APPLICATIONS OF THE GENERALISED RÉNYI DIVERGENCES IN TESTING HYPOTHESES ABOUT EXPONENTIAL MODELS I

LUCIE FAJFROVÁ

.

ABSTRACT. We present the first part of report about the research concerning applications of Rényi divergences in testing hypotheses about exponential models. We focus on testing simple hypotheses about parameters in the case when our observations are independent and governed by a distribution of an exponential type. One can find here the theoretic part and a description of the testing algorithm. In the theoretic part we summarise basic terms and statements concerning testing using the Rényi divergence statistics. The description of the algorithm then can be used as a simplified software documentation. Moreover, a detailed list of treated distributions can be find in the Appendix.

1. INTRODUCTION

In this report we deal with divergences of real orders $r > 0$ introduced by Rényi [R61] for probability measures P, P_0 with densities f, f_0 on a σ-finite measure space $(\mathfrak{X}, \mathcal{A}, \mu)$ by the formulas:

$$D_r(P, P_0) = \frac{1}{r-1} \ln \int f^r f_0^{1-r} \, d\mu,$$

for $r \neq 1$, and

$$D_1(P, P_0) = \int f \ln \frac{f}{f_0} \, d\mu \tag{1.1}$$

where the integration extends over $\{x \in \mathfrak{X} : f(x) + f_0(x) > 0\}$ and the integrand is assumed to be infinite if the numerator is positive and denominator is zero. Under some restrictions on P, P_0 he established the continuity

$$\lim_{r \to 1} D_r(P, P_0) = D_1(P, P_0).$$

Later, Liese and Vajda [LV87] generalised the Rényi divergences to all real orders $r \in \mathbb{R}$ by the formulas (1.1) and

$$D_0(P, P_0) = D_1(P_0, P) \tag{1.2}$$

$$D_r(P, P_0) = \frac{1}{r(r-1)} \ln \int f^r f_0^{1-r} \, d\mu, \tag{1.3}$$

for $r \neq 1, r \neq 0$, where the convention about integrands are: $0 \cdot y = 0$ if $0 \leq y \leq \infty$ and $y/0 = \infty$ if $0 < y \leq \infty$. These authors also precised and extended the continuity law to all P, P_0 as follows

$$\lim_{r \nearrow 1} D_r(P, P_0) = D_1(P, P_0)$$

and noticed the skew symmetry about the order $r = 1/2$,

$$D_r(P_0, P) = D_{1-r}(P, P_0), \quad r \in \mathbb{R}.$$

The only Rényi divergence which is symmetric for all P, P_0 is that of the order $r = 1/2$,

$$D_{1/2}(P_0, P) = D_{1/2}(P, P_0) = 4 \ln \int \sqrt{f f_0}\, d\mu.$$

It was introduced already by Bhattacharyya [Bh46] and is known as the Bhattacharyya distance, see e.g. Zachs [Z71]. Another known members of the generalised Rényi divergence class are the information distances (Kullback distances) (1.2) introduced by Kullback and Leibler [KL51].

The Rényi divergences are traditionally applied in the testing hypotheses about statistical models with independent observations (see e.g. [MPV97] and references therein), about random processes (see e.g. [P74] or [V90]) and also about random fields (see e.g. [J88]).

Morales et al. [MPV00] for the first time pointed out that the Rényi divergences between theoretical and empirical distributions from general exponential models take on very simple form and that they are therefore suitable for testing hypotheses. They illustrated this by testing simple hypotheses about exponential random processes. In Morales et al. [MPPV04] this method was extended to composite hypotheses about exponential random processes, in particular about Lévy processes.

The aim of our research is to verify practical statistical applicability of the generalised Rényi divergences and the related Rényi statistics in testing hypotheses about independent exponential models, and also about exponential random processes and random fields. In this report we summarise the main results of this research achieved in the period 2005-2006. We expect to continue this effort in the future and summarise the more recent results in a subsequent reports.

Our research is mainly based on the results of [MPV00], [MPPV04] and on the theoretical background concerning statistical exponential families provided by the monographs of Brown [Br86] and Küchler and Sørensen [KS97]. For the sake of completeness, in Sections 2-4 we summarise the basic concepts and facts established in these references which are relevant for our research. In Section 5 the detailed results of our research, focused on models with independent observations, are presented. At the end of this report an appendix containing a list of basic exponential distributions is attached.

2. Natural exponential families

We consider exponential families in the so-called natural forms. Let $\mathfrak{X}$ be a metric observation space with the Borel field $\mathcal{A}$ of events and μ a σ-finite measure on $\mathcal{A}$. Further, let $T = (T_1, ..., T_d)$ be a measurable mapping from $\mathfrak{X}$ to $\mathbb{R}^d$ and $\Theta \subset \mathbb{R}^d$ the set of all $\theta = (\theta_1, ..., \theta_d)$ for which

$$c(\theta) := \int_{\mathfrak{X}} \exp\{\theta' T(x)\}\, d\mu \;<\; \infty$$

where $'$ denotes here and in the sequel the vector or matrix transpose and we assume column vectors. As well known, cf. [Br86], the set of natural parameters Θ is convex and the *cumulant generating function*

$$\kappa(\theta) = \ln c(\theta)$$

is convex on Θ and infinitely differentiable in the interior Θ.

By an *exponential family of densities* we mean the set $\mathcal{E} = \{f_\theta : \theta \in \Theta\}$ of probability densities given by formula

$$f_\theta(x) = \exp\{\theta' T(x) - \kappa(\theta)\} \tag{2.1}$$

for every $x \in \mathfrak{X}$.

Exponential families $\mathcal{E}$ of densities are specified by pairs (T, μ). We restrict ourselves in this paper to the pairs for which $\mathcal{E}$ are minimal in the sense that densities corresponding to different parameters cannot μ-a.s. coincide, see e.g. [Br86, p.2]. In these families the model is identifiable by the parameter and the function κ is strictly convex on Θ. By [Br86, Th.1.9] this restriction means no loss of generality. Further, we consider only the families $\mathcal{E}$ which are regular in the sense that Θ is open, see again [Br86, p.2].

Denote by

$$\dot{\kappa} = \Big(\frac{\partial \kappa}{\partial \theta_1}, ..., \frac{\partial \kappa}{\partial \theta_d}\Big)'$$

the vector of first derivatives of κ, $\dot{\kappa} : \Theta \to \mathbb{R}^d$, and by

$$\ddot{\kappa} = \Big(\frac{\partial^2 \kappa}{\partial \theta_i \, \partial \theta_j} : 1 \le i, j \le d\Big)$$

the matrix of second derivatives of κ, $\ddot{\kappa} : \Theta \to \mathbb{R}^{2d}$.

Lemma 2.1. *If* $\mathbf{X} \sim f_\theta$ *and* $f_\theta \in \mathcal{E} \equiv (T, \mu)$ *then*

$$\mathsf{E}\, T(\mathbf{X}) = \dot{\kappa}(\theta) \qquad \textit{and} \qquad \mathsf{Var}\, T(\mathbf{X}) = \ddot{\kappa}(\theta). \tag{2.2}$$

Proof. By [Br86, Th.2.2], the conditions for differentiation behind the integrals

$$\int_{\mathfrak{X}} e^{\theta' T(x) - \kappa(\theta)} \, d\mu \qquad \text{and} \qquad \int_{\mathfrak{X}} (T(x) - \dot{\kappa}(\theta)) e^{\theta' T(x) - \kappa(\theta)} \, d\mu$$

are satisfied. Since f_θ is a probability density, $\int_{\mathfrak{X}} e^{\theta' T(x) - \kappa(\theta)} \, d\mu = 1$ and then

$$\int_{\mathfrak{X}} (T(x) - \dot{\kappa}(\theta)) e^{\theta' T(x) - \kappa(\theta)} \, d\mu = 0 \in \mathbb{R}^d.$$

The last identity provides the first relation of (2.2). By applying operator $(\frac{\partial}{\partial \theta_1}, \ldots$ $\ldots, \frac{\partial}{\partial \theta_d})$ on the last identity - componentwise, we obtain the second relation. □

Since κ is strictly convex on Θ, the mapping $\dot{\kappa} : \Theta \to \mathbb{R}^d$ is invertible, let us denote by $\dot{\kappa}^{-1}$ its inverse, and the *Fisher information* $\ddot{\kappa}(\theta)$ is positive semidefinite. Taking into account the minimality of $\mathcal{E}$, one obtains that it is, in addition, positive definite and therefore $\mathcal{K}$, the range of $\dot{\kappa}$, is an open convex subset of $\mathbb{R}^d$. By [Br86, Th.3.6] and [Br86, (2) on p.145], the observation $x \in \mathfrak{X}$ with $T(x)$ not in closure of $\mathcal{K}$ are of μ-measure zero, i.e.

$$\mu(\, T \in \mathbb{R}^d \backslash \mathrm{cl}\, \mathcal{K}\,) = 0.$$

Of particular interest are the families $\mathcal{E}$ which satisfy the stronger condition: $\mu(\, T \in \mathbb{R}^d \backslash \mathcal{K}\,) = 0$. In such families

$$\mathsf{P}(\, T(\mathbf{X}) \in \mathrm{bd}\, \mathcal{K}\,) = 0 \quad \text{for} \quad \mathbf{X} \sim f_\theta \quad \text{for each} \quad \theta \in \Theta. \tag{2.3}$$

Differences between probability densities from an arbitrary family $(f_\theta : \theta \in \Theta)$ defined w.r.t. a σ-finite measure μ on $\mathfrak{X}$ can be characterised by the Rényi divergences, cf. Rényi [R61], Csiszár [C95]. As started in Section 1, Liese and Vajda [LV87] introduced the extended class of Rényi divergences for densities f_θ, f_{θ_0} given by formulas (1.1), (1.2) and (1.3). They proved that for densities $f_\theta, f_{\theta_0} \in \mathcal{E} \equiv (T, \mu)$ with the cumulant generating function κ the Rényi divergences are expressed by

$$D_r(\theta, \theta_0) = \begin{cases} \dfrac{1}{r(r-1)} \big[\kappa\big(r\theta + (1-r)\theta_0\big) - r\kappa(\theta) - (1-r)\kappa(\theta_0)\big]\,, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad r\theta + (1-r)\theta_0 \in \Theta \\ \infty \qquad\qquad\qquad\qquad\qquad\quad\;\; \text{otherwise} \end{cases} \tag{2.4}$$

for all real $r \neq 0$, $r \neq 1$, and

$$D_1(\theta, \theta_0) = \dot{\kappa}(\theta)'\,(\theta - \theta_0) + \kappa(\theta_0) - \kappa(\theta). \tag{2.5}$$

For example,

$$\frac{e^{r\theta+(1-r)\theta_0} - re^{\theta} - (1-r)e^{\theta_0}}{r(r-1)} = \frac{\lambda^r\lambda_0^{1-r} - r\lambda - (1-r)\lambda_0}{r(r-1)}$$

is the Rényi divergence $D_r(\theta, \theta_0)$ of Poisson distributions P, P_0 (where $\lambda = e^{\theta}$ and $\lambda_0 = e^{\theta_0}$ are the corresponding rates), for all r different from 0 and 1, and

$$(\theta - \theta_0)e^{\theta} + e^{\theta_0} - e^{\theta} = \lambda\left(\frac{\lambda_0}{\lambda} - 1 - \ln\frac{\lambda_0}{\lambda}\right)$$

is the Kullback-Leibler divergence $D_1(\theta, \theta_0)$.

If θ_0 is in interior Θ then the distance

$$\triangle(\theta_0) = \inf_{\theta\notin\Theta} \|\theta_0 - \theta\|$$

of θ_0 from the complement of Θ is positive (infinite if $\Theta = \mathbb{R}^d$). The following assertion is obvious.

Lemma 2.2. *If $\theta = \theta_0$ then the upper formula in (2.4) holds for all $r \in \mathbb{R} \setminus \{0, 1\}$. If $\theta \neq \theta_0$ then the same formula holds for all $r \in \mathbb{R} \setminus \{0, 1\}$ which satisfy the condition*

$$-\frac{\triangle(\theta_0)}{\|\theta_0 - \theta\|} < r < \frac{\triangle(\theta_0)}{\|\theta_0 - \theta\|}.$$

3. Convergent exponential families

We shall consider exponential experiments such that the size t of observations may come from an arbitrary directed set $\mathcal{T}$. Set $\mathcal{T}$ is called directed if it is partially ordered and every finite subset is dominated by an element from $\mathcal{T}$. Typical examples are $\mathcal{T} = \{0, 1, 2, ...\}$ or $\mathcal{T} = [0, \infty)^d$ or $\mathcal{T} = \{..., -1, 0, 1, ...\}^d$. For a generalised sequence $(x_t)_{t\in\mathcal{T}}$ with values in $\mathbb{R}$ we define $\lim_t x_t = x$ if for every $\varepsilon > 0$ there exists $t_\varepsilon \in \mathcal{T}$ for which $t \geq t_\varepsilon$ implies $\|x_t - x\| < \varepsilon$.

Let $\mathcal{E}_t \equiv (T_t, \mu_t)$ be a generalised sequence of exponential families assumed in Section 2, with the corresponding measure spaces $(\mathfrak{X}_t, \mathcal{A}_t, \mu_t)$, functions κ_t and with a common natural parameter space $\Theta = \Theta_t$. For simplicity we shall assume that also the range $\mathcal{K} \in \mathbb{R}^d$ of functions $\dot{\kappa}_t$ is common. This way we may represent observations on random sequences, processes and fields. This will be apparent from the examples bellow.

As argued in the previous section, $\dot{\kappa}_t$ is a homeomorphism of Θ and $\mathcal{K}$, where $\operatorname{cl}\mathcal{K}$ is the support of statistics $T_t = T_t(\mathbf{X}_t)$ for observations $\mathbf{X}_t$ distributed by any density $f_{\theta,t}$ from $\mathcal{E}_t$. By choosing a fixed $\theta_* \in \mathbb{R}^d$ and putting

$$\dot{\kappa}_t^{-1}(y) = \theta_* \qquad \text{for all } y \in \mathbb{R}^d \setminus \mathcal{K} \tag{3.1}$$

we obtain an extension of functions $\dot{\kappa}_t^{-1}$ on the whole space $\mathbb{R}^d$.

Define the estimator

$$\hat{\theta}_t = \hat{\theta}_t(T_t) = \dot{\kappa}_t^{-1}(T_t) \qquad \text{for } T_t = T_t(\mathbf{X}_t). \tag{3.2}$$

As easy to verify, $f_{\hat{\theta}_t,t}(\mathbf{X}_t) = \max_\theta f_{\theta,t}(\mathbf{X}_t)$ whenever the maximum exists. Thinking about density $f_{\theta,t}$ as about a function of $\theta \in \Theta$ then we call it the *likelihood* function. Therefore (3.2) can be viewed as a *maximum likelihood estimator* (MLE) of the parameter θ_0 figuring in the distribution density of the observation $\mathbf{X}_t$.

Obviously, if family $\mathcal{E}_t$ satisfies (2.3) then (by results mentioned in previous section) $P(T_t \notin \mathcal{K}) = 0$. Since the MLE for data $\mathbf{X}_t$ exists if and only if $T_t = T_t(\mathbf{X}_t) \in \mathcal{K}$, it implies that $\hat{\theta}_t$ maximises the likelihood almost surely. We shall assume a weaker condition than (2.3) is, namely,

$$\lim_{t\to\infty} \mathsf{P}(\,T_t \in \mathrm{bd}\,\mathcal{K}\,) = 0 \quad \text{for} \quad T_t = T_t(\mathbf{X}_t),\ \mathbf{X}_t \sim f_{\theta_0,t} \in \mathcal{E}_t. \tag{3.3}$$

Under this assumption,

$$\lim_{t\to\infty} \mathsf{P}\left(f_{\hat{\theta}_t,t}(\mathbf{X}_t) = \max_{\theta} f_{\theta,t}(\mathbf{X}_t) \right) = 1.$$

Let us now suppose that there exists a generalised sequence $\gamma_t \nearrow \infty$ for which

$$\lim_t \frac{\kappa_t(\theta)}{\gamma_t} = \kappa(\theta) \qquad \text{for all } \theta \in \Theta. \tag{3.4}$$

By [Ro70, Th.10.8], $\kappa(\theta)$ is convex on Θ and the convergence is locally uniform. By [Ro70, Th.25.7], if $\kappa(\theta)$ is differentiable then also

$$\lim_t \frac{\dot{\kappa}_t(\theta)}{\gamma_t} = \dot{\kappa}(\theta) \qquad \text{for all } \theta \in \Theta, \text{where } \dot{\kappa} = \left(\frac{\partial\kappa}{\partial\theta_1}, ..., \frac{\partial\kappa}{\partial\theta_d}\right)' \tag{3.5}$$

and this convergence is also locally uniform. If $\kappa(\theta)$ is moreover strictly convex and infinitely differentiable on Θ, then $\dot{\kappa}(\theta)$ is invertible and differentiable on Θ. In the following definition we consider $\dot{\kappa}^{-1}$ extended on $\mathbb{R}^d$ by means of the same θ_* as in (3.1), and the matrices $\ddot{\kappa}(\theta)$ and their inverses $\ddot{\kappa}(\theta)^{-1}$.

Definition 3.1. *A generalises sequence $\mathcal{E}_t \equiv (T_t, \mu_t)$, $t \in \mathcal{T}$, of exponential models under consideration is said to be convergent if (3.3) and (3.5) hold for γ_t and $\kappa(\theta)$ specified above and, moreover, take place the locally uniform convergences*

$$\lim_t \frac{\ddot{\kappa}_t(\theta)}{\gamma_t} = \ddot{\kappa}(\theta) \qquad \textit{for all } \theta \in \Theta \tag{3.6}$$

and

$$\lim_t \dot{\kappa}_t^{-1}(\gamma_t\, y) = \dot{\kappa}^{-1}(y) \qquad \textit{for all } y \in \mathbb{R}^d, \tag{3.7}$$

as well as the law of large numbers

$$\lim_t \frac{T_t - \mathsf{E}T_t}{\gamma_t} = 0 \qquad \textit{in probability} \tag{3.8}$$

and the central limit theorem

$$\lim_t \frac{T_t - \mathsf{E}T_t}{\sqrt{\gamma_t}} = \mathsf{N}(0, \ddot{\kappa}(\theta_0)) \qquad \textit{in distribution.} \tag{3.9}$$

For convergent exponential models $\mathcal{E}_t$, (3.6) implies

$$\lim_t \gamma_t\, \ddot{\kappa}_t(\theta)^{-1} = \ddot{\kappa}(\theta)^{-1} \qquad \text{for all } \theta \in \Theta \tag{3.10}$$

in the locally uniform sense. Further, for these models (3.3) implies

$$\lim_t P(T_t \notin \mathcal{K}) = 0.$$

Since this report presents only asymptotic results, we are interested in "large" t, for which events $[T_t \notin \mathcal{K}]$ are "almost impossible". Therefore in the rest of the report we neglect these events, i.e. in all formulas and arguments we tacitly assume $T_t \in \mathcal{K}$. This considerably simplifies the following text. On the other hand, in each case the eventuality $T_t \notin \mathcal{K}$ can be discussed separately, and an obvious additional effort leads to the extension or standart modification of formulas and arguments valid also under this eventuality. For example, in this sense $\hat{\theta}_t$ is considered to be MLE, as explained already after its definition (3.2).

Lemma 3.2. *If the exponential model $\mathcal{E}_t$ is convergent then the MLE given by (3.2) is consistent, i.e.*

$$\lim_t \|\hat\theta_t - \theta_0\| = 0 \qquad \text{in probability.}$$

Proof. By Lemma 2.1, (3.3) and (3.8)

$$\lim_t \frac{T_t - \dot\kappa_t(\theta_0)}{\gamma_t} = 0 \qquad \text{in probability.}$$

Hence (3.5) implies

$$\lim_t \frac{T_t}{\gamma_t} - \dot\kappa(\theta_0) = 0 \qquad \text{in probability.}$$

It follows from here and from the locally uniform convergence in (3.7) that

$$\dot\kappa_t^{-1}(T_t) = \dot\kappa_t^{-1}\left(\gamma_t\left(\dot\kappa(\theta_0) + \frac{T_t}{\gamma_t} - \dot\kappa(\theta_0)\right)\right)$$

converges in probability to θ_0. □

Lemma 3.3. *For convergent exponential models the MLE (3.2) is asymptotically normal in the sense that*

$$\lim_t \sqrt{\gamma_t}\,(\hat\theta_t - \theta_0) = \mathsf{N}(0, \ddot\kappa(\theta_0)^{-1}) \qquad \text{in distribution.}$$

Proof. By Lemma 3.2 and (3.2)

$$Z_t := \frac{T_t - \mathsf{E}T_t}{\sqrt{\gamma_t}} \qquad \text{and} \qquad Y_t := \frac{\dot\kappa(\hat\theta_t) - \dot\kappa(\theta_0)}{\sqrt{\gamma_t}}$$

coincide. By the mean value theorem there exists $\tilde\theta_t$ on the line joining θ_0 and $\hat\theta_t$ such that

$$Y_t = \frac{\ddot\kappa(\tilde\theta_t)\,(\hat\theta_t - \theta_0)}{\sqrt{\gamma_t}} = \sqrt{\gamma_t}\,\frac{\ddot\kappa(\tilde\theta_t)\,(\hat\theta_t - \theta_0)}{\gamma_t}$$

i.e.

$$\sqrt{\gamma_t}\,(\hat\theta_t - \theta_0) = \gamma_t\,\ddot\kappa(\tilde\theta_t)^{-1}\,Z_t$$

where $\tilde\theta_t$ tends by Lemma 3.2 in probability to θ_0. The desired assertion thus follows from the locally uniform convergence in (3.10) and from (3.9). □

Let k_t be an increasing sequence of natural numbers. We say that an estimator θ_t^* is *consistent of order* k_t if

$$\lim_{\upsilon\to\infty} \limsup_t P(k_t\|\theta_t^* - \theta_0\| > \upsilon) = 0.$$

For example, the MLE $\hat\theta_t$ is under assumptions of Lemma 3.3 consistent of order $k_t = \sqrt{\gamma_t}$.

Lemma 3.4. *Let the models under consideration be convergent and let $\theta_t^1 = \theta_t^1(\boldsymbol{X}_t)$, $\theta_t^2 = \theta_t^2(\boldsymbol{X}_t)$ be two estimators consintent of order k_t. Then for every $r \in \mathbb{R}$ the Rényi divergence $D_r = D_r^{(t)}$ satisfies the asymptotic relation*

$$\lim_t k_t^2\left[\frac{2}{\gamma_t} D_r(\theta_t^1, \theta_t^2) - (\theta_t^1 - \theta_t^2)'\ddot\kappa(\theta_0)(\theta_t^1 - \theta_t^2)\right] = 0 \qquad \text{in probability.}$$

Proof. Let us consider $r \in \mathbb{R} \setminus \{0, 1\}$. From the consistence of estimators and from Lemma 2.2, it follows that $D_r(\theta_t^1, \theta_t^2) = D_r^{(t)}(\theta_t^1, \theta_t^2)$ is given by the upper formula in (2.4), where $\kappa = \kappa_t$ now depends on t, with probability tending to 1.

In the rest of proof we assume that the upper formula in (2.4) holds. By the mean value theorem and Taylor expansion it holds for every $h \in \mathbb{R}^d$

$$\kappa_t(\theta_0 + h) = \kappa_t(\theta_0) + h'\dot\kappa_t(\theta_0) + \frac{1}{2}h'\ddot\kappa_t(\theta_h)h$$

where, here and in the sequel, θ_h denotes a point from the line joining θ_0 and θ_0+h. Then (2.4) implies for $\theta_t^1 = \theta_0 + \alpha_t$ and $\theta_t^2 = \theta_0 + \beta_t$, and for $\xi_t = r\alpha_t + (1-r)\beta_t$

$$\frac{D_r(\theta_t^1, \theta_t^2)}{\gamma_t} = \frac{1}{r(r-1)}\left[\frac{\xi_t'\,\ddot\kappa_t(\theta_{\xi_t})\,\xi_t}{2\gamma_t} - r\frac{\alpha_t'\,\ddot\kappa_t(\theta_{\alpha_t})\,\alpha_t}{2\gamma_t} - (1-r)\frac{\beta_t'\,\ddot\kappa_t(\theta_{\beta_t})\,\beta_t}{2\gamma_t}\right].$$

Let Z_t differs from $D_r(\theta_t^1, \theta_t^2)/\gamma_t$ by replacing all information matrices $\ddot\kappa_t(\cdot)/\gamma_t$ in the last formula by the differences $\ddot\kappa_t(\cdot)/\gamma_t - \ddot\kappa(\theta_0)$. Then

$$\frac{D_r(\theta_t^1, \theta_t^2)}{\gamma_t} = Z_t + \frac{(\alpha_t - \beta_t)'\,\ddot\kappa(\theta_0)\,(\alpha_t - \beta_t)}{2}$$

If $\max\{\|\alpha_t\|, \|\beta_t\|\} \le \varepsilon_t$, and if $\Gamma_t(\varepsilon) = \sup\{\ddot\kappa_t(\cdot)/\gamma_t - \ddot\kappa(\theta_0) \ : \ \|\theta - \theta_0\| \le \varepsilon\}$ then

$$|Z_t| \le \Gamma_t(\varepsilon_t)\frac{\|r\alpha_t + (1-r)\beta_t\|^2 + |r|\|\alpha_t\|^2 + |1-r|\|\beta_t\|^2}{2|r|\,|1-r|}$$

By the Minkowski inequality

$$\|r\alpha_t + (1-r)\beta_t\|^2 \le r^2\|\alpha_t\|^2 + (1-r)^2\|\beta_t\|^2 \le \left(r^2 + (1-r)^2\right)\varepsilon_t^2,$$

so that

$$|Z_t| \le \varepsilon_t^2\Gamma_t(\varepsilon_t)\frac{1}{2}\left(\frac{|r|+1}{|r-1|} + \frac{|r-1|+1}{|r|}\right).$$

If

$$\varepsilon_t = \frac{y}{k_t}$$

for y positive and C_r stands for the constant on the right-hand side of the last upper bound for $|Z_t|$, then

$$k_t^2|Z_t| \le y^2\Gamma_t\left(\frac{y}{k_t}\right)C_r \qquad \text{for all } y \text{ positive.}$$

The locally uniform convergence in (3.6) implies $\lim_t \Gamma_t(y/k_t) = 0$ for all $y > 0$. The consistence of order k_t means that, by selecting sufficiently large $y > 0$, one can keep the probabilities of the event

$$\max\{\|\alpha_t\|, \|\beta_t\|\} \le \frac{y}{k_t}, \quad \text{i.e.} \quad \max\{k_t\|\theta_t^1 - \theta_0\|,\, k_t\|\theta_t^2 - \theta_0\|\} \le y,$$

arbitrarily close to 1 uniformly for all t large enough. By combining these facts one obtains that for every $\delta > 0$ there are $y > 0$ and t_0 such that, with probability arbitrarily close to 1, $t > t_0$ implies $k_t^2|Z_t| < \delta$. The desired assertion is clear from here. □

Example 3.5. I.i.d exponential observations.

Let $\mathcal{T} = \{1, 2, ...\}$ and $\mathcal{E} \equiv (T, \mu)$ be an experiment considered in Section 2 with $T : \mathfrak{X} \to \mathbb{R}$. If $(\mathfrak{X}_t, \mathcal{A}_t) = (\mathfrak{X}^t, \mathcal{A}^t)$ then for $\mathcal{E}_t \equiv (T_t, \mu_t)$ with $\mu_t = \mu^t$ and

$$T_t(\mathbf{X}_t) = \sum_{i=1}^t T(X_i) \qquad \text{for } \mathbf{X}_t = (X_1, ..., X_t)$$

one obtains

$$\kappa_t(\theta) = t\kappa(\theta), \qquad \dot\kappa_t(\theta) = t\dot\kappa(\theta) \quad \text{and} \quad \ddot\kappa_t(\theta) = t\ddot\kappa(\theta).$$

The convergence conditions thus hold for $\gamma_t = t$. The MLE in this case satisfies the relation

$$\hat\theta_t = \dot\kappa^{-1}\left(\frac{1}{t}\sum_{i=1}^t T(X_i)\right), \tag{3.11}$$

provided the arithmetic mean is not in the boundary of range $\dot\kappa$. ○

Example 3.6. Dependent Gaussian observations.

Consider a sequence of real numbers $y_1, y_2, \ldots$ and a Gaussian random sequence $(Z_i : i \geq 1)$ with zero means and regular covariances C_{ij}. Let $P_\theta,\ \theta \in \mathbb{R}$, be probability distributions of observations

$$X_i = \theta y_i + Z_i,\ i \geq 1,$$

on the space $\mathfrak{X} = \mathbb{R} \times \mathbb{R} \times \ldots$. Let $t \in \mathcal{T} = \{1, 2, \ldots\}$, $\mathbf{y}_t = (y_1, \ldots, y_t)$, $\mathbf{Z}_t = (Z_1, \ldots, Z_t)$ and $C_t = (C_{ij})_{1 \leq i,j \leq t}$. Then one easily obtains for the distributions $P_{\theta,t}$ of $\mathbf{X}_t = (X_1, \ldots, X_t)$

$$\frac{dP_{\theta,t}}{dP_{0,t}}(\mathbf{X}_t) = \exp\left\{-\frac{1}{2}\left[(\mathbf{X}_t - \theta\mathbf{y}_t)'\, C_t^{-1}\,(\mathbf{X}_t - \theta\mathbf{y}_t) - \mathbf{X}_t'\, C_t^{-1}\, \mathbf{X}_t\right]\right\}.$$

Thus these models are exponential with the observation spaces $\mathfrak{X}_t = \mathbb{R}^t$, $\mu_t = P_{0,t}$ and

$$\begin{aligned} T_t = T_t(\mathbf{X}_t) &= \mathbf{y}_t'\, C_t^{-1}\, \mathbf{X}_t \\ \kappa_t(\theta) &= \frac{\theta^2}{2}\mathbf{y}_t'\, C_t^{-1}\, \mathbf{y}_t \end{aligned}$$

Let

$$\rho_t := \mathbf{y}_t'\, C_t^{-1}\, \mathbf{y}_t \ \to \infty \qquad \text{as } t \to \infty.$$

Then

$$\dot{\kappa}_t(\theta) = \theta\, \rho_t \qquad \text{and} \qquad \ddot{\kappa}_t(\theta) = \rho_t,$$

and strong regularity conditions hold for $\gamma_t = \rho_t$, $\kappa(\theta) = \theta^2/2$, $\dot{\kappa}(\theta) = \theta$ and $\ddot{\kappa}(\theta) = 1$. The MLE is in this case given by the formula

$$\hat{\theta}_t = \frac{T_t}{\rho_t} = \frac{\mathbf{y}_t' C_t^{-1} \mathbf{X}_t}{\rho_t}.$$

If $\mathbf{X}_t$ is distributed by $P_{\theta_0,t}$ then

$$T_t = \theta_0 \rho_t + \mathbf{y}_t'\, C_t^{-1}\, \mathbf{Z}_t$$

where $\mathbf{y}_t'\, C_t^{-1}\, \mathbf{Z}_t \sim \mathsf{N}(0, \rho_t)$. It follows from here that the models $\mathcal{E}_t$ are convergent with $\gamma_t = \rho_t$. ∘

Example 3.7. Diffusion processes.

Let $\mathfrak{X} = C[0, \infty)$ and let $P_\theta, \theta \in \mathbb{R}$, be distributions of diffusion processes $(X_s : s \geq 0)$ defined by the stochastic differential equation

$$dX_s = \theta\, a(s) ds + dW_s, \quad s \geq 0,$$

where $a : [0, \infty) \to \mathbb{R}$ is continuous and W_s is a standard Wiener process. For each $t \in \mathbb{R}^+$ and for the distributions $P_{\theta,t}$ of segments $\mathbf{X}_t = (X_s : 0 \leq s \leq t)$ one obtains

$$\frac{dP_{\theta,t}}{dP_{0,t}}(\mathbf{X}_t) = \exp\left\{\theta \int_0^t a(s) dX_s - \frac{\theta^2}{2}\int_0^t a^2(s) ds\right\}.$$

Thus the models with

$$T_t = T_t(\mathbf{X}_t) = \int_0^t a(s) dX_s, \qquad \mu_t = P_{0,t} \qquad \text{and} \qquad \kappa_t(\theta) = \frac{\theta^2}{2}\int_0^t a^2(s) ds$$

are exponential. The situation is similar as in the previous example. If we put $\rho_t = \int_0^t a^2(s) ds$ then we obtain the same formulas for $\dot{\kappa}_t$ and $\ddot{\kappa}_t$ as before. Also if $\lim_t \rho_t = \infty$ then the convergences considered in (3.6) and (3.7) take place for the same $\gamma_t, \kappa, \dot{\kappa}$ and $\ddot{\kappa}$ as before. The MLE is given by the formula

$$\hat{\theta}_t = \frac{T_t}{\rho_t} = \frac{\int_0^t a(s) dX_s}{\rho_t}.$$

If $\mathbf{X}_t$ is distributed by $P_{\theta_0,t}$ then

$$T_t = \theta_0 \rho_t + \int_0^t a(s) dW_s$$

where $\int_0^t a(s)dW_s \sim \mathsf{N}(0, \rho_t)$. Since $\int_0^t a(s)dW_s/\rho_t$ tends in probability to zero, (3.8) and (3.9) hold. So the experiments $\mathcal{E}_t$ are convergent with $\gamma_t = \rho_t$. ∘

Example 3.8. Poisson processes.

Let us consider $\mathcal{T} = [0, \infty)$ and let $\mathbf{X} = (X_s : s \geq 0)$ be the Poisson process with $X_0 = 0$, intensity e^θ for $\theta \in \Theta = \mathbb{R}$. Let P_θ be the probability distribution of $\mathbf{X}$ on the Skorokhod space $(\mathfrak{X}, \mathcal{A})$ of realisations of this process. Finally, consider the right-continuous filtration $(\mathcal{A}_t)$ generated by $\mathbf{X}$ and denote by $P_{\theta,t}$ the restriction of P_θ on $\mathcal{A}_t$. Then the distribution density of the observations $\mathbf{X}_t = (X_s : 0 \leq s \leq t)$ can be specified by the formula

$$\frac{dP_{\theta,t}}{dP_{0,t}} = \exp\left\{\theta\, X_t - t(e^\theta - 1)\right\}$$

(see e.g. [GS75]). Hence we face here the system of exponential experiments $\mathcal{E}_t \equiv (T_t, P_{0,t}), t \in \mathcal{T}$, with

$$T_t(\mathbf{X}_t) = X_t \qquad \text{and} \qquad \kappa_t(\theta) = t(e^\theta - 1).$$

It is easy to see that

$$\hat{\theta}_t = \hat{\theta}_t(\mathbf{X}_t) = \ln\left(\frac{X_t}{t}\right)$$

is the MLE of the true parameter $\theta_0 \in \mathbb{R}$. If $\mathbf{X}_t$ is distributed by $P_{\theta_0,t}$ then

$$X_t \sim \mathsf{Po}(e^{\theta_0} t).$$

It follows from here that (3.6) - (3.9) hold and the experiments $\mathcal{E}_t$ are convergent with $\gamma_t = t$. ∘

4. Simple hypotheses

The *hypotheses testing problem* studied in this section can be formulated as follows. Let the observed data $\mathbf{X}_t$ be distributed by a density $f_{\theta_0,t}$ from an exponential family $\mathcal{E}_t \equiv (T_t, \mu_t)$ and let the generalised sequence $\mathcal{E}_t, t \in \mathcal{T}$, be convergent in the sense of Section 3. Let the tested null hypothesis H_0 be specified as a subset Θ_0 of the parameter space Θ common for all families $\mathcal{E}_t$. The problem is to find a generalised sequence of test statistics

$$S_t = S_t(\mathbf{X}_t) : \mathfrak{X}_t \to \mathbb{R}$$

with a known asymptotic distribution F on $\mathbb{R}$, identical for all hypothetical values $\theta_0 \in \Theta_0$.

In this section we solve the problem of testing simple hypotheses $H_0 \equiv \{\theta_0\}$ by using the collection of *Rényi statistics*

$$D_{r,t} = 2D_r^{(t)}(\hat{\theta}_t, \theta_0) = 2\gamma_t D_r(\hat{\theta}_t, \theta_0) \tag{4.1}$$

where $D_r^{(t)}(\hat{\theta}_t, \theta_0)$ are the Rényi divergences, given by (2.4) and (2.5) with $\kappa = \kappa_t$ depending now on t, between the most likely distribution $P_{\hat{\theta}_t,t}$ defined by (3.2) and the hypothetical distribution $P_{\theta_0,t}$. The third term in (4.1) shows that the Rényi statistics can be expressed also using the parameter γ_t of convergence figuring in (3.6) - (3.10) where $D_r(\cdot,\cdot)$ are given by (2.4) and (2.5) with κ the limit function presented in (3.4).

In the following theorem, χ^2_k stands for the χ^2-*distributed* random variable with k degrees of freedom.

Theorem 4.1. *Let $\mathcal{E}_t, t \in \mathcal{T}$, be convergent. Then under any simple hypothesis $H_0 \equiv \{\theta_0\} \subset \Theta$, where $\Theta \subseteq \mathbb{R}^d$, all Rényi statistics (4.1) converge in distribution to χ^2_d.*

Proof. By Lemma 3.3, $\hat{\theta}_t$ is consistent of order $\sqrt{\gamma_t}$. The "estimator" $\tilde{\theta}_t := \theta_0$ is consistent of the same order. Therefore, in view of Lemma 3.4, it is sufficient to prove that under H_0

$$Y_t := \gamma_t(\hat{\theta}_t - \theta_0)' \, \ddot{\kappa}(\theta_0) \, (\hat{\theta}_t - \theta_0)$$

converges in distribution to χ^2_d.

By Lemma 3.3, $\sqrt{\gamma_t}(\hat{\theta}_t - \theta_0)$ converges under H_0 in distribution to $\mathsf{N}(0, \ddot{\kappa}(\theta_0)^{-1})$. This means that

$$\lim_{t\to\infty} Y_t = \sum_{i=1}^{d} Z_i \qquad \text{in distribution,}$$

where Z_i are independent and χ^2_1-distributed random variables. □

Theorem 4.1 provides a continuum of asymptotically α-level tests of the hypothesis $H_0 \equiv \{\theta_0\}$. The test with statistic $D_{r,t}$ and critical value c_α equal to the $(1-\alpha)$-quantile of χ^2_d for $\alpha \in (0,1)$ will be called *Rényi (r, α)-test*. If the asymptotic size α is fixed in advance then we speak simply about *Rényi r-test*.

Remark 4.2. *The well known likelihood ratio test of $H_0 \equiv \{\theta_0\}$ uses the generalised likelihood ratio statistic*

$$Q_t = Q_t(\mathbf{X}_t) = -2 \ln \frac{f_{\theta_0,t}(\mathbf{X}_t)}{f_{\hat{\theta}_t,t}(\mathbf{X}_t)},$$

where $\hat{\theta}_t = \hat{\theta}_t(\mathbf{X}_t)$ is the MLE of θ_0. Using the explicit formulas for densities $f_{\theta_0,t}$ and $f_{\hat{\theta}_t,t}$ from $\mathcal{E}_t$ we obtain that

$$Q_t = 2\left[\kappa_t(\theta_0) - \kappa_t(\hat{\theta}_t) + T_t'(\hat{\theta}_t - \theta_0)\right].$$

Since $T_t = \dot{\kappa}(\hat{\theta}_t)$ it follows from (4.1) and from formula (2.5) for $D_1^{(t)}(\hat{\theta}_t, \theta_0)$ that

$$Q_t = D_{1,t}.$$

This means that for any $\alpha \in (0,1)$ the Rényi $(1,\alpha)$-test coincides with the asymptotically α-level generalised likelihood ratio test. Note that in accordance with simplification which we formed under definition 3.1, we assume here for simplicity that $T_t = T_t(\mathbf{X}_t)$ with $\mathbf{X}_t \sim f_{\theta_0,t}$ takes on the values in set $\mathcal{K}$ considered in (3.3), the probability that T_t is not in $\mathcal{K}$ is negligible for $t \in \mathcal{T}$ of our interest.

The choice of the most appropriate r-test must be based on additional optimality criteria, e.g. on the test powers under local alternatives. Below we introduce an approach based on calculation and comparison of test powers for alternatives from a neighbourhood of H_0.

5. Algorithm for a simple hypothesis and independent samples

Let us consider a statistical model with independent identically distributed random variables $X_1, ..., X_n$. Assume that the distribution of X_1 is from an exponential family $\mathcal{E} \equiv (T, \mu)$ how they were introduced in Section 2 with $\mathfrak{X} = \mathbb{R}^k$ or $\mathfrak{X} = \mathbb{N}^k$ (k stands for the dimension of random value X_1). Recall that for this case

$T : \mathfrak{X} \to \mathbb{R}^d$ and μ is a measure on $\mathfrak{X}$,

$\kappa : \mathbb{R}^d \to \mathbb{R}, \quad \kappa(\theta) = \ln \int_{\mathfrak{X}} \exp\{\theta' \, T(x)\} \, \mu(dx)$ for every $\theta \in \Theta \subseteq \mathbb{R}^d$.

We can also consider whole observation $\mathbf{X}_n = (X_1, ..., X_n)'$ as one vector and use the notation from Example 3.5 for a sequence of exponential families $\mathcal{E}_n$.

Our aim is to design an algorithm which, for a data coming from the described model $X_1, ..., X_n$, makes the following statistical decisions:

1. to estimate parameter θ
2. to test a simple hypothesis $H_0 \equiv \{\theta_0\}$ against the double-sided alternative
3. to identify the model; i.e. to choose the most fitting one from a given finite set of possible parameters.

The first task consists simply in finding the MLE $\hat{\theta}_n$ which we obtain as the solution of the following equation

$$\dot{\kappa}(\theta) = \frac{1}{n}\sum_{i=1}^{n} T(X_i), \tag{5.1}$$

c.f. (3.11).

The main task (the second one) is the construction of Rényi (r, α)- tests when an asymptotic test level $\alpha \in (0, 1)$, a finite set $\mathcal{R}$ of Rényi orders and a hypothetical value $\theta_0 \in \Theta$ are given. This task also includes an optimisation among these (r, α)-tests with respect to $r \in \mathcal{R}$ for a fixed level α. We have chosen an approach based on a comparison of exact test powers which are calculated for a representative set of alternatives Θ_1 from a neighbourhood of $H_0 \equiv \{\theta_0\}$. Let us denote by

$$\begin{aligned}
\mathbf{a} &= \mathbf{a}(\alpha, r, \theta_0, n) = P_{\theta_0,n}(D_{r,n} > c_\alpha), \\
\mathbf{b} &= \Big(\mathbf{b}(\alpha, r, \theta, n) : \theta \in \Theta_1\Big) = \Big(P_{\theta,n}(D_{r,n} > c_\alpha) : \theta \in \Theta_1\Big)
\end{aligned}$$

exact test levels and exact test powers, respectively, where c_α stands for the critical value (equal to the $(1-\alpha)$-quantile of χ^2_d) and $D_{r,n}$ are the appropriate Rényi statistics (4.1). Either we are able to find values $\mathbf{a}, \mathbf{b}$ for the given distribution or we numerically evaluate $\mathbf{a}, \mathbf{b}$ by using statistical simulations (Monte Carlo method). If we choose set of alternatives Θ_1 of size m and we set

$$\Theta_1 = \{\theta_1, ..., \theta_m\}$$

then we have for each order $r \in \mathcal{R}$ a vector $\big(\mathbf{a}(r), \mathbf{b}(r, \theta_1), ..., \mathbf{b}(r, \theta_m)\big)$ to compare. We employ the *relative inefficiency method* which says:

- compute relative inefficiency $\eta(r, \theta_i) = \sup_{r^* \in \mathcal{R}} \mathbf{b}(r^*, \theta_i) - \mathbf{b}(r, \theta_i)$ for each $r \in \mathcal{R}$ and for each $1 \leq i \leq m$
- then compute the maximum over alternatives from Θ_1, $\eta(r) = \sup_{1 \leq i \leq m} \eta(r, \theta_i)$ for each $r \in \mathcal{R}$,
- and, finally, find the value r^* which minimises $\{\eta(r) : r \in \mathcal{R}\}$.

On the basis of this optimisation we select the Rényi (r^*, α)-test which we use for the final decision:

we reject hypothesis $H_0 \equiv \{\theta_0\}$ on the asymptotic level α if $D_{r^*,n} > c_\alpha$.

The third task is independent on the second one. It represents a classification problem since it assumes that we have an a priori knowledge about the real parameter, particularly assumes that the true parameter is from a set

$$\Theta_2 = \{\theta_1, ...\theta_e\}.$$

We can compute for every $r \in \mathcal{R}$ and every $\theta_i \in \Theta_2$ the Rényi divergence $D_r(\hat{\theta}_n, \theta_i)$, given by (2.4), between the most likely distribution using MLE $\hat{\theta}_n$ and the hypothetical distribution with parameter θ_i. Let us choose, for each fixed order $r \in \mathcal{R}$, the

Rényi divergence with the minimal value. It means, we obtain a mapping

$$\begin{array}{rccc} s: & r & \mapsto & i \\ & \mathcal{R} & \rightarrow & 1,...,e \end{array} \tag{5.2}$$

selecting for each order of Rényi divergences one parameter from Θ_2. If a majority of Rényi divergences will be minimal for the same one of parameters from Θ_2, i.e.

$$\text{there exists } i \text{ such that } \#\{r \in \mathcal{R} : s(r) = i\} > \#\{r \in \mathcal{R} : s(r) \neq i\}, \tag{5.3}$$

then we decide that

the date come from the model with parameter θ_i.

Our algorithm solve the described tasks for the following list of distributions:

1 - Bernoulli
2 - Binomial
3 - Poisson
4 - Geometrical
5 - Negative binomial
6 - Multinomial
7 - Exponential
8 - Gaussian
9 - Gaussian mean-known
10 - Gaussian variance-known
11 - Gamma
12 - Gamma p-known
13 - Beta
14 - Lognormal
15 - Double exponential
16 - Weibull
17 - Reyliegh
18 - Maxwell
19 - Pareto
20 - Modular
21 - Inverse Gaussian
22 - Inverse gamma
23 - χ^2_k
24 - Dirichlet
25 - Bivariate Gaussian.

See Appendix for a list of these distributions with a detailed description of their parameters and their exponential representations.

In the rest of this section we introduce successively

* constants
* input variables
* derived variables
* calculation
* outputs

of the algorithm.

CONSTANTS

$\mathsf{l} = 11$	...	the number of used Rényi orders;
$\mathsf{R} = (-2; -1.5; -1; \ldots; 2; 2.5; 3)$	...	the vector of used Rényi orders, size: $1 \times \mathsf{l}$.

INPUT VARIABLES

x	...	the data, size: $\mathsf{k} \times \mathsf{n}$ for distributions $6, 24, 25$, $1 \times \mathsf{n}$ otherwise;
dis	...	the type of distribution, dis $\in \{1, ..., 25\}$, see the list above;
p0	...	the hypothetical parameter when the density is considered in its original form (see Appendix), size: $\mathsf{d} \times 1$;
α	...	the asymptotic test level, $\alpha \in (0, 1)$, the default value is 0.05;
p1	...	the user-choice of alternatives from a neighbourhood of p0 (acts in the calculation of test powers), size: $\mathsf{d} \times \mathsf{m}$, for default values see h1;
p2	...	parameters for the identification of model, size: $\mathsf{d} \times \mathsf{e}$, (again consider the density in its original form);
next	...	for some of distributions (dis $\in \{2, 5, 6, 9, 10, 12, 16, 19\}$) a known value of a next parameter is needed, size: 1×1.

DERIVED VARIABLES

n	...	the size of data;
k	...	the dimension of data, k is arbitrary positive natural for dis $= 6$ or 24, k $= 2$ for dis $= 25$ and k $= 1$ otherwise;
m	...	the number of alternatives for which test powers are calculated, the default value is 20;
e	...	the number of parameters from which the model is identified;
$\tilde{\mathsf{m}}$	...	m $+ 1$, the number of alternatives plus hypothesis;
h0	...	the hypothetical parameter θ_0 if the density in its exponential form is considered, a transformation (see Appendix) of p0, size: d $\times$ 1;
h	...	the borders of interval Θ of possible values for parameter θ of the given distribution, size: d $\times$ 2;
h1	...	the choice of alternatives, h1 is either a transformation of p1 or a default choice obtained as a function of dis, h0 and h, size: d $\times$ m;
h2	...	a transformation of parameters p2 into the form of exponential density, size: d $\times$ e;
hh	...	alternatives h1 together with hypothetical parameter h0, size: d $\times$ $\tilde{\mathsf{m}}$.

CALCULATION

At this moment, when a distribution dis is given, we need to know a particular form of appropriate functions $T, \kappa, \dot{\kappa}$ which appear in the exponential density of the given distribution and also a particular transformation function π which transforms parameters of the density in its exponential form into parameters of the original density. It means we need a mapping which for each value dis picks up these appropriate functions. In the algorithm, there are used following algorithm–functions. An input variable each of them is surely variable dis and therefore it is no need to emphasise this fact in what follows.

BASIC FUNCTIONS

function TE**:** input has size k $\times$ n, output has size d $\times$ n,
purpose: on each column of input (of data nature) function T is applied and the result is put down in the appropriate column of the output;

function KAPPA**:** input has size d $\times$ w, output has size $1 \times$ w, w $\in \mathbb{N}$,
purpose: on each column of input (of parameter nature) function κ is applied and the result is put down in the appropriate column of the output;

function DKAPPA**:** input has size d $\times$ w, output has size d $\times$ w, w $\in \mathbb{N}$,
purpose: on each column of input (of parameter nature) function $\dot{\kappa}$ is applied and the result is put down in the appropriate column of the output;

function PI**:** input has size d $\times$ w, output has size d $\times$ w, w $\in \mathbb{N}$,
purpose: on each column of input (of parameter nature) function π is applied and the result is put down in the appropriate column of the output;

function INVPI**:** input has size d $\times$ w, output has size d $\times$ w, w $\in \mathbb{N}$,
purpose: on each column of input (of parameter nature) function π^{-1} is applied and the result is put down in the appropriate column of the output;

function MEAN**:** input has size k $\times$ n, output has size d $\times$ 1,
purpose: firstly, function TE is applied and then, for the result of d $\times$ n size, a mean over each row is calculated, i.e. $\mathsf{MEAN}_i = \frac{1}{\mathsf{n}} \sum_{j=1}^{\mathsf{n}} \mathsf{TE}_{ij}$

function MEAN2**:** input has size k $\times$ (n $\cdot$ w), output has size d $\times$ w,
purpose: on each (i-th) n-tuple of columns the function MEAN is applied and the result is put down in i-th column of the output, for any $1 \leq i \leq$ w;

function MLE**:** input has size $\mathsf{d} \times \mathsf{w}$, output has size $\mathsf{d} \times \mathsf{w}$, $\mathsf{w} \in \mathbb{N}$,
purpose: on each column of input (typically, it is output of function MEAN) the inverse of function $\dot{\kappa}$ is applied (or analogical equation (5.1) is solved) and the result is put down in the appropriate column of the output.

FUNCTION OF RÉNYI STATISTICS

function DE

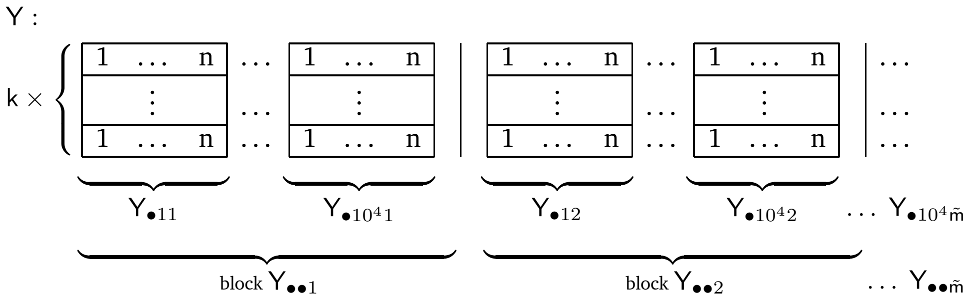
input: two matrices $\mathsf{U}^1, \mathsf{U}^2$, each has size $\mathsf{d} \times \mathsf{w}$ (typically, matrix U^1 is an output of function MLE), it is also allowed that one of the matrices has size $\mathsf{d} \times 1$ although $\mathsf{w} > 1$. In this case the missing columns are automatically filled by copies of this only column to reach $\mathsf{d} \times \mathsf{w}$ size,

output: has size $\mathsf{w} \times \mathsf{l}$,

purpose: each (i-th) row of the output is a vector of Rényi statistics $(D_{r,n} : r \in \mathsf{R})$ for i-th column of matrix U^1 and i-th column of matrix U^2, i.e.
DE_{ij} is Rényi statistic $D_{r,n}(\mathsf{U}^1_{:i}, \mathsf{U}^2_{:i})$ of order $r = \mathsf{R}_j$ given by formula (4.1).

FUNCTIONS FOR OPTIMISATION

Let us fix the value of variable dis which indicates the given distribution and fix asymptotic level α. In order to evaluate test powers of Rényi (r, α)-tests under local alternatives h1 we use Monte Carlo simulation method which works very universally. It is based on a simulation of a big data file from the distribution dis. Let us put these simulated values into the following special-formed matrix Y of size $\mathsf{k} \times (\mathsf{n} \cdot 10^4 \cdot \tilde{\mathsf{m}})$ by analogy with the size of the real-input data



where each 'block' $\mathsf{Y}_{\bullet\bullet i}$, for arbitrary $1 \le i \le \tilde{\mathsf{m}}$, containts in its columns independent samples from distribution dis with a fixed parameter given by i-th column of matrix hh. Furthemore, each block is divided into 'packages' $\mathsf{Y}_{\bullet ji}$, for $1 \le j \le 10^4$, where one package means n samples.

We have chosen the number 10^4 of simulated packages per block in order to obtain a given accuracy. We mean the accuracy of an evaluation of probability $P(f(Z) > c)$ where Z is a random variables, f is a real function on the values of Z and c is a constant. We estimate this probability by

$$\hat{p} = \frac{\#\{1 \le i \le 10^4 : f(z_i) > c\}}{10^4}.$$

where $z_1, ..., z_{10^4}$ are simulated values from the distribution of Z. It follows from the central limit theorem for independent Bernoulli samples that size 10^4 is sufficient to guarantee that the error which we make using $\hat{p}$ instead of exact probability $P(f(Z) > c)$ is less than 0.01 with assurance 95%.

In our case, we are interested in probabilities $P_i = P_{\theta_i,n}(D_{r,n}(\hat{\theta}_n, \theta_0) > c_\alpha)$ for any $1 \le i \le \tilde{\mathsf{m}}$ and we estimate them by

$$\hat{p}_i = \frac{\#\{1 \le j \le 10^4 : D_{r,n}(\text{“}\mathsf{Y}_{\bullet ji}\text{”}, \mathsf{h0}) > c_\alpha\}}{10^4},$$

using simulated values in matrix Y. Here $D_{r,n}(\text{“}\mathsf{Y}_{\bullet ji}\text{”}, \mathsf{h0})$ means the Rényi statistic of order r for an empirical distribution derived from the simulated data given by block $\mathsf{Y}_{\bullet ji}$ of matrix Y (note that size of this block is the same as observed data size) and for the hypothetical distribution with parameter h0. Using algorithm-functions

$$D_{r,n}(\text{“}\mathsf{Y}_{\bullet ji}\text{”}, \mathsf{h0}) = \mathsf{DE}\Big(\mathsf{MLE}\big(\mathsf{MEAN}(\mathsf{Y}_{\bullet \mathsf{ji}})\big), \mathsf{h0}\Big)$$

for every $1 \le j \le 10^4$ and every $1 \le i \le \tilde{\mathsf{m}}$.

We can apply algorithm-functions on whole matrix Y directly. Then each (j-th) column of matrix

$$\mathsf{DE}\Big(\ \mathsf{MLE}\big(\ \mathsf{MEAN2}(\mathsf{Y})\ \big),\ \mathsf{h0}\ \Big) \tag{5.4}$$

corresponds to the Rényi statistic of ‘j-th’ order, i.e. order R_j, and each (i-th) 10^4-tuple of rows is used for calculation of estimate $\hat{p}_i$ of probability $P_{\theta_i,n}(D_{r,n}(\hat{\theta}_n, \theta_0) > c_\alpha)$ where $r = \mathsf{R}_j$, $n = \mathsf{n}$ and $\theta_i = \mathsf{hh}_i$. In this context, let us introduce

function POWER
- input: has size $(10^4 \cdot \hat{\mathsf{m}}) \times \mathsf{l}$, it is typically the output of function DE applied on Y as described in (5.4),
- output: has size $\hat{\mathsf{m}} \times \mathsf{l}$,
- *purpose:* POWER_{ij} is an estimate of probability $P_{\mathsf{hh}_i,\mathsf{n}}(D_{\mathsf{R}_j,\mathsf{n}} > c_\alpha)$;

function OPT
- input: has size $\hat{\mathsf{m}} \times \mathsf{l}$, it is typically the output of function POWER,
- output: is scalar from set $\{1, ..., \mathsf{l}\}$,
- *purpose:* it indicates a ‘winning’ order of Rényi statistics following from the minimax optimisation described in the previous (page 11).

DECISION FUNCTIONS

function TEST_HYP: Inputs of this function are hypothetical parameter h0, data x and the output of function OPT. Output of this function is one of the decisions:

REJECT hypothesis H_0 on level α	if $\mathsf{DE}\big(\mathsf{MLE}(\mathsf{MEAN}(\mathsf{x})), \mathsf{h0}\big)_{\mathsf{OPT}} > c_\alpha$,
DON'T REJECT hypothesis H_0 on level α	if $\mathsf{DE}\big(\mathsf{MLE}(\mathsf{MEAN}(\mathsf{x})), \mathsf{h0}\big)_{\mathsf{OPT}} \le c_\alpha$.

function IDENTIF: Inputs of this function are data x and collection h2 of the only possible parameters of the given distribution dis. Output of this function is the most suitable parameter from this collection or the decision that the identification is impossible. In the first step we calculate Rényi divergences for the empirical distribution derived from data x and for different hypothetical distributions where the appropriate parameters arise in columns of h2. Using algorithm-functions, let us set

$$\mathsf{TAB} = \mathsf{DE}\Big(\mathsf{MLE}\big(\mathsf{MEAN}(\mathsf{x})\big), \mathsf{h2}\Big)$$

a matrix of size $\mathsf{e} \times \mathsf{l}$. Then, for each Rényi order $r \in \mathsf{R}$ (appropriate column of TAB), we compare e values of divergences (in rows of TAB) and choose the parameter (i.e. index of the row) corresponding to the minimal divergence.

Thus we obtain a vector $(1 \times \mathsf{I})$ of naturals from $\{1, \ldots, \mathsf{e}\}$, cf. function s in (5.2). If the frequency of the most frequently chosen parameter has absolute majority, i.e. (5.3) holds, then output is equal to this parameter, otherwise we can not identify the model. Let this frequency is the second output of function IDENTIF.

OUTPUTS

For task 1.: The maximum likelihood estimate of parameter is

$\mathsf{MLE}\big(\mathsf{MEAN}(\mathsf{x})\big)$	when the exponential form of density is considered,
$\mathsf{PI}\big(\mathsf{MEAN}(\mathsf{MLE}(\mathsf{x}))\big)$	when the original form of density is considered.

For task 2.: The test based on the Rényi divergence of order OPT was chosen among other tests based on Rényi divergences of orders $r \in \mathsf{R}$ as the most optimal. Depending on Boolean value of function TEST_HYP, this test $\left\{ \begin{array}{l} \text{REJECTS} \\ \text{DOES NOT REJECT} \end{array} \right\}$ hypothesis that the parameter is equal to h0 on the asymptotic level α. See also the table POWER of evaluated test powers and a test level.

For task 3.: By comparison of Rényi divergences computed for the empirical distribution and different hypothetical distributions corresponding to parameters from the given set $\Theta_2 \equiv \mathsf{h2}$ we can decide that

$(+)$ data comes from the model with parameter IDENTIF. For this parameter Rényi divergences were minimal in $\mathsf{IDENTIF}(2)$ cases from I;

$(-)$ since there is no majority of Rényi divergences choosing the same parameter we can not decide for one of these parameters.

Here $(+)$ and $(-)$ are two disjoint variants of output for the third task.

Appendix

A list of (supported) distributions of a real random variable with an exponential form of the density is presented in this appendix. For each distribution the following characteristics are prescribed:

- state space $\mathfrak{X}$ of the random variable;
- an original (it means commonly used) formula for the density with an (original) notation of parameters; in the algorithm-part, we denote by p the unknown parameter and by next the known parameter if any;
- an exponential form of this density with unknown (d-dimensional) parameter denoted in this report by θ (in the algorithm-part by h·);
- function π which provides the transformation of 'exponential' parameters into 'original' parameters and vice versa (π^{-1});
- boundaries of convex set Θ of possible values for 'exponential' parameter, written into a $d \times 2$ matrix;
- function T which appears in the exponential form of the density;
- cumulant function κ and its derivative;
- the explicit formula for the maximum likelihood estimator MLE or an appropriate equation which uniquely determines the MLE.

1. Bernoulli distribution

$\mathfrak{X} = \{0, 1\}$;
original density (with respect to the counting measure on $\mathfrak{X}$):

$$f(x) = p^x(1-p)^{1-x};$$

its exponential representation:

$$f_\theta(x) = \frac{1}{1+e^\theta}\, e^{\theta x}$$

with respect to measure $\mu: \; x \in \mathfrak{X} \longmapsto 1$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \frac{e^\theta}{1+e^\theta} \qquad \& \qquad \pi^{-1}(p) = \ln\left(\frac{p}{1-p}\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} -\infty & \infty \end{pmatrix}$;
$\mathrm{T}(x) = x$
$\kappa(\theta) = \ln(1+e^\theta)$

$$\dot{\kappa}(\theta) = \frac{e^\theta}{1+e^\theta} \qquad \& \qquad \mathrm{MLE} = \ln\left(\frac{\sum_{i=1}^n X_i}{n - \sum_{i=1}^n X_i}\right).$$

2. Binomial distribution

known parameter: $\tilde{n} \in \mathbb{N}^+$;
$\mathfrak{X} = \{0, 1, ..., \tilde{n}\}$;
original density (with respect to the counting measure on $\mathfrak{X}$):

$$f(x) = \binom{\tilde{n}}{x} p^x(1-p)^{\tilde{n}-x};$$

its exponential representation:

$$f_\theta(x) = \frac{1}{(1+e^\theta)^{\tilde{n}}}\, e^{\theta x}$$

with respect to measure $\mu: \; x \in \mathfrak{X} \longmapsto \binom{\tilde{n}}{x}$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \frac{e^\theta}{1+e^\theta} \qquad \& \qquad \pi^{-1}(p) = \ln\left(\frac{p}{1-p}\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} -\infty & \infty \end{pmatrix}$;
$\mathrm{T}(x) = x$
$\kappa(\theta) = \tilde{n}\ln(1+e^\theta)$

$$\dot{\kappa}(\theta) = \frac{\tilde{n}e^\theta}{1+e^\theta} \qquad \& \qquad \mathrm{MLE} = \ln\left(\frac{\sum_{i=1}^n X_i}{n\tilde{n} - \sum_{i=1}^n X_i}\right).$$

3. Poisson distribution

$\mathfrak{X} = \mathbb{N}$;
original density (with respect to the counting measure on $\mathfrak{X}$):

$$f(x) = \frac{\lambda^x}{x!}\, e^{-\lambda};$$

its exponential representation:

$$f_\theta(x) = \frac{1}{\exp(e^\theta)}\, e^{\theta x}$$

with respect to measure $\mu:\ x \in \mathfrak{X} \longmapsto \frac{1}{x!}$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = e^\theta \qquad \& \qquad \pi^{-1}(\lambda) = \ln(\lambda);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} -\infty & \infty \end{pmatrix}$;
$\mathrm{T}(x) = x$
$\kappa(\theta) = e^\theta$

$$\dot{\kappa}(\theta) = e^\theta \qquad \& \qquad \mathrm{MLE} = \ln\left(\frac{1}{n}\sum_{i=1}^n X_i\right).$$

4. Geometrical distribution

$\mathfrak{X} = \mathbb{N}$;
original density (with respect to the counting measure on $\mathfrak{X}$):

$$f(x) = p(1-p)^x;$$

its exponential representation:

$$f_\theta(x) = (1 - e^{-\theta})e^{-\theta x}$$

with respect to measure $\mu:\ x \in \mathfrak{X} \longmapsto 1$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = 1 - e^{-\theta} \qquad \& \qquad \pi^{-1}(p) = \ln\left(\frac{1}{1-p}\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = -x$
$\kappa(\theta) = -\ln(1 - e^{-\theta})$

$$\dot{\kappa}(\theta) = \frac{-e^{-\theta}}{1 - e^{-\theta}} \qquad \& \qquad \mathrm{MLE} = \ln\left(\frac{\sum_{i=1}^n X_i + n}{\sum_{i=1}^n X_i}\right).$$

5. Negative binomial distribution

known parameter: $r \in \mathbb{N}^+$;
$\mathfrak{X} = \mathbb{N}$;

original density (with respect to the counting measure on $\mathfrak{X}$):

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x;$$

its exponential representation:

$$f_\theta(x) = (1-e^{-\theta})^r \; e^{-\theta x}$$

with respect to measure $\mu: \; x \in \mathfrak{X} \longmapsto \binom{r+x-1}{x}$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = 1 - e^{-\theta} \qquad \& \qquad \pi^{-1}(p) = \ln\left(\frac{1}{1-p}\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = -x$
$\kappa(\theta) = -r\ln(1-e^{-\theta})$

$$\dot\kappa(\theta) = \frac{-re^{-\theta}}{1-e^{-\theta}} \qquad \& \qquad \mathrm{MLE} = \ln\left(\frac{\sum\limits_{i=1}^{n} X_i + rn}{\sum\limits_{i=1}^{n} X_i}\right).$$

6. Multinomial distribution

known parameter: $\tilde n \in \mathbb{N}^+$;
$\mathfrak{X} = \{\mathbf{x} \in \{0,1,...,\tilde n\}^{k+1} : x_1 + ... + x_{k+1} = \tilde n\} = \{0,1,...,\tilde n\}^k \times \{\tilde n - x_1 - ... - x_k\}$;
original density (with respect to the counting measure on $\{0,1,...,\tilde n\}^k$):

$$f(\mathbf{x}) = \frac{\tilde n!}{x_1!...x_k!(\tilde n - x_1 - ... - x_k)!} \, p_1^{x_1} \; ... \; p_k^{x_k} (1 - p_1 - ... - p_k)^{\tilde n - x_1 - ... - x_k};$$

its exponential representation:

$$f_\theta(\mathbf{x}) = \frac{\exp(\theta_1 x_1 + ... + \theta_k x_k)}{(1 + e^{\theta_1} + ... + e^{\theta_k})^{\tilde n}}$$

with respect to measure $\mu: \; \mathbf{x} \in \{0,1,...,\tilde n\}^k \longmapsto \frac{\tilde n!}{x_1!...x_k!(\tilde n - x_1 - ... - x_k)!}$;
dimension of the parameter: $d = k$;
parametrisation:

$$\pi(\theta) = \pi((\theta_1,...,\theta_k)) = \left(\frac{e^{\theta_1}}{1+e^{\theta_1}+...+e^{\theta_k}}, \; ..., \; \frac{e^{\theta_k}}{1+e^{\theta_1}+...+e^{\theta_k}}\right)$$

$$\pi^{-1}(p_1,...,p_k) = \left(\frac{p_1}{1-p_1-...-p_k}, \; ..., \; \frac{p_k}{1-p_1-...-p_k}\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} -\infty & \infty \\ \vdots & \vdots \\ -\infty & \infty \end{pmatrix}$;

$\mathrm{T}(\mathbf{x}) = \mathbf{x}$
$\kappa(\theta) = \tilde n \ln(1 + e^{\theta_1} + ... + e^{\theta_k})$

$$\dot\kappa(\theta) = \left(\frac{\tilde n e^{\theta_1}}{1+e^{\theta_1}+...+e^{\theta_k}}, \; ..., \; \frac{\tilde n e^{\theta_k}}{1+e^{\theta_1}+...+e^{\theta_k}}\right)$$

$$\text{MLE} = \left(\ln\Big(\frac{\sum_{i=1}^{n} X_1^{(i)}}{n\tilde{n} - \sum_{i=1}^{n} X_k^{(i)}}\Big),\ \ldots,\ \ln\Big(\frac{\sum_{i=1}^{n} X_1^{(i)}}{n\tilde{n} - \sum_{i=1}^{n} X_k^{(i)}}\Big) \right).$$

where $X^{(1)}, \ldots, X^{(n)}$ is a sample.

7. Exponential distribution

$\mathfrak{X} = (0, \infty)$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \lambda \exp(-\lambda x);$$

its exponential representation:

$$f_\theta(x) = \theta \exp(-\theta x)$$

with respect to measure μ on $\mathfrak{X}$ which is again the restricted Lebesgue measure;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \lambda \qquad \& \qquad \pi^{-1}(\lambda) = \theta;$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = -x$
$\kappa(\theta) = -\ln(\theta)$

$$\dot{\kappa}(\theta) = -\frac{1}{\theta} \qquad \& \qquad \text{MLE} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

8. Gaussian distribution

$\mathfrak{X} = \mathbb{R}$;
original density (with respect to the Lebesgue measure on $\mathfrak{X}$):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right);$$

its exponential representation:

$$f_\theta(x) = \frac{1}{\sqrt{\frac{\pi}{\theta_2}} \exp(\frac{\theta_1^2}{4\theta_2})} \exp\left(-x^2\theta_2 + x\theta_1\right)$$

with respect to measure μ on $\mathfrak{X}$ which is again the Lebesgue measure;
dimension of the parameter: $d = 2$;
parametrisation:

$$\pi(\theta_1, \theta_2) = \left(\frac{\theta_1}{2\theta_2}, \frac{1}{2\theta_2}\right) \qquad \& \qquad \pi^{-1}(m, \sigma^2) = \left(\frac{m}{\sigma^2}, \frac{1}{2\sigma^2}\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} -\infty & \infty \\ 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = (x, -x^2)$

$$\kappa(\theta_1, \theta_2) = \frac{\theta_1^2}{4\theta_2} + \frac{1}{2}\ln\left(\frac{\pi}{\theta_2}\right) \qquad \& \qquad \dot{\kappa}(\theta_1, \theta_2) = \left(\frac{\theta_1}{2\theta_2}, -\frac{\theta_1^2}{4\theta_2^2} - \frac{1}{2\theta_2}\right)$$

$$\mathrm{MLE} = \left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2}, \frac{\frac{1}{2}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2} \right).$$

9. Gaussian mean-known distribution

known parameter: $m \in \mathbb{R}$;

$\mathfrak{X} = \mathbb{R}$;
original density (with respect to the Lebesgue measure on $\mathfrak{X}$):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right);$$

its exponential representation:

$$f_\theta(x) = \frac{1}{\sqrt{\frac{\pi}{\theta}}} \exp\left(-\theta(x-m)^2\right)$$

with respect to measure μ which is again the Lebesgue measure on $\mathfrak{X}$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \frac{1}{2\theta} \qquad \& \qquad \pi^{-1}(\sigma^2) = \frac{1}{2\sigma^2};$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = -(x-m)^2$
$\kappa(\theta) = \frac{1}{2}\ln\left(\frac{\pi}{\theta}\right)$

$$\dot{\kappa}(\theta) = -\frac{1}{2\theta} \qquad \& \qquad \mathrm{MLE} = \frac{n}{2\sum_{i=1}^{n}(X_i - m)^2}.$$

10. Gaussian variance-known distribution

known parameter: $\sigma^2 > 0$;

$\mathfrak{X} = \mathbb{R}$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right);$$

its exponential representation:

$$f_\theta(x) = \exp\left(\frac{x}{\sigma^2}\theta - \frac{\theta^2}{2\sigma^2}\right)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the Lebesgue measure with density $\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{x^2}{2\sigma^2})$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \theta \qquad \& \qquad \pi^{-1}(m) = m;$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} -\infty & \infty \end{pmatrix}$;
$\mathrm{T}(x) = \frac{x}{\sigma^2}$

$\kappa(\theta) = \frac{\theta^2}{2\sigma^2}$

$$\dot{\kappa}(\theta) = \frac{\theta}{\sigma^2} \qquad \& \qquad \mathrm{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

11. Gamma distribution

$\mathfrak{X} = \mathbb{R}^+$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1};$$

its exponential representation:

$$f_\theta(x) = \frac{1}{\Gamma(\theta_1)/\theta_2^{\theta_1}} \exp(\theta_1 \ln x - \theta_2 x)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $1/x$;
dimension of the parameter: $d = 2$;
parametrisation:

$$\pi(\theta_1, \theta_2) = (\theta_1, \theta_2) \qquad \& \qquad \pi^{-1}(p, a) = (p, a);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \\ 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = (\ln x, -x)$
$\kappa(\theta_1, \theta_2) = \ln\Gamma(\theta_1) - \theta_1 \ln\theta_2$

$$\dot{\kappa}(\theta_1, \theta_2) = \left(\frac{\Gamma'(\theta_1)}{\Gamma(\theta_1)} - \ln\theta_2,\, -\frac{\theta_1}{\theta_2}\right)$$

MLE is the solution of the following equations

$$\begin{aligned} \frac{\Gamma'(\theta_1)}{\Gamma(\theta_1)} - \ln\theta_1 &= \frac{1}{n}\sum_{i=1}^{n} \ln X_i - \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ \theta_2 &= \frac{\theta_1}{\frac{1}{n}\sum\limits_{i=1}^{n} X_i}. \end{aligned}$$

12. Gamma p-known distribution

$p > 0$ is a known parameter;
$\mathfrak{X} = \mathbb{R}^+$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1};$$

its exponential representation:

$$f_\theta(x) = \frac{1}{\Gamma(p)/\theta^p} \exp(-\theta x)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $1/x^{p-1}$;
dimension of the parameter: $d = 1$;

parametrisation:

$$\pi(\theta) = \theta \qquad \& \qquad \pi^{-1}(a) = a;$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;

$\mathrm{T}(x) = -x$

$\kappa(\theta) = \ln \Gamma(p) - p \ln \theta$

$$\dot{\kappa}(\theta) = -\frac{p}{\theta} \qquad \& \qquad \mathrm{MLE} = \frac{p}{\frac{1}{n}\sum\limits_{i=1}^{n} X_i}$$

13. Beta distribution

$\mathfrak{X} = (0, 1)$;

original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)};$$

its exponential representation:

$$f_\theta(x) = \frac{1}{B(\theta_1, \theta_2)} \exp(\theta_1 \ln x + \theta_2 \ln(1-x))$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $\frac{1}{x(1-x)}$;

dimension of the parameter: $d = 2$;

parametrisation:

$$\pi(\theta_1, \theta_2) = (\theta_1, \theta_2) \qquad \& \qquad \pi^{-1}(a, b) = (a, b);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \\ 0 & \infty \end{pmatrix}$;

$\mathrm{T}(x) = (\ln x, \ln(1-x))$

$\kappa(\theta_1, \theta_2) = \ln(B(\theta_1, \theta_2)) = \ln \Gamma(\theta_1) + \ln \Gamma(\theta_2) - \ln \Gamma(\theta_1 + \theta_2)$

$$\dot{\kappa}(\theta_1, \theta_2) = \left(\frac{\Gamma'(\theta_1)}{\Gamma(\theta_1)} - \frac{\Gamma'(\theta_1 + \theta_2)}{\Gamma(\theta_1 + \theta_2)}, \frac{\Gamma'(\theta_2)}{\Gamma(\theta_2)} - \frac{\Gamma'(\theta_1 + \theta_2)}{\Gamma(\theta_1 + \theta_2)} \right)$$

MLE is the solution of the following equations

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) &= \frac{\Gamma'(\theta_1)}{\Gamma(\theta_1)} - \frac{\Gamma'(\theta_1 + \theta_2)}{\Gamma(\theta_1 + \theta_2)} \\ \frac{1}{n}\sum_{i=1}^{n} \ln(1 - X_i) &= \frac{\Gamma'(\theta_2)}{\Gamma(\theta_2)} - \frac{\Gamma'(\theta_1 + \theta_2)}{\Gamma(\theta_1 + \theta_2)}. \end{aligned}$$

14. Lognormal distribution

$\mathfrak{X} = \mathbb{R}^+$;

original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left(-\frac{(\ln x - m)^2}{2\sigma^2} \right);$$

its exponential representation:

$$f_\theta(x) = \frac{1}{\sqrt{\frac{\pi}{\theta_2}} \exp(\frac{\theta_1^2}{4\theta_2})} \exp\left(-(\ln x)^2 \theta_2 + \ln x \theta_1 \right)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $1/x$;
dimension of the parameter: $d = 2$;
parametrisation:

$$\pi(\theta_1, \theta_2) = \left(\frac{\theta_1}{2\theta_2}, \frac{1}{2\theta_2}\right) \qquad \& \qquad \pi^{-1}(m, \sigma^2) = \left(\frac{m}{\sigma^2}, \frac{1}{2\sigma^2}\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} -\infty & \infty \\ 0 & \infty \end{pmatrix}$;

$\mathrm{T}(x) = (\ln x, -(\ln x)^2)$

$$\kappa(\theta_1, \theta_2) = \frac{\theta_1^2}{4\theta_2} + \frac{1}{2}\ln\left(\frac{\pi}{\theta_2}\right) \qquad \& \qquad \dot{\kappa}(\theta_1, \theta_2) = \left(\frac{\theta_1}{2\theta_2}, -\frac{\theta_1^2}{4\theta_2^2} - \frac{1}{2\theta_2}\right)$$

$$\mathrm{MLE} = \left(\frac{\frac{1}{n}\sum_{i=1}^{n} \ln X_i}{\frac{1}{n}\sum_{i=1}^{n} (\ln X_i)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} \ln X_i\right)^2}, \frac{\frac{1}{2}}{\frac{1}{n}\sum_{i=1}^{n} (\ln X_i)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} \ln X_i\right)^2} \right).$$

15. Double exponential distribution

$\mathfrak{X} = \mathbb{R}$;
original density (with respect to the Lebesgue measure on $\mathfrak{X}$):

$$f(x) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right);$$

its exponential representation:

$$f_\theta(x) = \frac{\theta}{2}\exp(-\theta|x|)$$

with respect to measure μ which is again the Lebesgue measure on $\mathfrak{X}$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \frac{1}{\theta} \qquad \& \qquad \pi^{-1}(b) = \frac{1}{b};$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = -|x|$
$\kappa(\theta) = \ln 2 - \ln\theta$

$$\dot{\kappa}(\theta) = -\frac{1}{\theta} \qquad \& \qquad \mathrm{MLE} = \frac{n}{\sum_{i=1}^{n} |X_i|}.$$

16. Weibull distribution

known parameter: $p > 0$;
$\mathfrak{X} = \mathbb{R}^+$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = cpx^{p-1}\exp(-cx^p);$$

its exponential representation:

$$f_\theta(x) = \theta\exp(-\theta x^p)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density px^{p-1};
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \theta \qquad \& \qquad \pi^{-1}(c) = c;$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = -x^p$
$\kappa(\theta) = -\ln\theta$

$$\dot{\kappa}(\theta) = -\frac{1}{\theta} \qquad \& \qquad \mathrm{MLE} = \frac{n}{\sum_{i=1}^{n} X_i^p}.$$

17. Reyleigh distribution

$\mathfrak{X} = \mathbb{R}^+$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{x}{b^2}\exp\left(-\frac{x^2}{2b^2}\right);$$

its exponential representation:

$$f_\theta(x) = \theta\exp(-\theta x^2)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density x;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \frac{1}{2\theta} \qquad \& \qquad \pi^{-1}(b^2) = \frac{1}{2b^2};$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$;
$\mathrm{T}(x) = -x^2$
$\kappa(\theta) = -\ln\theta$

$$\dot{\kappa}(\theta) = -\frac{1}{\theta} \qquad \& \qquad \mathrm{MLE} = \frac{n}{\sum_{i=1}^{n} X_i^2}.$$

18. Maxwell distribution

$\mathfrak{X} = \mathbb{R}^+$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{2}{a^3\sqrt{2\pi}}\, x^2\exp\left(-\frac{x^2}{2a^2}\right);$$

its exponential representation:

$$f_\theta(x) = \frac{4}{\theta^{-3/2}\sqrt{\pi}}\exp(-x^2\theta)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density x^2;
dimension of the parameter: $d = 1$;

parametrisation:

$$\pi(\theta)=\sqrt{\frac{1}{2\theta}} \qquad \& \qquad \pi^{-1}(a)=\frac{1}{2a^2};$$

boundaries of Θ: $\mathsf{h}=\begin{pmatrix} 0 & \infty \end{pmatrix};$

$\mathrm{T}(x)=-x^2$

$\kappa(\theta)=-\frac{3}{2}\ln\theta-\ln(4/\sqrt{\pi})$

$$\dot\kappa(\theta)=-\frac{3}{2\theta} \qquad \& \qquad \mathrm{MLE}=\frac{3n}{2\sum\limits_{i=1}^{n}X_i^2}.$$

19. Pareto distribution

known parameter: $b>0$;

$\mathfrak{X}=(b,\infty)$;

original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x)=\frac{a}{b}\left(\frac{b}{x}\right)^{a+1};$$

its exponential representation:

$$f_\theta(x)=\theta b^\theta\exp(-\theta\ln x)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $1/x$;

dimension of the parameter: $d=1$;

parametrisation:

$$\pi(\theta)=\theta \qquad \& \qquad \pi^{-1}(a)=a;$$

boundaries of Θ: $\mathsf{h}=\begin{pmatrix} b & \infty \end{pmatrix};$

$\mathrm{T}(x)=-\ln x$

$\kappa(\theta)=-\ln\theta-\theta\ln b$

$$\dot\kappa(\theta)=-\frac{1}{\theta}-\ln b \qquad \& \qquad \mathrm{MLE}=\frac{n}{\sum\limits_{i=1}^{n}\ln\left(\frac{X_i}{b}\right)}.$$

20. Modular distribution

see Gaussian mean-known.

21. Inverse Gaussian distribution

$\mathfrak{X}=\mathbb{R}^+$;

original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x)=\sqrt{\frac{\lambda}{2\pi x^3}}\ \exp\left(-\frac{\lambda}{2m^2}\frac{(x-m)^2}{x}\right);$$

its exponential representation:

$$f_\theta(x)=\sqrt{\theta_2}\exp(-\theta_1\frac{x}{2}-\theta_2\frac{1}{2x}+\sqrt{\theta_1\theta_2})$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $1/\sqrt{2\pi x^3}$;

dimension of the parameter: $d = 2$;
parametrisation:

$$\pi(\theta_1, \theta_2) = \left(\sqrt{\frac{\theta_2}{\theta_1}},\, \theta_2\right) \qquad \& \qquad \pi^{-1}(m, \lambda) = \left(\frac{\lambda}{m^2},\, \lambda\right);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \\ 0 & \infty \end{pmatrix}$;

$\mathrm{T}(x) = \left(-\frac{x}{2}, -\frac{1}{2x}\right)$

$$\kappa(\theta_1, \theta_2) = -\sqrt{\theta_1\theta_2} - \frac{1}{2}\ln\theta_2 \qquad \& \qquad \dot\kappa(\theta_1, \theta_2) = \left(-\frac{1}{2}\sqrt{\frac{\theta_2}{\theta_1}},\, -\frac{1}{2}\left(\sqrt{\frac{\theta_1}{\theta_2}} + \frac{1}{\theta_2}\right)\right)$$

$$\mathrm{MLE} = \left(\frac{1}{(\frac{1}{n}\sum_{i=1}^n X_i)(\frac{1}{n}\sum_{i=1}^n \frac{1}{X_i}) - 1},\, \frac{\frac{1}{n}\sum_{i=1}^n X_i}{(\frac{1}{n}\sum_{i=1}^n X_i)(\frac{1}{n}\sum_{i=1}^n \frac{1}{X_i}) - 1}\right).$$

22. Inverse Gamma distribution

$\mathfrak{X} = \mathbb{R}^+$;
original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{a^p}{\Gamma(p)}\exp(-\frac{a}{x})x^{-p-1};$$

its exponential representation:

$$f_{\theta_1,\theta_2}(x) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\exp(-\theta_2\frac{1}{x} - \theta_1\ln x)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $1/x$;
dimension of the parameter: $d = 2$;
parametrisation:

$$\pi(\theta_1, \theta_2) = (\theta_1, \theta_2) \qquad \& \qquad \pi^{-1}(p, a) = (p, a);$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \\ 0 & \infty \end{pmatrix}$;

$\mathrm{T}(x) = \left(-\ln x, -\frac{1}{x}\right)$

$$\kappa(\theta_1, \theta_2) = \ln\Gamma(\theta_1) - \theta_1\ln\theta_2 \qquad \& \qquad \dot\kappa(\theta_1, \theta_2) = \left(\frac{\Gamma'(\theta_1)}{\Gamma(\theta_1)} - \ln\theta_2,\, -\frac{\theta_1}{\theta_2}\right)$$

MLE is the solution of the following equations

$$\begin{aligned} \ln\theta_1 - \frac{\Gamma'(\theta_1)}{\Gamma(\theta_1)} &= \frac{1}{n}\sum_{i=1}^n \ln(X_i) + \ln\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{X_i}\right) \\ \theta_2 &= \frac{\theta_1}{\frac{1}{n}\sum_{i=1}^n \frac{1}{X_i}}. \end{aligned}$$

23. χ^2 with k degrees of freedom distribution

$\mathfrak{X} = \mathbb{R}^+$;

original density (with respect to the Lebesgue measure restricted on $\mathfrak{X}$):

$$f(x) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})}\, x^{\frac{k}{2}-1} \exp(-\frac{x}{2});$$

its exponential representation:

$$f_\theta(x) = \frac{1}{2^{\theta/2}\Gamma(\frac{\theta}{2})} \exp\left(\theta \frac{\ln x}{2}\right)$$

with respect to measure μ on $\mathfrak{X}$ which is absolutely continuous with respect to the restricted Lebesgue measure with density $\frac{1}{x}\exp(-\frac{x}{2})$;
dimension of the parameter: $d = 1$;
parametrisation:

$$\pi(\theta) = \theta \qquad \& \qquad \pi^{-1}(k) = k;$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \end{pmatrix}$, (note that $\Theta = (0,\infty)$ is assumed);
$\mathrm{T}(x) = \frac{\ln x}{2}$

$$\kappa(\theta) = \frac{\theta}{2}\ln 2 + \ln\Gamma(\frac{\theta}{2}) \qquad \& \qquad \dot{\kappa}(\theta) = \frac{\ln 2}{2} + \frac{\Gamma'(\frac{\theta}{2})}{2\Gamma(\frac{\theta}{2})}$$

MLE is the solution of the following equation

$$\frac{1}{n}\sum_{i=1}^{n} \ln\frac{X_i}{2} \quad = \quad \frac{\Gamma'(\frac{\theta}{2})}{\Gamma(\frac{\theta}{2})}.$$

<u>24. Dirichlet distribution</u>

$\mathfrak{X} = \{\mathbf{x} \in (0,\infty)^{k+1} : x_1 + ... + x_{k+1} = 1\} = (0,\infty)^k \times \{1 - x_1 - ... - x_k\}$;
original density (with respect to the Lebesgue measure on $(0,\infty)^k$):

$$f(\mathbf{x}) = \frac{\Gamma(\sum_{j=1}^{k+1}\alpha_j)}{\prod_{j=1}^{k+1}\Gamma(\alpha_j)} x_1^{\alpha_1-1} ...\, x_k^{\alpha_k-1}\,(1 - x_1 - ... - x_k)^{\alpha_{k+1}-1};$$

its exponential representation:

$$f_\theta(\mathbf{x}) = \frac{\Gamma(\sum_{j=1}^{k+1}\theta_j)}{\prod_{j=1}^{k+1}\Gamma(\theta_j)} \exp\Big(\theta_1 \ln x_1 + ... + \theta_k \ln x_k + \theta_{k+1}\ln(1 - x_1 - ... - x_k)\Big)$$

with respect to measure μ on $(0,\infty)^k$ which is absolutely continuous with respect to the Lebesgue measure on $(0,\infty)^k$ with density $\frac{1}{x_1}\frac{1}{x_2}...\frac{1}{x_k}\frac{1}{1-x_1-...x_k}$;
dimension of the parameter: $d = k+1$;
parametrisation:

$$\pi(\theta_1,...,\theta_{k+1}) = (\theta_1,...,\theta_{k+1}) \qquad \& \qquad \pi^{-1}(\alpha_1,...,\alpha_{k+1}) = (\alpha_1,...,\alpha_{k+1});$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \\ \vdots & \vdots \\ 0 & \infty \end{pmatrix}$;

$\mathrm{T}(\mathbf{x}) = (\ln x_1, ..., \ln x_k, \ln(1 - x_1 - ... - x_k))$
$\kappa(\theta_1,...,\theta_{k+1}) = -\ln\Gamma(\sum_{j=1}^{k+1}\theta_j) + \sum_{j=1}^{k+1}\ln\Gamma(\theta_j)$

$$\dot{\kappa}(\theta_1,...,\theta_{k+1}) = \left(\frac{\Gamma'(\theta_j)}{\Gamma(\theta_j)} - \frac{\Gamma'(\sum_{j=1}^{k+1}\theta_j)}{\Gamma(\sum_{j=1}^{k+1}\theta_j)} : j = 1,...,k+1\right)$$

MLE is the solution of the following equations

$$\frac{1}{n}\sum_{i=1}^{n}\ln X_j^{(i)} = \frac{\Gamma'(\theta_j)}{\Gamma(\theta_j)} - \frac{\Gamma'(\sum_{l=1}^{k+1}\theta_l)}{\Gamma(\sum_{l=1}^{k+1}\theta_l)} \qquad \text{for } j = 1, ..., k+1$$

where $X^{(1)}, ..., X^{(n)}$ is a sample.

25. Bivariate Gaussian distribution

$\mathfrak{X} = \mathbb{R}^2$;
original density (with respect to the Lebesgue measure on $\mathfrak{X}$):

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} x-m_1 & y-m_2\end{pmatrix}\Sigma^{-1}\begin{pmatrix} x-m_1 \\ y-m_2\end{pmatrix}\right\} \\ &= \frac{1}{2\pi|\Sigma|^{1/2}} \frac{\exp\left\{-\frac{1}{2}\operatorname{Tr}\left(\Sigma^{-1}\binom{x}{y}(x\ \ y)\right) + (x\ \ y)\Sigma^{-1}\binom{m_1}{m_2}\right\}}{\exp\left\{\frac{1}{2}(m_1\ \ m_2)\ \Sigma^{-1}\binom{m_1}{m_2}\right\}} \end{aligned}$$

where $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$;
its exponential representation:

$$f_\theta(x,y) = \frac{\sqrt{\theta_1\theta_3 - \theta_2^2}}{2\pi}\ \frac{\exp\left\{-\theta_1\frac{x^2}{2} + \theta_2 xy - \theta_3\frac{y^2}{2} + \theta_4 x + \theta_5 y\right\}}{\exp\left\{\frac{1}{2}\ (\theta_4\ \ \theta_5)\begin{pmatrix}\theta_1 & -\theta_2 \\ -\theta_2 & \theta_3\end{pmatrix}^{-1}\binom{\theta_4}{\theta_5}\right\}}$$

with respect to measure μ which is again the Lebesgue measure on $\mathfrak{X}$;
dimension of the parameter: $d = 5$;
parametrisation:

$$\begin{pmatrix}\theta_1 & -\theta_2 \\ -\theta_2 & \theta_3\end{pmatrix} = \Sigma^{-1} \qquad \& \qquad \begin{pmatrix}\theta_4 \\ \theta_5\end{pmatrix} = \Sigma^{-1}\begin{pmatrix}m_1 \\ m_2\end{pmatrix},$$

i.e.

$$\begin{aligned} \pi(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5) &= \left(\frac{\theta_3}{\theta_1\theta_3-\theta_2^2}, \frac{\theta_1}{\theta_1\theta_3-\theta_2^2}, \frac{\theta_2}{\sqrt{\theta_1\theta_3}}, \frac{\theta_2\theta_5+\theta_3\theta_4}{\theta_1\theta_3-\theta_2^2}, \frac{\theta_1\theta_5+\theta_2\theta_4}{\theta_1\theta_3-\theta_2^2}\right) \\ \pi^{-1}(\sigma_1^2,\sigma_2^2,\rho,m_1,m_2) &= \left(\frac{1}{\sigma_1^2(1-\rho^2)}, \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}, \frac{1}{\sigma_2^2(1-\rho^2)},\right. \\ &\qquad \left.\frac{m_1}{\sigma_1^2(1-\rho^2)} - \frac{\rho m_2}{\sigma_1\sigma_2(1-\rho^2)}, \frac{m_2}{\sigma_2^2(1-\rho^2)} - \frac{\rho m_1}{\sigma_1\sigma_2(1-\rho^2)}\right); \end{aligned}$$

boundaries of Θ: $\mathsf{h} = \begin{pmatrix} 0 & \infty \\ -\infty & \infty \\ 0 & \infty \\ -\infty & \infty \\ -\infty & \infty \end{pmatrix}$;

$\mathrm{T}(x,y) = \left(-\frac{x^2}{2}, xy, -\frac{y^2}{2}, x, y\right)$

$$\kappa(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5) = \ln 2\pi - \frac{1}{2}\ln(\theta_1\theta_3 - \theta_2^2) + \frac{1}{2}\ (\theta_4\ \ \theta_5)\begin{pmatrix}\theta_1 & -\theta_2 \\ -\theta_2 & \theta_3\end{pmatrix}^{-1}\begin{pmatrix}\theta_4 \\ \theta_5\end{pmatrix}$$

let us set

$$\Delta := (\theta_4 \;\; \theta_5)\begin{pmatrix}\theta_1 & -\theta_2\\ -\theta_2 & \theta_3\end{pmatrix}^{-1}\begin{pmatrix}\theta_4\\ \theta_5\end{pmatrix} = \frac{1}{\theta_1\theta_3-\theta_2^2}\,(\theta_4 \;\; \theta_5)\begin{pmatrix}\theta_3 & \theta_2\\ \theta_2 & \theta_1\end{pmatrix}\begin{pmatrix}\theta_4\\ \theta_5\end{pmatrix} = \frac{\theta_3\theta_4^2+2\theta_2\theta_4\theta_5+\theta_1\theta_5^2}{\theta_1\theta_3-\theta_2^2}$$

then

$$\dot\kappa(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5) = \frac{1}{2}\frac{1}{\theta_1\theta_3-\theta_2^2}\big(-\theta_3+\theta_5^2-\theta_3\Delta,\; 2\theta_2+2\theta_4\theta_5+2\theta_2\Delta,\; -\theta_1+\theta_4^2-\theta_1\Delta, 2\theta_3\theta_4+2\theta_2\theta_5,\; 2\theta_2\theta_4+2\theta_1\theta_5\big)\,;$$

let us set

$$\Upsilon = \frac{1}{n^2}\sum_{i=1}^n X_i^2\sum_{i=1}^n Y_i^2 - \frac{1}{n^2}(\sum_{i=1}^n X_iY_i)^2 + \frac{1}{n^3}\sum_{i=1}^n X_i\sum_{i=1}^n Y_i\sum_{i=1}^n X_iY_i + -\frac{1}{n^2}\sum_{i=1}^n X_i\sum_{i=1}^n Y_i^2 - \frac{1}{n^2}\sum_{i=1}^n X_i^2\sum_{i=1}^n Y_i$$

then

$$\mathrm{MLE} = \frac{1}{\Upsilon}\Bigg(\frac{1}{n}\sum_{i=1}^n Y_i^2 - (\frac{1}{n}\sum_{i=1}^n Y_i)^2,\; \frac{1}{n}\sum_{i=1}^n X_iY_i - \frac{1}{n^2}\sum_{i=1}^n X_i\sum_{i=1}^n Y_i,\; \frac{1}{n}\sum_{i=1}^n X_i^2 - (\frac{1}{n}\sum_{i=1}^n X_i)^2,\; \frac{1}{n^2}\sum_{i=1}^n X_i\sum_{i=1}^n Y_i^2 - \frac{1}{n^2}\sum_{i=1}^n X_iY_i\sum_{i=1}^n Y_i,\; \frac{1}{n^2}\sum_{i=1}^n X_i^2\sum_{i=1}^n Y_i - \frac{1}{n^2}\sum_{i=1}^n X_iY_i\sum_{i=1}^n X_i\Bigg).$$

REFERENCES

[Bh46] A. Bhattacharyya. On some analogues to the amount of information and their uses in statistical estimation. *Sankhya*, 8: 1 – 14, 1946.

[Br86] L.D. Brown. *Fundamentals of statistical exponential families.* Lectures Notes 9, inst. of Math. Statistics, Hayward, California, 1986.

[C95] I. Csiszár. Generalized cutoff rates and Rényi's information measures. *Transactions of IEEE on Information Theory*, 41: 26 – 34, 1995.

[J88] M. Janžura. Divergences of Gauss-Markov random field with application to statistical inference. *Kybernetika*, 24: 401 – 412, 1988.

[GS75] I.I. Gihman, A.V. Skorokhod. *The Theory of Stochastic Processes*, vol.2 Springer, Berlin, 1975.

[KL51] S. Kullback, R. Leibler. On information and sufficiency. *Annn.Math.Statist.*, 22: 79 – 86, 1951.

[KS97] U. Küchler, M. Sørensen. *Exponential Families of Stochastic Processes*. Springer, Berlin, 1997.

[LV87] F. Liese, I. Vajda. *Convex statistical distances.* Teubner, Leipzig, 1987.

[MPV97] D. Morales, L. Pardo, I. Vajda. Some new statistics for testing hypothesis in parametric models. *Journal of Multivariate Analysis*, 62: 137 – 168, 1997.

[MPV00] D. Morales, L. Pardo, I. Vajda. Rényi statistics in directed families of exponential experiments. *Statistics*, 34: 151 – 174, 2000.

[MPPV04] D. Morales, L. Pardo, M.C. Pardo, I. Vajda. Rényi statistics for testing composite hypothesis in general exponential models. *Statistics*, 38, No.2: 133 – 147, 2004.

[P74] A. Perez. Generalization of Chernoff's result on the asymptotic discernibility of two random processes. *Colloguia Math.Soc. J.Bolyai.*, 9: 619 – 632, 1974.

[R61] A. Rényi. On measures of entropy and information. *Proc. 4th Berkeley Symp.Math.Statist.Probab.*, 1: 547 – 561, 1961.

[Ro70] R.T. Rockefellar. *Convex Analysis.* Princeton University Press, New Jersey, 1970.

[V90] I. Vajda. Distances and discrimination rates for stochastic processes. *Stoch.Processes and Appl.*, 35: 47 – 57, 1990.

[Z71] S. Zacks. *The Theory of Statistical Inference.* Wiley, New York, 1971.

E-mail address: `fajfrova@utia.cas.cz`

ÚTIA AV ČR, POD VODÁRENSKOU VĚŽÍ4, PRAHA 8, 182 08