Chapter 7

Efficiency via asymptotic minimax bounds

SECTION 1 reminds you of the Fisherian concept of efficiency, which is invalid. It points the way to the rescue of the concept via requirements that estimators behave reasonably at local alternatives.

SECTION 2 establishes a general asymptotic lower bound for minimax risk of a sequence of statistical models.

SECTION 3 solves the minimax problem for Gaussian shift families and a large class of loss functions.

SECTION 4 proves Anderson's lemma.

1. Efficiency: sunk and rescued

Recall from Chapter 1 the concept of efficiency according to Fisher (1922, page 277): "The criterion of efficiency is satisfied by those statistics which, when derived from large samples, tend to a normal distribution with the least possible standard deviation." More formally, for a sequence of models $\mathcal{P}_n := \{\mathbb{P}_{n,\theta} : \theta \in \Theta\}$, with Θ a subset of the real line, reasonable estimators $\widehat{\theta}_n$ for θ were supposed to have the property

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightsquigarrow N(0, \sigma_\theta^2) \qquad \text{under } \mathbb{P}_{n,\theta}.$$

The variance σ_θ^2 was supposed to be larger than $\mathbb{I}(\theta)^{-1}$, the inverse of the Fisher information. Efficient estimators were those for which $\sigma_\theta^2 = \mathbb{I}(\theta)^{-1}$.

As you learned in Section 1.4, the unqualified assertion about the limit distributions is not valid. There exist estimators, such as the Hodges estimator, that beat the efficiency bound. You also learned that the Hodges construction to improve the asymptotic behaviour of an estimator at a fixed θ_0 has unwelcome consequences at sequences of ***local alternatives*** θ_n that converge to θ_0 at a $1/\sqrt{n}$ rate.

By excluding estimators with the unwelcome local behavior we can rescue the Fisherian concept of efficiency. In this Chapter you will learn about one rescue method, due to Hájek (1972) and Le Cam (1972), which excludes superefficency by requiring reasonable behaviour of estimators uniformly over a range of parameter values. Chapter 8 will present an alternative approach (the Hájek-Le Cam convolution theorem), which derives a stronger form of efficiency under stronger assumptions.

Consider the meaning of efficiency at a fixed θ_0 in $\Theta \subseteq \mathbb{R}^d$. We wish to make asymptotic assertions about the behavior of $\Lambda_n(\widehat{\theta}_n - \theta_0)$, for some sequence of rescaling matrices $\{\Lambda_n\}$. For example, we could choose $\Lambda_n = \sqrt{n}\mathbb{I}_{\theta_0}^{1/2}$ to give efficient estimators a simple $N(0, I_d)$ limiting distribution. The appropriate local alternatives will turn out to be values θ_n of the form $\theta_0 + \Lambda_n^{-1}h$ for a fixed h. That is, we need to study the behavior of estimators under $\mathbb{Q}_{n,h} := \mathbb{P}_{n,\theta_0+\Lambda_n^{-1}h}$, for fixed (or bounded) h. The minimax theory will refer to the local model $\mathcal{Q}_n := \{\mathbb{Q}_{n,h} : h \in \mathbb{R}^d\}$.

> REMARK. Here I am tacitly assuming that $\theta_0 + \Lambda_n^{-1}h \in \Theta$ for all h. To avoid minor notational problems, you may take $\mathbb{P}_{n,\theta}$ to equal $\mathbb{P}_{n,\theta_0}$ if $\theta \notin \Theta$. If θ_0 is an interior point of Θ, the redefinition will eventually be unimportant, because the main result will be stated in terms of limiting behavior for finite subsets of Θ.

To each estimator $\widehat{\theta}_n$ for θ there is a corresponding "estimator" $\widehat{t}_n := \Lambda_n(\widehat{\theta}_n - \theta_0)$ for h. Of course $\widehat{t}_n$ is not a true estimator (because it depends on θ_0) but the distinction will have no effect on the subsequent analysis. Similarly, for each randomized estimator σ for θ there is a randomized estimator τ for h, defined by

<1> $$\tau_y^t g(t) = \sigma_y^z g\left(\Lambda_n(z - \theta_0)\right) \qquad \text{for } g \in \mathcal{M}^+(\mathbb{R}^d).$$

That is, to generate an observation t from τ_y, we first generate z from σ_y then put $t := \lambda_n(z - \theta_0)$.

It also makes sense to incorporate the rescaling into the loss function for $\mathcal{P}_n$, by defining $L_{n,\theta}(z) := \rho\left(\Lambda_n(z - \theta)\right)$, for a fixed ρ, in order to compare the behavior of estimators that converge to θ_0 at a Λ_n^{-1} rate. We then have

$$\mathbb{P}_{n,\theta_n}^y \sigma_y^z L_{n,\theta_n}(z) = \mathbb{Q}_{n,h}\tau_y^t \rho(t - h) \qquad \text{for } \theta_n := \theta_0 + \Lambda_n^{-1}h \text{ and } \tau \text{ as in <1>}.$$

That is, $R_n(\theta_0 + \Lambda_n^{-1}h, \sigma) = R(\tau, h)$, where the risk for σ uses the loss function L_n and the risk for τ uses the loss function $L_h(t) := \rho(t - h)$.

Of course all the rescue attempts require some assumptions about the asymptotic behaviour of the $\mathcal{P}_n$ models. The traditional approach imposes classical regularity assumptions: independent observations, existence of smooth densities, and requirements that derivatives can be taken under integral signs. The assumptions are used to construct approximations to likelihood ratios, which could be invoked to establish convergence in Le Cam's sense for models indexed by finite sets of h values.

The efficiency argument become more transparent if, instead of working via the classical assumptions, we assume directly the Le Cam convergence for the $\mathcal{Q}_n$ models. More precisely, we need only assume existence of a limit model $\mathcal{Q} := \{\mathbb{Q}_h : h \in \mathbb{R}^d\}$ for which $\delta(\mathcal{Q}(S), \mathcal{Q}_n(S)) \to 0$ for each finite subset S of $\mathbb{R}^d$. The Hájek result corresponds to the case where $\mathcal{Q}$ is a ***Gaussian shift family***, that is, where $\mathbb{Q}_h := N(h, \Gamma)$ for a fixed matrix Γ. In particular, it covers the case of product measures $\mathbb{Q}_{n,h} := P_{\theta_0+h/\sqrt{n}}^n$ under an LAN assumption. Limit theory for the $\mathcal{Q}_n$ models, at a fixed θ_0, is called ***local***, because it deals with behavior of estimators under alternatives in shrinking neighborhoods of θ_0. When specialized to models obtained by local reparametrizations, the general result in Section 2 will provide a ***local asymptotic minimax bound***.

2. Asymptotic minimax lower bound

Le Cam realized that central idea behind Hájek's proof could be reinterpreted as a simple semicontinuity property (under the Δ metric) for risk functions. The Gaussian form of the limit is needed only to calculate a neat expression for the asymptotic lower bound, under a mild assumption on the loss function.

For the general case, I will write the asymptotic lower bound in a form that makes clear the role of the Le Cam convergence. Actually, we do not need convergence in the Δ metric, even for the submodels, because the main Theorem provides only a one-sided bound. As shown in Section 3, the lower bound simplifies significantly when $\mathcal{Q}$ is a Gaussian shift and the loss function satisfies a mild convexity requirement.

In what follows, $\mathbb{D}$ will denote a fixed decison space, and $L_h(\cdot)$ will be a nonnegative loss function, not necessarily of the form $\rho(t-h)$. The index set H can be arbitrary, and not just $\mathbb{R}^d$. For each finite subset S of H, write $\mathcal{Q}(S)$ for the submodel of $\mathcal{Q}$ indexed by S. Define $\mathcal{Q}_n(S)$ similarly. The models $\mathcal{Q}_n$ will live on spaces $(\mathcal{Y}_n, \mathcal{B}_n)$, and the limit $\mathcal{Q} := \{\mathbb{Q}_h : h \in H\}$ will live on $(\mathcal{Y}, \mathcal{B})$. For the Gaussian shift family, $\mathcal{Y}$ is a Euclidean space $\mathbb{R}^d$. The risk for a randomized procedure $\tau \in \Re(\mathcal{Y}, \mathbb{D})$ under the model $\mathcal{Q}$ is defined as $R(\tau, h) := (\tau\mathbb{Q}_h)^t\, L_h(t)$ for $h \in H$. The maximum risk for a procedure τ is defined as $\sup_{h\in H} R(\tau, h)$. A minimax procedure, if it exists, has the smallest possible the maximum risk.

Remember that $\|K\mu - K\nu\| \le \|\mu - \nu\|$ for each K in $\Re(\mathcal{Y}, \mathcal{Y}_n)$ and all measures μ, ν in $\mathbb{L}^+_\sigma(\mathcal{Y})$.

<2> **Theorem.** *Suppose $\delta(\mathcal{Q}(S), \mathcal{Q}_n(S)) \to 0$ for each finite subset S of H. Define*

$$R_{\mathcal{Q}} := \sup_{C,S} \inf_\tau \max_{h\in S} (\tau\mathbb{Q}_h)^t \left(L_h(t) \wedge C\right),$$

the supremum running over all constants $C \in \mathbb{R}^+$ and all finite subsets S of H, and the infimum running over all randomized procedures τ in $\Re(\mathcal{Y}, \mathbb{D})$. Then, for each sequence of randomized estimators, with $\tau_n \in \Re(\mathcal{Y}_n, \mathbb{D})$,

$$\liminf_n \sup_{h\in H} R(\tau_n, h) \ge R_{\mathcal{Q}}.$$

More precisely, for each constant $R < R_{\mathcal{Q}}$, there exists a finite subset S of H, a finite constant C, and an n_0 such that

$$\inf_{\tau_n} \max_{h\in S} \left(\tau_n\mathbb{Q}_{n,h}\right)^t \left(L_h(t) \wedge C\right) > R \qquad \text{when } n \ge n_0,$$

the infimum running over all τ_n in $\Re(\mathcal{Y}_n, \mathbb{D})$, for each $n \ge n_0$.

REMARK. The result is really just a trivial consequence of the fact that

$$\{\tau_n K_n : \tau_n \in \Re(\mathcal{Y}_n, \mathbb{D})\} \subseteq \Re(\mathcal{Y}, \mathbb{D}) \qquad \text{for each } K_n \text{ in } \Re(\mathcal{Y}, \mathcal{Y}_n).$$

For models $\mathcal{Q}_n$ obtained by local reparametrization, the stronger form of the result makes it clear that good behavior is required only at local alternatives $\{\theta_0 + \Lambda_n^{-1}h : h \in S\}$, for finite S.

Proof. Choose a constant R' with $R_{\mathcal{Q}} > R' > R$. By definition of $R_{\mathcal{Q}}$, there exists a constant C and a finite S for which $\inf_\tau (\tau\mathbb{Q}_h)^t\,(L_h(t) \wedge C) > R'$. These are the C and S of the stronger assertion; they stay fixed for the remainder of the proof. Abbreviate $L_h(t) \wedge C$ to $\ell_h(t)$. For each n there exist randomizations $K_n \in \mathfrak{R}(\mathcal{Y}, \mathcal{Y}_n)$ for which

$$\epsilon_n := \tfrac{1}{2}\max_{h\in S} \|K_n\mathbb{Q}_h - \mathbb{Q}_{n,h}\| \to 0 \qquad \text{as } n \to \infty.$$

For every τ_n in $\mathfrak{R}(\mathcal{Y}_n, \mathbb{D})$, and $h \in S$, and g in $\mathbb{M}^+(\mathcal{Y})$ with $0 \le g \le 1$,

$$|\tau_n K_n \mathbb{Q}_h g - \tau_n \mathbb{Q}_{n,h} g| \le \tfrac{1}{2}\|\tau_n K_n \mathbb{Q}_h - \tau_n \mathbb{Q}_{n,h}\| \le \tfrac{1}{2}\|K_n\mathbb{Q}_h - \mathbb{Q}_{n,h}\| \le \epsilon_n.$$

In particular, taking g equal to ℓ_h/C we get, for every τ_n,

$$\left(\tau_n \mathbb{Q}_{n,h}\right)^t \ell_h(t) \ge (\tau_n K_n \mathbb{Q}_h)^t\,\ell_h(t) - C\epsilon_n.$$

The lower bound is $\ge R' - C\epsilon_n$ because $\tau_n K_n$ is a randomized procedure in $\mathfrak{R}(\mathcal{Y}, \mathbb{D})$.
□ For all n large enough, $R' - C\epsilon_n > R$.

Things to do:

- Asymptotic uniqueness under LAN. See Hájek (1972) and Le Cam (1979).

3. Gaussian shift families

For simplicity, let $\mathcal{Q} := \{\mathbb{Q}_h : h \in \mathbb{R}^d\}$ be a standardized Gaussian shift family on $\mathbb{R}^d$. That is, $\mathbb{Q}_t := N(t, I_d)$. For local reparametrizations under LAN, the more general case can be reduced to the standard case by absorbing an extra factor of $\Gamma^{1/2}$ into the rescaling matrix Λ_n. Assume $\mathbb{D} = H = \mathbb{R}^d$, so that the τ's from Section 2 become randomized estimators for the parameter. Local compactness of $\mathbb{R}^d$ and domination of $\mathcal{Q}$ ensure that each τ in $\mathfrak{R}(\mathcal{Y}, \mathbb{D})$ can be represented by a Markov kernel (Section 4.3). For a wide class of loss functions, the infimum in the definition of $R_{\mathcal{Q}}$ will be achieved by the very simple, nonrandom estimator $T_0(y) \equiv y$.

<3> **Definition.** *Say that a nonnegative function ρ on $\mathbb{R}^d$ is* ***bowl-shaped*** *if each of the sets $\{x : \rho(x) \le c\}$, for $c \in \mathbb{R}^+$, is convex and symmetric about the origin.*

For example, $\rho_1(x) := |x|^2$, and $\rho_2(x) := \min(1, x'Gx)$, with G a positive definite matrix, and $\rho_1 \vee \rho_1$, are all bowl-shaped. For the remainder of the Section, assume $L_h(t) := \rho(t-h)$, with ρ a bowl-shaped function.

For the lower bound $R_{\mathcal{Q}}$, we seek a Markov kernel τ to minimize

$$\max_{h\in S} \mathbb{Q}_h^y \tau_y^t\,(\rho(t-h) \wedge C)\,.$$

As you will see, the estimator T_0 minimizes for every finite C and finite S. The proof will use the fact that $\mathcal{Q}$ is a shift family, together with the invariance property of Lebesgue measure $\mathfrak{m}$ on $\mathbb{R}^d$,

<4> $$\mathfrak{m}^h g(h+y) = \mathfrak{m}^h g(h) \qquad \text{for all } y \in \mathbb{R}^d \text{ and } g \in \mathcal{M}^+(\mathbb{R}^d).$$

If $\mathfrak{m}$ were a finite measure, the proof would be very easy. To get some understanding for the method, consider the heuristic based on the false assumption that $\mathfrak{m}\mathbb{R}^d = 1$. The formal proof will use a limiting form of the same idea, replacing $\mathfrak{m}$ by a sequence a distributions behaving more and more like "uniform probability distributions on $\mathbb{R}^d$".

<5> **Example.** The proof that T_0 is minimax would be easy if Lebesgue measure were a probability. For the remainder of this Example, please suspend your disbelief and pretend that $\mathfrak{m}\mathbb{R}^d = 1$. For simplicity, take d equal to 1.

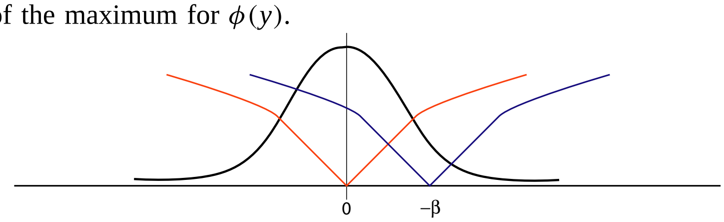
First note that $\mathbb{Q}_h^y \rho\left(T_0(y) - h\right) = \mathbb{Q}_0^y \rho(y)$ for all h, so that the supremum over h has no effect when we calculate the maximum risk for T_0. However, for a general randomized estimator τ,

$$\begin{aligned} \sup_h R(\tau,h) &\geq \mathfrak{m}^h R(\tau,h) && \text{because } \mathfrak{m}\mathbb{R} = 1 \text{ ?!} \\ &= \mathfrak{m}^h \mathbb{Q}_0^y \tau_{y+h}^t \rho(t-h) && \text{shift family} \\ &= \mathbb{Q}_0^y \mathfrak{m}^h \tau_{y+h}^t \rho(t-h) && \text{by Fubini} \\ &= \mathbb{Q}_0^y \mathfrak{m}^h \tau_h^t \rho(t-h+y) && \text{by <4>} \\ &= \mathfrak{m}^h \tau_h^t \left(\mathbb{Q}_0^y \rho(t-h+y)\right) && \text{by Fubini.} \end{aligned}$$

Consider the innermost integral. I hope you will find it obvious—if not, you could resort to Calculus—that the value of

$$\mathbb{Q}_0^y \rho(y+\beta) = \mathfrak{m}^y \left(\phi(y)\rho(y+\beta)\right) \qquad \text{where } \phi := N(0,1) \text{ density,}$$

is minimized when β equals 0, for then the minimum of $\rho(y+\beta)$ occurs at the location of the maximum for $\phi(y)$.



For fixed t and h, the innermost integral, $\mathbb{Q}_0^y \rho(t-h+y)$, is smallest when $t = h$. Thus the maximium risk for τ is greater than

$$\mathfrak{m}^h \tau_h^t \mathbb{Q}_0^y \rho(y) = \mathbb{Q}_0^y \rho(y) = \sup_h r(h, T_0) \qquad \text{because } \tau_h\mathbb{R} = 1 = \mathfrak{m}\mathbb{R} \text{ ?!}$$

□ Thus T_0 is the minimax estimator.

The same heuristic argument would work in higher dimensions, although the fact about minimization of the inner integral might no longer seem so obvious.

<6> **Lemma.** *For ρ bowl-shaped, $\beta \mapsto \mathbb{Q}_0^y \rho(y+\beta)$ is minimized when β equals* 0.

A rigorous proof of the Lemma, due to Anderson (1955), is not too difficult. The complete argument appears in Section 4.

To replace the false heuristic assumption that $\mathfrak{m}\mathbb{R}^d = 1$ by a rigorous argument, we must work with a sequence of probability measures that are almost invariant in an asymptotic sense. Let $B_r(z)$ denote the closed ball of radius r and center z. Write $\mathfrak{U}_\sigma$ for the uniform distribution on $B_\sigma(0)$. Then

$$\mathfrak{U}_\sigma^h g(h+y) = \mathfrak{m}^h g(\sigma h)\{h \in B_1(y/\sigma)\}/\mathfrak{m}B_1(0).$$

Intuitively, as $\sigma \to \infty$, the measure $\mathfrak{U}_\sigma$ behaves increasingly like the mythical uniform distribution. More precisely,

$$\sup_{|g|\le 1} |\mathfrak{U}_\sigma^h g(h+y) - \mathfrak{U}_\sigma^h g(h)|$$

$$= \sup_{|g|\le 1} |\mathfrak{m}^h g(\sigma h)\left(\{h \in B_1(y/\sigma)\} - \{h \in B_1(0)\}\right)|/\mathfrak{m}B_1(0)$$

<7> $$\le \mathfrak{m}|B_1(y/\sigma) - B_1(0)|/\mathfrak{m}B_1(0) =: \psi(y/\sigma).$$

The function $\psi(t) := \mathfrak{m}|B_1(t) - B_1(0)|/\mathfrak{m}B_1(0)$ quantifies how far $\mathfrak{U}_\sigma$ is from having the real invariance property. Notice that $2 \ge \psi(t) \downarrow 0$ as $|t| \downarrow 0$.

equality?

<8> **Theorem.** *Let $L_h(t) := \rho(t-h)$, with $\rho(\cdot)$ bowl-shaped. Then $R_\mathbb{Q} \ge \mathbb{Q}_0^z \rho(z)$.*

Proof. Consider first the case where ρ is bounded ($0 \le \rho \le C$) and uniformly continuous: $|\rho(x+\delta) - \rho(x)| \le \alpha(|\delta|) \to 0$ as $|\delta| \to 0$. In this case, the risk is also uniformly continuous in h, uniformly in τ:

$$|R(\tau, h+\delta) - R(\tau, h)|$$

$$= |\mathbb{Q}_{h+\delta}^y \tau_y^t \rho(t-h-\delta) - \mathbb{Q}_h^y \tau_y^t \rho(t-h)|$$

$$\le \mathbb{Q}_{h+\delta}^y \tau_y^t |\rho(t-h-\delta) - \rho(t-h)| + |\mathbb{Q}_\delta^y \tau_{h+y}^t \rho(t-h) - \mathbb{Q}_0^y \tau_{h+y}^t \rho(t-h)|$$

<9> $$\le \alpha(\delta) + C\|\mathbb{Q}_\delta - \mathbb{Q}_0\| \to 0 \qquad \text{as } \delta \to 0.$$

The uniform continuity allows us to reduce suprema over continuous ranges to maxima over discrete subranges with only a small decrease in maximum risk.

Given an $\epsilon > 0$, choose σ so large that $C\mathbb{Q}_0^y \psi(y/\sigma) < \epsilon$ and find δ so that the expression in the last line of <9> is also $< \epsilon$. Let S_σ be a δ-net for the ball $B_\sigma(0)$. Then, for every τ,

$$\epsilon + \max_{h \in S_\sigma} R(\tau, h) \ge \sup_{h \in B_\sigma(0)} R(\tau, h).$$

Now argue as in Example <5>, but with $\mathfrak{m}$ replaced by $\mathfrak{U}_\sigma$ to reduce the lower bound to

$$\mathbb{Q}_0^y \mathfrak{U}_\sigma^h \tau_{y+h}^t \rho(t-h) \ge \mathbb{Q}_0^y \left(\mathfrak{U}_\sigma^h \tau_h^t \rho(t-h+y) - C\psi(y/\sigma)\right) \qquad \text{cf. <7>}$$

$$\ge \mathfrak{U}_\sigma^h \tau_h^t \left(\mathbb{Q}_0^y \rho(t-h+y)\right) - \epsilon$$

$$\ge \mathfrak{U}_\sigma^h \tau_h^t \left(\mathbb{Q}_0^y \rho(y)\right) - \epsilon \qquad \text{by Lemma <6>}.$$

The assertion of the Theorem follows, at least when ρ is bounded and uniformly continuous.

For the general case, invoke the result from Problem [1] to express ρ as a pointwise increasing limit of bounded, uniformly continuous, bowl-shaped loss functions $\{\rho_i\}$. By Monotone Convergence, $\mathbb{Q}_0\rho_i \uparrow \mathbb{Q}_0\rho$. Fix an i such that $\mathbb{Q}_0\rho_i > \mathbb{Q}_0\rho - \epsilon$, then argue as above for ρ_i. □

*4. Anderson's lemma

The result from Lemma <6> is a special case of a general assertion about integrals.

<10> **Definition.** *Say that f is symmetric (about the origin) if $f(-x) = f(x)$ for all x. Say that a set C is symmetric (about the origin) if $-x \in C$ whenever $x \in C$. (That is, C is symmetric if its indicator is a symmetric function.)*

Write $\mathfrak{m}$ for Lebesgue measure on $\mathbb{R}^d$.

<11> **Theorem.** *Let f and g be nonnegative measurable functions on $\mathbb{R}^d$ such that both $\{f \geq t\}$ and $\{g \leq t\}$ are symmetric, convex sets for each $t \in \mathbb{R}^+$. Then the function*

$$G(\beta) := \mathfrak{m}^y \left(g(y+\beta) f(y)\right) = \mathfrak{m}^y \left(g(y) f(y-\beta)\right)$$

is minimized at $\beta = 0$.

The proof of the Theorem depends upon a strange-looking inequality involving sums of sets, $A \oplus B := \{x + y : x \in A, y \in B\}$. Even if both A and B are Lebesgue measurable, there is no guarantee that $A \oplus B$ is also Lebesgue measurable. To be precise, I should use inner measure with respect to $\mathfrak{m}$.

<12> **Brunn-Minkowski Inequality.** *For measurable subsets A and B of $\mathbb{R}^d$,*

$$(\mathfrak{m}(A \oplus B))^{1/d} \geq (\mathfrak{m}A)^{1/d} + (\mathfrak{m}B)^{1/d}.$$

Proof. Consider the case $d = 2$. The proof for higher dimensions differs only in notational details.

The argument develops in four steps. First dispose of the easy case where both A and B are open rectangles, with sides parallel to the coordinate axes. Call such sets *coordinate rectangles*. Then argue by induction to extend the result to finite unions of coordinate rectangles. Approximate by finite open covers to extend to the case of compact A and B, then complete the proof by an inner approximation.

CASE (i) Suppose A and B are both coordinate rectangles, with side lengths ℓ_a, w_a and ℓ_b, w_b. The sum $A \oplus B$ is a rectangle with side lengths $\ell_a + \ell_b, w_a + w_b$. The asserted inequality, $\sqrt{(\ell_a + \ell_b)(w_a + w_b)} \geq \sqrt{\ell_a w_a} + \sqrt{\ell_b w_b}$, is then a special case of the Cauchy-Schwarz inequality.

CASE (ii) Suppose A is a union of m disjoint open rectangles with sides parallel to the coordinate axes, and B is a similar union of n rectangles. Argue by induction on $m + n$. Case (i) covers $m + n = 2$. So suppose $m > 1$.

If two open coordinate rectangles do not intersect, they must lie on opposite sides of some line parallel to either the x- or y-axis. Thus we may assume, with no loss of generality, that at least one of the A rectangles lies in each of the open half-space to the left and right of a line $x = x_A$. Define

$$A^- := A \cap \{x < x_A\}, \qquad \text{and} \qquad A^+ := A \cap \{x > x_A\} \qquad \text{and} \qquad \theta = \frac{\mathfrak{m}A^+}{\mathfrak{m}A}.$$

Choose x_B so that the sets $B^- := B \cap \{x < x_B\}$ and $B^+ := B \cap \{x > x_B\}$ also have Lebesgue measures with $\mathfrak{m}B^+ = \theta\mathfrak{m}B$. That is, the line $x = x_B$ divides B in the same proportions as the line $x = x_A$ divides A.

Notice that each of A^- and A^+ is a disjoint union of at most $m - 1$ open rectangles; each of B^- and B^+ is a disjoint union of at most n open rectangles.

The sum $A^- \oplus B^-$ lies in $\{x < x_A + x_B\}$; the sum $A^+ \oplus B^+$ lies in the disjoint halfspace $\{x > x_A + x_B\}$; and each is a subset of $A \oplus B$. By disjointness and the inductive hypothesis,

$$\begin{aligned}
\mathfrak{m}(A \oplus B) &\ge \mathfrak{m}(A^- \oplus B^-) + \mathfrak{m}(A^+ \oplus B^+) \\
&\ge \left(\sqrt{\mathfrak{m}A^-} + \sqrt{\mathfrak{m}B^-}\right)^2 + \left(\sqrt{\mathfrak{m}A^+} + \sqrt{\mathfrak{m}B^+}\right)^2 \\
&= \left(\sqrt{1-\theta}\sqrt{\mathfrak{m}A} + \sqrt{1-\theta}\sqrt{\mathfrak{m}B)}\right)^2 + \left(\sqrt{\theta}\sqrt{\mathfrak{m}A} + \sqrt{\theta}\sqrt{\mathfrak{m}B}\right)^2 \\
&= (1 - \theta + \theta)\left(\sqrt{\mathfrak{m}A} + \sqrt{\mathfrak{m}B}\right)^2.
\end{aligned}$$

Take square roots to complete the argument.

Case (iii) Suppose both A and B are compact. Approximate by sets of the form appearing in case (ii) by arguing as follows. Fix $\delta > 0$. Let

$$A^\delta = \{z : d(z, A) < \delta\}.$$

Cover A by open coordinate rectangles all lying within A^δ. Extract a finite subcover. Express the union of the covering rectangles as a finite union of disjoint open rectangles plus a finite number of line segments. Let A^* be the union of just the disjoint rectangles. Notice that $\mathfrak{m}A^\delta \ge \mathfrak{m}A^* \ge \mathfrak{m}A$. In similar fashion find a finite union B^* of disjoint open coordinate rectangles in B^δ such that $\mathfrak{m}B^\delta \ge \mathfrak{m}B^* \ge \mathfrak{m}B$. All points of $A^\delta \oplus B^\delta$ lie within 2δ of $A \oplus B$. Thus

$$\begin{aligned}
\sqrt{\mathfrak{m}\left((A \oplus B)^{2\delta}\right)} &\ge \sqrt{\mathfrak{m}(A^* \oplus B^*)} \\
&\ge \sqrt{\mathfrak{m}A^*} + \sqrt{\mathfrak{m}B^*} \qquad \text{by Case (ii)} \\
&\ge \sqrt{\mathfrak{m}A} + \sqrt{\mathfrak{m}B}.
\end{aligned}$$

As $\delta \to 0$, the set $(A \oplus B)^{2\delta}$ shrinks down to the compact set $A \oplus B$, and

$$\mathfrak{m}(A \oplus B)^{2\delta} \downarrow \mathfrak{m}(a \oplus B).$$

Case (iv) Choose compact $\{A_i\}$ with $A_1 \subseteq A_2 \subseteq \ldots \subseteq A$ and $\mathfrak{m}A_i \uparrow \mathfrak{m}_*A$, and similarly for $\{B_i\}$. Then

$$\begin{aligned}
\sqrt{\mathfrak{m}_*(A \oplus B)} &\ge \sqrt{\mathfrak{m}(A_i \oplus B_i)} \\
&\ge \sqrt{\mathfrak{m}A_i} + \sqrt{\mathfrak{m}B_i} \to \sqrt{\mathfrak{m}_*A_i} + \sqrt{\mathfrak{m}_*B_i},
\end{aligned}$$

□ which completes the proof.

<13> **Corollary.** *For symmetric convex subsets C and D of $\mathbb{R}^d$, the function $h(\beta) := \mathfrak{m}\left((\beta \oplus C) \cap D\right)^{1/d}$ is maximized at $\beta = 0$.*

Proof. Write C_β for the translated set $\beta \oplus C$. By symmetry of both C and D, the map $x \mapsto -x$ takes $C_\beta \cap D$ onto $C_{-\beta} \cap D$. The two sets are reflections of each other. They have the same Lebesgue measure. The function $h(\cdot)$ is symmetric.

Convexity of both C and D ensures that

$$(C_\beta \cap D) \oplus (C_{-\beta} \cap D) \subseteq 2(C \oplus D).$$

It follows from the Theorem that

$$(h(\beta) + h(-\beta))^d \leq \mathfrak{m}\left(2(C \oplus D)\right) = 2^d \mathfrak{m}(C \oplus D) = (2h(0))^d .$$

□ That is, $h(\beta) = \frac{1}{2}\left(h(\beta) + h(-\beta)\right) \leq h(0)$, as asserted.

Proof of Theorem <11>. Supose first that the symmetric, convex set $C_t := \{f \geq t\}$ is bounded for each $t > 0$. Let $D_s := \{g \leq s\}$. Each C_t and D_s is a symmetric, convex set, with $\mathfrak{m}(C_t) < \infty$ for each $t > 0$. By Corollary <13>, for each s and t the d-dimensional volume

$$\mathfrak{m}\left((\beta \oplus C_t) \cap D_s^c\right) = \mathfrak{m}C_t - \mathfrak{m}\left((\beta \oplus C_t) \cap D_s\right)$$

is minimized at $\beta = 0$. It follows that

$$\begin{aligned} G(\beta) &= \mathfrak{m}^z \mathfrak{m}^s \mathfrak{m}^t \left(\{0 < t \leq f(z - \beta)\}\{0 < s < g(z)\}\right) \\ &= \mathfrak{m}^s \mathfrak{m}^t \mathfrak{m}^z \left(\{z \in \beta \oplus C_t\}\{z \in D_s^c\}\right) \\ &= \mathfrak{m}^s \mathfrak{m}^t \left(\mathfrak{m}\left((\beta \oplus C_t) \cap D_s^c\right)\right) \end{aligned}$$

is minimized when $\beta = 0$.

For the general case, replace f by the function $z \mapsto f(z)\{|z| \leq R\}$, to which
□ the special case applies, and then pass to the limit as $R \to \infty$.

5. Problems

[1] Show that every (nonnegative) bowl-shaped function ρ can be expressed as an almost sure (Lebesgue) pointwise limit of an increasing sequence of bounded, uniformly continuous, bowl-shaped functions, by following these steps.

Write T for the set of all nonnegative rational numbers. For each t in T, define C_t as the convex, symmetric set $\{\rho \leq t\}$.

(i) Define $f_{m,t}(x) := t \wedge (md(x, C_t))$, for $t \in T$ and $m \in \mathbb{N}$. Show that each $f_{m,t}$ is a bounded, uniformly continuous, bowl-shaped function for which $f_{m,t} \leq \rho$. Hint: Show that $f_{m,t}(x) = 0$ if $\rho(x) \leq t$.

(ii) Show that the boundary of each C_t has zero Lebesgue measure. Write $\mathcal{N}$ for the Lebesgue-negligible set $\cup_{t \in T} \partial C_t$.

(iii) Show that $\sup_{m,t} f_{m,t}(x) = \rho(x)$ for each x in $\mathcal{N}^c$. Hint: If $\rho(x) > t \in T$, show that $x \notin C_t \cup \partial C_t$. Deduce that $d(x, C_t) > 0$, and hence $\sup_m f_{m,t}(x) = t$.

(iv) Show that a pointwise maximum of any two bowl-shaped functions is also bowl-shaped.

(v) Use (iv) to build the required increasing sequence from the countable collection of bowl-shaped functions $\{f_{m,t} : m \in \mathbb{N},\ t \in T\}$.

6. Notes

Check Anderson (1955). The proof of the Brunn-Minkowski inequality, in Section 4, comes from Federer (1969, page 277).

References

Anderson, T. W. (1955), 'The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities', *Proceedings of the American Mathematical Society* **6**, 170–176.

Federer, H. (1969), *Geometric Measure Theory*, Springer-Verlag.

Fisher, R. A. (1922), 'On the mathematical foundations of theoretical statistics', *Philosophical Transactions of the Royal Society of London, A* **222**, 309–368.

Hájek, J. (1972), Local asymptotic minimax and admissibility in estimation, *in* L. Le Cam, J. Neyman & E. L. Scott, eds, 'Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability', Vol. I, University of California Press, Berkeley, pp. 175–194.

Le Cam, L. (1972), Limits of experiments, *in* L. Le Cam, J. Neyman & E. L. Scott, eds, 'Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability', Vol. I, University of California Press, Berkeley, pp. 245–261.

Le Cam, L. (1979), On a theorem of J. Hájek, *in* J. Jurečková, ed., 'Contributions to Statistics–Hájek Memorial Volume', Akademia, Prague, pp. 119–135.