1 Introduction

Inference for treatment effects is a major concern of empirical work in economics, biostatistics, and many other fields. Often, analysis of treatments is motivated by the need to provide policy prescriptions, for example guidelines on whether to assign certain individuals to a job training program, or whether provide a certain medical therapy to patients. Yet, much of the methodological and empirical work on treatment effects focuses solely on estimation and hypothesis testing, and does not explicitly consider the basic treatment assignment problem driving the analysis. Manski (2000, 2002, 2004) and Dehejia (1999) point out that the problem of optimally choosing a treatment, based on empirical data, is distinct from the problem of estimating the treatment effect efficiently or testing hypotheses about a treatment effect, and advocate formal analysis of decision procedures which map empirical data into policy recommendations.

In this paper, we follow Manski (2004) by considering statistical approaches to treatment assignment within the framework of statistical decision theory. Unfortunately, it is difficult to calculate risk functions except in very simple cases, and finding minmax decision rules when the class of rules is large, can be computationally infeasible.\(^1\) For example, Manski (2004) considers a randomized experiment with a discrete covariate, and a bounded continuous outcome. Even in this setting, it is difficult to calculate the risk functions of various possible rules. Manski develops bounds on risks over a restricted class of “empirical conditional success” rules.

Motivated by the difficulty of obtaining useful exact results in realistically complex settings, we consider large-sample approximations to statistical treatment assignment problems. We focus on obtaining statistical decision rules that are asymptotically minmax for certain loss functions

\(^1\)For applications of minmax statistical decision theory in other applications, see Chamberlain (2000).
that arise naturally within the treatment assignment setting. We consider three loss functions: a two-point loss function that has been used in hypothesis testing theory, a “social welfare” loss function, and a regret version of the social welfare loss that has been proposed by Manski.

We use Le Cam’s asymptotic extension of the Wald (1950) statistical decision theory framework. The key idea is to obtain an asymptotic approximation to the entire statistical decision problem, not just a particular estimator or test statistic. Often, the approximate decision problem is considerably simpler, and can be solved exactly. Then, one finds a sequence of rules in the original problem that asymptotically matches the optimal rule in the simple limiting version of the decision problem. In regular parametric models, the treatment assignment problem is asymptotically equivalent to a simpler problem, in which one observes a single draw from a multivariate normal distribution with unknown mean and known variance matrix, and must decide whether a linear combination of the elements of the mean vector is greater than zero. Not surprisingly, there is a close connection between the treatment assignment problem and a one-sided hypothesis testing problem, except that some of the loss functions we considered differ from standard loss functions associated with hypothesis testing. In the limiting version of the treatment assignment problem, we derive a simple extension of an essential complete class theorem due to Karlin and Rubin (1956), which allows us to restrict attention to relatively simple types of rules. Then, we obtain exact minmax bounds and the minmax rules for the three loss functions we are considering. Finally, we use these exact results in the simple multivariate normal case to provide asymptotic minmax bounds, and sequences of decision rules which achieve these bounds, in the original sequence of decision problems. For a symmetric version of the two-point loss, and for the minmax regret criterion, a simple rule based on an asymptotically efficient parameter estimator (such as the MLE), is asymptotically minmax. Although this rule has a very natural form, it implies less conservative decision making than hypothesis testing at conventional significance levels.

We then consider semiparametric models where the welfare gain of the treatment can be captured by a scalar parameter. Under regularity conditions that permit estimation of the scalar parameter at a $\sqrt{n}$ rate, we derive similar results to the parametric case. The analysis is complicated by the presence of a possibly infinite-dimensional nuisance parameter. We make use of a uniformly most powerful property of certain semiparametric one-sided tests shown by Choi, Hall, and Schick (1996) in our local asymptotic minmax result. [Note: this section is preliminary.]

Our results are local asymptotic results: we reparametrize the models so that the problem of determining whether to assign the treatment (conditional on a specific covariate value) does not become trivial as the sample size increases. This is the same parameter localization commonly used in hypothesis testing theory, where one considers Pitman alternatives to the null hypothesis. Alternatively, one could examine asymptotic minmaxity from a large deviations perspective: see for example Puhalskii and Spokoiny (1998).

In the next section, we set up the basic statistical treatment assignment problem. In Section 3,
motivated by the difficulty of solving the problem exactly, we take a local parameter approach and show the limiting Gaussian form of the treatment assignment problem. In Section 4, we solve the approximate treatment assignment problem according to the minmax criterion, and then apply the solution to obtain asymptotic minmax bounds on risk and asymptotic minmax rules in the original sequence of decision problems. Section 5 then develops a semiparametric extension of our results.

2 Statistical Treatment Assignment Problem

2.1 Known Outcome Distributions

Following Manski (2000, 2002, 2004), we consider a social planner, who assigns individuals to different treatments based on their observed background variables. Suppose that individuals have covariate $X$, a random variable on some space $X$, with marginal distribution $F_X$. For simplicity the set of possible treatment values is $T = \{0, 1\}$. (The analysis can be generalized to a finite number of possible treatments. The planner observes $X = x$, and assigns the individual to treatment 1 with some probability

$$\delta(x) = Pr(T = 1|X = x).$$

Let $Y_0$ and $Y_1$ denote potential outcomes for the individual, and let their distribution functions conditional on $X = x$ be denoted $F_0(\cdot|x)$ and $F_1(\cdot|x)$ respectively. Given the rule $\delta$, the outcome distribution conditional on $X = x$ is

$$F_\delta(\cdot|x) = \delta(x)F_1(\cdot|x) + (1 - \delta(x))F_0(\cdot|x).$$

For a given outcome distribution $F$, let the social welfare be

$$W(F) = \int w(y)dF(y)$$

where $w : \mathbb{R} \to \mathbb{R}$ is a strictly increasing, concave function.\(^2\)

Suppose that $F_0$ and $F_1$ are known. Then, let

$$W_0(x) = \int w(y)dF_0(y|x), \quad W_1(x) = \int w(y)dF_1(y|x).$$

The optimal rule would have, for $F_x$-almost all $x$,

$$\delta^*(x) = \begin{cases} 
1 & \text{if } W_0(x) < W_1(x) \\
0 & \text{if } W_0(x) > W_1(x) 
\end{cases}$$

\(^2\)One could imagine a more general definition of social welfare, permitting more general mappings $W$ from the space of probability distributions into the reals. This would allow for distributional concerns such as a preference for equity.
(For \(x\) such that \(W_0(x) = W_1(x)\), any value of \(\delta^*(x)\) is optimal.)

2.2 Unknown Outcome Distributions

If \(F_0\) and \(F_1\) are not known, then the optimal rule described above is not feasible. For simplicity, suppose that \(W_0(x) = 0\), and suppose that \(F_1 \in \mathcal{F}_1\), where \(\mathcal{F}_1\) is associated with a parameter space \(\Theta\).\(^3\) We assume that, for any \(x \in \mathcal{X}\) the treatment effect (in social welfare or utility terms) \(W_1(x) - W_0(x)\) can be expressed as a known function \(g(x, \theta)\). Assume that \(g\) is continuously differentiable in \(\theta\) for \(F_X\)-almost all \(x\).\(^4\)

Suppose we have some data that are informative about \(\theta\). For example, we might have run a randomized experiment in the past that is informative about the treatment effect. We can express this as \(Z_n \sim P_{\theta}^n\), where \(\{P_{\theta}^n, \theta \in \Theta\}\) is a collection of probability measures on some space \(Z^n\). Here, we interpret \(n\) as the sample size, and we will consider below a sequence of experiments \(E_n = \{P_{\theta}^n, \theta \in \Theta\}\) as the sample size grows.

**Example 1** Suppose that \(\mathcal{X} = \{1\}\), a singleton set, and that \(\theta \in \Theta \subset \mathbb{R}\) is corresponds to the expected treatment effect:

\[
\theta = W_1 - W_0.
\]

The infeasible rule would assign all individuals to treatment if \(\theta > 0\), and would assign all individuals to control if \(\theta < 0\). Suppose that \(\theta\) is not known, but we observe a single observation

\[
Z_1 = \theta + \epsilon,
\]

where \(\epsilon = N(0,1)\). For example, we might know that \(W_0 = 0\), and assign a single experimental individual to treatment. Then \(Z_1\) could represent the outcome (in utility terms) observed for the experimental individual.

A randomized statistical treatment rule is a mapping \(\delta : \mathcal{X} \times Z^n \rightarrow [0,1]\). We interpret it as the probability of assigning a (future) individual with covariate \(X = x\) to treatment, given past data \(Z_n = z\):

\[
\delta(x, z) = Pr(T = 1|X = x, Z_n = z).
\]

In order to implement the Wald statistical decision theory approach, we need to specify a loss function which connects actions with consequences. We consider three possible loss functions. The

\(^3\)The assumption that \(W_0(x)\) is known could be regarded as a normalization, but is not innocuous when considering minmax decision rules under one of the loss functions below, Loss B. However, we argue that minmax analysis does not yield sensible results for this particular loss function anyway, and focus on other forms of loss.

\(^4\)For a discussion of the relationship between the net social welfare and traditional measures of effects of treatments, see Dehejia (2003).
first is taken from standard hypothesis testing theory, and penalizes making the wrong choice by an amount that depends only on whether the optimal assignment is treatment or control:

**Loss A:**

\[
L^A(\delta, \theta, x) = \begin{cases} 
K_0 \cdot (1 - \delta) & \text{if } g(x, \theta) > 0 \\
K_1 \cdot \delta & \text{if } g(x, \theta) \leq 0 
\end{cases}
\]

\(K_0 > 0, \quad K_1 > 0\)

The next loss function is motivated by the interpretation of \(g(x, \theta)\) as social welfare:

**Loss B:**

\[
L^B(\delta, \theta, x) = -g(x, \theta) \cdot \delta.
\]

Finally, following Manski (2004), we consider a regret version of the welfare-based loss. Recall that the infeasible optimal treatment is \(\delta^*(x) = 1(g(x, \theta) > 0)\). The regret is the welfare loss of a rule, compared with the welfare loss of the infeasible optimal rule:

**Loss C:**

\[
L^C = L^B(\delta, \theta, x) - [-g(x, \theta)1(g(x, \theta) > 0)]
\]

\[= g(x, \theta)[1(g(x, \theta) > 0) - \delta].\]

The risk of a rule \(\delta(x, z)\) under loss \(L\) and given \(\theta\), is

\[
R(\delta, \theta) = EL(\delta(X, Z), \theta, X)
\]

\[= \int \int L(\delta(x, z), \theta, x)dP^\theta(z)dF_X(x)\]

A minmax decision rule over some class \(\Delta\) of decision rules, solves

\[
\inf_{\delta \in \Delta} \sup_{\theta \in \Theta} R(\delta, \theta).
\]

Notice that we can solve this problem for each fixed \(x\), and the resulting rule will be minmax for the overall problem (provided the solution is appropriately measurable with respect to \(X\)). Thus, in the remainder of the paper, we will focus on the “pointwise-in-\(X\)” minmax problem

\[
\inf_{\delta(x, \cdot) \in \Delta} \sup_{\theta \in \Theta} \int L(\delta(x, z), \theta, x)dP^\theta_0(z).
\]

### 3 Regular Parametric Models

In this section, we develop asymptotic results for regular (local asymptotic normal) parametric models. We argue that a local parametrization approach provides a useful way to make approximate
comparisons between treatment assignment rules, and we use the local asymptotic representation of regular parametric models by a Gaussian shift experiment to derive a simple approximate characterization of the decision problem.

### 3.1 Plug-in Rules and Local Parametrization

Suppose that $\Theta$ is an open subset of $\mathbb{R}^k$, and that the $\{P^n_\theta, \theta \in \Theta\}$ are dominated by some measure $\mu^n$ and satisfy conventional regularity conditions. Then a natural estimator of $\theta$ is the maximum likelihood estimator

$$\hat{\theta}(Z_n) = \arg \max_{\theta} \frac{dP^n_\theta}{d\mu}(Z_n),$$

and one possible treatment assignment rule is the “plug-in” rule

$$\hat{\delta}(x, Z_n) = 1(g(x, \hat{\theta}(Z_n)) > 0),$$

with associated risk

$$R_x(\hat{\delta}, \theta) = \int L(\hat{\delta}(x, z), \theta, x)dP^n_\theta(z)$$

$$= \int L(1(g(x, \hat{\theta}(z)) > 0), \theta, x)dP^n_\theta(z).$$

Except in certain special cases, the exact distribution of the MLE $\hat{\theta}$ under $\theta_0$ cannot be easily obtained, and as a consequence it is difficult to simply calculate the risk, much less find the rule which maximizes the worst-case risk. Although the exact distribution of the MLE is rarely known in a useful form, many asymptotic approximation results are available for the MLE and other estimators. This suggests that for reasonably large sample sizes, we may be able to use such approximations to study the corresponding decision rules.

First, consider the issue of consistency: As $n \to \infty$, the MLE and many other estimators satisfy $\hat{\theta} \overset{P}{\to} \theta_0$. This implies that the rule $\hat{\delta} = 1(g(x, \hat{\theta}) > 0)$ will be consistent, in the sense that if $g(x, \theta_0) > 0$, $\Pr(\hat{\delta} = 1) \to 1$, and if $g(x, \theta_0) < 0$, then $\Pr(\hat{\delta} = 0) \to 1$.

Although this is a useful first step, it does not permit us to distinguish between plug-in rules based on different consistent estimators, or to consider more general rules that do not have the plug-in form. We therefore focus on developing local asymptotic approximations to the distributions of decision rules.

For the MLE, under suitable regularity conditions we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{\theta_0}{\to} N(0, I^{-1}_{\theta_0}),$$

where we use $\overset{\theta_0}{\to}$ to denote convergence in distribution (weak convergence) under $\theta_0$, and $I_{\theta_0}$ is the
Fisher information matrix. By the delta method, we have that
\[
\sqrt{n}(g(x, \hat{\theta}) - g(x, \theta_0)) \xrightarrow{\theta_0} N(0, \hat{g}'I_{\theta_0}^{-1}\hat{g}),
\]
where \(\hat{g} = \frac{\partial}{\partial \theta} g(x, \theta_0)\). Consider an alternative estimator \(\tilde{\theta}\) with a larger asymptotic variance:
\[
\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\theta_0} N(0, \tilde{V}),
\]
where \(\tilde{V} - I_{\theta_0}^{-1}\) is positive definite. Then, since \(\tilde{\theta}\) is “noisier” than the MLE, we expect that the plug-in rule \(\tilde{\delta} = 1(g(x, \tilde{\theta}) > 0)\) should do worse than \(\hat{\delta}\). One way to make this reasoning formal, is to adopt the local parametrization (Pitman alternative) approach, which is commonly used in asymptotic analysis of hypothesis tests.\(^5\) In our setting, this means considering values for \(\theta\) such that \(g(x, \theta)\) is “close” to 0, so that there is a nontrivial difficulty in distinguishing between the effects of the two treatments as sample size grows. Specifically, assume that \(\theta_0\) is such that
\[
g(x, \theta_0) = 0,
\]
and consider parameter sequences of the form \(\theta_0 + \frac{h}{\sqrt{n}}\), for \(h \in \mathbb{R}^k\). For the MLE, it can typically be shown that.
\[
\sqrt{n}(\tilde{\theta} - \theta_0 - h/\sqrt{n}) \xrightarrow{\theta_0+h/\sqrt{n}} N(0, I_{\theta_0}^{-1}),
\]
where \(\xrightarrow{\theta_0+h/\sqrt{n}}\) denotes weak convergence under the sequence of probability measures \(P_{\theta_0+h/\sqrt{n}}\). We will sometimes abbreviate this as \(h\rightarrow\).

By assumption, for all \(h \in \mathbb{R}^k\), \(\sqrt{n}(g(x, \theta_0 + h/\sqrt{n}) - g(x, \theta_0)) \rightarrow \hat{g}'h\). Then by standard calculations
\[
P_{\theta_0+h/\sqrt{n}}(g(x, \hat{\theta}) > 0) \rightarrow 1 - \Phi \left( \frac{-\hat{g}'h}{\sqrt{\hat{g}'I_{\theta_0}^{-1}\hat{g}}} \right) = \Phi \left( \frac{\hat{g}'h}{\sqrt{\hat{g}'I_{\theta_0}^{-1}\hat{g}}} \right).
\]

As an illustration, consider Loss B, the social welfare-based loss function. In order to keep the loss from degenerating to 0 as sample size increases, we scale it up by the factor \(\sqrt{n}\), leading to a loss function
\[
L^B(\delta, h, x) = -\sqrt{n} \cdot g(x, \theta_0 + h/\sqrt{n})\delta(x, z).
\]
Then the risk can be written as
\[
R^B(\delta, h, x) = E_h L^B(\delta(x, Z), h, x)
\]
\(^5\) An alternative approach would be to use large-deviations asymptotics, in analogy with Bahadur efficiency of hypothesis tests. Manski (2003) uses finite-sample large-deviations results to bound the risk properties of certain types of treatment assignment rules in a binary-outcome randomized experiment. Puhalskii and Spokoiny (1998) develop a large-deviations version of asymptotic statistical decision theory and apply it to estimation and hypothesis testing.
= -\sqrt{n} \cdot g(x, \theta_0 + h/\sqrt{n}) \cdot E_h[\delta(x, Z)]

For the MLE plug-in rule,

\[ R^B(\hat{\delta}, h, x) = -\sqrt{n} \cdot g(x, \theta_0 + h/\sqrt{n}) \cdot P_h(g(x, \hat{\theta}) > 0) \]

\[ \rightarrow -\dot{g}'h \cdot \Phi\left(\frac{\dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}\dot{g}}}\right) \]

Now consider another estimator \( \tilde{\theta} \) with

\[ \sqrt{n}(\tilde{\theta} - \theta_0 - h/\sqrt{n}) \tilde{\theta}_0 + h/\sqrt{n} \overset{d}{\rightarrow} N(0, V), \]

where \( V - I_{\theta_0}^{-1} \) is positive semidefinite, and let \( \tilde{\delta} = 1(g(x, \tilde{\theta}) > 0) \). By straightforward calculations, it can be shown that for all \( h \),

\[ \lim_{n \to \infty} R^B(\tilde{\delta}, h, x) \geq \lim_{n \to \infty} R^B(\hat{\delta}, h, x). \]

Thus, \( \hat{\delta} \) asymptotically dominates \( \tilde{\delta} \), and the plug-in rule using the MLE is minmax among plug-in rules based on estimators which are asymptotically normal and unbiased. The other two loss functions also yield the same conclusion. However, this result is quite restricted. For example, a conventional hypothesis testing approach might choose the treatment only if \( g(x, \hat{\theta}) \) is greater than a strictly positive constant \( c \), rather than 0. In the next section, we examine the problem of finding asymptotically minmax decision rules without strong restrictions on the class of possible rules.

### 3.2 Limits of Experiments

In this section we use the Le Cam limits of experiments framework to examine the statistical treatment assignment problem. Although the Le Cam framework is typically applied to study point estimation and hypothesis testing, it applies much more broadly, to general statistical decision problems.

As before, we fix \( x \), and let \( \Theta \) be an open subset of \( \mathbb{R}^k \). Let \( \theta_0 \in \Theta \) satisfy \( g(x, \theta_0) = 0 \) for a given \( x \). Assume that the sequence of experiments \( E_n = \{P^n_{\theta}, \theta \in \Theta\} \) satisfies local asymptotic normality: for all sequences \( h_n \to h \) in \( \mathbb{R}^k \),

\[ \log \frac{dP^n_{\theta_0+h_n/\sqrt{n}}}{dP^n_{\theta_0}} = h'\Delta_n - \frac{1}{2}h'I_{\theta_0}h + o_P(1), \]

where \( \Delta \overset{d}{\to} N(0, I_{\theta_0}) \).\(^6\) Further, assume that \( I_{\theta_0} \) is nonsingular.

\(^6\)In the case where the \( P^n_{\theta} \) is the \( n \)-fold product measure corresponding to a random sample of size \( n \) from \( P_{\theta} \),
Then, by standard results, the experiments $\mathcal{E}_n$ converge weakly to the experiment

$$Z \sim N(h, I^{-1}_{\theta_0}).$$

By the asymptotic representation theorem, for any sequence of statistical decision rules $T_n$ which have limit distributions, there exists a feasible decision rule $T$ in the limit experiment such that $T_n \overset{h}{\sim} T$ for all $h$. In the case of statistical treatment rules, Theorem 9.4 and Corollary 9.5 of Van der Vaart (1998) can be specialized as follows:

**Proposition 1** Let $\Theta$ be an open subset of $\mathbb{R}^k$, with $\theta_0 \in \Theta$ such that $g(x, \theta_0) = 0$ for a given $x$, where $g(x, \theta)$ is differentiable at $x, \theta_0$. Let the sequence of experiments $\mathcal{E}_n = \{P^n_\theta, \theta \in \Theta\}$ satisfy local asymptotic normality with nonsingular information matrix $I_{\theta_0}$. Consider a sequence of treatment assignment rules $\delta(x, z_n)$ in the experiments $\mathcal{E}_n$, and let

$$\pi_n(x, h) = E_h[\delta_n(x, Z_n)].$$

Suppose $\pi_n(x, h) \to \pi(x, h)$ for every $h$. Then there exists a function $\delta(x, z)$ such that

$$\pi(x, h) = E_h[\delta(x, Z)] = \int \delta(x, z)dN(z|h, I^{-1}_{\theta_0}),$$

where $dN(z|h, I^{-1}_{\theta_0})/dz$ is the pdf of a multivariate normal distribution with mean $h$ and variance $I^{-1}_{\theta_0}$.

Proposition 1 shows that the simple multivariate normal shift experiment can be used to study the asymptotic behavior of treatment rules in parametric models. In particular, any asymptotic distribution of a sequence of treatment rules can be expressed as the (exact) distribution of a treatment rule in a simple Gaussian model with sample size one.

Before we consider the Gaussian limit experiment in detail, it is useful to examine the limiting behavior of the loss and risks functions, to provide heuristic guidance on the relevant forms of the loss functions in the limit experiment. All three loss functions we consider can be written in the form

$$L(\delta, \theta, x) = L(0, \theta, x) + \delta[L(1, \theta, x) - L(0, \theta, x)].$$

This linearity in $\delta$ makes it possible to define the asymptotic risk functions to have essentially the same form as the original risk functions.

For Loss A, $K_0 - K_1$ loss, and an estimator sequence $\delta_n$ with $\pi_n(h, x) \to \pi(x, h)$, the associated

---

then a sufficient condition for local asymptotic normality is differentiability in quadratic mean of the $P_\theta$. 

---
risk function can be written in terms of the local parameter \( h \) as

\[
R_n^A(\delta, h, x) = E_h \left[ L^A(\delta(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x) \right] = E_h \left[ L^A(0, \theta_0 + \frac{h}{\sqrt{n}}, x) + \delta(Z_n) \left( L^A(1, \theta_0 + \frac{h}{\sqrt{n}}, x) - L^A(0, \theta_0 + \frac{h}{\sqrt{n}}, x) \right) \right]
\]

By the differentiable continuity of \( g \) in \( \theta \),

\[
\lim_{n \to \infty} 1(g(x, \theta_0 + \frac{h}{\sqrt{n}} > 0)) = 1(\dot{g}'h > 0)
\]

and

\[
\lim_{n \to \infty} 1(g(x, \theta_0 + \frac{h}{\sqrt{n}} < 0)) = 1(\dot{g}'h < 0).
\]

The case \( \dot{g}'h = 0 \) presents a complication for taking limits as above, but since the loss function is bounded below by 0, we can express a lower bound on limiting risk as

\[
\lim \inf_{n \to \infty} R_n^A(\delta, h, x) \geq K_0 \cdot 1(\dot{g}'h > 0) + \pi(x, h) \left[ K_1 \cdot 1(\dot{g}'h < 0) - K_0 \cdot 1(\dot{g}'h > 0) \right].
\]

This is the risk function for the same loss, but using \( \dot{g}'h \) in place of \( g(x, \theta_0 + h/\sqrt{n}) \), and modified to be 0 when \( \dot{g}'h = 0 \). This suggests that analyzing the Gaussian shift limit experiment, with this modified version of the loss function, will yield asymptotic minmax bounds for the original treatment assignment problem.

For Loss B and C, the net welfare term \( g(x, \theta_0 + \frac{h}{\sqrt{n}}) \to g(x, \theta_0) = 0 \), so it is natural to renormalize the risk by a factor of \( \sqrt{n} \) to keep the limiting risk nondegenerate. The behavior of the loss at \( \dot{g}'h = 0 \) does not create a problem in these cases, and by simple calculations we have for Loss B:

\[
\lim_{n \to \infty} \sqrt{n} \cdot E_h \left[ L^B \left( \delta(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right] = - (\dot{g}'h) \cdot \pi(x, h).
\]

For Loss C, we have

\[
\lim_{n \to \infty} \sqrt{n} \cdot E_h \left[ L^C \left( \delta(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right] = (\dot{g}'h) \left[ 1(\dot{g}'h > 0) - \pi(x, h) \right].
\]

The forms of the limiting risk functions are the same as the original ones, except with \( g(x, \theta_0 + h/\sqrt{n}) \) replaced by \( \dot{g}'h \). This is intuitive: if \( \dot{g}'h > 0 \), then for sufficiently large \( n \), the treatment effect \( g(x, \theta_0 + h/\sqrt{n}) \) will be positive, and likewise if \( \dot{g}'h < 0 \) the treatment effect will eventually be negative. Combined with Proposition 1, this suggests that we can study a simplified version of the original treatment assignment problem, in which the only data is a single draw from a multivariate normal distribution with unknown mean, and the treatment effect of interest is a simple linear function of the mean.
4 Minmax Treatment Rules in Gaussian Shift Experiments, and Asymptotic Minmax Rules

We have argued that the original sequence of treatment assignment problems can be approximated by an analogous treatment assignment problem in the simple Gaussian shift model. In this section, we consider the Gaussian shift experiment, and derive an essential complete class theorem and exact solutions to the minmax problem for the three loss functions. This leads to a local asymptotic minmax theorem for treatment assignment rules.

Suppose that \( Z \sim N(h, I_{\theta_0}^{-1}) \), \( h \in \mathbb{R}^k \). Let \( \hat{g} \in \mathbb{R}^k \) satisfy \( \hat{g}'I_{\theta_0}^{-1}\hat{g} > 0 \). We are interesting in deciding whether \( \hat{g}'h \) is positive or negative. (Here, \( \hat{g} \) corresponds to the derivative of the function \( g(x, \theta_0) \) in the original sequence of experiments, but the results for the multivariate normal shift experiment simply treat it as a given constant vector.) The action space is \( A = \{0, 1\} \), and a randomized decision rule \( \delta \) is a map from \( \mathbb{R}^k \rightarrow [0, 1] \) with the interpretation that \( \delta(z) = Pr(T = 1|Z = z) \).

The following result says that rules of the form \( \delta_c = 1(\hat{g}' Z > c) \) form an essential complete class. This greatly simplifies the problem of finding minmax rules and bounds in the multivariate normal limit experiment, because we can limit our attention to the essential complete class rather than have to search over all possible decision rules.

**Proposition 2** Let the loss \( L(h, a) \) satisfy:

\[
[L(h, 1) - L(h, 0)] (\hat{g}'h) < 0
\]

for all \( h \) such that \( \hat{g}'h \neq 0 \), and \( L(h, 1) = L(h, 0) = 0 \) when \( \hat{g}'h = 0 \). For any randomized decision rule \( \tilde{\delta}(z) \), there exists a rule of the form

\[
\delta_c(z) = 1(\hat{g}' z > c)
\]

which is at least as good as \( \tilde{\delta} \).

**Proof:** see Appendix.

This result is modelled after the essential complete class theorem of Karlin and Rubin (1956), which applies to models with a scalar parameter satisfying the monotone likelihood ratio property. (See Schervish 1995, Theorem 4.68, p.244.) However, their proof is modified to make use of the properties of one-sided tests in the multivariate normal model, in particular the uniformly most powerful property shown in van der Vaart 1998, Proposition 15.2.

We now turn to the minmax problem in the limit experiment. We want to calculate the minmax risk, and a corresponding minmax rule (in the class \( \{\delta_c\} \)), under the three loss functions we are
working with. All of the risk functions turn out to be linear in \(E_h\delta\), so the following expression will play a key role in our risk computations.

\[
E_h(\delta, h) = \Pr_h\left(\frac{\gamma'(Z - h)}{\sqrt{\gamma' I_{\theta_0}^{-1} \gamma}} > \frac{c - \gamma' h}{\sqrt{\gamma' I_{\theta_0}^{-1} \gamma}}\right) = 1 - \Phi\left(\frac{c - \gamma' h}{\sqrt{\gamma' I_{\theta_0}^{-1} \gamma}}\right)
\]

For loss A in the limit experiment, the appropriate risk function is:

\[
R_A(\delta, h, x) = 1(\gamma' h > 0)[1 - E_h(\delta)]K_0 + 1(\gamma' h < 0)E_h(\delta)K_1.
\]

where \(K_0, K_1 > 0\). The limit experiment risk for losses B and C are as described before:

\[
R_B(\delta, h, x) = -(\gamma' h)E_h(\delta); \quad R_C(\delta, h, x) = (\gamma' h)[1(\gamma' h > 0) - E_h(\delta)].
\]

**Proposition 3** Suppose \(Z \sim N(h, I^{-1}_\theta)\) for \(h \in \mathbb{R}^k\). In each case below, the infimum is taken over all possible randomized decision rules, and \(\delta^*\) denotes a rule which attains the given bound:

(A) For Loss A,

\[
\inf_{\delta} \sup_h R_A(\delta(Z, x), h, x) = \frac{K_0 K_1}{K_0 + K_1}, \quad \delta^* = 1(\gamma' Z > c^*), \quad c^* = \sqrt{\gamma' I_{\theta_0}^{-1} \gamma} \Phi^{-1}\left(\frac{K_1}{K_0 + K_1}\right);
\]

(B) For Loss B,

\[
\inf_{\delta} \sup_h R_B(\delta(Z, x), h, x) = 0, \quad \delta^* = 0;
\]

(C) For Loss C,

\[
\inf_{\delta} \sup_h R_C(\delta(Z, x), h, x) = \tau^* \Phi(\tau^*)\sqrt{\gamma' I_{\theta_0}^{-1} \gamma}, \quad \tau^* = \arg \max_{\tau} \tau \Phi(-\tau), \quad \delta^* = 1(\gamma' Z > 0).
\]

**Proof:** see Appendix.

Loss A is well known from hypothesis testing theory, and the same bound is derived in the scalar case in Berger (1985), Section 5.3.2, Example 14. Our proof adapts Berger’s analysis for the multivariate normal case. The minmax results for Loss B and C appear to be new.

Minmax analysis of Loss B leads to excessively conservative rules; similar results have been noted
elsewhere and can be viewed as a motivation for the minmax regret approach, which corresponds to Loss C. Thus, in the remainder of the paper we focus on Loss A and Loss C.

Since the multivariate shift model provides an asymptotic version of the original problem, in the sense that any sequence of decision rules in the original problem with limit distributions is matched by a decision rule in the limit experiment, an application of the Portmanteau lemma allows us to use the exact bounds developed in Proposition 3 as asymptotic bounds in the original problem. This leads to the following result, which is our main finding in the paper:

**Theorem 1** Assume the conditions of Proposition 1, and suppose that $\delta_n$ is any sequence of treatment assignment rules that converge to limit distributions under $\theta_0 + \frac{h}{\sqrt{n}}$ for every $h \in \mathbb{R}^k$. Then, for Loss A,

$$
\liminf_{n \to \infty} \sup_{h \in \mathbb{R}^k} E_h \left[ L^A(\delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x) \right] \geq \frac{K_0 K_1}{K_0 + K_1},
$$

and the bound is attained by the decision rules

$$
\delta^*_n = 1(g(x, \hat{\theta}) > c^*), \quad c^* = \left( \sqrt{\hat{g}' I^{-1}_{\theta_0} \hat{g}} \right) \Phi^{-1} \left( \frac{K_1}{K_0 + K_1} \right),
$$

where $\hat{\theta}$ is an estimator sequence that satisfies $\sqrt{n}(\hat{\theta} - \theta_0 - \frac{h}{\sqrt{n}}) \overset{d}{\to} N(0, I_{\theta_0}^{-1})$ for every $h$.

For Loss C,

$$
\liminf_{n \to \infty} \sup_{h \in \mathbb{R}^k} \sqrt{n} \cdot E_n \left[ L^C(\delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x) \right] \geq \tau^* \Phi(\tau^*) \sqrt{\hat{g}' I_{\theta_0}^{-1} \hat{g}},
$$

where $\tau^* = \arg \max_{\tau} \tau \Phi(-\tau)$. The bound is attained by the rule $\delta^*_n = 1(g(x, \hat{\theta}) > 0)$

**Proof:** see Appendix.

For Loss A, note that if $K_0 = K_1$, then $c^* = 0$ so that the optimal rule is the same as for Loss C. In particular, the maximum likelihood “plug-in” approach yields decision rules which are optimal in this local asymptotic minmax risk sense. Although perhaps not surprising, this rule is distinct from the usual hypothesis testing approach, which would require that the estimated net effect be above some strictly positive cutoff determined by the level specified for the test.

### 5 Regular Semiparametric Models (Preliminary)

In this section, we consider an extension of the results for the parametric case, to semiparametric models where the parameter vector may have an infinite-dimensional nuisance component. As we have seen in the previous section, asymptotic minmax properties of certain treatment assignment
rules can be related to uniformly most powerful one-sided hypothesis tests. In semiparametric settings, results on optimal testing were obtained by Choi, Hall, and Schick (1996). We use essentially their framework, and show how to extend their results to obtain asymptotically minmax treatment assignment rules.

Let the data \( Z_n \) come from probability measures \( P_n, \theta \in \Theta \), where \( \theta = (\gamma, \eta) \). Here \( \gamma \) is assumed scalar, and \( \eta \) is possibly infinite dimensional, lying in a subset of a Hilbert space.

We fix \( \theta_0 = (0, \eta) \), and consider local alternatives \( \theta_n(h) = (\gamma_n(h), \eta_n(h)) \), where

\[
gamma_n(h) = 0 + \frac{h_{\gamma}}{\sqrt{n}},
\]

\[
\eta_n(h) = \eta + \frac{h_{\eta}}{\sqrt{n}},
\]

Let \( h = (h_{\gamma}, h_{\eta}) \). We assume \( h_{\gamma} \in \mathcal{H}_\gamma \), where \( \mathcal{H}_\gamma \) contains a neighborhood of 0, and \( h_{\eta} \in \mathcal{H}_\eta \), where \( \mathcal{H}_\eta \) is a Hilbert space.

Let the log likelihood process be

\[
L_n(h) = \log \frac{dP_{n, \theta_n(h)}}{dP_{n, \theta_0}} = S_n h - \frac{1}{2} \sigma^2(h) + r_n(h),
\]

(1)

where, under \( \theta_0 \), \( S_n = (S_n, S_m) \) is a random linear function that does not involve \( h \), which is asymptotically Gaussian with mean 0 and variance \( \sigma^2(h) = \langle \tilde{h}, B\tilde{h} \rangle \), where \( B \) is a positive definite self-adjoint bounded linear operator, and \( r_n(h) \xrightarrow{p} 0 \) for all \( h \) under \( \theta_0 \).

Partition \( B \) into \( (B_{ij}) \), for \( i, j = 1, 2 \), where \( B_{11} \) has dimension 1 and \( B_{22} \) has a bounded inverse \( B_{22}^{-1} \). Let

\[
B^* = B_{11} - B_{12}B_{22}^{-1}B_{21},
\]

and let the random operator \( S_n^* \) be defined by

\[
S_n^* a = S_n \begin{pmatrix} a \\ -B_{22}^{-1}B_{21} a \end{pmatrix} = S_n, a - S_m B_{22}^{-1} B_{21} a,
\]

for all \( a \in \mathbb{R} \). Let

\[
\xi_n(h) = (B^*)^{-1/2} S_n^*.
\]

Note that this depends on \( \eta \) which is part of \( \theta_0 = (0, \eta) \).

An efficient test statistic is a statistic (sequence) \( T_n \) s.t.

\[
T_n - \xi_n(h) \xrightarrow{p} 0,
\]

under \( \theta_0 = (0, \eta) \) for every \( \eta \). Construction of \( T_n \) for specific examples is discussed in Section 7.
of Choi, Hall, and Schick (1996). They show that one-sided tests based on $T_n$ are asymptotically uniformly most powerful. However, they impose the regularity condition that tests have the same level for all local nuisance parameters $h_\eta$. This makes it difficult to exactly use our previous argument for the parametric case. Our approach is to “slice” the problem by values of $h_\eta$ and derive the minmax rule for each slice. The same rule is minmax in every case, so we can then argue that it is minmax for the overall problem.

**Theorem 2** Assume that the experiments $E_n$ satisfy equation (1), with $B_{22}^{-1}$ having bounded inverse $B_{22}^{1}$. Suppose that $\delta_n$ is a sequence of treatment assignment rules such that

$$\pi_n(x, h) = E_h[\delta_n(x, Z_n)] \to \pi(x, h)$$

for every $h$. Suppose that an efficient test statistic $T_n$ exists. Then, for regret loss $C$,

$$\liminf_{n \to \infty} \sup_{h} \sqrt{n} \cdot E_h \left[ L^C \left( \delta(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right] \geq \liminf_{n \to \infty} \sup_{h} \sqrt{n} \cdot E_h \left[ L^C \left( 1(T_n > 0), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right].$$

6 Conclusion

Le Cam’s asymptotic statistical decision theory framework proposes to examine “limiting” versions of the original statistical decision problem, obtain optimal decision rules in the limit, and then find a sequence of decision rules in the original sequence of problems that matches the optimal rule asymptotically. This general approach has been used primarily for obtaining optimality results for point estimation and hypothesis testing, but it can be applied to a much wider range of statistical decision problems, such as the treatment assignment problem considered here.

We have examined asymptotic properties of treatment assignment rules in parametric and regular semiparametric settings, under standard regularity conditions. The limiting version of the decision problem is a treatment assignment problem involving a single observation from a Gaussian shift. Using simple extensions of classic results from the theory of one-sided tests, it turns out to be possible to exactly solve for minmax rules in this simple setting. This leads to local asymptotic minmax bounds on risk in the original sequence of models. Our sharpest results are for the social welfare regret loss (loss $C$), which has been emphasized by Manski (2004). We find that a plug-in rule, which uses an efficient estimator of the treatment effect, is locally asymptotically minmax. This rule is intuitive, but does lead to less conservative treatment assignment than typical applications of hypothesis testing, which would suggest to apply the treatment only if an estimator of the treatment effect was above some strictly positive cutoff.
7 Appendix: Proofs

Proof of Proposition 2:
For a decision rule $\delta$, the risk function is

$$R(h, \delta) = \int [\delta(z)L(h, 1) - (1 - \delta(z))L(h, 0)] f(z|h)dz$$

$$= L(h, 0) - [L(h, 1) - L(h, 0)] \int \delta(z)f(z|h)dz$$

So, for any two tests $\delta_1, \delta_2$,

$$R(h, \delta_1) - R(h, \delta_2) = [L(h, 1) - L(h, 0)] \int [\delta_1(z) - \delta_2(z)] f(z|h)dz$$

$$= [L(h, 1) - L(h, 0)] \{E_h[\delta_1(Z)] - E_h[\delta_2(Z)]\}.$$

Let $\tilde{\delta}$ be an arbitrary treatment assignment rule. We want to show that there is some $c$ such that

$$R(h, \tilde{\delta}) - R(h, \delta_c) \geq 0$$

for all $h$.

For $h = 0$, let $c$ satisfy

$$E_0[\delta_c(Z)] = E_0[\tilde{\delta}(Z)].$$

Note that $\bar{g}Z \sim N(0, \bar{g}^{-1}I_0 \bar{g})$, so

$$E_0[\delta_c(Z)] = Pr_0(\bar{g}Z > c)$$

$$= 1 - \Phi \left( \frac{c}{\sqrt{\bar{g}^{-1}I_0 \bar{g}}} \right).$$

It is easy to see that for any $\tilde{\delta}$, we can choose a $c$ to satisfy the requirement above.

By construction, $\delta_c$ has the same risk as $\tilde{\delta}$ when $\bar{g}h = 0$. Consider values of $h$ such that $\bar{g}h > 0$. By van der Vaart, Proposition 15.2, $\delta_c$ is uniformly most powerful for testing $H_0 : \bar{g}h = 0$ against $H_1 : \bar{g}h > 0$. This means that for all $h$ such that $\bar{g}h > 0$,

$$E_h[\delta_c(Z)] \geq E_h[\tilde{\delta}(Z)].$$

Thus,

$$R(h, \tilde{\delta}) - R(h, \delta_c) = [L(h, 1) - L(h, 0)] \left\{ E_h[\tilde{\delta}(Z)] - E_h[\delta_c(Z)] \right\}$$

$$\geq 0.$$
Next, consider values of $h$ such that $\dot{g}'h < 0$. Let $k = -\dot{g}$. Then again using van der Vaart, Proposition 15.2, the rule $\delta'_c = 1(k'h > 0)$ is UMP for testing $H_0 : k'h = 0$ vs $H_1 : k'h > 0$, so that $E_h[\delta'_c] \geq E_h[\tilde{\delta}]$. Noting that $\delta_c = 1 - \delta'_c$, we have $E_h[\delta'_c] = 1 - E_h[\delta_c]$, so that $E_h[\delta_c] \leq E_h[\tilde{\delta}]$. □

Proof of Proposition 3:

(A) Take a rule $\delta_c$, we want $\sup_h R^A(\delta_c, h, x)$.

$$
\sup_{h: \dot{g}'h > 0} R^A(\delta_c, h, x) = \sup_{h: \dot{g}'h > 0} [1 - E_h(\delta_c)]K_0 = \sup_{h: \dot{g}'h > 0} K_0 \Phi \left( \frac{c - \dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}} \right) = K_0 \Phi \left( \frac{c}{\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}} \right)
$$

$$
\sup_{h: \dot{g}'h < 0} R^A(\delta_c, h, x) = \sup_{h: \dot{g}'h < 0} E_h(\delta_c)K_1 = \sup_{h: \dot{g}'h < 0} K_1 \left[ 1 - \Phi \left( \frac{c - \dot{g}'h}{\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}} \right) \right] = K_1 \left[ 1 - \Phi \left( \frac{c}{\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}} \right) \right]
$$

For $h$ such that $\dot{g}'h = 0$, $R^A(\delta_c, h, x) = 0$. Hence,

$$
\sup_h R^A(\delta_c, h, x) = \max \left\{ K_0 \Phi \left( \frac{c}{\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}} \right) , K_1 \left[ 1 - \Phi \left( \frac{c}{\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}} \right) \right] \right\}.
$$

Then, $\inf_c \sup_h R^A(\delta_c, h, x)$ occurs when $c$ is chosen to set $\Phi(c/\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}) = K_1/(K_0 + K_1)$. By Proposition 2, it suffices to consider only the class of decision rules of the form $\delta_c$. Plugging in this minmax rule, the minmax value in the conclusion follows.

(B) For risk $R^B$, consider two cases:

(i) $\dot{g}'h = 0$ for all $h$. Then, $R^B(0, \delta, h) = 0$ for all $\delta$ and $h$. Hence every rule is minimax and the conclusion follows.

(ii) Suppose there exists $\tilde{h}$ such that $\dot{g}'\tilde{h} < 0$. Consider $\delta_c$ with $c = \infty$, ie $\delta = 0$. Then $R^B(0, h, x) = 0$ for all $h$. Take any other rule $\delta_c$ and note that $R^B(\delta_c, \tilde{h}, x) = -(\dot{g}'\tilde{h})(E_h \delta_c) \geq 0$. Hence $0 = \sup_h R^B(0, h, x) \leq R^B(\delta_c, \tilde{h}, x) \leq \sup_h R^B(\delta_c, h, x)$. Then $\delta_\infty = 0$ is the minimax rule and the result follows.

(Note that in general there could be a third case, where $\dot{g}'h \geq 0$ for all $h$, and $\dot{g}'h > 0$ for some $h$. However, since we are considering $h \in \mathcal{R}^k$, and assume $\dot{g}$ is nonzero (contrast with (i)), this case can be ruled out.)

(C) Consider $\delta_c$ with $c \geq 0$. Fix $\tilde{h}$ such that $\dot{g}'\tilde{h} > 0$. Then

$$
R^C(\delta_c, \tilde{h}, x) = (\dot{g}'\tilde{h}) \Phi \left( \frac{c - \dot{g}'\tilde{h}}{\sqrt{\dot{g}'I_{\theta_0}^{-1}} \dot{g}} \right)
$$

17
\[
\begin{align*}
&\geq (g'\tilde{h})\Phi \left( -\frac{c - g'\tilde{h}}{\sqrt{g'I_{\theta_0}^{-1}g}} \right) \\
&= (g'\tilde{h}) \left[ 1 - \Phi \left( \frac{c + g'\tilde{h}}{\sqrt{g'I_{\theta_0}^{-1}g}} \right) \right] \\
&= R^C(\delta_c, -\tilde{h}, x).
\end{align*}
\]

So, \(\sup_h R^C(\delta_c, h, x) = \sup_{h: g'h > 0} R^C(\delta_c, h, x)\).

Further, take \(c > 0\),

\[
R^C(\delta_0, h, x) = (g'\tilde{h})\Phi \left( -\frac{g'\tilde{h}}{\sqrt{g'I_{\theta_0}^{-1}g}} \right) < (g'\tilde{h})\Phi \left( \frac{c - g'\tilde{h}}{\sqrt{g'I_{\theta_0}^{-1}g}} \right) = R^C(\delta_c, \tilde{h}, x)
\]

Also, \(R^C(\delta_0, 0, x) = 0 = R^C(\delta_c, 0, x)\). So,

\[
\sup_h R^C(\delta_0, h, x) = \sup_{h: g'h > 0} R^C(\delta_0, h, x) \leq \sup_{h: g'h > 0} R^C(\delta_c, h, x) = \sup_h R^C(\delta_c, h, x),
\]

which shows that \(\delta_0\) is minmax over all rules \(\delta_c\) with \(c \geq 0\). The analogous argument for \(c \leq 0\) yields \(\delta_0\) as the minmax rule.

\(\square\)

**Proof of Theorem 1:**
Consider Loss A. Let \(Q_h\) denote the limit distributions of \(\delta_n\) under \(h\). Note that \(L^A\) is lower semicontinuous, so by the Portmanteau Lemma (e.g. van der Vaart and Wellner, 1996, Theorem 1.3.4),

\[
\liminf_{n \to \infty} \sup_{h \in \mathbb{R}^k} E_h \left[ L^A \left( \delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right] \geq \sup_{h \in \mathbb{R}^k} \int L(a, h, x) dQ_h(a).
\]

By Proposition 1, \(Q_h\) is the distribution of a randomized decision rule in the \(N(h, I_{\theta_0}^{-1})\) limit experiment. Hence the right hand side of the previous expression is bounded below by

\[
\inf_T \sup_{h \in \mathbb{R}^k} \int L^A( T(z), h, x) dN(z|h, I_{\theta_0}^{-1}),
\]

where the infimum is taken over all randomized decision rules \(T\). Proposition 3 provides a lower bound on the maximum risk of a slightly modified version of Loss A, call it \(\bar{L}^A\), which satisfies \(L^A \geq \bar{L}^A\). Thus the previous express is greater than

\[
\inf_T \sup_{h \in \mathbb{R}^k} \int \bar{L}^A( T(z), h, x) dN(z|h, I_{\theta_0}^{-1}) = \frac{K_0K_1}{K_0 + K_1},
\]

18
giving the local asymptotic minmax bound as stated. It is straightforward to verify that the rule \( \delta^*_n = 1(g(x, \hat{\theta}) > c^*) \) achieves this bound.

For Loss C, consider

\[
\liminf_{n \to \infty} \sup_{h \in \mathbb{R}^k} \sqrt{n} \cdot E_n \left[ L^C \left( \delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right].
\]

Note that

\[
\liminf_{n \to \infty} \sqrt{n} L^C(a_n, \theta_0 + \frac{h}{\sqrt{n}}, x) \geq L^C(a, h, x)
\]

for every sequence \( a_n \to a \). We can use this fact, and modify the proof of the Portmanteau lemma to obtain that

\[
\liminf_{n \to \infty} \sup_{h \in \mathbb{R}^k} \sqrt{n} \cdot E_n \left[ L^C \left( \delta_n(Z_n), \theta_0 + \frac{h}{\sqrt{n}}, x \right) \right] \geq \sup_{h \in \mathbb{R}^k} \int L^C (a, h, x) dQ_h(a).
\]

The remainder of the proof is analogous to the case for Loss A.
References


