Appendix D

Hypothesis Tests

This appendix presents theory on the construction and existence of exponentially powerful tests. The focus is towards applications to proofs of consistency and contraction rates of posterior distributions. The appendix starts with general theory, and next focuses on a number of different statistical models.

D.1 Minimax Risk

Let P be a probability measure and let $\mathcal{Q}$ be a collection of finite measures on a measurable space $(\mathfrak{X}, \mathscr{X})$. The *minimax risk for testing* P versus $\mathcal{Q}$, weighted by positive numbers a and b, is defined by

$$\pi(P, \mathcal{Q}) = \inf_{\phi}\Big(aP\phi + b \sup_{Q\in\mathcal{Q}} Q(1-\phi)\Big). \tag{D.1}$$

The infimum is taken over all *tests*, i.e. measurable functions $\phi: \mathfrak{X} \to [0, 1]$. The problem is to give a manageable bound on this risk, or equivalently on its two components, the probabilities of *errors of the first kind* $P\phi$ and *of the second kind* $Q(1-\phi)$. Consideration of the symmetric case $a = b$ and probability measures Q suffices for most applications, but considering weights and general finite measures is useful and not more difficult. For simplicity, we assume throughout the section that P and $\mathcal{Q}$ are dominated by a σ-finite measure μ, and denote by p and q the densities of the measures P and Q.

The *Hellinger transform* $\rho_\alpha(p; q)$ is defined in (B.5) for pairs of densities. We also write $\rho_\alpha(P; Q)$ and for ease of notation

$$\rho_\alpha(P; \mathcal{Q}) = \sup\{\rho_\alpha(P; Q): Q \in \mathcal{Q}\}. \tag{D.2}$$

Proposition D.1 (Minimax risk) *If P and $\mathcal{Q}$ are dominated, then the infimum in* (D.1) *is attained, and for every* $0 < \alpha < 1$,

$$\pi(P, \mathcal{Q}) = \sup_{Q\in\mathrm{conv}(\mathcal{Q})} \frac{1}{2}(a\|p\|_1 + b\|q\|_1 - \|ap - bq\|_1) \le a^\alpha b^{1-\alpha}\rho_\alpha(P; \mathrm{conv}(\mathcal{Q})).$$

Proof The set of test-functions ϕ can be identified with the nonnegative functions in the unit ball Φ of $\mathbb{L}_\infty(\mathfrak{X}, \mathscr{X}, \mu)$, which is dual to $\mathbb{L}_1(\mathfrak{X}, \mathscr{X}, \mu)$, since μ is σ-finite. The set Φ is compact and Hausdorff with respect to the weak*-topology, by the Banach-Alaoglu theorem (cf. Theorem 3.15 of Rudin 1973) and weak*-closure of the set of positive functions. Because the map $(\phi, Q) \mapsto aP\phi + bQ(1-\phi)$ from $\mathbb{L}_\infty(\mathfrak{X}, \mathscr{X}, \mu) \times \mathbb{L}_1(\mathfrak{X}, \mathscr{X}, \mu)$ to $\mathbb{R}$ is convex and weak*-continuous in ϕ and linear in Q, the infimum over ϕ is attained, and the minimax theorem, Theorem L.5, gives

$$\inf_{\phi\in\Phi}\sup_{Q\in\mathrm{conv}(\mathcal{Q})}(aP\phi+bQ(1-\phi))=\sup_{Q\in\mathrm{conv}(\mathcal{Q})}\inf_{\phi\in\Phi}(aP\phi+bQ(1-\phi)).$$

The expression on the left side is the minimax testing risk $\pi(P,\mathcal{Q})$. The infimum on the right side is attained for $\phi=\mathbb{1}\{ap<bq\}$, and the minimal value $aP(ap<bq)+bQ(ap\geq bq)=b\|q\|_1-\int(ap-bq)^-\,d\mu$ can be rewritten as in the equality in the lemma.

For the inequality, we write

$$aP(ap<bq)+bQ(ap\geq bq)=a\int_{ap<bq}p\,d\mu+b\int_{ap\geq bq}q\,d\mu,$$

and bound p in the first integral by $p^\alpha(bq/a)^{1-\alpha}$, and q in the second integral by $(ap/b)^\alpha q^{1-\alpha}$. □

In particular, for probability measures P and Q and $a=b=1$, the minimax risk can be written in the form

$$\pi(P,\mathcal{Q})=1-\tfrac{1}{2}\|P-\mathrm{conv}(\mathcal{Q})\|_1=1-d_{TV}(P,\mathrm{conv}(\mathcal{Q})). \tag{D.3}$$

This exact expression is often difficult to handle. In Section D.4 we shall see that the further bound using the Hellinger transform is easy to manipulate for product measures.

The Hellinger distance is closely related to testing through its link to the affinity (see Lemma B.5). Because $-\log\rho_{1/2}=R_{1/2}\geq d_H^2/2$, the preceding display shows that convex alternatives $\mathcal{Q}$ with $d_H(P,\mathcal{Q})>\epsilon$ can be tested with errors bounded by $e^{-\epsilon^2/2}$. That the Hellinger distance is a Hilbert space norm also makes it manageable for direct constructions of tests. In the following lemma, this is exploited to construct a likelihood ratio test $\phi=\mathbb{1}\{\bar{q}/\bar{p}>c^{-2}\}$ between Hellinger balls around two given densities p and q based on a "least-favorable" pair $\bar{p}$ and $\bar{q}$ of densities. This lemma will be the basis of constructing tests for non-identically distributed observations and Markov chains, later on.

Lemma D.2 (Basic Hellinger testing) *Given arbitrary probability densities p and q, there exist probability densities $\bar{p}$ and $\bar{q}$ such that for, any probability density r,*

$$R\sqrt{\frac{\bar{q}}{\bar{p}}}\leq 1-\frac{1}{6}d_H^2(p,q)+d_H^2(p,r),\qquad R\sqrt{\frac{\bar{p}}{\bar{q}}}\leq 1-\frac{1}{6}d_H^2(p,q)+d_H^2(q,r).$$

Proof The linear subspace $\{a\sqrt{p}+b\sqrt{q}\colon a,b\in\mathbb{R}\}$ of $\mathbb{L}_2(\nu)$ is isometric to $\mathbb{R}^2$ equipped with the inner product $\langle(a,b),(a',b')\rangle_\omega=(a,b)V_\omega(a',b')^\top$, for V_ω the nonnegative-definite matrix, with $\omega\in[0,\pi/2]$ such that $\rho_{1/2}(p;q)=\cos\omega$,

$$V_\omega=\begin{pmatrix}1&\cos\omega\\ \cos\omega&1\end{pmatrix}=L^\top L,\qquad L=\begin{pmatrix}1&\cos\omega\\ 0&\sin\omega\end{pmatrix}.$$

The collection of linear combinations $a\sqrt{p}+b\sqrt{q}$ with $\mathbb{L}_2(\nu)$-norm equal to 1 is represented by the vectors (a,b) such that $\langle(a,b),(a,b)\rangle_\omega=a^2+b^2+2ab\cos\omega=1$. In terms of the Choleski decomposition $V_\omega = L^\top L$ as given in the display this ellipsoid can also be parameterized as

$$\begin{pmatrix} a \\ b \end{pmatrix} = L^{-1}\begin{pmatrix} \cos t \\ \sin t \end{pmatrix} = \frac{1}{\sin\omega}\begin{pmatrix} \sin(\omega - t) \\ \sin t \end{pmatrix}, \qquad 0 \le t \le 2\pi.$$

Denote the element $a\sqrt{p} + b\sqrt{q} \in \mathbb{L}_2(\nu)$ corresponding to t by $g(t)$. Then $g(0) = \sqrt{p}$ and $g(\omega) = \sqrt{q}$, and the inner product $\int g(t_1)g(t_2)\,d\nu$, which is equal to the $\langle\cdot,\cdot\rangle_\omega$ inner product between the corresponding vectors (a_1, b_1) and (a_2, b_2), is given by the Euclidean inner product between the vectors $(\cos t_1, \sin t_1)$ and $(\cos t_2, \sin t_2)$, which is $\cos t_1 \cos t_2 + \sin t_1 \sin t_2 = \cos(t_1 - t_2)$.

We define $\bar{p}$ and $\bar{q}$ by their roots, through $\sqrt{\bar{p}} = g(\omega/3)$ and $\sqrt{\bar{q}} = g(2\omega/3)$, respectively. Then

$$\frac{\sqrt{\bar{q}}}{\sqrt{\bar{p}}} = \frac{g(2\omega/3)}{g(\omega/3)} = \frac{\sin(\omega/3)\sqrt{p} + \sin(2\omega/3)\sqrt{q}}{\sin(2\omega/3)\sqrt{p} + \sin(\omega/3)\sqrt{q}} \le \frac{\sin(2\omega/3)}{\sin(\omega/3)} = 2\cos(\omega/3).$$

The root $\sqrt{r}$ of a general element $r \in \mathbb{L}_2(\mu)$ can be decomposed as its projection onto $\mathrm{lin}(\sqrt{p}, \sqrt{q})$ and an orthogonal part. Because its projection integrates to at most 1, it corresponds to a point inside the ellipsoid and can be represented as $\theta g(\gamma)$, for some $\theta \in [0, 1]$ and $\gamma \in [0, 2\pi]$. It follows that $\int \sqrt{r}\, g(t)\,d\nu = \theta\cos(\gamma - t)$, for any $t \in [0, 2\pi]$.

Now

$$R\sqrt{\frac{\bar{q}}{\bar{p}}} = \int \sqrt{\frac{\bar{q}}{\bar{p}}}(\sqrt{r} - \sqrt{\bar{p}})^2\,d\nu + 2\int \sqrt{\bar{q}}\sqrt{r}\,d\nu - \int \sqrt{\bar{q}}\sqrt{\bar{p}}\,d\nu.$$

Bounding the first term on the right by $2\cos(\omega/3)d_H^2(r, \bar{p})$, using $d_H^2 = 2 - 2\rho_{1/2}$, and expressing the inner products in their angles as found previously, we see that this is bounded above by

$$\begin{aligned} &2\cos(\omega/3)(2 - 2\theta\cos(\gamma - \omega/3)) + 2\theta\cos(\gamma - 2\omega/3) - \cos(\omega/3) \\ &\qquad = 3\cos(\omega/3) - 2\theta\cos\gamma \le 3\Big[\tfrac{8}{9} + \tfrac{1}{9}\cos\omega\Big] - 2\theta\cos\gamma. \end{aligned}$$

Finally we substitute $\cos\omega = \int \sqrt{p}\sqrt{q}\,d\nu = 1 - d_H^2(p, q)/2$, and $\theta\cos\gamma = \int \sqrt{p}\sqrt{r}\,d\nu = 1 - d_H^2(p, r)/2$. This concludes the proof of the first inequality; the second follows by symmetry. □

D.2 Composite Alternatives

Proposition D.1 shows the importance of the convex hull of the alternatives $\mathcal{Q}$. Not the separation of $\mathcal{Q}$ from the null hypothesis, but the separation of its convex hull drives the error probabilities. Unfortunately the complement $\{Q: d(P, Q) > \epsilon\}$ of a ball around P is not convex, and hence even though it is separated from P by distance ϵ, the existence of good tests, even consistent ones in an asymptotic set-up, is not guaranteed for this alternative in general. This is true even for d the total variation distance!

To handle a nonconvex alternative using a metric structure, such an alternative may be covered by convex sets, and corresponding tests combined into a single overall test. The power will then depend on the number of sets needed in a cover, and their separation from the null hypothesis. In the following basic lemma we make this precise in a form suited to posterior analysis. The following definition is motivated by the fact that convex alternatives at Hellinger distance ϵ from the null hypothesis can be tested with error probabilities bounded by $e^{-\epsilon^2/2}$ (see the remarks following (D.3), or Lemma D.2).

Given a semimetric d on a collection of measures that contains null and alternative hypotheses (or an arbitrary function $Q \mapsto d(P, Q)$ from $\mathcal{Q}$ to $[0, \infty)$), and for given positive constants c and K, and every $\epsilon > 0$, define the *covering number for testing* $N_t(\epsilon, \mathcal{Q}, d)$ as the minimal number of sets $\mathcal{Q}_1, \ldots, \mathcal{Q}_N$ in a partition of $\{Q \in \mathcal{Q}: \epsilon < d(P, Q) < 2\epsilon\}$ such that for every partitioning set $\mathcal{Q}_l$ there exists a test ψ_l with

$$P\psi_l \le c\, e^{-K\epsilon^2}, \qquad \sup_{Q \in \mathcal{Q}_l} Q(1-\psi_l) \le c^{-1}\, e^{-K\epsilon^2}. \tag{D.4}$$

Lemma D.3 *If $N_t(\epsilon, \mathcal{Q}, d) \le N(\epsilon)$ for every $\epsilon > \epsilon_0 \ge 0$ and some nonincreasing function $N: (0, \infty) \to (0, \infty)$, then for every $\epsilon > \epsilon_0$ there exists a test ϕ such that, for all $j \in \mathbb{N}$,*

$$P\phi \le c\, N(\epsilon) \frac{e^{-K\epsilon^2}}{1 - e^{-K\epsilon^2}}, \qquad \sup_{Q \in \mathcal{Q}: d(P,Q) > j\epsilon} Q(1-\phi) \le c^{-1} e^{-K\epsilon^2 j^2}.$$

Proof For a given $j \in \mathbb{N}$, choose a minimal partition of $\mathcal{Q}_j := \{Q \in \mathcal{Q}: j\epsilon < d(P, Q) < 2j\epsilon\}$ as in the definition of $N_t(j\epsilon, \mathcal{Q}_j, d)$, and let $\phi_{j,l}$ be the corresponding tests. This gives $N_t(j\epsilon, \mathcal{Q}_j, d) \le N(j\epsilon) \le N(\epsilon)$ tests for every j. Let ϕ be the supremum of the countably many tests obtained in this way, when j ranges over $\mathbb{N}$. Then

$$P\phi \le \sum_{j=1}^{\infty} \sum_l c e^{-Kj^2\epsilon^2} \le c \sum_{j=1}^{\infty} N(\epsilon) e^{-Kj^2\epsilon^2} \le cN(\epsilon) \frac{e^{-K\epsilon^2}}{1 - e^{-K\epsilon^2}}$$

and, for every $j \in \mathbb{N}$,

$$\sup_{Q \in \cup_{i>j} \mathcal{Q}_i} Q(1-\phi) \le \sup_{i>j} c^{-1} e^{-Ki^2\epsilon^2} \le c^{-1} e^{-Kj^2\epsilon^2},$$

since for every $Q \in \mathcal{Q}_i$ there exists a test $\phi_{i,l}$ with $\phi \ge \phi_{i,l}$ that satisfies $Q(1 - \phi_{i,l}) \le c^{-1} e^{-Ki^2\epsilon^2}$ by construction. ☐

In view of Proposition D.1 (applied with $a = c^{-1}$ and $b = c$) the cover $\mathcal{Q}_1, \ldots, \mathcal{Q}_N$ of the alternative $\mathcal{Q}$ in the definition of the testing number $N_t(\epsilon, \mathcal{Q}, d)$ can consist of any sets with

$$\inf_{0<\alpha<1} \left[c^{1-2\alpha} \rho_\alpha(P; \mathrm{conv}(\mathcal{Q}_l))\right] \le e^{-K\epsilon^2}. \tag{D.5}$$

If restricted to probability measures the Hellinger transform satisfies $-\log \rho_{1/2} = R_{1/2} \ge d_H^2/2$, by Lemma B.5 (vi) and (i), and hence the left side of the display is bounded above by $\exp(-d_H^2(P, \mathrm{conv}(\mathcal{Q}_l)/2)$ (use $\alpha = 1/2$). Consequently *convex* sets $\mathcal{Q}_l$ with $d_H(P, \mathcal{Q}_l) \ge \sqrt{2K}\epsilon$ will do.[1]

[1] Because the log affinity responds better to operations such as forming product measures, (D.5) is better left in terms of the affinity than weakened to a bound in terms of the square Hellinger distance.

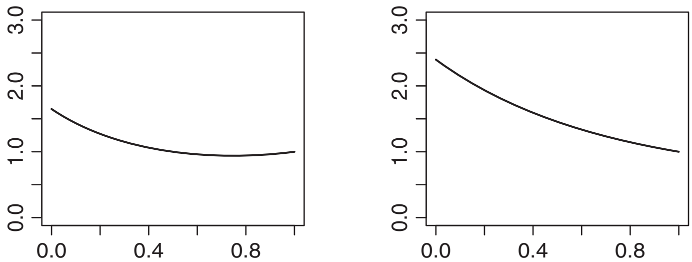

Figure D.1 The Hellinger transforms $\alpha \mapsto \rho_\alpha(p; q)$, for $P = \text{Nor}(0, 2)$ and Q the measure defined by $q = (d\text{Nor}(1, 1)/d\text{Nor}(0, 1))\, p$ (left) and $q = (d\text{Nor}(1, 1)/d\text{Nor}(3/2, 1))\, p$ (right). Intercepts with the vertical axis at the left and right of the graphs equal $\|q\|$ and $\|p\| = 1$, respectively. The slope at 1 equals $K(p; q)$, and has positive and negative in the two cases. In the left side the hypotheses p and q are testable versus each other, but in the right side they are not. (The measure $P^* = N(0, 1)$ is the point in the model minimizing $K(P; P^*)$ over the model $\{\text{Nor}(\theta, 1): \theta \in \mathbb{R}\}$; the slope at 1 is also positive if $N(1, 1)$ is replaced by $\text{Nor}(\theta, 1)$ for arbitrary θ. The measure $\text{Nor}(3/2, 1)$ is also in the model, but not the projection, which implies the existence of θ (such as $\theta = 1$) yielding a graph as on the right.)

By the same lemma, there is no point in using a Hellinger transform ρ_α for a value of α different from $1/2$ if the hypotheses involve only *probability* measures, but a different value $\alpha \in (0, 1)$ can be useful if the measures Q possess total mass bigger than 1. While the Hellinger transform $\rho_\alpha(p; q)$ of two *probability* densities $p \neq q$ is strictly smaller than 1 for any $0 < \alpha < 1$, it may assume values above 1 if q is a general density. For equivalent p and q the Hellinger transform $\alpha \mapsto \rho_\alpha(p; q)$ is a convex function, which takes the values $\|q\|_1$ at $\alpha = 0$ and $\|p\|_1 = 1$ at $\alpha = 1$. If $\|q\|_1 > 1 = \|p\|_1$, then it assumes values strictly smaller than 1 if and only if its left derivative at $\alpha = 1$, which is the Kullback-Leibler divergence $K(p; q)$, is strictly positive (see Lemma B.5). Figure D.1 illustrates that the Hellinger transform may be smaller than one for some α (left panel), or may be bigger than one for any $\alpha \in (0, 1)$. In the second case, the bound on the minimax testing risk given by Proposition D.1 is the useless bound 1.

If $K(p; q) > 0$, then a value of α close to 1 will be appropriate for constructing a test, even if q is not a probability density. By Lemma B.7 (with $p = p_0$), for any set of densities $\mathcal{Q}$,

$$\sup_{q \in \text{conv}(\mathcal{Q})} |1 - \rho_{1-\alpha}(p; q) - \alpha K(p; q)| \lesssim \alpha^2 \sup_{q \in \mathcal{Q}} \Big[d_H^2(p, q) + V_2(p; q) \Big].$$

If the right side can be controlled appropriately this means that the preceding argument can be made uniform in q, yielding that $\rho_\alpha(p; \mathcal{Q}) < 1$ for α sufficiently close to 1. The second inequality of the lemma allows this to extend to the convex hull of $\mathcal{Q}$.

We summarize the preceding in the following theorem.

Theorem D.4 *If there exists a nonincreasing function $N: (0, \infty) \to (0, \infty)$ such that for every $\epsilon > \epsilon_0 \geq 0$ the set $\{Q \in \mathcal{Q}: d(P, Q) < 2\epsilon\}$ can be covered with $N(\epsilon)$ sets $\mathcal{Q}_l$ satisfying* (D.5)*, then for every $\epsilon > \epsilon_0$ there exist a test ϕ such that, for all $j \in \mathbb{N}$,*

$$P\phi \le c\,N(\epsilon)\frac{e^{-K\epsilon^2}}{1-e^{-K\epsilon^2}}, \qquad \sup_{Q\in\mathcal{Q}:d(P,Q)>j\epsilon} Q(1-\phi) \le c^{-1}e^{-K\epsilon^2 j^2}.$$

Proof For $a^{-1} = c = b$, condition (D.5) says that $\inf_{0<\alpha<1}\left[a^\alpha b^{1-\alpha}\rho_\alpha(p, \mathrm{conv}(\mathcal{Q}_l))\right]$ is bounded above by $e^{-K\epsilon^2}$. Hence by Proposition D.1, there exists a test ψ_l satisfying (D.4). The theorem therefore follows from Lemma D.3. □

D.3 Testing and Metric Entropy

If the sets $\mathcal{Q}_l$ are constructed as balls relative to a metric, then the covering number for testing can be bounded by ordinary metric entropy. The following *basic testing assumption*, which will be verified for a number of statistical setups in the next sections, is instrumental. Suppose that d and e are semimetrics such that for universal constants $K > 0$ and $\xi \in (0, 1)$, there exists for every $\epsilon > 0$ and every $Q \in \mathcal{Q}$ with $d(P, Q) > \epsilon$ a test ϕ with

$$P\phi \le c\,e^{-K\epsilon^2}, \qquad \sup_{R\in\mathcal{Q}:e(R,Q)<\xi\epsilon} R(1-\phi) \le c^{-1}\,e^{-K\epsilon^2}. \tag{D.6}$$

In the common case that $d = e$, this requires that the null hypothesis P can be tested against any ball with errors that are exponential in minus a constant times the square distance of the ball to P, as illustrated in Figure D.2. It follows readily from the definitions that in this situation

$$N_t(\epsilon, \mathcal{Q}, d) \le N(\xi\epsilon, \{Q \in \mathcal{Q}: d(P, Q) < 2\epsilon\}, e), \qquad \epsilon > 0.$$

The quantity on the right is known as the *local covering number*, and its logarithm as the *Le Cam dimension* (at level ϵ) of $\mathcal{Q}$. Clearly the local covering number is upper bounded by the ordinary covering numbers $N(\xi\epsilon, \mathcal{Q}, e)$. In infinite-dimensional situations these often have the same order of magnitude as $\epsilon \downarrow 0$, and hence nothing is lost by this simplification. For finite-dimensional models, local covering numbers are typically of smaller order. The difference is also felt in asymptotics where the dimension of the model tends to infinity.

The Le Cam dimension $\log N(\epsilon, \{\theta \in \mathbb{R}^k: \|\theta\| < 2\epsilon\}, \|\cdot\|_2)$ of Euclidean space is bounded by a multiple of its dimension k, in view of Proposition C.2, uniformly in ϵ. This is one motivation for the term "dimension."

The following result is a consequence of the aggregation of basic tests and is the basis for nearly all theorems on posterior consistency and rates of contraction.

Theorem D.5 (Basic testing) *If the basic testing assumption* (D.6) *holds for arbitrary semimetrics d and e and $N(\xi\epsilon, \{Q \in \mathcal{Q}: d(P, Q) < 2\epsilon\}, e) \le N(\epsilon)$ for every $\epsilon > \epsilon_0 \ge 0$*

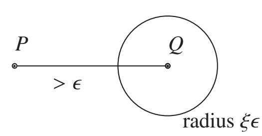

Figure D.2 Assumption (D.6) assumes existence of a test of a simple null hypothesis P against a ball of radius $\xi\epsilon$ with center Q located at distance ϵ from P.

and some nonincreasing function $N: (0, \infty) \to (0, \infty)$, *then for every* $\epsilon > \epsilon_0$ *there exist a test* ϕ *such that, for all* $j \in \mathbb{N}$,

$$P\phi \le c\, N(\epsilon) \frac{e^{-K\epsilon^2}}{1 - e^{-K\epsilon^2}}, \qquad \sup_{Q \in \mathcal{Q}: d(P,Q) > j\epsilon} Q(1-\phi) \le c^{-1} e^{-K\epsilon^2 j^2}.$$

D.4 Product Measures

For $i = 1, \ldots, n$, let P_i and $\mathcal{Q}_i$ be a probability measure and a set of probability measures on an arbitrary measurable space $(\mathfrak{X}_i, \mathscr{X}_i)$, and consider testing the product $\otimes_i P_i$ versus the set $\otimes_i \mathcal{Q}_i$ of products $\otimes_i Q_i$ with Q_i ranging over $\mathcal{Q}_i$.

The Hellinger transform factorizes on pairs of product measures. For set-valued arguments it is defined as a supremum over the convex hull. Then it does not factorize, but as shown in the following lemma, it is still "sub-multiplicative."

Lemma D.6 *For any* $0 < \alpha < 1$ *and any measures* P_i *and any sets* $\mathcal{Q}_i$ *of finite measures* $\rho_\alpha(\otimes_i P_i; \operatorname{conv}(\otimes_i \mathcal{Q}_i)) \le \prod_i \rho_\alpha(P_i; \operatorname{conv}(\mathcal{Q}_i))$.

Proof If suffices to give the proof for $n = 2$; the general case follows by repetition. Any measure $Q \in \operatorname{conv}(\mathcal{Q}_1 \times \mathcal{Q}_2)$ can be represented by a density of the form $q(x, y) = \sum_j \kappa_j q_{1j}(x) q_{2j}(y)$, for nonnegative constants κ_j with $\sum_j \kappa_j = 1$, and q_{ij} densities of measures belong to $\mathcal{Q}_i$. Then $\rho_\alpha(p_1 \times p_2; q)$ can be written in the form

$$\int p_1(x)^\alpha \Big(\sum_j \kappa_j q_{1j}(x)\Big)^{1-\alpha} \Big[\int p_2(y)^\alpha \Big(\frac{\sum_j \kappa_j q_{1j}(x) q_{2j}(y)}{\sum_j \kappa_j q_{1j}(x)}\Big)^{1-\alpha} d\mu_2(y)\Big]\, d\mu_1(x).$$

(If $\sum_j \kappa_j q_{1j}(x) = 0$, the quotient in the inner integral is interpreted as 0.) The inner integral is bounded by $\rho_\alpha(P_2; \operatorname{conv}(\mathcal{Q}_2))$ for every fixed $x \in \mathfrak{X}$. After substitution of this upper bound the remaining integral is bounded by $\rho_\alpha(P_1; \operatorname{conv}(\mathcal{Q}_1))$. $\square$

Combining Lemma D.6 and Proposition D.1 (with $\alpha = 1/2$), we obtain the following corollary.

Corollary D.7 *For any probability measures* P_i *and sets of densities* $\mathcal{Q}_i$ *and* $a, b > 0$ *there exists a test* ϕ *such that*

$$a\big(\bigotimes_{i=1}^n P_i\big)\phi + b \sup_{q_i \in \mathcal{Q}_i} \big(\bigotimes_{i=1}^n Q_i\big)(1-\phi) \le \sqrt{ab} \prod_{i=1}^n \rho_{1/2}(p_i; \operatorname{conv}(\mathcal{Q}_i)).$$

In particular, for identically distributed observations and null and alternative hypotheses P and $\mathcal{Q}$ there exists a test ϕ with (since $\rho_{1/2} \le e^{-d_H^2/2}$)

$$a P^n \phi + b \sup_{Q \in \mathcal{Q}} Q^n(1-\phi) \le \sqrt{ab}\, e^{-n d_H^2(p, \operatorname{conv}(\mathcal{Q}))/2}.$$

This shows that (D.6) is satisfied (with $c = 1$) for collections of product measures equipped with $d = e$ the Hellinger distance on the marginal distributions, with the constant K equal to $n(1-\xi)^2/2$. In the following theorem we record the more general fact that it is valid for any metric that generates convex balls and is bounded above by a multiple of the Hellinger distance. Examples are the total variation distance, and the $\mathbb{L}_2$-distance if the densities are uniformly bounded.

Proposition D.8 *Suppose that d is a semimetric that generates convex balls and satisfies $d(p,q) \le d_H(p,q)$ for every q. Then for every $c > 0$, $n \in \mathbb{N}$ and $\epsilon > 0$ and probability densities p and q with $d(p,q) > \epsilon$, there exists a test ϕ such that*

$$P^n\phi \le c\, e^{-n\epsilon^2/8}, \qquad \sup_{r: d(r,q)<\epsilon/2} R^n(1-\phi) \le c^{-1}\, e^{-n\epsilon^2/8}.$$

Proof If $d(p,q) > \epsilon$, then the d-ball B of radius $\epsilon/2$ around q has d-distance and hence Hellinger distance at least $\epsilon/2$ to p. Consequently $-\log \rho_{1/2}(p; B) \ge d_H^2(p,B)/2 \ge \epsilon^2/8$, by Lemma B.5(vi) and (i). Because B is convex, the result follows from the display preceding the statement of the lemma, applied with $\sqrt{b/a} = c$. ☐

By the same arguments the general result (D.6) can be verified relative to the Hellinger distance on general product measures, with possibly different components. However, in this situation a more natural distance is the *root average square Hellinger metric*

$$d_{n,H}\Big(\bigotimes_{i=1}^n p_i, \bigotimes_{i=1}^n q_i\Big) = \sqrt{\frac{1}{n}\sum_{i=1}^n d_H^2(p_i, q_i)}.$$

By Lemma B.8 (iv) this is an upper bound on the Hellinger distance on the product measures, but the two distances are not equivalent. The following lemma nevertheless guarantees the existence of tests as in (D.6).

Proposition D.9 *For every $c > 0$, $n \in \mathbb{N}$ and $\epsilon > 0$ and probability densities p_i and q_i with $d_{n,H}(\otimes_{i=1}^n p_i, \otimes_{i=1}^n q_i) > \epsilon$, there exists a test ϕ such that*

$$\Big(\bigotimes_{i=1}^n P_i\Big)\phi \le c\, e^{-n\epsilon^2/8}, \qquad \sup_{\bigotimes_{i=1}^n r_i: d_{n,H}(\bigotimes_{i=1}^n r_i, \bigotimes_{i=1}^n q_i)<\epsilon/5} \Big(\bigotimes_{i=1}^n R_i\Big)(1-\phi) \le c^{-1}\, e^{-n\epsilon^2/8}.$$

Proof Let $\phi = \mathbb{1}\{\otimes_{i=1}^n \sqrt{\bar q_i/\bar p_i} > c^{-1}\}$, for $\bar p_1, \ldots, \bar p_n$ and $\bar q_1, \ldots, \bar q_n$ the densities attached to $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ in Lemma D.2. By Markov's inequality

$$\Big(\bigotimes_{i=1}^n P_i\Big)\phi \le c \prod_{i=1}^n P_i \sqrt{\frac{\bar q_i}{\bar p_i}} \le c \prod_{i=1}^n \big(1 - \tfrac{1}{6} d_H^2(p_i, q_i)\big) \le c e^{-n d_{n,H}^2(\otimes p_i, \otimes q_i)/6},$$

$$\Big(\bigotimes_{i=1}^n R_i\Big)(1-\phi) \le c^{-1} \prod_{i=1}^n R_i \sqrt{\frac{\bar p_i}{\bar q_i}} \le c^{-1} \prod_{i=1}^n \big(1 - \tfrac{1}{6} d_H^2(p_i, q_i) + d_H^2(r_i, q_i)\big)$$
$$\le c^{-1} e^{-n d_{n,H}^2(\otimes p_i, \otimes q_i)/6 + n d_{n,H}^2(\otimes r_i, \otimes q_i)}.$$

For $d_{n,H}(\otimes p_i, \otimes q_i) > \epsilon$ and $d_{n,H}(\otimes r_i, \otimes q_i) < \epsilon/\sqrt{24}$, the right sides are bounded as in the lemma (as $1/6 - 1/24 = 1/8$). ☐

Corollary D.7 is in terms of the affinity or Hellinger distance between the null hypothesis and the *convex hull* of the alternative hypothesis. Analytic computations on this hull may be avoided by constructing tests of the hypotheses. Tests with uniform error probabilities automatically refer to the convex hull, and may be suggested by statistical reasoning. In view of the characterization of the minimax testing risk in terms of the $\mathbb{L}_1$-norm (see Theorem D.1) and the relation between this norm and the affinity (see Lemma B.5(ii)) these two approaches are equivalent. The next lemma makes this explicit.

Lemma D.10 (Aggregating tests) *If for each $i = 1, \ldots, n$, there exists a test ϕ_i such that*

$$P_i\phi_i \leq \alpha_i \leq \gamma_i \leq \inf_{Q_i \in \mathcal{Q}_i} Q_i\phi_i, \tag{D.7}$$

then for $\epsilon^2 = n^{-1}\sum_{i=1}^n (\gamma_i - \alpha_i)^2$ there exist a test ϕ such that $(\otimes_{i=1}^n P_i)\phi < e^{-n\epsilon^2/2}$ and $(\otimes_{i=1}^n Q_i)(1-\phi) < e^{-n\epsilon^2/2}$, for every $Q_i \in \mathcal{Q}_i$.

Proof The condition implies that $\|p_i - q_i\|_1/2 = d_{TV}(P_i, Q_i) \geq \gamma_i - \alpha_i$, for every $Q_i \in \mathrm{conv}(\mathcal{Q}_i)$. By Lemma B.5(ii), $\rho_{1/2}(p_i; q_i)^2 \leq 1 - \|p_i - q_i\|_1^2/4 \leq 1 - (\gamma_i - \alpha_i)^2$, for any $q_i \in \mathrm{conv}(\mathcal{Q}_i)$. Next apply Lemma D.6 to see that $\rho_{1/2}(\otimes_i P_i; Q)^2 \leq e^{-n\epsilon^2}$ for any $Q \in \mathrm{conv}(\otimes_i Q_i)$, and finally apply Proposition D.1 (with $a = b = 1$ and $\alpha = 1/2$). ☐

In the case of i.i.d. observations there is nothing special about exponentially small error probabilities: if a fixed set $\mathcal{Q}$ can be uniformly *consistently* tested versus P, then it can automatically be tested with exponentially small error probabilities.

Lemma D.11 (Fixed alternative) *If there exist tests ψ_n such that, for a given probability measure P and a set $\mathcal{Q}$ of probability measures, $P^n\psi_n \to 0$ and $\sup_{Q\in\mathcal{Q}} Q^n(1-\psi_n) \to 0$, then there exist tests ϕ_n and a constant $K > 0$ such that*

$$P^n\phi_n \leq e^{-Kn}, \qquad \sup_{Q\in\mathcal{Q}} Q^n(1-\phi_n) \leq e^{-Kn}.$$

Proof For any $0 < \alpha < \gamma < 1$, there exists n_0 such that $P^{n_0}\psi_{n_0} < \alpha < \gamma < Q^{n_0}\psi_{n_0}$. We apply Lemma D.10 with P_i and Q_i of the lemma equal to the present P^{n_0} and Q^{n_0}. For a given n we can construct $\lfloor n/n_0 \rfloor \asymp n$ blocks of size n_0, and obtain a test with error probabilities that are exponential in minus $\lfloor n/n_0\rfloor(\gamma - \alpha)^2/2$. ☐

D.5 Markov Chains

Let p be a (Markov) transition density $(x, y) \mapsto p(y|\, x)$ and $\mathcal{Q}$ a collection of Markov transition densities $(x, y) \mapsto q(y|\, x)$ from a given sample space $(\mathfrak{X}, \mathscr{X})$ into itself, relative to a dominating measure ν. Consider testing the distribution of a Markov chain $X_0, X_1, \ldots, X_n$ that evolves according to either p or some $q \in \mathcal{Q}$. Let an initial value X_0 be distributed according to a measure P_0 or Q_0, which may or may not be stationary distributions to the transition kernels p and q. Denote the laws of $(X_1, \ldots, X_n)$ under the null and alternative hypotheses by $P^{(n)}$ and $Q^{(n)}$.

The first result is a generalization of Lemma D.6. For given x let $\rho_\alpha(p; q|\,x)$ denote the Hellinger affinity between the probability densities $p(\cdot|\,x)$ and $q(\cdot|\,x)$.

Lemma D.12 *For any* $0 < \alpha < 1$ *and any transition density* p *and class* $\mathcal{Q}$ *of transition densities,*

$$\rho_\alpha(P^{(n)}; \mathrm{conv}(\mathcal{Q}^{(n)})) \le \Big(\sup_{x\in\mathfrak{X}} \sup_{q\in\mathrm{conv}(\mathcal{Q})} \rho_\alpha(p; q|\,x)\Big)^n.$$

Proof The proof can evolve as the proof of Lemma D.6, where we peel off observations in the order $X_n, X_{n-1}, \ldots, X_0$, each time bounding the integral

$$\int p^\alpha(x_i|\,x_{i-1})\Big(\frac{\sum_j \kappa_j q_j^{(i-1)}(x_1,\ldots,x_{i-1}) q_j(x_i|\,x_{i-1})}{\sum_j \kappa_j q_j^{(i-1)}(x_1,\ldots,x_{i-1})}\Big)^{1-\alpha}\,d\nu(x_i)$$

by the supremum over the convex hull of the transition densities $y \mapsto q(y|\,x_{i-1})$, and next the supremum over x_{i-1}, of the Hellinger affinity, yielding the n terms in the product on the right side. □

Combining Lemma D.12 and Proposition D.1, we can obtain the analog of Corollary D.7. This may next be translated into the following testing statement, which uses the *supremum Hellinger distance*, defined by

$$d_{H,\infty}(p,q) = \sup_{x\in\mathfrak{X}} d_H(p(\cdot|\,x), q(\cdot|\,x)).$$

Corollary D.13 *For any transition density* p *and convex set of transition densities* $\mathcal{Q}$ *with* $\inf_{x\in\mathfrak{X}} \inf_{q\in\mathcal{Q}} d_H^2(p(\cdot|\,x), q(\cdot|\,x)) \ge \epsilon$*, any initial distributions* P_0 *and* Q_0*, and any* $a, b > 0$*, there exists a test* ϕ *such that*

$$aP^{(n)}\phi + b\sup_{q\in\mathcal{Q}} Q^{(n)}(1-\phi) \le \sqrt{ab}\,\epsilon^{-n\epsilon^2/2}.$$

Furthermore, if $\mathcal{Q}$ *is not convex, then there still exists a test satisfying the preceding display with the right-hand side replaced by* $N(\epsilon/4, \mathcal{Q}, d_{H,\infty})e^{-n\epsilon^2/2}$.

The supremum Hellinger distance may be too strong, and the condition that the infimum of the Hellinger distances is bounded away from zero unworkable. An alternative is a *weighted Hellinger distance* of the form, for some measure μ,

$$d_{H,\mu}(q_1,q_2) = \Big(\int\!\!\int \Big(\sqrt{q_1(y|\,x)} - \sqrt{q_2(y|\,x)}\Big)^2\,d\nu(y)\,d\mu(x)\Big)^{1/2}. \tag{D.8}$$

We can use these distances for μ bounding the transition probabilities of the Markov chain as follows. For $n \in \mathbb{N}$ let $(x, A) \mapsto Q^n(A|\,x)$ be the n-step transition kernel, given recursively by

$$Q^1(A|\,x) = \int_A q(y|\,x)\,d\nu(y), \qquad Q^{n+1}(A|\,x) = \int Q^n(A|\,y)\,Q(dy|\,x).$$

Then assume that there exist some $k, l \in \mathbb{N}$ and measures $\underline{\mu}, \overline{\mu}$ such that for every element of $\{P\} \cup \mathcal{Q}$, every $x \in \mathfrak{X}$ and every $A \in \mathscr{X}$,

$$\underline{\mu}(A) \leq \frac{1}{k}\sum_{j=1}^{k} Q^j(A|\,x), \qquad Q^l(A|\,x) \leq \overline{\mu}(A). \tag{D.9}$$

This condition requires that the transitions out of the possible states x occur with a certain uniformity in the initial state, captured by the measures $\underline{\mu}$ and $\overline{\mu}$.

The following lemma, due to Birgé (1983a,b), shows that tests satisfying (D.6) exist for the semimetrics d_n and e_n equal to the weighted Hellinger distances $d_{H,\underline{\mu}}$ and $d_{H,\overline{\mu}}$, respectively.

Proposition D.14 *There exist a constant K depending only on (k, l) such that for every transition kernels p and q with $d_{H,\underline{\mu}}(p, q) > \epsilon$ there exist tests ϕ such that, for every $n \in \mathbb{N}$,*

$$P^{(n)}\phi \leq e^{-Kn\epsilon^2}, \qquad \sup_{r:d_{H,\overline{\mu}}(r,q)\leq\epsilon/5} R^{(n)}(1-\phi) \leq e^{-Kn\epsilon^2}.$$

Proof For $m = k + l$ partition the n observations in $N = \lfloor n/m \rfloor$ blocks of m consecutive observations, and a remaining set of observations, which are discarded. Let $I_1, \ldots, I_N$ be independent random variables, with I_j uniformly distributed on the set of the k last indices of the jth block: $\{(j-1)m+l+1, (j-1)m+l+2, \ldots, (j-1)m+l+k\}$. Then for $\bar{p}(\cdot|\,x)$ and $\bar{q}(\cdot|\,x)$ the densities attached to $p(\cdot|\,x)$ and $q(\cdot|\,x)$ in Lemma D.2, for every given x, define $\phi = \mathbb{1}\{\sum_{j=1}^N \log(\bar{q}/\bar{p})(X_{I_j}|\,X_{I_{j-1}}) > -\log(2c)\}$.

By Markov's inequality

$$P^{(n)}\phi \leq cP^{(n)}\Big[\prod_{j=1}^{N}\sqrt{\frac{\bar{q}}{\bar{p}}}(X_{I_j}|\,X_{I_j-1})\Big], \quad R^{(n)}(1-\phi) \leq c^{-1}R^{(n)}\Big[\prod_{j=1}^{N}\sqrt{\frac{\bar{p}}{\bar{q}}}(X_{I_j}|\,X_{I_j-1})\Big].$$

We evaluate these expressions by peeling off the terms of the products one-by-one from the right, each time conditioning on the preceding observations. For the error probability of the second kind, the peeling step proceeds as

$$\begin{aligned}
&\mathrm{E}_R\Big(\sqrt{\frac{\bar{p}}{\bar{q}}}(X_{I_j}|\,X_{I_j-1})|\,X_1, \ldots, X_{(j-1)m}\Big)\\
&= \frac{1}{k}\sum_{i=1}^{k}\int\!\!\int \sqrt{\frac{\bar{p}}{\bar{q}}}(y|\,x)\,r(y|\,x)\,d\nu(x)\,R^{l+i-1}(dx|\,X_{(j-1)m})\\
&\leq \frac{1}{k}\sum_{i=1}^{k}\Big(1-\tfrac{1}{6}d_H^2(p(\cdot|\,x), q(\cdot|\,x)) + d_H^2(r(\cdot|\,x), q(\cdot|\,x))\Big)\,R^{l+i-1}(dx|\,X_{(j-1)m}),
\end{aligned}$$

by Lemma D.2. By the Chapman-Kolmogorov equations and the bounds (D.9) on the transition kernels $R^{l+i}(A|\,x) = \int R^l(A|\,y)\,R^i(dy|\,x) \leq \overline{\mu}(A)$, for every $i \geq 0$; and also

$$k^{-1}\sum_{i=1}^{k} R^{l+i-1}(A|\,x) = k^{-1}\sum_{i=1}^{k}\int R^i(A|\,y)\,R^{l-1}(dy|\,x) \geq \underline{\mu}(A).$$

Therefore, the previous display is bounded above by $1 - \frac{1}{6}d^2_{H,\underline{\mu}}(p,q) + d^2_{H,\overline{\mu}}(r,q) < 1 - \epsilon^2/8$, for $d_{H,\underline{\mu}}(p,q) > \epsilon$ and $d_{H,\overline{\mu}}(r,q) < \epsilon/\sqrt{24}$. Peeling off all N terms in this manner, we obtain the upper bound $(1-\epsilon^2/8)^N$, which is bounded as desired.

The probability of an error of the first kind can be handled similarly. □

D.6 Gaussian Time Series

Let $P^{(n)}_f$ denote the distribution of $(X_1,\ldots,X_n)$ for a stationary Gaussian time series $(X_t\colon t\in\mathbb{Z})$ with mean zero and spectral density f. For $k\in\mathbb{Z}$ let $\gamma_f(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\,d\lambda$ define the corresponding autocovariance function.

The following proposition, whose proof can be found in Birgé (1983a,b), shows that (D.6) is satisfied under some restrictions, for d equal to the $\mathbb{L}_2[0,\pi]$-metric and e equal to the uniform metric.

Proposition D.15 *Let $\mathcal{F}$ be a set of measurable functions $f\colon[0,\pi]\to[0,\infty)$ such that $\|\log f\|_\infty \le M$ and $\sum_{h\in\mathbb{Z}}|h|\gamma_f^2(h)\le N$ for all $f\in\mathcal{F}$. Then there exist constants ξ and K depending only on M and N such that for every $\epsilon\ge n^{-1/2}$ and every $f_0, f_1\in\mathcal{F}$ with $\|f_1-f_0\|_2\ge\epsilon$,*

$$P^{(n)}_{f_0}\phi_n \le e^{-Kn\epsilon^2}, \qquad \sup_{f\in\mathcal{F}:\|f-f_1\|_\infty\le\xi\epsilon} P^{(n)}_f(1-\phi_n)\le e^{-Kn\epsilon^2}.$$

D.7 Gaussian White Noise

For $\theta\in\Theta\subset\mathbb{L}_2[0,1]$, let $P^{(n)}_\theta$ be the distribution on $\mathfrak{C}[0,1]$ of the stochastic process $X^{(n)} = (X^{(n)}_t\colon 0\le t\le 1)$ defined structurally relative to a standard Brownian motion W as

$$X^{(n)}_t = \int_0^t \theta(s)\,ds + \frac{1}{\sqrt{n}}W_t.$$

An equivalent experiment is obtained by expanding $dX^{(n)}$ on an arbitrary orthonormal basis $e_1, e_2, \ldots$ of $\mathbb{L}_2[0,1]$, giving the random vector $X_n = (X_{n,1}, X_{n,2},\ldots)$, for $X_{n,i} = \int e_i(t)\,dX^{(n)}_t$. The vector X_n is sufficient in the experiment consisting of observing $X^{(n)}$. Its coordinates are independent and normally distributed with means the coefficients $\theta_i := \int e_i(t)\theta(t)\,dt$ of θ relative to the basis and variance $1/n$.

Lemma D.16 *For any $\theta_0,\theta_1\in\Theta$ with $\|\theta-\theta_1\|\ge\epsilon$ the test $\phi_n = \mathbb{1}\{2\langle\theta_1-\theta_0, X^{(n)}\rangle > \|\theta_1\|^2-\|\theta_0\|^2\}$ satisfies*

$$P^{(n)}_{\theta_0}\phi_n \le 1-\Phi(\sqrt{n}\epsilon/2)\le e^{-n\epsilon^2/8},$$
$$\sup\nolimits_{\theta:\|\theta-\theta_1\|<\epsilon/4} P^{(n)}_\theta(1-\phi_n)\le 1-\Phi(\sqrt{n}\epsilon/4)\le e^{-n\epsilon^2/32}.$$

Proof The test ϕ_n rejects the null hypothesis for positive values of the statistic $T_n = \langle\theta_1-\theta_0, dX^{(n)}\rangle - \frac{1}{2}\|\theta_1\|^2+\frac{1}{2}\|\theta_0\|^2$. Under $P^{(n)}_\theta$ this possesses a normal distribution with mean $\langle\theta_1-\theta_0,\theta-\theta_1\rangle+\frac{1}{2}\|\theta_1-\theta_0\|^2$ and variance $\|\theta_1-\theta_0\|^2/n$. Under $P^{(n)}_{\theta_0}$ the mean is

$-\frac{1}{2}\|\theta_0 - \theta_1\|^2$, giving $P_{\theta_0}^{(n)}(T_n < 0) \le \Phi(-\sqrt{n}\|\theta_1 - \theta_0\|/2)$, whereas by the Cauchy-Schwarz inequality under $P_\theta^{(n)}$ for $\|\theta - \theta_1\| \le \|\theta_1 - \theta_0\|/4$, the mean is greater than $\|\theta_0 - \theta_1\|^2/4$, giving $P_\theta^{(n)}(T_n < 0) \le 1 - \Phi(\sqrt{n}\|\theta_1 - \theta_0\|/4)$. The second inequalities follow by Lemma K.6. □

D.8 Historical Notes

Proposition D.1 is due to Le Cam (1986). The form of Lemma D.2 presented here was personally communicated by Birgé, and is modified from Birgé (1979) and Birgé (1983b). The combining technique of Lemma D.3 is by Le Cam (1986). The generalization to non-probability measures for the use in misspecified models presented in Theorem D.4 is due to Kleijn and van der Vaart (2006). Lemma D.6 and Proposition D.6 are due to Le Cam (1986). Lemma D.9 is due to Birgé (1979) and Birgé (1983b). Lemma D.11 was observed by Le Cam (1986). Proposition D.14 is due to Birgé (1983b) and Proposition D.15 is obtained from Birgé (1983a).

Problems

D.1 (Ghosal et al. 1999b) Use Hoeffding's inequality to show that for every pair P_0 and P_1 of probability measures there exist tests ϕ_n such that (D.6) holds for the product measures P_0^n and P^n with $d = e$ the total variation distance d_{TV} on the marginal distributions. Identify the resulting constants.

D.2 Extend Lemma D.11 by weakening the assumption to the following condition: For some $m \in \mathbb{N}$, there exists a test $\phi = \phi(X_1, \ldots, X_m)$ such that

$$\sup_{P \in \mathcal{P}_0} P^m \phi + \sup_{P \in \mathcal{P}_1} P^m (1 - \phi) < 1.$$