Statistical Inference

Daniel M Wolcott, University of Central Missouri, Warrensburg, MO, United States
Adam Duarte, Oregon State University, Corvallis, OR, United States
Floyd W Weckerly, Texas State University, San Marcos, TX, United States

Glossary

Bias Processes or procedures that consistently result in measurements that differ from the true or correct value.
Inductive reasoning Making a scientifically valid conclusion about a large unknown entity from only a part of the unknown entity.
Maximum likelihood estimation The omnibus estimation approach that assumes large sample size.
Precision The repeatability of a measurement taken multiple times.
Probability The science of uncertainty with values that range from 0 to 1, inclusively.
Repeated measures Observations that are repeatedly measured through time and when subjected to different treatments.
Restricted maximum likelihood estimation Using maximum likelihood estimation on small data sets where estimation relies on using some of the observed data.
Statistics The science of analyzing data.

Introduction

Simply put, statistical methods are used by ecologists to detect patterns in data. If patterns are detected, then the ecologist might also want to predict. Prediction is particularly useful for visualizing the estimated relationships.

Fundamental to the correct application of statistical methods, is an understanding of words and terms such as observation, variable, sample, random sample, population, statistics, parameters, and sample error. An observation is the basic unit. By definition an observation is independent, meaning that each observation has its own information unconnected to the information in any other observation. Replicates or experimental units are synonymous with observations. Characteristics measured from observations are called variables. It is the numerical values of variables that are actually analyzed, and it is convenient to classify variables to assist in choosing the appropriate statistical method, which is largely governed by whether a variable is categorical or numeric. Numeric variables are further classified into continuous or discrete because of different mathematical characteristics (Fig. 1). Continuous variables have infinite possible values between end points and discrete variables are restricted to integer values (e.g., 0, 1, 2, etc.). The population (usually denoted as N) is all the observations of interest and can be quite large or infinite in size. Because it is often not practical to measure variables from all observations of a population, a random subset of observations (usually denoted as n) is selected. The random subset of observations that are drawn from the population is the sample. To, hopefully, obtain a sample that is a microcosm of the population, observations should be selected without any known bias (i.e., a random sample). For example, if we wanted to know if male and female deer were different in body mass at the Kerr Wildlife Management Area, the observation would be the individual deer, the variables measured from each deer would be sex and body mass, the sample would be the collection of deer that were randomly selected and measured, and the population would be all the deer on the Kerr Wildlife Management Area. Clear definitions of these basic terms are needed as the foundation to understand statistical inference.

The essence of statistical inference is making a conclusion about the large unknown and unknowable (population) from the known (random sample). As such, inductive reasoning is required so that a valid conclusion is obtained. Integral to making statistical inferences is accommodating sample error—which is from the sampling process—and using distributions (Fig. 1). The practical consequence is that sample statistics (e.g., mean and variance) almost always will not be identical, but are hopefully close to, their corresponding parameters (population mean and variance). In other words, the data analyst strives for statistical estimates that are precise and unbiased. Using a simple example, a valid inference about a population mean can be made from a 95% confidence interval of a sample mean. The confidence interval is patently estimated to accommodate sample error, and it is a representation of the uncertainty associated with the point estimate. Notably, the interval is more likely to correctly inform about the population mean when the sample data is unbiased and sufficiently large.

Distributions

There are a number of distributions that are used in statistical inference to model uncertainty. We describe six distributions that are often used (Fig. 1). The binomial and Poisson are discrete distributions. Values of these distributions have a lower limit of zero and upper limits that can be quite large. The values are integers. The binomial distribution is one of the most widely

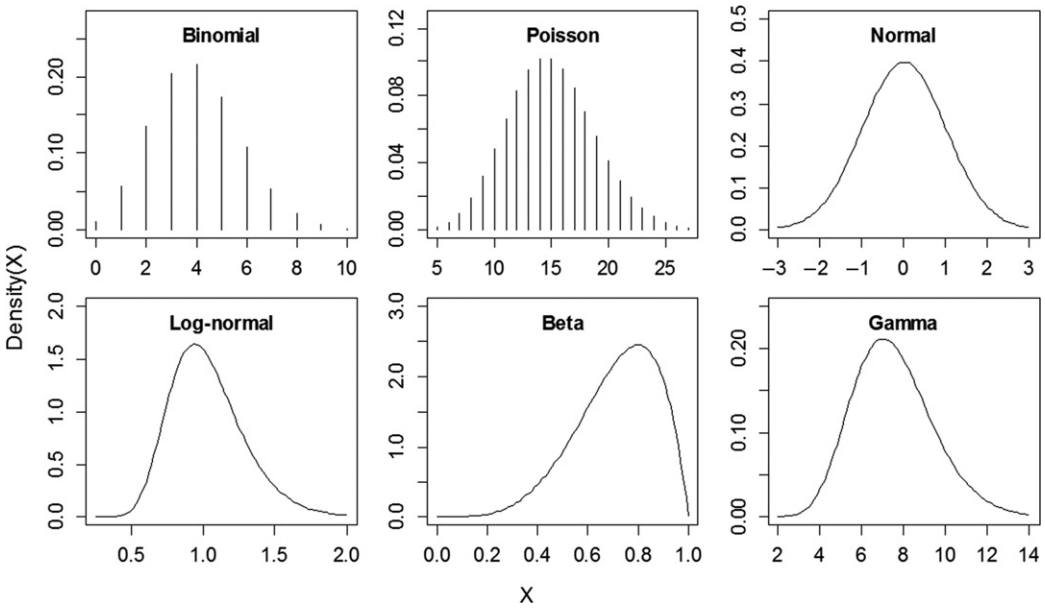

Fig. 1 Examples of six distributions used in statistical inference. Each distribution can have a variety of shapes. Two of the distributions are discrete and the remaining four are continuous. Because the discrete distributions can only have integers for *x* values, vertical lines were used instead of a smooth curve.

implemented distributions for ecological studies focused on the distribution and abundance of animal populations. The key property of the binomial are trials, which by definition are independent. Each trial has two mutually exclusive outcomes. Often, one of the outcomes is of interest and labeled a success. For example, a binomial distribution can be used to describe whether an individual female deer (the trial) gave birth to one or more young in a year (the success) or not. With knowledge about the probability of success it is possible to use the binomial distribution to predict the probability of success for a given set of scenarios. The Poisson distribution is sometimes labeled the distribution of rare events. Imagine you are counting fish at a weir and it is likely that there will be thousands of fish that pass through the weir in a day. One of your tasks is to note if the observed fish are tagged, but there might be three tagged fish counted in a day. For the binomial distribution, the probability of success would be very small and not a convenient application because of the extra variation in the data from the few successes relative to the number of trials. A convenient alternative is the Poisson distribution. The defining feature of the Poisson distribution is the equality between the mean and variance. This distribution is frequently used for modeling population abundances because there usually is small variance among counts of a small population and large variance among counts of an abundant population.

The remaining four distributions are continuous distributions. The normal or Gaussian distribution is a widely used distribution for at least three main reasons. One reason has to do with symmetry. If you superimpose a vertical line through the mean or the peak of the distribution, the right half is a mirror image of the left half. Second, values of the distribution can be negative or positive. Third, the bell-shape of the distribution is useful for describing the sampling process and it matches the distribution of numerous ecological patterns. The normal distribution can be described from knowing the mean and standard deviation. The remaining three continuous distributions are useful in particular settings. The lognormal distribution is a right skewed distribution that becomes more symmetrical with larger means. The lognormal is useful when populations grow exponentially or when the mean is close to zero but the values are constrained to be positive. The beta distribution ranges from 0 to 1 and has two shape parameters that are used to calculate mean and variance. Because the beta distribution ranges from 0 to 1, it is suited for modeling probabilities or proportions. Last but not least is the gamma distribution. The gamma distribution ranges from 0 to values that are quite large. It is a distribution of waiting times until a certain number of events occur. For example, the length of time it takes for three deer to die. The mean and variance are calculated from a shape parameter and a length per event or scale parameter.

Regression Models

General linear models are popular statistical methods (Bolker, 2008). Three conditions (assumptions) define a general linear model, which should be met in the response variable: independence, normality and homoscedasticity (similar variances). It is important to note that the assumption of normality is often misunderstood as a requirement for data to be distributed normally. Instead, normality refers to an error distribution (assessed from residuals, value of variable minus mean of variable) (see Kéry and Hatfield, 2003).

Table 1 Findings of the same general linear model expressed in two ways, as a two-sample *t*-test and as a simple linear regression

Two-sample t-*test*							*Simple linear regression*				
Sex	*n*	*Mean*	*Variance*	*t*	*df*	*P*	*Coefficient*	*Estimate*	*SE*	*t*	*P*
Female	51	2.64	0.39	**2.51**	122	**0.007**[a]	Intercept	2.64	0.09	28.78	
Male	73	2.94	0.46				Sex	0.30	0.12	**2.51**	**0.007**

[a]The *P* can be interpreted as, if you decide to conclude that males are indeed larger than females, there is a 0.007 probability that you are incorrect. Because of the small probability, the conclusion is that males are heavier than females.

The data set is birth mass (kg) of female and male white-tailed deer born to 2 year old mothers fed a pelleted-diet rich in nutrients. For the simple linear regression sex was coded 0 for females and 1 for males. Note the *t* and *P* values (in bold font) of the two-sample *t*-test and sex coefficient of the regression are identical. Because I expected males to be heavier than female a one-tailed alternative hypothesis was used instead of the default two-tailed alternative.

Examples of general linear models include *t*-tests, analysis of variance (ANOVA), simple and multiple regression. Simple regression has one predictor variable (sometimes referred to as an explanatory variable or independent variable) and multiple regression has two or more predictor variables. The choice of a general linear model depends on the number of groups or samples and the number of variables measured from observations. It might also be supposed that method of choice is dictated by whether you need to test for differences among means or estimate relationships. Yet, these boundaries that seemingly determine the kind of method to use are actually blurred (Draper and Smith, 1998). Realizing how the same data set can be analyzed by more than one general linear model is useful to grasping that categorical as well as numeric variables can be analyzed in a regression analysis. To illustrate, we compare findings from a two-sample *t*-test and a simple linear regression. The data was birth mass of female and male white-tailed deer (*Odocoileus virginianus*) born to 2-year-old mothers on a high nutritional plane. Mothers only gave birth to a single young (Wolcott *et al.*, 2015). Our goal was to test whether male birth mass was greater than female birth mass. In size-dimorphic ungulates, adult males are larger than adult females and the difference can manifest at birth (Wolcott *et al.*, 2015). Findings from the two-sample *t*-test indicated that males, on average, were heavier than females (**Table 1**). The *t* value was 2.51 and with 122 degrees of freedom the *P* was 0.007. A *P* is the probability of a type I error, which occurs when you conclude that there are differences when there actually are no differences. The critical or α value for a *P* (the cutoff in which you decide there is a difference or not) that is most often used in ecology is 0.05; however, 0.1 and 0.01 are also used. The other kind of error is a type II error, which is failing to detect differences that truly exist, a possibility that seems remote in this example due to the relatively large sample size, substantial size dimorphism and low variances between the sexes. To conduct a simple linear regression on this data, the categorical variable sex was coded as a dummy variable (Draper and Smith, 1998). Here, we arbitrarily assigned 0 to females and 1 to males. In the parlance of regression, the predictor was sex and the response variable was birth mass. Because females were coded 0, the intercept of the regression is the estimate of mean birth mass of females (**Table 1**). The estimated slope measures the difference in means between male and female birth masses because males were coded 1. On average, male birth mass was 0.30 kg heavier than females. If 0.30 is added to 2.64, the intercept, you arrive at the mean birth mass of males (2.94 kg). Thus, sometimes the distinction between two-sample *t*-test and simple linear regression is in name only.

A large part of the popularity with general linear models is the versatility to summarize complex relationships. When the response variable is continuous and there are multiple categorical, numeric, or both kinds of predictors; multiple regression is appropriate. To illustrate, we estimated relationships between body mass, age and liver mass of male white-tailed deer during the mating season (Parra *et al.*, 2014). Deer age was either young (1.5–2.5 years-old, $n = 26$) or prime-aged (4.5–6.5, 13). Body mass was continuous and age was categorical. During the mating season, prime-aged, but not young, males engage in physically demanding activities to attain copulations. The consequence is that energetic demands usually exceed energy intake in prime-aged males. To meet the energy demands, prime-aged males mobilize adipose tissue, which is a function of the liver that leads to a heavier liver mass because the workload of the liver increases (Parra *et al.*, 2014). There is a complication, however, in that prime-aged male deer are larger than young deer and also require larger livers simply due to their larger body size. Thus, the need to include body mass as a predictor to control for this effect. The estimated multiple regression revealed that body mass was positively related to liver mass and that prime-aged males had heavier livers (**Fig. 2**). Accounting for body mass, prime-aged males had livers that were, on average, 0.21 kg heavier than young males. The R^2, adjusted for number of predictors, (0.75) indicated that about 75% of the variation in liver mass was accounted for by the regression. In addition to versatility, general linear models are appealing because they can be used to predict and scatterplots help to visually display complex relationships in data.

Hierarchical Models

In many ecological studies, efficient sampling design often entails repeated or clustered measurements of subjects or observations. These types of sampling designs inherently create hierarchical structure within the data, which are often correlated spatially and temporally. Correlations between measurements violate one of the foundational assumptions that each observation is

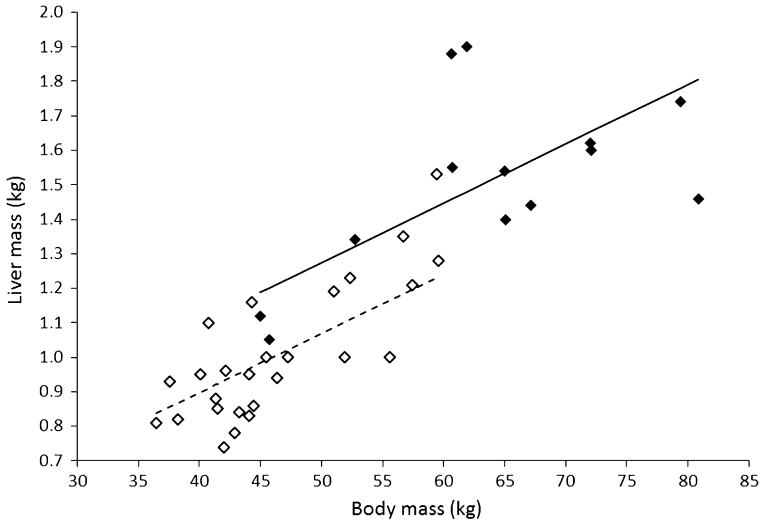

Fig. 2 Scatterplot of data and relationships between body (minus rumen-reticulum and liver organ masses) and liver masses of young (1.5–2.5 years of age, open diamonds, dashed line) and prime-aged (4.5–6.5, filled diamonds, solid line) male, white-tailed deer. The general linear model or multiple regression was: Liver mass $= 0.21 + 0.02$*body mass $+ 0.21$*age. Body mass and age were influential ($t \geq 2.7$, df $= 36$, $P \leq 0.011$). Young males were coded 0 and prime-aged males were coded 1. The adjusted R^2 was 0.75.

independent and identically distributed. Repeated measures analyses have been devised to account for the issue of dependence among observations, however, many of these analyses are not flexible about missing data (i.e., unequal sample size among levels) or uneven temporal spacing of repeated measurements. Fortunately, mixed-effect models account for these issues and are also more easily interpretable than most classical repeated measures analyses.

Mixed-effect modeling is a powerful tool that uses both fixed and random variables within the same model. Whether a variable should be considered as fixed or random is largely dependent on the question of interest and the number of levels within the variable. If the number of levels for the variable is representative of the population and the question of interest involves finding differences between levels, utilizing the variable as a fixed component is the recommended approach. If the levels of the variable are a randomly selected subset of all available levels and the question of interest involves understanding the variation within the population, treating the variable as a random variable is the recommended approach. Understanding the desired objective of the research question is of great importance when considering whether a variable should be treated as fixed or random. Further, while variables treated as fixed can be of several types (e.g., continuous, categorical, etc.) random variables must be categorical with a large enough number of levels to achieve an accurate representation of the variance in the population (usually >4 but preferably >10 levels, Zuur *et al.*, 2007).

Mixed-effect models demonstrate tremendous flexibility in how the random components can be modeled. For example, consider a study in which 10 individuals were repeatedly sampled across time to measure the influence of a continuous predictor (X) on the response variable (Y). We could simply regress X against Y at a cost of three parameters (one intercept, one slope of X, and residual error); however, we know that repeatedly sampling 10 individuals is most likely creating correlations within the residual error because individuals are more likely to be similar to themselves than the other nine individuals. We could account for this, in a fixed effect only framework, by incorporating the individual identification as a categorical covariate. Inclusion of this covariate would create a regression line with a different intercept but same slope for each one of the sampled individuals, which increases the number of parameters drastically to 12 (10 unique intercepts, 1 slope for all sampled individuals, and residual error). The inclusion of an individual-specific variable quickly inflates the number of parameters, which penalizes the ability to detect differences through the loss of degrees of freedom. Instead, we could include the individuals as a random effect when regressing X against Y. This essentially produces the same outcome as the above fixed effect only model except it only costs us four parameters (one intercept, one random effect to estimate variation of individuals around the intercept, one slope, and residual error). With this mixed-effect model we are able to reduce the residual error by accounting for the variation created by sampling the 10 individuals. Mixed-effect models are even more powerful than this simple problem. Random effects can be modeled to assess variation in intercepts only (like the above example), slopes only, intercepts and slopes, and even residual error itself. We strongly encourage the reader to continue to explore the many possible ways that mixed-effect modeling can be used to partition variance structures (see Pinheiro and Bates, 2000, and Zuur *et al.*, 2007 for more in-depth examples).

Invariably, mixed-effect model development will require some form of selection process to assess which fixed variables and random components are necessary to accurately describe reality, within a modeling framework, without overfitting (i.e., include too many predictors and random components) the data. The generally accepted approach consists of a top-down strategy beginning with a saturated model that includes all fixed factors of interest as well as any interactions between fixed factors that are thought to be biologically relevant (Zuur *et al.*, 2007). The next step is to assess which random error structure is most optimal (e.g.,

random intercept, slopes, or intercept and slopes, etc.) using restricted maximum likelihood estimation. Once the optimal random error structure is defined for the saturated model, assess which fixed variables should be kept to form an optimal model using maximum likelihood estimation. Finally, conclusions from the optimal model, using the estimated parameters and standard errors, should be presented and discussed using restricted maximum likelihood estimation to avoid biased variance estimates.

When comparing models with the same fixed effects and different random effects through hypothesis testing, it is important to note that significance values are only approximations, which should be assessed with caution if the value is on the cusp of the selected α level. This approximation is a consequence of two-tailed hypothesis testing during model comparison. Since variances, which are centered on zero, can only be positive values, a two-tailed hypothesis test inaccurately assesses the range of variance values yielding a significance value that is generally inflated higher than expected. There is limited concern for significance values that are very large or very small as compared to the α level; however, the concept of "testing on the boundary" should be considered when the significance value nears the value for α.

Bayesian Inference

Up until now we have described the process of statistical inference using what is known as a classical or frequentist paradigm. However, the use of Bayesian inference within ecological studies has increased in the past decade, so a chapter focused on statistical inference would be inadequate without some mention of it. For an in-depth description of the background information and application of Bayesian inference, we recommend readers consult Link and Barker (2010) and Hobbs and Hooten (2015). It should be stated that there is no such thing as a "Bayesian model" despite the term used within the peer-reviewed literature. As Kéry and Schaub (2012) discuss, ecologists develop models to represent hypotheses they want to explore in data, and they can choose to fit the model(s) to the data in either a Bayesian or frequentist framework. Indeed, the regression models described above as well as many complex hierarchical models can be fit under both paradigms (Royle and Dorazio, 2008; Kéry and Royle, 2016). Furthermore, when prior distributions are specified as diffuse or noninformative in Bayesian statistics there is a great deal of agreement in the output estimates from the two paradigms (Kéry, 2010).

So what is the difference in the two paradigms and why would you choose to fit models using one paradigm versus the other? In a frequentist paradigm, parameters are treated as fixed and unknown constants, and inferences are based on the notion of repeated experiments or the frequency of hypothetical replicates. Here, ecologists make probability statements about the data and the probability is based on the hypothetical replicates. Thus, a confidence interval is interpreted as "If I performed this experiment an infinite number of times, X% of the parameters would fall within this range." In contrast, under the Bayesian paradigm all parameters are treated as random variables and the unknown quantities are modeled using the statistical distributions introduced above, among others. Here, our knowledge of parameters is updated using data, rules of conditional probability, and Bayes' theorem. In other words, prior (knowledge) distributions and data are used to obtain or yield posterior distributions to make inferences. This is usually accomplished with the aid of simulation methods, such as Markov chain Monte Carlo (MCMC) algorithms, to characterize the posterior distribution of parameters. Because parameters are treated as random variables, under the Bayesian paradigm credible intervals, which are the Bayesian analog of confidence intervals in frequentist statistics, are interpretable as "I am X% confident the parameter is within this range." Although appealing from an interpretability standpoint, in practice this difference is of little consequence as confidence intervals and credible intervals are typically applied in the same manner.

As we mentioned earlier, Bayesian inference can accommodate previous knowledge from other studies via informative priors so that inferences are based on both the current and previous studies. Also, in Bayesian inference all results are exact even for small datasets because Bayesian analyses do not rely on asymptotic approximations that are used in frequentist analyses (Conroy and Peterson, 2013). This is particularly appealing because data are often limited in ecological studies. The BUGS language that is often used to fit models using a Bayesian analysis, offers a framework for better understanding the structure of statistical models and by extension a more transparent structure for accounting for sources of variation in the data. Once ecologists have gained the skills to develop simpler models in BUGS, more complicated models are a relatively simple extension that use the same rulesets. Thus, more complicated models can be fit using a Bayesian framework relatively easy, including models in which there is no frequentist method available. This has led to the recent increase in the use of integrated analysis of multiple but related datasets within the ecological literature. These integrated analyses represent an important advancement in the ecological sciences because they allow for the development of tailored analyses for specific datasets, which, in turn, allow for the incorporation and estimation of complex ecological relationships. Similarly, derived parameters (i.e., estimates that are calculated from other estimated parameters) that propagate the full uncertainty of estimates can be calculated in a straightforward manner using BUGS. In the frequentist paradigm, deriving fitted models requires using the delta method or the omission of parameter uncertainty.

It should be noted that there are also drawbacks to fitting models in a Bayesian framework. First, the BUGS programming language can admittedly be intimidating when getting started, particularly if you are not well acquainted with statistical distributions. Nonetheless, we stress that if fitting more complex models is a necessary objective, the time investment to overcome the initial learning curve is well worth it. Also, it should be noted that the number of iterations required during the MCMC simulations to achieve convergence when fitting more complex models to large data sets can take hours to weeks, whereas the same model using a frequentist analysis may be much quicker. We view the frequentist and Bayesian paradigms as useful tools for statistical inference and stress that ecologists do not have to choose one or the other. Instead, we recommend ecologist consider the pros and cons of each approach as they relate to the objectives of a particular study and choose the best tool for the task.

Model Selection

To aid in statistical inference, models are developed to mimic the underlying distribution of a population using empirical data. In an ecological context, most studies are considered to be observational with limited ability to set up control–treatment experiments that include proper randomization and replication. Since it is rarely possible to create a model that describes a population with complete accuracy, a suite of variables that are hypothesized to be responsible for some portion of the observed phenomenon are measured and multiple competing models are developed. Because of this, it is necessary to choose a model that best approximates the underlying distribution of the sampled population, without overfitting the available data, through a model selection process.

The coefficient of determination (R^2), as an example and demonstrated above, is a useful tool to assess how well a model replicates observed data. Although a useful tool to assess how well a single model explains the observed data, it is not useful for comparison between competing models. The R^2 value will almost always increase with each additional inclusion of a predictor variable, which leads to the selection of a complicated model that probably over fits the observed data. To account for this issue, various stepwise procedures have been developed to assess how strongly each variable influences the predictive power of a model using hypothesis tests.

Hypothesis testing procedures are popular techniques for model selection and include forward selection, backward elimination, or stepwise processes. The forward selection process begins with the most simplified (null) model and compares the null model to another model that includes a single added variable. These two models are then compared using an F-test. If the F statistic is larger than the set threshold, the model with the added variable has a stronger predictive power than the null model and is considered the null model thereafter. The procedure repeats itself by including additional variables and comparing the newest model to the previously selected model until a candidate variable no longer increases the F statistic above the accepted threshold. The backward elimination process is similar except that it begins with a global model that contains all predictor variables. This model is tested against another model that is reduced by a single variable with a subsequent F-test assessing if the variable should be included or not. Stepwise procedures were developed as an extension of forward and backward selection processes to remedy flexibility issues after a variable is considered in the analysis. In the stepwise process, all variables are considered, at each step, for inclusion or omission from the model even if they were previously included or excluded in a previous model. While these procedures are widely used it is important to note that they are not without flaws. Specifically, there is no clear criterion for model choice and the final model choice is usually a good model but not always the optimal model. An optimal model does not under- or over fit the data. Indeed, using forward and backward processes on the same dataset can often lead to the selection of very different models.

An information-theoretic approach to model selection seeks to find the optimal model that bests describes the relationship between the response variable and a set of predictors. Unlike hypothesis testing, it uses a criterion, involving the log likelihood (a maximum likelihood measure of model fit), to assess the distance (i.e., information lost) between a model and the observed data. Instead of focusing on the significance of individual predictor variables within a model, information-theoretic model selection assesses the likelihood of a model as a whole. The intent of multimodel inference in an information-theoretic framework is to be a confirmatory analysis rather than an exploratory analysis so the process of data dredging (assessing all possible models) is strongly discouraged (Burnham and Anderson, 2003). Instead, the goal is to develop models with sound biological reasoning for comparison through multimodel inference to determine the best approximating model given the data through strength of evidence. Yet, information-theoretic approaches have been applied to traditional hypothesis testing procedures such as the forward selection, backward elimination and stepwise processes because of simplicity and flexibility.

One of the most common forms of information-theoretic model selection in ecology, from the frequentist perspective, is the Akaike's Information Criterion (AIC). The criterion is simply defined and calculated as "deviance plus $2K$," where deviance is a measure of the information loss between the model and the observed data (-2 times the log likelihood) and $2K$ is a penalty for increasing model complexity (K is the number of parameters estimated in the model). A second-order bias correction (AICc) has also been developed for use when sample size is low (generally $n/K < 40$). Since AIC and AICc values converge as sample size increases, the use of AICc in all cases is common. Multimodel inference from AIC is very simple because the model with the lowest AIC value is the best approximating model. For competing models (models with <2 AIC value difference from the best model), it is possible to model average parameter estimates, however, parsimony should also be considered before model averaging competing models. Burnham and Anderson (2003) demonstrate the process, flexibility, and advantages to AIC model selection, which includes valuable resources necessary to calculate evidence ratios between models and the assessment of relative variable importance within the selected model. We advise reading this resource for further information.

An information-theoretic approach to model selection is also possible from the Bayesian perspective. The most commonly used criterion for Bayesian inference is the Deviance Information Criterion (DIC) because it is effective for models with informative prior information and is readily accessible as a function in the BUGS software. Although DIC has been widely used over the past decade, it has come under increasing criticism for several reasons (e.g., model selection is not always consistent and tends to produce over-fitted models). A relatively new addition to the expanding field of information criterion is the Watanabe–Akaike information criterion or widely applicable information criterion (WAIC). It has been shown to be asymptotically equal to Bayesian cross-validation procedures and is the most fully Bayesian process because it uses the full posterior distribution (rather than using a point estimate like AIC and DIC). WAIC is still not without limitations and has some difficulties with spatially structured data. For Bayesian inference, cross-validation is still the recommended method for model validation with WAIC used as a more computationally expedient option (Gelman *et al.*, 2014).

See also: Ecological Data Analysis and Modelling: Ecological Models: Model Development and Analysis; Model Types: Overview; Sensitivity, Calibration, Validation, Verification; Structural Dynamic Models. General Ecology: Principal Components Analysis

References

Bolker, B.M., 2008. Ecological models and data in R. Princeton, USA: Princeton University Press.
Burnham, K.P., Anderson, D., 2003. Model selection and multi-model inference: A practical information-theoretic approach, 3rd ed. New York, USA: Springer Science & Business Media.
Conroy, M.J., Peterson, J.T., 2013. Decision making in natural resource management: A structured, adaptive approach. Hoboken, USA: Wiley-Blackwell.
Draper, N.R., Smith, H., 1998. Applied regression models. New York, USA: John Wiley & Sons.
Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing 24, 997–1016.
Hobbs, N.T., Hooten, M.B., 2015. Bayesian models: A statistical primer for ecologists. Princeton, USA: Princeton University Press.
Kéry, M., 2010. Introduction to WinBUGS for ecologists: A Bayesian approach to regression, ANOVA, mixed models and related analyses. New York, USA: Academic Press.
Kéry, M., Hatfield, J.S., 2003. Normality of raw data in general linear models: The most widespread myth in statistics. The Bulletin of the Ecological Society of America 84, 92–94.
Kéry, M., Royle, J.A., 2016. Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS, volume 1: Prelude and static models. San Diego, USA: Academic Press.
Kéry, M., Schaub, M., 2012. Bayesian population analysis using WinBUGS: A hierarchical perspective. New York, USA: Academic Press.
Link, W.A., Barker, R.J., 2010. Bayesian inference with ecological applications. New York, USA: Academic Press.
Parra, C.A., Duarte, A., Luna, R.S., Wolcott, D.M., Weckerly, F.W., 2014. Body mass, age, and reproductive influences on liver mass of white-tailed deer (*Odocoileus virginianus*). Canadian Journal of Zoology 92, 273–278.
Pinheiro, J.C., Bates, D.M., 2000. Mixed-effects models in S and S-Plus. New York, USA: Springer Verlag.
Royle, J.A., Dorazio, R.M., 2008. Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities. New York, USA: Academic Press.
Wolcott, D.M., Reitz, R.L., Weckerly, F.W., 2015. Biological and environmental influences on parturition date and birth mass of a seasonal breeder. PLoS One 10.e0124431
Zuur, A.F., Ieno, E.N., Smith, G.M., 2007. Analyzing ecological data. New York, USA: Springer Science & Business Media.