On statistical experiments converging to exponential limits

A.A. Gushchin*
Steklov Mathematical Institute
Gubkina Str. 8
117966 Moscow GSP-1, Russia
E-mail: gushchin@genesis.mi.ras.ru

E. Valkeila†
Department of Mathematics, P.O. Box 4
FIN–00014 University of Helsinki, Finland
E-mail: Esko.Valkeila@Helsinki.Fi

Abstract

A characterization of a certain class of exponential experiments, so called E-experiments, is given. This allows us to give necessary and sufficient conditions for a sequence of experiments to converge to an E-experiment. The obtained results are valid for Gaussian shift experiments. Some asymptotic approximations using E-experiments are studied.

Key words: Kullback information, Hellinger integral, convergence of experiments, exponential limiting experiments

AMS-classification: 62B15, 62A99

*Supported by INTAS/RFBR-95-0061 and Suomalainen Tiedeakatemia.
†Supported by the Research Grants Committee of the University of Helsinki.

1 Introduction

0. (*Introduction*) Exponential families, especially the Gaussian shift experiments, often occur in asymptotic statistics as limit models. They are also used as approximating models for the original experiments. The main purpose of this paper is to give some description of the convergence to an experiment generated by an exponential family. Some asymptotic approximation properties are also studied. In order to achieve all this we prefer to work with a different, but related, class of exponential-type experiments, which we call E-experiments. To define this class of experiments we use a certain geometric property, which characterizes exponential families. This property does not depend on the dimensionality of the parameter space. This allows us to work with infinite dimensional parameter spaces, at least in principle. Finally, we do not assume pairwise equivalence of measures defining an E-experiment. This generalization seems to be of minor practical importance but it gives such a nice property as weak compactness of E-experiments.

We use the neutral term 'E-experiment' instead of 'exponential experiment', since the latter is used in the statistical literature not only in connection to exponential families but also to exponential distributions.

The results in this paper rely heavily on the convexity of the parameter space of the E-experiment. This means that these methods fit poorly to analyze curved exponential families. Curved exponential families often appear in limiting experiments of different models of stochastic processes, e.g. if the model satisfies the local asymptotic mixed normality condition or, more generally, if it is locally asymptotically quadratic. In this situation the above mentioned disadvantage of our approach can be avoided to some extent by modifying it in the context of filtered statistical experiments. Since this modified approach is quite complicated and technical — it is based on the heavy use of stochastic calculus — it is not even possible to give a summary of its main steps here.

1. (*Basic definitions*) A statistical experiment $\mathbb{E}$ is a collection $\mathbb{E} = \big(\Omega, F, (P_\theta)_{\theta\in\Theta}\big)$, where (Ω, F) is a measurable space and P_θ is a probability measure for each $\theta \in \Theta$. To indicate that the experiment depends on the parameter set Θ we also use the notation $\mathbb{E}(\Theta)$ instead of $\mathbb{E}$. Generally we will assume that the parameter set Θ is a convex set in a real linear space V.

$\mathbb{E}$ is a homogeneous experiment if all the measures P_θ, $\theta \in \Theta$, are pairwise equivalent.

If $I \subset \Theta$, then the experiment $\mathbb{E}_I = (\Omega, F, (P_\theta)_{\theta\in I})$ is the restriction of $\mathbb{E}$ to I.

If $\mathbb{E}$ and $\mathbb{E}'$ are two experiments, then the notation $\mathbb{E} \sim \mathbb{E}'$ means that the experiments are equivalent (see [9, p. 260]) and $\mathbb{E} \overset{(k)}{\sim} \mathbb{E}'$ means that the experiments are k-wise equivalent, i.e. $\mathbb{E}_J \sim \mathbb{E}'_J$ for all J such that $|J| \leq k$, where $|J|$ is the cardinality of the set J.

2. (*Hellinger integrals and transforms*) Let P, P' be two probability measures on (Ω, F). The Hellinger integral $H(\alpha; P, P')$ of order α, $\alpha \in \mathbb{R}$, is

$$H(\alpha; P, P') \doteq \int \left(\frac{dP}{dQ}\right)^{\alpha} \left(\frac{dP'}{dQ}\right)^{1-\alpha} dQ,$$

where Q dominates the measures P, P'. Above we used the conventions $0^0 = 1$, $0^\alpha = \infty$ for $\alpha < 0$ and $0 \cdot \infty = \infty \cdot 0 = 0$.

The Hellinger integral is a special case of the Hellinger transform. We say that a mapping $\alpha\colon \Theta \to \mathbb{R}$ is a multi-index if $I_\alpha \doteq \{\theta : \alpha(\theta) \neq 0\}$ is finite and $\sum_{\theta \in I_\alpha} \alpha(\theta) = 1$. $\mathcal{A}$ is the set of all multi-indices over Θ and $\mathcal{A}_+$ is the set of all non-negative multi-indices with values in $\mathbb{R}_+$. In contrast to standard definitions, we define Hellinger transforms not only for $\alpha \in \mathcal{A}_+$ but for all $\alpha \in \mathcal{A}$.

Let $\mathbb{E}$ be an experiment and α a multi-index. The Hellinger transform $H(\alpha; \mathbb{E})$ of order α is

$$H(\alpha; \mathbb{E}) \doteq \int \prod_{\theta \in I_\alpha} \left(\frac{dP_\theta}{dQ}\right)^{\alpha(\theta)} dQ,$$

where the measure Q dominates the family $(P_\theta)_{\theta \in I_\alpha}$ and the same conventions as above are used.

The Hellinger integrals and transforms are independent of the choice of the dominating measure Q.

Evidently, $0 \leq H(\alpha; \mathbb{E}) \leq \infty$; $H(\alpha; \mathbb{E}) \leq 1$ if $\alpha \in \mathcal{A}_+$. It is well known that experiments $\mathbb{E}(\Theta)$ and $\mathbb{E}'(\Theta)$ are equivalent iff $H(\alpha; \mathbb{E}) = H(\alpha; \mathbb{E}')$ for all $\alpha \in \mathcal{A}_+$ (see [9, p. 272]). Of course, if $\mathbb{E} \sim \mathbb{E}'$, then $H(\alpha; \mathbb{E}) = H(\alpha; \mathbb{E}')$ for all $\alpha \in \mathcal{A}$.

3. (*Geodesic curve*) Given two probability measures P and P', which are not singular we define the geodesic curve $G(u; P, P')$, $u \in (0, 1)$, between them by

$$dG(u; P, P') \doteq \frac{(dP)^{1-u} (dP')^{u}}{H(1-u; P, P')} \tag{1.1}$$

(compare also [2, p. 263]). Clearly the measures $G(u; P, P')$, $u \in (0, 1)$, are pairwise equivalent, but $G(u; P, P')$ is not necessarily equivalent to P or to P'.

4. (*Definition of an E-experiment*) We say that a statistical experiment

$$\mathbb{E} \doteq \left(\Omega, F, (P_\theta)_{\theta \in \Theta}\right)$$

is an E-experiment if Θ is a convex set in a real linear space V and

$$P_{(1-u)\theta+u\theta'} = G(u; P_\theta, P_{\theta'}) \tag{1.2}$$

for any $\theta, \theta' \in \Theta$ such that P_θ and $P_{\theta'}$ are not singular, and for any $u \in (0,1)$.

The following is easy to check from the definitions: If $\mathbb{E}$ is an E-experiment, and $\mathbb{E}'$ is another experiment with the property $\mathbb{E}' \overset{(3)}{\sim} \mathbb{E}$, then $\mathbb{E}'$ is also an E-experiment. In particular, the property of being an E-experiment is a property of equivalence classes of experiments.

A stronger result of this type will be proved in Section 4.

5. (*Reparameterization*) The definition of an E-experiment depends essentially from the parameterization. Hence the following problem is natural:

We have a family of probability measures $P \in \mathcal{P}$. When is it possible to define a one-to-one mapping $\psi: \mathcal{P} \to V$, where V is a linear space, such that the family $(P_\theta)_{\theta\in\Theta}$, where $P_\theta = \psi^{-1}(\theta)$ and $\Theta = \psi(\mathcal{P})$, is an E-experiment?

Clearly a necessary condition for this is that for any nonsingular pair $P, P' \in \mathcal{P}$ we have that $G(u; P, P') \in \mathcal{P}$, $u \in (0,1)$. This condition is also sufficient, if in addition the measures $P \in \mathcal{P}$ are pairwise equivalent. Indeed, let $\mathcal{A}$ be the set of multi-indices $\alpha: \mathcal{P} \to \mathbb{R}$. If α, α' are two multi-indices such that

$$\sum_{P\in I_\alpha} \alpha(P) = \sum_{P\in I_{\alpha'}} \alpha'(P)$$

and

$$\sum_{P\in I_\alpha} \alpha(P)\log\Big(\frac{dP}{dQ}\Big) = \sum_{P\in I_{\alpha'}} \alpha'(P)\log\Big(\frac{dP}{dQ}\Big) \quad Q\text{-a.s.},$$

where Q is a measure equivalent to the family $P \in \mathcal{P}$, then we say that the multi-indices α and α' are equivalent. This clearly defines an equivalence relation in $\mathcal{A}$; denote the corresponding equivalence class by $[\alpha]$ for $\alpha \in \mathcal{A}$. Put $[\alpha]+[\beta] \doteq [\alpha+\beta]$ and $c[\alpha] \doteq [c\alpha]$. It is easily checked that these definitions are correct. Moreover, the space $V \doteq \{[\alpha]: \alpha \in \mathcal{A}\}$ is linear. Define $\psi(P) \doteq [\mathbf{1}_{\{P\}}]$. From the condition $G(u; P, P') \in \mathcal{P}$ it follows that $\psi(\mathcal{P})$ is convex and with the above parameterization the family $\mathcal{P}$ is an E-experiment.

6. (*Structure of this paper*) We give some elementary properties of E-experiments in the next section. In particular, we show that two E-experiments are equivalent, if all their binary subexperiments are equivalent. The connections between E-experiments and exponential families and Gaussian shift experiments are discussed. In the next section we give

some basic properties for geodesic curves. The main result is Lemma 3.3, where a characterization for geodesic curves is proved. In the fourth section this allows us to characterize E-experiments, which in turn implies that the property of being an E-experiment is a property of binary subexperiments only. In the last section we prove the main results of the paper. We give necessary and sufficient conditions for the cluster points of a sequence of experiments to be E-experiments and for the convergence to an E-experiment, in particular for the asymptotic normality of experiments. At the end we discuss possibilities of asymptotic approximations of experiments by E-experiments.

2 First properties

1. (*Some elementary properties of E-experiments*)

Proposition 2.1 *Let $\mathbb{E}$ be an E-experiment and $\alpha \in \mathcal{A}$. Put $\theta(\alpha) \doteq \sum_{\theta \in I_\alpha} \alpha(\theta)\theta$.*

1 *If $\alpha \in \mathcal{A}_+$ and $H(\alpha;\mathbb{E}) > 0$, then $P_{\theta(\alpha)} \ll P_\theta$, $\theta \in I_\alpha$.*

2 *If $\theta(\alpha) \in \Theta$ and $P_{\theta(\alpha)} \ll P_\theta$, $\theta \in I_\alpha$, then*

$$dP_{\theta(\alpha)} = \prod_{\theta \in I_\alpha} (dP_\theta)^{\alpha(\theta)} \left[H\left(\alpha;\mathbb{E}\right)\right]^{-1}. \tag{2.3}$$

Proof Without loss of generality, we can assume that $|I_\alpha| \geq 2$.

Let us first consider the case $\alpha \in \mathcal{A}_+$. The proof of both statements goes by induction over $|I_\alpha|$. If $|I_\alpha| = 2$, the claim follows from (1.2). Let $|I_\alpha| = n > 2$ and $H(\alpha;\mathbb{E}) > 0$. Fix some $\eta \in I_\alpha$ and put

$$\beta(\theta) = \begin{cases} \alpha(\theta)/\left(1-\alpha(\eta)\right), & \text{if} \quad \theta \in I_\alpha \setminus \{\eta\}, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\beta \in \mathcal{A}_+$, $|I_\beta| = n-1$, and, evidently, $H(\beta;\mathbb{E}) > 0$. By the induction argument, $P_{\theta(\beta)} \ll P_\theta$, $\theta \in I_\beta$, and

$$dP_{\theta(\beta)} = \prod_{\theta \in I_\beta} (dP_\theta)^{\beta(\theta)} \left[H\left(\beta;\mathbb{E}\right)\right]^{-1}. \tag{2.4}$$

In particular, $P_{\theta(\beta)}$ and P_η are not singular in view of the assumption $H(\alpha;\mathbb{E}) > 0$. Since $\alpha(\eta)\eta + \left(1-\alpha(\eta)\right)\theta(\beta) = \theta(\alpha)$, it follows from (1.2) that $P_{\theta(\alpha)} \ll P_\eta$, $P_{\theta(\alpha)} \ll P_{\theta(\beta)}$, and

$$dP_{\theta(\alpha)} = (dP_\eta)^{\alpha(\eta)} \left(dP_{\theta(\beta)}\right)^{1-\alpha(\eta)} \left[H\left(\alpha(\eta); P_\eta, P_{\theta(\beta)}\right)\right]^{-1}. \tag{2.5}$$

Combining (2.4) and (2.5), we get

$$dP_{\theta(\alpha)} = \prod_{\theta \in I_\alpha} (dP_\theta)^{\alpha(\theta)} \left[H\left(\beta; \mathbb{E}\right)\right]^{-(1-\alpha(\eta))} \left[H\left(\alpha(\eta); P_\eta, P_{\theta(\beta)}\right)\right]^{-1}.$$

Now we arrive at (2.3) integrating over Ω.

Let us now prove the second statement if $\alpha \in \mathcal{A} \setminus \mathcal{A}_+$. One can find a number $a > 1$, multi-indices $\alpha^+ \in \mathcal{A}_+$ and $\alpha^- \in \mathcal{A}_+$ such that $I_{\alpha^+} \cup I_{\alpha^-} = I_\alpha$ and $\alpha = a\alpha^+ - (a-1)\alpha^-$. From the first part of the proof

$$dP_{\theta(\alpha^+)} = \prod_{\theta \in I_{\alpha^+}} (dP_\theta)^{\alpha^+(\theta)} \left[H\left(\alpha^+; \mathbb{E}\right)\right]^{-1}. \tag{2.6}$$

and

$$dP_{\theta(\alpha^-)} = \prod_{\theta \in I_{\alpha^-}} (dP_\theta)^{\alpha^-(\theta)} \left[H\left(\alpha^-; \mathbb{E}\right)\right]^{-1}. \tag{2.7}$$

In particular, $P_{\theta(\alpha)} \ll P_{\theta(\alpha^+)}$ and $P_{\theta(\alpha)} \ll P_{\theta(\alpha^-)}$.

On the other hand, (1.2) implies

$$dP_{\theta(\alpha^+)} = \left(dP_{\theta(\alpha)}\right)^{1/a} \left(dP_{\theta(\alpha^-)}\right)^{(a-1)/a} \left[H\left(1/a; P_{\theta(\alpha)}, P_{\theta(\alpha^-)}\right)\right]^{-1}. \tag{2.8}$$

Combining (2.6)–(2.8), we get

$$dP_{\theta(\alpha)} = \prod_{\theta \in I_\alpha} (dP_\theta)^{\alpha(\theta)} \left[H\left(\alpha^+; \mathbb{E}\right)\right]^{-a} \left[H\left(\alpha^-; \mathbb{E}\right)\right]^{a-1} \left[H\left(1/a; P_{\theta(\alpha)}, P_{\theta(\alpha^-)}\right)\right]^{a},$$

and the integration over Ω completes the proof. ∎

Corollary 2.1 *Let $\mathbb{E}$ and $\mathbb{E}'$ be homogeneous E-experiments. Assume that $\Theta \subseteq \operatorname{aff} I$ and $\mathbb{E}_I \sim \mathbb{E}'_I$. Then $\mathbb{E} \sim \mathbb{E}'$.*

The induction argument in the proof of Proposition 2.1 shows also that, If $\mathbb{E}$ is an E-experiment, the Hellinger transform $H(\alpha; \mathbb{E})$, $\alpha \in \mathcal{A}_+$ is determined by the values of the Hellinger integrals $H(\beta; P_\theta, P_{\theta'})$, $\beta \in (0,1)$, $\theta, \theta' \in \operatorname{conv} I_\alpha$ (in particular, $H(\alpha; \mathbb{E}) = 0$ iff there are $\theta, \theta' \in \operatorname{conv} I_\alpha$ such that P_θ and $P_{\theta'}$ are singular). Hence we obtain the following

Proposition 2.2 *If $\mathbb{E}$ and $\mathbb{E}'$ are E-experiments with $\mathbb{E} \overset{(2)}{\sim} \mathbb{E}'$, then $\mathbb{E} \sim \mathbb{E}'$.*

This result is strengthened in Section 4.

2. (*Exponential families*) Homogeneous E-experiments are closely connected with exponential families. Recall that a family $(P_\theta)_{\theta \in \Theta}$, $\Theta \subseteq \mathbb{R}^k$, on (Ω, F) is an exponential

family with canonical parameterization if there exist a σ-finite measure μ on (Ω, F) and measurable functions $h: \Omega \to \mathrm{I\!R}$ and $S: \Omega \to \mathrm{I\!R}^k$ such that $P_\theta \ll \mu$ and

$$\frac{dP_\theta}{d\mu} = C(\theta) h \exp(\langle \theta, S \rangle) \tag{2.9}$$

for all $\theta \in \Theta$ and suitable constants $C(\theta)$, where $\langle \cdot, \cdot \rangle$ is the scalar product in $\mathrm{I\!R}^k$ (see [1, Chapter 8] for additional information on exponential families).

It is evident that if Θ is a convex subset of $\mathrm{I\!R}^k$ and $(P_\theta)_{\theta \in \Theta}$ is an exponential family with canonical parameterization, then $\big(\Omega, F, (P_\theta)_{\theta \in \Theta}\big)$ is a homogeneous E-experiment.

Conversely, let $E = \big(\Omega, F, (P_\theta)_{\theta \in \Theta}\big)$ be a homogeneous E-experiment. Let $\varphi: \mathrm{I\!R}^k \to V$ be an affine mapping. The set $T \doteq \varphi^{-1}(\Theta)$ is convex. Take an arbitrary affine basis $I = \{t_0, \ldots, t_m\} \subset T$, $m \le k$, for $\operatorname{aff} T$, i.e. a set of affinely independent vectors whose affine hull is $\operatorname{aff} T$. Put $z_i = dP_{\varphi(t_i)}/dP_{\varphi(t_0)}$. By Proposition 2.1,

$$\frac{dP_{\varphi(t)}}{dP_{\varphi(t_0)}} = c(t) \exp\left(\sum_{i=1}^{m} \alpha_i(t) \log z_i \right), \quad t \in T,$$

where $\alpha_0(t)$, $\alpha_1(t)$, $\ldots$, $\alpha_m(t)$ are the barycentric coordinates of t in the basis $\{t_0, \ldots, t_m\}$ and $c(t)$ is a normalizing constant. Since $\alpha_i(t)$, $1 \le i \le m$, are linear functions of $t - t_0$, there are vectors $\eta_i \in \mathrm{I\!R}^k$ such that $\alpha_i(t) = \langle t - t_0, \eta_i \rangle$, $1 \le i \le m$, $t \in \operatorname{aff} I$, and we have the representation (2.9) for the family $\big(P_{\varphi(t)}\big)_{t \in T}$, i.e. $\big(P_{\varphi(t)}\big)_{t \in T}$ is an exponential family with canonical parameterization.

In particular, the class of homogeneous E-experiments over an open convex subset of $\mathrm{I\!R}^k$ coincides with the class of exponential experiments in the sense of Definition 26.1 in [9, p. 115].

Note also that the infinite dimensional generalization of exponential families suggested in [7] is an E-experiment.

3. (*Gaussian experiments*) An important example of an E-experiment is the Gaussian shift experiment. Recall that an experiment $\mathrm{I\!E}$ is a Gaussian experiment, if it is homogeneous and the process

$$\theta \mapsto \log \frac{dP_\theta}{dP}, \quad \theta \in \Theta,$$

is a Gaussian process with respect to the measure $P \doteq P_{\theta_0}$ for some $\theta_0 \in \Theta$. It is a Gaussian shift defined on a finite-dimensional Hilbert space H, if $\Theta = H$, and it is equivalent to the experiment $(H, \mathcal{B}(H), \{\mathcal{N}_H(\theta, I) : \theta \in H\})$, where $\mathcal{N}_H(\theta, I)$ is the Gaussian measure on H with mean θ and the identity covariance I. Finally, a Gaussian shift experiment

on an arbitrary Hilbert space H is an experiment $\mathbb{E}(H)$, whose restrictions to the finite-dimensional subspaces of H are Gaussian shifts [9, p.343]. All Gaussian shift experiments on H are equivalent.
Obviously Gaussian shifts are E-experiments. Conversely, let $\mathbb{E}$ be a Gaussian E-experiment defined on Θ, where $0 \in \Theta$ and $P_\theta \neq P_\eta$ for $\theta, \eta \in \Theta$, $\theta \neq \eta$. Then, one can construct a symmetric positive definite bilinear form $\langle \cdot, \cdot \rangle$ on the linear span $\mathrm{sp}(\Theta)$ such that

$$\langle \theta, \eta \rangle = \mathrm{Cov}_{P_0}\left(\log\left(\frac{dP_\theta}{dP_0}\right), \log\left(\frac{dP_\eta}{dP_0}\right)\right)$$

for $\theta, \eta \in \Theta$. For the corresponding norm $\|\theta\|^2 = \langle \theta, \theta \rangle$ the space $\mathrm{sp}(\Theta)$ may be completed to give a Hilbert space H. This allows us to construct a Gaussian shift on H such that $\mathbb{E}$ is its restriction on Θ.

3 The basic inequalities

1. (*Setup*) Let P and P' be two probability measures on a measurable space (Ω, F) and let Q dominate P and P'. Denote the densities by $z \doteq dP/dQ$ and $z' \doteq dP'/dQ$.

2. (*Kullback information*) The Kullback information $K(P', P)$ is defined as $K(P', P) \doteq E_Q(z' \log(z'/z))$, where $z' \log(z'/z) = 0$ if $z' = 0$ and $z' \log(z'/z) = \infty$ if $z = 0$, $z' > 0$. This definition does not depend on the choice of the dominating measure.
In particular, if $K(P', P) < \infty$, then $P' \ll P$. We have also that $0 \leq K(P', P) \leq \infty$, where the left-hand inequality follows from the inequality $y \log(y/x) \geq y - x$ valid for $x, y \geq 0$.

3. (*Hellinger integrals and Kullback information*) We discuss some known connections between Hellinger integrals and Kullback information (see, for example [5, Section 2]). To make this exposition self-contained we give the proofs for these facts.
Put $f(\alpha; x, y) \doteq x^\alpha y^{1-\alpha}$ for $x, y \geq 0$ and $\alpha \in (0,1)$; $f(0; x, y) \doteq \lim_{\alpha \downarrow 0} f(\alpha; x, y) = y \mathbf{1}_{\{x>0\}}$. The function $f(\alpha; x, y)$ is convex in α for $x, y \geq 0$. Using this and the representation $f(\alpha; x, y) = y e^{\alpha \log(x/y)}$, valid for $x, y > 0$ we obtain that for $0 < \tilde{\alpha} < \alpha$:

$$f'_\alpha(0; x, y) \leq \frac{f(\tilde{\alpha}; x, y) - f(0; x, y)}{\tilde{\alpha}} \leq \frac{f(\alpha; x, y) - f(0; x, y)}{\alpha}, \tag{3.1}$$

where $f'_\alpha(0; x, y)$ is the right-hand side derivative of $f(\alpha; x, y)$ in α at 0. Note also that for $x > 0$, $y > 0$ we have that

$$f'_\alpha(0; x, y) = -y \log \frac{y}{x}. \tag{3.2}$$

Lemma 3.1 *The following relations are valid:*

1 $\frac{1-H(\alpha;P,P')}{\alpha}$ *is a decreasing function in* α, $\alpha \in (0,1)$.

2 $\lim_{\alpha\downarrow 0} \frac{1-H(\alpha;P,P')}{\alpha} = K(P',P)$.

Proof Let $\gamma = P'\{z>0\}$. Then

$$\begin{aligned} 1 - H\left(\alpha;P,P'\right) &= 1-\gamma+\left(\gamma - H\left(\alpha;P,P'\right)\right) = 1-\gamma + E_Q\left(z'\mathbf{1}_{\{z>0\}} - z^{\alpha}{z'}^{1-\alpha}\right) \\ &= 1-\gamma + E_Q\left(f\left(0;z,z'\right) - f\left(\alpha;z,z'\right)\right). \end{aligned}$$

The first statement follows from the second inequality in (3.1).
By the monotone convergence theorem,

$$-\infty < \lim_{\alpha\downarrow 0} \frac{E_Q\left(f\left(0;z,z'\right)-f\left(\alpha;z,z'\right)\right)}{\alpha} = E_Q\left(z'\log\frac{z'}{z}\mathbf{1}_{\{z>0\}}\right).$$

Hence, if $\gamma < 1$,

$$\lim_{\alpha\downarrow 0}\frac{1-H\left(\alpha;P,P'\right)}{\alpha} = \infty.$$

But $\gamma < 1$ means that the measure P' is not absolutely continuous with respect to the measure P, and in this case we also have $K\left(P',P\right)=\infty$.
If $\gamma = 1$, we have that $\{z=0\}\subseteq\{z'=0\}$ Q-a.s., which implies

$$E_Q\left(z'\log\frac{z'}{z}\mathbf{1}_{\{z>0\}}\right) = E_Q\left(z'\log\frac{z'}{z}\right) = K\left(P',P\right).$$

This finishes the proof. ■

4. (*Properties of geodesic curves*) From now on we assume that the measures P and P' are not singular and $0<u<1$. Put $P_u \doteq G(u;P,P')$. Essentially, the next lemma is due to Čencov [2, p. 297].

Lemma 3.2 *Let* R *be an arbitrary probability measure on* (Ω,F). *Then*

$$K\left(R,P_u\right) = (1-u)\,K\left(R,P\right) + uK\left(R,P'\right) + \log H\left(1-u;P,P'\right). \tag{3.3}$$

Proof We may assume that $R \ll P_u$; otherwise, $K\left(R,P_u\right)=\infty$ and either $K\left(R,P\right)=\infty$ (if R is not absolutely continuous with respect to P) or $K\left(R,P'\right)=\infty$ (if R is not absolutely continuous with respect to P').

Put $z_R \doteq dR/dQ$. Then $\{z_R > 0\} \subseteq \{zz' > 0\}$ Q-a.s. and

$$\begin{aligned} K(R, P_u) &= E_Q\left(z_R \log \frac{z_R H(1-u; P, P')}{z^{1-u} z'^u} \mathbf{1}_{\{z_R>0\}}\right) = E_Q\left(z_R \log \left(\frac{z_R}{z}\right)^{1-u} \mathbf{1}_{\{z_R>0\}}\right) \\ &\quad + E_Q\left(z_R \log \left(\frac{z_R}{z'}\right)^{u} \mathbf{1}_{\{z_R>0\}}\right) + E_Q\left(z_R \log H\left(1-u; P, P'\right) \mathbf{1}_{\{z_R>0\}}\right) \\ &= (1-u) K(R, P) + uK\left(R, P'\right) + \log H\left(1-u; P, P'\right). \end{aligned}$$

The claim is proved. ■

Corollary 3.1 *With the above notation*

1 $(1-u) K(P_u, P) + uK(P_u, P') + \log H(1-u; P, P') = 0.$

2 *If* $(1-u) K(R, P) + uK(R, P') + \log H(1-u; P, P') \leq 0$, *then* $R = P_u$.

Lemma 3.3 *Let* P, P' *and* R *be arbitrary probability measures on* (Ω, F) *and fix* $\alpha, u \in (0,1)$. *Then*

$$H(\alpha; P, P') \geq \begin{cases} \dfrac{H\left(\frac{1-\alpha-u}{1-u}; P', R\right)^{\frac{(1-\alpha)(1-u)}{1-\alpha-u}}}{H\left(-\frac{1-\alpha-u}{u}; P, R\right)^{\frac{\alpha u}{1-\alpha-u}}} & \text{if} \quad \alpha + u < 1 \\ \exp[-(1-u)K(R,P) - uK(R,P')] & \text{if} \quad \alpha + u = 1 \\ \dfrac{H\left(\frac{\alpha+u-1}{u}; P, R\right)^{\frac{\alpha u}{\alpha+u-1}}}{H\left(-\frac{\alpha+u-1}{1-u}; P', R\right)^{\frac{(1-u)(1-\alpha)}{\alpha+u-1}}} & \text{if} \quad \alpha + u > 1 \end{cases} \tag{3.4}$$

The equality holds if and only if either $P \perp P'$ *or* $R = G(u; P, P')$.

Remark 3.1 If $H(-\varepsilon; P, R) < \infty$ (resp. $H(-\varepsilon; P', R) < \infty$) for some $\varepsilon > 0$, then the inequality for $\alpha + u = 1$ can be proved by taking limit in the inequality for $\alpha + u < 1$ (resp. $\alpha + u > 1$) as $\alpha \uparrow 1 - u$ (resp. $\alpha \downarrow 1 - u$). But in general it may happen that the right-hand side of the inequality is equal to 0 for all $\alpha < 1 - u$ and/or $\alpha > 1 - u$, while it is strictly positive for $\alpha = 1 - u$.

Proof The case $\alpha + u = 1$ was considered earlier. The case $\alpha + u > 1$ reduces to the case $\alpha + u < 1$ by interchanging P and P', α and $1 - \alpha$, u and $1 - u$.
So let $\alpha + u < 1$. Assume that $P \perp P'$, then $H(\alpha; P, P') = 0$. If $R \perp P'$, then $H\left(\frac{1-\alpha-u}{1-u}; P', R\right) = 0$, otherwise R is not absolutely continuous with respect to P and

hence $H\left(-\frac{1-\alpha-u}{u};P,R\right)=\infty$; in the both cases the right-hand side of the inequality is also equal to 0.

Let us now consider the case where P and P' are not singular. We shall also assume that $H\left(-\frac{1-\alpha-u}{u};P,R\right)<\infty$, otherwise the statement is trivial. In particular, we have then that $R\ll P$. Set $Q=(P+P')/2$ and put $z=dP/dQ$, $z'=dP'/dQ$ and $z_R=dR/dQ$. We have $\{z=0\}\subseteq\{z_R=0\}$ Q-a.s., therefore by Hölder's inequality

$$\begin{aligned}
H\left(\frac{1-\alpha-u}{1-u};P',R\right) &= E_Q\left[z'^{\frac{1-\alpha-u}{1-u}}z_R^{\frac{\alpha}{1-u}}\right]=E_Q\left[z'^{\frac{1-\alpha-u}{1-u}}z_R^{\frac{\alpha}{1-u}}\mathbf{1}_{\{z>0\}}\right]\\
&= E_Q\left[z'^{\frac{1-\alpha-u}{1-u}}z^{\frac{\alpha(1-\alpha-u)}{(1-\alpha)(1-u)}}z^{-\frac{\alpha(1-\alpha-u)}{(1-\alpha)(1-u)}}z_R^{\frac{\alpha}{1-u}}\mathbf{1}_{\{z>0\}}\right]\\
\leq \left[E_Q\left(z'^{1-\alpha}z^{\alpha}\right)\right]^{\frac{1-\alpha-u}{(1-\alpha)(1-u)}}&\left[E_Q\left(z^{-\frac{1-\alpha-u}{u}}z_R^{\frac{1-\alpha}{u}}\mathbf{1}_{\{z>0\}}\right)\right]^{\frac{\alpha u}{(1-\alpha)(1-u)}}\\
&= \left[H\left(\alpha;P,P'\right)\right]^{\frac{1-\alpha-u}{(1-\alpha)(1-u)}}\left[H\left(-\frac{1-\alpha-u}{u};P,R\right)\right]^{\frac{\alpha u}{(1-\alpha)(1-u)}}.
\end{aligned}$$

Moreover, the equality holds if and only if

$$z^{-\frac{1-\alpha-u}{u}}z_R^{\frac{1-\alpha}{u}}\mathbf{1}_{\{z>0\}}=Cz^{\alpha}z'^{1-\alpha}\quad Q\text{-a.s.}$$

i.e. if and only if $z_R=Cz^{\alpha}z'^{1-\alpha}$ Q-a.s. ■

4 Characterization of E-experiments

1. (*Characterization of E-experiments*) Let $\mathbb{E}$ be any experiment. From Lemma 3.3 we obtain the following characterization.

Theorem 4.1 *The following are equivalent*:

a $\mathbb{E}$ *is an E-experiment.*

b *For any $\theta,\eta\in\Theta$, $u\in(0,1)$ and for any $\alpha\in(0,1)$ we have equality in* (3.4) *with $P=P_\theta$, $P'=P_\eta$ and $R=P_{(1-u)\theta+u\eta}$.*

c *For any $\theta,\eta\in\Theta$, $u\in(0,1)$ and for some $\alpha\in(0,1)$ we have equality in* (3.4) *with $P=P_\theta$, $P'=P_\eta$ and $R=P_{(1-u)\theta+u\eta}$.*

Remark 4.1 If $\mathbb{E}$ is a homogeneous experiment, then the sufficient conditions in Theorem 4.1 for an experiment $\mathbb{E}$ to be an E-experiment can be weakened. Let $\Theta\subset\mathbb{R}$. It is not necessary to verify the condition **c** for all triples $(\theta, \eta, (1-u)\theta + u\eta)$, $\theta, \eta \in \Theta$, $u \in (0,1)$. It is sufficient to verify it only for such triples, that contain a fixed arbitrary pair of separated points or for triples, where the middle point is fixed.

According to Proposition 2.2 an E-experiment is completely described by its binary subexperiments. But Theorem 4.1 says that the property of being an E-experiment is also a property of binary subexperiments; thus we obtain

Theorem 4.2 *Let $\mathbb{E}$ be an E-experiment and $\mathbb{E}'$ any experiment. Suppose that $\mathbb{E} \overset{(2)}{\sim} \mathbb{E}'$. Then $\mathbb{E} \sim \mathbb{E}'$.*

2. (*Characterization of Gaussian shifts*) Since Gaussian shifts are the most important example of E-experiments, we reformulate Theorem 4.2 for them.

Theorem 4.3 *Assume that $\Theta \subset H$, where H is a Hilbert space with norm $\|\cdot\|$ and consider the experiment $\mathbb{E}(\Theta) \doteq \big(\Omega, F, (P_\theta)_{\theta\in\Theta}\big)$. The following are equivalent:*

a *For any $\theta, \eta \in \Theta$*

$$\mathcal{L}\left(\log \frac{dP_\eta}{P_\theta} \,\middle|\, P_\theta\right) = N\left(-\frac{\|\eta-\theta\|^2}{2}, \|\eta-\theta\|^2\right).$$

b *$\mathbb{E}$ is equivalent to the restriction to Θ of a Gaussian shift experiment on H.*

Remark 4.2 If, in addition to **a**, $\operatorname{aff}\Theta$ is dense in H, then one can construct a Gaussian shift on H whose restriction to Θ is $\mathbb{E}$.

Remark 4.3 One can further simplify the condition **a** in Theorem 4.3, similarly to Remark 4.1.

5 On convergence to exponential limits

1. (*Weak convergence of experiments*) In this section we deal with the weak convergence of experiments to E-experiments. For the notion of weak convergence, denoted by $\xrightarrow{w}$, we refer to [9, p. 302] or [3, p. 15]. Recall that $\mathbb{E}^n \xrightarrow{w} \mathbb{E}$ iff $H(\alpha;\mathbb{E}^n) \to H(\alpha;\mathbb{E})$ for all $\alpha \in \mathcal{A}_+$ ([9, p. 302]) and that from every sequence (or net) of experiments it is possible to extract a weakly convergent subnet ([9, p. 298]).
In this section we formulate the results for sequences of experiments. Note that in the next two subsections all the results could be as well formulated for nets of experiments without any changes in the proofs.

2. (*E-experiments as cluster points*) Recall that Θ is a convex set in a real linear space V. Let $\mathbb{E}^n\,(\Theta^n)$, $n \in \mathbb{N}$, $\Theta^n \uparrow \Theta$, be a sequence of statistical experiments.
In this subsection we give necessary and sufficient conditions for the cluster points of the sequence $(\mathbb{E}^n)_{n\in\mathbb{N}}$ to be E-experiments. As it can be expected from the previous considerations, these conditions are expressed in terms of binary subexperiments only.

Theorem 5.1 *Any cluster point of the sequence $\mathbb{E}^n$ is an E-experiment if and only if, for any $\theta, \theta' \in \Theta$, $u \in (0,1)$, we have*

$$\limsup_n \Bigg(\frac{1-(1-u)\,H\left(\alpha; P^n_\theta, P^n_{(1-u)\theta+u\theta'}\right) - uH\left(\alpha; P^n_{\theta'}, P^n_{(1-u)\theta+u\theta'}\right)}{\alpha} + \log H\left(1-u; P^n_\theta, P^n_{\theta'}\right)\Bigg) \le 0 \tag{5.5}$$

for all $\alpha \in (0,1)$.

Remark 5.1 The expression on the left in (5.5) is increasing when α decreases to 0 by Lemma 3.1, **1**. Thus (5.5) holds for all $\alpha \in (0,1)$ (or for sufficiently small α) if and only if

$$\lim_{\alpha\downarrow 0}\limsup_n \Bigg(\frac{1-(1-u)\,H\left(\alpha; P^n_\theta, P^n_{(1-u)\theta+u\theta'}\right) - uH\left(\alpha; P^n_{\theta'}, P^n_{(1-u)\theta+u\theta'}\right)}{\alpha} + \log H\left(1-u; P^n_\theta, P^n_{\theta'}\right)\Bigg) \le 0. \tag{5.6}$$

Proof Sufficiency. Let $\mathbb{E} = \left(\Omega, F, (P_\theta)_{\theta\in\Theta}\right)$ be any cluster point of the sequence $\mathbb{E}^n$, i.e. there is a subnet $(\mathbb{E}^{n_\lambda})_{\lambda\in L}$, where L is a directed set, such that

$$\mathbb{E}^{n_\lambda} \xrightarrow{w} \mathbb{E}.$$

Fix now $\theta, \theta' \in \Theta$ such that P_θ and $P_{\theta'}$ are not singular, and $u \in (0,1)$. Then, for any $\alpha \in (0,1)$,

$$\begin{aligned}
\lim_\lambda H\left(\alpha; P^{n_\lambda}_\theta, P^{n_\lambda}_{(1-u)\theta+u\theta'}\right) &= H\left(\alpha; P_\theta, P_{(1-u)\theta+u\theta'}\right),\\
\lim_\lambda H\left(\alpha; P^{n_\lambda}_{\theta'}, P^{n_\lambda}_{(1-u)\theta+u\theta'}\right) &= H\left(\alpha; P_{\theta'}, P_{(1-u)\theta+u\theta'}\right),\\
\lim_\lambda H\left(1-u; P^{n_\lambda}_\theta, P^{n_\lambda}_{\theta'}\right) &= H\left(1-u; P_\theta, P_{\theta'}\right),
\end{aligned}$$

so we have

$$\frac{1-(1-u)\,H\left(\alpha; P_\theta, P_{(1-u)\theta+u\theta'}\right) - uH\left(\alpha; P_{\theta'}, P_{(1-u)\theta+u\theta'}\right)}{\alpha} + \log H\left(1-u; P_\theta, P_{\theta'}\right) \le 0$$

for any $\alpha > 0$ by (5.5). Passing to the limit in the last inequality as $\alpha \downarrow 0$ we obtain in view of Lemma 3.1, **2**,

$$(1-u)\, K\left(P_{(1-u)\theta+u\theta'}, P_\theta\right) + uK\left(P_{(1-u)\theta+u\theta'}, P_{\theta'}\right) + \log H\left(1-u; P_\theta, P_{\theta'}\right) \leq 0,$$

which implies $P_{(1-u)\theta+u\theta'} = G(u; P_\theta, P_{\theta'})$ by Corollary 3.1, **2**. Hence the result.

Necessity. Assume that (5.5) is violated for some $\theta, \theta' \in \Theta$, $u \in (0,1)$, $\alpha \in (0,1)$. Then there is a subnet $(\mathbb{E}^{n_\lambda})_{\lambda \in L}$ such that

$$\lim_\lambda \left(\frac{1-(1-u)\, H\left(\alpha; P_\theta^{n_\lambda}, P_{(1-u)\theta+u\theta'}^{n_\lambda}\right) - uH\left(\alpha; P_{\theta'}^{n_\lambda}, P_{(1-u)\theta+u\theta'}^{n_\lambda}\right)}{\alpha} + \log H\left(1-u; P_\theta^{n_\lambda}, P_{\theta'}^{n_\lambda}\right) \right) > 0.$$

By weak compactness, take any convergent subnet of $(\mathbb{E}^{n_\lambda})$ and let $\mathbb{E} = \left(\Omega, F, (P_\theta)_{\theta\in\Theta}\right)$ be the limit experiment. Our assumption implies

$$\frac{1-(1-u)\, H\left(\alpha; P_\theta, P_{(1-u)\theta+u\theta'}\right) - uH\left(\alpha; P_{\theta'}, P_{(1-u)\theta+u\theta'}\right)}{\alpha} + \log H\left(1-u; P_\theta, P_{\theta'}\right) > 0$$

(in particular, $\log H\left(1-u; P_\theta, P_{\theta'}\right) > -\infty$ and P_θ and $P_{\theta'}$ are not singular). The left-hand side of the previous inequality is majorized by

$$(1-u)\, K\left(P_{(1-u)\theta+u\theta'}, P_\theta\right) + uK\left(P_{(1-u)\theta+u\theta'}, P_{\theta'}\right) + \log H\left(1-u; P_\theta, P_{\theta'}\right),$$

and by Corollary 3.1, **1**, $\mathbb{E}$ is not an E-experiment. We have arrived to a contradiction, and the necessity is proved. ■

Corollary 5.1 *Assume that, for any $\theta, \theta' \in \Theta$, $u \in (0,1)$*

$$\limsup_n \left\{ (1-u)\, K(P_{(1-u)\theta+u\theta'}^n, P_\theta^n) + uK(P_{(1-u)\theta+u\theta'}^n, P_{\theta'}^n) + \log H\left(1-u; P_\theta^n, P_{\theta'}^n\right) \right\} \leq 0.$$

Then any cluster point of the sequence $\mathbb{E}^n$ is an E-experiment.

Corollary 5.2 *Assume that $\mathbb{E}^n \xrightarrow{w} \mathbb{E}$, and, for any $\theta, \theta' \in \Theta$, $u \in (0,1)$*

$$\liminf_n \left\{ (1-u)\, K(P_{(1-u)\theta+u\theta'}^n, P_\theta^n) + uK(P_{(1-u)\theta+u\theta'}^n, P_{\theta'}^n) + \log H\left(1-u; P_\theta^n, P_{\theta'}^n\right) \right\} \leq 0.$$

Then $\mathbb{E}$ is an E-experiment.

Proofs The assertions follow from Theorem 5.1 and Lemma 3.1.

Corollary 5.3 *E-experiments are weakly compact.*

3. (*Convergence to E-experiments*) Let $(\mathbb{E}^n)_{n\in\mathbb{N}}$ be a sequence of experiments on Θ^n, where $\Theta^n \uparrow \Theta$. Here we study conditions for the weak convergence of $\mathbb{E}^n$ to an E-experiment $\mathbb{E}$.

Theorem 5.2 *Let* $\mathbb{E} = (\Omega, F, (P_\theta)_{\theta\in\Theta})$ *be an* E*-experiment. Then*

$$\mathbb{E}^n \overset{w}{\to} \mathbb{E} \qquad \textit{iff} \qquad (P_\theta^n, P_\eta^n) \overset{w}{\to} (P_\theta, P_\eta) \quad \textit{for all} \quad \theta, \eta \in \Theta.$$

Proof The proof follows in an obvious way from Theorem 4.2 using weak compactness arguments. ■

Let $(\mathbb{E}^n)_{n\in\mathbb{N}}$ be a sequence of experiments on H^n, where $H^n \uparrow H$ and H is a Hilbert space with norm $\|\cdot\|$. Recall that the sequence $\mathbb{E}^n$ is asymptotically normal, if it converges weakly to a Gaussian shift on H [9, p. 409].

Corollary 5.4 *A sequence* $\mathbb{E}^n$ *is asymptotically normal if and only if for any* $\theta, \eta \in H$

$$\mathcal{L}\left(\log \frac{dP_\eta^n}{P_\theta^n} \;\middle|\; P_\theta^n\right) \Rightarrow N\left(-\frac{\|\eta-\theta\|^2}{2}, \|\eta-\theta\|^2\right). \tag{5.7}$$

The results obtained in previous sections allow us to formulate other necessary and sufficient conditions on the convergence to an E-experiment. Recall that the sequence $\mathbb{E}^n$ is said to be contiguous, if for every pair $\theta, \eta \in \Theta$, the sequences of measures $(P_\theta^n)_{n\geq 1}$ and $(P_\eta^n)_{n\geq 1}$ are mutually contiguous [9, p.305].

Theorem 5.3 *Let* $\mathbb{E} = (\Omega, F, (P_\theta)_{\theta\in\Theta})$ *be a homogeneous* E*-experiment. Assume that* I *is a subset of* Θ *such that its affine hull* aff I *contains* Θ*. Then* $\mathbb{E}^n \overset{w}{\to} \mathbb{E}$ *if and only if the sequence* $\mathbb{E}^n$ *is contiguous, satisfies the conditions of Theorem* 5.1 *and*

$$\mathbb{E}_I^n \overset{w}{\to} \mathbb{E}_I.$$

Remark 5.2 There are modifications of Theorem 5.3 that might be useful. For instance: if $\mathbb{E}^n$ is contiguous, satisfies the conditions of Theorem 5.1 and $\mathbb{E}_I^n \xrightarrow{w} \mathbb{E}'$, where $\mathbb{E}' = \mathbb{E}'(I)$, then there is a homogeneous E-experiment $\mathbb{E} = \mathbb{E}(\Theta)$ such that $\mathbb{E}_I = \mathbb{E}'$ and $\mathbb{E}^n \xrightarrow{w} \mathbb{E}$.

Proof The proof follows from Theorem 5.1 and Corollary 2.1 using weak compactness arguments. ■

Corollary 5.5 *Let $H = \mathbb{R}^k$ and let $I = \{\theta_0, \dots, \theta_k\}$ be an affine basis for $\mathbb{R}^k$. The sequence $\mathbb{E}^n$ is asymptotically normal if and only if it is contiguous, satisfies the conditions of Theorem* 5.1 *and*

$$\mathbb{E}_I^n \xrightarrow{w} \mathbb{E}_I,$$

where $\mathbb{E}$ is a Gaussian shift experiment on $\mathbb{R}^k$.

4. (*Approximation by E-experiments*) Let $(\mathbb{E}^n)_{n\geq 1}$, where $\mathbb{E}^n = (\Omega^n, F^n, (P_\theta^n)_{\theta\in\Theta^n})$, be a contiguous sequence of experiments satisfying the conditions of Theorem 5.1; here $\Theta^n \uparrow \Theta \subseteq \mathbb{R}^k$. In this subsection we study the question when the sequence $\mathbb{E}^n$ can be approximated in a certain sense by a sequence $(\widetilde{\mathbb{E}}^n)_{n\geq 1}$ of E-experiments. Approximations by exponential families are well known in the case of local asymptotic normality.
We start with the following simple observation. Under the above assumptions, $\theta, \theta' \in \Theta^n$ and the measures P_θ^n and $P_{\theta'}^n$ are not singular for n large enough for an arbitrary pair $\theta, \theta' \in \Theta$. Thus we can construct the geodesic curve $G(u; P_\theta^n, P_{\theta'}^n)_{u\in(0,1)}$ for such n. It turns out that

$$\lim_{n\to\infty} \|P_{(1-u)\theta+u\theta'}^n - G(u; P_\theta^n, P_{\theta'}^n)\| = 0, \tag{5.8}$$

where $\|P - P'\|$ is the variational distance between the two probability measures P, P'. This follows easily from Theorem 5.1, Corollary 2.1 and weak compactness; see also the proof of Theorem 5.4 below. (In fact (5.6) and (5.8) are equivalent under the only assumption that $\limsup_{n\to\infty} \|P_\theta^n - P_{\theta'}^n\| < 2$.)
Therefore, if $\Theta' \subset \Theta$ and Θ' is the convex hull of a two-point set, we can construct a sequence $\widetilde{\mathbb{E}}^n = (\Omega^n, F^n, (\widetilde{P}_\theta^n)_{\theta\in\Theta'})$ of E-experiments such that

$$\lim_n \|\widetilde{P}_\theta^n - P_\theta^n\| = 0, \quad \theta \in \Theta',$$

i.e. $\widetilde{\mathbb{E}}^n$ approximates $\mathbb{E}^n$ on Θ'. Completely the same arguments show that this statement is true if Θ' is the convex hull of a finite set.
A natural question arises whether one can approximate $\mathbb{E}^n$ on the whole parameter set Θ by E-experiments. The complete answer to this question is not at our disposal. However,

when Θ is the convex hull of a countable set, in particular, if $\Theta \subseteq \mathbb{R}^1$ or Θ is an open set (e.g. $\Theta = \mathbb{R}^k$), then the approximation is possible, at least in a slightly weaker form: we either define $\widetilde{\mathbb{E}}^n$ with a parameter set $\widetilde{\Theta}^n$, $\widetilde{\Theta}^n \uparrow \Theta$, or extend the original experiments (as below), and approximate the extensions $\widehat{\mathbb{E}}^n \sim \mathbb{E}^n$ by E-experiments. We will omit the details of this approximation.

Another approach is possible, if the sequence $\mathbb{E}^n$ weakly converges to a homogeneous E-experiment $\mathbb{E}$. It is based on the fact that the desired kind of approximation always takes place if we also have so-called λ-convergence of $\mathbb{E}^n$ to $\mathbb{E}$. This convergence, introduced by Spokoĭnyĭ and Shiryaev [8] (a complete exposition will appear in [6]), plays an intermediate role between the strong and the weak convergence of experiments. In the proof below we just check that the weak convergence to a homogeneous E-experiment with a finite dimensional parameter set Θ implies the λ-convergence.

Recall that an experiment $\widehat{\mathbb{E}} = \left(\widehat{\Omega}, \widehat{F}, \left(\widehat{P}_\theta\right)_{\theta\in\Theta}\right)$ is an extension of an experiment $\mathbb{E} = \left(\Omega, F, (P_\theta)_{\theta\in\Theta}\right)$ if $\widehat{\Omega} = \Omega \times \Omega'$, $\widehat{F} = F \otimes F'$, $\widehat{P}_\theta(A \times B) = \int_A M(\omega, B) P_\theta(d\omega)$, $A \in F$, $B \in F'$, $\theta \in \Theta$, where (Ω', F') is a measurable space and $M(\omega, B)$ is a Markov kernel from (Ω, F) to (Ω', F'). If $\widehat{\mathbb{E}}$ is an extension of $\mathbb{E}$ then $\widehat{\mathbb{E}} \sim \mathbb{E}$.

Theorem 5.4 *Let* $(\mathbb{E}^n)_{n\geq 1}$, $\mathbb{E}^n = \left(\Omega^n, F^n, (P^n_\theta)_{\theta\in\Theta^n}\right)$, *be a sequence of experiments on* Θ^n, *where* $\Theta^n \uparrow \Theta \subseteq \mathbb{R}^k$. *Assume* $\mathbb{E}^n \xrightarrow{w} \mathbb{E}$, *where* $\mathbb{E} = \left(\Omega, F, (P_\theta)_{\theta\in\Theta}\right)$ *is a homogeneous* E*-experiment. Then there are extensions* $\widehat{\mathbb{E}}^n = \left(\widehat{\Omega}^n, \widehat{F}^n, \left(\widehat{P}^n_\theta\right)_{\theta\in\Theta^n}\right)$ *of* $\mathbb{E}^n$ *and a sequence* $\widetilde{\mathbb{E}}^n = \left(\widehat{\Omega}^n, \widehat{F}^n, \left(\widetilde{P}^n_\theta\right)_{\theta\in\Theta}\right)$, $n \geq 1$, *of* E*-experiments such that* $\widetilde{\mathbb{E}}^n \sim \mathbb{E}$, $n \geq 1$, *and*

$$\lim_n \|\widetilde{P}^n_\theta - \widehat{P}^n_\theta\| = 0, \qquad \theta \in \Theta.$$

Proof Let $I = \{\theta_0, \ldots, \theta_m\} \subset \Theta$, $m \leq k$, be an affine basis for aff Θ. Below we assume that n is large enough, namely $I \subset \Theta^n$.

Denote $z_\theta = dP_\theta / dP_{\theta_0}$ and let z^n_θ be the density of the absolutely continuous part of P^n_θ with respect to $P^n_{\theta_0}$. For a given θ there is a multi-index α such that $I_\alpha \subseteq I$ and $\theta = \sum_{i=0}^m \alpha(\theta)\theta_i$. By Proposition 2.1,

$$z_\theta = \prod_{i=0}^m (z_{\theta_i})^{\alpha(\theta)} \left[H(\alpha; \mathbb{E})\right]^{-1}.$$

Since $\mathcal{L}\left(z^n_\theta, z^n_{\theta_0}, \ldots, z^n_{\theta_m} \mid P^n_{\theta_0}\right) \Rightarrow \mathcal{L}\left(z_\theta, z_{\theta_0}, \ldots, z_{\theta_m} \mid P_{\theta_0}\right)$ by the assumptions, the preceding equality implies

$$z^n_\theta - \prod_{i=0}^m \left(z^n_{\theta_i}\right)^{\alpha(\theta)} \left[H(\alpha; \mathbb{E})\right]^{-1} \xrightarrow{P^n_{\theta_0}} 0.$$

In the terminology of [8], this property means that the experiments $\mathbb{E}^n$, $n \geq 1$, λ-converge to the experiment $\mathbb{E}$. Now our statement follows directly from Theorem and Corollary 1 in [8]. ■

References

[1] Barndorff-Nielsen, O. E., Information and Exponential Families in Statistical Theory, Wiley, New York, 1978.

[2] Čencov, N. N., Statistical Decision Rules and Optimal Inference, Translations of Mathematical Monographs, vol. 53, AMS, Providence, R.I., 1982.

[3] Janssen, A., Milbrodt, H. and Strasser, H., Infinitely Divisible Statistical Experiments, Lecture Notes in Statistics, vol. 27, Springer, Berlin, 1985.

[4] LeCam, L., Asymptotical Methods in Statistical Decision Theory, Springer, Berlin, 1986.

[5] Liese, F. and Vajda, I., Convex Statistical Distances, Teubner, Leipzig, 1987.

[6] Shiryaev, A. N. and Spokoiny, V. G., Statistical Experiments and Decisions: Asymptotic Theory, to appear.

[7] Soler, J.-L., Infinite dimensional exponential type statistical spaces (generalized exponential families). In: Recent Developments in Statistics, J. R. Barra et al. (eds.), North-Holland, Amsterdam, 1977, pp. 269–284.

[8] Spokoĭnyĭ, V. G. and Shiryaev, A. N., On the concept of the λ-convergence of statistical experiments, Proc. Steklov Inst. Math. vol. 202, no. 4, 1994, pp. 225–228.

[9] Strasser, H., Mathematical Theory of Statistics, de Gruyter, Berlin, 1985.

[10] Torgersen, E., Comparison of Statistical Experiments, Cambridge University Press, Cambridge, 1991.