On a $L_1$-test statistic of homogeneity

G. Biau, B. Cadre, L. Devroye and L. Györfi

University Paris VI — ENS Cachan — McGill University — UTE Budapest

NIPS, Whistler, December 2007
Outline

1. A $L_1$-test statistic for the two sample problem

2. Application to density model selection
Outline

1. A $L_1$-test statistic for the two sample problem

2. Application to density model selection
The problem

- Two mutually independent samples

\[ X_1, \ldots, X_n \quad \text{and} \quad X'_1, \ldots, X'_n \]

distributed according to unknown probability measures \( \mu \) and \( \mu' \) on \( \mathbb{R}^d \).

- We are interested in testing the null hypothesis that the two samples are homogeneous, that is

\[ \mathcal{H}_0 : \mu = \mu'. \]

- Such tests have been extensively studied (for an overview, see Gretton, Borgwardt, Rasch, Schölkopf and Smola, 2006).
The problem

- **Two** mutually independent samples

\[ X_1, \ldots, X_n \] and \[ X'_1, \ldots, X'_n \]

distributed according to **unknown** probability measures \( \mu \) and \( \mu' \) on \( \mathbb{R}^d \).

- **We are interested in testing the null hypothesis** that the two samples are **homogeneous**, that is

\[ \mathcal{H}_0 : \mu = \mu'. \]

- **Such tests have been extensively studied** (for an overview, see Gretton, Borgwardt, Rasch, Schölkopf and Smola, 2006).
The problem

- Two mutually independent samples
  
  \[ X_1, \ldots, X_n \text{ and } X'_1, \ldots, X'_n \]

  distributed according to unknown probability measures \( \mu \) and \( \mu' \) on \( \mathbb{R}^d \).

- We are interested in testing the null hypothesis that the two samples are homogeneous, that is
  
  \[ H_0 : \mu = \mu'. \]

- Such tests have been extensively studied (for an overview, see Gretton, Borgwardt, Rasch, Schölkopf and Smola, 2006).
The test statistic

- Based on a partition $\mathcal{P}_n = \{A_{n1}, \ldots, A_{nm_n}\}$ of $\mathbb{R}^d$, we let the test statistic be defined as

$$T_n = \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu_n'(A_{nj})|.$$ 

- Györfi and van der Meulen (1990) introduced a related goodness of fit test statistic $L_n$ defined as

$$L_n = \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu(A_{nj})|.$$
The test statistic

Based on a partition $P_n = \{A_{n1}, \ldots, A_{nm_n}\}$ of $\mathbb{R}^d$, we let the test statistic be defined as

$$T_n = \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu'_n(A_{nj})|.$$ 

Györfi and van der Meulen (1990) introduced a related goodness of fit test statistic $L_n$ defined as

$$L_n = \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu(A_{nj})|.$$
Asymptotic behavior of $L_n$

**Theorem (Devroye and Györfi, 2002)**

If

$$\lim_{n \to \infty} \frac{m_n}{n} = 0 \quad \text{and} \quad \lim_{n \to \infty} \max_{j=1, \ldots, m_n} \mu(A_{nj}) = 0,$$

then, for all $0 < \varepsilon < 2$,

$$\mathbb{P}\{L_n > \varepsilon\} = e^{-n(g_{L}(\varepsilon) + o(1))} \quad \text{as} \quad n \to \infty,$$

where

$$g_{L}(\varepsilon) = \inf_{0<p<1-\varepsilon/2} D(p \parallel p + \varepsilon/2),$$

and

$$D(\alpha \parallel \beta) = \alpha \ln \frac{\alpha}{\beta} + (1 - \alpha) \ln \frac{1 - \alpha}{1 - \beta}.$$
Asymptotic behavior of \( T_n \)

**Theorem**

*Under \( \mathcal{H}_0, \) for all \( 0 < \varepsilon < 2, \)*

\[
P\{ T_n > \varepsilon \} = e^{-n(g_T(\varepsilon) + o(1))} \quad \text{as} \ n \to \infty,
\]

*where*

\[
g_T(\varepsilon) = \left(1 + \frac{\varepsilon}{2}\right) \ln\left(1 + \frac{\varepsilon}{2}\right) + \left(1 - \frac{\varepsilon}{2}\right) \ln\left(1 - \frac{\varepsilon}{2}\right).
\]

- As \( \varepsilon \downarrow 0, \) \( g_T(\varepsilon) \approx \varepsilon^2/4, \) whereas \( g_L(\varepsilon) \approx \varepsilon^2/2. \)
- In contrast to \( g_T(\varepsilon), \) the rate function \( g_L(\varepsilon) \) is unbounded as \( \varepsilon \uparrow 2. \)
- **Conclusion:** \( L_n \) and \( T_n \) have different large deviation properties.
Asymptotic behavior of $T_n$

**Theorem**

*Under $\mathcal{H}_0$, for all $0 < \varepsilon < 2$,*

$$
\mathbb{P}\{T_n > \varepsilon\} = e^{-n(g_T(\varepsilon) + o(1))} \quad \text{as } n \to \infty,
$$

*where*

$$
g_T(\varepsilon) = (1 + \varepsilon/2) \ln(1 + \varepsilon/2) + (1 - \varepsilon/2) \ln(1 - \varepsilon/2).
$$

- As $\varepsilon \downarrow 0$, $g_T(\varepsilon) \approx \varepsilon^2/4$, whereas $g_L(\varepsilon) \approx \varepsilon^2/2$.
- In contrast to $g_T(\varepsilon)$, the rate function $g_L(\varepsilon)$ is unbounded as $\varepsilon \uparrow 2$.
- **Conclusion:** $L_n$ and $T_n$ have different large deviation properties.
Asymptotic behavior of $T_n$

**Theorem**

**Under $H_0$, for all $0 < \varepsilon < 2$,**

$$
\mathbb{P}\{T_n > \varepsilon\} = e^{-n(g_T(\varepsilon) + o(1))} \quad \text{as } n \to \infty,
$$

where

$$
g_T(\varepsilon) = (1 + \varepsilon/2) \ln(1 + \varepsilon/2) + (1 - \varepsilon/2) \ln(1 - \varepsilon/2).
$$

- As $\varepsilon \downarrow 0$, $g_T(\varepsilon) \approx \varepsilon^2/4$, whereas $g_L(\varepsilon) \approx \varepsilon^2/2$.
- In contrast to $g_T(\varepsilon)$, the rate function $g_L(\varepsilon)$ is unbounded as $\varepsilon \uparrow 2$.
- **Conclusion:** $L_n$ and $T_n$ have different large deviation properties.
Asymptotic behavior of $T_n$

**Theorem**

**Under $\mathcal{H}_0$, for all $0 < \varepsilon < 2$,**

$$\mathbb{P}\{T_n > \varepsilon\} = e^{-n(g_T(\varepsilon) + o(1))} \text{ as } n \to \infty,$$

where

$$g_T(\varepsilon) = (1 + \varepsilon/2) \ln(1 + \varepsilon/2) + (1 - \varepsilon/2) \ln(1 - \varepsilon/2).$$

- As $\varepsilon \downarrow 0$, $g_T(\varepsilon) \approx \varepsilon^2/4$, whereas $g_L(\varepsilon) \approx \varepsilon^2/2$.
- In contrast to $g_T(\varepsilon)$, the rate function $g_L(\varepsilon)$ is unbounded as $\varepsilon \uparrow 2$.
- **Conclusion:** $L_n$ and $T_n$ have different large deviation properties.
Rate functions $g_L$ and $g_T$

\[ g_L(\varepsilon/2) \leq g_T(\varepsilon) \leq g_L(\varepsilon) \]
Sketch of proof

- **Generating function** of the sequence \((T_n)\):

  \[
  \lambda_T(s) = \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}\{e^{snT_n}\}, \quad s > 0.
  \]

- By Scheffé’s theorem for partitions:

  \[
  T_n = \sum_{A \in \mathcal{P}_n} |\mu_n(A) - \mu'_n(A)| = 2 \max_{A \in \sigma(\mathcal{P}_n)} (\mu_n(A) - \mu'_n(A)).
  \]

- Thus,

  \[
  \mathbb{E}\{e^{snT_n}\} = \mathbb{E} \left\{ \max_{A \in \sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A) - \mu'_n(A))} \right\}
  \leq 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn(\mu_n(A) - \mu'_n(A))}\}
  \]

  \[
  = 2^{m_n} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn\mu_n(A)}\} \mathbb{E}\{e^{-2sn\mu'_n(A)}\}
  \]

  \[
  \leq 2^{m_n} \left[ \frac{1}{2} + \left( e^{2s} + e^{-2s} \right)/4 \right]^n.
  \]
Sketch of proof

**Generating function** of the sequence \((T_n)\):

\[
\lambda_T(s) = \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}\{e^{snT_n}\}, \quad s > 0.
\]

By **Scheffé’s theorem** for partitions:

\[
T_n = \sum_{A \in \mathcal{P}_n} |\mu_n(A) - \mu'_n(A)| = 2 \max_{A \in \sigma(\mathcal{P}_n)} (\mu_n(A) - \mu'_n(A)).
\]

Thus,

\[
\mathbb{E}\{e^{snT_n}\} = \mathbb{E}\left\{ \max_{A \in \sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A) - \mu'_n(A))} \right\}
\leq 2^{mn} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn(\mu_n(A) - \mu'_n(A))}\}
= 2^{mn} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn\mu_n(A)}\} \mathbb{E}\{e^{-2sn\mu'_n(A)}\}
\leq 2^{mn} \left[ 1/2 + (e^{2s} + e^{-2s})/4 \right]^n.
\]
Sketch of proof

- Generating function of the sequence \((T_n)\):
  \[
  \lambda_T(s) = \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}\{e^{snT_n}\}, \quad s > 0.
  \]

- By Scheffé’s theorem for partitions:
  \[
  T_n = \sum_{A \in \mathcal{P}_n} |\mu_n(A) - \mu'_n(A)| = 2 \max_{A \in \sigma(\mathcal{P}_n)} (\mu_n(A) - \mu'_n(A)).
  \]

- Thus,
  \[
  \mathbb{E}\{e^{snT_n}\} = \mathbb{E} \left\{ \max_{A \in \sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A) - \mu'_n(A))} \right\}
  \leq 2^{mn} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn(\mu_n(A) - \mu'_n(A))}\}
  = 2^{mn} \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{e^{2sn\mu_n(A)}\} \mathbb{E}\{e^{-2sn\mu'_n(A)}\}
  \leq 2^{mn} \left[ 1/2 + (e^{2s} + e^{-2s})/4 \right]^n.
  \]
Sketch of proof

This implies that

\[ \lambda_T(s) \leq \ln(1/2 + (e^{2s} + e^{-2s})/4). \]

Similarly,

\[
\mathbb{E}\{e^{snT_n}\} = \mathbb{E}\{ \max_{A \in \sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A) - \mu'_n(A))} \} \\
\geq \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{ e^{2sn(\mu_n(A) - \mu'_n(A))} \} \\
= \max_{A \in \sigma(\mathcal{P}_n)} \mathbb{E}\{ e^{2sn\mu_n(A)} \} \mathbb{E}\{ e^{-2sn\mu'_n(A)} \},
\]

which implies

\[ \lambda_T(s) \geq \ln(1/2 + (e^{2s} + e^{-2s})/4). \]

Conclusion by Gärtner-Ellis theorem:

\[ g_T(\varepsilon) = \max_{s>0} (s\varepsilon - \lambda_T(s)) . \]
Sketch of proof

This implies that

\[ \lambda_T(s) \leq \ln(1/2 + (e^{2s} + e^{-2s})/4). \]

Similarly,

\[ \mathbb{E}\{e^{snT_n}\} = \mathbb{E}\left\{ \max_{A \in \sigma(P_n)} e^{2sn(\mu_n(A) - \mu'_n(A))} \right\} \]

\[ \geq \max_{A \in \sigma(P_n)} \mathbb{E}\{e^{2sn(\mu_n(A) - \mu'_n(A))}\} \]

\[ = \max_{A \in \sigma(P_n)} \mathbb{E}\{e^{2sn\mu_n(A)}\} \mathbb{E}\{e^{-2sn\mu'_n(A)}\}, \]

which implies

\[ \lambda_T(s) \geq \ln(1/2 + (e^{2s} + e^{-2s})/4). \]

Conclusion by Gärtner-Ellis theorem:

\[ g_T(\varepsilon) = \max_{s > 0} (s\varepsilon - \lambda_T(s)). \]
Sketch of proof

This implies that

$$\lambda_T(s) \leq \ln(1/2 + (e^{2s} + e^{-2s})/4).$$

Similarly,

$$\mathbb{E}\{e^{snT_n}\} = \mathbb{E}\left\{ \max_{A \in \sigma(P_n)} e^{2sn(\mu_n(A) - \mu'_n(A))} \right\}$$

$$\geq \max_{A \in \sigma(P_n)} \mathbb{E}\{e^{2sn(\mu_n(A) - \mu'_n(A))}\}$$

$$= \max_{A \in \sigma(P_n)} \mathbb{E}\{e^{2sn\mu_n(A)}\}\mathbb{E}\{e^{-2sn\mu'_n(A)}\},$$

which implies

$$\lambda_T(s) \geq \ln(1/2 + (e^{2s} + e^{-2s})/4).$$

Conclusion by Gärtner-Ellis theorem:

$$g_T(\varepsilon) = \max_{s > 0} (s\varepsilon - \lambda_T(s)).$$
Sketch of proof

- This implies that
  \[ \lambda_T(s) \leq \ln(1/2 + (\text{e}^{2s} + \text{e}^{-2s})/4). \]

- Similarly,
  \[
  \mathbb{E}\{ e^{snT_n} \} = \mathbb{E}\{ \max_{A \in \sigma(P_n)} e^{2sn(\mu_n(A) - \mu'_n(A))} \}
  \geq \max_{A \in \sigma(P_n)} \mathbb{E}\{ e^{2sn(\mu_n(A) - \mu'_n(A))} \}
  = \max_{A \in \sigma(P_n)} \mathbb{E}\{ e^{2sn\mu_n(A)} \} \mathbb{E}\{ e^{-2sn\mu'_n(A)} \},
  \]

  which implies
  \[ \lambda_T(s) \geq \ln(1/2 + (\text{e}^{2s} + \text{e}^{-2s})/4). \]

- Conclusion by Gärtner-Ellis theorem:
  \[ g_T(\varepsilon) = \max_{s > 0} (s\varepsilon - \lambda_T(s)). \]
This technique yields a distribution-free strong consistent test of homogeneity, which rejects the null hypothesis if $T_n$ becomes large.

It means that both on $\mathcal{H}_0$ and on its complement the test makes a.s. no error after a random sample size.

In other words, we have

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

and

$$\mathbb{P}_1\{\text{accepting } \mathcal{H}_0 \text{ for only finitely many } n\} = 1.$$
This technique yields a distribution-free strong consistent test of homogeneity, which rejects the null hypothesis if $T_n$ becomes large.

It means that both on $\mathcal{H}_0$ and on its complement the test makes a.s. no error after a random sample size.

In other words, we have

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

and

$$\mathbb{P}_1\{\text{accepting } \mathcal{H}_0 \text{ for only finitely many } n\} = 1.$$
This technique yields a **distribution-free strong consistent test of homogeneity**, which rejects the null hypothesis if $T_n$ becomes large.

It means that both on $H_0$ and on its complement the test makes a.s. no error after a random sample size.

In other words, we have

$$P_0\{\text{rejecting } H_0 \text{ for only finitely many } n\} = 1$$

and

$$P_1\{\text{accepting } H_0 \text{ for only finitely many } n\} = 1.$$
A strong consistent test

Corollary

Consider the test which rejects \( \mathcal{H}_0 \) when

\[
T_n > c_1 \sqrt{\frac{m_n}{n}},
\]

where \( c_1 > 2\sqrt{\ln 2} \approx 1.6651 \). Assume that

\[
\lim_{n \to \infty} \frac{m_n}{n} = 0 \quad \text{and} \quad \lim_{n \to \infty} \frac{m_n}{\ln n} = \infty.
\]

Then, under \( \mathcal{H}_0 \), after a random sample size the test makes a.s. no error. Moreover, if \( \mu \neq \mu' \), and for any sphere \( S \) centered at the origin

\[
\lim_{n \to \infty} \max_{A_{nj} \cap S \neq 0} \text{diam}(A_{nj}) = 0,
\]

then after a random sample size the test makes a.s. no error.
Beirlant, Györfi and Lugosi (1994) proved that

\[ \sqrt{n} (L_n - \mathbb{E}\{L_n\}) / \sigma \xrightarrow{D} \mathcal{N}(0, 1), \]

where \( \sigma^2 = 1 - 2/\pi \).

Their technique involves a Poisson representation of the empirical process in conjunction with Bartlett’s (1938) idea of partial inversion for obtaining characteristic functions of conditional distributions.
Beirlant, Györfi and Lugosi (1994) proved that
\[ \sqrt{n}(L_n - \mathbb{E}\{L_n\}) / \sigma \xrightarrow{D} \mathcal{N}(0, 1), \]
where \( \sigma^2 = 1 - 2/\pi \).

Their technique involves a Poisson representation of the empirical process in conjunction with Bartlett’s (1938) idea of partial inversion for obtaining characteristic functions of conditional distributions.
Asymptotic normality

**Theorem**

If

\[
\lim_{n \to \infty} \frac{m_n}{n} = 0 \quad \text{and} \quad \lim_{n \to \infty} \max_{j=1,\ldots,m_n} \mu(A_{nj}) = 0,
\]

then, under $\mathcal{H}_0$, with a centering sequence $(C_n)$,

\[
\sqrt{n}(T_n - C_n) / \sigma \xrightarrow{D} \mathcal{N}(0, 1),
\]

where $\sigma^2 = 2(1 - 2/\pi)$. 

G. Biau (Université Paris VI)
**Sketch of proof**

- **Difficulty:** $T_n$ is a sum of dependent random variables.
- To overcome this problem, we use a ‘Poissonization’ argument.
- Denote by $N_n$ and $N'_n$ two independent Poisson $(n)$ random variables independent of $(X_i)_{i \geq 1}$ and $(X'_i)_{i \geq 1}$.
- The Poissonized version $\tilde{T}_n$ of $T_n$ is then defined by

$$\tilde{T}_n = \sum_{j=1}^{m_n} |\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|,$$

where, for any Borel subset $A$,

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

and, similarly,

$$\mu'_{N'_n}(A) = \frac{\# \{i : X'_i \in A, i = 1, \ldots, N'_n \}}{n}.$$
Sketch of proof

**Difficulty:** $T_n$ is a sum of dependent random variables.

To overcome this problem, we use a ‘Poissonization’ argument.

Denote by $N_n$ and $N'_n$ two independent Poisson $(n)$ random variables independent of $(X_i)_{i \geq 1}$ and $(X'_i)_{i \geq 1}$.

The Poissonized version $\tilde{T}_n$ of $T_n$ is then defined by

$$\tilde{T}_n = \sum_{j=1}^{m_n} |\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|,$$

where, for any Borel subset $A$,

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

and, similarly,

$$\mu'_{N'_n}(A) = \frac{\# \{ i : X'_i \in A, i = 1, \ldots, N'_n \}}{n}.$$
Sketch of proof

- **Difficulty**: $T_n$ is a sum of dependent random variables.
- To overcome this problem, we use a ‘Poissonization’ argument.
- Denote by $N_n$ and $N'_n$ two independent Poisson ($n$) random variables independent of $(X_i)_{i \geq 1}$ and $(X'_i)_{i \geq 1}$.
- The Poissonized version $\tilde{T}_n$ of $T_n$ is then defined by

$$\tilde{T}_n = \sum_{j=1}^{m_n} |\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|,$$

where, for any Borel subset $A$,

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

and, similarly,

$$\mu'_{N'_n}(A) = \frac{\# \{ i : X'_i \in A, i = 1, \ldots, N'_n \}}{n}.$$
**Sketch of proof**

- **Difficulty:** $T_n$ is a sum of dependent random variables.
- To overcome this problem, we use a ‘Poissonization’ argument.
- Denote by $N_n$ and $N'_n$ two independent Poisson $(n)$ random variables independent of $(X_i)_{i \geq 1}$ and $(X'_i)_{i \geq 1}$.
- The Poissonized version $\tilde{T}_n$ of $T_n$ is then defined by

$$\tilde{T}_n = \sum_{j=1}^{m_n} |\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|,$$

where, for any Borel subset $A$,

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

and, similarly,

$$\mu'_{N'_n}(A) = \frac{\#\{i : X'_i \in A, i = 1, \ldots, N'_n\}}{n}.$$
Sketch of proof

Setting

\[ Y_n = (n\mu_{N_n}(A_{n1}), \ldots, n\mu_{N_n}(A_{nm_n})) \]

and

\[ Y'_n = (n\mu'_{N'_n}(A_{n1}), \ldots, n\mu'_{N'_n}(A_{nm_n})) , \]

one shows that \( Y_n \) and \( Y'_n \) are independent vectors of independent random variables with

\[ (n\mu_{N_n}(A_{nj})) \overset{D}{=} (n\mu'_{N'_n}(A_{nj})) \overset{D}{=} \text{Poisson} \left( n\mu(A_{nj}) \right) . \]

Moreover,

\[ (Y_n|N_n = n) \overset{D}{=} (Y'_n|N'_n = n) \overset{D}{=} \text{Multinomial} \left( n; \mu(A_{n1}), \ldots, \mu(A_{nm_n}) \right) . \]

The key of the proof is the following property, which uses Fourier’s inversion formula.
Sketch of proof

- Setting

\[ Y_n = (n\mu_{N_n}(A_{n1}), \ldots, n\mu_{N_n}(A_{nm_n})) \]

and

\[ Y'_n = (n\mu'_{N'_n}(A_{n1}), \ldots, n\mu'_{N'_n}(A_{nm_n})) \],

one shows that \( Y_n \) and \( Y'_n \) are independent vectors of independent random variables with

\[ \left( n\mu_{N_n}(A_{nj}) \right) \overset{\mathcal{D}}{=} \left( n\mu'_{N'_n}(A_{nj}) \right) \overset{\mathcal{D}}{=} \text{Poisson} \left( n\mu(A_{nj}) \right). \]

- Moreover,

\[ (Y_n|N_n = n) \overset{\mathcal{D}}{=} (Y'_n|N'_n = n) \overset{\mathcal{D}}{=} \text{Multinomial} \left( n; \mu(A_{n1}), \ldots, \mu(A_{nm_n}) \right). \]

- The key of the proof is the following property, which uses Fourier’s inversion formula.
Sketch of proof

- Setting

\[ Y_n = (n\mu_{N_n}(A_{n1}), \ldots, n\mu_{N_n}(A_{nm_n})) \]

and

\[ Y'_n = (n\mu'_{N'_n}(A_{n1}), \ldots, n\mu'_{N'_n}(A_{nm_n})) \],

one shows that \( Y_n \) and \( Y'_n \) are independent vectors of independent random variables with

\[ (n\mu_{N_n}(A_{nj})) \overset{\mathcal{D}}{=} (n\mu'_{N'_n}(A_{nj})) \overset{\mathcal{D}}{=} \text{Poisson} \left( n\mu(A_{nj}) \right) . \]

- Moreover,

\[ (Y_n|N_n = n) \overset{\mathcal{D}}{=} (Y'_n|N'_n = n) \overset{\mathcal{D}}{=} \text{Multinomial} \left( n; \mu(A_{n1}), \ldots, \mu(A_{nm_n}) \right) . \]

- The key of the proof is the following property, which uses Fourier's inversion formula.
Proposition

Let \( g_{nj} (j = 1, \ldots, m_n) \) be real measurable functions, with

\[
\mathbb{E} \left\{ g_{nj} \left( \mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj}) \right) \right\} = 0,
\]

and let

\[
M_n = \sum_{j=1}^{m_n} g_{nj} \left( \mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj}) \right).
\]

Assume that

\[
\left( M_n, \frac{N_n - n}{\sqrt{n}}, \frac{N'_n - n}{\sqrt{n}} \right) \overset{D}{\to} \mathcal{N}_3(0, 0, 0, \sigma^2, 1, 1),
\]

as \( n \to \infty \), where \( \sigma \) is a positive constant. Then

\[
\frac{1}{\sigma} \sum_{j=1}^{m_n} g_{nj} \left( \mu_n(A_{nj}) - \mu'_n(A_{nj}) \right) \overset{D}{\to} \mathcal{N}(0, 1).
\]
Corollary

Put $\alpha \in (0, 1)$, $C^* = 0.7655$, and consider the test which rejects $\mathcal{H}_0$ when

$$T_n > c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha),$$

where $c_2 = 2/\sqrt{\pi} \approx 1.1284$. Then the test has asymptotic significance level $\alpha$. Moreover, under the additional condition

$$\lim_{n \to \infty} \max_{A_{nj} \cap S \neq 0} \text{diam}(A_{nj}) = 0,$$

the test is consistent.
1 A $L_1$-test statistic for the two sample problem

2 Application to density model selection
We wish to estimate a density $f$ on $\mathbb{R}^d$ that belongs to a parametric family, $\mathcal{F}_k$, where $k$ is unknown, but $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ for all $k$.

$$\mathcal{F} = \bigcup_{k \geq 1} \mathcal{F}_k.$$ 

Formally, we let the complexity associated with $f$ be defined as

$$k^* = \min\{k \geq 1 : f \in \mathcal{F}_k\}.$$
The problem

- We wish to estimate a density $f$ on $\mathbb{R}^d$ that belongs to a \textit{parametric family}, $\mathcal{F}_k$, where $k$ is unknown, but $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ for all $k$.

$$\mathcal{F} = \bigcup_{k \geq 1} \mathcal{F}_k.$$ 

- Formally, we let the \textit{complexity} associated with $f$ be defined as

$$k^* = \min\{k \geq 1 : f \in \mathcal{F}_k\}.$$
Objective

We wish to pick a density estimate $\hat{f}_{K_n}$ in $\mathcal{F}$ with

(i) $K_n \to k^*$ almost surely

(ii) and

$$\mathbb{E} \left\{ \int |\hat{f}_{K_n} - f| \right\} = O \left( \frac{1}{\sqrt{n}} \right).$$

$K_n$ is obtained by minimizing the $L_1$ error between candidate models and the empirical measure.

The model parameters are selected using the general combinatorial tools developed in Devroye and Lugosi (2001).
Objective

We wish to pick a density estimate $\hat{f}_{K_n}$ in $\mathcal{F}$ with

(i) $K_n \to k^*$ almost surely

(ii) and

$$\mathbb{E} \left\{ \int |\hat{f}_{K_n} - f| \right\} = O \left( \frac{1}{\sqrt{n}} \right).$$

$K_n$ is obtained by minimizing the $L_1$ error between candidate models and the empirical measure.

The model parameters are selected using the general combinatorial tools developed in Devroye and Lugosi (2001).
Objective

- We wish to pick a density estimate \( \hat{f}_{K_n} \) in \( \mathcal{F} \) with
  
  (i) \( K_n \to k^* \) almost surely
  
  (ii) and

  \[
  \mathbb{E} \left\{ \int |\hat{f}_{K_n} - f| \right\} = O \left( \frac{1}{\sqrt{n}} \right).
  \]

- \( K_n \) is obtained by minimizing the \( L_1 \) error between candidate models and the empirical measure.

- The model parameters are selected using the general combinatorial tools developed in Devroye and Lugosi (2001).
Objective

- We wish to pick a density estimate $\hat{f}_{K_n}$ in $F$ with
  
  (i) $K_n \to k^*$ almost surely

  (ii) and

  $$
  \mathbb{E} \left\{ \int |\hat{f}_{K_n} - f| \right\} = O \left( \frac{1}{\sqrt{n}} \right).
  $$

- $K_n$ is obtained by minimizing the $L_1$ error between candidate models and the empirical measure.

- The model parameters are selected using the general combinatorial tools developed in Devroye and Lugosi (2001).
Examples I

**Mixture classes.** Consider first the classes $\mathcal{F}_k$ of all mixtures of $k$ normal densities over $\mathbb{R}^d$,

$$f_k(x) = \sum_{i=1}^{k} \frac{p_i}{\sqrt{(2\pi)^d \det(\Sigma_i)}} \ e^{-\frac{1}{2}(x-m_i)^T \Sigma_i^{-1}(x-m_i)}.$$

→ **Bayesian literature**: Hurn, Justel and Robert (2003).
→ **Statistical learning literature**: Figueiredo and Jain (2002).
→ **Clustering literature**: Fukumizu (2002).
→ **Statistical literature**: Dacunha–Castelle and Gassiat (1997).
Mixture classes. Consider first the classes $\mathcal{F}_k$ of all mixtures of $k$ normal densities over $\mathbb{R}^d$,

$$f_k(x) = \sum_{i=1}^{k} \frac{p_i}{\sqrt{(2\pi)^d \det(\Sigma_i)}} e^{-\frac{1}{2}(x-m_i)^T \Sigma_i^{-1} (x-m_i)}.$$ 

→ **Bayesian literature**: Hurn, Justel and Robert (2003).
→ **Statistical learning literature**: Figueiredo and Jain (2002).
→ **Clustering literature**: Fukumizu (2002).
→ **Statistical literature**: Dacunha-Castelle and Gassiat (1997).
Mixture classes. Consider first the classes $\mathcal{F}_k$ of all mixtures of $k$ normal densities over $\mathbb{R}^d$,

$$f_k(x) = \sum_{i=1}^{k} \frac{p_i}{\sqrt{(2\pi)^d \det(\Sigma_i)}} e^{-\frac{1}{2}(x-m_i)^T\Sigma_i^{-1}(x-m_i)}.$$ 

→ **Bayesian literature**: Hurn, Justel and Robert (2003).
→ **Statistical learning literature**: Figueiredo and Jain (2002).
→ **Clustering literature**: Fukumizu (2002).
→ **Statistical literature**: Dacunha-Castelle and Gassiat (1997).
Mixture classes. Consider first the classes $\mathcal{F}_k$ of all mixtures of $k$ normal densities over $\mathbb{R}^d$,

$$f_k(x) = \sum_{i=1}^{k} \frac{p_i}{\sqrt{(2\pi)^d \det(\Sigma_i)}} e^{-\frac{1}{2}(x-m_i)^T \Sigma_i^{-1}(x-m_i)}.$$ 

→ Clustering literature: Fukumizu (2002).
Mixture classes. Consider first the classes $\mathcal{F}_k$ of all mixtures of $k$ normal densities over $\mathbb{R}^d$,

$$f_k(x) = \sum_{i=1}^{k} \frac{p_i}{\sqrt{(2\pi)^d \det(\Sigma_i)}} \ e^{-\frac{1}{2}(x-m_i)^T \Sigma_i^{-1}(x-m_i)}.$$ 

→ **Bayesian literature:** Hurn, Justel and Robert (2003).  
→ **Statistical learning literature:** Figueiredo and Jain (2002).  
→ **Clustering literature:** Fukumizu (2002).  
→ **Statistical literature:** Dacunha-Castelle and Gassiat (1997).
Increasing exponential families. Each density $f_k$ in an exponential family $\mathcal{F}_k$ may be written in the form

$$f_k(x) = c\alpha(\theta)\beta(x)e^{\sum_{i=1}^{k} \pi_i(\theta)\psi_i(x)}.$$ 

Examples of exponential families include classes of Gaussian, gamma, beta, Rayleigh, and Maxwell densities.

Other models are feasible: series estimates, neural network estimates, wavelets...

We require that the Vapnik-Chervonenkis dimension of $\mathcal{F}_k^*$ is finite.
Increasing exponential families. Each density $f_k$ in an exponential family $F_k$ may be written in the form

$$f_k(x) = c \alpha(\theta) \beta(x) e^{\sum_{i=1}^{k} \pi_i(\theta) \psi_i(x)}.$$

Examples of exponential families include classes of Gaussian, gamma, beta, Rayleigh, and Maxwell densities.

Other models are feasible: series estimates, neural network estimates, wavelets...

We require that the Vapnik-Chervonenkis dimension of $F_k$ is finite.
Examples II

- **Increasing exponential families.** Each density $f_k$ in an exponential family $\mathcal{F}_k$ may be written in the form

  $$f_k(x) = c \alpha(\theta) \beta(x) e^{\sum_{i=1}^{k} \pi_i(\theta) \psi_i(x)}.$$

- Examples of exponential families include classes of Gaussian, gamma, beta, Rayleigh, and Maxwell densities.

- Other models are feasible: series estimates, neural network estimates, wavelets...

- We require that the Vapnik-Chervonenkis dimension of $\mathcal{F}_{k^*}$ is finite.
Increasing exponential families. Each density $f_k$ in an exponential family $\mathcal{F}_k$ may be written in the form

$$f_k(x) = c\alpha(\theta)\beta(x)e^{\sum_{i=1}^{k} \pi_i(\theta)\psi_i(x)}.$$ 

Examples of exponential families include classes of Gaussian, gamma, beta, Rayleigh, and Maxwell densities.

Other models are feasible: series estimates, neural network estimates, wavelets...

We require that the Vapnik-Chervonenkis dimension of $\mathcal{F}_k^*$ is finite.
A closure condition

- Let $\mathcal{D}$ be the class of all density functions on $\mathbb{R}^d$ and $\hat{\mathcal{D}}$ the set of Fourier transforms $\hat{g}$.

Assumption

The set $\hat{F}_k$ is closed in $\hat{\mathcal{D}}$.

- By Paul Lévy’s theorem, this is equivalent to require that for any sequence $(g_n)$ in $F_k$ satisfying

$$\lim_{n \to \infty} \int g_n(x)\varphi(x) \, dx = \int g(x)\varphi(x) \, dx$$

for every bounded, continuous real function $\varphi$, one has in fact $g \in F_k$. 

G. Biau (Université Paris VI)
A closure condition

- Let $\mathcal{D}$ be the class of all density functions on $\mathbb{R}^d$ and $\hat{\mathcal{D}}$ the set of Fourier transforms $\hat{g}$.

**Assumption**

The set $\hat{F}_k$ is closed in $\hat{\mathcal{D}}$.

- By Paul Lévy's theorem, this is equivalent to require that for any sequence $(g_n)$ in $F_k$ satisfying

$$\lim_{n \to \infty} \int g_n(x) \varphi(x) \, dx = \int g(x) \varphi(x) \, dx$$

for every bounded, continuous real function $\varphi$, one has in fact $g \in F_k$. 
A closure condition

- Let $\mathcal{D}$ be the class of all density functions on $\mathbb{R}^d$ and $\hat{\mathcal{D}}$ the set of Fourier transforms $\hat{g}$.

**Assumption**

The set $\hat{\mathcal{F}}_k$ is closed in $\hat{\mathcal{D}}$.

- By Paul Lévy’s theorem, this is equivalent to require that for any sequence $(g_n)$ in $\mathcal{F}_k$ satisfying

\[
\lim_{n \to \infty} \int g_n(x) \varphi(x) \, dx = \int g(x) \varphi(x) \, dx
\]

for every bounded, continuous real function $\varphi$, one has in fact $g \in \mathcal{F}_k$. 
Split the sample into two subsamples:
\[ \{ X_1, \ldots, X_n \} \text{ and } \{ X'_1, \ldots, X'_n \} = \{ X_{n+1}, \ldots, X_{2n} \}. \]

Let \( P_n = \{ A_{nj} : j \geq 1 \} \) be a cubic partition of \( \mathbb{R}^d \) with volume \( h_n^d \).

Introduce the statistic
\[
d_{n,k} = \inf_{g \in F_k} \sum_{A \in P_n} \left| \int_A g - \mu_{2n}(A) \right|. \]

Let the threshold be
\[
T_n = \sum_{A \in P_n} |\mu_n(A) - \mu'_n(A)|. \]

Estimate of \( k^* \):
\[
K_n = \min\{ k \geq 1 : d_{n,k} \leq T_n \}. \]
Complexity estimation

- Split the sample into two subsamples:
  \[
  \{X_1, \ldots, X_n\} \quad \text{and} \quad \{X'_1, \ldots, X'_n\} = \{X_{n+1}, \ldots, X_{2n}\}.
  \]

- Let \( P_n = \{A_{nj} : j \geq 1\} \) be a cubic partition of \( \mathbb{R}^d \) with volume \( h^d_n \).
- Introduce the statistic
  \[
  d_{n,k} = \inf_{g \in \mathcal{F}_k} \sum_{A \in P_n} \left| \int_A g - \mu_{2n}(A) \right|.
  \]

- Let the threshold be
  \[
  T_n = \sum_{A \in P_n} |\mu_n(A) - \mu'_n(A)|.
  \]

- Estimate of \( k^* \):
  \[
  K_n = \min\{k \geq 1 : d_{n,k} \leq T_n\}.
  \]
Complexity estimation

- Split the sample into **two subsamples**:
  \[
  \{X_1, \ldots, X_n\} \quad \text{and} \quad \{X'_1, \ldots, X'_n\} = \{X_{n+1}, \ldots, X_{2n}\}.
  \]

- Let \(\mathcal{P}_n = \{A_{nj} : j \geq 1\}\) be a cubic partition of \(\mathbb{R}^d\) with volume \(h_n^d\).
- Introduce the statistic
  \[
  d_{n,k} = \inf_{g \in \mathcal{F}_k} \sum_{A \in \mathcal{P}_n} \left| \int_A g - \mu_{2n}(A) \right|.
  \]

- Let the **threshold** be
  \[
  T_n = \sum_{A \in \mathcal{P}_n} |\mu_n(A) - \mu'_n(A)|.
  \]

- **Estimate of \(k^*\):**
  \[
  K_n = \min\{k \geq 1 : d_{n,k} \leq T_n\}.
  \]
Complexity estimation

- Split the sample into two subsamples:
  \[
  \{X_1, \ldots, X_n\} \quad \text{and} \quad \{X'_1, \ldots, X'_n\} = \{X_{n+1}, \ldots, X_{2n}\}.
  \]

- Let \( \mathcal{P}_n = \{A_{nj} : j \geq 1\} \) be a cubic partition of \( \mathbb{R}^d \) with volume \( h_n^d \).
- Introduce the statistic
  \[
  d_{n,k} = \inf_{g \in \mathcal{F}_k} \sum_{A \in \mathcal{P}_n} \left| \int_A g - \mu_{2n}(A) \right|.
  \]

- Let the threshold be
  \[
  T_n = \sum_{A \in \mathcal{P}_n} |\mu_n(A) - \mu'_n(A)|.
  \]

- Estimate of \( k^* \):
  \[
  K_n = \min\{k \geq 1 : d_{n,k} \leq T_n\}.
  \]
Split the sample into two subsamples:
\[ \{X_1, \ldots, X_n\} \quad \text{and} \quad \{X'_1, \ldots, X'_n\} = \{X_{n+1}, \ldots, X_{2n}\}. \]

Let \( P_n = \{A_{nj} : j \geq 1\} \) be a cubic partition of \( \mathbb{R}^d \) with volume \( h_n^d \).

Introduce the statistic
\[
d_{n,k} = \inf_{g \in \mathcal{F}_k} \sum_{A \in P_n} \left| \int_A g - \mu_{2n}(A) \right|.\]

Let the threshold be
\[
T_n = \sum_{A \in P_n} |\mu_n(A) - \mu'_n(A)|.
\]

Estimate of \( k^* \):
\[
K_n = \min\{k \geq 1 : d_{n,k} \leq T_n\}.
\]
Complexity estimation

- Our estimate of $k^*$:

\[ K_n = \min\{k \geq 1 : d_{n,k} \leq T_n\}. \]

**Theorem**

Choose $h_n = n^{-\delta}$ with $0 < \delta < 1/d$. Then there exists a positive constant $\kappa$, depending on $f$, such that

\[ \mathbb{P}\{K_n \neq k^*\} \leq \exp\left(-\kappa n^{d\delta}\right), \]

and consequently, almost surely,

\[ K_n = k^* \]

for all $n$ large enough.
Our estimate of $k^*$:

$$K_n = \min\{k \geq 1 : d_{n,k} \leq T_n\}.$$ 

**Theorem**

Choose $h_n = n^{-\delta}$ with $0 < \delta < 1/d$. Then there exists a positive constant $\kappa$, depending on $f$, such that

$$\mathbb{P}\{K_n \neq k^*\} \leq \exp\left(-\kappa n^{d\delta}\right),$$

and consequently, almost surely,

$$K_n = k^*$$

for all $n$ large enough.
Fast density estimate

- Fix $k \geq 1$ and introduce the class of sets
  
  $A_k = \{ \{ x : g_1(x) > g_2(x) \} : g_1, g_2 \in \mathcal{F}_k \}$

  and the goodness criterion for a density $g \in \mathcal{F}_k$:
  
  $$\Delta_k(g) = \sup_{A \in A_k} \left| \int_A g - \mu_{2n}(A) \right|.$$ 

- The minimum distance estimate $\hat{f}_k$ minimizes the criterion $\Delta_k(g)$ over all $g$ in $\mathcal{F}_k$. 
Fix $k \geq 1$ and introduce the class of sets

$$A_k = \left\{ \{ x : g_1(x) > g_2(x) \} : g_1, g_2 \in \mathcal{F}_k \right\}$$

and the goodness criterion for a density $g \in \mathcal{F}_k$:

$$\Delta_k(g) = \sup_{A \in \mathcal{A}_k} \left| \int_A g - \mu_{2n}(A) \right|.$$

The minimum distance estimate $\hat{f}_k$ minimizes the criterion $\Delta_k(g)$ over all $g$ in $\mathcal{F}_k$. 
Fast density estimate

For the elected minimum distance estimate $\hat{f}_k$, we have [Devroye and Lugosi (2001)]

$$\int |\hat{f}_k - f| \leq 3 \inf_{g \in \mathcal{F}_k} \int |g - f| + 4\Delta_k(f) + \frac{3}{2n}.$$ 

The minimum distance estimate $\hat{f}_{Kn}$ is a natural candidate for the estimation of $f$.

We deduce that

$$\mathbb{E}\left\{ \int |\hat{f}_{Kn} - f| \right\} \leq 4\mathbb{E}\{\Delta_k^*(f)\} + \frac{3}{2n} + 2\exp\left(-\kappa n^{d\delta}\right),$$

where $\Delta_k^*(f) = \sup_{A \in \mathcal{A}_k^*} \left| \int_A f - \mu_{2n}(A) \right|$. 

G. Biau (Université Paris VI)
Fast density estimate

- For the elected minimum distance estimate $\hat{f}_k$, we have [Devroye and Lugosi (2001)]

$$\int |\hat{f}_k - f| \leq 3 \inf_{g \in \mathcal{F}_k} \int |g - f| + 4\Delta_k(f) + \frac{3}{2n}.$$ 

- The minimum distance estimate $\hat{f}_{Kn}$ is a natural candidate for the estimation of $f$.

- We deduce that

$$\mathbb{E}\left\{ \int |\hat{f}_{Kn} - f| \right\} \leq 4\mathbb{E}\{\Delta_k^*(f)\} + \frac{3}{2n} + 2 \exp\left(-\kappa n^{d_\delta}\right),$$

where $\Delta_k^*(f) = \sup_{A \in \mathcal{A}_k^*} \left| \int_A f - \mu_{2n}(A) \right|$. 
For the elected minimum distance estimate \( \hat{f}_k \), we have \cite{Devroye:2001}

\[
\int |\hat{f}_k - f| \leq 3 \inf_{g \in \mathcal{F}_k} \int |g - f| + 4\Delta_k(f) + \frac{3}{2n}.
\]

The minimum distance estimate \( \hat{f}_{Kn} \) is a natural candidate for the estimation of \( f \).

We deduce that

\[
\mathbb{E}\left\{ \int |\hat{f}_{Kn} - f| \right\} \leq 4\mathbb{E}\{\Delta_k^*(f)\} + \frac{3}{2n} + 2 \exp\left(-\kappa n^{d\delta}\right),
\]

where \( \Delta_k^*(f) = \sup_{A \in \mathcal{A}_k^*} \left| \int_A f - \mu_{2n}(A) \right| \).
Link with the VC theory

- If $A_{k^*}$ has Vapnik-Chervonenkis dimension $V_{k^*}$, then

$$
\mathbb{E} \{ \Delta_{k^*}(f) \} \leq C \sqrt{\frac{V_{k^*}}{n}}.
$$

- Consequently,

$$
\mathbb{E} \left\{ \int |\hat{f}_{Kn} - f| \right\} \leq 4C \sqrt{\frac{V_{k^*}}{n}} + \frac{3}{2n} + 14 \exp \left( -\kappa n^{d\delta} \right).
$$

- In particular,

$$
\mathbb{E} \left\{ \int |\hat{f}_{Kn} - f| \right\} = O \left( \frac{1}{\sqrt{n}} \right).
$$
Link with the VC theory

- If $A_{k^*}$ has Vapnik-Chervonenkis dimension $V_{k^*}$, then

$$
\mathbb{E}\{\Delta_{k^*}(f)\} \leq C\sqrt{\frac{V_{k^*}}{n}}.
$$

- Consequently,

$$
\mathbb{E}\left\{ \int |\hat{f}_{Kn} - f| \right\} \leq 4C\sqrt{\frac{V_{k^*}}{n}} + \frac{3}{2n} + 14 \exp\left(-\kappa n^{d\delta}\right).
$$

- In particular,

$$
\mathbb{E}\left\{ \int |\hat{f}_{Kn} - f| \right\} = O\left(\frac{1}{\sqrt{n}}\right).
$$
Link with the VC theory

- If \( \mathcal{A}_{k^*} \) has Vapnik-Chervonenkis dimension \( V_{k^*} \), then

\[
\mathbb{E}\{ \Delta_{k^*}(f) \} \leq C \sqrt{\frac{V_{k^*}}{n}}.
\]

- Consequently,

\[
\mathbb{E}\left\{ \int |\hat{f}_{Kn} - f| \right\} \leq 4C \sqrt{\frac{V_{k^*}}{n}} + \frac{3}{2n} + 14 \exp \left( -\kappa n^{d*} \right).
\]

- In particular,

\[
\mathbb{E}\left\{ \int |\hat{f}_{Kn} - f| \right\} = O \left( \frac{1}{\sqrt{n}} \right).
\]
Examples

→ \( V_k = O(k^4) \) for the univariate Gaussian mixtures.

→ \( V_k \leq k + 1 \) for the exponential families.

Case \( f \notin \mathcal{F} \)? It seems that we can have an error bound of the order of the bound for \( T_n \).
Examples

→ $V_k = O(k^4)$ for the univariate Gaussian mixtures.

→ $V_k \leq k + 1$ for the exponential families.

Case $f \notin \mathcal{F}$? It seems that we can have an error bound of the order of the bound for $T_n$. 
Examples

$V_k = O(k^4)$ for the univariate Gaussian mixtures.

$V_k \leq k + 1$ for the exponential families.

Case $f \notin \mathcal{F}$? It seems that we can have an error bound of the order of the bound for $T_n$. 
On the closure condition

For any set of parameters $\Theta \subset \mathbb{R}^s$ and any density $\psi(., \theta)$ defined on $\mathbb{R}^d \times \Theta$, let the collection $\mathcal{C}_\psi$ be

$$\mathcal{C}_\psi = \{ \psi(., \theta) : \theta \in \Theta \}.$$

**Proposition**

Assume that

(i) For all $t \in \mathbb{R}^d$, $\hat{\psi}(t,.)$ is continuous on $\Theta$.

(ii) For all $\theta_0 \in \bar{\Theta} \setminus \Theta$ and any sequence $(\theta_n)$ in $\Theta$ with $\theta_n \to \theta_0$, one has

$$\limsup_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{\mathcal{D}} \quad \text{or} \quad \liminf_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{\mathcal{D}}.$$

Then $\hat{\mathcal{C}}_\psi$ is closed in $\hat{\mathcal{D}}$. 

G. Biau (Université Paris VI)
On the closure condition

For any set of parameters $\Theta \subset \mathbb{R}^s$ and any density $\psi(., \theta)$ defined on $\mathbb{R}^d \times \Theta$, let the collection $C_\psi$ be

$$C_\psi = \{ \psi(., \theta) : \theta \in \Theta \}.$$

**Proposition**

Assume that

(i) For all $t \in \mathbb{R}^d$, $\hat{\psi}(t, .)$ is continuous on $\Theta$.

(ii) For all $\theta_0 \in \bar{\Theta} \setminus \Theta$ and any sequence $(\theta_n)$ in $\Theta$ with $\theta_n \to \theta_0$, one has

$$\limsup_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D} \quad \text{or} \quad \liminf_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D}.$$

Then $\hat{C}_\psi$ is closed in $\hat{D}$. 
The mixture case

Let $\psi(., \theta)$ be a density defined on $\mathbb{R}^d \times \Theta$.

**Proposition**

Suppose that $\hat{C}_\psi$ is closed in $\hat{D}$, and consider the $k$-th mixture class associated with $\psi$ defined by

\[ F_k = \left\{ \sum_{i=1}^{k} p_i \psi(., \theta_i) : p_i \geq 0, \sum_{i=1}^{k} p_i = 1, \theta_i \in \Theta \right\}. \]

Then $\hat{F}_k$ is closed in $\hat{D}$.

True if $t \to \hat{\psi}(t,.)$ is continuous on $\Theta$ and for all $\theta_0 \in \Theta \setminus \Theta$ and any sequence $(\theta_n)$ in $\Theta$ with $\theta_n \to \theta_0$, one has

\[ \limsup_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D} \quad \text{or} \quad \liminf_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D}. \]
Let \( \psi(., \theta) \) be a density defined on \( \mathbb{R}^d \times \Theta \).

**Proposition**

Suppose that \( \hat{C}_\psi \) is closed in \( \hat{D} \), and consider the \( k \)-th mixture class associated with \( \psi \) defined by

\[
F_k = \left\{ \sum_{i=1}^{k} p_i \psi(., \theta_i) : p_i \geq 0, \sum_{i=1}^{k} p_i = 1, \theta_i \in \Theta \right\}.
\]

Then \( \hat{F}_k \) is closed in \( \hat{D} \).

**True** if \( t \to \hat{\psi}(t,.) \) is continuous on \( \Theta \) and for all \( \theta_0 \in \bar{\Theta} \setminus \Theta \) and any sequence \( (\theta_n) \) in \( \Theta \) with \( \theta_n \to \theta_0 \), one has

\[
\lim \sup_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D} \quad \text{or} \quad \lim \inf_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D}.
\]
Let \( \psi(., \theta) \) be a density defined on \( \mathbb{R}^d \times \Theta \).

**Proposition**

Suppose that \( \hat{C}_\psi \) is closed in \( \hat{D} \), and consider the \( k \)-th mixture class associated with \( \psi \) defined by

\[
F_k = \left\{ \sum_{i=1}^{k} p_i \psi(., \theta_i) : p_i \geq 0, \sum_{i=1}^{k} p_i = 1, \theta_i \in \Theta \right\}.
\]

Then \( \hat{F}_k \) is closed in \( \hat{D} \).

**True** if \( t \to \hat{\psi}(t, .) \) is continuous on \( \Theta \) and for all \( \theta_0 \in \bar{\Theta} \setminus \Theta \) and any sequence \( (\theta_n) \) in \( \Theta \) with \( \theta_n \to \theta_0 \), one has

\[
\limsup_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D} \quad \text{or} \quad \liminf_{n \to \infty} \hat{\psi}(., \theta_n) \notin \hat{D}.
\]