Chapter 4

Lower bounds for testing

4.1 Statistical tests in the white noise model

In this chapter we consider testing problems in the context of the signal in white noise model (2.1). Specifically, we assume that we observe a stochastic process $X = (X_t : t \in [0,1])$ satisfying

$$dX_t = f(t)\,dt + \frac{1}{\sqrt{n}} dW_t$$

for some $f \in L^2[0,1]$. We mostly consider tests for a simple null hypothesis of the form $H_0 : f = f_0$ for some fixed $f_0 \in L^2[0,1]$, against an alternative H_1 which can be either simple or composite. Formally, a (non-randomized) test in this context is a $\{0,1\}$-valued random variable φ that is a function of the observation X. The variable $\varphi = \varphi(X)$ should simply be viewed as a decision rule. If we observe X, then $\varphi(X)$ says whether we should reject the null or not. If $\varphi(X) = 0$ we don't reject H_0, and if $\varphi(X) = 1$, then we reject H_0 and accept H_1. If the hypotheses are of the form $H_0 : f = f_0$ and $H_1 : f \in \mathcal{F}$, then in this formalism for describing tests, the error of the first kind, i.e. the probability that H_0 is rejected while it is true, is equal to $\mathrm{E}_{f_0}\varphi$. Similarly, the error of the second kind is $\sup_{f\in\mathcal{F}} \mathrm{E}_f(1-\varphi)$. The total error of the test, or simply the *error* of the test φ is defined as the sum $\mathrm{E}_{f_0}\varphi + \sup_{f\in\mathcal{F}} \mathrm{E}_f(1-\varphi)$ of the two types of errors. Recall that for $\alpha \in (0,1)$ a test is said to be of *level* α if the error of the first kind is bounded by α. In the case of a simple null as above this means that $\mathrm{E}_{f_0}\varphi \le \alpha$.

We will be interested in the existence of *consistent tests*. A sequence of tests φ_n is called consistent if the total error of the tests, as just defined, vanishes as $n \to \infty$. Existence of a consistent test means that asymptotically, we can distinguish observations from the null and from the alternative hypothesis without error. If, on the other hand, the error of a sequence of tests is bounded away from 0, then using these tests it is not possible to distinguish the hypotheses. If that holds uniformly for all tests φ, then there does not exists any test that can asymptotically discriminate between the hypotheses without error.

The fundamental difficulty of the testing problem $H_0 : f = f_0$ against $H_1 : f \in \mathcal{F}$ can therefore be assessed by studying lower bounds for the *minimax risk* for testing f_0 against $\mathcal{F}$, defined as

$$\inf_{\varphi} \Big(\mathrm{E}_{f_0}\varphi + \sup_{f\in\mathcal{F}} \mathrm{E}_f(1-\varphi) \Big),$$

where the infimum is over all tests φ. The larger $\mathcal{F}$ is, the larger this quantity becomes. It is also intuitively clear that the 'closer' $\mathcal{F}$ is to f_0, the more difficult the testing problem is. In this chapter we derive results that make these intuitive statements mathematically precise and that allow us to study the difficulty of certain specific testing problems in the white noise model.

Testing problems of the form $H_0 : f = 0$ against $H_1 : f \in \mathcal{F}$ can be viewed as signal detection problems. We will see that for a consistent test to exist for such a problem, the set of signals $\mathcal{F}$ has to be sufficiently separated from 0. The idea is simply that for consistent testing for the presence of a signal to be possible, the signal should be 'large' or 'strong' enough in some sense. The concrete form that this requirement can take depends on the type of signals that are considered. In this chapter we will consider detection of Hölder or Sobolev smooth signals. It turns out that the strength of the signal that is necessary for consistent testing to be possible in these cases, depends on the degree of smoothness and on the type of norm that is considered to quantify the strength.

4.2 Two simple hypotheses

Suppose we want to test the two simple hypotheses $H_0 : f = f_0$ and $H_1 : f = f_1$ against each other. There exists a consistent sequence of tests for these hypotheses if and only if the infimum over all possible tests φ of the total error $\mathrm{E}_{f_0}\varphi + \mathrm{E}_{f_1}(1-\varphi)$ vanishes as $n \to \infty$. Using the Neyman-Pearson lemma we can actually compute the minimax risk for testing f_0 against f_1 explicitly.

We first recall the Neyman-Pearson lemma, which takes the following form in the white noise model.

Theorem 4.2.1 (Neyman-Pearson)**.** *Consider an observation X from the white noise model (2.1) and two simple hypotheses $H_0 : f = f_0$, $H_1 : f = f_1$, for $f_0, f_1 \in L^2[0,1]$. For $\alpha \in (0,1)$, the most powerful test φ of level α, i.e. the test of level α for which $\mathrm{E}_{f_1}\varphi$ is maximal, is the likelihood ratio test*

$$\varphi = 1_{p_{f_1}(X)/p_{f_0}(X) > c_\alpha}, \tag{4.1}$$

where p_f is the density given by (2.7) and

$$c_\alpha = \exp\Big(\sqrt{n}\|f_1 - f_0\|_2 \xi_{1-\alpha} - \frac{1}{2} n \|f_1 - f_0\|_2^2\Big),$$

with $\xi_{1-\alpha}$ the $1-\alpha$ quantile of the standard normal distribution.

Proof. We first note that the test indeed has level α. To see this, we use the expression (2.6) for the likelihood ratio. It implies that for W a P_{f_0}-Brownian motion,

$$\begin{aligned}\mathrm{E}_{f_0}\varphi &= \mathrm{P}_{f_0}\Big(e^{\sqrt{n}\int(f_1-f_0)\,dW-\frac{1}{2}n\|f_1-f_0\|_2^2} > c_\alpha\Big)\\ &= \mathrm{P}_{f_0}\Big(\int(f_1-f_0)\,dW > \|f_1-f_0\|_2\xi_{1-\alpha}\Big)\\ &= \alpha,\end{aligned}$$

where the last equality follows from the properties of the Wiener integral (see Theorem 2.2.2).

Now let φ' be another test of level α. Then by separately considering the events that $\varphi = 1$ and $\varphi = 0$ we see that it almost surely holds that

$$(\varphi' - \varphi)(c_\alpha p_{f_0}(X) - p_{f_1}(X)) \geq 0.$$

By taking the expectation under P_0 and using the fact that $p_f = d\mathrm{P}_f/d\mathrm{P}_0$ it follows that

$$c_\alpha(\mathrm{E}_{f_0}\varphi' - \mathrm{E}_{f_0}\varphi) - (\mathrm{E}_{f_1}\varphi' - \mathrm{E}_{f_1}\varphi) \geq 0.$$

Rearranging this and using that $\mathrm{E}_{f_0}\varphi' \leq \alpha = \mathrm{E}_{f_0}\varphi$ we obtain the desired inequality $\mathrm{E}_{f_1}\varphi' \leq \mathrm{E}_{f_1}\varphi$. □

We can now compute the minimax risk for testing two simple hypothesis against each other. As usual, Φ denotes the standard normal distribution function.

Proposition 4.2.2. *For all $f_0, f_1 \in L^2[0,1]$,*

$$\frac{1}{2}\inf_\varphi(\mathrm{E}_{f_0}\varphi + \mathrm{E}_{f_1}(1-\varphi)) = 1 - \Phi\Big(\frac{1}{2}\sqrt{n}\|f_1-f_0\|_2\Big),$$

where the infimum is over all tests φ.

Proof. By writing the collection of all tests as a union of collections of tests of a fixed level we see that

$$\inf_\varphi(\mathrm{E}_{f_0}\varphi + \mathrm{E}_{f_1}(1-\varphi)) = \inf_{\alpha\in(0,1)}\Big(\alpha + \inf_{\varphi \text{ of level } \alpha}\mathrm{E}_{f_1}(1-\varphi)\Big).$$

By the Neyman-Pearson lemma, the most powerful test of level α is the likelihood ratio test (4.1). For that test we have, by (2.6) and with W a P_{f_1}-Brownian motion, that

$$\begin{aligned}&\mathrm{E}_{f_1}(1-\varphi) = \mathrm{P}_{f_1}(p_{f_0}(X)/p_{f_1}(X) \geq 1/c_\alpha)\\ &= \mathrm{P}_{f_1}(\log(p_{f_0}(X)/p_{f_1}(X)) \geq -\log c_\alpha)\\ &= \mathrm{P}_{f_1}\Big(\sqrt{n}\int(f_0-f_1)\,dW \geq n\|f_1-f_0\|_2^2 - \xi_{1-\alpha}\sqrt{n}\|f_1-f_0\|_2\Big)\\ &= 1 - \Phi\Big(\sqrt{n}\|f_1-f_0\|_2 - \xi_{1-\alpha}\Big).\end{aligned}$$

Hence we obtain

$$\inf_{\varphi}(\mathrm{E}_{f_0}\varphi + \mathrm{E}_{f_1}(1-\varphi)) = \inf_{\alpha\in(0,1)} \Big(\alpha + 1 - \Phi\Big(\sqrt{n}\|f_1 - f_0\|_2 - \xi_{1-\alpha}\Big)\Big).$$

The proof is completed by minimizing over α (see Exercise 4.1). □

The proposition implies that two simple hypotheses corresponding to functions f_0 and f_1 are asymptotically distinguishable by a consistent sequence of tests if and only if $\sqrt{n}\|f_1 - f_0\|_2 \to \infty$. The next examples treat some concrete testing problems.

Example 4.2.3 (Detecting a spike at a fixed point)**.** Suppose we want to test $H_0 : f = 0$ against $H_1 : f = f_\sigma$, where f_σ is a 'spike' at $1/2$ given by

$$f_\sigma(x) = K\Big(\frac{x - 1/2}{\sigma}\Big), \qquad x \in [0, 1],$$

where K is the tent-shaped function $K(x) = (1 - |x|)1_{|x|\le 1}$ and $\sigma > 0$ is a parameter that determines how narrow the spike is. For the L^2-norm of f_σ we have

$$\|f_\sigma\|_2^2 = \int_0^1 K^2\Big(\frac{x-1/2}{\sigma}\Big)\,dx = \sigma \int_{-1/(2\sigma)}^{1/(2\sigma)} K^2(x)\,dx.$$

Due to the compactness of the support of K, $\|f_\sigma\|_2^2$ behaves like a constant times σ for $\sigma \to 0$, hence $\sqrt{n}\|f_\sigma\|_2 \sim \text{const} \times \sqrt{n\sigma}$. For the existence of a consistent test for the presence of the spike this quantity should tend to $+\infty$. We see that this is the case if and only if $n\sigma \to \infty$. Hence, a spike with compact support is only detectable in Gaussian white noise if its width is of larger order than the square $1/n$ of the 'magnitude' of the noise. ■

Whether certain smooth bumps can be detected in Gaussian white noise depends on the relation between the smoothness and the size of the bump. The smoother the bump, the easier it is to detect it.

Example 4.2.4 (Detecting a smooth bump at a fixed point)**.** Consider the function

$$f_\sigma(x) = \sigma^\beta K\Big(\frac{x - 1/2}{\sigma}\Big), \qquad x \in [0, 1],$$

where K is C^∞ and has compact support, and $\beta, \sigma > 0$. This function has C^β-norm of the order 1, in the sense that the supremum of the C^β-norm over bounded σ is bounded, and its squared L^2-norm of the order $\sigma^{1+2\beta}$ (see Exercise 4.2). So the spike is consistently detectable if and only if $n\sigma^{1+2\beta} \to \infty$, that is, if and only the height $f_\sigma(1/2)$ of the bump is of of larger order than $n^{-\beta/(1+2\beta)}$.
■

The last example implies that we are testing $H_0 : f = 0$ against an alternative that contains a smooth bump at a fixed location with height bounded by $n^{-\beta/(1+2\beta)}$, then no consistent test exists. See Exercise 4.3 for an example.

4.3 Composite alternative hypothesis

In this section we consider the problem of testing a simple null hypotheis $H_0 : f = f_0$ against a composite alternative of the form $H_1 : f \in \{f_1, \dots, f_k\}$. The error of a test φ is in this case given by $\mathrm{E}_{f_0}\varphi + \max_{i=1,\dots,k} \mathrm{E}_{f_i}(1-\varphi)$. Again we are interested in conditions for the existence or non-existence of consistent tests. It is in the present setting no longer possible to derive an explicit expression for the minimax testing risk. It is however still possible to derive useful sufficient conditions.

It is intuitively clear perhaps that it should be possible to relate the difficulty of the testing problem to the average of the likelihood ratios $d\mathrm{P}_{f_i}/d\mathrm{P}_{f_0}$. Indeed, if this average is 'large', then one of the likelihoods $d\mathrm{P}_{f_i}/d\mathrm{P}_{f_0}$ must be large as well. This means that data arising from the model with signal f_0 can with substantial probability be generated by the model with signal f_i too, and hence the two hypotheses will be difficult to distinguish.

The following lemma make this reasoning precise and asserts that consistent testing is impossible if under the null, the average likelihood is bounded away from 0 with sufficient probability.

Lemma 4.3.1. *Let $f_0, f_1, \dots, f_k \in L^2[0,1]$. If for some $c, \varepsilon \in (0,1)$, independent of n, it holds that*

$$\mathrm{P}_{f_0}\Big(\frac{1}{k}\sum_{i=1}^{k}\frac{d\mathrm{P}_{f_i}}{d\mathrm{P}_{f_0}} > c\Big) \geq \varepsilon, \tag{4.2}$$

then there exists no consistent test for the hypotheses $H_0 : f = f_0$ against $H_1 : f \in \{f_1, \dots, f_k\}$. A sufficient condition for (4.2) is that

$$\mathrm{E}_{f_0}\Big(\frac{1}{k}\sum_{i=1}^{k}\frac{d\mathrm{P}_{f_i}}{d\mathrm{P}_{f_0}} - 1\Big)^2 \leq \delta$$

for some $\delta \in (0,1)$, independent of n.

Proof. Observe that for an arbitrary test φ,

$$\max_{i=1,\dots,k}\mathrm{E}_{f_i}(1-\varphi) \geq \frac{1}{k}\sum_{i=1}^{k}\mathrm{E}_{f_i}(1-\varphi) = \frac{1}{k}\sum_{i=1}^{k}\mathrm{E}_{f_0}\frac{d\mathrm{P}_{f_i}}{d\mathrm{P}_{f_0}}(1-\varphi) = \mathrm{E}_{f_0}L(1-\varphi),$$

where L is the average likelihood defined by $L = k^{-1}\sum_{i=1}^{k} dP_{f_i}/dP_{f_0}$. Hence for $c \in (0,1)$ the testing error $\mathrm{E}_{f_0}\varphi + \max_{i=1,\dots,k}\mathrm{E}_{f_i}(1-\varphi)$ is bounded from below by

$$\begin{aligned}\mathrm{E}_{f_0}(\varphi + L(1-\varphi)) &= \mathrm{E}_{f_0}(1_{\varphi=1} + L1_{\varphi=0})\\ &\geq \mathrm{E}_{f_0}(1_{\varphi=1} + L1_{\varphi=0})1_{L>c}\\ &\geq c\mathrm{P}_{f_0}(L>c).\end{aligned}$$

It follows that under the condition of the lemma, the error of any test is bounded away from 0.

To prove the second statement of the lemma note that

$$\mathrm{P}_{f_0}(L>c) \geq 1 - \mathrm{P}_{f_0}(|L-1| \geq 1-c) \geq 1 - \frac{\mathrm{E}_{f_0}(L-1)^2}{(1-c)^2},$$

by Markov's inequality. If $\delta \in (0,1)$, there is a $c \in (0,1)$ such that $\delta < (1-c)^2$. This completes the proof. □

The preceding lemma is general in that it does not use the fact that we are in the setting of the white noise model. The following proposition gives conditions specific to our situation.

Proposition 4.3.2. *Let $f_0, f_1, \dots, f_k \in L^2[0,1]$. If for some $\delta \in (0,1)$, independent of n,*

$$\frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} e^{n\langle f_i - f_0, f_j - f_0\rangle} \leq 1 + \delta, \tag{4.3}$$

then there exists no consistent test for the hypotheses $H_0 : f = f_0$ against $H_1 : f \in \{f_1, \dots, f_k\}$.

Proof. Denote the average likelihood by L again. By (2.6) we have

$$L^2 = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} e^{\sqrt{n}\int (f_i - f_0)\,dW - \frac{1}{2}n\|f_i - f_0\|_2^2}\, e^{\sqrt{n}\int (f_j - f_0)\,dW - \frac{1}{2}n\|f_j - f_0\|_2^2},$$

where W is a P_{f_0}-Brownian motion. It follows from the properties of the Wiener integral that the P_{f_0}-expectation of term (i,j) in the sum equals $\exp(n\langle f_i - f_0, f_j - f_0\rangle)$ (check!). Hence

$$\mathrm{E}_{f_0}(L-1)^2 = \mathrm{E}_{f_0}L^2 - 1 = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} e^{n\langle f_i - f_0, f_j - f_0\rangle} - 1,$$

and the proposition follows from Lemma 4.3.1. □

Observe that for $k = 1$ the proposition implies that there exists no consistent test for the simple hypotheses $H_0 : f = f_0$ against $H_1 : f = f_1$ if $n\|f_1 - f_0\|_2^2 \le \log(1+\delta)$. This matches the sharp result that we obtained above from Proposition 4.2.2.

The proof of Proposition 4.3.2 shows that the sum in (4.3) is always bounded from below by 1 (this can also be seen directly using Jensen's inequality, for instance). The sum can be decomposed as

$$\frac{1}{k^2}\sum_{i=1}^k e^{n\|f_i-f_0\|_2^2} + \frac{1}{k^2}\sum_{i=1}^k\sum_{j\ne i} e^{n\langle f_i-f_0, f_j-f_0\rangle}.$$

The first term quantifies the L^2-distance of the functions in the alternative to f_0. It is small if all the f_i are close to f_0. The second term somehow quantifies the 'dimension' of the alternative. It is small if many of the 'directions' $f_i - f_0$ are (nearly) orthogonal to each other. So we see that the testing problem $H_0 : f = f_0$ against $H_1 : f \in \mathcal{F}$ becomes hard if the alternative $\mathcal{F}$ contains many functions that are in L^2-sense close to f_0 and that are at orthogonal directions from f_0. See Exercise 4.4 for a precise result of this type.

Example 4.3.3 (Detecting a smooth bump at one of many locations)**.** Fix $0 < u < v < 1$. For $x_1, \dots, x_k \in (u, v)$, consider the bump functions $f_{\sigma,i}$ defined by

$$f_{\sigma,i}(x) = \varepsilon\sigma^\beta K\Big(\frac{x-x_i}{\sigma}\Big), \qquad x \in [0,1],$$

where K is a C^∞ function with compact support and $\beta, \sigma, \varepsilon > 0$. As in Example 4.2.4, the C^β-norm of these functions is uniformly bounded by 1 if ε is small enough and their squared L^2-norm is bounded by $C\varepsilon^2\sigma^{1+2\beta}$ for some $C > 0$. Since K has compact support, we can position $k \sim 1/\sigma$ points x_i in (u, v) in such a way that the k resulting bump functions have disjoint supports, hence are orthogonal in $L^2[0,1]$. It follows from Proposition 4.3.2, see also Exercise 4.4, there exists no consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \{f_{\sigma,1}, \dots, f_{\sigma,k}\}$ if

$$\sigma e^{nC\varepsilon^2\sigma^{1+2\beta}} \le \delta$$

for some $\delta \in (0,1)$. This condition is fulfilled for ε small enough and $\sigma = (n/\log n)^{-1/(1+2\beta)}$, which corresponds to $k \sim (n/\log n)^{1/(1+2\beta)}$.

So to be able to detect a bump at one of the locations x_i in the interval $[u, v]$, the height $f_{\sigma,i}(x_i)$ of the bumps has to be a large enough multiple of $(n/\log n)^{-\beta/(1+2\beta)}$. This is a (slightly) more stringent condition than the condition that we obtained in Example 4.2.4 for detecting a single bump, reflecting the fact that statistically it is more difficult to detect a bump at one of many possible locations than at a single fixed location. ■

Using this last example we can obtain the uniform norm analogue of the result of Exercise 4.3, see Exercise 4.5. Our next goal is to understand how the results of Examples 4.2.4 and 4.3.3 and Exercises 4.3 and 4.5 change if we replace Hölder regularity by Sobolev regularity and the pointwise and uniform risks by L^2-risk. It will turn out that the result looks rather different in this case. To prove this we need a different construction than the orthogonal bump functions.

For concreteness we fix an orthonormal basis e_j of $L^2[0,1]$. The Fourier coefficients of a function $f \in L^2[0,1]$ are denoted by f_j, i.e. $f_j = \langle f, e_j \rangle$. For $\beta > 0$, the associated Sobolev norm $\|f\|_{H^\beta}$ of f is defined by

$$\|f\|_{H^\beta}^2 = \sum f_j^2 j^{2\beta}.$$

Recall that the Sobolev space $H^\beta[0,1]$ is the space of all functions in $L^2[0,1]$ for which this norm is finite and that the unit ball of $H^\beta[0,1]$ is denoted by $H_1^\beta[0,1]$. Following the same line of reasoning as before we now construct a finite collection of functions in $H_1^\beta[0,1]$ that are difficult to distinguish from 0.

Example 4.3.4 (Detecting Sobolev smooth signals)**.** Let $k \sim n^{1/(1/2+2\beta)}$. For $\varepsilon > 0$ and $\theta \in \{-1,+1\}^k$, define

$$f_\theta = \varepsilon k^{-1/2-\beta} \sum_{j=1}^k \theta_j e_j.$$

We have $\langle f_\theta, f_\psi \rangle = \varepsilon^2 k^{-1-2\beta} \sum \theta_j \psi_j$. In particular, $\|f_\theta\|_2^2 = \varepsilon^2 k^{-2\beta} \sim \varepsilon^2 n^{-2\beta/(1/2+2\beta)}$. For the Sobolev norm we have

$$\|f_\theta\|_{H^\beta}^2 = \varepsilon^2 k^{-1-2\beta} \sum_{j=1}^k j^{2\beta} \theta_j^2 = \text{const.} \times \varepsilon^2.$$

So for ε small enough all functions f_θ belong to the Sobolev ball $H_1^\beta[0,1]$ and their squared L^2-norm is of the order $n^{-2\beta/(1/2+2\beta)}$.

The condition (4.3) for non-testability of $H_0 : f = 0$ against $H_1 : f \in \{f_\theta : \theta \in \{-1,+1\}^k\}$ reads in this case

$$\frac{1}{4^k} \sum_\theta \sum_\psi e^{n\varepsilon^2 k^{-1-2\beta} \sum \theta_j \psi_j} \leq 1 + \delta$$

for some $\delta \in (0,1)$. Now let $R_1, \ldots, R_k$ and $S_1, \ldots, S_k$ be two independent sequences of k independent Rademacher variables, each variable having values ± 1 with probability $1/2$. Then the quantity on the left-hand side of the display equals

$$\mathrm{E} e^{n\varepsilon^2 k^{-1-2\beta} \sum R_j S_j} = \left(\mathrm{E} e^{n\varepsilon^2 k^{-1-2\beta} R_1 S_1} \right)^k$$

(check!). But $R_1 S_1$ is again a Rademacher variable, so there exists no consistent test if

$$k \log \mathrm{E} e^{n\varepsilon^2 k^{-1-2\beta} R} \leq \delta'$$

for some $\delta' \in (0, 1/2)$, where R is a Rademacher variable. It is easily verified that this condition is fulfilled (see Exercise 4.6) if ε is small enough. We conclude that it is not possible to distinguish whether the signal is 0 or one of the $f_\theta \in H_1^\beta[0,1]$ if ε is small enough. ■

In Example 4.3.4 we obtain a different rate than in the preceding Examples 4.2.4 and 4.3.3, namely $n^{-\beta/(1/2+2\beta)}$ instead of $n^{-\beta/(1+2\beta)}$ (with or without a logarithmic factor). As a result, the separation relative to the L^2-norm that is necessary for consistent testing of $f = 0$ against Sobolev smooth alternatives is different from the pointwise or uniform separation that is necessary for consistent testing of $f = 0$ against Hölder smooth alternatives. See Exercise 4.7.

4.4 Testing multiple hypotheses

So far we have been considering tests for two hypotheses H_0 and H_1, where H_0 is a simple hypothesis and H_1 might be composite. In this section we study tests for the situation that we have multiple simple hypotheses $H_0, H_1, \dots, H_k$ of the form $H_i : f = f_i$, for some collection of functions $\{f_0, f_1, \dots, f_k\} \subset L^2[0,1]$. Now the problem is not to only decide wether f_0 is the truth or not, but which of the f_i's is in fact the true parameter. In this setting a test is formally a $\{0, 1, \dots, k\}$-valued measurable function $\varphi = \varphi(X)$ of the observation X and we want all of the error probabilities $\mathrm{P}_{f_i}(\varphi \neq i)$ to be small. In this situation we say that there exists a consistent test if

$$\inf_\varphi \max_i \mathrm{P}_{f_i}(\varphi \neq i) \to 0$$

as $n \to \infty$, where the infimum is over all possible tests. It is intuitively clear that this might be more demanding than testing f_0 against the union of the other f_i, and that we might get different results than before.

The following lemma gives a sufficient condition for non-testability in terms of likelihood ratios, in the spirit of Lemma 4.3.1.

Lemma 4.4.1. *Let $f_0, f_1, \dots, f_k \in L^2[0,1]$. If for some $i \in \{0, 1, \dots, k\}$ and $\varepsilon \in (0,1)$, independent of n, and $c > 0$ such that ck is bounded away from 0 it holds that*

$$\frac{1}{k} \sum_{j \neq i} \mathrm{P}_{f_j}\Big(\frac{d\mathrm{P}_{f_i}}{d\mathrm{P}_{f_j}} \geq c\Big) \geq \varepsilon, \tag{4.4}$$

then there exists no consistent multiple hypothesis test for the hypotheses $H_0 : f = f_0$, $H_1 : f = f_1, \dots, H_k : f = f_k$.

Proof. For $j = 1, \dots, k$, define the events $A_j = \{d\mathrm{P}_{f_0}/d\mathrm{P}_{f_j} \geq c\}$. Then we have

$$\begin{aligned}
\mathrm{P}_{f_0}(\varphi \neq 0) &= \sum_{j\neq 0} \mathrm{P}_{f_0}(\varphi = j) \\
&= \sum_{j\neq 0} \mathrm{E}_{f_j} \frac{d\mathrm{P}_{f_0}}{d\mathrm{P}_{f_j}} 1_{\varphi=j} \\
&\geq c \sum_{j\neq 0} \mathrm{P}_{f_j}(\varphi = j, A_j) \\
&\geq c \sum_{j\neq 0} \Big(\mathrm{P}_{f_j}(\varphi = j) - \mathrm{P}_{f_j}(A_j^c) \Big).
\end{aligned}$$

Writing

$$s = \frac{1}{k} \sum_{j=1}^{k} \mathrm{P}_{f_j}(\varphi = j), \qquad t = \frac{1}{k} \sum_{j=1}^{k} \mathrm{P}_{f_j}\Big(\frac{d\mathrm{P}_{f_0}}{d\mathrm{P}_{f_j}} < c \Big),$$

it follows that

$$\begin{aligned}
\max_j \mathrm{P}_{f_j}(\varphi \neq j) &= \max\Big\{ \mathrm{P}_{f_0}(\varphi \neq 0), \max_{j\neq 0} \mathrm{P}_{f_j}(\varphi \neq j) \Big\} \\
&\geq \max\Big\{ \mathrm{P}_{f_0}(\varphi \neq 0), \frac{1}{k} \sum_{j=1}^{k} \mathrm{P}_{f_j}(\varphi \neq j) \Big\} \\
&\geq \max\Big\{ ck(s-t), 1-s \Big\}.
\end{aligned}$$

By considering the functions $s \mapsto ck(s-t)$ and $s \mapsto 1-s$ on $(0,1)$ we see that the maximum is bounded from below by $(1-t)ck/(1+ck)$ (check!), i.e.

$$\max_j \mathrm{P}_{f_j}(\varphi \neq j) \geq \frac{ck}{1+ck} \frac{1}{k} \sum_{f\neq f_0} \mathrm{P}_f\Big(\frac{d\mathrm{P}_{f_0}}{d\mathrm{P}_f} \geq c \Big)$$

The proof of the lemma is completed by noting that we can replace f_0 by any other of the f_i's. □

The following proposition translates the requirements of the general lemma into specific conditions in the setting of the white noise model. Note that the condition is very close to the condition obtained Exercise 4.4 for non-testability of 0 against a composite alternative consisting of k orthogonal functions. In the present setting of multiple hypothesis testing, there is no orthogonality requirement.

Proposition 4.4.2. *Let $f_0, f_1, \dots, f_k \in L^2[0,1]$. If there exist a constant $c > 0$, independent of n, such that*

$$n \max_{i,j} \|f_i - f_j\|_2^2 \leq c + 2\log k$$

for n large enough, then there exists no consistent multiple hypothesis test for the hypotheses $H_0 : f = f_0$, $H_1 : f = f_1, \dots, H_k : f = f_k$.

Proof. By lower bounding the maximum of a sequence by the mean we see from the preceding lemma that a sufficient condition for non-testability is that for some $c > 0$ such that ck is bounded away from 0 it holds that

$$\frac{1}{k(k+1)} \sum_i \sum_{j \neq i} \mathrm{P}_{f_i}\Big(\frac{d\mathrm{P}_{f_j}}{d\mathrm{P}_{f_i}} \geq c\Big)$$

is bounded away from 0. By the properties of the Wiener integral (see Theorem 2.2.2),

$$\mathrm{P}_{f_i}\Big(\frac{d\mathrm{P}_{f_j}}{d\mathrm{P}_{f_i}} \geq c\Big) = \mathrm{P}_{f_i}\Big(\log\frac{d\mathrm{P}_{f_j}}{d\mathrm{P}_{f_i}} \geq \log c\Big) = 1 - \Phi\Big(\frac{\frac{1}{2}n\|f_i - f_j\|_2^2 + \log c}{\sqrt{n}\|f_i - f_j\|_2}\Big).$$

For $\log c = -\max_{i,j} n\|f_i - f_j\|_2^2/2$ this is bounded from below by 1/2, and therefore also

$$\frac{1}{k(k+1)} \sum_i \sum_{j \neq i} \mathrm{P}_{f_i}\Big(\frac{d\mathrm{P}_{f_j}}{d\mathrm{P}_{f_i}} \geq c\Big) \geq 1/2.$$

For c as defined above, the condition that ck is bounded away from 0 is exactly the condition of the proposition. □

The following example shows that the condition for non-testability in the present multiple hypothesis setting is indeed less restrictive than what we had before. Compare with Example 4.3.4.

Example 4.4.3 (Distinguishing Sobolev smooth functions)**.** Set $k = n^{1/(1+2\beta)}$ and consider the first k elements $e_1, \dots, e_k$ of an orthonormal basis e_i of $L^2[0,1]$. For $\varepsilon > 0$ and $\theta \in \{-1, +1\}^k$, define

$$f_\theta = \varepsilon k^{-1/2-\beta} \sum_{i=1}^k \theta_i e_i.$$

We have

$$\|f_\theta - f_\psi\|_2^2 = \varepsilon^2 k^{-1-2\beta}\|\theta - \psi\|^2.$$

Since θ, ψ are vectors of ± 1's, their squared distance equals 4 times the number $d_{\mathrm{ham}}(\theta, \psi)$ of indices at which they differ. In other words, $\|f_\theta - f_\psi\|_2^2 =$

$4\varepsilon^2 k^{-1-2\beta} d_{\text{ham}}(\theta, \psi)$. We also see that $\|f_\theta\|_2^2 = \varepsilon^2 k^{-2\beta} = \varepsilon^2 n^{-2\beta/(1+2\beta)}$. Moreover, for the β-Sobolev norm (relative to the basis e_i) we have

$$\|f_\theta\|_{H^\beta}^2 = \varepsilon^2 k^{-1-2\beta} \sum_{i=1}^{k} i^{2\beta} \theta_i^2 = \text{const.} \times \varepsilon^2.$$

We have constructed 2^k different functions and

$$n \max_{\theta,\psi} \|f_\theta - f_\psi\|_2^2 = 4\varepsilon^2 n k^{-2\beta} = 4\varepsilon^2 k.$$

This implies that the condition of Proposition 4.4.2 is fulfilled if ε is small enough. ■

It is interesting to compare this example with the result of Exercise 4.7. By part (ii) of that exercise, we can consistently test whether there is a non-zero signal that belongs to the Sobolev ball $H_1^\beta[0, 1]$ as soon as that signal has L^2-norm of larger order than $n^{-\beta/(1/2+2\beta)}$. Example 4.4.3 shows however that there are signals in $H_1^\beta[0, 1]$ with L^2-norm between $n^{-\beta/(1/2+2\beta)}$ and $n^{-\beta/(1+2\beta)}$ such that we can not consistently distinguish between them. In other words, there are collections of signals for which we *can* consistently decide that *one* of them generated the data, but for which we *can not* say *which one* of them did so. This hints at the fact that in this setting, signal detection is in some sense easier than signal estimation. The latter will be studied in the next chapter.

4.5 Exercises

Exercise 4.1. Complete the proof of Proposition 4.2.2.

Exercise 4.2 (L^2- and C^β-norms of smooth bump functions)**.** Prove the claims about the C^β- and L^2-norms of the function f_σ in Example 4.2.4.

Exercise 4.3 (Testing against Hölder alternatives, pointwise separation)**.** Let $\beta > 0$. Let $C_1^\beta[0,1]$ be the unit ball of the Hölder space $C^\beta[0,1]$.

(i) Show that in the white noise model, there exists no consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in C_1^\beta[0,1]\backslash\{0\}$.

(ii) Let $t_0 \in (0,1)$. Show that if r_n is of smaller order than $n^{-\beta/(1+2\beta)}$, then there exists no consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \{f \in C_1^\beta[0,1] : |f(t_0)| \geq r_n\}$.

(iii) Let $t_0 \in (0,1)$. Show that if r_n is of larger order than $n^{-\beta/(1+2\beta)}$, then there does exist a consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \{f \in C_1^\beta[0,1] : |f(t_0)| \geq r_n\}$. (Hint: use the estimator of Example 3.4.3 as test statistic and use the result (3.12) to bound the errors of the test.)

Exercise 4.4. Show that in the white noise model there exists no consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \mathcal{F}_n$ if $\mathcal{F}_n$ contains k_n different orthogonal functions f_i such that $k_n^{-1}\exp(n \max_i \|f_i\|_2^2) \leq \delta$ for some $\delta \in (0,1)$.

Exercise 4.5 (Testing against Hölder alternatives, uniform separation)**.** Let $\beta > 0$. Let $C_1^\beta[0,1]$ be the unit ball of the Hölder space $C^\beta[0,1]$.

(i) Show that if r_n is of smaller order than $(n/\log n)^{-\beta/(1+2\beta)}$, then there exists no consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \{f \in C_1^\beta[0,1] : \sup_{u<t<v} |f(t)| \geq r_n\}$.

(ii) Show that if r_n is of larger order than $(n/\log n)^{-\beta/(1+2\beta)}$, then there does exist a consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \{f \in C_1^\beta[0,1] : \sup_{u<t<v} |f(t)| \geq r_n\}$. (Hint: use the estimator of Example 3.4.4 as test statistic and use the result (3.13) to bound the errors of the test.)

Explain the differences with the results of Exercise 4.3.

Exercise 4.6. Prove the last statements in Example 4.3.4.

Exercise 4.7 (Testing against Sobolev alternatives, L^2-separation)**.** Let $\beta > 0$. Let $H_1^\beta[0,1]$ be the unit ball of the Sobolev space $H^\beta[0,1]$.

(i) Show that if r_n is of smaller order than $n^{-\beta/(1/2+2\beta)}$, then there exists no consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \{f \in H_1^\beta[0,1] : \|f\|_2 \geq r_n\}$.

(ii) Show that if r_n is of larger order than $n^{-\beta/(1/2+2\beta)}$, then there does exist a consistent test for the hypotheses $H_0 : f = 0$ against $H_1 : f \in \{f \in H_1^\beta[0,1] : \|f\|_2 \geq r_n\}$. (Hint: use a test that rejects H_0 for large values of $T = \sum_{j \leq J}(Y_j^2 - 1/n)$, for $Y_j = \int e_j\, dX$ and $J \sim n^{1/(1/2+2\beta)}$.)

Chapter 5

Lower bounds for estimation

5.1 Minimax lower bounds for estimation

After studying the fundamental possibilities and limitations in various testing problems in the white noise model in Chapter 4 we return to estimation problems in this chapter. In Chapter 3 we introduced and studied various regularization methods for nonparametric estimation of the signal f from observations $X = (X_t : t \in [0,1])$ satisfying

$$dX_t = f(t)\,dt + \frac{1}{\sqrt{n}} dW_t$$

We found in particular that under the assumption that f is β-smooth in Hölder or Sobolev sense, there exist estimators that, perhaps up to a logarithmic factor, achieve a rate of convergence of the order $n^{-\beta/(1+2\beta)}$ relative to the pointwise, supremum or L^2-norm. Moreover, such results hold uniformly over Hölder or Sobolev balls. Specifically, we have the upper bounds (3.12), (3.13), (3.14) for appropriately tuned kernel or projection estimators.

The cited rate of convergence results are all risk upper bounds of the form

$$\sup_{f\in\mathcal{F}} \mathrm{E}_f \ell(\hat{f}, f) \leq r_n,$$

where $\hat{f}$ is a specific estimator, $\mathcal{F} \subset L^2[0,1]$ is a class of signals of interest, ℓ is a choice of *loss function*, such as pointwise loss, uniform loss, or squared L^2-loss, and r_n is a sequence converging to 0 as $n \to \infty$. Such results quantify how well a specific estimator performs for a certain class of truths. As we have seen, the bounds typically depend both on the class $\mathcal{F}$ and on the type of loss ℓ that is considered.

In this chapter we consider the converse question. We ask what the best possible performance of *any* estimator is over a given class of signals $\mathcal{F}$. More precisely, we are interested in proving so-called *minimax lower bounds* for estimation of the form

$$\inf_{\hat{f}} \sup_{f\in\mathcal{F}} \mathrm{E}_f \ell(\hat{f}, f) \geq r_n,$$