E-Values / Exponential families

Peter Grünwald

Joint work with Yunda Hao and Tyron Lardy

CWI

Universiteit Leiden

Menu

1. GRO E-values for Exponential Family Nulls: the **Simple** Case

2. E-values for Exponential Family Nulls: the **Anti-Simple** Case
   - GRO and Conditional E-Values

3. Asymptotic growth difference for **UI** vs **GRO** vs **sequential GRO** e-variables/processes

Starter

- U : random quantity ; $X = t(U) \in \mathbb{R}^d$
- Q : distribution for U with density q
- $\mu^* := \mathrm{E}_Q[X]$; we assume Q has a moment generating function
- $\mathcal{P} = \{P_\beta : \beta \in B\}$: d-dimensional regular exponential family for U with sufficient statistic $X = t(U)$, and densities

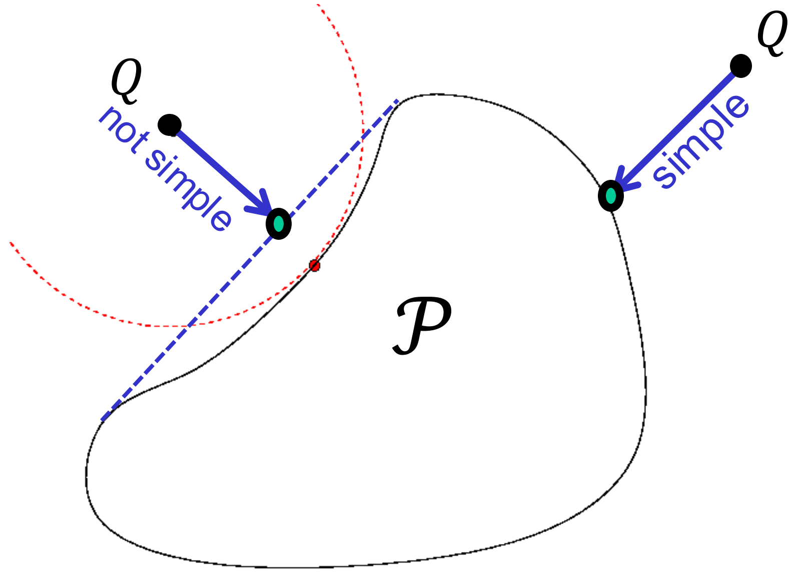
$$p_\beta(U) = \frac{1}{\boldsymbol{Z_p}(\beta)} \cdot \exp(\beta^T X) p_0(U)$$

  such that $0 \in B$ and $E_{P_0}[X] = \mu^*$.
- Then $\min_{\beta \in B} D(Q \,||\, P_\beta)$ uniquely achieved for $\beta = 0$.

- In "**simple**" case (which is very pleasant),

$$\min_{P \in \mathrm{conv}(\mathcal{P})} D(Q\,||P) = \min_{P \in \mathcal{P}} D(Q||P) \;\; (= D(Q||P_0)\,)$$

- Then P_0 is Reverse Information Projection (RIPr) on $\mathrm{conv}(\mathcal{P})$ so $q(U)/p_0(U)$ is Q-GRO e-variable
- First part of talk: generic condition under which we are in simple case



A Second Exponential Family

- Let $f(\beta) = \log \mathrm{E}_{P_\beta}\left[\frac{q(U)}{p_0(U)}\right]$
- Simple e-variable if $\boldsymbol{f(\beta) \leq 0, \forall \beta \in B}$. How to investigate this?

A Second Exponential Family

- Let $f(\beta) = \log \mathrm{E}_{P_\beta}\left[\frac{q(U)}{p_0(U)}\right]$
- Simple e-variable if $\boldsymbol{f(\beta) \leq 0, \forall \beta \in B}$. How to investigate this?

- First little surprise: $f(\beta) = \log Z_q(\beta) - \log Z_p(\beta)$
  with $Z_q(\beta)$ normalizer of another exponential family $\mathcal{Q}$,

  $$q_\beta(U) = \frac{1}{\boldsymbol{Z_q}(\beta)} \cdot \exp(\beta^T X) q(U) \;\; ; \;\; Z_q(\beta) = \int \exp(\beta^T X) q(U)$$

- $\mathcal{Q}$ has same sufficient statistic as $\mathcal{P}$ but different carrier

Intermezzo:
Exponential Family Duality Facts

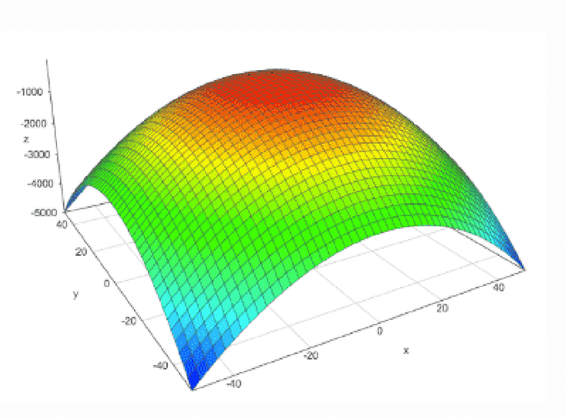
- P°_{μ} used to denote mean-value parameterization
- convex duality: $\beta^T \mu \leq \log \mathrm{Z}(\beta) - D(P^{\circ}_{\mu} \| P^{\circ}_{\mu^*})$,
  with equality iff $\mu = \mu(\beta)$
- ….where $\mu(\beta) \coloneqq E_{P_\beta}[X] = \nabla \log \mathrm{Z}(\beta)$
- $\Sigma(\beta) \coloneqq$ cov matrix of P_β = Hessian of $\log Z(\beta)$; $\Sigma(\beta)_{ij} = \frac{\partial^2}{\partial \beta_i \, \partial \beta_j} \log Z(\beta)$
- $\beta^{\circ}(\mu) \coloneqq$ inverse of $\mu(\beta)$ = $\nabla\, D(P^{\circ}_{\mu} \,||P^{\circ}_{\mu^*})$
- $\Sigma^{\circ}(\mu) \coloneqq \frac{\partial^2}{\partial \mu_i \, \partial \mu_j} \, D(P^{\circ}_{\mu}||P^{\circ}_{\mu^*})$ = Fisher inf matrix of P°_{μ} = $\left(\Sigma\big(\beta^{\circ}(\mu)\big)\right)^{-1}$

Local E-Variables

- Let $f(\beta) = \log \mathrm{E}_{P_\beta}\left[\frac{q(U)}{p_0(U)}\right]$. Simple e-variable if $f(\beta) \leq 0, \forall \beta \in B$.

- $f(\beta) = \log Z_q(\beta) - \log Z_p(\beta)$
- $f(0) = 0$, $\nabla f(\beta)_{|\beta=0} = \mu^* - \mu^* = 0$
- So q/p_0 is "local" e-variable if Hessian of $f(\beta)$, i.e. $\Sigma_q(\beta) - \Sigma_p(\beta)$ is negative definite at $\beta = 0$, i.e. if

  $\boldsymbol{\Sigma}_p^\circ(\boldsymbol{\mu}^*) - \boldsymbol{\Sigma}_q^\circ(\boldsymbol{\mu}^*)$ is positive definite!



Simple E-Variable Theorem, Simplest Version

- Start with exp family $\mathcal{P}$ and distr Q generating $\mathcal{Q}$ as before
- Let M_q, M_p and B_q, B_p be mean-value and canonical parameter spaces, respectively
- **Theorem:** suppose $M_q = M_p$ and $B_p \subseteq B_q$. Then:

q_μ / p_μ is (**GRO)** e-variable for **all** $\mu \in M_q$

iff

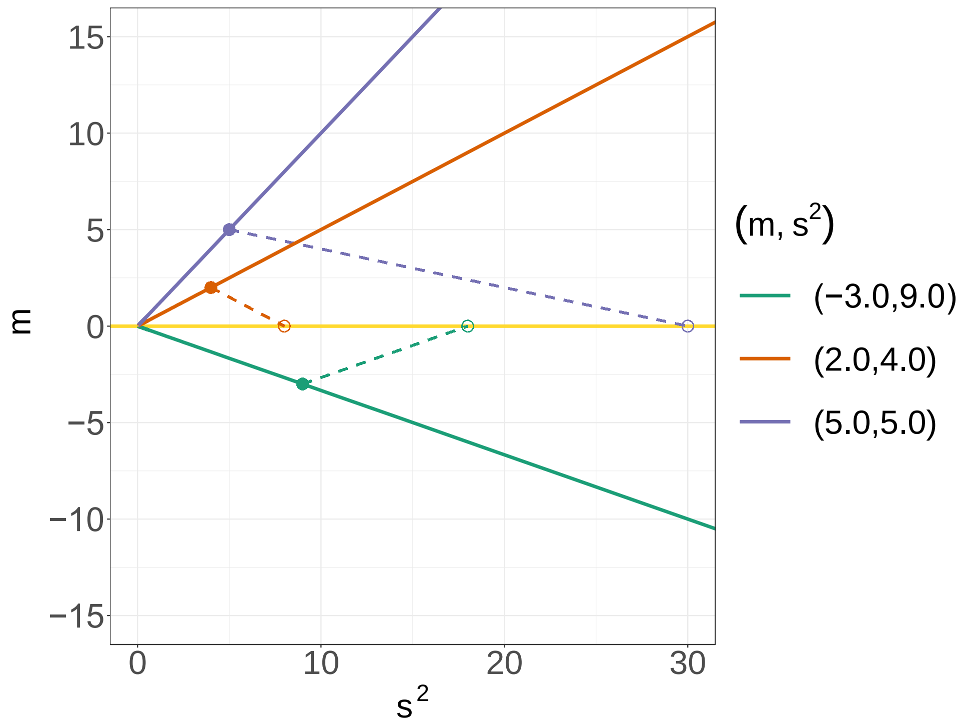
$\Sigma_p^\circ(\mu) - \Sigma_q^\circ(\mu)$ positive definite for **all** $\mu \in M_q$

Example 1: Gauss vs Gauss

- Let $U \sim Q = N(m, s^2)$, $X = U^2$. So $\mu^* = m^2 + s^2$
- $\mathcal{P}$: Gaussian scale family, $U \sim N(0, \sigma^2)$, $\sigma^2 > 0$.

- Preconditions on theorem and positive definiteness condition holds:
  $\Sigma_p^\circ(\mu) - \Sigma_q^\circ(\mu)$ positive definite for all $\mu \in M_q$

- Every choice m, s^2 determines family $\mathcal{Q}$, itself a 1-dimensional subset of the full Gaussian family, such that the projection of every member of $\mathcal{Q}$ onto $\mathcal{P}$ induces this very same family $\mathcal{Q}$



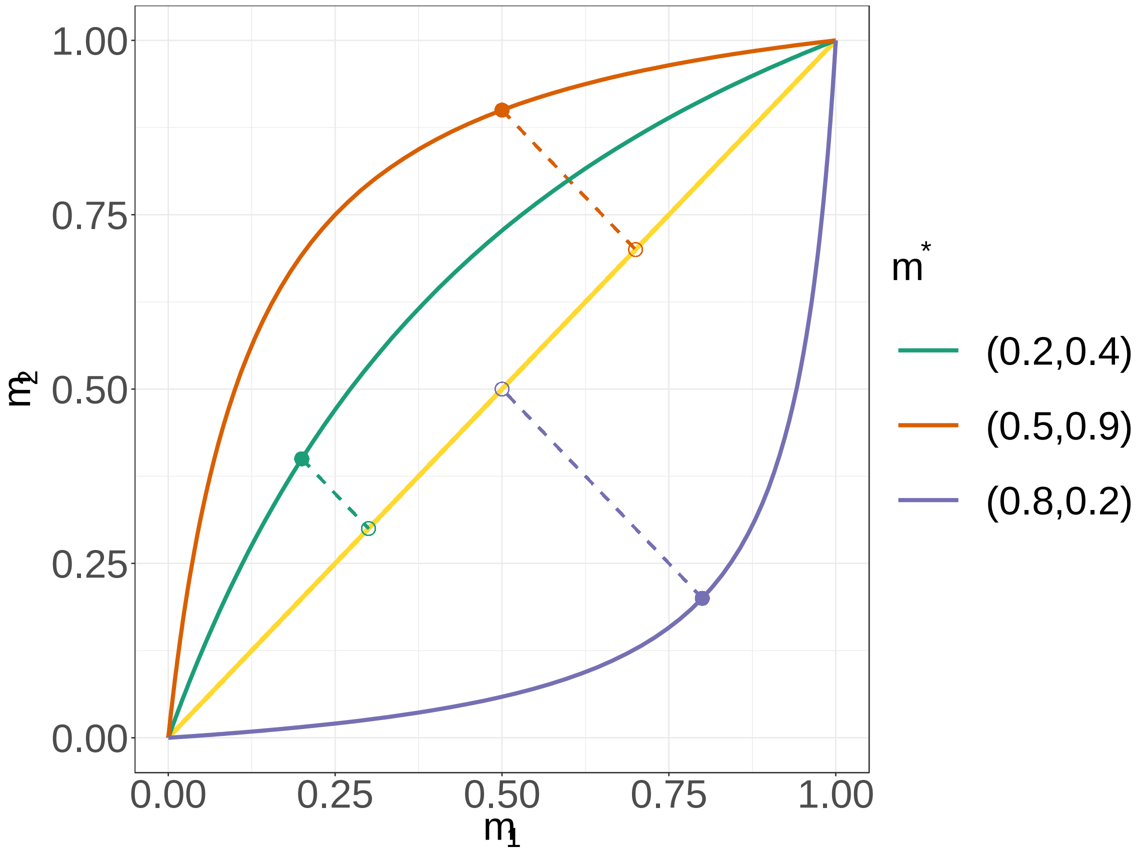
Example 2: k-Sample Bernoulli Test

- $Q: U = (U_1, U_2)\,, U_1 \sim \mathrm{Ber}(\mu_1^*), U_2 \sim \mathrm{Ber}(\mu_2^*)$, independent

- $\mathcal{P}$: $(U_1, U_2) \sim \text{iid}\,\mathrm{Ber}(\mu),\ \mu \in [0,1]$.

- Take $X = U_1 + U_2$ $(\mu^* = \mu_1^* + \mu_2^*)$

- Precondition and pos def conditions hold
- Every choice μ_1^*, μ_2^* determines family $\mathcal{Q}$, itself a 1-dimensional subset of the full 2x2 family



Proof Sketch

Thm: q_μ / p_μ is e-variable for **all** $\mu \in M_q \Leftrightarrow$

$\sigma_p^2(\mu) - \sigma_q^2(\mu)$ positive definite for **all** $\mu \in M_q$ [leave out $\circ$ for convenience]

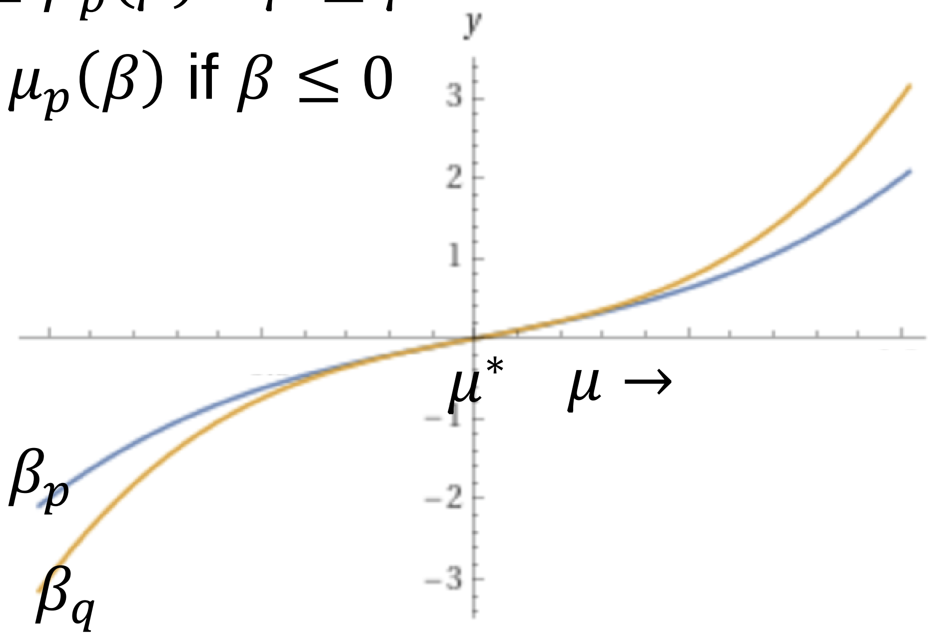
- note: condition equivalent to higher variance in $\mathcal{P}$
- " $\Rightarrow$" : follows directly from earlier reasoning (any e-variable is also a "local" e-variable)
- difficult part is " $\Leftarrow$ "
- Note: once we show " $\Leftarrow$ " it follows that q_μ / p_μ is GRO by Corollary 1 of G., De Heide, Koolen, JRSSB 2024.

Proof Sketch 1-d Case

- condition equivalent to higher variance in $\mathcal{P}$: $\sigma_p^2(\mu) \geq \sigma_q^2(\mu) \;\; \forall \mu \in M_q$, i.e. $I_q(\mu) \geq I_p(\mu)$
- Since $I(\mu)$ = $(\mathrm{d}/\mathrm{d}\mu)\ \beta(\mu)$ and $\beta_q(\mu^*) = \beta_p(\mu^*) = 0$, this implies $\forall \mu \in M_q$:

$$\beta_q(\mu) \geq \beta_p(\mu) \text{ if } \mu \geq \mu^* \text{ ; } \beta_q(\mu) \leq \beta_p(\mu) \text{ if } \mu \leq \mu^*$$

so $$\mu_q(\beta) \leq \mu_p(\beta) \text{ if } \beta \geq 0, \mu_q(\beta) \geq \mu_p(\beta) \text{ if } \beta \leq 0$$



Proof Sketch 1-d Case

- condition equivalent to higher variance in $\mathcal{P}$: $\sigma_p^2(\mu) \geq \sigma_q^2(\mu) \;\; \forall \mu \in M_q$, i.e. $I_q(\mu) \geq I_p(\mu)$
- Since $I(\mu)$ = $(\mathrm{d}/\mathrm{d}\mu)\,\beta(\mu)$ and $\beta_q(\mu^*) = \beta_p(\mu^*) = 0$, this implies $\forall \mu \in M_q$:

$$\beta_q(\mu) \geq \beta_p(\mu) \text{ if } \mu \geq \mu^* \;;\; \beta_q(\mu) \leq \beta_p(\mu) \text{ if } \mu \leq \mu^*$$

so $$\mu_q(\beta) \leq \mu_p(\beta) \text{ if } \beta \geq 0, \mu_q(\beta) \geq \mu_p(\beta) \text{ if } \beta \leq 0$$

- $\mu(\beta) = (\mathrm{d}/\mathrm{d}\beta)\log Z(\beta)$ now implies $\log Z_p(\beta) \geq \log Z_q(\beta)$ for all $\beta \in B_p$
- hence $f(\beta) \leq 0$ for all $\beta \in B_p$

Part II: The Anti-Simple Case

- What if opposite condition holds?
- First consider sweet multivariate Gaussian location case.
- Fix positive definite $d \times d$ matrices Σ_q, Σ_p .
- Let $Q = N(\mu, \Sigma_q)$ with density q_μ; $\mathcal{P} = \{ N(\mu, \Sigma_p) : \mu \in \mathbb{R}^d \}, U = X$
- Note $\Sigma_q(\mu)$ and $\Sigma_p(\mu)$ are constant as function of μ
- If $\Sigma_p - \Sigma_q$ positive semidefinite then by our theorem, q_μ / p_μ is e-variable.

Part II: The Anti-Simple Case

- What if opposite condition holds?
- First consider sweet multivariate Gaussian location case.
- Fix positive definite $d \times d$ matrices Σ_q, Σ_p .
- Let $Q = N\left(\mu, \Sigma_q\right)$ with density q_μ; $\mathcal{P} = \{ N(\mu, \Sigma_p) : \mu \in \mathbb{R}^d\}, U = X$
- Note $\Sigma_q(\mu)$ and $\Sigma_p(\mu)$ are constant as function of μ
- If $\Sigma_p - \Sigma_q$ positive semidefinite then by our theorem, q_μ / p_μ is e-variable

- If $\Sigma_p - \Sigma_q$ **negative** semidefinite (anti-simple case) then by our theorem q_μ / p_μ is **not** an e-variable $\Rightarrow$ need to consider $\mathrm{conv}(\mathcal{P})$ / mixtures

The Anti-Simple I.I.D. Case

- Now taking into account sample size becomes essential!
- RIPr of $Q_{\mu^*}^{(n)}$ onto $\mathcal{P}^{(n)}$ "must" be Bayes marginal P_W over $\mathcal{P}^{(n)}$ based on some prior W. Which one (guess!)?

The Anti-Simple Case

- Now taking into account sample size becomes essential!
- RIPr of $Q_\mu^{(n)}$ onto $\mathcal{P}^{(n)}$ "must" be Bayes marginal P_W over $\mathcal{P}^{(n)}$ based on some prior W. Which one?

The Anti-Simple Case

- Now taking into account sample size becomes essential!
- The RIPr of $Q_{\mu^*}^{(n)}$ onto $\mathcal{P}^{(n)}$ "must" be Bayes marginal P_W over $\mathcal{P}^{(n)}$ based on some prior W.
- It turns out $W = N\left(\mu^*, \frac{\Sigma_q - \Sigma_p}{n}\right)$

The Anti-Simple Case

- Now taking into account sample size becomes essential!
- The RIPr of $Q_{\mu^*}^{(n)}$ onto $\mathcal{P}^{(n)}$ "must" be some Bayes marginal over $\mathcal{P}^{(n)}$ based on some prior W.
- It turns out $W = N\left(\mu^*, \frac{\Sigma_q - \Sigma_p}{n}\right)$
- Proof Idea: in Gaussian anti-simple case, the GRO e-variable must be **equal** to the **COND**itional e-variable defined as

$$E_{\text{cond}}(X_1, \dots, X_n) := \frac{q_{\mu^*}(U^n \mid \hat{\mu}_p = n^{-1} \sum_{i=1..n} X_i)}{p_{\dots}(U^n \mid \hat{\mu}_p = n^{-1} \sum_{i=1..n} X_i)}$$

Proof Sketch

For arbitrary prior W,

$$\frac{q(U^n)}{p_W(U^n)} = \frac{q\left(U^n|\hat{\mu}_p=\bar{X}\right)}{p_W\left(U^n|\hat{\mu}_p=\bar{X}\right)} \cdot \frac{q^{[\hat{\mu}_p]}(\bar{X})}{p_W^{[\hat{\mu}_q]}(\bar{X})} = E_{\text{cond}} \cdot \frac{q^{[\hat{\mu}_p]}(\bar{X})}{p_W^{[\hat{\mu}_q]}(\bar{X})}$$

Marginal distributions of Gaussians with Gaussian priors are Gaussian...

Specifically if we plug in $W = N(\mu^*, (\Sigma_q - \Sigma_p)/n)$ then $Q^{[\hat{\mu}_p]} = P_W{}^{[\hat{\mu}_p]}$ and second factor is 1…so q/p_W coincides with E_{cond} and is e-variable

By Corollary 1 of Theorem 1 of GHK24, there can be at most one e-variable of form q/p_W and if it exists, it must be GRO!

Resulting GROwth for Gaussian Nulls

Anti-Simple Case:

one now gets with W set to RIPr prior relative to $Q = N\left(\mu^*, \Sigma_q\right)$:

$$\mathbf{E}_Q\left[\log\ E_{\mathrm{gro}}\right] = \ \mathbf{E}_Q\left[\log \frac{q(U^n)}{p_W(U^n)}\right] = \mathbf{E}_Q[E_{\mathrm{cond}}] = (n - \mathbf{1}) D_{\mathrm{Gauss}}(\Sigma_q \Sigma_p^{-1})$$

where $D_{\mathrm{Gauss}}(\Sigma_q \Sigma_p^{-1})$ is KL divergence between two 0-mean Gaussians with covariance matrices Σ_q and Σ_p respectively:

$$D_{\mathrm{Gauss}}(B) = \frac{1}{2}\left(-\log\det(B) - \left(d\ - \mathrm{tr}(B)\right)\right)$$

Resulting GROwth for Gaussian Nulls

Anti-Simple Case:

one now gets with W set to RIPr prior relative to $Q = N\left(\mu^*, \Sigma_q\right)$:

$$\mathbf{E}_Q\left[\log\ E_{\text{gro}}\right] = \mathbf{E}_Q\left[\log \frac{q(U^n)}{p_W(U^n)}\right] = \mathbf{E}_Q[E_{\text{cond}}] = (n - \mathbf{1}) D_{\text{Gauss}}(\Sigma_q \Sigma_p^{-1})$$

where $D_{\text{Gauss}}(\Sigma_q \Sigma_p^{-1})$ is KL divergence between two 0-mean Gaussians with covariance matrices Σ_q and Σ_p respectively:

$$D_{\text{Gauss}}(B) = \frac{1}{2}\left(-\log \det(B) - \left(d\ - \text{tr}(B)\right)\right)$$

Simple Case: $\mathbf{E}_Q\left[\log\ E_{\text{gro}}\right] = \mathbf{E}_Q\left[\log \frac{q(U^n)}{p_{\mu^*}(U^n)}\right] = n\ D_{\text{Gauss}}(\Sigma_q \Sigma_p^{-1})$

Resulting GROwth for Gaussian Nulls

Anti-Simple Case:

one now gets with W set to RIPr prior relative to $Q = N(\mu^*, \Sigma_q)$:

$$\mathbf{E}_Q[\log\ E_{\text{gro}}] = \mathbf{E}_Q\left[\log \frac{q(U^n)}{p_W(U^n)}\right] = \mathbf{E}_Q[E_{\text{cond}}] = (n - \mathbf{1}) D_{\text{Gauss}}(\Sigma_q \Sigma_p^{-1})$$

where $D_{\text{Gauss}}(\Sigma_q \Sigma_p^{-1})$ is KL divergence between two 0-mean Gaussians with covariance matrices Σ_q and Σ_p respectively:

$$D_{\text{Gauss}}(B) = \frac{1}{2}\left(-\log \det(B) - \left(d\ - \operatorname{tr}(B)\right)\right)$$

Simple Case: $\mathbf{E}_Q[\log\ E_{\text{gro}}] = \mathbf{E}_Q\left[\log \frac{q(U^n)}{p_{\mu^*}(U^n)}\right] = n\ D_{\text{Gauss}}(\Sigma_q \Sigma_p^{-1})$

Anti-Simple Case with Composite Alternative

- What if we take $Q = \{N(\mu, \Sigma_q) : \mu \in \mathbb{R}\}$ as our composite alternative and handle it by the method of mixtures, i.e. put a Gaussian prior $N(\mu^*, \Pi_q)$ on Q ?

- By the same reasoning as before, one finds the RIPr is p_W with $W = N\left(\mu^*, \Pi_q + \frac{\Sigma_q - \Sigma_p}{n}\right)$ and the resulting **Bayes factor** is again equal to the conditional e-variable!
- **…prior on null is almost the same as prior on alternative!**

Simple & Anti-Simple, Composite Alternative General Exponential Families

Put prior W_1 with pos. cont. density w_1 on alternative $\mathcal{Q} = \{Q_\mu : \mu \in \mathrm{M}_q\}$ one now gets with $W_{\mathrm{ripr,n}}$ set to RIPr prior relative to Q_{W_1} uniformly for all μ^* in any fixed compact subset of M_q:

$$\mathbf{E}_{Q_{\mu^*}}\left[\log\ E_{\mathrm{gro}}^{(n)}\right] = \mathbf{E}_Q\left[\log \frac{q_{W_1}(U^n)}{p_{W_{\mathrm{ripr},n}}(U^n)}\right] = \mathbf{E}_{Q_\mu^*}\left[\log E_{\mathrm{cond}}^{(n)}\right] + o(1) =$$

$$\mathbf{E}_Q\left[\log \frac{q_{W_1}(U^n)}{p_{W_1}(U^n)}\right] + o(1) = (n-1)D(Q \,||\, P_{\mu^*}) + o(1).$$

Note that $E_{\mathrm{cond}}^{(n)}$ **can be calculated without resorting to a prior or plug-in estimator**; there is no visible 'learning'. This is a remarkable result!