Information tools for analyzing tests

The complexities of batch to sequential settings

Aurélien Garivier

ENS Lyon

November 23rd, 2023

Batch Testing

Large deviation bounds

Lower bounds : TV and KL

Sequential Testing

A sequential lower bound

An "optimal" sequential test

$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{B}(p)$

$$H_- : p = \frac{1}{2} - \epsilon \qquad H_+ : p = \frac{1}{2} + \epsilon$$

Simple test on space $\mathcal{X} = \{0,1\}^n$ with sigma-field $\mathfrak{A} = \{A : A \subset \mathcal{X}\}$ of hypothesis $P_- = \mathcal{B}\left(\frac{1}{2} - \epsilon\right)^{\otimes n}$ versus $P_+ = \mathcal{B}\left(\frac{1}{2} + \epsilon\right)^{\otimes n}$ $X_i(\omega) = \omega_i$, $\mathbb{P}_p = \mathcal{B}(p)^{\otimes n}$

Test statistic $T = t(X_1, \dots, X_n)$ with range $\{0,1\}$

δ-correct: $P_-(T = 0) \geq 1 - \delta$ and $P_+(T = 1) \geq 1 - \delta$

Sample complexity: minimal value of n such that there exists a δ-correct test T

Idea 0: $T = \mathbb{1}\left\{\bar{X}_n > \frac{1}{2}\right\}$

δ-correct if $P_-(T=1) = \mathbb{P}_{\frac{1}{2}-\epsilon}\left(\bar{X}_n \geq \frac{1}{2}\right) \leq \delta$ and $P_+(T=0) = \mathbb{P}_{\frac{1}{2}+\epsilon}\left(\bar{X}_n < \frac{1}{2}\right) \leq \delta$

Chernoff's bound: if $0 < p \leq x < 1$, $\mathbb{P}_p\big(\bar{X}_n \geq x\big) \leq \exp\big(-n\,\mathrm{kl}(x,p)\big)$ where

$\mathrm{kl}(p,x) = x \log\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p}$

Pinsker Bernoulli: By Taylor's formula, since $f = \mathrm{kl}(\cdot, p)$ satisfies $f(p) = f'(p) = 0$ and $f''(x) = [x(1-x)]^{-1} \geq 4$, $\mathrm{kl}(x,p) = \dfrac{(x-p)^2}{2}\displaystyle\int_0^1 f''((1-s)p+sx)\,2s\,ds \geq 2(x-p)^2$

Hoeffding's inequality: if $0 < p \leq x < 1$, $\mathbb{P}_p\big(\bar{X}_n \geq p+\epsilon\big) \leq \exp\big(-2n\epsilon^2\big)$

Upper bound on the sample complexity $P_-(T=1) \leq \exp\big(-2n\epsilon^2\big) \leq \delta$ for $n \geq \dfrac{\log\frac{1}{\delta}}{2\epsilon^2}$

Similarly, if $0 < x \leq p < 1$, $\mathbb{P}_p\big(\bar{X}_n \leq x\big) \leq \exp\big(-n\,\mathrm{kl}(x,p)\big)$...

Let $L_n = \ln \frac{d\mathbb{P}_{x+\alpha}}{d\mathbb{P}_p}(X1, \ldots, X_n) = \sum_{i=1}^n \ln \frac{d\mathcal{B}(x+\alpha)}{d\mathcal{B}(p)}(X_i)$. For all $\alpha, \beta > 0$,

$$\begin{aligned}\mathbb{P}_p(\bar{X}_n \geq x) &= \mathbb{E}_{x+\epsilon}\left[\mathbb{1}\{\bar{X}_n \geq x\} e^{-L_n}\right] \\ &\geq e^{-n\left(\mathrm{kl}(x+\alpha,p)+\beta\right)} \mathbb{E}_{x+\epsilon}\left[\mathbb{1}\{\bar{X}_n \geq x\} \mathbb{1}\left\{\frac{L_n}{n} \leq \mathbb{E}_{x+\alpha}\left[\frac{d\mathcal{B}(x+\epsilon)}{d\mathcal{B}(p)}(X_i)\right] + \beta\right\}\right] \\ &= e^{-n\left(\mathrm{kl}(x+\alpha,p)+\beta\right)}\left(1 - o_n(1)\right)\end{aligned}$$

by the law of large numbers. Hence, large deviation principle: as $n \to \infty$,

$$\frac{1}{n} \ln \mathbb{P}_p\left(\bar{X}_n \geq x\right) \to -\mathrm{kl}(x, p)$$

"at exponential scale, Chernoff's bound is tight"

Batch Testing

Large deviation bounds

Lower bounds : TV and KL

Sequential Testing

A sequential lower bound

An "optimal" sequential test

"the test cannot discriminate between P_- and P_+ if it sees the same observation"

Coupling: Q probability on $\mathcal{X} \times \mathcal{X}$ such that $Q(\cdot, \mathcal{X}) = P_-$ and $Q(\mathcal{X}, \cdot) = P_+$.

Maximal coupling: Denoting $\Delta = \{(x, y) \in \mathcal{X} \times \mathcal{X} : x \neq y\}$, one can construct Q st $Q(\Delta) = \mathrm{TV}(P_-, P_+) \stackrel{\text{def}}{=} \sup_{A \in \mathfrak{A}} P_-(A) - P_+(A) = P_-(A^*) - P_+(A^*)$ for $A^* = \left\{ \frac{dP_+}{dP_-} \leq 1 \right\}$

Let $R = \frac{dP_+}{dP_-} \wedge 1$ and $r = P_-(Q) = 1 - \mathrm{TV}(P_-, P_+)$. Let then $Z \sim \frac{R}{q} dP_0$ and $Z_0 \sim (1-R)/(1-r)dP_-$, $Z_1 \sim (\frac{dP_+}{dP_-} - Q)/(1-q)dP_-$, and $B \sim \mathcal{B}(q)$. The pair $(Z, Z)U + (Z_1, Z_2)(1-U)$ has the right distribution.

$$\delta \geq P_+(T = 0) = Q(\mathcal{X}, \{T = 0\}) \geq Q(\{T = 0\}, \mathcal{X}) - Q(\Delta) \geq 1 - \delta - \mathrm{TV}(P_-, P_+)$$

ie $\mathrm{TV}(P_-, P_+) \geq 1 - 2\delta$

Other divergences like Kullback-Leibler better handle tensorization:

$$\mathrm{KL}(P, Q) = \begin{cases} \int \ln \frac{dP}{dQ}\, dP \text{ if } P \ll Q \\ +\infty \text{ otherwise} \end{cases}$$

Tensorization: $\mathrm{KL}(P_1 \otimes P_2, Q_1 \otimes Q_2) = \mathrm{KL}(P_1, Q_1) + \mathrm{KL}(P_2, Q_2)$

Contraction: (data-processing inequality) for every measurable map

$f : (\mathcal{X}, \mathfrak{A}) \to (\mathcal{Y}, \mathfrak{B})$, $\mathrm{KL}(P^f, Q^f) \leq \mathrm{KL}(P, Q)$ where $\forall B \in \mathfrak{B}, P^f(B) = P(f^{-1}(B))$

(pushforward measure)

Pinsker: $\mathrm{KL}(P, Q) \geq 2\,\mathrm{TV}(P, Q)^2$.

if $\mathrm{TV}(P, Q) = P(A^*) - Q(A^*)$, $\mathrm{KL}(P, Q) \geq \mathrm{KL}\left(P^{\mathbb{1}_{A^*}}, Q^{\mathbb{1}_{A^*}}\right) = \mathrm{kl}\left(P(A^*), Q(A^*)\right) \geq 2\big(P(A^*), Q(A^*)\big)^2 = 2\,\mathrm{TV}(P, Q)^2$

For every $\alpha > 0$, and for $f(x_1, \ldots, x_n) = \mathbb{1}\{x_1 + \cdots + x_n \geq nx\}$,

$$n\,\mathrm{kl}(x+\alpha, p) = \mathrm{KL}(\mathbb{P}_{x+\epsilon}, \mathbb{P}_p) \geq \mathrm{KL}\left(\mathbb{P}^f_{x+\epsilon}, \mathbb{P}^f_p)\right)$$

$$= \mathrm{kl}\left(\mathbb{P}_{x+\epsilon}(\bar{X}_n \geq x), \mathbb{P}_p(\bar{X}_n \geq x)\right) \geq \mathbb{P}_{x+\epsilon}(\bar{X}_n \geq x) \log \frac{1}{\mathbb{P}_p(\bar{X}_n \geq x)} - \ln(2)$$

since $\mathrm{kl}(p, q) = p \log \frac{1}{q} - h(p) + (1-p) \log \frac{1}{1-q} \geq p \log \frac{1}{q} - \ln(2)$,

$$\mathbb{P}_p(\bar{X}_n \geq x) \geq \exp\left(\frac{-n\ \mathrm{kl}(x+\alpha) - \ln(2)}{1 - e^{-2n\alpha^2}}\right)$$

Pinsker's inequality yields:

$$1 - 2\delta \le \mathrm{TV}(P_-, P_+) \le \sqrt{\frac{\mathrm{KL}(P_-, P_+)}{2}} = \sqrt{\frac{n\,\mathrm{kl}\left(\frac{1}{2} - \epsilon, \frac{1}{2} - \epsilon\right)}{2}}$$

and hence $n \ge \dfrac{2}{k}(1-2\delta)^2$ where $k = \mathrm{kl}\left(\frac{1}{2} - \epsilon, \frac{1}{2} - \epsilon\right) = 2\epsilon \ln \frac{1+2\epsilon}{1-2\epsilon} \le \frac{8\epsilon^2}{1-2\epsilon}$

In order to catch the good dependency in δ, need a tighter inequality for large KL like

Bretagnolle-Huber's inequality $\mathrm{TV}(P_-, P_+) \le \sqrt{1 - e^{-\mathrm{KL}(P_-,P_+)}} \le 1 - \dfrac{1}{2} e^{-\mathrm{KL}(P_-,P_+)}$

which yields

$$n \ge \frac{\log \frac{1}{4\delta}}{k} \ge \frac{(1-2\epsilon)\log \frac{1}{4\delta}}{8\epsilon^2}$$

Good dependency in ϵ and δ (but a factor 4 too large)

By contraction,

$$n\,\mathrm{kl}\left(\frac{1}{2}-\epsilon,\frac{1}{2}+\epsilon\right)=\mathrm{KL}(P_-,P_+)\geq\mathrm{KL}\left(P_-^T,P_+^T\right)$$
$$=\ \mathrm{kl}\left(P_-(T=0),P_+(T=0)\right)\geq\mathrm{kl}(1-\delta,\delta)\geq\ln\frac{1}{2.4\delta}$$

and hence

$$n\geq\frac{\ln\frac{1}{2.4\delta}}{k}\geq\frac{(1-2\epsilon)\ln\frac{1}{2.4\delta}}{8\epsilon^2}$$

Same bound by direct KL (information) manipulations.

Can we get rid of the factor 4?

$$n\,\mathrm{kl}\left(\frac{1}{2},\frac{1}{2}+\epsilon\right) \geq \mathbb{P}_{1/2}(T=0)\ln\frac{1}{P_+(T=0)} - \ln(2) \geq \mathbb{P}_{1/2}(T=0)\ln\frac{1}{\delta} - \ln(2)\,,$$
$$n\,\mathrm{kl}\left(\frac{1}{2},\frac{1}{2}-\epsilon\right) \geq \mathbb{P}_{1/2}(T=1)\ln\frac{1}{P_-(T=1)} - \ln(2) \geq \mathbb{P}_{1/2}(T=1)\ln\frac{1}{\delta} - \ln(2)$$

and hence, with $\kappa = \mathrm{kl}\left(\frac{1}{2},\frac{1}{2}+\epsilon\right) = \mathrm{kl}\left(\frac{1}{2},\frac{1}{2}-\epsilon\right) = \frac{1}{2}\ln\frac{1}{1-4\epsilon^2} \leq \frac{2\epsilon^2}{1-4\epsilon^2}$,
$2n\kappa \geq \ln\frac{1}{\delta} - 2\ln(2) = \ln\frac{1}{4\delta}$ that is $n \geq \dfrac{\ln\frac{1}{4\delta}}{2\kappa} \geq \dfrac{(1-4\epsilon^2)\log\frac{1}{4\delta}}{4\epsilon^2}$ still sub-optimal.

With $L_n = \ln \frac{dP_+}{d\mathbb{P}_{\frac{1}{2}}}(X1, \dots, X_n) = \sum_{i=1}^n \ln \frac{d\mathcal{B}(\frac{1}{2}+\epsilon)}{d\mathcal{B}(\frac{1}{2})}(X_i)$. For all $\beta > 0$,

$$\begin{aligned}\delta &\geq P_+(T=0) = \mathbb{E}_{\frac{1}{2}}\left[\mathbb{1}\{T=0\}\, e^{-L_n})\right] \\ &\geq e^{-n\left(\mathrm{kl}\left(\frac{1}{2},\frac{1}{2}+\epsilon\right)+\alpha\right)}\mathbb{E}_{\frac{1}{2}}\left[\mathbb{1}\{T=0\}\,\mathbb{1}\left\{\frac{L_n}{n} \leq \mathrm{kl}\left(\frac{1}{2},\frac{1}{2}+\epsilon\right)\right\}\right] \\ &= e^{-n(\kappa+\alpha)}\,\mathbb{P}_{\frac{1}{2}}(T=0)\left(1-o_n(1)\right)\end{aligned}$$

and similarly $\delta \geq e^{-n(\kappa+\alpha)}\,\mathbb{P}_{\frac{1}{2}}(T=1)\big(1-o_n(1)\big)$.

Hence $2\delta \geq e^{-n(\kappa+\alpha)}\big(1-o_n(1)\big)$, so that $n(\delta)$ can give a δ-correct test only if

$$\liminf_{\delta\to 0} \frac{n(\delta)}{\log\frac{2}{\delta}} \geq \frac{1}{\kappa} \geq \frac{1-4\epsilon^2}{2\epsilon^2}$$

Batch Testing

- Large deviation bounds
- Lower bounds : TV and KL

Sequential Testing

- A sequential lower bound
- An "optimal" sequential test

Possibility to interrupt the experiment at stopping time $\tau \leq n$ wrt the filtration $\mathcal{F}_t = \big(\sigma(X_1, \dots, X_t)\big)_{t \leq n}$ if sufficient evidence has been gathered. The test T then has to be $\mathcal{F}_\tau$-measurable.

Since the increments $\ln \frac{d\mathcal{B}\left(\frac{1}{2}-\epsilon\right)}{d\mathcal{B}\left(\frac{1}{2}+\epsilon\right)}(X_s) - \mathrm{kl}\left(\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon\right)$ are iid,

$$M_t = \sum_{s=1}^{t} \left(\ln \frac{d\mathcal{B}\left(\frac{1}{2} - \epsilon\right)}{d\mathcal{B}\left(\frac{1}{2} + \epsilon\right)}(X_s) - \mathrm{kl}\left(\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon\right)\right) \mathbb{1}\{s \geq \tau\} \text{ is a } P_-\text{-martingale.}$$

The restrictions P_-^τ (resp. P_+^τ) of P_- (resp. P_+) to $(\mathcal{X}, \mathcal{F}_\tau)$ satisfy $\mathrm{KL}(P_0^\tau, P_1^\tau) = \mathbb{E}_-\left[M_\tau\right] = \mathbb{E}_-[\tau]\,\mathrm{kl}(p, q)$. By concentration,

$$\begin{aligned}
\mathbb{E}_-[\tau]\,\mathrm{kl}\left(\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon\right) &= \mathrm{KL}(P_-^\tau, P_+^\tau) \geq \mathrm{KL}\left((P_-^\tau)^T, (P_+^\tau)^T\right) \\
&= \mathrm{kl}\left(P_-^\tau(T = 0), P_+^\tau(T = 0)\right) \geq \mathrm{kl}(1 - \delta, \delta) \geq \ln \frac{1}{2.4\delta}
\end{aligned}$$

Batch Testing

Large deviation bounds

Lower bounds : TV and KL

Sequential Testing

A sequential lower bound

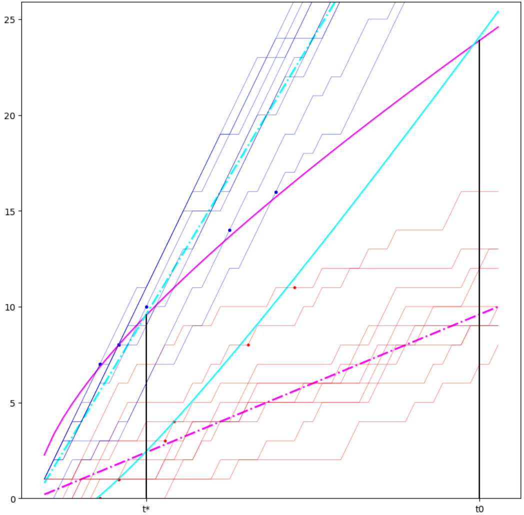
An "optimal" sequential test

This time, the bound is tight!

$$\tau_0 = \inf\left\{t \geq 1 : \bar{X}_t \geq \frac{1}{2} - \epsilon + \sqrt{\frac{\ln \frac{t^2}{2\delta}}{2t}}\right\}$$

$$\tau_1 = \inf\left\{t \geq 1 : \bar{X}_t \leq \frac{1}{2} + \epsilon - \sqrt{\frac{\ln \frac{t^2}{2\delta}}{2t}}\right\}$$

$$\tau = \min\{\tau_1, \tau_2\} \qquad T = \mathbb{1}\{\tau = \tau_1\}$$



Prop: $P_-(\tau_0 < \infty) \leq \delta$ and $P_+(\tau_1 < \infty) \leq \delta$, hence T is δ-correct

Prop: $\tau \leq \bar{t} \stackrel{\text{def}}{=} \frac{\ln \frac{1}{2\delta}}{2\epsilon^2} + \frac{2 \ln \frac{\ln \frac{2}{\delta}}{\epsilon^2}}{\epsilon^2}$ almost surely

Proof: let $u_t = \frac{1}{2} - \epsilon + \sqrt{\frac{\ln \frac{t^2}{2\delta}}{2t}}$ and $\ell_t = \frac{1}{2} + \epsilon - \sqrt{\frac{\ln \frac{t^2}{2\delta}}{2t}}$. For $t \geq \bar{t}$, $u_t \leq \ell_t$ and hence either $\tau_0 \leq \bar{t}$ or $\tau_1 \leq \bar{t}$. Indeed,

$$u_t \leq \ell_t \iff 2\epsilon \geq 2\sqrt{\frac{\ln \frac{t^2}{2\delta}}{2t}} \iff t \geq \frac{1}{2\epsilon^2} \ln \frac{1}{2\delta} + \frac{1}{\epsilon^2} \ln t$$

and we conclude since for all $\gamma \geq \alpha \geq 1/2$, $t \geq \gamma + 2\alpha \ln(2\gamma) \implies t \geq \gamma + \alpha \ln t$

Indeed, if $f(t) = t - \gamma - \alpha \ln(t)$, then $f'(t) = 1 - \alpha/t \geq 0$ iff $t \geq \alpha$. Hence, for all $t \geq t_0 \stackrel{\text{def}}{=} \gamma + 2\alpha \ln(2\gamma) \geq \alpha$, $f(t) \geq f(t_0) = \gamma + 2\alpha \ln(2\gamma) - \gamma - \alpha \ln\left(\gamma + 2\alpha \ln(2\gamma)\right) = \alpha \ln(4\gamma) - \alpha \ln\left(1 + \frac{2\alpha}{\gamma} \ln(2\gamma)\right)$ which has the same sign as $4\gamma - 1 - \frac{2\alpha}{\gamma} \ln(2\gamma) \geq 4\gamma - 1 - \frac{4\alpha\gamma}{e\,\gamma} \geq 4\alpha - 1 - 2\alpha \geq 0$

Prop: $\limsup_{\delta\to 0} \frac{\mathbb{E}_+[\tau]}{\ln\frac{1}{\delta}} \leq \frac{1}{8\epsilon^2}$ when $\delta \leq 4\%$

Proof: if $t^\star = \frac{\ln\frac{1}{2\delta}}{8\epsilon^2} + \frac{\ln\frac{\ln\frac{2}{\delta}}{4\epsilon^2}}{2\epsilon^2}$ then as above $t \geq t^\star \implies u_t \leq \left(\frac{1}{2}+\epsilon\right) t$

If $\alpha > 0$, $t \geq (1+\alpha)t^\star \implies u_t \leq \left(\frac{1}{2}+\epsilon - g(\alpha)\right) t$ for some $g(\alpha) > 0$, and hence

$$P_+\left(\tau_0 > t\right) \leq P_+\left(\bar{X}_t \leq \left(\frac{1}{2}+\epsilon-g(\alpha)\right)\right) \leq e^{-2tg(\alpha)^2}$$

and hence $\mathbb{E}_+[\tau] \leq (1+\alpha)t^\star + \sum_{t=(1+\alpha)t^\star}^{\infty} e^{-2tg(\alpha)^2} \leq (1+\alpha)t^\star + \frac{1}{1-e^{-2g(\alpha)^2}}$ which entails that $\liminf_{\delta\to 0} \frac{\mathbb{E}_+[\tau]}{\ln\frac{1}{\delta}} \leq \frac{1+\alpha}{8\epsilon^2}$ for all $\alpha > 0$.

With some more work, one can obtain an explicit non-asymptotic bound on $\mathbb{E}_+[\tau]$

- similar analysis for "$p \leq \frac{1}{2} - \epsilon$" vs "$p \geq \frac{1}{2} + \epsilon$" (where the sequential approach is even more relevant)
- generalize to other Bernoulli parameters
- generalize to other (one-parameter exponential) families of distributions
- non-parametric approach via the Empirical Likelihood method
- generalize to "$p \geq \frac{1}{2} + \epsilon$" vs "$p \leq \frac{1}{2} + \epsilon$" (more interesting)
- Best-arm identification in bandit models: generalize to *active* sequential sampling

- Non-Asymptotic Sequential Tests for Overlapping Hypotheses and application to near optimal arm identification in bandit models, *Aurélien Garivier, Emilie Kaufmann*, https://www.tandfonline.com/doi/abs/10.1080/07474946.2021.1847965
- A short note on an inequality between KL and TV, *Clément L. Canonne*, https://arxiv.org/abs/2202.07198v2
- Optimal Best Arm Identification with Fixed Confidence, *Aurélien Garivier, Emilie Kaufmann*; http://proceedings.mlr.press/v49/