Le Cam's comparison of experiments

Paul Delatte
delatte@usc.edu
University of Southern California

Last updated: 08 November, 2022

Comparison of statistical experiments (a.k.a. statistical models), which form the fundamental stratum of statistics, has a naturally intuitive justification. It is more precisely warranted from two directions (see vdV S.7. in The Statistical Work of Lucien Le Cam): (a) between two statistical models for the same inferential parameter, we want to know which one contains more information on the parameter (so as to work directly with the most informative one); (b) given a possibly complex statistical model, we want to approximate it (in the limit) by some simpler model for the same parameter (ideally, we would like some notion of asymptotic equivalence for experiments with strong guarantees in the complex models for rules derived in the simpler model of the equivalence class). The two objectives can be satisfied simultaneously by introducing properly quantified notions of comparison and closeness between statistical experiments (which should be intuitively based on "information"). This was done by Le Cam from the 1950s (with a paper published finally in 1964) building on ideas of Bohneblust, Shapley, Blackwell, Sherman, and Stein (among others): the starting point is the realization that one experiment is "better" than another experiment if the latter can be perfectly approximated, in the sense of total variation, by a randomization of the former; this yields by symmetrization the subsequent idea that "two experiments are close if each is well approximated, in the sense of total variation, by a randomization of the other" (Pollard), with the limiting case of equivalence by perfect approximation of each by the other (which is equivalent to each experiment being "better" than the other).

Remark. The idea of randomization is already central in the early ideas for the comparison of statistical models (the same idea is at the basis of sufficiency for statistics). Randomization should be thought "as a mechanism to convert observations from one distribution into observations from another distribution: sample a y from $\mathbb{Q}$, crank up the randomizer, then out pops an observation x from $\mathbb{P}$" (Pollard); for instance, an experiment $\mathscr{F}$ can be obtained from randomization of $\mathscr{E}$ if "$\mathscr{F}$ is reproducible from $\mathscr{E}$ by "tossing coins"" (Le Cam). This naturally translates in terms of probability kernels (a.k.a. Markov kernels): that is, we say that $\mathscr{E}$ can be obtained from randomization from $\mathscr{F}$ if there is a kernel ρ such that

$$\rho\mathbb{P} = \mathbb{Q}$$

where

$$\rho\mathbb{P}(A) = \int_E \rho(x, A)\, d\mathbb{P}(x)$$

for all $A \in \mathcal{F}$. The problem is that the theory of comparison of experiments does not develop well with probability kernels as tool for randomization due to technical reasons of a measure-theoretic nature. This forced Le Cam to define randomization procedures in greater generality (see vdV SWofLC and Pollard Paris2001 "Randomization") up to the complete abandonment of the sample spaces. For our purpose (we follow Mariucci), it is sufficient to impose further constraint on the experiments for the technical difficulties to disappear. One such solution that cover many cases is to consider experiments where the underlying space is a Polish metric space and the measures are all commonly dominated by a σ-finite measure. We should call such experiments **Polish dominated** and denote the set of all such experiments on a given parameter space Θ by $S_p(\Theta)$. The equivalence between Markov kernels and more general randomization procedures à la Le Cam is then guaranteed (see P.9.2. in Nussbaum (1996)). Note that other topological conditions can be imposed to guarantee the same equivalence (see Nussbaum (1996) or Pollard Paris2001 "Randomization"). Even if we state all notions for Polish dominated experiments, we should keep in mind that many of the definitions and results generalize for arbitrary sample spaces when considering the right notion of randomization procedures.

1 Deficiency and Le Cam distance

Definition 1 (Deficiency)**.** Let Θ be arbitrary set and $\mathscr{E} = \{E, \mathcal{E}, \{\mathbb{P}_\theta : \theta \in \Theta\}\}$ and $\mathscr{F} = \{F, \mathcal{F}, \{\mathbb{Q}_\theta : \theta \in \Theta\}\}$ dominated Polish statistical experiments. The **deficiency** of $\mathscr{E}$ relative to $\mathscr{F}$ is given by[1]

$$\delta(\mathscr{E}, \mathscr{F}) = \inf_{\rho} \sup_{\theta \in \Theta} d_{TV}(\rho \mathbb{P}_\theta, \mathbb{Q}_\theta)$$

where the infimum runs over all Markov kernels $\rho \colon E \times \mathcal{F} \to [0, 1]$ from $(E, \mathcal{E})$ to $(F, \mathcal{F})$.

Definition 2 (Le Cam Distance)**.** Let Θ be arbitrary set and $\mathscr{E} = \{E, \mathcal{E}, \{\mathbb{P}_\theta : \theta \in \Theta\}\}$ and $\mathscr{F} = \{F, \mathcal{F}, \{\mathbb{Q}_\theta : \theta \in \Theta\}\}$ dominated Polish statistical experiments. The **Le Cam distance** or **$\boldsymbol{\Delta}$-distance** between $\mathscr{E}$ and $\mathscr{F}$ is given by

$$\Delta(\mathscr{E}, \mathscr{F}) = \max\{\delta(\mathscr{E}, \mathscr{F}), \delta(\mathscr{F}, \mathscr{E})\}.$$

Proposition 3. *The Le Cam distance is a pseudo-metric on the set of all Polish dominated experiments with parameter space Θ.*

Proof. See S.9. in Nussbaum (1996) which refers to 59.2 in Strasser (1985). □

Definition 4. Let $\mathscr{E}$ and $\mathscr{F}$ be Polish dominated statistical experiments for the same parameter space. Then:

(a) $\mathscr{E}$ is said to be **better** or **more informative** than $\mathscr{F}$ if $\delta(E, F) = 0$;

(b) $\mathscr{E}$ and $\mathscr{F}$ are said to be **equivalent** if $\Delta(\mathscr{E}, \mathscr{F}) = 0$.

Definition 5. Let $(\mathscr{E}_n)_{n\in\mathbb{N}}$ and $(\mathscr{F}_n)_{n\in\mathbb{N}}$ be sequences of Polish dominated statistical experiments for the same parameter space. The (sequences of) experiments are said to be **asymptotically equivalent** if $\lim_{n\to\infty} \Delta(\mathscr{E}_n, \mathscr{F}_n) = 0$.

[1] Factors 2 or $1/2$ may appear in the definitions and results of other authors involving deficiency, depending on which convention for total variation is used. In particular, vdV (2002) scales the TV by 2 and hence the deficiency, whereas Le Cam (1986) scales up the TV by 2 but divides the defiency by $1/2$ to be equal to ours.

The (pseudo)metrics Δ naturally induces a topology on the set $S_p(\Theta)$ of Polish dominated experiments with parameter space Θ. Convergence with respect to this topology can then be considered. In particular, if $(\mathscr{E}_n)_{n\in\mathbb{N}}$ and $(\mathscr{E})_{n\in\mathbb{N}}$ are sequences in $S_p(\Theta)$ such that $\lim_{n\to\infty}\Delta(\mathscr{E}_n,\mathscr{E})=0$, then the sequence of experiments $(\mathscr{E}_n)_{n\in\mathbb{N}}$ is said to be **converge** to the experiment $\mathscr{E}$, or, by abuse of language, that $\mathscr{E}_n$ converges to $\mathscr{E}$. In our terminology, this is equivalent to the sequences $(\mathscr{E}_n)_{n\in\mathbb{N}}$ and $(\mathscr{E})_{n\in\mathbb{N}}$ being asymptotically equivalent.

2 Interpretation of $\delta(\mathscr{E},\mathscr{F})$ via risk functions

Theorem 6. *Let $\mathscr{E}$ and $\mathscr{F}$ be Polish dominated statistical experiments for the same parameter space Θ. Let $\varepsilon \geq 0$ and $(\mathbb{A},\mathcal{A})$ be an action space where $\mathbb{A}$ is a Polish metric space and $\mathcal{A}$ its Borel σ-algebra. Then*

$$\delta(\mathscr{E},\mathscr{F}) \leq \varepsilon$$

if and only if for every decision rule[2] *$\sigma\colon F\times\mathcal{A}\longrightarrow[0,1]$ on $\mathscr{F}$, and every loss function L with values in $[0,1]$, there is a decision rule $\rho\colon E\times\mathcal{A}\longrightarrow[0,1]$ on $\mathscr{E}$ such that for every $\theta\in\Theta$,*

$$R(\theta,\rho;\mathscr{E},L)\leq R(\theta,\sigma;\mathscr{F},L)+\varepsilon,$$

where $R(\theta,\rho;\mathscr{E},L)=\int_E\int_{\mathbb{A}}L(\theta,a)\,d\rho(x,a)\,d\mathbb{P}_\theta(x)$ is the risk function on $\mathscr{E}$ relative to L and $R(\theta,\sigma;\mathscr{E},L)=\int_F\int_{\mathbb{A}}L(\theta,a)\,d\rho(x,a)\,d\mathbb{Q}_\theta(x)$ is the risk function associated with $\mathscr{F}$ relative to L. Moreover, if $\delta(\mathscr{E},\mathscr{F})<\varepsilon$, then ρ may be chosen independently of L.

Proof. T.2. in Le Cam (1986), which is already in Le Cam (1964) and equivalently stated in C.6.4.4. in Torgersen (1991) (with the additional close on L), for general spaces and general randomizations. Our formulation is the application to Markov kernels under Polish domination restriction as in T.2.7. in Mariucci (2016) or in T.7.2 in vdV (2002). (Mariucci does not state the additional L clause whereas vdV directly state the L clause in the equivalence, which is simply a weaker result). See the previous footnote for explanations of the differences by factors of 2 or 1/2 in the result. □

This implies that

$$\delta(\mathscr{E},\mathscr{F})=\sup_\sigma\inf_\rho\sup_L\sup_\theta|R(\theta,\rho;\mathscr{E},L)-R(\theta,\sigma;\mathscr{F},L)|$$

and, subsequently, that[3]

$$\Delta(\mathscr{E},\mathscr{F})=\max\Big\{\sup_\sigma\inf_\rho\sup_L\sup_\theta|R(\theta,\rho;\mathscr{E},L)-R(\theta,\sigma;\mathscr{F},L)|,\\ \sup_\rho\inf_\sigma\sup_L\sup_\theta|R(\theta,\sigma;\mathscr{F},L)-R(\theta,\rho;\mathscr{E},L)|\Big\}.$$

[2] Given a statistical model $\mathscr{F}=\{F,\mathcal{F},\{\mathbb{Q}_\theta:\theta\in\Theta\}\}$ and an action space $(\mathbb{A},\mathcal{A})$, a decision rule σ on $\mathscr{F}$ is defined as a Markov kernel from $(F,\mathcal{F})$ to $(\mathbb{A},\mathcal{A})$. Some authors call them randomized decision rules to distinguish them from deterministic decision rules which are the special case when for every $x\in F$, $A\mapsto\rho(x,A)$ is the Dirac measure at some point $T(x)$ where $T\colon F\longrightarrow\mathbb{A}$ is some measurable function.

[3] This is the definition of the Le Cam distance in Giné&Nickl MFIDS p.9.

3 Strong and weak topology of experiments

The topology induced by the Le Cam (pseudo)distance Δ on the set of (Polish dominated) experiments $S_P(\Theta)$ is said to be the **strong topology** on $S_P(\Theta)$. As showed in the previous result, this topology provides excellent guarantees in terms of risk function. It is, however, possible to define a weak(er) topology on $S_P(\Theta)$, which naturally provides weaker decision-theoretic guarantees but is sufficient for many important statistical results (e.g., Hajek–Le Cam convolution and LAM theorems) and appears more often in practice. The main reason for the usefulness of this weaker topology is that it coincides with the strong topology when the parameter space Θ is finite, so that the strong decision-theoretic guarantees then apply.

Definition 7 (Weak Topology of Experiments)**.** Let Θ be an arbitrary set and $S_P(\Theta)$ the set of Polish dominated experiments with parameter space Θ. Denote $A(\Theta)$ the family of finite subsets of Θ and $\mathscr{E}_\alpha$ the restriction of an experiment $\mathscr{E}$ to the parameter space $\alpha \in A(\Theta)$. The topology induced by the family[4] of pseudometrics $((\mathscr{E}, \mathscr{F}) \mapsto \Delta(\mathscr{E}_\alpha, \mathscr{F}_\alpha))_{\alpha \in A(\Theta)}$ is said to be the **weak topology** on $S_P(\Theta)$.

Reference. See S.3.4. in Le Cam (1986), D.59.10. in Strasser (1985), T.7.4.12. in Torgersen (1991), or Pollard's Thoughts (2000) p.2. See also remarks in vdV AS p.137.

Proposition 8. *The weak topology on $S_p(\Theta)$ is weaker than the strong topology on $S_p(\Theta)$. If Θ is finite, then the two topologies coincide.*

Proof. T.7.4.12. in Torgersen (1991) or T.60.2. in Strasser (1985) p.302. □

The weak topology naturally induces a notion of convergence for (nets and) sequences of experiments. In particular, a sequence of experiments $(\mathscr{E}_n)_{n\in\mathbb{N}}$ in $S_P(\Theta)$ is said to **weakly converge** to an experiment $\mathscr{E}$ in $S_P(\Theta)$ if $(\mathscr{E}_n)_{n\in\mathbb{N}}$ converge to $\mathscr{E}$ with respect to the weak topology on $S_P(\Theta)$. As shown next, weak convergence can be expressed through the restriction of Δ to all finite subsets of Θ. Naturally, this invites us to define weak version of the notions of better experiments, equivalence, and asymptotic equivalence through the restriction of δ and Δ to all finite subsets of Θ.

Proposition 9. *A sequence of experiments $(\mathscr{E}_n)_{n\in\mathbb{N}}$ in $S_P(\Theta)$ weakly converges to an experiment $\mathscr{E}$ in $S_P(\Theta)$ if and only if*

$$\lim_{n\to\infty} \Delta(\mathscr{E}_{n,\alpha}, \mathscr{E}_\alpha) = 0$$

for all $\alpha \in A(\Theta)$.

Proof. T.7.4.12. in Torgersen (1991) or T.60.2. in Strasser (1985) p.302. □

This equivalence works only on the condition that the parameter space Θ is invariant. If we allow Θ to be a sequence $(\Theta_n)_{n\in\mathbb{N}}$ such that $\Theta_n \to \Theta$, then weak convergence and the Δ_α characterization need not coincide (see S.60. in Strasser (1985)).

[4] A topology induced by a family of functions, also know as an initial topology, is the weakest topology that makes all the functions in the family continuous. A topology induced by a family of pseudometrics $d \in D$ on a set X is the initial topology for the family $(x \mapsto d(x, y))_{y\in X,\, d\in D}$ (see uniform spaces; e.g., p.467 in Strasser (1985)).

Proposition 10. *A sequence of experiments $(\mathscr{E}_n)_{n\in\mathbb{N}}$ in $S_P(\Theta)$ weakly converges if and only if $(\mathscr{E}_n,\alpha)_{n\in\mathbb{N}}$ converges for every $\alpha \in A(\Theta)$. In this case,* $(\lim_{n\to\infty}\mathscr{E}_n)_\alpha$ *is equivalent to* $\lim_{n\to\infty}\mathscr{E}_{n,\alpha}$ *for every $\alpha \in A(\Theta)$.*

Proof. T.60.2. in Strasser (1985) p.302. □

4 Control of the Le Cam distance

The explicit computation of the Le Cam distance is in general difficult, hence the need for simpler control procedures. We partly follow Mariucci (2016).

4.1 Total variation upper bound

If the statistical models have same sample space, then a simple upper bound in total variation obtained by taking the Dirac measure in the definition of the deficiency often suffices for the control of the Le Cam distance. Results on the total variation distance then come to help to further bound the Le Cam distance.

Proposition 11. *Let $\mathscr{E} = \{E, \mathcal{E}, \{\mathbb{P}_\theta : \theta \in \Theta\}\}$ and $\mathscr{F} = \{F, \mathcal{F}, \{\mathbb{Q}_\theta : \theta \in \Theta\}\}$ be Polish dominated statistical experiments for the same parameter space Θ and defined for the same sample space $(E, \mathcal{E}) = (F, \mathcal{F})$. Then*

$$\Delta(\mathscr{E}, \mathscr{F}) \leq \sup_{\theta\in\Theta} d_{TV}(\mathbb{P}_\theta, \mathbb{Q}_\theta).$$

Proof. This follows directly by taking the kernel given by the indicator function in the definition of the deficiency. See P.3.1. in Mariucci (2016) or C.6.2.5. in Torgersen (1991) or P.59.6. in Strasser p.297. See also (1.13) in Giné&Nickl MFIDS p.9 for the proof starting with the equivalent characterization of the deficiency in terms of risk functions. □

4.2 Control by the likelihood ratio process

It turns out that equivalence of experiments can be linked directly to the behavior of some stochastic processes defined from the likelihood, namely likelihood ratio processes. This connection is particularly useful due to all the known results on likelihood ratios for regular enough models. The connection becomes essential when considering the weak topology of experiments, for then we have complete equivalence between weak convergence of experiments and weak convergence of the finite-dimensional distributions of likelihood ratio processes.

Definition 12 (Likelihood Ratio Process)**.** Let Θ be an arbitrary set. Let $\mathscr{E} = \{E, \mathcal{E}, \{\mathbb{P}_\theta : \theta \in \Theta\}\}$ be a statistical experiment and $\vartheta \in \Theta$. The process $(L_{\theta,\vartheta})_{\theta\in\Theta}$ given by

$$L_{\theta,\vartheta} = \frac{d\mathbb{P}_\theta}{d\mathbb{P}_\vartheta}$$

is said to be the **likelihood ratio process** from $(\mathbb{P}_h)_{h\in H}$ in base ϑ.

If $\mathbb{P}_\theta$ is absolutely continuous with respect to $\mathbb{P}_\vartheta$, then $d\mathbb{P}_h/d\mathbb{P}_\vartheta$ is the Radon–Nikodym derivative (also known as the likelihood ratio in statistics). If it is not, then the Lebesgue decomposition theorem guarantees that $\mathbb{P}_\vartheta = \mathbb{P}_0 + \mathbb{P}_1$ with $\mathbb{P}_\theta \ll \mathbb{P}_0$ and $\mathbb{P}_\theta \perp \mathbb{P}_1$, so we define $d\mathbb{P}_\theta / d\mathbb{P}_\vartheta := d\mathbb{P}_\theta / d\mathbb{P}_0$. In each case, the likelihood ratio $L_{\theta,\vartheta} \colon E \longrightarrow [0, +\infty)$ for a given θ can be interpreted as a random variable on $(E, \mathcal{E}, \mathbb{P}_\vartheta)$. Moreover, in both cases (absolute continuity or not), it is always possible to find a finite measure μ such that $\mathbb{P}_\theta$ and $\mathbb{P}_\vartheta$ are absolutely continuous with respect to μ (for instance, $\mu = \mathbb{P}_\theta + \mathbb{P}_\vartheta$). Then, if we denote $p_\theta := d\mathbb{P}_\theta / d\mu$ and $p_\vartheta := d\mathbb{P}_\vartheta / d\mu$ their respective density with respect to μ, we have (see L.6.2. in vdV AS p.86 or E.1.4. in Strasser MS p.2 for the general case and P.3.9. in Folland RA p.91 for the absolute continuity case) that[5]

$$L_{\theta,\vartheta} = \frac{d\mathbb{P}_\theta}{d\mathbb{P}_\vartheta} = \frac{p_\theta}{p_\vartheta} \quad (\mathbb{P}_\vartheta\text{-a.e.}).$$

In both cases (absolute continuity or not) and any two formulations (with or without densities), the likelihood ratio is only uniquely defined $\mathbb{P}_\vartheta$-a.e..

Proposition 13. *Let Θ be an arbitrary set. Let $\mathscr{E} = \{E, \mathcal{E}, \{\mathbb{P}_\theta : \theta \in \Theta\}\}$ and $\mathscr{F} = \{F, \mathcal{F}, \{\mathbb{Q}_\theta : \theta \in \Theta\}\}$ be statistical experiments in $S_P(\Theta)$. Then $\mathscr{E}$ and $\mathscr{F}$ are equivalent if and only if $\frac{d\mathbb{P}_\theta}{d\mathbb{P}_\vartheta}$ and $\frac{d\mathbb{Q}_\theta}{d\mathbb{Q}_\vartheta}$ have same law for all $\vartheta \in \Theta$.*

Proof. C.25.9. in Strasser (1985) p.114 (see also P.3.6. in Mariucci (2016) p.7). □

Proposition 14. *Let Θ be an arbitrary set. Let $(\Lambda^E_{\theta,\vartheta})_{\theta\in\Theta}$ and $(\Lambda^F_{\theta,\vartheta})_{\theta\in\Theta}$ be stochastic processes, where $\Lambda^E_{\theta,\vartheta}$ is defined on $(E, \mathcal{E}, \mathbb{P}_\vartheta)$ and $\Lambda^F_{\theta,\vartheta}$ on $(F, \mathcal{F}, \mathbb{Q}_\vartheta)$. Suppose there is some coupling Γ for $(\Lambda^E_{\theta,\vartheta})_{\theta\in\Theta}$ and $(\Lambda^F_{\theta,\vartheta})_{\theta\in\Theta}$. If $(\Lambda^E_{\theta,\vartheta})_{\theta\in\Theta}$ and $(\Lambda^F_{\theta,\vartheta})_{\theta\in\Theta}$ are equal to the likelihood ratio processes of some experiments $\mathscr{E} = (E, \mathcal{E}, \{\mathbb{P}_\theta : \theta \in \Theta\})$ and $\mathscr{F} = \{F, \mathcal{F}, \{\mathbb{Q}_\theta : \theta \in \Theta\}\}$ in $S_P(\Theta)$, then*

$$\Delta(\mathscr{E}, \mathscr{F}) \leq \sup_{\theta\in\Theta} \mathbb{E}_\Gamma \Big| \Lambda^E_{\theta,\vartheta} - \Lambda^F_{\theta,\vartheta} \Big|.$$

Proof. L.6. in Le Cam&Yang (2000) p.30 (see also P.3.7. in Mariucci (2016) p.7). □

Proposition 15. *Let Θ be an arbitrary set. A sequence of experiments $(\mathscr{E}_n)_{n\in\mathbb{N}} = ((E_n, \mathcal{E}_n, \{\mathbb{P}_{\theta,n} : \theta \in \Theta\}))_{n\in\mathbb{N}}$ in $S_P(\Theta)$ weakly converges to an experiment $\mathscr{E} = \{E, \mathcal{E}, \{\mathbb{P}_\theta : \theta \in \Theta\}\}$ in $S_P(\Theta)$ if and only if*

$$\left(\frac{d\mathbb{P}_{\theta,n}}{d\mathbb{P}_{\vartheta,n}}\right)_{\theta\in\alpha} \rightsquigarrow \left(\frac{d\mathbb{P}_\theta}{d\mathbb{P}_\vartheta}\right)_{\theta\in\alpha}$$

for all $\alpha \in A(\Theta)$ and all $\vartheta \in \alpha$, where $A(\Theta)$ denotes the family of all finite subsets of Θ.

Proof. T.60.3. in Strasser (1985). See also, for Θ finite, L.5. in Le Cam&Yang (2000) p.29 (with a standard proof) and L.1. in Pollard's Thoughts (2000) p.3 (with a different proof than by the canonical representation); the finite result can then be applied to $\alpha(\Theta)$ and used with the characterization of weak convergence of experiments in terms of the restriction of Δ to $\mathscr{E}_{n,\alpha}$. See also S.5. in vdV (2002) for some context. □

[5] The result is the same $\mathbb{P}_\vartheta$-a.e. for any dominating σ-finite measure μ.

References

FOLLAND, G. B. (1999): *Real analysis: modern techniques and their applications*, vol. 40. John Wiley & Sons.

GINE, E., AND R. NICKL (2021): *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press.

LE CAM, L. (1986): *Asymptotic methods in statistical decision theory*. Springer.

LE CAM, L., AND G. L. YANG (2000): *Asymptotics in statistics: some basic concepts*. Springer.

MARIUCCI, E. (2016): "Le Cam theory on the comparison of statistical models," *Preprint*.

NUSSBAUM, M. (1996): "Asymptotic equivalence of density estimation and Gaussian white noise," *The Annals of Statistics*, 24(6), 2399–2430.

POLLARD, D. (2000): "Some thoughts on Le Cam's statistical decision theory," *Preprint*.

——— (2001): "Lecture in Paris on Le Cam's theory," *Lecture notes*.

STRASSER, H. (1985): *Mathematical theory of statistics: statistical experiments and asymptotic decision theory*, vol. 7. Walter de Gruyter.

TORGERSEN, E. (1991): *Comparison of statistical experiments*, vol. 36. Cambridge University Press.

VAN DER VAART, A. (2002): "The statistical work of lucien le cam," *The Annals of Statistics*, 30(3), 631–682.

VAN DER VAART, A. W. (1998): *Asymptotic statistics*. Cambridge University Press.