Asymptotic power of goodness of fit tests based on the Wasserstein distance

Hélène Boistard
joint work with Eustasio del Barrio
Universidad de Valladolid

23 de mayo de 2006

Asymptotic Decision Theory

(Le Cam, Hájek, Ibragimov, Hasminskii, Strasser, . . .)

Statistical Experiment: $T \neq \emptyset$ set of "parameters", $E \in \mathcal{E}(T)$ if $E = (\Omega, \mathcal{A}, \{P_t : t \in T\})$ P_t probability measures over $(\Omega, \mathcal{A})$.

Decision Problem: (T, D, W); D space of decisions (a topological space); $W = (W_t)_{t \in T}$ loss function, $W_t : D \to \mathbb{R}$.

Decision function: procedure for making a (possibly randomized) decision based on the outcome of the experiment:

$$\rho : \Omega \times \mathcal{B}_0(D) \to [0, 1],$$

$\mathcal{B}_0(D)$ Baire σ-field; ρ a Markov kernel ($\rho(\cdot, A)$ measurable for fixed A, $\rho(\omega, \cdot)$ a probability for fixed ω)

$\mathcal{R}(E, D) = \{$decision functions for E and $D\}$

Generalized decision functions

$\rho \in \mathcal{R}(E, D)$ induces a bilinear form $\beta_\rho : \mathcal{C}_b(D) \times L(E) \to \mathbb{R}$

$$\beta_\rho(f, \mu) = \int\int_{D\times\Omega} f(x)\rho(\omega, dx)\mu(d\omega) \quad \text{satisfying}$$

(1) $|\beta_\rho(f, \mu)| \leq \|f\|\|\mu\|$; **(2)** $\beta_\rho(f, \mu) \geq 0$ si $f \geq 0, \mu \geq 0$;
(3) $\beta_\rho(1, \mu) = \mu(\Omega)$

$(L(E) = \left(\{P_t : t \in T\}^\perp\right)^\perp$; $L(E) \simeq L_1(\Omega, \mathcal{A}, \nu)$ if $P_t \ll \nu)$

$\mathcal{B}(E, D) = \{$bilinear forms on satisfying **(1)**,**(2)**,**(3)**$\}$

$\mathcal{R}(E, D) \subset \mathcal{B}(E, D)$

$\mathcal{B}(E, D)$ compact (weak topology) and convex; $\mathcal{R}(E, D)$ dense in $\mathcal{B}(E, D)$.

Risk functions

Assume $W = (W_t)_{t \in T}$ continuous loss function, $\rho \in \mathcal{R}(E, D)$;

$\int_D W_t(x)\rho(\omega, dx)$ (average) loss for the decision $\rho(\omega, \cdot)$

$$\text{risk function } : t \mapsto \beta_\rho(W_t, P_t) = \int\int_{D \times \Omega} W_t(x)\rho(\omega, dx)P_t(d\omega)$$

(expected loss when *true* parameter is t).

$\beta_\rho(W_t, P_t)$ generalizes easily to lower semicontinuous W

Comparison of experiments

Consider E, $F \in \mathcal{E}(T)$; (T, D, W) a decision problem

$E \underset{(D,W)}{\supseteq} F$ (E is more informative than F) $\Leftrightarrow \forall \beta_2 \in \mathcal{B}(F, D)\, \exists \beta_1 \in \mathcal{B}(E, D)$

$$\text{such that} \qquad \beta_1(W_t, P_t) \leq \beta_2(W_t, Q_t), \quad t \in T$$

$E \supseteq F \Leftrightarrow E \underset{(D,W)}{\supseteq} F \; \forall \, (D, W)$; if $E \supseteq F$ and $F \supseteq E$ then $E \sim F$.

Theorem $E = (\Omega_1, \mathcal{A}_1, \{P_t : t \in T\})$, $F = (\Omega_2, \mathcal{A}_2, \{Q_t : t \in T\})$ are equivalent iff

$$\mathcal{L}\left(\left(\frac{dP_t}{dP_s}\right)_{t \in T} \middle| P_s\right) = \mathcal{L}\left(\left(\frac{dQ_t}{dQ_s}\right)_{t \in T} \middle| Q_s\right)$$

Topological spaces of experiments. Deficiency.

$E \overset{\varepsilon}{\supseteq} F$ if $\forall\,(D, W)$ and $\beta_2 \in \mathcal{B}(F, D)\,\exists \beta_1 \in \mathcal{B}(E, D)$

$$\text{such that} \qquad \beta_1(W_t, P_t) \leq \beta_2(W_t, Q_t) + \varepsilon \|W_t\|, \quad t \in T$$

For $E, F \in \mathcal{E}(T)$

$$\delta(E, F) = \inf\{\varepsilon > 0 : \; E \overset{\varepsilon}{\supseteq} F\}; \quad \Delta(E, F) = \max\{\delta(E, F), \delta(F, E)\}$$

Δ is a distance on $\mathcal{E}(T)/\sim$; $(\mathcal{E}(T)/\sim, \Delta)$ (strong topology) is complete

Weak topology generated by $\Delta(E_\alpha, F_\alpha)$, $\alpha \subset T$ finite; makes $\mathcal{E}(T)/\sim$ compact

Weak convergence of experiments & decision functions

$T_n \uparrow T$; $E_n \in \mathcal{E}(T_n)$, $E \in \mathcal{E}(T)$

$$E_n \to E \text{ weakly iff } \Delta(E_{n,\alpha}, E_\alpha) \to 0, \quad \alpha \text{ finite}$$

Theorem $E_n \to E$ weakly iff $\mathcal{L}\left(\left(\frac{dP_{n,t}}{dP_{n,s}}\right)_{t\in T} \middle| P_{n,s}\right) \to_w \mathcal{L}\left(\left(\frac{dP_t}{dP_s}\right)_{t\in T} \middle| P_s\right)$

If $E_n \to E$ weakly, $\beta_n \in \mathcal{B}(E_n, D)$ and $\beta \in \mathcal{B}(E, D)$ then

$$\beta_n \to \beta \text{ in distribution iff } \beta_n(f, P_{n,t}) \to \beta(f, P_t), \quad f \in \mathcal{C}_b(D)$$

For testing problems this means pointwise convergence of power functions

Theorem (LeCam) Every sequence β_n has accumulation points

Any sequence of testing procedures can be judged by the properties of the limiting testing procedures that it defines for the limiting experiment.

Testing in the framework of Gaussian shift experiments

$(H, \langle , \rangle)$ a real Hilbert space. An experiment $(\Omega, \mathcal{A}, \{P_h, h \in H\})$ on H is a Gaussian shift experiment, if and only if
1) for all h, $P_h \ll P_0$ and
2) the process $(L(h))_{h \in H}$ defined by the log-likelihood ratio

$$\log \frac{dP_h}{dP_0} = L(h) - \frac{\|h\|^2}{2},$$

is a standard Gaussian process under P_0, i.e.: it is centered and for any $h_1, h_2 \in H$

$$Cov(L(h_1), L(h_2)) = \langle h_1, h_2, \rangle.$$

Spectral decomposition

The null hypothesis: $H_0 \subset H$ a linear subspace of H (e.g. $H_0 = \{0\}$).
$\varphi : \Omega \to [0,1]$ a α H_0-similar test function for H_0.

Taylor expansion of the power of φ under a straight line of alternatives directed by some $h \in H \backslash \{0\}$, near 0:

$$E_{th}\varphi = \alpha + b(h)t + a(h)\frac{t^2}{2} + o(t^2), t \to 0. \tag{1}$$

Theorem 1) $\exists h_0 \in H$ gradient such that:

$$b(h) = \langle h, h_0 \rangle, \forall h \in H.$$

2) $\exists T : H \to H$ a self-adjoint Hilbert-Schmidt operator, an ortonormal system (h_i) and eigenvectors (λ_i) such that:

$$a(h) = \langle h, T(h) \rangle \forall h \in H, \text{ and } T = \sum_{i=1}^{+\infty} \lambda_i \langle ., h_i \rangle.$$

Theorem: Properties based on the generalized Neyman Pearson lemma.
1.$\|h_0\| \le \phi(\Phi^{-1}(1-\alpha))$.
The equality holds iff

$$\varphi = \mathbf{1}_{\{L(h_0) > \Phi^{-1}(1-\alpha)\|h_0\|\}}.$$

2.The largest eigenvalue λ_1 of T satisfies the inequality

$$|\lambda_{\mathbf{1}}| \le 2\phi\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)\Phi^{-1}\left(1-\frac{\alpha}{2}\right).$$

The equality holds iff

$$\varphi = \mathbf{1}_{\{|L(h_1)| > \Phi^{-1}(1-\frac{\alpha}{2})\}}.$$

Asymptotic relative local efficiency

One-sided tests: φ a test for the Gaussian shift, with $E_{P_0}\varphi = \alpha$.

Alternatives: $K \setminus \{0\}$, where K is a cone.

$$ARE_L^{(1)}(\varphi, h) = \left(\frac{\langle h, h_0 \rangle}{\|h\| \phi\left(\Phi^{-1}\left(1-\alpha\right)\right)} \right)^2.$$

Two-sided tests: φ an unbiased test at level α for $H_0 = \{0\}$ against $H \setminus \{0\}$, i.e.:

$$E_0\varphi = \alpha \text{ and } E_h\phi \geq \alpha \text{ for } h \in H \setminus \{0\}.$$

Then $\lambda_i \geq 0 \; \forall i \geq n$ and $h_0 = 0$.

$h \mapsto a(h) = \langle h, T(h) \rangle$ is the curvature of the power function at 0 in direction h.

$$ARE_L^{(2)}(\varphi, h) = \frac{\langle h, T(h) \rangle}{2\|h\|^2 \phi\left(\Phi^{-1}\left(1-\alpha/2\right)\right) \Phi^{-1}\left(1-\alpha/2\right)}.$$

Remind that $\frac{\langle h, T(h) \rangle}{\|h\|^2} \leq \lambda_1$.

Power properties expected for a test in a Gaussian shift experiment

Global power function (Janssen) φ a level α test for the null hypothesis $H_0 = \{0\}$ of the Gaussian shift. It has non-vanishing power only on a finite-dimensional subspace of alternatives. Namely: $\forall \epsilon > 0$, $K > 0$, there exists a linear subspace $V \subset H$ of finite dimension such that

$$\sup\{|E_h \varphi - \alpha| : h \in V^\perp, \|h\| \leq K\} \leq \epsilon,$$

the dimension of V being bounded by $1 + \epsilon^{-1}\alpha(1-\alpha)(\exp(K^2) - 1)$.

Small levels (Rahnenführer) When $\alpha \to 0$, the largest eigenvalue $\lambda_1 = \lambda_{1,\alpha}$ converges to its maximal possible value $m_\alpha = 2\phi\left(\Phi^{-1}(1-\alpha/2)\right)\Phi^{-1}(1-\alpha/2)$, while all the other eigenvalues vanish:

$$\lim_{\alpha \to 0} \frac{\lambda_{1,\alpha}}{m_\alpha} \to 1, \quad \lim_{\alpha \to 0} \frac{\lambda_{i,\alpha}}{m_\alpha} \to 0, \quad \text{for } i \geq 1.$$

Example: Kolmogorov-Smirnov tests for the signal detection problem with Brownian bridge as noise term

B_0 a Brownian bridge, $L_2^0(0,1) = \{h \in L_2(0,1) : \int_0^1 h(u)du = 0\}$, consider the regression problem

$$B_h(t) = \int_0^t h(s)ds + B_0(t),$$

where $h \in L_2^0(0,1)$. P_0 the law of B_0 on the space $\mathcal{C}(0,1)$, P_h the law of B_h.

By Girsanov's theorem:

$$\log\left(\frac{dP_h}{dP_0}\right) = \int_0^1 h(s)dB_0(s) - \frac{\|h\|^2}{2}.$$

One-sided Kolmogorov-Smirnov test for $H_0 = \{0\}$

$$\begin{aligned} \varphi_\alpha : \mathcal{C}(0,1) &\longrightarrow [0,1] \\ x \mapsto \varphi_\alpha(x) &= \mathbf{1}_{\{\sup_{0\le t\le 1} x(t) > c_{1-\alpha}\}}, \end{aligned}$$

with $c_{1-\alpha}$ such that $E_0\varphi_\alpha(B_0) = \alpha$.

Explicit calculus of the gradient $h_{0,\alpha}$ of φ_α by Hájek and Śidák.

Result: $\frac{h_{0,\alpha}}{\phi(\Phi^{-1}(1-\alpha))} \longrightarrow h_0$ in $L_2(0,1)$ when $\alpha \longrightarrow 0$, where

$$h_0(u) = sign(2u-1).$$

Consequence:

$$ARE_L^{(1)}(\varphi_\alpha, h_0) \longrightarrow 1, \alpha \longrightarrow 0$$

and behavior of φ_α similar to that of the median test

$$\varphi_0 = \mathbf{1}_{\left\{2x(1/2) > \Phi^{-1}(1-\alpha)\right\}}$$

Two-sided Kolmogorov-Smirnov test for $H_0 = \{0\}$

$$\begin{aligned} \psi_\alpha : \mathcal{C}(0,1) &\longrightarrow [0,1] \\ x \mapsto \psi_\alpha(x) &= \mathbf{1}_{\{\sup_{0\le t\le 1} |x(t)|>k_{1-\alpha}\}}, \end{aligned}$$

$\lambda_{1,\alpha}$, (resp. $\lambda_{i,\alpha}$) the largest (resp. the ith largest) eigenvalue of the spectral decomposition of the curvature of ψ_α.

Results: (Janssen, Rahnenführer)
1) For $h \in L_2^0(0,1)$,

$$\lim_{\alpha\downarrow 0} ARE_L^{(2)}(\Psi_\alpha, h) = \frac{\langle h, h_0\rangle^2}{\|h\|^2}.$$

2) As $\alpha \downarrow 0$,

$$\frac{\lambda_{1,\alpha}}{2\phi\left(\Phi^{-1}\left(1-\alpha/2\right)\right)} \to 1 \text{ and } \frac{\lambda_{i,\alpha}}{2\phi\left(\Phi^{-1}\left(1-\alpha/2\right)\right)} \to 0.$$

Wasserstein goodness of fit test

Wasserstein distance on $\mathcal{P}_2(\mathbb{R}) =: \{\text{Probability measures on } \mathbb{R} \text{ with finite second moment}\}$

L_2 Wasserstein distance between $P_1, P_2 \in \mathcal{P}_2(\mathbb{R})$:

$$\mathcal{W}(P_1, P_2) = \inf \left\{ \left[E(X_1 - X_2)^2 \right]^{1/2} : \mathcal{L}(X_1) = P_1, \mathcal{L}(X_2) = P_2 \right\}.$$

Expression with the quantile functions F_1^{-1} and F_2^{-1}:

$$\mathcal{W}(P_1, P_2) = \left(\int_0^1 \left(F_1^{-1}(t) - F_2^{-1}(t) \right)^2 dt \right)^{1/2}.$$

Empirical version to test fit to a distribution P_F with quantile function F^{-1}: P_n the empirical measure associated to the sample $(X_1, \ldots, X_n)$, F_n^{-1} the empirical quantile function.

$$r_n := \mathcal{W}^2(P_n, P_F) = \int_0^1 \left(F_n^{-1}(t) - F^{-1}(t) \right)^2 dt.$$

Distance to a location-scale family

A location-scale family is obtained from a law with mean 0, variance 1 and distribution function F as follows:

$$\mathcal{H}_F = \left\{ H : H(x) = F\left(\frac{x-\mu}{\sigma}\right), \mu \in \mathbb{R}, \sigma > 0 \right\}.$$

If $P \in \mathcal{P}_2(\mathbb{R})$ with distribution function F_0, standard deviation σ_0:

$$\begin{aligned} \mathcal{W}^2(P, \mathcal{H}_F) &= \inf \left\{ \mathcal{W}^2(P, H) : H \in \mathcal{H}_F \right\} \\ &= \sigma_0^2 - \left(\int_0^1 F_0^{-1}(t) F^{-1}(t) dt \right)^2. \end{aligned}$$

Normalized empirical version to test fit to the family $\mathcal{H}_F$:

$$\mathcal{R}_n := \frac{\mathcal{W}^2(F_n, \mathcal{H}_F)}{S_n^2} = 1 - \frac{\left(\int_0^1 F_n^{-1}(t) F^{-1}(t) dt \right)^2}{S_n^2},$$

where S_n^2 is the empirical variance.

Fit to a specified distribution, asymptotic distribution under null hypothesis

H_0: $(X_1, \ldots, X_n)$ i.i.d. $\sim P_F$, with distribution function F, density function f, quantile function F^{-1}. F_n^{-1} denotes the empirical quantile function.

Hypothesis on P_F: regularity conditions and the condition on extremes:

$$\int_0^1 \int_0^1 \frac{(s \wedge t - st)^2}{f^2(F^{-1}(s)) f^2(F^{-1}(t))} ds dt < +\infty.$$

$$nr_n - \int_{1/n}^{1-1/n} \frac{t(1-t)}{f^2(F^{-1}(t))} dt \to \int_0^1 \frac{B^2(t) - t(1-t)}{f^2(F^{-1}(t))} dt.$$

Key: $nr_n = \int_0^1 \left(\frac{\rho_n(t)}{f(F^{-1}(t))} \right)^2 dt$, where

$$\rho_n(t) = \sqrt{n} f\left(F^{-1}(t)\right) \left(F_n^{-1}(t) - F^{-1}(t)\right)$$

is the empirical quantile process.

Asymptotic distribution under a type of local alternatives

Alternative H_n: $(X_1, \ldots, X_n)$ i.i.d. with quantile function Φ_n^{-1} such that for some $h \in L_2(0,1)$,

$$\int_0^1 \left(\sqrt{n}\left(\Phi_n^{-1} - F^{-1}\right) - h\right)^p \to 0, n \to \infty, \text{ for some } p > 2.$$

$$nr_n - \int_{1/n}^{1-1/n} \frac{t(1-t)}{f^2(F^{-1}(t))} dt \to \int_0^1 \left[\left(\frac{B(t)}{f(F^{-1}(t))} + h(t)\right)^2 - \frac{t(1-t)}{f^2(F^{-1}(t))}\right] dt.$$

Representation of the limiting distribution in a unified way

Denote by K the covariance kernel of $B(t)/f\left(F^{-1}(t)\right)$.

$$K(s,t) = \frac{s \wedge t - st}{f(F^{-1}(s))f(F^{-1}(t))}, \quad (s,t) \in (0,1) \times (0,1).$$

By Mercer's theorem: there exist eigenvalues $\lambda_j, j \geq 0$ and normalized eigenfunctions $f_j, j \geq 0$ such that

$$K(s,t) = \sum_{j=0}^{+\infty} \lambda_j f_j(s) f_j(t).$$

For each j, define g_j such that: $f_j(x) = \frac{1}{\lambda_j} \frac{\int_0^x g_j(s)ds}{f(F^{-1}(x))}$.

Define the new kernel

$$\tilde{K}(s,t) = \sum_{j=0}^{+\infty} \lambda_j g_j(s) g_j(t) = C(F) + \int_{F(0)}^{s\wedge t} \frac{2x-1}{f^2(F^{-1}(x))dx} - \int_{s\wedge t}^{s \vee t} \frac{1-x}{f^2(F^{-1}(x))}.$$

$$\int_0^1 \frac{B^2(t) - t(1-t)}{f^2(F^{-1}(t))} dt = \int_0^1 \int_0^1 \tilde{K}(s,t) dW(s) dW(t),$$

where W is the Brownian motion related to B by: $B(t) = W(t) - tW(1)$.

When h (parameter of the local alternative) is

$$h(t) = \frac{\int_0^t g(s)ds}{f(F^{-1}(t))}, \text{ for some } g \in L_2^0(0,1),$$

$$\int_0^1 \left[\left(\frac{B(t)}{f(F^{-1}(t))} + h(t) \right)^2 - \frac{t(1-t)}{f^2(F^{-1}(t))} \right] dt = \int_0^1 \int_0^1 \tilde{K}(s,t) dW_g(s) dW_g(t), \text{ where}$$

$$W_g : t \rightarrow W(t) + \int_0^t g(s)ds.$$

Wasserstein goodness of fit test

Parametrization of the probability measures which are absolutely continuous with respect to a given probability

$(\Omega, \mathcal{A}, P_0)$ a probability space.

$$H = \{g \in L_2(P_0) : \int g dP_0 = P_0 g = 0\}, M = \{g \in H : P_0 g^2 \leq 4\}.$$

For $g \in M$, define $P_g \ll P_0$ the probability measure given by

$$\frac{dP_g}{dP_0} = \left(\frac{1}{2} g + \sqrt{1 - \frac{1}{4} P_0 g^2} \right)^2 .$$

$$\{P_g : g \in M\} = \{ \text{ probability measures } P \text{ on } (\Omega, \mathcal{A}) \text{ such that } P \ll P_0\}.$$

Gaussian shift experiment

A way to parametrize a local alternative:

$H_n = \{g \in H : \frac{g}{\sqrt{n}} \in M\}$. $H_n \uparrow H$.

An i.i.d. sample $(X_1, \ldots, X_n)$ with common distribution $P_{g/\sqrt{n}}$ corresponds to the experiment

$$E_n = \left(\Omega^n, \mathcal{A}^n, \left\{P^n_{g/\sqrt{n}} : g \in H_n\right\}\right).$$

E_n converges as an experiment to a Gaussian shift, since

$$\log \frac{dP^n_{g/\sqrt{n}}}{dP^n_0}(\omega) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\omega_i) - \frac{\|g\|^2}{2} + o_{P_0}(1).$$

Modified test

We define the empirical process for $t \in (0,1)$ by:

$$\alpha_n(t) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{F(X_i)\leq t\}} - t\right).$$

We introduce the two tests: (with critical value C_α given by the Wasserstein critical value)

$$\varphi_n^1 = \mathbf{1}_{\int\int_{s\neq t} \tilde{K}(s,t)d\rho_n(s)d\rho_n(t) > C_\alpha},$$

$$\varphi_n^2 = \mathbf{1}_{\int\int_{s\neq t} \tilde{K}(s,t)d\alpha_n(s)d\alpha_n(t) > C_\alpha}.$$

In the Gaussian shift experiment, there is convergence of the decision functions in distribution iff there is convergence in distribution of the test statistics under the null and the (local) alternatives.

Study of the last test for which the convergence is conjectured.

Case of a location-scale family

$nS_n^2\mathcal{R}_n$ has correction terms due to the estimation of the parameters in the location scale family.

$$
\begin{aligned}
nS_n^2\mathcal{R}_n = \quad & n\int_0^1 \left(F_n^{-1}(t) - F^{-1}(t)\right)^2 dt \\
& -n\left(\int_0^1 \left(F_n^{-1}(t) - F^{-1}(t)\right) dt\right)^2 \\
& -n\left(\int_0^1 \left(F_n^{-1}(t) - F^{-1}(t)\right) F^{-1}(t)dt\right)^2
\end{aligned}
$$

Same treatment, with a new kernel $\tilde{\tilde{K}}$.

Example of the standard normal distribution

$$K(s,t) = \sum_{j=0}^{+\infty} \lambda_j f_j(s) f_j(t), \text{ where}$$

$f_j(s) = h_j\left(\Phi^{-1}(s)\right)$, h_j denotes the normalized Hermite polynomial with degree j ($Eh_j^2(X) = 1$ for X r.v. with standard normal distribution), $\lambda_j = 1/(j+1)$.

$g_j = f_{j+1}$ and $\tilde{K}(s,t) = \sum_{j=0}^{+\infty} \lambda_j f_{j+1}(s) f_{j+1}(t)$.

Calculus of the gradient and curvature of the asymptotic test for the direction g at level $\alpha \in (0,1)$:

$$h_{0,\alpha} = 0, T_\alpha(g) = \sum_{i=1}^{+\infty} \mu_{\alpha,i} \langle f_i, g \rangle, \text{ with an known expression for } \mu_{\alpha,i}\ .$$

Extension: focused tests for a specified set of alternatives

$h_1, \ldots, h_r$ some fixed directions, $V = \langle h_1, \ldots, h_r \rangle$ linear subspace of interesting alternatives.

Construct $K_V = \sum_{1=1}^{+\infty} \lambda_i h_i$ ($h_{r+1}, \ldots,$ basis for $V^{\perp}$).

The test $\int \int K_V(s,t) d\alpha_n(s) d\alpha_n(t)$ has curvature operator $\sum \mu_i \langle ., h_j \rangle$.

Choose eigenvalues $\lambda_1, \ldots, \lambda_r, \lambda_{r+1}, \ldots$ such that $\mu_1, \ldots, \mu_r$ are big enough.