Introduction: In progress...

Background: Let $\mathcal{X}$ be the observation space, Θ be the parameter space, and $\mathcal{A}$ be the action space. For simplicity, we allow all of these spaces to be discrete. The observations $x \in \mathcal{X}$ are connected to the parameter $\theta \in \Theta$ by the probability mass function $p(x|\theta)$, referred to as the data-generating process (DGP) [1]. In a discrete setting, the DGP describes the probability of an observation $x \in \mathcal{X}$ under a given parameter θ. The primary objective of statistical inference is to infer underlying properties of the DGP [2]. From the perspective of decision theory, the decision $a \in \mathcal{A}$ will be to propose a function $\alpha(x)$ to estimate the parameter θ as precisely as possible. To illustrate this concept, suppose we are flipping a fair coin and wish to recover the parameter θ corresponding to the proportion of heads, $\theta = 0.5$. Thus, the DGP $p(x|\theta)$ is a Bernoulli distribution with parameter θ. One action $a_1 \in \mathcal{A}$ is to propose the estimator $\alpha_1(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$ (where n is the number of flips) whereas $a_2 \in \mathcal{A}$ is to naively propose $\alpha_2(x) = 1$ (every flip is heads). It can be shown[1] that a_1 proposes an estimator which maximizes the likelihood of the observed data under the DGP [3], whereas a_2's estimator is clearly biased, thus trivially $a_1 \succ a_2$. Unless necessarily distinct, we henceforth use estimators α and the actions a proposing them interchangeably.

To quantify the preference orderings beyond the simple heuristics mentioned in the coin-flipping case, statisticians leverage loss functions [4], which we denote $\mathcal{L}(\theta, \alpha)$. The loss function represents the error associated with proposing a "bad" evaluation of the θ (or function of θ) of interest. Thus, the best evaluation of this function is a zero loss; therefore $\mathcal{L}(\theta, \alpha) \geq 0$ [5]. From a decision-theoretic perspective, the objective of the decision-maker is to propose an estimator α which minimizes this loss. Since the true value of parameter θ is often unknown, the preference ordering of estimators is often dictated by their *expected* loss. However, exactly how we define *expected* relies upon whether one takes a frequentist or Bayesian approach.

Frequentism and Minimax: Under the frequentist paradigm, the data $x \in \mathcal{X}$ are considered random because it arises from repeated sampling via the DGP $p(x|\theta)$. Meanwhile, θ is treated as a fixed but unknown constant in the parameter space Θ. In the coin-flipping example, a frequentist would assume that the coin has a fixed (unknown) probability θ of landing heads, and each flip outcome is then governed by $p(x|\theta)$. Thus, to evaluating a proposed estimator α, the frequentist approach focuses on expected loss, akin to how Peterson [6] considers the expected utility. Specifically, we define the expected loss (EL) as the product of the probability of observing $x \in \mathcal{X}$ and the loss associated with estimating θ with $\alpha(x)$,

$$\text{EL}(\theta, \alpha) = \mathbb{E}_\theta[\mathcal{L}(\theta, \alpha)] = \sum_{x \in \mathcal{X}} \mathcal{L}\big(\theta, \alpha(x)\big)p(x|\theta) \tag{1}$$

The above is also referred to as a risk function [7]. From this definition of expected loss, we introduce the concept of "minimax" through a game-theoretic analogy of a game against Nature. In this framework, our goal is to select an estimator $\alpha \in \mathcal{A}$ that *minimizes* our expected loss. Meanwhile, Nature acts as an adversary, selecting a parameter $\theta \in \Theta$ (i.e. a "state of the world") in an attempt to *maximize* our expected loss [8]. The expected loss in

[1]Given $p(x|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$, the log-likelihood of the n observations is $\ell(x, \theta) = \log(\theta)\sum_{i=1}^{n} x_i + \log(1-\theta)\sum_{i=1}^{n}(1-x_i)$. Maximizing wrt θ yields $\hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \alpha_1(x)$.

such a game is known as the “minimax risk”, which we define as

$$\overline{R} = \min_{\alpha \in \mathcal{A}} \max_{\theta \in \Theta} \mathrm{EL}(\theta, \alpha) \tag{2}$$

The estimator/decision rule $\alpha \in \mathcal{A}$ that achieves the minimax risk is known as the minimax estimator. While the minimax risk $\overline{R}$ is occasionally criticized as being overly conservative [6], the ability of an estimator to be the best in the worst case scenario (which we refer to as the “minimax guarantee”) is desirable for many real-world applications including management of financial portfolios [9].

Having defined the minimax risk in Equation (2) and the corresponding guarantee, we now turn to *The Bayesian Choice* [5], in which Christian Robert demonstrates that under certain “least favorable” priors, Bayesian procedures achieve a Bayes risk that coincides with (or even improves upon) this frequentist minimax bound.

Remaining Work

1. Introduce Robert’s Argument and proof. (Bayesianism, Integrated Risk, proof of Integrated Risk $\leq$ Minimax Risk using weighted sum vs. set maxima)

2. Introduce Stark’s Counterargument: How the prior $\pi(\theta)$ is subjective, and Robert’s proof is trivial since you are “adding information” to the risk problem which was previously constrained by objectivity.

3. Introduce the Bayesian Rebuttal: Namely, the subjectivity of choice of loss function $\mathcal{L}(\dots)$ implies the frequentist construction of the problem isn’t operating under such “objective constraints,” so given that subjective claims need to be made on the state of Nature, a Bayesian approach gives provable optimality.

4. Conclusion and Introduction

References

[1] Jun Tu and Guofu Zhou. Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics*, 72(2):385–421, 2004.

[2] Graham Upton and Ian Cook. *Oxford Dictionary of Statistics*. Oxford University Press, 2008.

[3] Richard J. Rossi. *Mathematical Statistics: An Introduction to Likelihood Based Inference*. John Wiley & Sons, New York, 2018. p. 227.

[4] Abraham Wald. *Statistical Decision Functions*. Wiley, Oxford, England, 1950. Includes 76-item bibliography. PsycINFO Database Record (c) 2016.

[5] Christian P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. Springer-Verlag New York, 2 edition, 2007.

[6] Martin Peterson. *An Introduction to Decision Theory*. Cambridge University Press, New York, 2nd edition, 2017.

[7] M. S. Nikulin. Risk of a statistical procedure. In *Encyclopedia of Mathematics*. EMS Press, 2001. Originally published in 1994.

[8] V. Ulansky and A. Raza. Generalization of minimax and maximin criteria in a game against nature for the case of a partial a priori uncertainty. *Heliyon*, 7(7):e07498, Jul 2021.

[9] Xiao-Tie Deng, Zhong-Fei Li, and Shou-Yang Wang. A minimax portfolio selection strategy with equilibrium. *European Journal of Operational Research*, 166(1):278–292, 2005. Metaheuristics and Worst-Case Guarantee Algorithms: Relations, Provable Properties and Applications.