquake prediction, rat behavior, fruit fly behavior, strength of material, educational value of certain curricula and evaluation methods, land subsidence, malaria prevalence, and the effects of monetary policy were among the chosen topics. Most of these student-generated projects resulted in a presentation in their research group's lab meeting.

To get a sense of the source and motivation of this transformation, we are now systematically gathering information on changes in student and educator attitude as a consequence of a variety of experiences, including this course. Analysis of these data will bring further insights into the type of attitudes that students who choose this course have and how their attitude is impacted by the experiences of the course.

DISCUSSION

The National Research Council Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century (2003) produced *BIO2010: Transforming Undergraduate Education for Future Research Biologists* for the National Research Council of the National Academies. They remark that "Mathematics teaching presents a special case. Most biology majors take no more than one year of calculus, although some also take an additional semester of statistics. Very few are exposed to discrete mathematics, linear alge

bra, probability, and modeling topics, which could greatly enhance their future research careers." The issues of the synergies of these mathematics subjects are left largely unexplored in *BIO2010*. Some connections are well known and have become standard in the curriculum. Knowledge of linear algebra is certainly essential in learning both linear models in statistics and multidimensional differential equations. Students with a solid background in discrete mathematics have an advantage in the early stages of a course in probability. Any additional course will bring its complement of tools for a modeling course. Here we explore the how the combined effect of the unfolding understanding of biology and proficiency in calculus impact the students' ability to gain a firm foundation in statistical science.

As the course described in this article settles on an approach and a curriculum, we can now see that the uses of calculus in these contexts turn out to be neither particularly difficult nor novel for a student comfortable with the subject. In addition, after a couple of frustrating weeks with syntax, students continue acquiring skills in the use of statistical software. What is difficult for the students is the increased maturity necessary to apprehend the breadth of applicability that statistics brings to any carefully conceived data-rich exploration. In the end, having the tools of algebra and calculus, with probability as a frame of reference and ready access to computational software, students can kindle excitement in their new discoveries in the life science and bring themselves to a level of understanding that is not accessible absent the combined use of these quantitative tools.

Despite the aspiration inherent in the title *BIO2010*, we still have much work to do to create the type of curriculum in the quantitative sciences that suits the demands of the next generation of life science researchers. This effort toward a calculus-based statistics course as a small contribution among many other efforts seems to be bringing us closer to that goal.

ACKNOWLEDGMENTS

I thank the Howard Hughes Medical Institute (HHMI) Biomath Committee at the University of Arizona for many thought-provoking discussions. Student work and their feedback were essential in setting and refining the pedagogical approaches in the course. A special note of thanks goes to Christopher Bergevin, who was co-instructor the first time the course was taught, and to Carol Bender, the Director of the Undergraduate Biology Research Program, whose efforts have provided many undergraduate students with the opportunity to experience an authentic research experience. The activities described in this article were supported in part by a grant to the University of Arizona from the HHMI (52005889).

REFERENCES

Agresti, A, and Franklin, C. (2008). Statistics: The Art and Science of Learning from Data, Upper Saddle River, NJ: Pearson Prentice Hall.

Bevington, P. R., and Robinson, D. K. (2002). Data Reduction and Error Analysis for the Physical Sciences, New York: McGraw-Hill.

Diamond, J, and Bellwood, P. (2003). Farmers and their language: the first expansion. Science *300*, 597–603.

Efron, B, and Tibshirani, R. J. (1994). An Introduction to the Bootstrap, New York and London: Chapman and Hall.

Hall, M. R., and Rowell, G. (2008). Undergraduate statistics education and the National Science Foundation. J. Statistics Ed. *16*, http://www.amstat.org/publications/jse/v16n2/rowell1.html

Hoel, P. G., Port, S. C., and Stone, C. J. (1972). Introduction to Statistical Theory, Boston: Brooks Cole.

Hogg, R. V., and Tanis, E. (2009). Probability and Statistical Inference, Upper Saddle River, NJ: Prentice Hall.

Lansing, J. S., *et al*. (2007). Coevolution of languages and genes on the island of Sumba, eastern Indonesia. Proc. Natl. Acad. Sci. USA *104*, 16022–16026.

Meyer, S. L. (1975). Data Analysis for Scientists and Engineers, New York: Wiley.

Moore, D. S., McCabe, G. P., and Craig, B. (2007). Introduction to the Practice of Statistics, New York: W. H. Freeman.

National Research Council. (2003). BIO2010, Transforming Undergraduate Education for Future Research Biologists. Washington, DC: National Academies Press.

Nelson, D. L., and Cox, M. M. (2005). Principles of Biochemistry, New York: W. H. Freeman.

Nolan, D., and Lang, D. T. (2009). Approaches to broadening the statistics curricula. In: Quality Research in Literacy and Science Education: International Perspectives and Gold Standards, ed. M. C. Shelley, L. D. Yore, and B. B. Hand. The Netherlands: Springer.

Nolan, D., and Speed, T. P. (1999). Teaching statistics theory through applications. Am. Statistician *53*, 370–376.

Nolan, D., and Speed, T. P. (2000). Stat Labs: Mathematical Statistics Through Applications, New York: Springer.

Piegorsch, W. W., and Bailer, A. J. (2005). Analyzing Environmental Data, Hoboken, NJ: John Wiley & Sons.

Pitman, J. (1999). Probability, New York: Springer.

Polya, G. (1920) Uber den zentralen Grenzwertzsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. Mathematische Zeitschrift *8*, 171–181.

Powell, L. A. (2007). Approximating variance of demographic parameters using the delta method: a reference for avian biologists. Condor *109*, 950–955.

R Development Core Team (2009). R: A Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing.

Ross, S. (2009). First Course in Probability, Upper Saddle River, NJ: Prentice Hall.

Rossman, A., and Chance, B. (2004). A data-oriented, active-learning, post-calculus introduction to statistical concepts, applications, and theory. In: ASE 2004 Roundtable on Curricular Development in Statistics Education.

Siegrist, K. (2004). The probability/statistics object library. J. Online Math. Applications *4*, http://mathdl.maa.org/mathDL/4/?pa=content&sa=viewDocument&nodeId=381.

Sokal, R. R., and Rohlf, F. J. (1994). Biometry: The Principles and Practices of Statistics in Biological Research, New York: W. H. Freeman.

Walton, D. Brian (2002). Analysis of single-molecule kinesin assay data by hidden Markov model filtering, PhD Thesis, University of Arizona.

Wiehe, T, and Stephan, W. (1993). Analysis of genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. *10*, 842–854.

Zweig, M. H., and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. *39*, 705–742.