Lecture 3: Growth rate optimality

Eugenio Clerico

November 2025

Testing a null against an alternative. When testing a null hypothesis $\mathcal{H}$, the first problem one faces is simply deciding *which* test to use. Even in very classical settings this choice is not always obvious. Consider, for instance, the problem of testing whether the mean of a Gaussian random variable is zero. If the data suggest a positive mean should we be more inclined to reject than if they suggest a negative one? Or should our testing rule be completely symmetric? In practice, asymmetry can matter: in some applications positive deviations may pose a major issue than negative ones, and one might wish to design a test that is particularly sensitive in that direction.

From the Neyman-Pearson perspective, this is formalised by specifying a "most relevant" alternative distribution $Q^\star$. If the data suggest that $Q^\star$ is the true distribution, we want to make sure to reject the null $\mathcal{H}$. A natural paradigm is then the following: *among all tests of significant level α for $\mathcal{H}$, one chooses the test with maximal power against $Q^\star$*, where *power* means the rejection probability when the data are actually generated from $Q^\star$. This is a fundamental idea, which has shaped much of the classical theory of testing.

Yet, this notion of power does not feel right when working with e-values. Suppose that for each α we select an optimal (in the above sense) level-α test ϕ_α[1] for the alternative $Q^\star$. One can form an e-variable by setting

$$E_\alpha = \frac{\phi_\alpha}{\alpha}\,,$$

and this is an optimal choice under the Neyman-Pearson paradigm. However, this construction immediately exposes its own limitations. First, it depends explicitly on the significance level: for each α we obtain a different e-variable. This clashes with one of the motivations for e-values, namely allowing post-hoc alpha tuning. Second, E_α might take the value 0 with non-zero probability under the alternative. This is not good for when we want to accumulate or aggregate evidence. For instance, when multiplying e-values from independent repetitions to deal with sequential data, a single zero erases all accumulated information, making rejection impossible even if the remaining data are overwhelmingly in favour of the alternative.

Numéraire property. For the reasons outlined above, when dealing with a fixed alternative $Q^\star$, the classical notion of power is not the most informative way to compare the usefulness of different e-variables. Recall that an e-variable is a function that, given a data set x, returns the amount of evidence we have against the null: larger values correspond to stronger evidence. Thus, if for two e-variables E_1 and E_2 we observe $E_1(x) > E_2(x)$, then E_1 provides strictly more evidence on that particular data set.

A natural possible way to compare two e-variables is through their ratio. If $E_2(x)/E_1(x)$ is small, then E_1 outperforms E_2 on the data set x. The issue, however, is that before seeing the data we do not know which x will be observed. For some data set x we may have $E_2(x)/E_1(x) < 1$, while for another the inequality may be reversed. To compare e-variables in a meaningful, data-independent way, an option is to weight each possible data set by how likely it is under the alternative $Q^\star$. This suggests that E_1 should be preferred to E_2 whenever

$$\mathbb{E}_{Q^\star}\left[\frac{E_2}{E_1}\right] \leq 1\,.$$

Quite remarkably, whenever we fix a null hypothesis $\mathcal{H}$ and an alternative $Q^\star$, there exists[2] an

[1] Here a test ϕ is a binary random variables such that $\phi(x) = 1$ if the test rejects after observing the data set x, and 0 otherwise.

[2] The proof is rather technical and beyond the scope of these notes; see Lemma 2.10 in *The numeraire e-variable and reverse information projection* by Martin Larsson, Aaditya Ramdas, and Johannes Ruf.

e-variable $E^\star$, unique up to $Q^\star$-null sets and strictly positive $Q^\star$-almost surely, such that

$$\mathbb{E}_{Q^\star}\left[\frac{E}{E^\star}\right] \le 1, \tag{1}$$

for every e-variable E. This property is known as the *numéraire* property, and it identifies $E^\star$ as a natural optimal choice when testing $\mathcal{H}$ against $Q^\star$.

GRO e-variables. The formulation above does not immediately suggest how to *find* $E^\star$, unlike the classical Neyman-Pearson setting, which is stated as a maximisation problem. However, $E^\star$ admits an equivalent characterisation as the maximiser of an expected-log criterion. In particular, $E^\star$ satisfies (1) if, and only if,

$$E^\star \in \arg\max_{E\in\mathcal{E}} \mathbb{E}_{Q^\star}[\log E]\,, \tag{2}$$

where $\mathcal{E}$ denotes the set of all e-variables for $\mathcal{H}$. The functional $E \mapsto \mathbb{E}_{Q^\star}[\log E]$ plays the role of power in the e-value framework: among all the e-variables, one chooses the one maximising the expected log-evidence. For this reason $\mathbb{E}_{Q^\star}[\log E]$ is often called the *e-power* of E under $Q^\star$.

Overall, maximising e-power is a rather natural choice. A key motivation comes from sequential or repeated experiments: when several independent e-values are collected over time, the total evidence is typically computed as the product of the individual e-values. Since logarithms turn products into sums, e-power is an additive quantity across experiments. $\mathbb{E}_{Q^\star}[\log E]$ describes the long-run exponential rate at which evidence accumulates under the alternative, and for this reason $E^\star$ is referred to as the *growth-rate optimal* (GRO) e-variable against $Q^\star$. This is entirely analogous to the classical Kelly criterion in information theory and gambling, where one chooses the strategy that maximises the expected logarithmic growth of wealth. Finally, $\log E$ also resolves the issue mentioned earlier: using $\mathbb{E}_{Q^\star}[\log E]$ as an objective strongly penalises $E = 0$, automatically discarding e-variables that take the value 0 with positive $Q^\star$-probability.

Now, let us show the equivalence between (1) and (2). Suppose first that $E^\star$ satisfies (1). Then for any $E \in \mathcal{E}$, by Jensen's inequality,

$$\mathbb{E}_{Q^\star}[\log E] - \mathbb{E}_{Q^\star}[\log E^\star] = \mathbb{E}_{Q^\star}\left[\log\frac{E}{E^\star}\right] \le \log \mathbb{E}_{Q^\star}\left[\frac{E}{E^\star}\right] \le 0\,,$$

which shows that $E^\star$ is a maximiser of e-power. Conversely, suppose now that (2) holds. Fix any $E \in \mathcal{E}$ and define $E_t = (1-t)E^\star + tE$ (which is still an e-variable), for $t \in (0, 1/2)$. Noting that $-2\log 2 \le \log(E_t/E^\star)/t \to E/E^\star - 1$ as $t \to 0$, Fatou's lemma brings that

$$\mathbb{E}_{Q^\star}\left[\frac{E}{E^\star}\right] - 1 = \mathbb{E}_{Q^\star}\left[\lim_{t\to 0}\frac{1}{t}\log\frac{E_t}{E^\star}\right] \le \liminf_{t\to 0}\frac{1}{t}\,\mathbb{E}_{Q^\star}\left[\log\frac{E_t}{E^\star}\right] \le 0\,,$$

which is (1).

Reverse information projection. The e-power of a GRO e-variable also admits a natural information-geometric interpretation: it coincides with the minimal relative entropy between $Q^\star$ and a suitably *extended* version of the null hypothesis. We now make this precise.

Define the *effective null* as the set

$$\mathcal{H}^\star = \left\{P \text{ sub-probability measure} : \mathbb{E}_P[E] \le 1 \text{ for all } E \in \mathcal{E}\right\},$$

where a sub-probability is a non-negative measure of total mass at most 1, and where $\mathbb{E}_P[E]$ denotes the usual expectation $\int_{\mathcal{X}} E(x)\mathrm{d}P(x)$. Geometrically, this set $\mathcal{H}^\star$ is called the *bipolar* of $\mathcal{E}$. It is the largest (sub-probabilistic) hypothesis that generates exactly the same class of e-variables as $\mathcal{H}$. We remark that the characterisation (1) of the GRO e-variable tells us that every $E \in \mathcal{E}$ is also an e-variable for the measure $P^\star$ defined via

$$\frac{dP^\star}{dQ^\star} = \frac{1}{E^\star}\,, \tag{3}$$

as we have that $\mathbb{E}_{P^\star}[E] = \mathbb{E}_{Q^\star}[E/E^\star] \le 1$. In particular, we see that $P^\star \in \mathcal{H}^\star$.

We define the *relative entropy* of $Q^\star$, with respect to a sub-probability P, as

$$D(Q^\star \mid P) = \mathbb{E}_{Q^\star}\left[\log\frac{dQ}{dP}\right]\,,$$

if $Q^\star \ll P$,[3] while $D(Q \mid P) = \infty$ if $Q \not\ll P$. A fundamental result in the theory of e-variables is that[4]

$$\sup_{E\in\mathcal{E}} \mathbb{E}_{Q^\star}[\log E] = \mathbb{E}_{Q^\star}[\log E^\star] = D(Q^\star \mid P^\star) = \inf_{P\in\mathcal{H}^\star} D(Q^\star \mid P)\,, \tag{4}$$

where $P^\star$, defined via (3), is the unique minimiser of $P \mapsto D(Q^\star \mid P)$ for P in the effective null hypothesis $\mathcal{H}^\star$. For this reason, $P^\star$ is commonly called the *reverse information projection* (RIPr) of $Q^\star$ onto $\mathcal{H}^\star$. (4) shows that the maximal e-power, namely $\mathbb{E}_{Q^\star}[\log E^\star]$, coincides with the relative entropy $D(Q^\star \mid P^\star)$ between the alternative and the closest element of the extended null $\mathcal{H}^\star$. This provides a clear interpretation: the larger the maximal e-power, the more easily $Q^\star$ can be distinguished from the null, and the further it lies from $\mathcal{H}^\star$ in an information-geometric sense.

It is also helpful to recall that the bipolar $\mathcal{H}^\star$ can be viewed as a form of *convex closure* of the original hypothesis. In particular, testing with e-variables cannot distinguish the null from any alternative $Q^\star$ that is itself a mixture of elements of $\mathcal{H}$. Such alternatives belong to $\mathcal{H}^\star$ and therefore have zero "information distance" to the null, which explains why testing via e-variables is powerless against them.

Simple e-variables. The simplest situation arises when testing a simple null $\mathcal{H} = \{P\}$ against a simple alternative $Q^\star$, with $Q^\star \ll P$. In this case, the likelihood ratio $\frac{dQ^\star}{dP}$ is always an e-variable for $\mathcal{H}$, because

$$\mathbb{E}_P\left[\frac{dQ^\star}{dP}\right] = \mathbb{E}_{Q^\star}[1] = 1\,.$$

Moreover, $\frac{dQ^\star}{dP}$ is in fact a GRO e-variable.

This follows from a more general observation. Suppose $\mathcal{H}$ is any hypothesis and that there exists $P \in \mathcal{H}^\star$ such that $Q^\star \ll P$ and $\frac{dQ^\star}{dP}$ is an e-variable for $\mathcal{H}$. Then $\frac{dQ^\star}{dP}$ is automatically a GRO e-variable for testing $\mathcal{H}$ against $Q^\star$. Indeed, let $E^\star \in \arg\max_{E\in\mathcal{E}} \mathbb{E}_{Q^\star}[\log E]$ be a maximiser of the e-power. Then

$$\mathbb{E}_{Q^\star}[\log E^\star] \geq \mathbb{E}_{Q^\star}\left[\log \frac{dQ^\star}{dP}\right] = D(Q^\star \mid P) \geq D(Q^\star \mid P^\star) = \mathbb{E}_{Q^\star}[\log E^\star]\,,$$

where $P^\star$ is the RIPr. All inequalities are therefore equalities, so

$$\mathbb{E}_{Q^\star}\left[\log \frac{dQ^\star}{dP}\right] = \mathbb{E}_{Q^\star}[\log E^\star]\,,$$

and by uniqueness of the GRO e-variable (up to $Q^\star$-null sets) we conclude that $\frac{dQ^\star}{dP} = E^\star$, $Q^\star$-almost surely.

If $E = \frac{dQ^\star}{dP}$ is an e-variable and $P \in \mathcal{H}$ (not only in $\mathcal{H}^\star$), E is often called a *simple e-variable*. Clearly, a simple e-variable is always GRO. Note, however, that not every hypothesis admits a simple e-variable: although a GRO e-variable always exists and can be written as $\frac{dQ^\star}{dP^\star}$, one may have $P^\star \in \mathcal{H}^\star \setminus \mathcal{H}$.

Next, we discuss a non-trivial setting where, under suitable assumptions, a simple e-variable can be explicitly found.

Regular exponential families. Let $\mathcal{X} = \mathbb{R}^d$, and let $U \subseteq \mathbb{R}^d$ be an open convex set containing 0. Let P_0 be a probability measure on $\mathcal{X}$ such that, for every $u \in U$,

$$A(u) = \log \mathbb{E}_{P_0}[\exp(u \cdot X)] < \infty\,,$$

and $u \mapsto A(u)$ is continuously differentiable. For each $u \in U$ we can define a probability measure P_u via

$$\frac{\mathrm{d}P_u}{\mathrm{d}P_0}(x) = \exp\left(u \cdot x - A(u)\right).$$

The set of all these measures is called a *regular exponential family*.

We remark that the mean of P_u can be easily computed. Indeed, we have that

$$\mathbb{E}_{P_u}[X] = \int_{\mathcal{X}} x e^{u\cdot x - A(u)} \mathrm{d}P_0(x) = \frac{\int_{\mathcal{X}} x e^{u\cdot x} \mathrm{d}P_0(x)}{\int_{\mathcal{X}} e^{u\cdot x} \mathrm{d}P_0(x)} = \nabla_u \log \mathbb{E}_{P_0}[e^{u\cdot X}] = \nabla A(u)\,.$$

[3]$Q^\star \ll P$ means that $Q^\star$ is absolutely continuous with respect to P, namely whenever for a measurable set A we have $P(A) = 0$, then $Q^\star(A) = 0$.

[4]See Theorem 3.6 in *The numeraire e-variable and reverse information projection* (by Martin Larsson, Aaditya Ramdas, and Johannes Ruf) for a proof and more details.

It is a well known fact that for a regular exponential family the set

$$M = \{\mathbb{E}_{P_u}[X] \,:\, u \in U\} = \nabla A(U)$$

is convex and open, and $\nabla A : U \to M$ is a bijection. This means that for every $m \in M$, there is a unique $u \in U$ such that P_u has mean m.

Notably, the choice of P_0 as *base measure* for the exponential family is rather arbitrary. Indeed, pick any $u \in U$. Let $V = U - u = \{u' - u \,:\, u' \in U\}$, and for $v \in V$ define $\hat{P}_v = P_{v+u}$. Then, $\hat{P}_0 = P_u$. We have that

$$\frac{\mathrm{d}\hat{P}_v}{\mathrm{d}\hat{P}_0}(x) = \frac{\mathrm{d}P_{u+v}}{\mathrm{d}P_u}(x) = \frac{\exp\big((u+v)\cdot x - A(u+v)\big)}{\exp\big(u\cdot x - A(u)\big)} = \exp\big(v\cdot x - (A(u+v) - A(u)\big)\,.$$

This show that the original exponential family can be re-parameterised around $\hat{P}_0$, using V instead of U as parameter set, and $\hat{A} : v \mapsto A(u+v) - A(u)$ instead of A.

GRO for a regular exponential family (part 1). We consider the case where

$$\mathcal{H} = \{P_u \,:\, u \in U\}$$

is a regular exponential family. Let M be the set of means for this family, and consider an alternative $Q^\star$ such that

$$\mathbb{E}_{Q^\star}[X] = \mu \in M\,.$$

We noted earlier that (up to possibly shifting the parametrisation) we can choose as base measure any element of $\mathcal{H}$. Thus, we can assume without loss of generality that

$$\mathbb{E}_{P_0}[X] = \mu\,.$$

We suppose that, for every $u \in U$, we have

$$B(u) = \log \mathbb{E}_{Q^\star}[\exp(u\cdot X)] < \infty\,,$$

and that B is differentiable with continuous derivatives. Then, we claim that

$$A(u) - B(u) \geq 0\,, \quad \forall u \in U \quad \iff \quad \frac{\mathrm{d}Q^\star}{\mathrm{d}P_0} \text{ is GRO.}$$

Recall that to show that $\frac{\mathrm{d}Q^\star}{\mathrm{d}P_0}$ is a GRO e-variable is enough to show that it is an e-variable, as in this case it is a *simple* e-variable. Define the function $h : U \to \mathbb{R}$ as

$$h(u) = \log \mathbb{E}_{P_u}\left[\frac{\mathrm{d}Q^\star}{\mathrm{d}P_0}\right]\,.$$

It is clear that $\mathrm{d}Q^\star/\mathrm{d}P_0$ is an e-variable if, and only if, $\sup_{u\in U} h(u) \leq 0$. We note that

$$h(u) = \log \mathbb{E}_{P_0}\left[\tfrac{\mathrm{d}Q^\star}{\mathrm{d}P_0}(X)\exp\big(u\cdot X - A(u)\big)\right] = \log \mathbb{E}_{Q^\star}\left[e^{u\cdot X}\right] - A(u) = B(u) - A(u)\,.$$

It is now clear that if $A - B$ is everywhere non-negative, than $h(u) \leq 0$ for all $u \in U$, and so $\frac{\mathrm{d}Q^\star}{\mathrm{d}P_0}$ is a (simple) e-variable.

The condition $A \leq B$ admits a natural information-geometric interpretation. It is a known fact that, among all distributions in an exponential family, the one that minimises the relative entropy to a fixed distribution is the member whose mean matches that of the target. Since P_0 and $Q^\star$ share the same mean μ, the only candidate for the RIPr *within* $\mathcal{H}$ is therefore P_0. The question becomes whether the global RIPr $P^\star$ lies inside $\mathcal{H}$ (in which case $P^\star = P_0$ and the GRO e-variable is simple) or whether $P^\star$ lies in the larger bipolar $\mathcal{H}^\star \setminus \mathcal{H}$, where it is a non-trivial mixture. Now observe that A and B are the log-moment generating functions of P_0 and $Q^\star$, and as such they encode their cumulants. Consequently, A is larger than B when the element of the exponential family $\mathcal{H}$ are "more spread out" than $Q^\star$. In that case, replacing P_0 by a mixture of P_us (i.e. moving away from $\mathcal{H}$ into $\mathcal{H}^\star \setminus \mathcal{H}$) would likely only increase the "distance" from $Q^\star$, and thus the RIPr is actually P_0. This corresponds exactly to the situation $A \geq B$, where $\mathrm{d}Q^\star/\mathrm{d}P_0$ is an e-variable and therefore GRO. Conversely, when $Q^\star$ is less concentrated than the measures in the exponential family, one can expect a well chosen mixture of the P_u to approximate $Q^\star$ better than P_0, making the RIPr shift into $\mathcal{H}^\star \setminus \mathcal{H}$. We try to make this intuition more clear in the next example.

Example: testing unit variance normals. As a simple illustration, consider the null hypothesis

$$\mathcal{H} = \{P_u = \mathcal{N}(u,1) : u \in \mathbb{R}\}\,,$$

and let the alternative be the centred normal distribution $Q^\star = \mathcal{N}(0,\sigma^2)$. The null $\mathcal{H}$ is a regular exponential family with natural parameter space $U = \mathbb{R}$, base measure $P_0 = \mathcal{N}(0,1)$, and log-partition function $A(u) = u^2/2$. For the alternative one readily checks that

$$B(u) = \log \mathbb{E}_{Q^\star}[e^{uX}] = \frac{\sigma^2 u^2}{2}\,.$$

Hence

$$A(u) - B(u) = \frac{1-\sigma^2}{2}\, u^2\,.$$

If $\sigma^2 < 1$, then $A - B$ is non-negative, and the criterion established above applies: $\frac{dQ^\star}{dP_0}$ is an e-variable and therefore the GRO e-variable. Its e-power is

$$\mathbb{E}_{Q^\star}\left[\log \tfrac{dQ^\star}{dP_0}\right] = \frac{\sigma^2 - 1}{2} - \log\sigma = D(Q^\star \mid P_0) > 0\,.$$

If instead $\sigma^2 > 1$, the situation reverses: $B(u) > A(u)$ for any $u \neq 0$, so $\frac{dQ^\star}{dP_0}$ is not an e-variable for $\mathcal{H}$. In this regime, however, something more can be said. A classical convolution identity shows that $\mathcal{N}(0,\sigma^2)$ is a mixture of unit-variance normals: for every measurable $S \subseteq \mathbb{R}$,

$$Q^\star(S) = \frac{1}{\sqrt{2\pi(\sigma^2-1)}} \int_{\mathbb{R}} \exp\left(-\frac{u^2}{2(\sigma^2-1)}\right) P_u(S)\, du\,.$$

Thus $Q^\star \in \mathcal{H}^\star$, and in particular no e-value test can distinguish $\mathcal{H}$ from $Q^\star$: the reverse information projection of $Q^\star$ onto $\mathcal{H}^\star$ is simply $P^\star = Q^\star$, and the corresponding growth-rate optimal e-variable is the trivial one, $E^\star \equiv 1$.

GRO for a regular exponential family (part 2). Returning to the notation introduced earlier, we now give two conditions that are sufficient (but in general not necessary) for $\mathrm{d}Q^\star/\mathrm{d}P_0$ to be a (simple) e-variable. To simplify the discussion, we assume that $U = \mathbb{R}^d$, that both A and B are twice continuously differentiable, and that $M = \nabla A(\mathbb{R}^d) = \mathbb{R}^d$ and $M_Q = \nabla B(\mathbb{R}^d) = \mathbb{R}^d$.

A first way to ensure that $\sup_{u\in\mathbb{R}^d} h(u) \leq 0$ is to assume that

$$\nabla^2 A(u) \succeq \nabla^2 B(u)\,, \qquad \forall u \in \mathbb{R}^d, \tag{5}$$

where $a \succeq b$ means that $a - b$ is non-negative definite. Under this assumption, $\nabla^2 h = \nabla^2 B - \nabla^2 A$ is everywhere non-positive definite, and hence h is concave. Since $h(0) = 0$ and

$$\nabla h(0) = \nabla B(0) - \nabla A(0) = \mathbb{E}_{Q^\star}[X] - \mathbb{E}_{P_0}[X] = \mu - \mu = 0\,,$$

the point 0 is a global maximiser of h. Thus $\sup_u h(u) = h(0) = 0$, and therefore $\mathrm{d}Q^\star/\mathrm{d}P_0$ is an e-variable.

An alternative sufficient condition result stems from classical convex duality theory. Since A and B are log-partition functions they are both convex and satisfy $A(0) = B(0) = 0$. Moreover, by construction $\nabla A(0) = \nabla B(0) = \mu$. Let $A^\star$ and $B^\star$ denote the convex conjugates of A and B. Then $A^\star(\mu) = B^\star(\mu) = 0$, and $\nabla A^\star(\mu) = \nabla B^\star(\mu) = 0$.

Because $M = M_Q = \mathbb{R}^d$, the gradient maps $\nabla A : \mathbb{R}^d \to \mathbb{R}^d$ and $\nabla B : \mathbb{R}^d \to \mathbb{R}^d$ are differentiable bijections. Let us denote their inverses by $\hat{u}$ and $\hat{u}_Q$, respectively. Hence for each $m \in \mathbb{R}^d$,

$$\nabla A(\hat{u}(m)) = m\,, \qquad \nabla B(\hat{u}_Q(m)) = m\,.$$

A standard identity in convex analysis gives, for all $m \in \mathbb{R}^d$,

$$\nabla^2 A^\star(m) = \left(\nabla^2 A(\hat{u}(m))\right)^{-1}, \qquad \nabla^2 B^\star(m) = \left(\nabla^2 B(\hat{u}_Q(m))\right)^{-1}.$$

Therefore, the assumption

$$\nabla^2 A(\hat{u}(m)) \succeq \nabla^2 B(\hat{u}_Q(m))\,, \qquad \forall m \in \mathbb{R}^d \tag{6}$$

is equivalent to $\nabla^2 B^\star(m) - \nabla^2 A^\star(m) \succeq 0$ for all m. Thus $B^\star - A^\star$ is convex with a global minimum at $m = \mu$, where it equals zero. Consequently $A^\star \geq B^\star$ everywhere, which by duality implies $B \geq A$ everywhere. Again, this shows that $\mathrm{d}Q^\star/\mathrm{d}P_0$ is an e-variable.

Although the conditions above may at first appear technical, they admit a natural interpretation. One may construct the exponential family $(Q_u)_{u\in\mathbb{R}^d}$ with base measure $Q_0 = Q^\star$ and log-partition function B. Then $\nabla^2 A(u)$ and $\nabla^2 B(u)$ are the covariance matrices of P_u and Q_u, respectively. Condition (5) requires that, for every canonical parameter u, the member P_u has covariance at least as large as the corresponding Q_u. Condition (6) compares the two families not at the same canonical parameter, but at the points $\hat{u}(m)$ and $\hat{u}_Q(m)$ that produce the same mean m: it states that for each mean vector m, the element of the P-family with mean m has covariance dominating that of the mean m element of the Q-family.

Bibliography. Most of the discussion on the GRO and numéraire e-variables is taken from *The numeraire e-variable and reverse information projection* by Martin Larsson, Aaditya Ramdas, and Johannes Ruf. Yet, the idea of using growth-rate optimality as a fundamental principle for constructing e-variables was fist advocated in *Safe testing*, by Peter Grünwald, Rianne de Heide, and Wouter Koolen. Finally, the discussion on testing exponential families is inspired by *Optimal E-Values for Exponential Families: the Simple Case*, by Peter Grünwald, Tyron Lardy, Yunda Hao, Shaul Bar-Lev, and Martijn de Jong.