Game-theoretic statistics and sequential anytime-valid inference (SAVI)

Aaditya Ramdas (aramdas@cmu.edu), Carnegie Mellon University

A brief recap of key definitions and facts in the non-sequential setting. References are from the E-book.

- **Hypothesis.** A hypothesis (null or alternative) is a set of distributions defined on a common underlying measure space $(\Omega, \mathcal{F})$. A hypothesis is simple if the set is a singleton, otherwise it is composite.
- **E-variables and e-values.** Given a null hypothesis $\mathcal{P}$, an e-variable E is a nonnegative random variable such that $\mathbb{E}_{\mathbb{P}}[E] \leq 1$ for all $\mathbb{P} \in \mathcal{P}$. Its realization is called an e-value.
- **Betting interpretation.** An e-variable is a bet against the null hypothesis. The corresponding e-value is the return received on each dollar that was bet. By definition, if the null is true, it is not profitable to bet against it.
- **Hypothesis test.** A level-α test for $\mathcal{P}$ is a $\{0,1\}$- or $[0,1]$-valued function ϕ such that $\mathbb{E}_{\mathbb{P}}[\phi] \leq \alpha$ for all $\mathbb{P} \in \mathcal{P}$.
- **From e to test.** Every e-variable E can be converted to a level-α test $\mathbb{I}(E \geq 1/\alpha)$, or simply $(\alpha E) \wedge 1$. The test's validity follows by *Markov's inequality*: $\mathbb{P}(E \geq 1/\alpha) \leq \alpha$ for all $\mathbb{P} \in \mathcal{P}$.
- **Universality.** *Every* level-α test ϕ can be recovered by thresholding an e-variable E_ϕ at $1/\alpha$. Indeed, $E_\phi := \phi/\alpha$. Thus, e-variables are fundamental, and the study of e-variables for $\mathcal{P}$ is implicitly the study of tests for $\mathcal{P}$.
- **P-variables and p-values.** Given a null hypothesis $\mathcal{P}$, a p-variable P is a $[0,1]$-valued random variable such that $\mathbb{P}(P \leq t) \leq t$ for all $\mathbb{P} \in \mathcal{P}$ and $t \in [0,1]$. Its realization is called an p-value.
- **Calibrator (Ch. 2.3).** If E is an e-variable, then $P = 1/E$ is a p-variable. Similarly, any p-variable P can be converted to an e-variable $E = h(P)$ using a *calibrator*: a decreasing function h such that $\int_0^1 h(u)du = 1$.
- **Combining e-values.** The only admissible way to combine arbitrarily dependent e-values is to take their (weighted) average (with nonnegative weights). Products of independent e-values are e-values.

The above concepts did not utilize the alternative hypothesis. Let's now bring that in.

- **E-power (Ch. 3.3).** Given a simple alternative $\mathbb{Q}$, the e-power of an e-variable E (for $\mathcal{P}$) is defined as $\mathbb{E}_{\mathbb{Q}}[\log E]$.
- **Numeraire, GRO, log-optimal e-variable (Ch. 6.1).** When testing $\mathcal{P}$ against $\mathbb{Q}$, the numeraire E^* is an e-variable for $\mathcal{P}$ such that, for every other e-variable E, we have $\mathbb{E}_{\mathbb{Q}}[E/E^*] \leq 1$. It always exists and is $\mathbb{Q}$-a.s. unique. It is also called the log-optimal or GRO e-variable, and it maximizes e-power (which could be infinite).
- **Reverse information projection (RIPr, Ch. 6.3).** The sub-probability measure $\mathbb{P}^* \ll \mathbb{Q}$ defined by $d\mathbb{P}^*/d\mathbb{Q} = 1/E^*$ is called the RIPr due to the strong duality $\mathbb{E}_{\mathbb{Q}}[\log E^*] = \sup_E \mathbb{E}_{\mathbb{Q}}[\log E] = \inf_{\mathbb{P}} \mathrm{KL}(\mathbb{Q}, \mathbb{P}) = \mathrm{KL}(\mathbb{Q}, \mathbb{P}^*)$, where the inf is taken over the *bipolar* of $\mathcal{P}$ (the *effective null*) and sup is over all e-variables for $\mathcal{P}$.
- **Likelihood ratio (Ch. 3.5, 6.4, 6.5).** When testing $\mathbb{P}$ against $\mathbb{Q} \ll \mathbb{P}$, the numeraire is their likelihood ratio $d\mathbb{Q}/d\mathbb{P}$. For composite $\mathcal{P}$, the *universal inference* e-variable defined by $\inf_{\mathbb{P}\in\mathcal{P}} d\mathbb{Q}/d\mathbb{P}$ is often easier to compute and is asymptotically log-optimal, but is dominated by the numeraire. The numeraire $E^* = d\mathbb{Q}/d\mathbb{P}^*$ is the only e-variable that can be expressed as a likelihood ratio between $\mathbb{Q}$ and some element of the effective null hypothesis.

The sequential setting (Ch. 7) considers distributions $\mathcal{P}$ on filtered spaces $(\Omega, \mathcal{F})$, where $\mathcal{F} = (\mathcal{F}_t)$ is a filtration.

- **Test supermartingales for $\mathcal{P}$.** A test supermartingale for $\mathcal{P}$ is a nonnegative $\mathcal{F}$-adapted process M such that $M_0 = 1$ and $\mathbb{E}_{\mathbb{P}}[M_t|\mathcal{F}_{t-1}] \leq M_{t-1}$, $\mathbb{P}$-a.s. for every $\mathbb{P} \in \mathcal{P}$. (It is a test martingale if $=$ replaces $\leq$.) Every test (super)martingale can be written as the product of (sequentially dependent) e-variables $E_t = M_t/M_{t-1}$ (0/0=0).
- **E-processes for $\mathcal{P}$ (Ch. 7.2).** An e-process for $\mathcal{P}$ is a nonnegative $\mathcal{F}$-adapted process E such that $\mathbb{E}_{\mathbb{P}\in\mathcal{P}}[E_\tau] \leq 1$ for all $\mathcal{F}$-stopping times τ. (We can restrict to bounded or finite stopping times.) By the *optional stopping theorem*, all test supermartingales are e-processes, but not vice versa. Under weak assumptions, admissible e-processes for $\mathcal{P}$ can be written as $E = \inf_{\mathbb{P}\in\mathcal{P}} M^{\mathbb{P}}$, where $M^{\mathbb{P}}$ is a test martingale for $\mathbb{P}$.
- **Sequential test.** A level-α sequential test for $\mathcal{P}$ is a binary adapted sequence ϕ_t such that $\mathbb{P}(\exists t \geq 1 : \phi_t = 1) \leq \alpha$ for all $\mathbb{P} \in \mathcal{P}$. Equivalently, for every stopping time τ, $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(\phi_\tau = 1) \leq \alpha$.

- **Ville's inequality** states that for any e-process M for $\mathcal{P}$, $\mathbb{P}(\exists \geq 1 : M_t \geq 1/\alpha) \leq \alpha$ for all $\mathbb{P} \in \mathcal{P}$. This implies that $\phi_t = \mathbb{I}(M_t \geq 1/\alpha)$ yields a level-α sequential test.
- **Universality.** *Every* level-α sequential test for $\mathcal{P}$ can be recovered by thresholding an e-process for $\mathcal{P}$ at $1/\alpha$. The same does not hold true for test supermartingales. A more general concept than e-processes is not required.
- **Log-optimality (Ch. 7.5).** One can define log-optimal (or numeraire) e-processes, and show that the likelihood ratio test martingale is log-optimal amongst all e-processes for sequentially testing $\mathbb{P}$ against $\mathbb{Q}$.
- **Method of mixtures (Ch. 3.7).** For composite alternatives $\mathcal{Q}$, a single e-process cannot be simultaneously log-optimal for all $\mathbb{Q} \in \mathcal{Q}$. The *method of mixtures* is a general technique that delivers *asymptotic* log-optimality (and regret bounds). The mixture e-process for $\mathcal{P}$ against $\mathcal{Q}$ is formed by considering base e-processes for $\mathcal{P}$ against $\mathbb{Q}$, and considering their weighted average (an integral wrt some mixture distribution over $\mathcal{Q}$).
- **UI vs RIPr (Ch. 7.7, 7.9).** The sequence of universal inference e-variables is actually an e-process, but the sequence of RIPr e-variables is in general not an e-process, but can be "time-mixed" to yield an e-process.
- **Confidence sequences (Ch. 13.1).** A $(1-\alpha)$-confidence sequence for a functional ψ (like the mean/median) is an adapted sequence of sets C_t such that for every $\mathbb{P} \in \mathcal{P}$, $\mathbb{P}(\forall t \geq 1 : \psi(\mathbb{P}) \in C_t) \geq 1-\alpha$, or equivalently that $\mathbb{P}(\psi(\mathbb{P}) \notin C_t) \leq \alpha$ for all stopping times τ. A universal way to construct these is to define an e-process E^θ to test $\{\mathbb{P} : \psi(\mathbb{P}) = \theta\}$, for every possible value of θ, and define $C_t = \{\theta : E_t^\theta < 1/\alpha\}$.

For multiple testing (Chapter 9), we consider testing K different null hypotheses $\mathcal{P}_1, \ldots, \mathcal{P}_K$.

- **Compound e-values.** A set of nonnegative random variables $E_1, \ldots, E_K$ are called compound e-variables for $\mathcal{P}_1, \ldots, \mathcal{P}_K$ if for all $\mathbb{P} \in \bigcup_k \mathcal{P}_k$, we have $\sum_{k:\mathbb{P}\in\mathcal{P}_k} \mathbb{E}_\mathbb{P}[E_k] \leq K$.
- **e-BH procedure.** Given compound e-values $E_1, \ldots, E_K$, the e-BH procedure runs the Benjamini-Hochberg procedure on $(1/E_1, \ldots, 1/E_K)$ to make $k^* := \max\{k : E_{[k]} \geq K/(\alpha k)\}$ rejections, where $E_{[k]}$ is the k-th largest e-value. This controls the false discovery rate below α under arbitrary dependence between the e-values.
- **E-collection.** $\{E_S\}_{S\subseteq[K]}$ is an e-collection if E_S is an e-variable for $\cap_{k\in S}\mathcal{P}_k$ for all $S \subseteq [K] := \{1, \ldots, K\}$. For example, given e-values $E_1, \ldots, E_K$, define the mean e-collection $E_S = \sum_{k\in S} E_k/|S|$.
- **Closed e-BH procedure.** Given an e-collection $\{E_S\}_{S\subseteq[K]}$, define $\mathcal{C} = \{R \subseteq [K] : E_A \geq |A \cap R|/(|R|\alpha)\}$, treating 0/0=0. The closed e-BH procedure rejects any largest set R^* in $\mathcal{C}$. This can be shown to control FDR below α under arbitrary dependence within the e-collection.
- **Universality.** Every procedure that controls FDR (in any setting, under any assumptions) can be expressed as an instance of the e-BH procedure applied to certain compound e-variables, and also as an instance of the closed e-BH procedure applied to some e-collection. When the mean e-collection is used, e-BH is identical to closed e-BH if $E_1, \ldots, E_K$ are compound e-values, but if they are e-values, closed e-BH improves e-BH.

There are several other miscellaneous topics worth mentioning.

- **Randomization (Ch. 2.4, 9.5)** can improve the power of (multiple) testing. The randomized Markov's inequality states that if E is an e-variable for $\mathbb{P}$ and U is a uniform $[0,1]$ random variable independent of E, then $\mathbb{P}(E \geq U/\alpha) \leq \alpha$. In other words, U/E is also a valid p-value. Finally, given any e-value E and any grid of positive reals, it is possible to *stochastically round* E to this grid while maintaining the e-value property.
- **Asymptotic e-values (Ch. 10)** It is possible to define approximate and asymptotic e-values in a natural way that requires the e-value property to only hold asymptotically. It possible to construct asymptotic e-values in settings where one cannot construct any nontrivial nonasymptotic e-value. In particular, one can construct asymptotic compound e-values using *Empirical Bayes* techniques and compound decision theory.
- **Post-hoc validity (Ch. 4).** When used for (multiple) testing, e-values not only guarantee type-I error (or FDR) control at a fixed level α, but allow for choosing α in a data-dependent fashion. In fact, any (multiple) test with such level-post-hoc control must be based on e-values.