Data-processing inequality with expectations of random variables

(C'est le nom du lemme-clé de cet exposé)

Gilles Stoltz

+ Aurélien Garivier, Sébastien Gerchinovitz, Pierre Ménard (UPS Toulouse)

Data-processing inequality with expectations of random variables

Data-processing inequality $\qquad \mathrm{KL}(\mathbb{P}^X, \mathbb{Q}^X) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$

X is any random variable $\Omega \to E$

Our version $\qquad \mathrm{KL}\Big(\mathrm{Ber}\big(\mathbb{E}_{\mathbb{P}}[X]\big), \mathrm{Ber}\big(\mathbb{E}_{\mathbb{Q}}[X]\big)\Big) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$

for $X : \Omega \to [0, 1]$

KL is the Kullback-Leibler divergence between two distributions,

$$\mathrm{KL}(\mathbb{P}, \mathbb{Q}) = \begin{cases} +\infty & \text{if } \mathbb{P} \not\ll \mathbb{Q} \\ \displaystyle\int_\Omega \left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \ln \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \right) \mathrm{d}\mathbb{Q} & \text{if } \mathbb{P} \ll \mathbb{Q} \end{cases}$$

Proof of $\mathrm{KL}(\mathbb{P}^X, \mathbb{Q}^X) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$ — part 1/2

Proof: We may assume that $\mathbb{P} \ll \mathbb{Q}$, otherwise $\mathrm{KL}(\mathbb{P},\mathbb{Q}) = +\infty$ and the inequality is true. We shows that we then have

$$\mathbb{P}^X \ll \mathbb{Q}^X, \quad \text{with} \quad \frac{d\mathbb{P}^X}{d\mathbb{Q}^X} = \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\middle|\, X = \cdot \right] \overset{\text{not.}}{=} \gamma$$

i.e., $\gamma(x) = \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\middle|\, X \right]$.

Indeed, for all $B \in \mathcal{F}'$:

$$\mathbb{P}^X(B) = \mathbb{P}\{X \in B\} = \int_\Omega \mathbb{1}_B(X) \frac{d\mathbb{P}}{d\mathbb{Q}}\, d\mathbb{Q} \overset{\text{tower rule}}{=} \int_\Omega \mathbb{1}_B(X)\, \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\middle|\, X\right] d\mathbb{Q}$$

$$\overset{\text{not.}}{=} \int_\Omega \mathbb{1}_B(X)\, \gamma(X)\, d\mathbb{Q} \underset{\text{by definition of } \mathbb{Q}^X}{=} \int_{\Omega'} \mathbb{1}_B\, \gamma\, d\mathbb{Q}^X.$$

Proof of $\mathrm{KL}\big(\mathbb{P}^X, \mathbb{Q}^X\big) \leqslant \mathrm{KL}(\mathbb{P}, \mathbb{Q})$ — part 2/2

Therefore,

$$\mathrm{KL}(\mathbb{P}^X, \mathbb{Q}^X) = \int_{\Omega'} \gamma \ln \gamma \, d\mathbb{Q}^X = \int_{\Omega} \gamma(x) \ln \gamma(x) \, d\mathbb{Q}$$

definition of γ

$$= \int_{\Omega} \left(\mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X \right] \ln \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X \right] \right) d\mathbb{Q}$$

conditional version of Jensen's inequality

$$\leqslant \int_{\Omega} \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} \,\Big|\, X \right] d\mathbb{Q}$$

tower rule

$$= \int_{\Omega} \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} \right) d\mathbb{Q} = \mathrm{KL}(\mathbb{P}, \mathbb{Q})$$

Reference: Ali and Silvey [1966]; implies joint convexity of KL

Proof of $\mathrm{KL}\Big(\mathrm{Ber}\big(\mathbb{E}_{\mathbb{P}}[X]\big),\, \mathrm{Ber}\big(\mathbb{E}_{\mathbb{Q}}[X]\big)\Big) \leqslant \mathrm{KL}(\mathbb{P},\mathbb{Q})$

$$\mathrm{KL}\Big(\underbrace{(\mathbb{P}\otimes m)^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{P}\otimes m(E))},\ \underbrace{(\mathbb{Q}\otimes m)^{\mathbb{1}_E}}_{\mathrm{Ber}(\mathbb{Q}\otimes m(E))} \Big) \leqslant \mathrm{KL}(\mathbb{P}\otimes m,\ \mathbb{Q}\otimes m) \underset{\text{cf. product distributions}}{=} \mathrm{KL}(\mathbb{P},\mathbb{Q}) + \mathrm{KL}(m,m) = \mathrm{KL}(\mathbb{P},\mathbb{Q})$$

Thus: $\quad \mathrm{KL}\big(\mathrm{Ber}(\mathbb{P}\otimes m(E)),\ \mathrm{Ber}(\mathbb{Q}\otimes m(E))\big) \leqslant \mathrm{KL}(\mathbb{P},\mathbb{Q})$

The proof is concluded by picking $E \in \mathcal{F}\otimes\mathcal{B}([0,1])$ such that

$$\mathbb{P}\otimes m(E) = \mathbb{E}_{\mathbb{P}}[X] \quad \text{and} \quad \mathbb{Q}\otimes m(E) = \mathbb{E}_{\mathbb{Q}}[X]$$

Namely, $\quad E = \big\{ (\omega, x) \in \Omega\times[0,1] : \ x \leqslant X(\omega) \big\}$

By Tonelli's theorem:

$$\mathbb{P}\otimes m(E) = \int_\Omega \left(\int_{[0,1]} \mathbb{1}_{\{x \leqslant X(\omega)\}} \, dm(x) \right) d\mathbb{P}(\omega) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \mathbb{E}_{\mathbb{P}}[X]$$

and a similar equality for $\mathbb{Q}\otimes m(E)$.

First application: K–armed bandits

K probability distributions $\nu_1, \ldots, \nu_K$
with expectations $\mu_1, \ldots, \mu_K$ $\longrightarrow$ $\mu^\star = \max_{k=1,\ldots,K} \mu_k$

At each round $t = 1, 2, \ldots,$
1. Statistician picks arm $I_t \in \{1, \ldots, K\}$, possibly using U_{t-1}
2. She gets a reward Y_t with law ν_{I_t} given I_t
3. This is the only feedback she receives

$\longrightarrow$ Exploration–exploitation dilemma
estimate the ν_k vs. get high rewards Y_t

Regret:

$$R_T = \sum_{t=1}^{T} \bigl(\mu^\star - \mathbb{E}[Y_t]\bigr) = \sum_{k=1}^{K} \left((\mu^\star - \mu_k)\, \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}_{\{I_t = k\}} \right] \right)$$

Indeed, $Y_t \,|\, I_t \sim \nu_{I_t}$, thus $\mathbb{E}\bigl[Y_t \,|\, I_t\bigr] = \mu_{I_t}$

Summary:

At each round, pick I_t (based on U_{t-1} + past) and get $Y_t \mid I_t \sim \nu_{I_t}$

Control the regret $R_T = \sum_{k=1}^{K} (\mu^\star - \mu_k)\, \mathbb{E}\big[N_k(T)\big]$, where $N_k(T) = \sum_{t=1}^{T} \mathbb{I}_{\{I_t = k\}}$

Lower bound $R_T \iff$ Lower bound $\mathbb{E}\big[N_k(T)\big]$ for $\mu_k < \mu^\star$

Randomized strategy $\psi = (\psi_t)_{t \geqslant 0}$: measurable functions

$$\psi_t : H_t = \big(U_0, Y_1, U_1, \ldots, Y_t, U_t\big) \longmapsto \psi_t(H_t) = I_{t+1}$$

Take $U_0, U_1, \ldots$ iid $\sim \mathcal{U}_{[0,1]}$ and denote by $\mathfrak{m}$ the Lebesgue measure

Transition kernel (conditional distributions):

$$\mathbb{P}\big(Y_{t+1} \in B,\ U_{t+1} \in B' \,\big|\, H_t\big) = \nu_{\psi_t(H_t)}(B)\, \mathfrak{m}(B')$$

Summary: history $H_t = \bigl(U_0, Y_1, U_1, \ldots, Y_t, U_t\bigr)$ and $I_{t+1} = \psi_t(H_t)$

Lower bound $\mathbb{E}\bigl[N_k(T)\bigr]$ for $\mu_k < \mu^\star$, where $N_k(T) = \sum \mathbb{I}_{\{I_t = k\}}$

Transition kernel: $\mathbb{P}\bigl(Y_{t+1} \in B,\ U_{t+1} \in B' \bigm| H_t\bigr) = \nu_{\psi_t(H_t)}(B)\,\mathfrak{m}(B')$

Change of measure: $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ vs. $\underline{\nu}' = (\nu'_1, \ldots, \nu'_K)$

Fundamental inequality: quantifies the impact of such a change

For all Z taking values in $[0,1]$ and $\sigma(H_T)$–measurable,

$$\sum_{k=1}^{K} \mathbb{E}_{\underline{\nu}}\bigl[N_k(T)\bigr]\,\mathrm{KL}(\nu_k, \nu'_k) = \mathrm{KL}\bigl(\mathbb{P}_{\underline{\nu}}^{H_T}, \mathbb{P}_{\underline{\nu}'}^{H_T}\bigr)$$

$$\geqslant \mathrm{KL}\Bigl(\mathrm{Ber}\bigl(\mathbb{E}_{\underline{\nu}}[Z]\bigr), \mathrm{Ber}\bigl(\mathbb{E}_{\underline{\nu}'}[Z]\bigr)\Bigr)$$

Inequality holds by the data-processing inequality with expectations of r.v.

Later use: $\underline{\nu}'$ only differ from $\underline{\nu}$ at k and $Z = N_k(T)/T$

Proof of the equality: chain rule for KL

$H_{t+1} = \bigl(H_t, (Y_{t+1}, U_{t+1})\bigr)$ and $\mathbb{P}\bigl(Y_{t+1} \in B,\ U_{t+1} \in B' \,\big|\, H_t\bigr) = \nu_{\psi_t(H_t)}(B)\,\mathfrak{m}(B')$

$$
\begin{aligned}
&\mathrm{KL}\Bigl(\mathbb{P}_{\underline{\nu}}^{H_{t+1}},\, \mathbb{P}_{\underline{\nu}'}^{H_{t+1}}\Bigr) \\
&= \mathrm{KL}\bigl(\mathbb{P}_{\underline{\nu}}^{H_t},\, \mathbb{P}_{\underline{\nu}'}^{H_t}\bigr) + \mathrm{KL}\Bigl(\mathbb{P}_{\underline{\nu}}^{(Y_{t+1},U_{t+1})\,|\,H_t},\, \mathbb{P}_{\underline{\nu}'}^{(Y_{t+1},U_{t+1})\,|\,H_t}\Bigr) \\
&= \mathrm{KL}\bigl(\mathbb{P}_{\underline{\nu}}^{H_t},\, \mathbb{P}_{\underline{\nu}'}^{H_t}\bigr) + \mathbb{E}_{\underline{\nu}}\Bigl[\mathbb{E}_{\underline{\nu}}\Bigl[\mathrm{KL}\bigl(\nu_{\psi_t(H_t)} \otimes \mathfrak{m},\, \nu'_{\psi_t(H_t)} \otimes \mathfrak{m}\bigr) \,\Big|\, H_t\Bigr]\Bigr] \\
&= \mathrm{KL}\bigl(\mathbb{P}_{\underline{\nu}}^{H_t},\, \mathbb{P}_{\underline{\nu}'}^{H_t}\bigr) + \mathbb{E}_{\underline{\nu}}\Bigl[\mathbb{E}_{\underline{\nu}}\Bigl[\mathrm{KL}\bigl(\nu_{\psi_t(H_t)},\, \nu'_{\psi_t(H_t)}\bigr) \,\Big|\, H_t\Bigr]\Bigr] \\
&= \mathrm{KL}\bigl(\mathbb{P}_{\underline{\nu}}^{H_t},\, \mathbb{P}_{\underline{\nu}'}^{H_t}\bigr) + \mathbb{E}_{\underline{\nu}}\Biggl[\sum_{k=1}^{K} \mathrm{KL}(\nu_k, \nu'_k)\, \mathbb{I}_{\{I_{t+1}=k\}}\Biggr]
\end{aligned}
$$

By induction: $$\mathrm{KL}\bigl(\mathbb{P}_{\underline{\nu}}^{H_T},\, \mathbb{P}_{\underline{\nu}'}^{H_T}\bigr) = \sum_{k=1}^{K} \mathbb{E}_{\underline{\nu}}\bigl[N_k(T)\bigr]\, \mathrm{KL}(\nu_k, \nu'_k)$$

References: already present in Auer, Cesa-Bianchi, Freund and Schapire [2002]

Fundamental inequality: For all $Z \in [0,1]$ and $\sigma(H_T)$–measurable,

$$\sum_{k=1}^{K} \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\,\mathrm{KL}(\nu_k, \nu'_k) \geqslant \mathrm{kl}\big(\mathbb{E}_{\underline{\nu}}[Z],\, \mathbb{E}_{\underline{\nu}'}[Z]\big)$$

How to use it?

Bandit problem $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ where k is suboptimal: $\mu_k < \mu^\star$

Pick $Z = N_k(T)/T$

Pick $\underline{\nu}'$ that only differs from $\underline{\nu}$ at k:

$$\underline{\nu}' = (\nu_1, \ldots, \nu_{k-1}, \nu'_k, \nu_{k+1}, \ldots, \nu_K)$$

Then

$$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\,\mathrm{KL}(\nu_k, \nu'_k) \geqslant \mathrm{KL}\Big(\mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}}[N_k(T)/T]\big),\, \mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}'}[N_k(T)/T]\big)\Big)$$

To lower bound $R_T = \sum_{k=1}^{K} (\mu^\star - \mu_k)\, \mathbb{E}\big[N_k(T)\big]$, lower bound $\mathbb{E}\big[N_k(T)\big]$ for $\mu_k < \mu^\star$

Bandit model $\mathcal{D}$: where the $\nu_1, \ldots, \nu_K$ may lie in

Assumption (UFC – uniform fast convergence on $\mathcal{D}$)

The strategy ψ is such that:
For all bandit problems $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ in $\mathcal{D}$, for all $\mu_k < \mu^\star$,

$$\forall\, \alpha \in (0, 1], \qquad \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] = o(T^\alpha)$$

Bandit problem $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ where k is suboptimal: $\mu_k < \mu^\star$

Pick $\nu'_k \in \mathcal{D}$ with expectation $\mu'_k > \mu^\star$

Form $\underline{\nu}'$ that only differs from $\underline{\nu}$ at k:

$$\underline{\nu}' = (\nu_1, \ldots, \nu_{k-1}, \nu'_k, \nu_{k+1}, \ldots, \nu_K)$$

Then $\quad \mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] = o(T) \quad$ and $\quad T - \mathbb{E}_{\underline{\nu}'}\big[N_k(T)\big] = o(T^\alpha)$

$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]\,\mathrm{KL}(\nu_k,\nu'_k) \geqslant \mathrm{KL}\Big(\mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}}[N_k(T)/T]\big),\,\mathrm{Ber}\big(\mathbb{E}_{\underline{\nu}'}[N_k(T)/T]\big)\Big)$

$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] = o(T)$ and $T - \mathbb{E}_{\underline{\nu}'}\big[N_k(T)\big] = o(T^\alpha)$ for a strategy ψ UFC on $\mathcal{D}$

Also, $\mathrm{KL}\big(\mathrm{Ber}(p),\,\mathrm{Ber}(q)\big) = p\ln\frac{p}{q} + (1-p)\ln\frac{1-p}{1-q} \geqslant (1-p)\ln\dfrac{1}{1-q} - \ln 2$

Combining all these elements:

$$
\begin{aligned}
&\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] \\
&\geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu'_k)}\,\mathrm{kl}\Big(\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big],\,\mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big]\Big) \\
&\geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu'_k)}\left(-\ln 2 + \Big(1-\mathbb{E}_{\underline{\nu}}\big[N_k(T)/T\big]\Big)\ln\frac{1}{1-\mathbb{E}_{\underline{\nu}'}\big[N_k(T)/T\big]}\right) \\
&\geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu'_k)}\left(-\ln 2 + \big(1-o(1)\big)\ln\frac{1}{T^{\alpha-1}}\right)
\end{aligned}
$$

Thus, $\forall\,\alpha\in(0,1]$, $\qquad \displaystyle\liminf_{T\to\infty}\frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathrm{KL}(\nu_k,\nu'_k)}\,\frac{\ln T^{1-\alpha}}{\ln T}$

That is, for all models $\mathcal{D}$ (for the first time, no assumption on $\mathcal{D}$)

for all strategies ψ UFC on $\mathcal{D}$ (this is not a real restriction)

for all bandit problems $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ in $\mathcal{D}$

for $\mu_k < \mu^\star$

Lemma

for all ν'_k in $\mathcal{D}$ with $\mu'_k > \mu^\star$,

$$\liminf_{T \to \infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathrm{KL}(\nu_k, \nu'_k)}$$

Theorem (see Lai and Robbins [1985], Burnetas and Katehakis [1996])

$$\liminf_{T \to \infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}$$

where $\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D}) = \inf\big\{\mathrm{KL}(\nu_k, \nu'_k) : \ \nu'_k \in \mathcal{D} \text{ with } \mu'_k > \mu^\star\big\}$

This distribution-dependent bound is asymptotically optimal:

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \geqslant \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}$$

I.e., at least for well-behaved models $\mathcal{D}$, we can exhibit a matching upper bound:

$$\limsup_{T\to\infty} \frac{\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big]}{\ln T} \leqslant \frac{1}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})}$$

See Lai and Robbins [1985], Burnetas and Katehakis [1996], Honda and Takemura [2010–2015], Cappé, Garivier, Maillard, Munos and Stoltz [2013], etc.

Replacing the $o(T^\alpha)$ in the definition of UFC by a $O(\ln T)$:

$$\mathbb{E}_{\underline{\nu}}\big[N_k(T)\big] \geqslant \frac{\ln T}{\mathcal{K}_{\inf}(\nu_k, \mu^\star, \mathcal{D})} - \Omega\big(\ln(\ln T)\big)$$

Cf. the upper bound of Honda and Takemura [2015]:
This second-order term $-\ln(\ln T)$ is optimal

Second application: Fano's inequality

For collections of distributions $\mathbb{P}_\theta$, $\mathbb{Q}_\theta$ and r.v. Z_θ:

$$\mathrm{KL}\left(\mathrm{Ber}\left(\int_\Theta \mathbb{E}_{\mathbb{P}_\theta}\big[Z_\theta\big]\,\mathrm{d}\nu(\theta)\right),\ \mathrm{Ber}\left(\int_\Theta \mathbb{E}_{\mathbb{Q}_\theta}\big[Z_\theta\big]\,\mathrm{d}\nu(\theta)\right)\right)$$
$$\leqslant \int_\Theta \mathrm{KL}\left(\mathrm{Ber}\left(\mathbb{E}_{\mathbb{P}_\theta}\big[Z_\theta\big]\right),\ \mathrm{Ber}\left(\mathbb{E}_{\mathbb{Q}_\theta}\big[Z_\theta\big]\right)\right)\mathrm{d}\nu(\theta)$$
$$\leqslant \int_\Theta \mathrm{KL}(\mathbb{P}_\theta, \mathbb{Q}_\theta)\,\mathrm{d}\nu(\theta)$$

Thus,

$$\int_\Theta \mathbb{E}_{\mathbb{P}_\theta}\big[Z_\theta\big]\,\mathrm{d}\nu(\theta) \leqslant \frac{\displaystyle\int_\Theta \mathrm{KL}(\mathbb{P}_\theta, \mathbb{P}_0)\,\mathrm{d}\nu(\theta) + \ln 2}{\displaystyle -\ln \int_\Theta \mathbb{E}_{\mathbb{P}_0}\big[Z_\theta\big]\,\mathrm{d}\nu(\theta)}$$

In particular,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{P}_i(A_i) \leqslant \frac{\displaystyle\frac{1}{N}\sum_{i=1}^{N} \mathrm{KL}(\mathbb{P}_i, \mathbb{P}_0) + \ln 2}{\displaystyle -\ln \frac{1}{N}\sum_{i=1}^{N} \mathbb{P}_0(A_i)}$$

We obtained, with no real assumptions:

$$\int_\Theta \mathbb{E}_{\mathbb{P}_\theta}\big[Z_\theta\big]\,\mathrm{d}\nu(\theta) \leqslant \frac{\displaystyle\int_\Theta \mathrm{KL}(\mathbb{P}_\theta,\mathbb{P}_0)\,\mathrm{d}\nu(\theta)}{\displaystyle -\ln \int_\Theta \mathbb{E}_{\mathbb{P}_0}\big[Z_\theta\big]\,\mathrm{d}\nu(\theta)}$$

To get lower bounds on the posterior concentration rate, we use

$$Z_\theta = \mathbb{P}_\pi\big(\theta' : \|\theta'-\theta\|_2 \leqslant \varepsilon_n \,\big|\, X_{1:n}\big)$$

E.g., when $\mathbb{P}_\theta = \mathcal{N}(\theta,\sigma^2)^{\otimes n}$:

$$\inf_{\theta\in\Theta} \mathbb{E}_\theta\Big[\mathbb{P}_\pi\big(\theta' : \|\theta'-\theta\|_2 \leqslant \varepsilon_n \,\big|\, X_{1:n}\big)\Big] \leqslant \frac{n\rho^2\varepsilon_n^2/(2\sigma^2)+\ln 2}{d\ln\rho} \leqslant c < 1$$

for ε_n of order $\sigma\sqrt{d/n}$

We took ν as the uniform distribution over $B(0,\rho\varepsilon_n)$