Efficiency in Asymptotic Shift Experiments*

Keisuke Hirano
Department of Economics
University of Miami
khirano@miami.edu

Jack R. Porter
Department of Economics
Harvard University
jporter@harvard.edu

August 25, 2003

Abstract

This paper generalizes classical results on efficient point estimation in parametric models. The asymptotic optimality of maximum likelihood estimators is, generally, tied to both the regularity of the statistical model and symmetry of the loss function in the efficiency criterion. We examine the possibility of locally shifting or "bias-correcting" the MLE to achieve efficiency. In parametric models which asymptotically have a full shift form, including local asymptotic normal models, a feasible locally shifted MLE is optimal.

*We are grateful to Gary Chamberlain, Thomas Rothenberg, Peter Bickel, and numerous seminar participants for comments. We thank the National Science Foundation for research support under grants SES-9985257 (Hirano) and SES-0112095 (Porter).

1 Introduction

Our goal in this paper is to generalize standard results on efficient point estimation in parametric models, and develop modified versions of the maximum likelihood estimator (MLE) that are efficient under more general conditions. In regular parametric models, it is well known that the MLE is asymptotically efficient.[1] Efficiency can be defined in a number of different ways, but conventional results are based, at least implicitly, on a symmetric loss function. If the underlying model is not regular, or if the loss function is not symmetric, then the simple MLE will not generally be optimal.

It is known that Bayes estimators are asymptotically efficient in a number of situations where the optimality of MLE fails ((Ibragimov and Hasminskii 1981), (Strasser 1982), and (Hirano and Porter 2002)). In some specific cases, a shifted or bias-corrected version of MLE is also known to be efficient. In regular models with asymmetric loss, Cavanagh, Jones, and Rothenberg (1990) showed that a locally shifted MLE is efficient among asymptotically normal estimators. In certain nonregular models, a shifted version of MLE is known to be equivalent to the efficient Bayes estimator. In this paper, we show that this phenomenon holds for a fairly wide set of models. Specifically, for models which asymptotically have a shift form (including all regular models), we show that a locally shifted version of MLE is asymptotically efficient under asymmetric loss. Here, asymptotic efficiency can be defined as an estimator achieving asymptotic minimal risk among all regular estimators, or having local asymptotic minmax risk among all estimators. We also show how to construct the optimal shifted MLE based on the loss function and the features of the model.

We make use of Le Cam's limits of experiments theory to develop these results. Recent expositions of the limits of experiments theory include Van der Vaart (1991) and Van der Vaart (1998). In this approach, we approximate the model of interest by a simpler model, called the limit experiment. For any estimator sequence in the original model, its asymptotic distribution is equal to the exact distribution of *some* estimator in the limit experiment. So the limit experiment provides a characterization of the whole class of attainable limit distributions of estimators in the original model. A key advantage of this approach is that the limit experiment often has a quite simple form, for which optimality can be studied straightforwardly. We find optimal estimators in the limit experiment, and then construct estimators in the original model that asymptotically "match" these optimal estimators.

We consider cases where the limit experiments are full shift models, in which the observation has a distribution known up to an additive shift. This holds for all regular models, and some interesting nonregular models that have arisen in recent applications, including some search and

[1] Throughout the paper, we follow conventional statistical usage of the term "regular." When referring to models, regular means locally asymptotically normal. When referring to estimators or statistics, regular means asymptotically equivariant-in-law.

auction models as well as certain explosive growth models. One of our main illustrative examples is a procurement auction model from Paarsch (1992).[2]

Other authors have emphasized the usefulness of full shift models, and more general shift-invariant models, in asymptotic statistics. Le Cam ((1986), Chapter 8.4) shows that, under a continuity condition, every limit experiment is expressable as a shift-invariant experiment at almost every parameter value. Strasser (1982) shows that for a general class of models which are asymptotically shift-invariant, Bayes-type estimators are typically locally asymptotically minmax. Van der Vaart (1996) examines the more specialized case of full shift models and obtains convolution theorems and local asymptotic minmax bounds.

In this paper, we emphasize the relationship between asymptotic shift models and maximum likelihood estimation. In full shift models, the MLE is equivariant, and any other equivariant estimator can be obtained by adding a constant to the MLE. Since the risk is constant for any equivariant estimator in the shift model, we can easily find the additive shift to the MLE which minimizes risk. Thus an appropriately shifted MLE is the best equivariant estimator, and by the Hunt-Stein theorem, this estimator is also minmax among all estimators. We then construct a locally shifted version of MLE in the original model, which converges to the optimally shifted MLE in the limit experiment and is asymptotically best equivariant, and locally asymptotically minmax.

We also extend our results to conditional shift experiments, where the observation has an additive shift, and its distribution may also depend on another observed random variable. A special case is the limit experiment from locally asymptotically mixed normal (LAMN) models, which arise in some nonstationary time series settings. A similar equivariance argument works in the conditional shift case, except that the minmax shift may need to be conditioned upon the other observed random variable.

In the next section, we briefly review the classical efficiency theory for regular parametric models, and discuss the role of regularity (local asymptotic normality) and symmetric loss functions in the classical theory. Section 3 considers the full shift model, and shows that shifted MLE (and Bayes) is best equivariant and minmax. Section 4 connects the results in the earlier sections to asymptotic theory for a large class of models, using the limits of experiments theory. It provides asymptotic theory for the MLE and shows how to construct the optimal shifted MLE. Our expression for the optimal shift depends only on the likelihood, the loss, and the MLE itself. We give some examples of how to construct the optimal shift in specific applications. An extension of these results to

[2] Not all structural auction models are asymptotically of the full shift form. Donald and Paarsch (1993) and Chernozhukov and Hong (2001) discuss MLE, Bayes, and related estimators in a larger class of nonregular statistical models that includes many parametric auction and search models. Hirano and Porter (2002) show that the limit experiments for some of these models are of a generalized shift form. For other special cases, in particular the continuous covariates case, the limit experiment is apparently unknown.

conditional shift experiments is given in Section 5. Throughout, we use high-level assumptions that allow for broad application of our results and a simpler exposition. Section 6 concludes. All proofs are contained in the Appendix.

2 Classical Efficiency Results for Maximum Likelihood and the LAN Limit Experiment

Hájek (1970, 1972) established the conventional criteria by which MLE is asymptotically efficient in regular models. In a regular model with parameter θ, the limiting distribution of the MLE is given by $\sqrt{n}(\hat{\theta}_{ML} - \theta) \rightsquigarrow N(0, I_\theta^{-1})$, where I_θ is the Fisher information matrix, and $\rightsquigarrow$ denotes weak convergence (convergence in distribution).

Hájek's (1970) convolution theorem shows that the limiting distribution of any regular estimator T_n can be written as a convolution of $N(0, I_\theta^{-1})$ and "noise". Specifically, there exists a probability measure M_θ such that

$$\sqrt{n}(T_n - \theta) \rightsquigarrow N(0, I_\theta^{-1}) * M_\theta.$$

It follows that the variance of the limiting distribution of T_n is bounded below by I_θ^{-1}, the asymptotic variance of the MLE. Variances only capture one aspect of these limiting distributions. More generally, efficiency reflects the notion that one estimator is more concentrated around θ than another. In that spirit, a loss function L is introduced and risk comparisons can be made. According to Anderson's Lemma, for any loss function L which is "bowl-shaped" (symmetric, with closed convex upper level sets), the following inequality holds:

$$\int L d[N(0, I_\theta^{-1}) * M_\theta] \geq \int L dN(0, I_\theta^{-1}). \tag{1}$$

This implies that for bowl-shaped loss functions, the MLE will have lower asymptotic risk than any other regular estimator.

A second efficiency criteria, local asymptotic minimaxity, is given in Hájek (1972). For *any* estimator T_n and bowl-shaped L,

$$\sup_c \liminf_{n \longrightarrow \infty} \sup_{\|h\| \leq c, \theta + h/\sqrt{n} \in \Theta} E_{\theta + h/\sqrt{n}} L(\sqrt{n}(T_n - (\theta + h/\sqrt{n})) \geq \int L dN(0, I_\theta^{-1}) \tag{2}$$

This result gives a bound on the maximum risk of any estimator T_n over an asymptotically shrinking neighborhood of θ. Moreover, the bound established in (2) is attained by the MLE, showing its asymptotic efficiency.

In the classic results of Hájek, it is evident that both normality and symmetry play an important role. For example, if one uses an asymmetric loss function, such as a "check-function" loss, it is no longer clear that the $N(0, I_\theta^{-1})$ distribution is the "best" distribution. One may prefer a distribution with mean different from zero, if the loss function penalizes positive and negative estimation errors differently. There are also nonregular models where the limit distribution of the MLE is nonnormal. A well known example is i.i.d. sampling from the uniform distribution on $[0, \theta]$. In that case, the limit distribution of the normalized MLE is biased, and is not optimal even for conventional symmetric loss functions.

Le Cam (1972) explains the efficiency of MLE in regular models using the concept of a limit experiment. He shows that regular (local asymptotic normal, LAN) models behave asymptotically like the experiment of observing a single draw from the simple "shift" model

$$Z = h + N(0, I_\theta^{-1}),$$

where h is an unknown location parameter, and I_θ^{-1} is known. More precisely, according to the Asymptotic Representation Theorem ((Le Cam 1972), (Van der Vaart 1991)), any limit distribution of a sequence of estimators in the original model is matched by the distribution of a (possibly randomized) estimator in the limit experiment. Thus in the LAN case, for any estimator sequence T_n with limit distributions under local alternatives $\theta + h/\sqrt{n}$, its distribution is the distribution of some randomized estimator based on $Z \sim N(h, I_\theta^{-1})$. In the normal shift model, it can be shown that any equivariant estimator is equal to Z plus a random variable. This immediately implies the Hájek convolution theorem. It can also be shown that the simple estimator $\hat{h} = Z$ is minmax, which provides intuition for the Hájek local asymptotic minmax theorem.

Le Cam's framework makes it possible to extend the classical efficiency results to other forms of loss functions, and to nonregular models. Nonregular models have limit experiments that are different from the normal shift model, but in a number of cases they are still of a simple "full shift" form. In the next section, we consider full shift experiments where the disturbance term is not necessarily normally distributed. We also allow for asymmetric loss functions, and show that it is possible to obtain simple risk bounds and show efficiency of a shifted form of the MLE and the Bayes estimator for a flat prior. Moreover, under conditions described in section 4, maximum likelihood and Bayes estimators are matched by maximum likelihood and (flat-prior) Bayes estimators in the limit experiment. Hence, for statistical models with full shift limit experiments, we are able show that locally shifted MLE and Bayes estimators are optimal according to both of Hájek's asymptotic efficiency criteria.

3 Efficient Estimation in Shift Experiments

In this section, we study full shift experiments and the behavior of shifted maximum likelihood estimators and Bayes estimators in these experiments. We show that a certain shifted maximum likelihood estimator is best equivariant and optimal in the minmax sense, and show how to construct the optimal shift. We also show that for a particular prior distribution, the flat-prior Bayes estimator is optimal. The simple shift model will be shown to be the limit experiment for a large class of statistical models, including regular parametric models. So the findings in this section will be useful for the asymptotic efficiency results in section 4.

3.1 Full Shift Experiments

A *statistical experiment* is a measurable space $(\mathcal{X}, \mathcal{B})$, together with a collection of probability measures $\{P_h : h \in H\}$ defined on it. Here, the space $(\mathcal{X}, \mathcal{B})$ is intepreted as the sample space, h is interpreted as the parameter, and H is the parameter space. The observation X is a random variable distributed according to P_h for some h. We are interested in estimating h based on the observation X.

Suppose that the sample space is $(\mathbb{R}^k, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}^k$, and that the parameter space H is equal to $\mathbb{R}^k$. Let $f(\cdot)$ be a fixed Lebesgue probability density on $\mathbb{R}^k$, and suppose that the probability measures $\{P_h : h \in \mathbb{R}^k\}$ specify that X has density $f(x - h)$. Van der Vaart (1996) calls such a statistical experiment a *full shift experiment.* It can be interpreted as specifying that $X = h + V$, where V has a fixed distribution F with density f.

Example 1 *Suppose that the distribution of X under h is $N(h, I_\theta^{-1})$, where I_θ is a symmetric, positive definite $k \times k$ matrix which is known. This experiment arises as the limit experiment in regular parametric models; see Example 3 below. The "natural" estimator $\hat{h} = X$ is unbiased for h, and is minmax for loss functions which are symmetric about zero and have convex upper level sets, such as squared error loss and absolute error loss.*

Example 2 *Suppose that the distribution of X under h has density*

$$f(x|h) = \exp\left(\frac{h - x}{\theta}\right) \mathbf{1}\{x - h \geq 0\}/\theta.$$

for a known θ. This is called the shifted exponential model*, because $X = h + \epsilon$, where ϵ is an exponentially distributed random variable with mean θ. It arises as the limit experiment corresponding to the uniform model $U[0, \theta]$, and to other nonregular models where the support of the observations depends on model parameters ((Hirano and Porter 2002)); see Examples 4 and 5.*

3.2 Decision Problem

We focus on using the observation X to estimate the unknown parameter h. A *nonrandomized estimator* is a Borel-measurable function $T : \mathcal{X} \to H$ that maps observations on the random variable X into point estimates of the parameter h. In full shift experiments, this means that T is a function from $\mathbb{R}^k$ into $\mathbb{R}^k$. In order to establish optimality results, it is useful to allow for a broader class of estimators which allows for randomization. A *randomized estimator* is a Borel-measurable function $T(X, U)$ of X and U into $\mathbb{R}^k$, where U has a uniform distribution independent of X.[3]

In order to evaluate the quality of estimators, we introduce a *loss function* $L : H \times H \to [0, \infty)$. The quantity $L(a, h)$ is interpreted as the loss arising from choosing an estimate a when the true parameter is h. We will make the following assumptions on the loss function:

Assumption 1 *(a) The loss function has the form $L(a - h)$, is continuous, and $L(0) = 0$. Also, for all $\tau \in [0, \infty)$, the sets $\{w \in \mathbb{R}^k : L(w) \leq \tau\}$ are compact;*
(b) There exist η and D_0 such that for all $D \geq D_0$, $\inf_{\|x\| \geq D^\eta} L(x) - \sup_{\|x\| \leq D} L(x) \geq 0$ with strict inequality for some $D_1 \geq D_0$.

Note we do not require that the loss is symmetric about zero, but we do require that it depend on a and h only through their difference $a - h$. The last part of (a) the assumption is used in the application of the Hunt-Stein theorem below (Proposition 2). Part (b) ensures that loss is nondecreasing in its tails and is useful for the asymptotic results of section 4.

An estimator T has *risk function*

$$\begin{aligned} R(T, h) &= E_h[L(T(X, U) - h)] \\ &= \int_{[0,1]} \int_{\mathbb{R}^k} L(T(x, u) - h) dP_h(x) d\lambda(u) \\ &= \int \int L(T(x, u) - h) f(x - h) dx du. \end{aligned}$$

The risk is the expected loss of T under h. Risk comparisons are used to determine optimality. Since the loss function is bounded from below, the risk is always well defined.

3.3 Equivariant Estimators

Before considering optimality among all estimators it is useful to first consider optimality within a smaller class of estimators, called equivariant estimators. In full shift models, all equivariant esti-

[3] U does not need to be uniform; for example, it could be replaced by any random variable that is continuously distributed and independent of X.

mators take the simple form of the observation plus some constant. Specifically, a nonrandomized estimator is *equivariant* if, for all $g, x \in \mathbb{R}^k$,

$$T(x) + g = T(x + g).$$

Let $g = -x$. Then

$$T(x) = T(0) + x.$$

It follows that any equivariant estimator must be of the form $T(X) = X + a$, where a is some constant.

The notion of equivariance extends to randomized estimators as well. A randomized estimator is *equivariant-in-law* if, for every $h \in \mathbb{R}^k$, the law of $T - h$ under h is the same:

$$L = \mathcal{L}_h(T - h), \qquad \forall h \in \mathbb{R}^k,$$

where $\mathcal{L}_h(\cdot)$ is used to denote the law of a given random variable under the probability measure with parameter h. The convolution theorem below, which is proven in Van der Vaart (1996), generalizes our previous characterization of equivariant nonrandomized estimators, and shows that every equivariant-in-law randomized estimator in a shift experiment can be written as the sum of X and a random variable independent of X.

Theorem 1 *(Convolution Theorem) Let T be an equivariant-in-law randomized estimator in a full shift experiment. Let L be the null law $L = \mathcal{L}_h(T - h)$. Then L can be written as*

$$L = \mathcal{L}(V + W)$$

where $V = X - h$ and W is a random variable independent of V.

This result implies that any equivariant-in-law estimator can be written as $T = X + W$, where W is independent of X.

3.4 Best Equivariant and Minmax Estimators

Suppose that T is an equivariant-in-law randomized estimator, so that it can be written as $T = X + W$. Let μ_W denote the marginal distribution of W. Then the risk of T can be written as

$$R(T, h) = E_h[L(T - h)] = \int \int L(x + w - h) f(x - h) dx d\mu_W(w).$$

Setting $v = x - h$, we have

$$R(T, h) = \int \int L(v + w) f(v) dv d\mu_W(w),$$

which does not depend on h. Since the risk of an equivariant-in-law estimator is a constant, it is meaningful to look for a *minimum risk* or *best equivariant-in-law* estimator, which chooses μ_W to minimize the preceding expression. If we restrict attention to nonrandomized equivariant estimators, then the risk is of the form

$$R(T, h) = \int L(v + a) f(v) dv,$$

and the best nonrandomized rule chooses $a \in \mathbb{R}^k$ to minimize this risk. We will make the following assumption:

Assumption 2 *The minimization problem*

$$\min_{a \in \mathbb{R}^k} \int L(v + a) f(v) dv$$

has a unique solution a^*.

The uniqueness of a^* is not necessary for all of the results to follow, but it will be required for the asymptotic results of Section 4. In practice, a^* will be straightforward to calculate given L and f.

The following result shows that for minimizing risk among equivariant estimators, it is enough to consider nonrandomized estimators:

Proposition 1 *Suppose that Assumptions 1(a) and 2 hold. Then the risk of the best nonrandomized equivariant estimator $T = X + a^*$ is less than or equal to the risk of any equivariant-in-law randomized estimator.*

(All proofs are in the Appendix.)

When the full shift model is a limit experiment, the class of equivariant-in-law estimators includes the limits of all regular estimators, as defined in section 4, in the limit experiment. Proposition 1 will then yield the limit distribution of the best regular estimator.

An estimator T is minmax if it minimizes the maximum risk with respect to h:

$$\sup_h R(T, h) = \inf_{\tilde{T}} \sup_h R(\tilde{T}, h).$$

For equivariant estimators, the risk does not depend on h. So the best equivariant estimator is minmax among equivariant estimators. By an application of the Hunt-Stein theorem, it is also minmax among all estimators.

Proposition 2 *Suppose we have a full shift experiment and Assumptions 1(a) and 2 hold. Then* $T = X + a^*$ *is minmax among all randomized estimators.*

3.5 Maximum Likelihood and Bayes Estimators

Next, we consider some specific estimators. For a full shift experiment, the *maximum likelihood* estimator $\hat{h}$ can be defined as

$$\hat{h} = \arg\max_{h \in \mathbb{R}^k} f(X - h).$$

The following condition is usually easy to check for the particular density f under consideration.

Assumption 3 *(a) The density* $f(v)$ *has a unique maximum* v^**;*
(b) the maximum satisfies

$$f(v^*) > \sup_{v \in G^c} f(v),$$

for every G^c *equal to the complement of an open set* G *containing* v^*.

Part (b) of this assumption assures that the maximum is well-separated and will be useful for the asymptotic results in section 4.

Comparing the definition of the MLE and the definition of v^* as the $\arg\max_v f(v)$, it is easy to see that under this assumption, the MLE will be unique and equivariant:

Proposition 3 *Assume that we have a full shift model and that Assumption 3(a) holds. Then the MLE* $\hat{h}$ *is unique almost surely and equals*

$$\hat{h} = X - v^*,$$

where v^* *is defined in Assumption 3.*

Remark: The automatic equivariance of MLE holds in other group families as well; Eaton (1989) extends the result above to general group-invariant statistical models.

Since the MLE is equivariant, a shifted version of it will be best equivariant and minmax. This result is stated in the following proposition which is an immediate consequence of Propositions 1, 2, and 3.

Proposition 4 *Suppose we have a full shift experiment, and Assumptions 1(a), 2 and 3(a) hold. Let* $a_m = a^* + v^*$*, where* a^* *is defined in Assumption 2 and* v^* *is defined in Assumption 3. Then* $\hat{h} + a_m$ *is best equivariant and minmax.*

So we have constructed a shift, based on the loss function L and the density f, that makes the maximum likelihood estimator optimal in both the minmax and the best equivariant senses. This construction of the shift will be used in Section 4, to construct a locally shifted version of the maximum likelihood estimator that is locally asymptotically minmax.

Another estimator is the *Bayes estimator*. To define the Bayes estimator, let Π be a probability measure on the parameter space. The Bayes procedure T_B minimizes the average risk with respect to Π:

$$\int R(T_B, h) d\Pi(h) \leq \int R(T, h) d\Pi(h), \qquad \forall T.$$

This definition of the Bayes estimator as minimizing ex ante average risk, rather than posterior expected loss, was emphasized by Wald (1950), and turns out to be analytically convenient. Suppose that Π has a density π with respect to Lebesgue measure. We can regard $f(x-h)\pi(h)$ as a joint density for (x, h), and use Bayes' rule to calculate the posterior density

$$p(h|x) = c \cdot f(x-h)\pi(h),$$

where c is a normalizing constant. Then, to minimize ex ante expected risk

$$\int R(T, h) d\Pi(h) = \int \int L(T(x) - h) f(x-h)\pi(h) dx dh,$$

we can interchange the order of the integrals, and minimize

$$\int L(T(x) - h) f(x-h)\pi(h) dh$$

for each x. Since the posterior expected loss differs from this last expression only by a multiplicative constant, it is clear that minimizing expected loss with respect to $p(h|x)$ will yield an equivalent estimator.

It is useful to consider *generalized Bayes estimators*, which take Π in the preceding definition to be a σ-finite measure (not necessarily a probability measure). A special case of a generalized Bayes estimator is the *flat prior Bayes estimator*, or the *Pitman estimator*, which takes Π to be Lebesgue measure on $\mathbb{R}^k$. A simple argument shows that the flat prior Bayes estimator is also equivariant.

Proposition 5 *Suppose we have a full shift experiment and Assumptions 1(a) and 2 hold. Then,* $T_B = X + a^*$ *is a flat prior Bayes estimator, and* T_B *is best equivariant and minmax.*

So the flat-prior Bayes estimator is also best equivariant and minmax in shift experiments. Other choices for the prior Π will not lead to Bayes estimators which are equivariant or optimal in the minmax sense in general, although by construction they are optimal for average risk with respect to Π. In Section 4, we will see that asymptotically the influence of the prior becomes negligible, and that Bayes estimators for fairly arbitrary choices of the prior, behave asymptotically like flat prior Bayes estimators. This will lead to the conclusion that Bayes estimators are also asymptotically optimal.

4 Asymptotic Shift Experiments and Asymptotically Efficient Estimation

The results of Section 3 show that for shift experiments, a shifted version of the MLE is best equivariant and minmax, and that the flat prior Bayes estimator is equal to the shifted MLE. These results are useful even when the experiment of interest is not a shift experiment, because many parametric experiments behave asymptotically like shift experiments, in a sense to be made precise below. Thus, intuition for the form of optimal estimators in the shift experiment carries over into statements about asymptotically optimal estimators in a broad class of statistical models. In particular, a locally shifted version of the MLE will be shown to be best regular and locally asymptotically minmax, and asymptotically equivalent to the Bayes estimator.

4.1 Basic Setup

Suppose we observe $Z_n \sim P_{n,\theta}$, for some $\theta \in \Theta$, where Θ is a compact convex subset of $\mathbb{R}^k$ and Z_n is a random variable supported on some measurable space $(\mathcal{Z}_n, \mathcal{B}_n)$. We will denote this experiment as $\mathcal{E}_n$. Here, n is interpreted as the sample size. For example, if $Y_i \overset{\text{i.i.d.}}{\sim} P_\theta$, for $i = 1, \ldots, n$, then we can take $Z_n = (Y_1, \ldots, Y_n)'$ and $P_{n,\theta}$ is the product of n copies of P_θ.

Typically, there exist a number of estimators $\hat{\theta}(Z_n)$ such that $\hat{\theta} \xrightarrow{\text{p}} \theta$ as $n \longrightarrow \infty$ under $P_{n,\theta}$. So, to make useful comparisons among estimators, we fix θ and consider "local" alternatives $\theta + \varphi_n h$, for local parameters $h \in H_n = \{h \in \mathbb{R}^k : \theta + \varphi_n h \in \Theta\}$ and an appropriate normalization sequence $\varphi_n \longrightarrow 0$. This normalization sequence is chosen to ensure that the resulting limit distributions are not degenerate; in general, φ_n can be calculated by taking the Hellinger metric between two parameter points. This is the same kind of local asymptotic parametrization that is used in asymptotic power calculations for hypothesis tests; there the h are sometimes referred to as Pitman alternatives.

For example, in regular parametric models we take $\varphi_n = 1/\sqrt{n}$, and consider the limiting behavior of

$$\sqrt{n}\left(\hat{\theta}(Z_n) - \theta - \frac{h}{\sqrt{n}}\right) = \sqrt{n}\left(\hat{\theta}(Z_n) - \theta\right) - h$$

under the sequence of measures $P_{n,\theta+h/\sqrt{n}}$. As before, we use a loss function to assess the performance of estimators. The loss function $L : \mathbb{R}^k \to \mathbb{R}_+$ defines the risk

$$\int L\left(\varphi_n^{-1}\left(\hat{\theta} - \theta - \varphi_n h\right)\right) dP_{n,\theta+\varphi_n h}.$$

An estimator $\tilde{\theta}$ is called *regular* at θ if the limiting distribution of $\varphi_n^{-1}\left(\tilde{\theta} - (\theta + \varphi_n h)\right)$, under the $P_{n,\theta+\varphi_n h}$, exists and does not depend on h. That is,

$$\varphi_n^{-1}\left(\tilde{\theta} - (\theta + \varphi_n h)\right) \overset{\theta+\varphi_n h}{\rightsquigarrow} Q_\theta^{\tilde{\theta}}$$

where $\overset{\theta}{\rightsquigarrow}$ denotes weak convergence under $P_{n,\theta}$, and the limit law of $\tilde{\theta}$ at θ, $Q_\theta^{\tilde{\theta}}$, does not depend on h. This property is the asymptotic counterpart to the equivariance-in-law notion defined in section 3.3.

In regular models, asymptotic efficiency results are often derived by a comparison of the asymptotic variances of regular estimators. Of course, the asymptotic variance is only one feature of a limit distribution. More generally, we can make risk comparisons to see which limiting distribution is more concentrated around zero. Such an asymptotic efficiency notion extends to nonregular models. The *best regular* estimator $\hat{\theta}$ is the regular estimator that minimizes asymptotic risk,

$$\int L dQ_\theta^{\hat{\theta}} = \inf_{\tilde{\theta}} \int L dQ_\theta^{\tilde{\theta}}$$

where the infimum is taken over all regular estimator sequences $\tilde{\theta}$.

An estimator $\hat{\theta}$ is *locally asymptotically minmax* if, for $I \subset \mathbb{R}^k$ finite and *all* estimator sequences T_n,[4]

$$\begin{aligned}
&\sup_I \liminf_{n\to\infty} \sup_{h\in I} \int L\left(\varphi_n^{-1}\left(\hat{\theta} - \theta - \varphi_n h\right)\right) dP_{n,\theta+\varphi_n h} \\
&= \inf_{T_n} \sup_I \liminf_{n\to\infty} \sup_{h\in I} \int L\left(\varphi_n^{-1}\left(T_n - \theta - \varphi_n h\right)\right) dP_{n,\theta+\varphi_n h}.
\end{aligned}$$

[4]Measurability of the estimators will be assumed throughout the paper. In the locally asymptotically minmax definition, this assumption could be weakened to asymptotic measurability as defined in Van der Vaart and Wellner (1996).

4.2 Limits of Experiments Theory

The likelihood ratio processes of the experiments $\mathcal{E}_n$ are defined as

$$\Lambda_{n,h_0} = \left(\frac{dP_{n,\theta+\varphi_n h}}{dP_{n,\theta+\varphi_n h_0}}(Z_n) \right)_{h \in \mathbb{R}^k}.$$

When working with Λ_{n,h_0}, we always take its distribution under the localized probability measure $P_{n,\theta+\varphi_n h_0}$.

The experiments $\mathcal{E}_n$ are said to converge weakly to an experiment $\mathcal{E} = ((\mathcal{X}, \mathcal{B}), F_{\theta,h} : h \in \mathbb{R}^k)$ if, for every finite subset $I \subset \mathbb{R}^k$, and every local parameter h_0,

$$\left(\frac{dP_{n,\theta+\varphi_n h}}{dP_{n,\theta+\varphi_n h_0}}(Z_n) \right)_{h \in I} \overset{\theta+\varphi_n h_0}{\rightsquigarrow} \left(\frac{dF_{\theta,h}}{dF_{\theta,h_0}}(X) \right)_{h \in I},$$

where X is distributed F_{θ,h_0} on $(\mathcal{X}, \mathcal{B})$. This statement says that the likelihood ratio process satisfies finite-dimensional weak convergence to the likelihood ratio for some experiment, under each h_0.

We will assume that the sequence of experiments converges to a shift experiment:

Assumption 4 *$\mathcal{E}_n$ converges weakly to $\mathcal{E} = ((\mathcal{X}, \mathcal{B}), F_{\theta,h} : h \in \mathbb{R}^k)$, where $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^k, \mathcal{B})$, and where $F_{\theta,h}$ are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^k$ with densities $f_\theta(x-h)$ for a fixed probability density function f_θ.*

Under this assumption, we can write

$$\frac{dF_{\theta,h}}{dF_{\theta,h_0}}(X) = \frac{f_\theta(X-h)}{f_\theta(X-h_0)},$$

where X is distributed as F_{h_0} and thus has density $f_\theta(\cdot - h_0)$. In general, the limit experiment will depend on the value of θ used as the centering point for the local parameters.

The following examples show that in a number of leading cases, including all "regular" (locally asymptotically normal) settings, such weak convergence to a shift experiment occurs.

Example 3 *(LAN) Parametric models which satisfy the standard regularity conditions are* locally asymptotically normal*: under $P_{n,\theta}$, the log of their likelihood ratios can be written as*

$$\log \left[\frac{dP_{n,\theta+\varphi_n h}}{dP_{n,\theta}} \right] = h'\Delta_n - \frac{1}{2} h' I_\theta h + o_p(1),$$

where I_θ is a nonrandom positive definite matrix, and $\Delta_n \rightsquigarrow N(0, I_\theta^{-1})$. We can then obtain that *for each h_0,*

$$\left(\frac{dP_{n,\theta+\varphi_n h}}{dP_{n,\theta+\varphi_n h_0}}\right)_{h\in I} \overset{\theta+\varphi_n h_0}{\rightsquigarrow} \left(\exp\left[(h-h_0)'\Delta - \frac{1}{2}(h-h_0)'I_\theta(h-h_0)\right]\right)_{h\in I}$$

where $\Delta \sim N(h_0, I_\theta^{-1})$. The expression on the right has the same distribution as

$$\left(\frac{dN(h, I_\theta^{-1})}{dN(h_0, I_\theta^{-1})}\right)_{h\in I}.$$

So for LAN models, there is a limit experiment consisting of observing $X = h + V$, where $V \sim N(0, I_\theta^{-1})$, which is the same experiment as Example 1.

Example 4 *(Uniform Distribution) Suppose that we observe $Z_n = (Y_1, \ldots, Y_n)$, where the Y_i are i.i.d. $U[0, \theta]$. For this nonregular model, the natural scaling is $\varphi_n = n^{-1}$ (rather than $n^{-1/2}$). Hirano and Porter (2002) discusses this model in detail and shows that its limit experiment is a shifted exponential, as in Example 2.*

Example 5 *(Parametric Auction Model) Paarsch (1992) develops an econometric specification for a first-price procurement auction under the independent private values paradigm. There are m bidders whose valuations c are drawn from an exponential density with mean θ. The symmetric equilibrium bid function can be shown to be $b = c + \theta/(m-1)$. Paarsch (1992) derives the density of the winning bid to be*

$$p_\theta(b) = \frac{m}{\theta}\exp\left(-\frac{m}{\theta}\left(b - \frac{\theta}{m-1}\right)\right)\mathbf{1}\left\{b \geq \frac{\theta}{m-1}\right\}.$$

The natural scaling for this nonregular model is $\varphi_n = n^{-1}$. The likelihood ratio process for a sample of n independent winning bids can then be used to derive the limit experiment.

$$\frac{dP_{n,\theta+h/n}(b_1, \ldots, b_n)}{dP_{n,\theta+h_0/n}(b_1, \ldots, b_n)} \overset{\theta+h_0/n}{\rightsquigarrow} \exp\left(\frac{m(h-h_0)}{\theta(m-1)}\right)\mathbf{1}\{X \geq h\}, \tag{3}$$

where X is distributed as $h_0 + V$, and V has an exponential distribution with mean $\frac{m-1}{m}\theta$. The expression on the right is the likelihood ratio for the experiment consisting of observing $X = h + V$, where V has an exponential distribution, which is the same experiment as in Example 2.

Convergence of experiments allows us to neatly characterize asymptotic risk bounds for the original sequence of experiments. The asymptotic representation theorem ((Van der Vaart 1991)) says that for any estimator sequence T_n in the experiments $\mathcal{E}_n$ with limit distribution Q_h under h, there exists a randomized estimator T in the limit experiment with the same law Q_h for every h. Thus, the results we developed for shift experiments can be translated into asymptotic bounds on the performance of estimators, and provide guidance on how to construct asymptotically optimal estimators.

Recall that the best equivariant-in-law estimator and the minmax estimator in the full shift experiment had risk

$$R^* = \int L(v + a^*) f_\theta(v) dv \equiv \min_a \int L(v + a) f_\theta(v) dv.$$

It follows that this risk provides the best regular and the local asymptotic minmax bounds.

Theorem 2 *Under Assumptions 1(a), 2, and 4, R^* provides the asymptotic risk bound for best regular estimators and the local asymptotic minmax bound. In particular, for any regular estimator sequence $\tilde{\theta}$ with limit law $Q_\theta^{\tilde{\theta}}$ at θ,*

$$\int L dQ_\theta^{\tilde{\theta}} \geq R^* \tag{4}$$

and for any *estimator sequence $\hat{\theta}$ and $I \subset \mathbb{R}^k$ finite,*

$$\sup_I \liminf_{n\to\infty} \sup_{h\in I} \int L\left(\varphi_n^{-1}\left(\hat{\theta} - \theta - \varphi_n h\right)\right) dP_{n,\theta+\varphi_n h} \geq R^*. \tag{5}$$

This theorem suggests that either of Hájek's asymptotic efficiency notions will lead to the same conclusions for the general class of statistical models considered here. That the asymptotic risk bound is exactly the same for both regular estimators and for all estimators in a local asymptotic minimax sense may seem at first surprising. However, the connection is made clearer through a limit experiment characterization. Regular estimators correspond to the equivariant-in-law (randomized) estimators in the limit experiment. So a best regular bound is matched by a bound for equivariant-in-law estimators in the limit experiment. When the limit experiment is of a full shift form, best equivariant estimators are also minmax by the Hunt-Stein Theorem. Finally, local asymptotic minmaxity corresponds to minmaxity in the limit experiment.

In section 3.5, we saw that Bayes and shifted ML estimators were best equivariant and minmax in shift experiments. In many cases, we could expect Bayes and ML estimators in the original statistical model to behave asymptotically like their counterparts in the shift experiments and hence attain the risk bound R^* asymptotically. This idea is studied in the next two sections.

4.3 Asymptotic Behavior of the MLE

In the previous section, we assumed only finite-dimensional weak convergence of the likelihood ratio process. This condition was enough to allow us to use the limit experiments to provide risk bounds. Under a stronger convergence of the likelihood ratio process, these risk bounds are guaranteed to be attainable, and likelihood-based estimators in the original experiment are "matched" by their counterparts in the limit experiment. As a consequence, the distribution of ML will follow fairly easily from the form of the limit experiment, and a shifted version of ML will be asymptotically efficient. In addition, we will show how to construct the appropriate shift term, leading to a feasible asymptotically optimal estimator.

In order to strengthen the marginal weak convergence, we make a standard asymptotic equicontinuity assumption.[5]

Assumption 5 *The processes $\Lambda_{n,h_0}(h)$ satisfy asymptotic equicontinuity: for every h_0, and every $\epsilon, \eta > 0$, there exists a finite partition $(T_1, \ldots, T_m)$ of $\mathbb{R}^k$ such that*

$$\limsup_{n\to\infty} P_{n,h_0}\left[\sup_j \sup_{h_s,h_t\in T_j} |\Lambda_{n,h_0}(h_s) - \Lambda_{n,h_0}(h_t)| \geq \epsilon\right] \leq \eta.^6$$

By Assumption 4 and Theorem 18.14 of Van der Vaart (1998), the asymptotic equicontinuity condition is both necessary and sufficient to conclude that the stochastic process $\Lambda_{n,h_0}(h)$, regarded as a random element of $l^\infty(\mathbb{R}^k)$ (the space of functions on $\mathbb{R}^k$ endowed with the supremum norm), converges weakly in distribution to the process $\Lambda_{h_0}(h) := f_\theta(X-h)/f_\theta(X-h_0)$ on $\mathbb{R}^k$.

Let $\hat\theta_n$ denote the maximum likelihood estimator for θ. For fixed θ, we can write

$$\hat\theta_n = \theta + \varphi_n \hat h_n,$$

so that we can equivalently focus on the ML estimator for h,

$$\hat h_n = \varphi_n^{-1}(\hat\theta_n - \theta).$$

This is simply the usual normalized form of the MLE, and we will examine its behavior under the local sequence $P_{n,\theta+\varphi_n h_0}$. Under $P_{n,\theta+\varphi_n h_0}$, the (near-) maximizer of the likelihood function will also (nearly) maximize the likelihood ratio process $\Lambda_{n,h_0}(h)$, so we can define the MLE in the following way:

[5] See Andrews (1994) for an extensive discussion of this condition. Another approach to obtaining asymptotic results for argmax estimators is developed in Knight (2000).

[6] This form of the condition assumes measurability of the process Λ_{n,h_0}. The assumption could be generalized to allow for non-measurable processes by replacing P_{n,h_0} with the corresponding outer probability measure.

Assumption 6 *$\hat{h}_n$ satisfies*

$$\Lambda_{n,h_0}(\hat{h}_n) \geq \sup_h \Lambda_{n,h_0}(h) - o_p(1).$$

Proposition 6 *Let Assumptions 3- 6 hold. Then*

$$\left\{\varphi_n^{-1}\left(\hat{\theta}-\theta\right)\right\} \overset{\theta+\varphi_n h_0}{\rightsquigarrow} X - v^*,$$

where X has density $f_\theta(x-h_0)$, and v^* is defined in Assumption 3.

We have stated these results for the MLE under very high level assumptions, because our results are intended to apply to a wide range of models. For specific models, these would be verified from more primitive conditions, or the limit distribution of the MLE would be calculated directly from the primitive conditions.

We can also construct the asymptotically optimal estimator, by locally shifting the MLE in an appropriate way. Recall that a^* and v^* can be calculated from knowledge of L and f_θ. For example, in the LAN case, f_θ is the multivariate normal density with mean 0 and variance I_θ^{-1}. In the exponential limit experiment corresponding to the $U[0,\theta]$ model, f_θ is the exponential density with mean θ.

Since θ is unknown, we cannot calculate a^* and v^* exactly. Instead we use the MLE $\hat{\theta}$ to define the local shift in analogy to the definitions of a^* and v^*. Let

$$\hat{a} = \arg\min_a \int L(v+a) f_{\hat{\theta}}(v)dv,$$

and

$$\hat{v} = \arg\max_v f_{\hat{\theta}}(v).$$

In practice these problems will be easy to solve, because the f will be well understood densities such as the normal or exponential.

Recall that in the full shift (limit) experiment, the MLE can be written as $\hat{h} = X - v^*$, and that $X + a^* = \hat{h} + (a^* + v^*)$ is best equivariant. So it is natural to construct the *shifted ML estimator* as

$$\hat{\theta}_S = \hat{\theta}_n + \varphi_n(\hat{a} + \hat{v}).$$

In the next theorem, we show that $\hat{h}_n + (\hat{a}+\hat{v}) \rightsquigarrow \hat{h} + (a^*+v^*)$, which is optimal. From Proposition 6, we know that $\hat{h}_n \rightsquigarrow \hat{h}$, so consistency of $\hat{a}$ and $\hat{v}$ will yield the desired result.

The following assumptions, while at a high level, will usually be relatively easy to check because the density f_θ in the limit experiment will already have been calculated. Since $\hat{a}$ and $\hat{v}$ are based on $f_{\hat{\theta}}$, smoothness in limit experiment density with with respect to θ is needed.

Assumption 7 *For any $\delta > 0$, there exists $\epsilon > 0$ such that,*

$$\sup_{\theta' \in B_\theta(\epsilon)} \sup_v |f_{\theta'}(v) - f_\theta(v)| \leq \delta$$

where $B_\theta(\epsilon)$ denotes the closed ball of radius ϵ centered at θ.

The next assumption ensures that the limit experiment has positive density at the true parameter value. This condition is used to show that $\hat{v}$ can only take on extreme values with low probability.

Assumption 8 *Let $\mathcal{N}$ be an open neighborhood of zero.*[7] *If V has density f_θ, then $P_\theta(V \in \mathcal{N}) > 0$.*

We also require that the integral yielding $\hat{v}$ is negligibly determined by value of the integrand in the tails.

Assumption 9 *For any sequence $M_n \longrightarrow \infty$, some $\epsilon > 0$, and any compact K,*

$$\sup_{a \in K, \theta' \in B_\theta(\epsilon)} \int_{\|v\| \geq M_n} L(v + a) f_{\theta'}(v) dv \longrightarrow 0.$$

Theorem 3 *Let Assumptions 1 - 9 hold. Then, under $P_{n,\theta+\varphi_n h_0}$,*

$$\left\{ \varphi_n^{-1} \left(\hat{\theta}_S - \theta \right) \right\} \rightsquigarrow X + a^*,$$

where X is distributed as $f_\theta(x - h_0)$. Further, $\hat{\theta}_S$ is the best regular estimator and locally asymptotically minmax.

As an illustration of these results, consider again the procurement auction model.

Example 5 *(Auction Model, continued) As previously noted, the limit experiment has $X = h + V$, where V has density*

$$f_\theta(v) = \frac{m}{\theta(m-1)} \exp\left(-\frac{m}{\theta(m-1)} v \right) \mathbf{1}\{v \geq 0\}.$$

Notice that $f_\theta(v)$ achieves its maximum at $v^* = 0$ for all θ. So by Proposition 6, under $P_{\theta+h_0/n}$,

$$n(\hat{\theta} - \theta) \rightsquigarrow X,$$

[7] Note that it would suffice to have positive density around any point in $\{x : \|x\| < D_1\}$, not necessarily zero.

where X has density $\frac{m}{\theta(m-1)} \exp\left(-\frac{m}{\theta(m-1)}(v-h_0)\right) \mathbf{1}\{v \geq h_0\}$. (This result can also be derived directly from results on order statistics.)

To illustrate the optimal shifted MLE, suppose we use absolute error loss $L(a-h) = |a-h|$. The solution to

$$\min_a \int L(v+a) f_\theta(v) dv = \min_a \int_0^\infty |v+a| \frac{m}{\theta(m-1)} \exp\left(-\frac{m}{\theta(m-1)} v\right) dv$$

is $a^* = -(\ln 2)(m-1)\theta/m$. So we set $\hat{a} = -(\ln 2)(m-1)\hat{\theta}/m$, and $\hat{v} = 0$, and the optimal shifted ML estimator is

$$\begin{aligned} \hat{\theta}_S &= \hat{\theta} + \varphi_n(\hat{a} + \hat{v}) \\ &= \hat{\theta}\left(1 - \frac{(m-1)\ln 2}{nm}\right). \end{aligned}$$

4.4 Asymptotic Behavior of the Bayes Estimator

As the sample size increases the influence of a fixed prior on a Bayes estimator becomes negligible. Correspondingly, the flat prior Bayes estimator examined in section 3.5 is the limit experiment counterpart to generalized Bayes estimators in the experiments $\mathcal{E}_n$.

Given a prior density π on the parameter space Θ, the Bayes estimator is any solution to

$$\min_{\tilde{\theta}} \int_{H_n} L(\varphi_n^{-1}(\tilde{\theta} - (\theta + \varphi_n h)) dP_{n,\theta+\varphi_n h} \pi(\theta + \varphi_n h) dh$$

or equivalently take $\tilde{h}_n = \varphi_n^{-1}(\tilde{\theta}_B - \theta)$ to minimize

$$\psi_{n,\theta}(h') = \frac{\int_{H_n} L(h'-h) \Lambda_{n,0}(h) \pi(\theta + \varphi_n h) dh}{\int_{H_n} \Lambda_{n,0}(\bar{h}) \pi(\theta + \varphi_n \bar{h}) d\bar{h}}.$$

Generally, establishing the limiting distribution of a Bayes estimator requires less stringent conditions than for the MLE. In particular, weak convergence of the likelihood ratio *process*, as implied by Assumption 5, is not necessary for weak convergence of expected posterior risk. Here we use the following conditions.

Assumption 10 *For every compact K, $\psi_{n,\theta} \overset{\theta}{\rightsquigarrow} \psi_\theta$ in $l^\infty(K)$, where*

$$\psi_\theta(h') = \frac{\int L(h'-h) \left(\frac{f_\theta(X-h)}{f_\theta(X)}\right) dh}{\int \left(\frac{f_\theta(X-\bar{h})}{f_\theta(X)}\right) d\bar{h}}.$$

Assumption 11 *For any $M_n \longrightarrow \infty$,*

$$\frac{\int_{\|h\| \geq M_n} L(-h)\Lambda_{n,0}(h)\pi(\theta + \varphi_n h)dh}{\int_{H_n} \Lambda_{n,0}(\bar{h})\pi(\theta + \varphi_n \bar{h})d\bar{h}} \xrightarrow{p} 0$$

Primitive conditions for this assumption are readily available in the literature on Bayesian estimation. Ibragimov and Hasminskii (1981) assume bounds on the tail behavior of the loss, prior, and likelihood ratio process, so that expected posterior loss is well approximated by integrals over large bounded regions.

Assumption 12 *There exists $\epsilon > 0$ such that for all n large enough,*

$$\frac{\int_B \Lambda_{n,0}(h)\pi(\theta + \varphi_n h)dh}{\int_{H_n} \Lambda_{n,0}(\bar{h})\pi(\theta + \varphi_n \bar{h})d\bar{h}} \geq \epsilon$$

almost surely, where $B = \{x : \|x\| \leq D_1\}$ for D_1 defined in Assumption 1.

This assumption presumes positive posterior probability on the set B, which implicitly requires positive prior density at the true parameter θ.

Under these assumptions and the argmax theorem, the Bayes estimator in the experiment $\mathcal{E}_n$ converges in distribution to the Bayes estimator in the limit experiment, which by Proposition 5 is optimal.

Theorem 4 *Let Assumptions 1, 2, 4, and 9 - 12 hold. Then, under $P_{n,\theta+\varphi_n h_0}$,*

$$\left\{\varphi_n^{-1}\left(\tilde{\theta}_B - \theta\right)\right\} \rightsquigarrow X + a^*,$$

where X is distributed as $f_\theta(x - h_0)$. Further, $\tilde{\theta}_B$ is best regular and locally asymptotically minmax.

Example 5 *(Auction Model, continued) To illustrate a Bayes estimator, take absolute error loss, as before, and an inverted gamma prior distribution, $\pi(\theta) = (\tau^{j-1}/(j-2)!)\theta^{-j}\exp(-\tau/\theta)$. Given a sample of n i.i.d. winning bids, the posterior density is*

$$p(\theta|b_1, \ldots, b_n) \propto \frac{1}{\theta^{n+j}} \exp\left(-\frac{\tau + mn\bar{b}}{\theta}\right) \mathbf{1}\{0 \leq \theta \leq (m-1)b_{(1)}\}$$

where $\bar{b}$ is the average of the winning bid observations and $b_{(1)}$ is the lowest order statistic. Hence, the posterior is simply a truncated inverted gamma distribution. The Bayes estimator minimizes expected posterior risk, so $\tilde{\theta}_B$ =median$(P(\theta))$. So if $(m-1)b_{(1)}$ is the x^{th} quantile of the inverted *gamma,* $f(\theta) = [(\tau + mn\bar{b})^{n+j-1}/(\theta^{n+j}(n+j-2)!)] \exp\left(-\frac{\tau+mn\bar{b}}{\theta}\right) \mathbf{1}\{\theta \geq 0\}$*, then* $\tilde{\theta}_B$ *is the* $(x/2)^{th}$ *quantile of the same distribution.*

5 Efficiency in Conditional Shift and Asymptotic Conditional Shift Experiments

A useful generalization of the full shift model studied in Section 3, is a shift model where the distribution of the observation can also depend on another observed random variable. For such experiments, much of the intuition from the full shift model carries over; in particular, a conditionally shifted maximum likelihood estimator is best equivariant and minmax. For models which are asymptotically of the conditional shift form, arguments similar to those in Section 4 can be used to show asymptotic effiency of conditionally shifted MLE and Bayes estimators. Since much of the reasoning for the conditional shift cases is similar to the full shift cases, we present the main results briefly, omitting some of the proofs.

5.1 Conditional Shift Experiments

Suppose we have an experiment with sample space $(\mathbb{R}^{k+m}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra, and with parameter space $\mathbb{R}^k$. Let $f(\cdot,\cdot)$ be a fixed Lebesgue probability density on $\mathbb{R}^{k+m}$, and suppose that the probability measures $\{P_h : h \in \mathbb{R}^k\}$ specify that (X, J) has density $f(x-h, j)$. We call such a statistical experiment a *conditional shift experiment*. It can be interpreted as specifying that $(X, J) = (V + h, J)$, where (V, J) has a fixed distribution F with density $f(v, j)$. We will use $f(v|j)$ to denote the conditional density of V given J, and $g(j)$ to denote the marginal density of J.

Example 6 *(Mixed Normal Experiment) Suppose that J is a random $k \times k$ matrix which is almost always symmetric and positive definite. Suppose that the distribution of V given J is $N(0, J)$. Then $X|J \sim N(h, J)$. This is called a* mixed normal *experiment, and arises as the limit experiment in some nonstationary time series models, see Example 7.*

A nonrandomized estimator in this experiment is a Borel-measurable function $T : \mathbb{R}^{k+m} \to \mathbb{R}^k$. A randomized estimator is a Borel-measurable function of (X, J) and U into $\mathbb{R}^k$, where U has a uniform distribution independent of (X, J). As before, we take a loss function satisfying Assumption 1, and we define the risk of an estimator as

$$\begin{aligned} R(T,h) &= E_h[L(T(X,J,U)-h)] \\ &= \int_{[0,1]} \int_{\mathbb{R}^{k+m}} L(T(x,j,u)-h) dP_h(x,j) d\lambda(u) \end{aligned}$$

$$= \int\int\int L(T(x,j,u)-h)f(x-h,j)dxdjdu.$$

A nonrandomized estimator is equivariant if for all $c, x \in \mathbb{R}^k$ and $j \in \mathbb{R}^m$,

$$T(x+c,j) = T(x,j)+c.$$

By the same argument as before, such estimators can be written as

$$T(x,j) = x + a(j),$$

where $a(j) = T(0,j)$ is measurable with respect to $\mathcal{B}(\mathbb{R}^m)$. The risk of an equivariant estimator can therefore be written as

$$\begin{aligned} R(T,h) &= \int\int L(T(x,j)-h)f(x-h|j)g(j)dxdj \\ &= \int\int L(x+a(j)-h)f(x-h|j)dxg(j)dj \\ &= \int\int L(a(j)+v)f(v|j)g(j)dvdj, \end{aligned}$$

So the risk is constant over h, and we can find the best nonrandomized equivariant estimator $T = X + a(j)$, which solves

$$a^*(j) = \arg\min_a \int L(a+v)f(v|j)dv,$$

for (almost) all j. We shall assume that there is a unique measurable solution $a^*(j)$ for almost all j.

A randomized estimator is *conditionally equivariant-in-law* if the distribution of $T-h$ conditional on J does not depend on h:

$$L(J) = \mathcal{L}(T(V+h,J,U)-h|J) \ \text{ a.s. } J \qquad \forall h \in \mathbb{R}^k$$

The convolution theorem can be extended to conditional shift models as well, showing that equivariant-in-law estimators can be written as a convolution of X and a random variable W whose distribution is independent of V (hence X) and U. However, the distribution of W will depend on J in general.

Theorem 5 *(Conditional Convolution Theorem) Let T be a conditional equivariant-in-law randomized estimator for h in the conditional shift experiment $\varepsilon = \{\mathbb{R}^{k+m}, \mathcal{B}, \mathcal{L}(V+h,J) : h \in \mathbb{R}^k\}$, where (V,J) is a fixed Lebesgue absolutely continuous random vector in $\mathbb{R}^{k+m}$. Let L be the null* *law* $L(J) = \mathcal{L}_h(T - h|J)$*. Then* L *can be written as*

$$L(J) = \mathcal{L}(V + W|J)$$

where $V = X - h$ *and* W *is a Borel-measurable random variable independent of* V *conditional on* J*.*

By the conditional convolution theorem, it follows that the best nonrandomized equivariant estimator is also best among conditionally equivariant-in-law randomized estimators.

Now consider the maximum likelihood estimator, defined as

$$\hat{h} = \arg\max_h f(X - h, J) = \arg\max_h f(X - h|J)g(J).$$

So, for $J = j$, $\hat{h} = \arg\max_h f(X - h|j)$. If $f(v|j)$ has a unique maximum $v^*(j)$ for each j, then $\hat{h} = X - v^*(J)$ is equivariant, and

$$\hat{h} + v^*(J) + a^*(J) = X + a^*(J)$$

is best equivariant, and hence minmax.[8]

As an illustration, consider the mixed normal model of Example 6. Conditional on J, the distribution of $X = h + V$ is $N(h, J)$, so the maximum likelihood estimator $\hat{h} = X$ has $v^* = 0$. For loss functions which are symmetric and have convex upper level sets,

$$\int L(a + v)f(v|j)dv$$

has a minimum at $a^* = 0$ for each j. In consequence, the MLE is best equivariant and minmax for bowl-shaped loss. For asymmetric loss functions, however, MLE is not optimal.

5.2 Asymptotic Conditional Shift Experiments

Consider the same framework as described in section 4.1. We observe Z_n from a sequence of experiments $\mathcal{E}_n = \{P_{n,\theta+\varphi_n h} : h \in H\}$. Assume $\mathcal{E}_n$ converges weakly to a conditional shift limit experiment $\mathcal{E}$. For every finite subset $I \subset \mathbb{R}^k$, and every local parameter h_0,

$$\left(\frac{dP_{n,\theta+\varphi_n h}}{dP_{n,\theta+\varphi_n h_0}}(Z_n) \right)_{h \in I} \overset{\theta+\varphi_n h_0}{\rightsquigarrow} \left(\frac{f_\theta(X - h|J)}{f_\theta(X - h_0|J)} \right)_{h \in I},$$

where $X = V + h_0$ and $V|J$ has a fixed conditional density f_θ and J has density g_θ.

[8]Formally, this follows by the generalized Hunt-Stein theorem given in Wesler (1959).

Let

$$\bar{R}^* = \int\int L(v + a^*(j)) f_\theta(v|j) g_\theta dv dj.$$

If $\tilde{\theta}_n$ is a regular estimator sequence in $\mathcal{E}_n$ and $\tilde{h}_n = \varphi_n^{-1}(\tilde{\theta}_n - \theta)$, then

$$\varphi_n^{-1}(\tilde{\theta}_n - (\theta + \varphi_n h)) = \tilde{h}_n - h \overset{\theta+\varphi_n h}{\rightsquigarrow} Q_\theta^{\tilde{\theta}_n}$$

for some limit law $Q_\theta^{\tilde{\theta}_n}$ that does not depend on h. By the asymptotic representation theorem, there exists a randomized estimator $\tilde{h}$ in $\mathcal{E}$ such that $\tilde{h}_n \overset{\theta+\varphi_n h}{\rightsquigarrow} \tilde{h}$. Hence, $\tilde{h} - h$ is distributed as $Q_\theta^{\tilde{\theta}_n}$ under h. Since $\tilde{h}$ is a randomized estimator in $\mathcal{E}$, we can write $\tilde{h}(V + h, J, U)$ and note that the distribution of $\tilde{h}(V + h, J, U) - h$ does not depend on h. It follows that $\tilde{h}$ is conditionally equivariant-in-law, and $\bar{R}^*$ then provides a best regular asymptotic risk bound.

From above $\bar{R}^*$ also provides a minmax bound in the limit experiment $\mathcal{E}$. By the same argument as in the proof of Theorem 2, this value also yields a local asymptotic minmax bound.

As in section 4.3, if the likelihood ratio process $\Lambda_{n,h_0}(h)$ converges weakly to $\Lambda_{h_0}(h) = f_\theta(X - h|J)/f_\theta(X - h_0|J)$ as stochastic processes in $l^\infty(\mathbb{R}^k)$, then the standardized MLE in $\mathcal{E}_n$, $\hat{h}_n = \varphi_n^{-1}(\hat{\theta}_n - \theta)$, converges in distribution to the MLE in $\mathcal{E}$, $\hat{h} = X - v^*(J)$. In general this maximum likelihood estimator is not optimal. We consider a local conditional shift of $\hat{\theta}_n$ which will result in an asymptotically efficient estimator.

We already have $\hat{h}_n \rightsquigarrow X - v^*(J)$. To shift this estmator optimally, an observable counterpart to J in the sequence of experiments $\mathcal{E}_n$ is needed. Suppose there exists $J_{n,\theta}$ in $\mathcal{E}_n$ such that $(\hat{h}_n, J_{n,\theta}) \rightsquigarrow (X - v^*(J), J)$. For instance, in LAMN experiments, the quadratic term in the expansion of the log likelihood ratio process is $J_{n,\theta}$. A specific example is given below.

Similar to the full shift case, $a^*(\cdot)$ and $v^*(\cdot)$ may depend on θ, so we solve for $\hat{a}(\cdot)$ and $\hat{v}(\cdot)$ from

$$\hat{a}(j) = \arg\min_a \int L(a + v) f_{\hat{\theta}_n}(v|j) dv$$

and

$$\hat{v}(j) = \arg\max_v f_{\hat{\theta}_n}(v|j).$$

$J_{n,\theta}$ is written to note its potential dependence on the unknown θ, so we use $J_{n,\hat{\theta}_n}$ which would typically differ from $J_{n,\theta}$ by a factor of $o_p(1)$. Given sufficient smoothness of the risk and conditional density in both θ and j, we obtain

$$\varphi_n^{-1}(\hat{\theta}_S - \theta) \rightsquigarrow X + a^*(J)$$

where $\hat{\theta}_S = \hat{\theta}_n + \hat{a}(J_{n,\hat{\theta}_n}) + \hat{v}(J_{n,\hat{\theta}_n})$. This locally shifted MLE attains the asymptotic risk bound

$\bar{R}^*$ and hence is both best regular and locally asymptotically minmax. A Bayes estimator is also feasible in this context and presents an asymptotically efficient alternative.

Example 7 *Consider the AR(1) model $y_i = \rho y_{i-1} + \epsilon_i$, where the ϵ_i are i.i.d. standard normal, and where $|\rho| > 1$ and y_0 is fixed. With a scaling of $\varphi_n = |\rho|^n$, the limit experiment (X, J) for this model is the mixed normal experiment of Example 6 with $X|J \overset{h}{\sim} N(h, J^{-1})$. Hence, this AR(1) model is locally asymptotically mixed normal.*

The MLE is

$$\hat{\rho}_n = \frac{\sum_{i=1}^n y_i y_{i-1}}{\sum_{i=1}^n y_{i-1}^2}$$

with $\hat{h}_n = |\rho|^n(\hat{\rho}_n - \rho)$. By the symmetry of the conditional limit experiment distribution, $v^*(\cdot) = 0$. From the quadratic term in the log likelihood ratio expansion, let $J_{n,\rho} = |\rho|^{-2n} \sum_{i=1}^n y_{i-1}^2$. Then, $(\hat{h}_n, J_{n,\rho}) \overset{\rho+|\rho|^{-n}h}{\rightsquigarrow} (X, J)$.

To obtain the optimal local shift to the MLE, note first that $\hat{v}(j) = 0$. Given a loss function L, let

$$\hat{a}(j) = \arg\min_a \int L(v + a) dN(0, j^{-1})(v).$$

For instance, if L is symmetric, then $\hat{a}(j) = 0$ for all j, and the MLE is automatically efficient. If L is a check function, then $\hat{a}(j)$ is a quantile of a normal distribution with variance j^{-1}. Or, more simply, given the corresponding quantile a_0 ($= \arg\min_a \int L(v + a) dN(0, 1)(v) dv$) of a standard normal distribution, $\hat{a}(j) = j^{-1/2} a_0$. Since the convergence rate in this example depends on ρ, an estimated convergence rate must be used to scale $\hat{a}$ for a feasible locally shifted MLE:

$$\hat{\rho}_S = \hat{\rho}_n + |\hat{\rho}_n|^{-n} \hat{a}(J_{n,\hat{\rho}_n})$$

With check function loss, this expression would simplify to $\hat{\rho}_S = \hat{\rho}_n + |\hat{\rho}_n|^{-n} J_{n,\hat{\rho}_n}^{-1/2} a_0$. Noting that $(\hat{\rho}/\rho)^n \xrightarrow{p} 1$, it is straightforward to conclude that $|\rho|^n(\hat{\rho}_S - \rho) = |\rho|^n(\hat{\rho}_n - \rho) + J_{n,\rho}^{-1/2} a_0 + o_p(1) \rightsquigarrow X + a^*(J)$. Hence this locally shifted MLE is asymptotically efficient.

6 Conclusion

Full shift and conditional shift limit experiments generalize the standard local asymptotic normal limit experiment, which arises in regular models. Examples of nonregular models which have a shift limit include certain auction models, and some nonstationary time series models. Despite this generality, the full shift and conditional shift models have a very tractable structure, and elementary arguments can be used to construct best equivariant and minmax estimators.

Earlier work by Le Cam, Van der Vaart (1996), and Strasser (1982) have noted the usefulness of shift experiments and related models for constructing convolution theorems and showing efficiency of Bayes-type estimators. In this paper, we have shown that a shifted MLE is also optimal in full shift and conditionally shifted models, and further, under suitable conditions, a locally shifted MLE is asymptotically optimal in parametric models if the limit experiment has a shift form. The optimal shift term can be constructed based on the loss function, the limiting form of the likelihood ratio process, and the MLE itself.

A Proofs

PROOF OF THEOREM 1: See Van der Vaart (1996). $\square$

PROOF OF PROPOSITION 1: Since a^* minimizes $\int L(v+a)f(v)dv$, it follows that for any law μ_W,

$$\int\int L(v+w)f(v)dvd\mu_W(w) \geq \int\int L(v+a^*)f(v)dvd\mu_W(w) = \int L(v+a^*)f(v)dv.$$

$\square$

PROOF OF PROPOSITION 2: On the sample space $\mathbb{R}^k$, define the group of transformations $\mathcal{G} = \{g_c : c \in \mathbb{R}^k\}$, where $g_c(x) = x + c$. We can regard $\mathcal{G}$ as the Euclidean space $\mathbb{R}^k$ with the usual topology. The group $\mathcal{G}$ is also defined on the parameter space and the action space. To apply the generalized version of the Hunt-Stein theorem due to Wesler (1959), the following suffices. The distributions in the full shift experiment are dominated by a σ-finite measure. The action space is separable since it is Euclidean, and loss satisfies Assumption 1(a). Since $\mathcal{G}$ is a Euclidean space, $\mathcal{G}$ is a locally compact, σ-compact topological group with its Borel σ-algebra generated by the compact subsets of $\mathcal{G}$. Since $\mathcal{G}$ is also abelian ($g_c \circ g_d = g_d \circ g_c$), it satisfies a condition known as amenability. Under these conditions, the Hunt-Stein theorem yields that the minimal risk equivariant estimator is also minmax as stated in the proposition.

Note also that the version of the Hunt-Stein theorem in Wesler (1959) requires enlarging the class of estimators to all Markov kernels. However, for full shift models, Theorem 12.4 of Van der Vaart (1996) shows that there is no loss of generality in restricting attention to randomized estimators. $\square$

PROOF OF PROPOSITION 5: The flat prior Bayes estimator T_B minimizes

$$\begin{aligned}\int R(T_B, h)d\Pi(h) &= \int\int L(T_B - h)f(x-h)dxdh \\ &= \int\int L(T_B - h)f(x-h)dhdx\end{aligned}$$

Thus, for each x we can choose $T_B(x)$ to minimize the inner integral

$$\int L(T_B(x) - h)f(x-h)dh = \int L(T_B(x) + v - x)f(v)dv,$$

where $v = x - h$. Let $a(x) = T_B(x) - x$. Then we can equivalently choose $a(x)$ to minimize

$$\int L(v + a(x))f(v)dv.$$

This can be minimized by setting $a(x) = a^*$ for all x, so we can set

$$T_B(x) = x + a^*.$$

The last conclusion follows directly from Propositions 1 and 2. $\square$

PROOF OF THEOREM 2: The asymptotic representation theorem given in Van der Vaart (1991) can be applied to a sequence of standardized estimators as follows. Given the regular estimator $\tilde{\theta}_n$, set $\tilde{h}_n = \varphi_n^{-1}\left(\tilde{\theta}_n - \theta\right)$. Then there exists a randomized estimator $\tilde{h}$ in the full shift limit experiment such that $\tilde{h}_n \overset{\theta+\varphi_n h}{\rightsquigarrow} \tilde{h}$. It follows that $\tilde{h} - h$ is distributed as $Q_\theta^{\tilde{\theta}_n}$ under h in the limit experiment. Hence, $\tilde{h}$ is an equivariant-in-law estimator, and so the best equivariant-in-law estimator gives a bound on the best regular estimators. The conclusion in (4) follows by Proposition 1.

The conclusion in (5) is a direct consequence of Van der Vaart (1996) Theorem 5.9, since loss is subcompact by the last part of Assumption 1(a) and the limit experiment is dominated by the absolute continuity of $F_{\theta,h}$ in Assumption 4. $\square$

PROOF OF PROPOSITION 6: Assumptions 4 and 5 imply that Λ_{n,h_0} converges weakly to Λ_{h_0} in $l^\infty(\mathbb{R}^k)$, by Theorem 18.14 of Van der Vaart (1998). By Assumption 3, Λ_{h_0} has a unique, well-separated maximum almost surely under h_0. Then by Assumption 6, Lemma 3.2.1 in Van der Vaart and Wellner (1996), and the Portmanteau theorem, conclude that

$$\hat{h}_n \overset{\theta+\varphi_n h_0}{\rightsquigarrow} \arg\max_h \frac{f_\theta(X-h)}{f_\theta(X-h_0)},$$

where X has density $f_\theta(x - h_0)$. Under h_0, the maximizer of this last expression is almost surely the same as $\arg\max_h f_\theta(X - h)$. The result then follows from Proposition 3. $\square$

PROOF OF THEOREM 3: By Proposition 6, $\hat{\theta} \xrightarrow{p} \theta$. Then under Assumption 7,

$$\sup_v |f_{\hat{\theta}}(v) - f_\theta(v)| \xrightarrow{p} 0$$

Since the maximum of f_θ is well-separated by Assumption 3, Van der Vaart and Wellner (1996) Corollary 3.2.3(i) yields $\hat{v} \xrightarrow{p} v^*$.

For D_1 defined in Assumption 1, let $B = \{x : \|x\| \leq D_1\}$ and

$$\delta = \inf_{\|x\| \geq D_1^\eta} L(x) - \sup_{\|x\| \leq D_1} L(x) > 0.$$

For $\|a\| \geq D_1^\eta + D_1$ and $v \in B$, $\|v + a\| \geq D_1^\eta$ and so $L(v + a) - L(v) \geq \delta$.

Let $C_n = \{x : \|x\| \leq M_n\}$. For large enough M_n, $\|a\| \geq M_n^\eta + M_n$ and $v \in C_n$, $\|v + a\| \geq M_n^\eta$ and so $L(v + a) - L(v) \geq 0$. Hence, for $\|a\| \geq M_n^\eta + M_n$,

$$\begin{aligned}
&\int L(v+a) f_{\hat{\theta}}(v) dv - \int L(v) f_{\hat{\theta}}(v) dv \\
= \; & \int_B [L(v+a) - L(v)] f_{\hat{\theta}}(v) dv + \int_{C_n \cap B^c} [L(v+a) - L(v)] f_{\hat{\theta}}(v) dv + \int_{C_n^c} [L(v+a) - L(v)] f_{\hat{\theta}}(v) dv \\
\geq \; & \delta \int_B f_{\hat{\theta}}(v) dv + 0 + \int_{C_n^c} -L(v) f_{\hat{\theta}}(v) dv
\end{aligned}$$

By Assumption 9, $\int_{C_n^c} -L(v) f_{\hat{\theta}}(v) dv \xrightarrow{p} 0$. By Assumption 7, $\int_B f_{\hat{\theta}}(v) dv \xrightarrow{p} \int_B f_\theta(v) dv > 0$. This last expression is positive and bounded away from zero by Assumption 8. Thus, $\int L(v + a) f_{\hat{\theta}}(v) dv - \int L(v) f_{\hat{\theta}}(v) dv > 0$ has a positive limit almost surely for $\|a\| \geq M_n^\eta + M_n$. It follows that $\Pr(\|\hat{a}\| \geq M_n^\eta + M_n) \longrightarrow 0$, so $\hat{a} - a$ is uniformly tight.

From Assumption 9, for any compact K,

$$\sup_{a \in K} \left| \int L(v + a) f_{\hat{\theta}}(v) dv - \int L(v + a) f_\theta(v) dv \right| \xrightarrow{p} 0.$$

Also, by continuity of L, Assumption 9, and the dominated convergence theorem, $\int L(v+a) f_\theta(v) dv$ is continuous (and hence lower semi-continuous). So, by Van der Vaart and Wellner (1996) Corollary 3.2.3(ii), $\hat{a} \xrightarrow{p} a^*$. This verifies the first conclusion of the theorem.

From Theorem 2, the last conclusion follows directly from the given convergence in distribution expression. □

PROOF OF THEOREM 4: To apply the argmax theorem for a uniformly convergent stochastic process $\psi_{n,\theta}$ on compacta, we need to establish uniform tightness of the Bayes estimator. Let δ and C_n be defined as in the proof of Theorem 3. Then following a similar argument as in that proof, for $\|h'\| \geq M_n^\eta + M_n$,

$$\psi_{n,\theta}(h') - \psi_{n,\theta}(0) \geq \delta \frac{\int_B \Lambda_{n,0}(h) \pi(\theta + \varphi_n h) dh}{\int_{H_n} \Lambda_{n,0}(\bar{h}) \pi(\theta + \varphi_n \bar{h}) d\bar{h}} - \frac{\int_{C_n^c} L(-h) \Lambda_{n,0}(h) \pi(\theta + \varphi_n h) dh}{\int_{H_n} \Lambda_{n,0}(\bar{h}) \pi(\theta + \varphi_n \bar{h}) d\bar{h}}$$

By Assumption 11, the last term converges to zero in probability. Then by Assumption 12, $\psi_{n,\theta}(h') - \psi_{n,\theta}(0)$ has a positive limit almost surely. Hence, $\tilde{h}_n$ is uniformly tight. Further, ψ_θ is lower semi-continuous almost surely by Assumption 9 and the continuity of the loss function given in Assumption 1. By the argmax theorem, see Van der Vaart and Wellner (1996) Theorem 3.2.2, $\tilde{h}_n = \varphi_n^{-1}\left(\tilde{\theta}_B - \theta\right) \rightsquigarrow \arg\min_h \psi_\theta(h)$. By Assumption 2 and Proposition 5, this minimizer is unique and equal to $X + a^*$ almost surely. The asymptotic efficiency conclusions then follow from Theorem 2. $\square$

PROOF OF THEOREM 5: Let P denote the joint law of (V, J, U) on $\Omega = \mathbb{R}^k \times \mathbb{R}^m \times [0, 1]$ with Borel σ-algebra $\mathcal{F}$. Note that $\omega \in \Omega$ could also be denoted (v, j, u).

Let Y be the vector of coordinate projections given by (V, U), i.e. $Y(v, j, u) = (v, u)$. Let $\mathcal{D} = Y^{-1}(\mathcal{F})$ and $\mathcal{C} = J^{-1}(\mathcal{F})$. By Dudley (2002) Theorem 10.2.2, there exists a conditional distribution $P_{Y|\mathcal{C}}(D, \omega)$ for $D \in \mathcal{D}$. For all $D \in \mathcal{D}$, $P_{Y|\mathcal{C}}(D, \cdot)$ is $\mathcal{C}$-measurable, so we can write it as simply a function of j rather than $\omega = (v, j, u)$. Let $P_j(D) = P_{Y|\mathcal{C}}(D, (v, j, u))$. For almost all j, $P_j(\cdot)$ is a probability measure on $\mathcal{D}$. We can make this a probability measure for all j. For the exceptional set of measure zero, just choose a probability measure associated with another j.

For each j, define T_j as $T_j(V, U) = T(V, j, U)$. The family of probability measures indexed by $h \in \mathbb{R}^k$ and given by the law $\mathcal{L}(V + h|J = j)$ constitutes a full shift experiment, and T_j is an equivariant-in-law randomized estimator on this experiment. So by Theorem 1, there exists W_j independent of V conditional on $J = j$ such that $L(j) = \mathcal{L}(V + W_j|J = j)$. W_j exists on the probability space $(\Omega, \mathcal{F}, P)$, and by the construction in the proof of the convolution theorem, we can assume that as a mapping on Ω, $W_j(v, l, u)$ does not depend on l. We can then define a random variable W on Ω by $W(v, j, u) = W_j(v, j, u)$ $(= W_j(v, l, u)$ for all $l)$. By its definition, W satisfies the conclusion of this theorem. $\square$

References

ANDREWS, D. (1994): "Emprical Process Methods in Econometrics," in *Handbook of Econometrics, Vol. 4*. North-Holland, New York.

CAVANAGH, C. L., L. E. JONES, AND T. J. ROTHENBERG (1990): "Efficient Estimation under Asymmetric Loss," manuscript, University of California.

CHERNOZHUKOV, V., AND H. HONG (2001): "Likelihood Inference with Density Jump," manuscript.

DONALD, S., AND H. PAARSCH (1993): "Maximum Likelihood Estimation When the Support of the Distribution Depends Upon Some or All of the Unknown Parameters," manuscript.

DUDLEY, R. M. (2002): *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, second edn.

EATON, M. L. (1989): *Group Invariance Applications in Statistics*, Regional Conference Series in Probability and Statistics. IMS-ASA.

HÁJEK, J. (1970): "A Characterization of Limiting Distributions of Regular Estimates," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14, 323–330.

——— (1972): "Local Asymptotic Minmax and Admissibility in Estimation," *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 175–194.

HIRANO, K., AND J. PORTER (2002): "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," forthcoming, *Econometrica*.

IBRAGIMOV, I., AND R. HASMINSKII (1981): *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York.

KNIGHT, K. (2000): "Epi-convergence in Distribution and Stochastic Equi-semicontinuity," manuscript.

LE CAM, L. (1972): "Limits of Experiments," in *Proceedings of the Sixth Berkeley Symposium of Mathematical Statistics*, vol. 1, pp. 245–261.

——— (1986): *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

PAARSCH, D. (1992): "Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics*, 51, 191–215.

Strasser, H. (1982): "Local asymptotic minimax properties of Pitman estimates," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 60, 223–247.

Van der Vaart, A. W. (1991): "An Asymptotic Representation Theorem," *International Statistical Review*, 59, 97–121.

——— (1996): "Limits of Experiments," manuscript.

——— (1998): *Asymptotic Statistics.* Cambridge University Press, New York.

Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Wald, A. (1950): *Statistical Decision Functions.* Wiley, New York.

Wesler, O. (1959): "Invariance Theory and a Modified Minimax Principle," *Annals of Mathematical Statistics*, 30, 1–20.