Combining statistical inference and decisions in ecology

PERRY J. WILLIAMS[1, 2, 4] AND MEVIN B. HOOTEN[1, 3]

[1]*Department of Statistics, 102 Statistics Building, Colorado State University, Fort Collins, Colorado 80523 USA*
[2]*Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology,*
201 J.V.K. Wagar Building, 1484 Campus Delivery, Colorado State University, Fort Collins, Colorado 80523 USA
[3]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and*
Conservation Biology, 201 J.V.K. Wagar Building, 1484 Campus Delivery, Colorado State University, Fort Collins,
Colorado 80523 USA

Abstract. Statistical decision theory (SDT) is a sub-field of decision theory that formally incorporates statistical investigation into a decision-theoretic framework to account for uncertainties in a decision problem. SDT provides a unifying analysis of three types of information: statistical results from a data set, knowledge of the consequences of potential choices (i.e., loss), and prior beliefs about a system. SDT links the theoretical development of a large body of statistical methods, including point estimation, hypothesis testing, and confidence interval estimation. The theory and application of SDT have mainly been developed and published in the fields of mathematics, statistics, operations research, and other decision sciences, but have had limited exposure in ecology. Thus, we provide an introduction to SDT for ecologists and describe its utility for linking the conventionally separate tasks of statistical investigation and decision making in a single framework. We describe the basic framework of both Bayesian and frequentist SDT, its traditional use in statistics, and discuss its application to decision problems that occur in ecology. We demonstrate SDT with two types of decisions: Bayesian point estimation and an applied management problem of selecting a prescribed fire rotation for managing a grassland bird species. Central to SDT, and decision theory in general, are loss functions. Thus, we also provide basic guidance and references for constructing loss functions for an SDT problem.

Key words: Bayesian risk; Bayes rule; frequentist risk; loss function; optimal posterior estimator; statistical decision theory.

INTRODUCTION

Understanding ecological complexity based on empirical data has led to the proliferation of advanced statistical methods for ecological inference. Coinciding with these developments is a need for formal, rigorous methods for using these analyses to make decisions. Subsequent to a statistical investigation, an investigator has several choices to make. Consider two scenarios: (1) an investigator deciding how to summarize the results of an analysis to report in a scientific journal or technical report, and (2) a manager deciding how the results of a statistical analysis translate into a choice of management action. In each example, the decision maker is trying to achieve a (perhaps implicit) objective. Minimizing the amount of information lost from the data to the reported statistic is a possible objective in the first scenario, and maximizing some resource is a possible objective in the second scenario. Statistical theory alone does not provide guidance for these choices. In the first scenario, suppose the investigator has collected survival data on a critically endangered species and estimated a posterior

distribution for the probability of survival. Standard practice suggests the posterior mean or median is a conventional statistic to report. Is this choice of estimator arbitrary, or is it explicitly linked to an objective? In the second scenario, suppose a refuge manager has collected data on abundance of a species of concern and its relationship to prescribed fire frequency. The manager would like to use the information to maximize cumulative abundance of the species through time. We consider statistical decision theory (SDT) as a single framework to address both of these questions, and more generally, to formally link decisions to three sources of information: statistical results from a data set, knowledge of the consequences of potential choices (i.e., loss), and prior beliefs about a system (Fig. 1).

Decision theory is broadly defined as the theory of objective-oriented behavior in the presence of choices. SDT is a sub-field of decision theory concerned with using the results of a statistical investigation to reduce uncertainty of a decision problem with the ultimate goal of helping a decision maker choose the best available action under a specified objective (Berger 1985). The theory and application of SDT originated from a shift in the field of statistics in which statistical inference was regarded as a branch of decision theory; the focus of inference was the decision to be made (Neyman , and Pearson 1928, 1933,
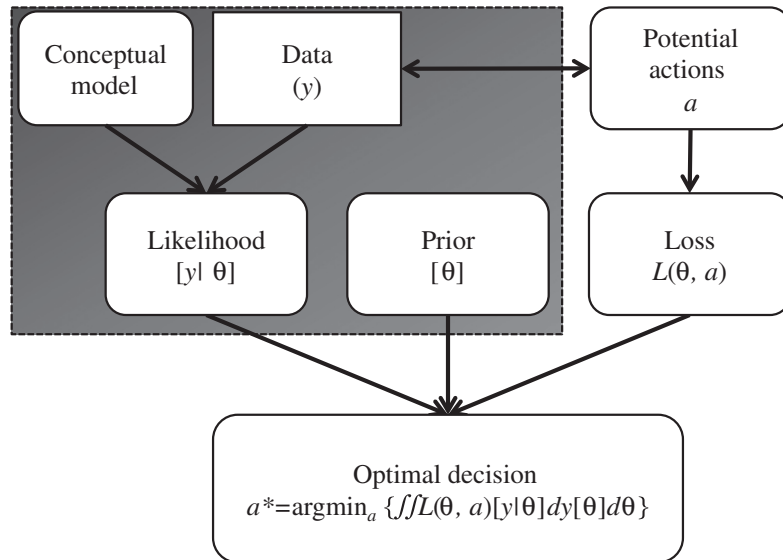
Fig. 1. Schematic of statistical decision theory (SDT). Conventional statistical inference (shaded region) is often performed without explicitly considering decisions or associated loss, whereas SDT formally incorporates decisions and loss into the framework. Equation 4 demonstrates the combination of the likelihood, the prior, and the loss.

Ramsey 1931, Wald 1950, Savage 1954, Ferguson 1967, Lindley 1971). On the importance of this movement, Savage (1962:161) said, "decision theory is the best and most stimulating, if not the only, systematic model of statistics." Although there were critics of the decision-theoretic model of statistical inference (e.g., Fisher 1955, Cox 1958, Tukey 1960, Birnbaum 1977), the shift had profound impacts on the field of statistics. Two impacts of particular relevance for ecological decisions were the revitalization of inverse probability from Bayes and Laplace (Ramsey 1931, De Finetti 1937, Savage 1954, DeGroot 1970, Lindley 1971) and the development of methods to combine data with utility theory to inform decisions (Wald 1950, Lindley 1953, Barnard 1954, DeGroot 1970, Akaike 1973). Bayesian probability re-emerged as a viable paradigm, in part, because of its compatibility with decision theory (Savage 1954, 1962). Decision theory fits naturally within Bayesian inference, and SDT and Bayesian inference are often presented in the same volume to be studied concurrently (e.g., Berger 1985, Pratt et al. 1995). We focus mainly on Bayesian SDT due to the added flexibility and its coherence with decision analysis, with the exception of basic definitions of frequentist SDT and its relationship to Bayesian SDT. We discuss the relationship between frequentist and Bayesian SDT to provide basic reference information and to highlight similarities and differences in paradigms. Bayesian methods have proven to be important tools for ecological inference. The additional concept of loss in an analysis allows Bayesian methods to be naturally extended to assist ecological decision making (Dorazio and Johnson 2003).

We summarize SDT and demonstrate its application for ecologists who must base decisions on data, in the presence of uncertainty, and prior information. We also emphasize the concept of loss and its implications for ecological science, statistical inference, and decision making in general. We focus on two scales of problems that are relevant for ecology: optimal point estimation and optimal natural resource management. Optimal point estimation has been covered in many statistical texts (e.g., Berger 1985, Casella and Berger 2002, Lehmann and Casella 1998), but has been underrepresented in the ecological literature. Optimal point estimation is important not only for choosing estimates to report in scientific journals, but for natural resource management, directly. Many natural resource decisions are based only on point estimates and do not consider estimates of uncertainty (e.g., Pacific Flyway Council 1999). In these cases, the choice of point estimator affects which management actions are ultimately implemented. Dorazio and Johnson (2003) discussed SDT as a framework for decision making in natural resource management and provide an example of using SDT for waterfowl management. Our objective was to provide a general overview of the concepts of SDT, their applicability to point estimation, and how the traditional use of SDT for point estimation can be extended to address problems in natural resource management. We summarize SDT for point estimation and provide an example of a natural resource management problem that involves selecting a prescribed fire burn rotation for Henslow's Sparrows (*Ammodramus henslowii*).

BASIC ELEMENTS OF SDT

Following Berger (1985), we begin with an introduction to the basic elements of SDT. The premise of

TABLE 1.   Notation and description of components in statistical decision theory (SDT)

Notation	Description
θ	The unknown, true state of nature
Θ	The support of θ (i.e., the potential values that θ could be)
$\mathcal{Y}$	The sample space or support of the data
Y	A random variable of interest to the decision maker that depends on θ (i.e., a function from the sample space $\mathcal{Y}$ into the real numbers)
y	Data; the data are realizations of the random variable Y from a scientific investigation carried out to provide information to the decision maker about the value of θ to assist in the decision to be made
a	An action the decision maker can take
$\mathcal{A}$	The set of potential actions or action set from which a decision maker can choose
$\delta(y)$	A decision rule that is a function of the observed data y. A decision rule is a map from the observed data to the action a. For example, if a scientific investigation is performed to inform the decision and $Y=y$ is the realization of the sample information, then the resulting decision to be made is $\delta(y)=a$. Decision rules are a frequentist concept
$L(\theta,\delta(y))$ or $L(\theta,a)$	The loss function; a function determined by the decision maker that describes the loss if action a is taken (or decision rule $\delta(y)$ is used) and θ is the true state of nature. The loss function is defined for all $(\theta,a)\in\Theta\times\mathcal{A}$. The loss function is analogous to a utility function (i.e., loss = -utility) or objective function in decision theory
$[y\|\theta]$	The probability density or mass of the data y given the true state of nature (or parameter) θ
$[\theta]$	The prior distribution of θ
$[\theta\|y]$	The posterior distribution of θ given the data y
$R(\theta,\delta)$	Frequentist risk; a function of a decision rule $(\delta(y))$ and θ that describes the expected loss to the decision maker if (s)he used $\delta(Y)$ a large number of times, for varying realizations of $Y=y$, and for any possible value of $\theta\in\Theta$
$\rho(a)$	Bayesian expected loss; the expected loss of each action a, given the loss function and either a prior distribution (a problem with no data) or a posterior distribution
$r(a)$	Bayesian risk; the Frequentist risk averaged over a prior distribution

SDT is that a decision maker chooses from a set of potential actions. The quality of the choice depends on an unknown, true state of nature; certain choices will be better for the decision maker under different potential states of nature. We assume the uncertainty in the state of nature is epistemic and reducible through scientific investigation. The exact nature of the uncertainty is problem specific. We focus on the uncertainty inherent in our ability to characterize the true state of nature in our examples (e.g., uncertainty in the value of a parameter), but SDT is sufficiently general to include other types of uncertainty (e.g., structural uncertainty of a process related to how a system will respond to decisions). Concise treatment of SDT requires defining

notation associated with its basic elements (Table 1). The true and unknown state of nature is represented by θ and all possible states of nature are represented by Θ. The list of potential actions a decision maker can choose is represented by $\mathcal{A}$, in which each potential action is represented by a. For example, suppose a natural resource manager must decide on a prescribed fire regime for a refuge containing eight grasslands. The manager must decide on a fire-return interval of 1, 2, 3, or 4 years to be applied to all grasslands to be implemented in a management plan that will span the next 20 years. The manager would like to maximize the cumulative number of Henslow's Sparrows on the refuge during that period. The manager can select the fire-return interval (i.e., $a \in \{1,2,3,4\}$) as an action. The cumulative number of Henslow's Sparrows over 20 years (i.e., $N_{a,\theta}$) that will result given each management action is unknown and a function of model parameters θ.

The link between statistical inference and decision making occurs through the specification of a loss function, $L(\theta,a)$, not to be confused with a likelihood function, (Fig. 1). Loss functions are synonymous with objective functions or utility functions (loss functions = negative utility functions) in other fields. Loss functions map actions and states of nature to a real number that represents some (not necessarily monetary) cost associated with the action and true state of nature. Mathematically, a loss function returns a value for every combination of the potential true states of nature $\theta \in \Theta$ and any action $a\in\mathcal{A}$ before a decision is made (Fig. 2). A loss function can be associated with any decision (including functions of parameters and/or predictions). Returning to our example, suppose fire is critical for Henslow's Sparrow habitat, but implementing burns is expensive. Decreasing the burn interval increases the financial cost of Henslow's Sparrow management. Thus, a manager can express loss as a function of the burn interval a, and the unknown cumulative abundance $N_{a,\theta}$ for the upcoming 20-year management period, a function of θ.

When a decision maker conducts a statistical investigation to provide information about θ, the data gathered are represented by y. The data are assumed to be a realization of a random variable Y having a probability distribution that depends on θ (i.e., $[y\|\theta]$). All possible realizations of the data (the sample space) are represented by $\mathcal{Y}$. In our example, the refuge manager, in preparation for choosing a management strategy, has recorded the population size of Henslow's Sparrows in each of the eight grasslands for four years following a prescribed fire on each grassland. Therefore, the manager possesses information describing how annual abundance is associated with the covariate x = summers-post-burn (1, 2, 3, or 4). This information will be used to inform the cumulative 20-year abundance for any of the potential burn intervals. Finally, the Bayesian framework formally incorporates a distribution for the prior probability distribution of θ (i.e., $[\theta]$), thereby incorporating information other than the sample information and a loss function (e.g.,
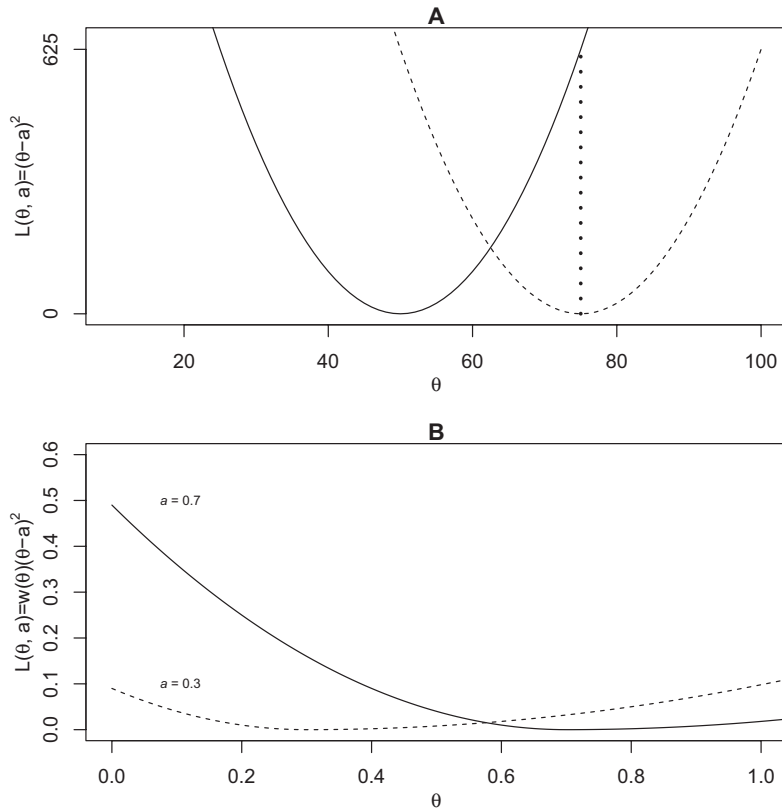
Fig. 2. (A) Squared-error loss function for the action (i.e., choice of estimator) *a*=50. If the true value of θ=75 (dotted vertical line), then $L(\theta=75, a=50)=625$. Similarly, if *a*=75, $L(75,75)=0$. (B) Asymmetric loss function with weights w(θ) for *a*=0.3 and *a*=0.7, an estimate of population growth rate for a critically endangered species. If *a* overestimates θ (i.e., θ<*a*) the loss is larger than if θ>*a*, representing preference for a conservative estimate of θ.

information from other studies on the response of Henslow's Sparrow abundance to prescribed fire; Herkert and Glass 1999). SDT proceeds by combining these sources of information to identify actions that minimize the expected loss of an action (i.e., risk).

Expected Loss (Risk)

A loss function is a function of the action of a decision maker and the unknown θ. Because θ is unknown at the time of decision, it is impossible to calculate the actual loss that will be incurred for each action. Instead, the expected loss, or risk, is calculated for each action and for each possible value of θ in the frequentist view (Fig. 3A). A decision maker proceeds by selecting the action with the smallest expected loss (in the Bayesian case) or some function of the expected loss (in the frequentist case; e.g., the minimax function). We begin our description of expected loss by describing decision rules and frequentist risk and then extend these concepts to Bayesian risk.

Frequentist risk

A decision rule δ(*y*) is a function that maps the sample space of *y* to an action *a*. That is, for any realization of

the data *y* that could occur, the decision rule prescribes the action to take. For example, a frequentist hypothesis test is an example of a decision rule. Frequentist risk is the evaluation of how much the decision maker would expect to lose if (s)he used the decision rule δ(*y*) repeatedly for different realizations of *Y*=*y*. Because θ is unknown, risk is calculated for each possible θ ∈ Θ (but see frequentist arguments against this approach in Spanos 2012). Risk is defined, for the continuous case, as the convolution of loss and likelihood, or alternatively, the expectation with respect to *y* over all samples,

$$R(\theta,\delta) = E_y(L(\theta,\delta(y))) = \int_y L(\theta,\delta(y))[y|\theta]dy, \quad (1)$$

where the term [*y*|θ] in Eq. 1 represents the probability density or mass function of *y* given the value of θ. Equation 1 is calculated for all values of θ and all decision rules δ(*y*). Because θ is unknown at the time of decision, the decision maker's choice of decision rule is equivocal (Fig. 3A). Frequentist risk provides information about the best choice of action for each θ ∈ Θ, and thus, after the risk is found, the decision maker is tasked with choosing among the decisions that are optimal for any given value of θ (i.e., admissible decision rules; Fig. 3A). A decision maker can use prior information (implicitly)
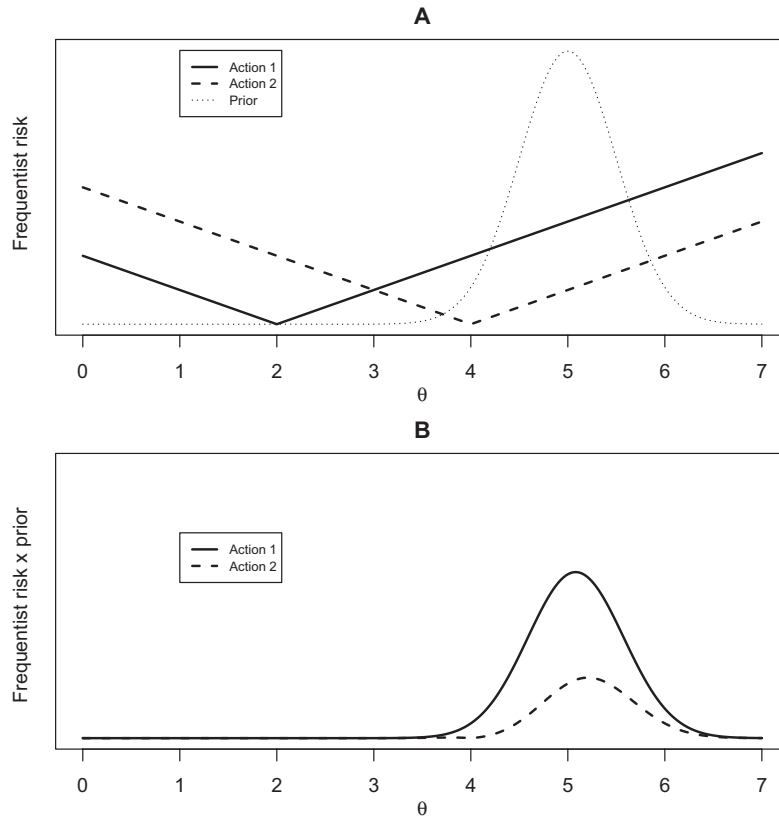
A

B

Fig. 3. (A) Frequentist risk functions for two actions. Action 1 has lower risk for values of $\theta < 3$, while action 2 has lower risk for values of $\theta > 3$. Because the value of θ is unknown, the optimal action is equivocal. Also shown is a prior distribution for θ (i.e., $[\theta]$) representing the *a priori* probability of θ. The prior can be used implicitly for frequentist risk, or explicitly for Bayesian risk. (B) Convolution of frequentist risk functions with prior distribution shown in (A) using Eq. 5. Bayes' risk for each action is found by integrating each line with respect to θ. Bayes' rule is the action with the smaller Bayes' risk. The dashed line has a smaller integral and is therefore Bayes' rule.

to identify which range of θ is most likely and choose the corresponding decision rule. Alternatively, the decision maker can consider various concepts of frequentist choice to narrow the decision space (e.g., admissibility, unbiasedness, equivariance, minimaxity).

Many frequentist and Bayesian inference procedures can be formally framed as expected loss (e.g., null hypothesis significance testing). In doing so, an investigator can use the techniques of decision theory to make a choice (Berger 1985, Ferguson 1967, Lehmann and Romano 2008).

Bayesian expected loss and Bayesian risk

Bayesian expected loss is different than frequentist risk, but mathematically related. In Bayesian expected loss, an investigator does not consider loss over hypothetical samples from the population to evaluate uncertainty in θ. Instead, an investigator assigns a probability distribution to θ. The probability distribution can be assigned without collecting new data (i.e., by using only the prior probability distribution $[\theta]$), or by combining new data, a likelihood (i.e., $[y|\theta]$), and a prior distribution

using Bayes' Theorem to calculate the posterior distribution $[\theta|y]$. Whether an investigator uses new data or not, Bayesian expected loss provides an explicit mechanism for inclusion of a prior distribution for the unknown θ; a component that is usually present, but not explicitly included in frequentist SDT. Bayesian expected loss using only prior information is defined as the average loss with respect to prior information,

$$E_\theta L(\theta, a) = \int_\Theta L(\theta, a)[\theta] d\theta. \qquad (2)$$

For the Bayesian case, we replace $\delta(y)$ with a to differentiate between Bayesian and frequentist expected loss. For a decision problem in which a scientific investigation has been conducted to collect data on the process affecting the decision, the Bayesian expected loss, or posterior expected loss is defined as

$$\rho(a) = E_{\theta|y} L(\theta, a) = \int_\Theta L(\theta, a)[\theta|y] d\theta, \qquad (3)$$

where $[\theta|y]$ relies on an assumed prior $[\theta]$ and likelihood $[y|\theta]$. A difference between Bayesian expected loss and frequentist risk is that, while frequentist risk results in a

value for risk for each possible value of $\theta \in \Theta$, Eq. 2 or 3 result in a single value of expected loss for each action and are not functions of θ after θ is integrated out. Thus, after a decision maker completes an analysis of Bayesian expected loss for each action, (s)he can select the action with the smallest expected loss, called Bayes' rule (not to be confused with Bayes' theorem). Bayes' rule is defined as the action or decision rule that minimizes the Bayesian expected loss: $a^* = \mathrm{argmin}_a(\rho(a))$. Note that the average frequentist risk over a prior distribution is

$$r(a) = \int_\Theta \int_y L(\theta,a)[y|\theta]dy[\theta]d\theta, \qquad (4)$$

and therefore, the unknown θ is integrated out of the equation. Using Fubini's theorem, and the fact that $[y|\theta][\theta]=[\theta|y][y]$, we can rearrange Eq. 4 into the form

$$r(a) = \int_y \left[\int_\theta L(\theta,a)[\theta|y]d\theta \right] [y]dy. \qquad (5)$$

The value in the large square brackets of Eq. 5 is the Bayesian expected loss, and Eq. 5 is known as Bayes risk. It can be shown that, in most non-pathological situations, minimizing the Bayesian expected loss also minimizes Eq. 5 (Berger 1985). Heuristically, Bayesian risk can be thought of as the frequentist risk averaged over the prior.

Identifying Bayes rule provides a method for selecting among actions that formally incorporates data from a statistical investigation, prior knowledge about the process, and the loss incurred for each decision as specified in the loss function (Fig. 1). This methodology can be applied to almost any situation that requires a defensible decision and allows loss to be reasonably captured in a loss function. Next, we demonstrate the application of SDT to two separate decision problems.

Point estimation using Bayesian posterior distributions

First, we discuss the class of problems dealing with finding optimal estimators for Bayesian point estimation. Suppose we collect data y to learn about an unknown parameter θ. Bayesian inference is concerned with finding the probability distribution of θ given the data (i.e., the posterior distribution $= [\theta|y]$). To summarize the posterior distribution, we often reduce the information in the posterior to a point estimate. In fact, many natural resource management decisions are based exclusively on a point estimate (e.g., Pacific Flyway Council 1999, Williams, 2016). The point estimate for θ can be any value, but ideally, we want to minimize the loss of information when using a point estimate to summarize a distribution. An application of SDT provides a method for selecting the optimal point estimator associated with a posterior distribution. Based on the principles described in *Expected loss (risk)*: *Bayesian expected loss and risk*, we use data, a loss function, and prior information to find Bayes' rule for point estimation, highlighting that

Bayes' rule explicitly depends on the choice of loss function. If we choose an estimator without explicitly considering loss, we are implicitly, and possibly inadvertently, assuming a specific form of loss, regardless of its appropriateness for our situation (c.f. the utility theorem of Morgenstern and Von Neumann 1953).

Squared-error loss.—The most ubiquitous loss function in statistics is squared-error loss, defined as $L(\theta,a)=(\theta-a)^2$. In an estimation problem, a is an estimate of θ chosen by the decision maker. Using the Henslow's Sparrow example in a different context, suppose we are trying to guess the abundance of Henslow's Sparrows in the upcoming year before the birds arrive from their wintering areas, and guesses will be penalized using squared-error loss relative to the abundance calculated after birds arrive. The guess is the action and the abundance calculated after the birds arrive is the truth (assuming we can perfectly estimate abundance). If the abundance is 75 ($\theta=75$) and the manager guesses 50 ($a=50$) the squared-error loss is $(75-50)^2=625$ (Fig. 2A). Similarly, if the decision maker correctly chooses $a = 75$, no loss would be incurred (Fig. 2A). In this context, θ represents the true abundance and is calculated after birds arrive in the spring. In the Henslow's Sparrow example, we are interested in $N_{a,\theta}$, the cumulative abundance over a 20-yr period, a function of unknown model parameters θ. The symbol θ represents different unknowns in each problem.

Notable properties of squared-error loss are that large errors are penalized relatively more than small errors, and the penalty is symmetric about a. These properties may or may not be appropriate for the decision problem. For example, if θ represented population growth rate of a critically endangered species, it would be unwise to overestimate θ, which could lead to the erroneous conclusion that the population was growing, when in fact it might have been declining. In such a situation, more loss should be given to overestimation than underestimation (Fig. 2B), and squared-error loss would not appropriately represent the objectives of the decision maker. The popularity of squared-error loss stems from its relationship to least squares theory and the normal distribution, its use for considering unbiased estimators of θ, and its relative computational ease (Berger 1985). The Bayesian risk for the squared-error loss function is

$$\rho(a) = \int_\theta (\theta-a)^2[\theta|y]d\theta. \qquad (6)$$

Note that Eq. 6 includes the investigator's understanding of loss (and ability to quantify loss in a mathematical function) and the posterior distribution, which is a function of data, the likelihood, and the investigator's belief about the prior distribution. To find Bayes' rule, we differentiate Eq. 6 with respect to a, set it equal to 0, and solve for a, resulting in the estimator $a^* = E(\theta|y)$ (see Appendix S1 for details); the results of this

optimization reveal that Bayes' rule for squared-error loss is $E(\theta|y)$, the posterior mean. Putting this in context, assuming squared-error loss is the appropriate loss function, the optimal estimator of θ (i.e., the estimator that minimizes the expected squared-error loss), among all possible estimators (e.g., mean, median, mode), is the posterior mean.

In the opposite direction, a more general result holds for the form of the loss function given the choice of the posterior mean as the point estimator of θ. If investigators choose the posterior mean as a summary statistic, they are implicitly choosing the broader class of loss functions described by

$$L(\theta,a) = f'(a)(a-\theta) - f(a) + g(\theta), \qquad (7)$$

(known as the Bregman function; Banerjee et al. 2005), where $f(a)$ is twice differentiable and independent of θ, $g(\theta)$ is independent of a, and $L(\theta,a)$ is convex to assure a global minimum of Bayes' rule (see Appendix S1 for details). Function 7 becomes squared-error loss when $f(a) = a^2$ and $g(\theta) = \theta^2$. The appropriateness of assuming Eq. 7 is context dependent and might not always hold. Thus, if an investigator chooses the posterior mean to summarize the data and prior, (s)he should ensure a specific form of Eq. 7 satisfies his/her perception of loss (i.e., it is consistent sensu Gneiting 2011). We provide general guidance for choosing loss functions in *Constructing loss functions*.

Absolute loss.—An assumption of squared-error loss is that an investigator wants to penalize large values of $(\theta-a)$ relatively more than small values. If the investigator wants to penalize larger deviations less, one option is the absolute error loss, $|\theta-a|$. The Bayesian risk for the absolute error loss function is $\rho(a) = \int_\Theta |\theta-a|[\theta|y]d\theta$, of which we find the value of a that minimizes $\rho(a)$ (see Appendix S1 for details). The posterior median is Bayes' rule for absolute error loss.

0–1 loss.—Zero-one loss is appropriate for the estimator selection problem when a decision maker desires the same penalty for an estimate whenever the estimate does

not equal the true value of θ, regardless of how far the estimate is from the true value. A loss of 0 is given when the estimate equals the true value of θ and a loss of 1 is given otherwise. The Bayesian risk for 0-1 loss results in the estimator $a^* = \text{mode}(\theta|y)$ (see Appendix S1 for details). The posterior mode is Bayes' rule for 0–1 loss. Kadane and Dickey (1980) demonstrate an extension of 0–1 loss and its relationship to Bayes' factors as Bayes' rule. In the case of symmetric, uni-modal posterior distributions, the mean, median, and mode are all equivalent. Thus, careful consideration of loss functions is critical for cases when the posterior distribution is multi-modal, skewed, or has other complex properties. In addition to the Bayes rules here, Table 2 provides a list of Bayes rules for other common loss functions.

The choice of loss function is critical for deciding on the appropriate inference. Gneiting (2011) demonstrates the importance of using the Bayes rule, or selecting consistent loss functions to evaluate forecasts, and that grossly misguided inference might result if loss functions and estimators are not carefully matched. In the next example, we consider a generalization of SDT from statistical inference procedures to a situation that may represent a common scenario in applied ecological research and management.

Optimal prescribed fire frequency for Henslow's Sparrows

Our example revisits the scenario described earlier and involves a refuge manager deciding on a prescribed burn interval to implement in a management plan for Henslow's Sparrows on eight grasslands. This management situation at Big Oaks National Wildlife Refuge in southeastern Indiana, USA, includes (1) an action set consisting of four burn intervals $a \in \{1,2,3,4\}$; one of which will be chosen for all grasslands for the next 20 yr, (2) an unknown cumulative abundance of Henslow's Sparrows for a 20-yr interval ($N_{a,\theta}$) that depends on unknown model parameters θ (i.e., θ represents the model parameters β, η_j, and σ^2 in Eq. 8) and the choice of management action a, (3) data from an investigation designed to estimate the effect of fire on Henslow's Sparrow abundance (y) at eight grasslands for a 4-yr

TABLE 2. Common loss functions, their mathematical formula, Bayes' rule, and whether the functions are symmetric, and their shape

Loss function	Formula	Bayes' rule	Symmetric	Shape
Squared-error loss	$(\theta-a)^2$	posterior mean	yes	convex
Weighted squared-error loss	$w(\theta)(\theta-a)^2$	$\frac{\int_\Theta \theta w(\theta)[y|\theta][\theta]d\theta}{\int_\Theta w(\theta)[y|\theta][\theta]d\theta}$	no provided $w(\theta) \neq 1 \forall \theta$	convex
Absolute-error loss	$|\theta-a|$	posterior median	yes	piece-wise linear
Linear loss	$c_1(\theta-a), \theta > a$ $c_2(a-\theta), \theta < a$	$\frac{c1}{c1+c2}$ quantile of posterior distribution	no provided $c1 \neq c2$	linear
0–1 loss	$0, \theta \in a$ $1, \theta \notin a$	posterior mode		piece-wise constant
Linex loss	$e^{c(a-\theta)} - c(a-\theta) - 1, c > 0$	$\frac{-\log \int_\Theta e^{-c\theta}[\theta|y]d\theta}{c}$	no (c controls asymmetry)	convex

period following prescribed fire, (4) prior information from another study on the relationship between Henslow's Sparrow abundance and the time since prescribed fire (e.g., Herkert and Glass 1999), and (5) a specification of loss designed to capture the expense of the management action and the importance of Henslow's Sparrows to managers. We begin by constructing a loss function that depends on the management action (a) and the unknown state of nature (i.e., the cumulative abundance of birds over a 20-yr period; $N_{a,\theta}$), but first some notational clarification is required.

Until now, we have described the loss function as a function of the action a and unknown value θ. In our example, the unknown cumulative abundance $N_{a,\theta}$ is a function of unknown model parameters θ (see Eq. 9), and $N_{a,\theta}$ is itself unknown. Thus, the loss function could still be described as $L(\theta, a)$, but it is perhaps more natural to interpret loss as a function of the unknown cumulative abundance $N_{a,\theta}$, instead of model parameters θ. Therefore, we write $L(N, a)$ instead of $L(\theta, a)$, for clarity.

To describe loss in terms of the management action, we first developed several axioms the loss function should meet, then developed a quantitative loss function that met all of the axioms. The first axiom was that frequent fire intervals are more costly than infrequent intervals and therefore, all else being equal, frequent fire intervals have higher loss. Second, if cumulative abundance of Henslow's Sparrows increases, loss decreases. For our third axiom, we assumed the manager had a dedicated budget for Henslow's Sparrow management; if the manager meets the abundance objective (or comes close to meeting the objective), the amount spent is proportionately less important than if the manager was far from meeting the objective. If the manager does not meet the objective, the amount spent is wasted and has proportionately higher loss than larger abundances. This reflects diminishing marginal returns of saving money as the true cumulative abundance increases. Thus our axiom was when the cumulative abundance of Henslow's Sparrows increases, cost becomes less important. Given these axioms, we developed a simple quantitative expression for the loss function as

$$L(N,a) = \begin{cases} \alpha_0(a) + \alpha_1(a)N_{a,\theta}, & N_{a,\theta} < 1835 \\ 0, & N_{a,\theta} \geq 1835 \end{cases}$$

(Fig. 5A). The loss function is a piecewise function with the first component being a line with negative slope (α_1) and intercept (α_0) that depends on a and the second component equal to 0 when the abundance is greater than 1835 birds (i.e., the population objective). We chose the intercepts (1, 0.9, 0.8, and 0.7 for 1, 2, 3, and 4 yr burn intervals, respectively) so that more frequent burn intervals would have higher loss, and we scaled the slope ($\alpha_1(a) = \frac{-\alpha_0(a)}{1835}$) so the loss would be 0 if the average annual population size reached 1835 birds. Thus, cost was incorporated in the differing slopes and intercepts for each action. We selected our management objective by identifying what the maximum cumulative number of

Henslow's Sparrows could be, given our model (this process is described in more detail below).

A hierarchical Bayesian statistical model can provide inference for the unknown cumulative 20-yr abundance of Henslow's Sparrows for each burn interval such that

$$\begin{aligned} y_{j,t} &\sim \text{Poisson}(A_j \lambda_{j,t}), \\ \log(\lambda_{j,t}) &= \boldsymbol{x}'_{j,t}\boldsymbol{\beta} + \eta_j, \\ \boldsymbol{\beta} &\sim \text{Normal}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}), \\ \eta_j &\sim \text{Normal}(0,1), \end{aligned} \quad (8)$$

where $y_{j,t}$ are the counts of Henslow's Sparrows at site $j = 1, \ldots, 8$ during years $t = 1, \ldots, T = 4$, A_j is the area of site j, $\lambda_{j,t}$ is the unknown density of Henslow's Sparrows at site j in time t and is a function of $\boldsymbol{\theta}' \equiv (\boldsymbol{\beta}, \boldsymbol{\eta})$, and $\boldsymbol{x}_{j,t}$ represents the categorical explanatory variable *summers-post-burn* in site j, year t. The η_j ($j = 1, \ldots, 8$) account for differences in densities among sites. We assumed η_j had mean 0 and variance equal to one to reflect the variation in densities among sites. We choose one as the variance because past estimates of densities at Big Oaks were usually between 0 and 2 birds/ha. The mean vector $\boldsymbol{\mu}' = (-5.0, 2.5, 0.2, 0.2)$ for the prior distribution of β was obtained by scaling density estimates of Herkert and Glass (1999, obtained from their Fig. 1) to densities for our study design. We let $\sigma^2 = 10$ to reflect our uncertainty in $\boldsymbol{\mu}$ because Herkert and Glass (1999) focused on a study site in a different state and during a different time period. The model was fit using an MCMC algorithm in R version 3.0.2 (R Core Team 2013; Software S1).

The estimated posterior distributions for β are shown in Fig. 4. We calculated the posterior distributions for the cumulative 20-yr abundance of Henslow's Sparrows across the eight sites using the derived quantity

$$N_{a,\theta} = \lim_{\tilde{T} \to \infty} \frac{20 \sum_{j=1}^{8} \sum_{t=T+1}^{\tilde{T}} A_i \lambda_{j,t}(a,\theta)}{\tilde{T} - T} \quad (9)$$

(Fig. 5B). Equation 9 calculates the expected annual abundance across the eight sites over an infinite time horizon, for each management action, and multiplies the expected annual abundance by 20 to scale it to the relevant management time frame. The limit in Eq. 9 represents annual abundance for each potential action and is multiplied by 20 to avoid an incomplete cycle for a 3-yr burn rotation. The management objective of 1835 individuals was chosen because it represented a large but attainable value of $N_{a,\theta}$. We calculated the Bayesian expected loss in Eq. 3 for each burn interval (see Software S2 for details) as

$$\rho(a) = E_{N_{a,\theta}|y} L(N,a) = \int_N L(N,a)[N_{a,\theta}|y] dN.$$

The resulting posterior risk for the burn intervals of 1, 2, 3, and 4 was 0.65, 0.27, 0.34, and 0.26 respectively. These results indicate that, despite a 2-yr burn interval appearing to produce the largest number of birds in the observed data (Fig. 5B), when including our loss
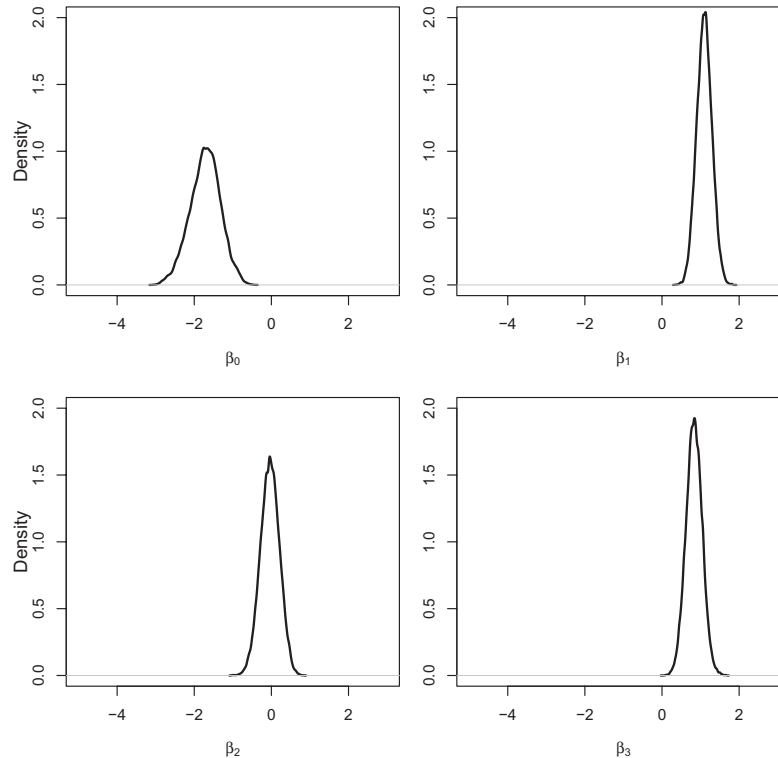
Fig. 4. Posterior distributions of β obtained from fitting Eq. 8 to Henslow's Sparrow data using the MCMC algorithm. The log density of site j, t years after a burn was: $\log(\lambda_{j,t}) = \beta_0 + \eta_j$ for $t=1$ and $\log(\lambda_{j,t}) = \beta_0 + \beta_{t-1} + \eta_j$ for $t=2,3,4$. The posterior distributions of β were used to derive cumulative Henslow's Sparrow abundance over a 20-yr period using Eq. 9.

functions associated with increasing financial cost, a 4-yr burn interval was Bayes' rule for management of grasslands for Henslow's Sparrows.

Constructing Loss Functions

In the previous sections, we developed or assumed various formulations of loss functions. Because statistical inference can be linked to decision theory through the incorporation of a loss function, guidance for constructing loss functions is important. In applied settings, practitioners of SDT assume functions for $[y|\theta]$, $[\theta]$, and $L(\theta,a)$. Ecologists are likely more comfortable with assuming relationships for the first two terms than the last term, and there are a large number of references that provide guidance on choice of the likelihood and prior distributions (e.g., Royle and Dorazio 2008, Rohde 2014, Gelman et al. 2014, Hobbs and Hooten 2015). In the ecological literature, however, much less guidance is available for constructing a loss function.

Hennig and Kutlukaya (2007:21) wrote, "the task of choosing a loss function is about the translation of an informal aim or interest that a researcher may have in the given application into the formal language of mathematics." Ultimately, the choice of a loss function is as subjective as a likelihood or prior because it reflects the knowledge of the decision maker. Different decision makers may likely construct different loss functions for the same problem. There are no generically optimal loss functions because optimizing a loss function would require specifying another loss function over which to optimize. The challenge for the decision maker is to translate their perception of loss (or utility) into a mathematical formula, which is usually not a trivial task. Due to the difficulty in translating a researcher's knowledge into a mathematical equation, the vast majority of applications of SDT, and decision theory in general, rely on some standard form of loss function. Hennig and Kutlukaya (2007:22–23) report that the majority of applications of statistical prediction and point estimation problems use versions of the squared-error loss function due to the simplicity of mathematics and "the self-confirming nature of the frequent use of certain 'standard' methods in science." We contend that when there is not a clear path forward for the development of an application-specific loss function, and a standard loss function does not contradict existing knowledge and goals, then using a standard loss function is a starting point. At least an explicit choice is made rather than acceptance of one that may contradict existing knowledge or goals.

We have collected the most popular loss functions used in academic research in Table 2. We arranged these in terms of two basic properties of each loss function: symmetry and relative shape. These properties can assist in
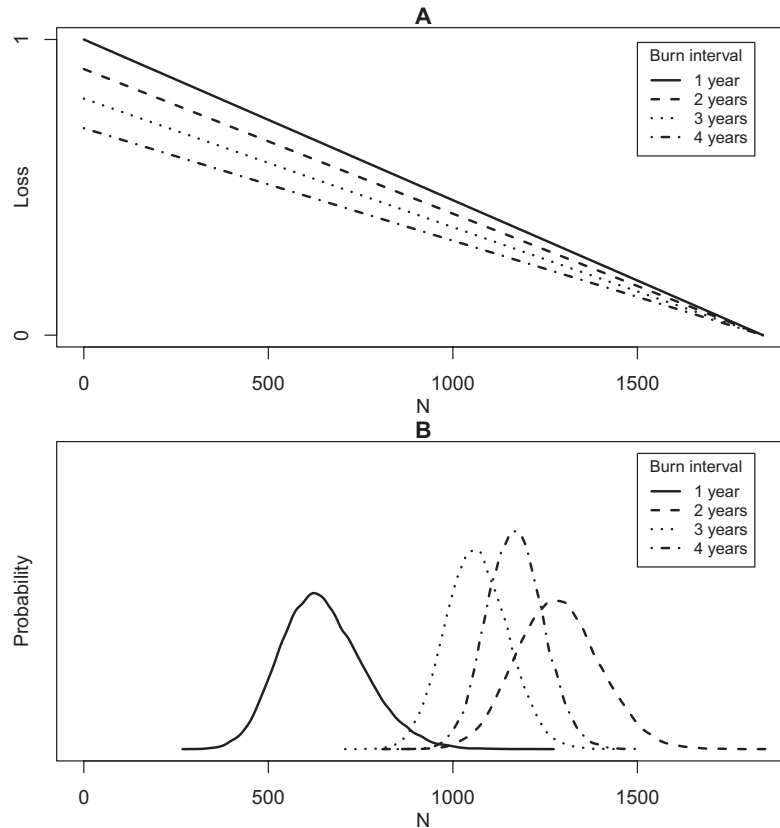
FIG. 5. (A) Loss functions for the example of four prescribed-fire frequencies considered for managing grasslands for Henslow's Sparrow habitat. (B) Posterior probability distributions $[N_{a,\theta}|y]$ of average annual abundance of Henslow's Sparrows given the decision of burn interval. Posterior risks were calculated by convolution of the loss functions in (A) with the posterior distribution in (B) for each interval. The posterior risks for burn intervals 1, 2, 3, and 4 were: 0.65, 0.27, 0.34, and 0.26, respectively. Thus, Bayes' rule was a 4-yr burn interval.

selecting from among standard loss functions. As noted previously, a symmetrical loss function penalizes underestimation of the true state of nature the same as overestimation. In many cases, symmetrical loss functions are appropriate, for example, when estimating animal locations from telemetry or satellite data. Usually, there is no reason to systematically penalize location error in one direction over another for every location in a data set (but see Brost et al. 2015). Although symmetric loss functions are the most ubiquitous in statistics, we find it easier to envisage ecological examples in which asymmetric loss functions are more appropriate. We described two scenarios in the cases of population growth rate of an endangered species and fire management for Henslow's Sparrows. Other examples include estimation of maximum dispersal distance of invasive species, estimating minimum habitat requirements of species for the development of protected areas, and modeling an animal's behavior of switching among mating and eating. In the first two cases, underestimating the true value would have severe consequences; invasive species could colonize areas thought to be beyond its dispersal distance, and resources invested in developing protected areas would be wasted on an area too small for a species to persist. In the third case, animals must allocate resources to both their probability of surviving and reproducing (e.g., songbirds choosing either to sing or collect food). The loss associated with starvation is likely larger than reduction of reproductive potential. Ver Hoef and Jansen (2007) use the asymmetric linex loss to correct for prediction bias in a space–time model of Harbor seal counts.

In addition to symmetry, a second important consideration is the relative shape of the loss function. The relative shape describes the penalty for increasingly large differences between the action and the true state of nature. How should a decision maker proceed in determining the curvature of the loss function? There are no easy rules for choosing a shape. However, one important aspect, whether the function is concave or convex, can be determined by axioms of the decision problem (as we described in the Henslow's Sparrow example). Most common loss functions used in applications are convex or linear (Table 2). These shapes assume that large errors are penalized relatively more than small errors, or relatively equally to small errors, respectively. Although

concave loss functions are rare in practice, it does not preclude their consideration from use. Concave loss functions assume that increasing error has diminishing marginal loss. After an error threshold has been reached, increasing error only has an arbitrarily small increase in loss. That is, if your choice of action is wrong, it might as well be very wrong; in this sense, concave loss functions are similar to 0–1 loss.

It is the decision maker's responsibility to specify the shape of loss and provide an explanation for their choice (as with study design, likelihood, and priors). When detailed loss information is lacking, we recommend that decision makers begin by considering standard loss functions like those reported in Table 2. A decision maker can narrow their choices by considering the simple properties of symmetry and relative shape which reflect the importance of over-estimation vs. under-estimation, and the relative penalty for increasing distance between actions and the true state of nature, respectively. The more the loss function reflects the decision maker's objectives, the more satisfied (s)he will be with the resulting decisions. Regardless of the initial choice of loss function, if the decision maker's choices are well-articulated and transparent, they lend themselves to rational debate among collaborators and peers. This transparency will lead to improvements in loss function specifications over time (c.f. double-loop learning, e.g., Johnson 2006).

Discussion

Traditional statistical inference was dedicated to learning about a process from data collected during a statistical investigation, without regard for how the inference would be used (Berger 1985). Fisher (1955:77) stated as statisticians "we aim, in fact, at methods of inference which should be equally convincing to all rational minds, irrespective of any intentions they may have in utilizing the knowledge inferred." In contrast, SDT is an extension of traditional statistical inference that pairs inference with the motives of a decision maker in a decision theoretic framework (Wald 1950, Savage 1954). This pairing is natural in applied ecology because data are often collected with the explicit purpose to inform decisions. SDT provides the formal link between advanced statistical methods for ecological inference and ecological decision making. The pairing is made through specification and integration of a loss function. We provided two different problems (point estimation and Henslow's Sparrow management) with varying implications for ecological investigation.

Our first problem of choosing point estimators for Bayesian posterior distributions illustrates two important messages. First, the choice of point estimator for posterior distributions is not arbitrary. Optimal estimators exist, given the choice of a loss function. Second, whether an investigator uses SDT or not, given the choice of estimator, there is an underlying assumed form of the loss function that is optimal. This notion is closely related

to the utility theory developed by Morgenstern and Von Neumann (1953) who proved that given a decision maker met a set of axioms, there existed a utility function representing their preferences. Generally, a class of loss functions has the same Bayes' rules for point estimation, and a choice of estimator implies an investigator's belief in the class of loss functions. These concepts also apply to evaluating forecast data (Gneiting 2011).

Our second problem demonstrates the flexibility of SDT as a general framework for applied decision problems. Natural resource agencies regularly collect information about processes for which they must make management decisions. How to link that information with the decision is often not well understood and the tasks of data collection and decision making are ad hoc and done in two unlinked steps. There are several potential issues with a two-step approach. First, relevant information collected on a process might be lost when using an ad hoc approach. Thus, there is no guarantee the decision will be optimal. Second, optimality is only defined with respect to a loss function and therefore, without a loss function, there are no optimal decisions. Third, how the decision maker came to their decision can be opaque without a transparent process of finding an optimal decision. SDT provides such a process for optimizing actions given data.

In each of these SDT problems, the decision maker has the additional task of explicitly defining a loss function. For point estimation, loss was described by a function quantifying an incorrect point estimate. For the Henslow's Sparrows, loss was described by a function quantifying the objectives of minimizing cost and maximizing cumulative abundance. Several authors have commented on the difficulty of choosing loss functions and how this difficulty often precludes the implementation of SDT (e.g., Fisher 1935, Tiao and Box 1973, Spanos 2012). The complexities of specifying or choosing a reasonable loss function are not trivial, and it is unlikely that loss functions can be developed for every problem. In some cases, there are obvious choices of loss functions (e.g., minimizing bias). In other situations, loss can be based on a set of pre-defined objectives. For example, adaptive harvest management of mallards (*Anas platyrhynchos*) in North America relies on a loss function based on the two objectives of maximizing long-term cumulative harvest and maintaining a population size >8 100 000 individuals (Johnson et al. 1997, Nichols et al. 2007). As a general principle for developing a loss function, we emphasize that decision makers first clearly articulate a set of axioms based on their objectives that the loss functions should meet and develop their loss function relative to those axioms. In this sense, developing a loss function is analogous to developing statistical models based on hypotheses of ecological process; the loss function is a model for true loss.

Many other decision frameworks are closely related to SDT. Structured decision making, adaptive management, and game theory each concern analytic tools of

evaluating decisions based on expected loss. For example, Markov decision processes (MDPs) solved using stochastic dynamic programming are used for making state-dependent decisions by calculating expected loss through time, with action-specific time-varying transition probabilities (Puterman 2014). In contrast to SDT, problems addressed using MDPs assume the true state of nature is known at the time of the decision; a difficult assumption to validate for many ecological problems. Partially observable MDPs (POMDPs) account for uncertainty in the true state of nature (Williams 2009, 2011) and are generalizations of SDT for recurrent decisions.

Other important concepts related to ecological inference with critical ties to SDT are model selection (Akaike 1973, Gelfand and Ghosh 1998, Hooten and Hobbs 2015, Williams 2016) and adaptive monitoring designs (Wikle and Royle 1999, 2004, Hooten et al., 2009, 2012). Akaike's information criterion was derived in a decision theoretic framework and is based on choosing a model that minimizes the approximated expected Kullback–Leibler loss function (Akaike 1973). The Kullback–Leibler loss function is attractive because it provides a theoretical basis for model selection (Burnham and Anderson 2002). Gelfand and Ghosh (1998) use SDT to address model selection for a more general class of loss functions. In adaptive sampling, the predictive variance of a process of interest or some other design criterion is the loss function and the sampling design that minimizes the expected loss is the optimal action (sampling design) chosen (Wikle and Royle 1999, Hooten et al., 2009, 2012).

Bayesian SDT provides the capability of explicitly incorporating prior information into the decision. Additionally, computational methods common for fitting Bayesian models (i.e., MCMC) can easily be extended to calculate Bayesian risk and select the Bayes rule for applied decision problems (as demonstrated in Software S2). This relatively simple extension of Bayesian analysis provides a framework for thinking about and analyzing almost any ecological decision problem.

Literature Cited

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov, and F. Csákieditors. Second International Symposium on Information Theory. Akadémiai Kiadó, Budapest, Hungary.

Banerjee, A., X. Guo, and H. Wang. 2005. On the optimality of conditional expectation as a Bregman predictor. IEEE Transactions on Information Theory, 51:2664–2669.

Barnard, G. A. 1954. Simplified decision functions. Biometrika 41:241–251.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer, New York, New York, USA.

Birnbaum, A. 1977. The Neyman-Pearson theory as decision theory, and as inference theory; with a criticism of the Lindley-Savage argument for Bayesian theory. Synthese 36:19–49.

Brost, B. M., M. B. Hooten, E. M. Hanks, and R. J. Small. 2015. Animal movement constraints improve resource selection inference in the presence of telemetry error. Ecology 96:2590–2597.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York, New York, USA.

Casella, G., and R. L. Berger. 2002. Statistical inference. Duxbury, Pacific Grove, California, USA.

Cox, D. R. 1958. Some problems connected with statistical inference. Annals of Mathematical Statistics 29:357–372.

De Finetti, B. 1937. La prévision: ses lois logiques, ses sources subjectives. Annales de l'institu Henri Poincaré 7:1–68.

DeGroot, M. H.1970. Optimal statistical decisions. John Wiley & Sons, Hoboken, New Jersey, USA.

Dorazio, R. M., and F. A. Johnson. 2003. Bayesian inference and decision theory: a framework for decision making in natural resource management. Ecological Applications 13:556–563.

Ferguson, T. S. 1967. Mathematical statistics: a decision theoretic approach. Academic Press, New York, New York, USA.

Fisher, S. R. A. 1935. The design of experiments. Oliver and Boyd, Edinburgh, UK.

Fisher, R. A. 1955. Statistical methods and scientific induction. Journal of the Royal Statistical Society B 17:69–78.

Gelfand, A. E., and S. K. Ghosh. 1998. Model choice: a minimum posterior predictive loss approach. Biometrika 85:1–11.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2014. Bayesian data analysisCRC Press, Boca Raton, Florida, USA.

Gneiting, T. 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106: 746–762.

Hennig, C., and M. Kutlukaya. 2007. Some thoughts about the design of loss functions. REVSTAT – Statistical Journal 5:19–39.

Herkert, J. R., and W. D. Glass. 1999. Henslow's sparrow response to prescribed fire in an Illinois prairie remnant. Studies in Avian Biology 19:160–164.

Hobbs, N. T., and M. B. Hooten. 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton, New Jersey, USA.

Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. Ecological Monographs 85:3–28.

Hooten, M. B., C. K. Wikle, S. L. Sheriff, and J. W. Rushin. 2009. Optimal spatio-temporal hybrid sampling designs for ecological monitoring. Journal of Vegetation Science 20:639–649.

Hooten, M. B., B. E. Ross, and C. K. Wikle, 2012. Optimal spatio-temporal monitoring designs for characterizing population trends. Pages 443–459 in R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, editors. Design and analysis of long-term ecological monitoring studies. Cambridge University Press, Cambridge, UK.

Johnson, F. A. 2006. Adaptive harvest management and double-loop learning. Transactions of the Seventy-first North American Wildlife and Natural Resources Conference 71:197–213.

Johnson, F. A., C. T. Moore, W. L. Kendall, J. A. Dubovsky, D. F. Caithamer, J. R. Kelley Jr, and B. K. Williams. 1997. Uncertainty and the management of mallard harvests. Journal of Wildlife Management 61:202–216.

Kadane, J. B., and J. M. Dickey, 1980. Bayesian decision theory and the simplification of models. Pages 245–268 in Evaluation of econometric models. Academic Press, Waltham, Massachusetts, USA.

Lehmann, E. L., and G. Casella. 1998. Theory of point estimation. Springer, New York, New York, USA.

Lehmann, E. L., and J. P. Romano. 2008. Testing statistical hypotheses. Springer, New York, New York, USA.

Lindley, D. V. 1953. Statistical inference. Journal of the Royal Statistical Society B 15:30–76.

Lindley, D. V. 1971. Bayesian statistics: a review. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

Morgenstern, O., and J. Von Neumann. 1953. Theory of games and economic behavior. Princeton University Press, Princeton, New Jersey, USA.

Neyman, J., and E. S. Pearson. 1928. On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika 20A:175–240.

Neyman, J., and E. S. Pearson. 1933. On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society A 231:289–337.

Nichols, J. D., M. C. Runge, F. A. Johnson, and B. K. Williams. 2007. Adaptive harvest management of North American waterfowl populations: a brief history and future prospects. Journal of Ornithology 148:343–349.

Pacific Flyway Council. 1999. Pacific Flyway management plan for the cackling Canada goose. Cackling Canada Goose subcommittee, Pacific Flyway Study committee, Portland, Oregon, USA.

Pratt, J. W., H. Raiffa, and R. Schlaifer. 1995. Introduction to statistical decision theory. MIT Press, Cambridge, Massachusetts, USA.

Puterman, M. L. 2014. Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons, Hoboken, New Jersey, USA.

R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Ramsey, F. P. 1931. Truth and probability (1926). Pages 156–198 in R. B. Braithwaite, editor. The foundations of mathematics and other logical essays. Harcourt, Brace, and Company, New York, New York, USA.

Rohde, C. A. 2014. Introductory statistical inference with the likelihood function. Springer, New York, New York, USA.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press, Waltham, Massachusetts, USA.

Savage, L. J. 1954. The foundations of statistics. Wiley, New York, New York, USA.

Savage, L. J., 1962. Bayesian statistics. Pages 161–194 in R. Machol, and P. Gray, editors. Recent developments in information and design processes. Macmillan Company, New York, New York, USA.

Spanos, A. 2016. Why the decision theoretic perspective misrepresents frequentist inference: 'buts and bolts' vs. learning from data. arXiv:1211.0638v2.

Tiao, G. C., and G. E. Box. 1973. Some comments on Bayes estimators. American Statistician 27:12–14.

Tukey, J. W. 1960. Conclusions vs decisions. Technometrics 2:423–433.

Ver Hoef, J. M., and J. K. Jansen. 2007. Space–time zero-inflated count models of Harbor seals. Environmetrics 18:697–712.

Wald, A. 1950. Statistical decision functions. Wiley, New York, New York, USA.

Wikle, C. K., and J. A. Royle. 1999. Space–time dynamic design of environmental monitoring networks. Journal of Agricultural, Biological, and Environmental Statistics 4:489–507.

Wikle, C. K., and J. A. Royle. 2004. Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. Environmetrics 16:507–522.

Williams, B. K. 2009. Markov decision processes in natural resources management: observability and uncertainty. Ecological Modelling 220:830–840.

Williams, B. K. 2011. Resolving structural uncertainty in natural resources management using POMDP approaches. Ecological Modelling 222:1092–1102.

Williams, P. J. 2016. Methods for incorporating population dynamics and decision theory in cackling goose management. Dissertation. Colorado State University, Fort Collins, Colorado, USA.

SUPPORTING INFORMATION