Chapter 22
Strongly Consistent Detection for Nonparametric Hypotheses

László Györfi and Harro Walk

Abstract Consider two robust detection problems formulated by nonparametric hypotheses such that both hypotheses are composite and indistinguishable. Strongly consistent testing rules are shown.

22.1 Composite Hypotheses Defined by Half Spaces of Distributions

Let ν_0, ν_1 be fixed distributions on $\mathbb{R}^d$ which are the nominal distributions under two hypotheses. Let

$$V(\nu, \mu) = \sup_{A \subseteq \mathbb{R}^d} |\nu(A) - \mu(A)|$$

denote the total variation distance between two distributions ν and μ, where the supremum is taken over all Borel sets of $\mathbb{R}^d$.

Let $X, X_1, X_2, \ldots$ be i.i.d. random vectors according to a common distribution μ. We observe $X_1, \ldots, X_n$. Under the hypothesis H_j $(j = 0, 1)$ the distribution μ is a distorted version of ν_j. Formally define the two hypotheses by

$$H_0 = \{\mu : V(\mu, \nu_0) < V(\mu, \nu_1)\},$$

The research reported here was supported in part by the National Development Agency (NFÜ, Hungary) as part of the project Introduction of Cognitive Methods for UAV Collision Avoidance Using Millimeter Wave Radar (grant no.: KMR-12-1-2012-0008).

L. Györfi (✉)
Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Stoczek u. 2, Budapest 1521, Hungary
e-mail: gyorfi@cs.bme.hu

H. Walk
Department of Mathematics, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany
e-mail: walk@mathematik.uni-stuttgart.de

© Springer International Publishing Switzerland 2015

V. Vovk et al. (eds.), *Measures of Complexity*,
DOI 10.1007/978-3-319-21852-6_22

and

$$H_1 = \{\mu : V(\mu, \nu_1) < V(\mu, \nu_0)\}.$$

Our aim is to construct a distribution-free strongly consistent test, which makes an error only finitely many times almost surely (a.s.). The concept of strongly consistent test is quite unusual: it means that both on H_0 and on H_1 the test makes a.s. no error after a random sample size. In other words, denoting by $\mathbb{P}_0$ and $\mathbb{P}_1$ the probability under the hypotheses H_0 and H_1, we have

$$\mathbb{P}_0\{\text{rejecting } H_0 \text{ for only finitely many } n\} = 1$$

and

$$\mathbb{P}_1\{\text{rejecting } H_1 \text{ for only finitely many } n\} = 1.$$

In a real-life problem, for example, when we get the data sequentially, one gets data just once, and should make good inferences from these data. Strong consistency means that the single sequence of inference is a.s. perfect if the sample size is large enough. This concept is close to the definition of discernability introduced by Dembo and Peres [5]. For a discussion and references, we refer the reader to Biau and Györfi [3], Devroye and Lugosi [7], Gretton and Györfi [10], and Györfi and Walk [15].

Motivated by a related goodness of fit test statistic of Györfi and van der Meulen [14], we put

$$L_{n,0} = \sum_{j=1}^{m_n} |\mu_n(A_{n,j}) - \nu_0(A_{n,j})|,$$

and

$$L_{n,1} = \sum_{j=1}^{m_n} |\mu_n(A_{n,j}) - \nu_1(A_{n,j})|,$$

where μ_n denotes the empirical measures associated with the sample $X_1, \ldots, X_n$, so that

$$\mu_n(A) = \frac{\#\{i : X_i \in A, i = 1, \ldots, n\}}{n}$$

for any Borel subset A, and $\mathcal{P}_n = \{A_{n,1}, \ldots, A_{n,m_n}\}$ is a finite partition of $\mathbb{R}^d$.

We introduce a test such that the hypothesis H_0 is accepted if

$$L_{n,0} < L_{n,1}, \tag{22.1}$$

and rejected otherwise.

The sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots$ is called asymptotically fine if for any sphere S centered at the origin

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n, A\cap S\neq\emptyset} diam(A) = 0.$$

Theorem 22.1 *Assume that the sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots$ is asymptotically fine and*

$$\lim_{n\to\infty} \frac{m_n}{n} = 0. \tag{22.2}$$

Then the test (22.1) is strongly consistent.

Proof Assume H_0 without loss of generality. Then the error event means that

$$L_{n,0} \geq L_{n,1}.$$

Thus,

$$\begin{aligned} 0 &\leq \sum_{j=1}^{m_n} |\mu_n(A_{n,j}) - \nu_0(A_{n,j})| - \sum_{j=1}^{m_n} |\mu_n(A_{n,j}) - \nu_1(A_{n,j})| \\ &\leq 2L_n + \sum_{j=1}^{m_n} |\mu(A_{n,j}) - \nu_0(A_{n,j})| - \sum_{j=1}^{m_n} |\mu(A_{n,j}) - \nu_1(A_{n,j})|, \end{aligned}$$

where

$$L_n = \sum_{j=1}^{m_n} |\mu_n(A_{n,j}) - \mu(A_{n,j})|.$$

Introduce the notation

$$\epsilon = -(V(\mu, \nu_0) - V(\mu, \nu_1)) > 0.$$

The sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots$ is asymptotically fine, which implies that

$$\begin{aligned} &\lim_{n\to\infty} \left(\sum_{j=1}^{m_n} |\mu(A_{n,j}) - \nu_0(A_{n,j})| - \sum_{j=1}^{m_n} |\mu(A_{n,j}) - \nu_1(A_{n,j})| \right) \\ &= 2\,(V(\mu, \nu_0) - V(\mu, \nu_1)) \\ &= -2\epsilon, \end{aligned}$$

(cf. Biau and Györfi [3]). Thus, for all n large enough,

$$P_{e,n} = \mathbb{P}\{\text{error}\} \leq \mathbb{P}\{0 \leq 2L_n - \epsilon\}.$$

Beirlant et al. [2] and Biau and Györfi [3] proved that, for any $\delta > 0$,

$$\mathbb{P}\{L_n > \delta\} \leq 2^{m_n} e^{-n\delta^2/2}.$$

Therefore

$$P_{e,n} \leq 2^{m_n} e^{-n\epsilon^2/8}.$$

Because of (22.2),

$$\sum_{n=1}^{\infty} P_{e,n} < \infty,$$

and so the Borel-Cantelli lemma implies that a.s.

$$L_{n,0} - L_{n,1} < 0$$

for all n large enough, i.e., the error

$$L_{n,0} - L_{n,1} \geq 0$$

occurs a.s. for only finitely many n. Thus, strong consistency is proved. □

In a straightforward way, the proof of Theorem 22.1 can be extended to infinite partitions if we assume that for each sphere S centered at the origin

$$\lim_{n\to\infty} \frac{|\{j : A_{n,j} \cap S \neq \emptyset\}|}{n} = 0.$$

Next, a variant of the test (22.1) with much smaller computational complexity will be defined. The test statistic is based on a recursive histogram. In this section we assume that the partitions are infinite and all cells of the partitions have finite and positive Lebesgue measure λ. Let $A_n(x)$ denote the cell of $\mathcal{P}_n$ to which x belongs. The density estimate

$$f_n(x) := \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{I}_{\{X_i \in A_i(x)\}}}{\lambda(A_i(x))}$$

is called a recursive histogram.

For $A \in \mathcal{P}_n$, introduce the estimate

$$\mu_n^*(A) := \int_A f_n(x)dx.$$

Notice that $\mu_n^*(A)$ can be calculated in a recursive way, which is important in on-line applications. These definitions imply that

$$\mu_n^*(A) = \frac{1}{n}\sum_{i=1}^{n}\int_A \frac{\mathbb{I}_{\{X_i \in A_i(x)\}}}{\lambda(A_i(x))}dx = \frac{1}{n}\sum_{i=1}^{n}\int_A \frac{\mathbb{I}_{\{x \in A_i(X_i)\}}}{\lambda(A_i(X_i))}dx$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda(A \cap A_i(X_i))}{\lambda(A_i(X_i))}.$$

If the sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots$ is nested, i.e., the sequence of σ-algebras $\sigma(\mathcal{P}_n)$ is non-decreasing, then for $A \in \mathcal{P}_n$ let the ancestor $B_A^{(i)} \in \mathcal{P}_i$ be such that $A \subseteq B_A^{(i)}$ $(i \le n)$. One can check that for nested partitions the estimate has the following form:

$$\mu_n^*(A) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{\{X_i \in B_A^{(i)}\}}\frac{\lambda(A)}{\lambda(B_A^{(i)})}.$$

Put

$$L_{n,j}^* := \sum_{A \in \mathcal{P}_n} |\mu_n^*(A) - \nu_j(A)|$$

$(j = 0, 1)$. We introduce a test such that the hypothesis H_0 is accepted if

$$L_{n,0}^* < L_{n,1}^*, \tag{22.3}$$

and rejected otherwise.

Theorem 22.2 *Assume that the sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots$ is asymptotically fine such that*

$$\sum_{n=1}^{\infty}\frac{1}{n^2 \inf_j \lambda(A_{n,j})} < \infty.$$

Further suppose that μ has a density. Then the test (22.3) is strongly consistent.

Proof Assume H_0 without loss of generality. One notices

$$L_{n,0}^* - L_{n,1}^* \le 2L_n^* + Q_n^*,$$

where

$$L_n^* = \sum_{A \in \mathcal{P}_n} |\mu_n^*(A) - \mu(A)|,$$

and

$$Q_n^* = \sum_{A \in \mathcal{P}_n} |\mu(A) - \nu_0(A)| - \sum_{A \in \mathcal{P}_n} |\mu(A) - \nu_1(A)|.$$

By Biau and Györfi [3],

$$Q_n^* \to 2(V(\mu, \nu_0) - V(\mu, \nu_1)) < 0,$$

the latter because of H_0. Next $L_n^* \to 0$ a.s. will be shown. Denote the density of μ by f. Thus

$$L_n^* = \sum_{A \in \mathcal{P}_n} \left| \int_A f_n(x)dx - \int_A f(x)dx \right| \leq \int |f_n(x) - f(x)|dx.$$

Therefore we have to prove the strong L_1-consistency of the recursive histogram. Consider the bias part. Introduce the ordinary histogram:

$$\tilde{f}_n(x) := \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{I}_{\{X_i \in A_n(x)\}}}{\lambda(A_n(x))},$$

and put

$$\bar{f}_n(x) := \mathbb{E}\{\tilde{f}_n(x)\} = \frac{\mu(A_n(x))}{\lambda(A_n(x))}.$$

According to the Abou-Jaoude theorem, if the sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \dots$ is asymptotically fine, then

$$\int |\bar{f}_n - f| \to 0$$

(cf. Devroye and Györfi [6]). Thus, for the bias term of the recursive histogram, we get

$$\int |\mathbb{E}\{f_n\} - f| = \int \left| \frac{1}{n} \sum_{i=1}^{n} \bar{f}_i - f \right| \leq \frac{1}{n} \sum_{i=1}^{n} \int |\bar{f}_i - f| \to 0. \tag{22.4}$$

For the variation term of the recursive histogram, we apply the generalized theorem of Kolmogorov. Let U_n, $n = 1, 2, \dots$ be an L_2-valued sequence of independent, zero-mean random variables such that

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\{\|U_n\|_2^2\}}{n^2} < \infty$$

where $\|\cdot\|_2$ denotes the L_2 norm. Then

$$\lim_{n \to \infty} \left\| \frac{1}{n} \sum_{i=1}^{n} U_i \right\|_2 = 0$$

a.s. (cf. Györfi et al. [11]). For

$$U_n := \frac{\mathbb{I}_{\{X_n \in A_n(\cdot)\}}}{\lambda(A_n(\cdot))} - \mathbb{E}\left\{ \frac{\mathbb{I}_{\{X_n \in A_n(\cdot)\}}}{\lambda(A_n(\cdot))} \right\},$$

one has to verify the condition of the generalized Kolmogorov theorem:

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{\mathbb{E}\left\{\left\| \frac{\mathbb{I}_{\{X_n \in A_n(\cdot)\}}}{\lambda(A_n(\cdot))} - \mathbb{E}\left\{\frac{\mathbb{I}_{\{X_n \in A_n(\cdot)\}}}{\lambda(A_n(\cdot))}\right\}\right\|_2^2\right\}}{n^2} &\leq \sum_{n=1}^{\infty} \frac{\mathbb{E}\left\{\left\| \frac{\mathbb{I}_{\{X_n \in A_n(\cdot)\}}}{\lambda(A_n(\cdot))}\right\|_2^2\right\}}{n^2} \\ &= \sum_{n=1}^{\infty} \frac{\mathbb{E}\left\{\int \frac{\mathbb{I}_{\{X_n \in A_n(x)\}}}{\lambda(A_n(x))^2} dx\right\}}{n^2} \\ &= \sum_{n=1}^{\infty} \frac{\mathbb{E}\left\{\int \frac{\mathbb{I}_{\{x \in A_n(X_n)\}}}{\lambda(A_n(X_n))^2} dx\right\}}{n^2} \\ &= \sum_{n=1}^{\infty} \frac{\mathbb{E}\left\{\frac{1}{\lambda(A_n(X_n))}\right\}}{n^2} \\ &\leq \sum_{n=1}^{\infty} \frac{1}{n^2 \inf_j \lambda(A_{n,j})} < \infty,\end{aligned}$$

by the condition of the theorem, and so

$$\int |f_n - \mathbb{E}\{f_n\}|^2 \to 0 \tag{22.5}$$

a.s. From Lemma 3.1 in Györfi and Masry [13] we get that the limit relations (22.4) and (22.5) imply

$$\int |f_n - f| \to 0$$

a.s. Therefore a.s.

$$L_{n,0}^* - L_{n,1}^* < 0$$

for all n large enough, and so strong consistency is proved. □

22.2 Composite Hypotheses Defined by Half Spheres of Distributions

Again, under the hypothesis H_j' ($j = 0, 1$) the distribution μ is a distorted version of ν_j. In this section we assume that the true distribution lies within a certain total variation distance of the underlying nominal distribution.

We formally define the two hypotheses by

$$H_j' = \left\{\mu : V(\mu, \nu_j) < \Delta\right\}, \quad j = 0, 1, \tag{22.6}$$

where

$$\Delta := (1/2)V(\nu_0, \nu_1).$$

Because of

$$H'_j \subseteq H_j, \quad j = 0, 1 ,$$

the test (22.1) in the previous section is strongly consistent. In this section we introduce a simpler test. For the notations

$$f = \frac{d\nu}{d(\nu + \mu)} \qquad \text{and} \qquad g = \frac{d\mu}{d(\nu + \mu)},$$

the general version of Scheffé's theorem implies that

$$V(\nu, \mu) = \nu(A^*) - \mu(A^*),$$

where

$$A^* = \{x : f(x) > g(x)\} .$$

Introduce the notation

$$A_{0,1} = \{x : f_0(x) > f_1(x)\} = \{x : f_0(x) > 1/2\} \ ,$$

where

$$f_0 = \frac{d\nu_0}{d(\nu_0 + \nu_1)} \qquad \text{and} \qquad f_1 = \frac{d\nu_1}{d(\nu_0 + \nu_1)}.$$

The proposed test is the following: accept hypothesis H'_0 if

$$\mu_n(A_{0,1}) \geq \frac{\nu_0(A_{0,1}) + \nu_1(A_{0,1})}{2}, \tag{22.7}$$

and reject otherwise.

Then, we get that

Theorem 22.3 *The test (22.7) is strongly consistent.*

Proof Assume H_0 without loss of generality. Put

$$\epsilon = \Delta - V(\mu, \nu_0) > 0.$$

Observe that by the Scheffé theorem [22],

$$\begin{aligned}
\nu_0(A_{0,1}) - \mu(A_{0,1}) &\leq V(\nu_0, \mu) \\
&= \Delta - \epsilon \\
&= \frac{1}{2}V(\nu_0, \nu_1) - \epsilon
\end{aligned}$$

$$= \frac{1}{2}\left(\nu_0(A_{0,1}) - \nu_1(A_{0,1})\right) - \epsilon.$$

Rearranging the obtained inequality, we get that

$$\mu(A_{0,1}) \geq \frac{\nu_0(A_{0,1}) + \nu_1(A_{0,1})}{2} + \epsilon\,. \tag{22.8}$$

Therefore, (22.8) and Hoeffding's inequality [16] imply that

$$\begin{aligned}\mathbb{P}\{\text{error}\} &= \mathbb{P}\left\{\mu_n(A_{0,1}) < \frac{\nu_0(A_{0,1}) + \nu_1(A_{0,1})}{2}\right\}\\ &\leq \mathbb{P}\left\{\mu(A_{0,1}) - \mu_n(A_{0,1}) > \epsilon\right\}\\ &\leq e^{-2n\epsilon^2}.\end{aligned}$$

Therefore the Borel-Cantelli lemma implies strong consistency. □

22.3 Discussion

22.3.1 Indistinguishability

For the hypotheses H_0 and H_1 there is no positive margin, because the gap between H_0 and H_1 is just the hyperplane

$$\{\mu : V(\mu, \nu_0) = V(\mu, \nu_1)\}\,.$$

Moreover, the margin is zero:

$$\inf_{\mu\in H_0, \nu\in H_1} V(\mu, \nu) = 0.$$

Without any positive margin condition it is impossible to derive a uniform bound on the error probabilities. The pair (H_0, H_1) of hypotheses is called distinguishable if there is a sequence of uniformly consistent tests, which means that the errors of the first and second kind tend to zero uniformly. For a test T_n with sample size n, let $\alpha_{n,\mu}(T_n)$ and $\beta_{n,\mu}(T_n)$ denote the errors of the first and second kind, resp. Put

$$\alpha_n(T_n, H_0) = \sup_{\mu\in H_0} \alpha_{n,\mu}(T_n), \qquad \beta_n(T_n, H_1) = \sup_{\mu\in H_1} \beta_{n,\mu}(T_n).$$

A sequence of tests T_n, $n = 1, 2, \ldots$ is called uniformly consistent if

$$\lim_{n\to\infty} (\alpha_n(T_n, H_0) + \beta_n(T_n, H_1)) = 0.$$

It is known that a necessary condition of the distinguishable property is that for any distribution μ

$$\max\left\{\inf_{\nu\in H_0} V(\mu,\nu),\ \inf_{\nu\in H_1} V(\mu,\nu)\right\} > 0.$$

(See Barron [1], Ermakov [9], Hoeffding and Wolfowitz [17], Kraft [18], LeCam [19], LeCam and Schwartz [20], Schwartz [23].) Obviously, this necessary condition is not satisfied when $\mu^* = (\nu_1+\nu_2)/2$. Because of

$$\max\left\{\inf_{\nu\in H_0'} V(\mu^*,\nu),\ \inf_{\nu\in H_1'} V(\mu^*,\nu)\right\} = 0,$$

the pair (H_0', H_1') of hypotheses is indistinguishable, too.

22.3.2 Computation

The hypothesis testing method (22.7) proposed above is computationally quite simple. The set $A_{0,1}$ and the nominal probabilities $\nu_0(A_{0,1})$ and $\nu_1(A_{0,1})$ may be computed and stored before seeing the data. Then one merely needs to calculate $\mu_n(A_{0,1})$.

22.3.3 Hypotheses Formulated by Densities

Devroye et al. [8] formulated a special case of hypotheses (H_0', H_1'), when μ, ν_0, and ν_1 have densities f, f_0, and f_1. Under some mild margin condition they proved uniform exponential bounds for the probability of failure for $k \geq 2$ hypotheses. Moreover, they illustrated robustness of these bounds under additive noise, and showed examples where the test (22.7) is consistent and the maximum likelihood test does not work. Formally, the maximum likelihood test T_n is defined by

$$T_n = \begin{cases} 0 & \text{if } \sum_{i=1}^n (\log f_0(X_i) - \log f_1(X_i)) > 0 \\ 1 & \text{otherwise.} \end{cases}$$

For $f \in H_0'$, the strong law of large numbers implies the strong consistency of the maximum likelihood test if both integrals $\int f \log f_0$ and $\int f \log f_1$ are well defined, and

$$\int f \log f_0 > \int f \log f_1.$$

22.3.4 Robustness

Note that Theorem 22.3 does not require any assumptions about the nominal distributions. The test is robust in a very strong sense: we obtain consistency under the sole assumption that the distorted distribution remains within a certain total variation distance of the nominal distribution. For example, if μ is either $(1-\delta)\nu_0 + \delta\tau$, or $(1-\delta)\nu_1 + \delta\tau$ with arbitrary "strange" distribution τ such that $\delta < \Delta$, then we have (22.6):

$$\begin{aligned} V(\mu, \nu_0) &= V((1-\delta)\nu_0 + \delta\tau, \nu_0) \\ &= V(\delta\tau, \delta\nu_0) \\ &\leq \delta \\ &< \Delta. \end{aligned}$$

The outliers' distribution τ is really arbitrary. For example, it may not have expectations, or may even be a discrete distribution. The probability of outlier δ can be at most equal to Δ. The outliers can be formulated such that we are given three independent i.i.d. sequences $\{U_i\}$, $\{V_i\}$, $\{I_i\}$, where $\{U_i\}$, $\{V_i\}$ are $\mathbb{R}^d$-valued, and $\{I_i\}$ are binary. Put

$$X_n = (1 - I_n)U_n + I_n V_n.$$

If U_n is ν_0 distributed, V_n is τ distributed, $\mathbb{P}\{I_n = 1\} = \delta$, then we get the previous scheme. Other application include the case of censored observations, when V_n is a distortion of U_n such that some components of the vector U_n are censored. In this scheme δ is the probability of censoring. Notice that in order to estimate the distribution from censored observations one needs samples $\{(X_i, I_i)\}_{i=1}^n$ (cf. Györfi et al. [12]), while for detection it is enough to have $\{X_i\}_{i=1}^n$.

22.3.5 Open Problems

1. Characterize the distributions $\mu \in H_0 \backslash H_0'$ where the simple test (22.7) is strongly consistent. As in the proof of Theorem 22.3, strong consistency can be verified if

   $$\mu(A_{0,1}) > \frac{\nu_0(A_{0,1}) + \nu_1(A_{0,1})}{2}.$$

   We are interested in non-consistent examples, too.
2. Maybe one can improve the test (22.1), since in the construction of the partitions we don't take into account the properties of ν_0 and ν_1. For example, we can include somehow the set $A_{0,1}$.

22.3.6 Sequential Tests

We dealt with sequences of nonparametric tests with increasing sample size n, where almost surely type I and II errors occur only for finitely many n. One has to distinguish them from nonparametric sequential tests with power one (cf. Darling and Robbins [4], Sect. 6 in Robbins [21], Sect. 9.2 in Sen [24]). Such tests almost surely terminate at a random sample size with rejection of a null hypothesis H_0 after finitely many observations, if the alternative hypothesis is valid, and with positive probability do not terminate if H_0 is valid (open-ended procedures). In the latter case an upper bound of the complementary probabilities is an upper bound for the type I error probability.

References

1. Barron, A.R.: Uniformly powerful goodness of fit tests. Ann. Stat. **17**(1), 107–124 (1989)
2. Beirlant, J., Devroye, L., Györfi, L., Vajda, I.: Large deviations of divergence measures on partitions. J. Stat. Plan. Inference **93**(1–2), 1–16 (2001)
3. Biau, G., Györfi, L.: On the asymptotic properties of a nonparametric l_1-test statistic of homogeneity. IEEE Trans. Inf. Theory **51**(11), 3965–3973 (2005)
4. Darling, D.A., Robbins, H.: Some nonparametric sequential tests with power one. Proc. Natl. Acad. Sci. USA **61**(3), 804–809 (1968)
5. Dembo, A., Peres, Y.: A topological criterion for hypothesis testing. Ann. Stat. **22**(1), 106–117 (1994)
6. Devroye, L., Györfi, L.: Nonparametric Density Estimation: The L_1 View. Wiley, New York (1985)
7. Devroye, L., Lugosi, G.: Almost sure classification of densities. J. Nonparametric Stat. **14**(6), 675–698 (2002)
8. Devroye, L., Györfi, L., Lugosi, G.: A note on robust hypothesis testing. IEEE Trans. Inf. Theory **48**(7), 2111–2114 (2002)
9. Ermakov, M.: On distinguishability of hypotheses. Technical report (2013). arXiv:1308.4295
10. Gretton, A., Györfi, L.: Consistent nonparametric tests of independence. J. Mach. Learn. Res. **11**, 1391–1423 (2010)
11. Györfi, L., Györfi, Z., Vajda, I.: A strong law of large numbers and some applications. Stud. Sci. Math. Hung. **12**, 233–244 (1977)
12. Györfi, L., Kohler, M., Krzyżak, A., Walk, H.: Distribution-Free Theory of Nonparametric Regression. Springer, New York (2002)
13. Györfi, L., Masry, E.: The L_1 and L_2 strong consistency of recursive kernel density estimation from dependent samples. IEEE Trans. Inf. Theory **36**(3), 531–539 (1990)
14. Györfi, L., van der Meulen, E.C.: A consistent goodness of fit test based on the total variation distance. In: Roussas, G. (ed.) Nonparametric Functional Estimation and Related Topics, pp. 631–645. Kluwer, Dordrecht (1990)
15. Györfi, L., Walk, H.: Strongly consistent nonparametric tests of conditional independence. Stat. Probab. Lett. **82**(6), 1145–1150 (2012)
16. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. **58**(301), 13–30 (1963)
17. Hoeffding, W., Wolfowitz, J.: Distinguishability of sets of distributions. Ann. Math. Stat. **29**(3), 700–718 (1958)
18. Kraft, C.: Some conditions for consistency and uniform consistency of statistical procedures. Univ. Calif. Publ. Stat. **2**, 125–142 (1955)

19. Le Cam, L.: Convergence of estimates under dimensionality restrictions. Ann. Stat. **1**(1), 38–53 (1973)
20. Le Cam, L., Schwartz, L.: A necessary and sufficient condition for the existence of consistent estimates. Ann. Math. Stat. **31**(1), 140–150 (1960)
21. Robbins, H.: Statistical methods related to the law of the iterated logarithm. Ann. Math. Stat. **41**(5), 1397–1409 (1970)
22. Scheffé, H.: A useful convergence theorem for probability distributions. Ann. Math. Stat. **18**(3), 434–438 (1947)
23. Schwartz, L.: On Bayes procedures. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **4**(1), 10–26 (1965)
24. Sen, P.K.: Sequential Nonparametrics: Invariance Principles and Statistical Inference. Wiley, New York (1981)