Saddlepoint Approximations of Cumulative Distribution Functions of Sums of Random Vectors

Dadja Anade, Jean-Marie Gorce, Philippe Mary, Samir M. Perlaza

Saddlepoint Approximations of Cumulative Distribution Functions of Sums of Random Vectors

Dadja Anade, Jean-Marie Gorce, Philippe Mary, and
Samir M. Perlaza

ISSN 0249-6399   ISRN INRIA/RR--9388--FR+ENG

RESEARCH REPORT
N° 9388
February 2021
Project-Teams MARACAS and NEO

Saddlepoint Approximations of Cumulative Distribution Functions of Sums of Random Vectors

Dadja Anade, Jean-Marie Gorce, Philippe Mary, and Samir M. Perlaza

Project-Teams MARACAS and NEO

Research Report n° 9388 — version 1 — initial version February 2021 — revised version Mars 2021 — 30 pages

Abstract: In this report, a real-valued function that approximates the cumulative distribution function (CDF) of a finite sum of real-valued independent and identically distributed random vectors is presented. The approximation error is upper bounded and thus, as a byproduct, an upper bound and a lower bound on the CDF are obtained. Finally, it is observed that in the case of lattice and absolutely continuous random variables, the proposed approximation is identical to the saddlepoint approximation of the CDF.

Key-words: Sums of independent random vectors, Saddlepoint approximations, Gaussian approximations

---

Dadja Anade and Jean-Marie Gorce are with the Laboratoire CITI, a joint laboratory between the Institut National de Recherche en Informatique et en Automatique (INRIA), the Université de Lyon and the Institut National de Sciences Apliquées (INSA) de Lyon. 6 Av. des Arts 69621 Villeurbanne, France. ({dadja.anade-akpo, jean-marie.gorce}@inria.fr)

Philippe Mary is with the Laboratoire IETR and the Institut National de Sciences Apliquées (INSA) de Rennes. (philippe.mary@insa-rennes.fr)

Samir M. Perlaza is with INRIA, Centre de Recherche de Sophia Antipolis - Méditerranée, équipe-projet NEO; and also with the Electrical Engineering Department, Princeton University, Princeton, NJ 08544, USA. (samir.perlaza@inria.fr)

This work was partially funded by the French National Agency for Research (ANR) under grant ANR-16-CE25-0001. Parts of this work have been submitted for publication to the IEEE International Symposium on Information Theory (ISIT), Melbourne, Australia, Jun. 2021.

RESEARCH CENTRE GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Résumé : Dans ce rapport, une fonction qui approxime la fonction de répartition d'une somme de vecteurs aléatoires indépendants et identiquement distribués est présentée. L'erreur d'approximation est majorée, et par consequent, une borne supérieure et une borne inférieure sur la fonction de répartition sont obtenues. Finalement, pour des vecteurs aléatoires absolument continues ou lattices, l'approximation proposée est identique à l'approximation du point de selle de la fonction de répartition.

Mots-clés : Approximation du point de selle, approximation Gaussienne

Contents

1 Notation

The real numbers are denoted by $\mathbb{R}$, and the natural numbers are denoted by $\mathbb{N}$. In particular, $0 \notin \mathbb{N}$. The Borel sigma field on $\mathbb{R}^k$, with $k \in \mathbb{N}$, is denoted by $\mathscr{B}\left(\mathbb{R}^k\right)$. The Lebesgue measure on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$ is denoted by ν_k. Given a discrete set $\mathcal{K}$, the biggest sigma field, i.e., the set of all its subsets, is denoted by $2^{\mathcal{K}}$. The Euclidian norm in $\mathbb{R}^k$ is denoted by $\|\cdot\|$. Given a set $\mathcal{A} \subseteq \mathbb{R}^k$, the closure of the set $\mathcal{A}$, denoted by $\mathrm{clo}\mathcal{A}$, is defined by $\mathrm{clo}\mathcal{A} \triangleq \left\{\boldsymbol{x} \in \mathbb{R}^k : \forall r > 0,\, \exists \boldsymbol{y} \in \mathcal{A},\, ||\boldsymbol{x} - \boldsymbol{y}|| < r\right\}$. A diagonal matrix whose diagonal is the vector $\boldsymbol{x} \in \mathbb{R}^k$ is denoted by $\mathrm{diag}\,(\boldsymbol{x})$.

2 Introduction

Let n be a finite integer, with $n > 1$, and let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ be independent random vectors such that each of them induces the probability measure $P_{\boldsymbol{Y}}$ on the measurable space $\left(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k)\right)$, with $k \in \mathbb{N}$. Denote by $K_{\boldsymbol{Y}} : \mathbb{R}^k \to \mathbb{R}$ the cumulant generating function (CGF) of each of these random variables. That is, for all $\boldsymbol{t} \in \mathbb{R}^k$,

$$K_{\boldsymbol{Y}}(\boldsymbol{t}) = \ln\left(\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\exp\left(\boldsymbol{t}^{\mathsf{T}}\, \boldsymbol{Y}\right)\right]\right). \tag{1}$$

The gradient of the CGF $K_{\boldsymbol{Y}}$ is a function denoted by $K^{(1)}_{\boldsymbol{Y}} : \mathbb{R}^k \to \mathbb{R}^k$. More specifically, for all $\boldsymbol{t} \in \mathbb{R}^k$,

$$K^{(1)}_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}_{P_{\boldsymbol{Y}}}\left[\boldsymbol{Y} \exp\left(\boldsymbol{t}^{\mathsf{T}}\boldsymbol{Y} - K_{\boldsymbol{Y}}(\boldsymbol{t})\right)\right]. \tag{2}$$

The Hessian of the CGF $K_{\boldsymbol{Y}}$ is a function denoted by $K^{(2)}_{\boldsymbol{Y}} : \mathbb{R}^k \to \mathbb{R}^{k \times k}$. That is, for all $\boldsymbol{t} \in \mathbb{R}^k$,

$$K^{(2)}_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}_{P_{\boldsymbol{Y}}}\left[\left(\boldsymbol{Y} - K^{(1)}_{\boldsymbol{Y}}(\boldsymbol{t})\right)\left(\boldsymbol{Y} - K^{(1)}_{\boldsymbol{Y}}(\boldsymbol{t})\right)^{\mathsf{T}} \exp\left(\boldsymbol{t}^{\mathsf{T}}\boldsymbol{Y} - K_{\boldsymbol{Y}}(\boldsymbol{t})\right)\right]. \tag{3}$$

Note that $K^{(2)}_{\boldsymbol{Y}}(\boldsymbol{0})$ is the covariance matrix of the random vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$. In the following, $K^{(2)}_{\boldsymbol{Y}}(\boldsymbol{0})$ is assumed to be positive definite (instead of positive semidefinite).

Let also

$$\boldsymbol{X}_n \triangleq \sum_{t=1}^{n} \boldsymbol{Y}_t \tag{4}$$

be a random vector that induces the probability measure $P_{\boldsymbol{X}_n}$ on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$, with cumulative distribution function (CDF) denoted by $F_{\boldsymbol{X}_n}$.

Often, the calculation of the CDF $F_{\boldsymbol{X}_n}$ requires elaborated numerical methods. From this perspective, approximations to the CDF $F_{\boldsymbol{X}_n}$, e.g., Gaussian approximations and saddlepoint approximations [1–4], are rather popular in the realm of applied mathematics. In the particular case of information theory, Gaussian and saddlepoint approximations play central roles in the approximation of the fundamental limits of data transmission, c.f., [5–12].

When for all $i \in \{1, 2, \ldots, n\}$ the random vector $\boldsymbol{Y}_i$ in (4) is absolutely continuous and its corresponding CGF $K_{\boldsymbol{Y}}$ is such that the set

$$\mathcal{C}_{\boldsymbol{Y}} \triangleq \left\{\boldsymbol{\theta} \in \mathbb{R}^k : K_{\boldsymbol{Y}}(\boldsymbol{\theta}) < \infty\right\} \cap]-\infty, 0[^k \tag{5}$$

is not empty, the CDF $F_{\boldsymbol{X}_n}$ can be written as a complex integral [1]. In particular, for all $\boldsymbol{x} \in \mathbb{R}^k$,

$$F_{\boldsymbol{X}_n}(\boldsymbol{x}) = \int_{\boldsymbol{c}-\mathrm{i}\,\boldsymbol{e}}^{\boldsymbol{c}+\mathrm{i}\,\boldsymbol{e}} \frac{\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\tau}) - \boldsymbol{\tau}^{\mathsf{T}}\boldsymbol{x}\right)}{(2\pi\mathrm{i})^k \prod_{t=1}^{k} \tau_t} \mathrm{d}\boldsymbol{\tau}, \tag{6}$$

where i is the imaginary unit; $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_k)$; the constant $\boldsymbol{c}$ is arbitrarily chosen to satisfy $\boldsymbol{c} \in \mathcal{C}_{\boldsymbol{Y}}$; and the vector $\boldsymbol{e} = (e_1, e_2, \ldots, e_k)$ is such that for all $t \in \{1, 2, \ldots, k\}$, $e_t = +\infty$.

The complex integral in (6) results from the multivariate Laplace inverse transform [13], and can be approximated with high precision, as shown hereunder. Denote by $\mathcal{D}$ the following set

$$\mathcal{D} \triangleq \left\{\boldsymbol{u} \in \mathbb{R}^k : \exists \boldsymbol{t} \in]-\infty, 0[^k,\; nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{t}) = \boldsymbol{u}\right\}, \tag{7}$$

and denote by $\boldsymbol{\tau}_0 \in \mathbb{R}^k$ the unique solution in $\boldsymbol{\tau}$ to

$$K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\tau}) = \frac{1}{n}\boldsymbol{x}. \tag{8}$$

For all $\boldsymbol{x} \in \mathcal{D}$, a Taylor series expansion of $nK_{\boldsymbol{Y}}(\boldsymbol{\tau}) - \boldsymbol{\tau}^{\mathsf{T}}\boldsymbol{x}$ in the neighborhood of $\boldsymbol{\tau}_0$, leads to the following asymptotic expansion of the integral in (6):

$$F_{\boldsymbol{X}_n}(\boldsymbol{x}) = \hat{F}_{\boldsymbol{X}_n}(\boldsymbol{x}) + O\left(\frac{\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\tau}_0) - \boldsymbol{\tau}_0^{\mathsf{T}}\,\boldsymbol{x}\right)}{\sqrt{n}}\right), \tag{9}$$

where the function $\hat{F}_{\boldsymbol{X}_n} : \mathcal{D} \to \mathbb{R}$ is

$$\hat{F}_{\boldsymbol{X}_n}(\boldsymbol{x}) = \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\tau}_0) - \boldsymbol{\tau}_0^{\mathsf{T}}\,\boldsymbol{x} + \frac{n\boldsymbol{\tau}_0^{\mathsf{T}}\,K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\,\boldsymbol{\tau}_0}{2}\right) F_{\boldsymbol{G}_n^{(\boldsymbol{\tau}_0)}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_0\right), \tag{10}$$

and the function $F_{\boldsymbol{G}_n^{(\boldsymbol{\tau}_0)}} : \mathbb{R}^k \to [0, 1]$ is the CDF of a Gaussian random vector with mean vector $(0, 0, \ldots, 0)$ and covariance matrix $nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)$.
The vector $\boldsymbol{\tau}_0$ and the function $\hat{F}_{\boldsymbol{X}_n}$ in (10) are respectively referred to as the *saddlepoint* and the *saddlepoint approximation* of the CDF $F_{\boldsymbol{X}_n}$. In [1], it is shown that the approximation $\hat{F}_{\boldsymbol{X}_n}$ in (10) also holds for the case in which for all $i \in \{1, 2, \ldots, n\}$ the vector $\boldsymbol{Y}_i$ in (4) is a lattice random vector. Moreover, when for all $i \in \{1, 2, \ldots, n\}$ the random vector $\boldsymbol{Y}_i$ in (4) is a Gaussian random vector, then *saddlepoint approximation* is exact. That is, $\hat{F}_{\boldsymbol{X}_n}$ and $F_{\boldsymbol{X}_n}$ are identical.
The main drawback of saddlepoint approximations, despite their well known precision, c.f., [3] and [4], is that the approximation error lacks of a tight upper bound. This is the main motivation of this work, whose main contributions are:

(a) A real-valued function that approximates the CDF $F_{\boldsymbol{X}_n}$ is presented. This approximation turns out to be identical to the saddlepoint approximation $\hat{F}_{\boldsymbol{X}_n}$ in (9) when $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, $\ldots$, $\boldsymbol{Y}_n$ are either absolutely continuous or lattices random vectors; and

(b) an upper bound on the error induced by the proposed approximation is also presented. The asymptotic behaviour with n of the proposed upper bound is consistent with the one suggested in (9).

This work is structured as follows. Section 3 introduces the Gaussian approximation to $F_{\boldsymbol{X}_n}$, which is used as a benchmark. Section 4 introduces the main results of this report, mainly an approximation to the measure $P_{\boldsymbol{X}_n}$; and an approximation to the CDF $F_{\boldsymbol{X}_n}$. Section 5 presents an example and numerical results. Section 6 concludes this work with some final remarks and a discussion on the main results.

3 Gaussian Approximation

Let $\boldsymbol{\mu}_{\boldsymbol{X}_n} \in \mathbb{R}^k$ and $\underline{\boldsymbol{v}}_{\boldsymbol{X}_n} \in \mathbb{R}^{k\times k}$ be the mean vector and covariance matrix of the random vector $\boldsymbol{X}_n$ in (4). The Gaussian approximation of the measure $P_{\boldsymbol{X}_n}$ induced by $\boldsymbol{X}_n$ is the probability measure induced by a Gaussian vector with mean vector $\boldsymbol{\mu}_{\boldsymbol{X}_n}$ and covariance matrix $\underline{\boldsymbol{v}}_{\boldsymbol{X}_n}$. The following theorem, known as the multivariate Berry–Esseen theorem [14], introduces an upper bound on the approximation error.

Theorem 1 ([14, Theorem 1.1]) *Assume that the measure $P_{\boldsymbol{Y}}$ induced by each of the random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_n$ in* (4) *satisfies,*

$$\mathbb{E}_{P_{\boldsymbol{Y}}}[\boldsymbol{Y}] = (0,0,\ldots,0), \text{ and} \tag{11}$$

$$\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{T}}\right] = \frac{1}{n}\mathrm{diag}(1,1,\ldots,1). \tag{12}$$

Let $P_{\boldsymbol{Z}_n}$ be the probability measure induced on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$ by a Gaussian random vector $\boldsymbol{Z}_n$ with mean vector $(0,0,\ldots,0)$ and covariance matrix $\mathrm{diag}(1,1,\ldots,1)$. Then,

$$\sup_{\mathcal{A}\in\mathcal{C}_k} \left|P_{\boldsymbol{X}_n}(\mathcal{A}) - P_{\boldsymbol{Z}_n}(\mathcal{A})\right| \leqslant \min\left(1, c(k)\, n\, \mathbb{E}_{P_{\boldsymbol{Y}}}\left[\|\boldsymbol{Y}\|^3\right]\right), \tag{13}$$

where $\mathcal{C}_k$ is the collection of all convex sets in $\mathscr{B}(\mathbb{R}^k)$; and the function $c : \mathbb{N} \to \mathbb{R}$ satisfies for all $k \in \mathbb{N}$,

$$c(k) = 42k^{\frac{1}{4}} + 16. \tag{14}$$

The measure $P_{\boldsymbol{Z}_n}$ in (13) is often referred to as the Gaussian approximation of $P_{\boldsymbol{X}_n}$. Similarly, $F_{\boldsymbol{Z}_n}$, the CDF of the measure $P_{\boldsymbol{Z}_n}$, is referred to as the Gaussian approximation of $F_{\boldsymbol{X}_n}$. Theorem 1 leads to the following inequalities for all $\boldsymbol{x} \in \mathbb{R}^k$,

$$\underline{\Sigma}(n,\boldsymbol{x}) \leqslant F_{\boldsymbol{X}_n}(\boldsymbol{x}) \leqslant \bar{\Sigma}(n,\boldsymbol{x}), \tag{15}$$

where,

$$\bar{\Sigma}(n,\boldsymbol{x}) \triangleq F_{\boldsymbol{Z}_n}(\boldsymbol{x}) + \min\left(1, c(k)\, n\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\|\boldsymbol{Y}\|^3\right]\right), \text{ and} \tag{16a}$$

$$\underline{\Sigma}(n,\boldsymbol{x}) \triangleq F_{\boldsymbol{Z}_n}(\boldsymbol{x}) - \min\left(1, c(k)\, n\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\|\boldsymbol{Y}\|^3\right]\right). \tag{16b}$$

That is, the functions $\underline{\Sigma}(n,\cdot)$ and $\bar{\Sigma}(n,\cdot)$ are respectively a lower and an upper bound on the CDF $F_{\boldsymbol{X}_n}$.

4 Saddlepoint Approximation

This section introduces two central results. First, given a convex set $\mathcal{A}$ in $\mathscr{B}(\mathbb{R}^k)$, the probability $P_{\boldsymbol{X}_n}(\mathcal{A})$, with $P_{\boldsymbol{X}_n}$ the probability measure induced by the random vector $\boldsymbol{X}_n$ in (4), is approximated by a function that is a measure but not necessary a probability measure. Second, using the first result, the CDF of $\boldsymbol{X}_n$ is approximated by a function that is not necessarily a CDF. Both functions, the one that approximates the measure and the one that approximates the CDF, are parametrized by a vector in $\mathbb{R}^k$ that can be arbitrarily chosen to locally minimize the approximation error. Additionally, an upper bound on the approximation error induced by both functions is provided. As a by product, an upper bound and a lower bound are provided for both the measure $P_{\boldsymbol{X}_n}$ and the CDF $F_{\boldsymbol{X}_n}$.

4.1 Approximation of the Measure

Given $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$, with

$$\Theta_{\boldsymbol{Y}} \triangleq \{\boldsymbol{t} \in \mathbb{R}^k : K_{\boldsymbol{Y}}(\boldsymbol{t}) < \infty\}, \tag{17}$$

let $\boldsymbol{Y}_1^{(\boldsymbol{\theta})}$, $\boldsymbol{Y}_2^{(\boldsymbol{\theta})}$, ..., $\boldsymbol{Y}_n^{(\boldsymbol{\theta})}$ be independent random vectors such that each of them induces the probability measure $P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}$ on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$ that satisfies for all $\boldsymbol{y} \in \mathbb{R}^k$,

$$\frac{\mathrm{d}P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}{\mathrm{d}P_{\boldsymbol{Y}}}(\boldsymbol{y}) = \exp\left(\boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{y} - K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right). \tag{18}$$

That is, the probability measure $P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}$ is an exponentially tilted measure with respect to $P_{\boldsymbol{Y}}$. Denote by $P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}$ and $P_{\boldsymbol{Y}_1\boldsymbol{Y}_2\ldots\boldsymbol{Y}_n}$ the joint probability measures respectively induced by the independent random vectors $\boldsymbol{Y}_1^{(\boldsymbol{\theta})}$, $\boldsymbol{Y}_2^{(\boldsymbol{\theta})}$, ..., $\boldsymbol{Y}_n^{(\boldsymbol{\theta})}$ and $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_n$ in (4) on the measurable space $\left(\mathbb{R}^{k\times n}, \mathscr{B}\left(\mathbb{R}^{k\times n}\right)\right)$. Then, for all $j \in \{1, 2, \ldots, n\}$ and for all $\boldsymbol{y}_j \in \mathbb{R}^k$, it holds that

$$\frac{\mathrm{d}P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}{\mathrm{d}P_{\boldsymbol{Y}_1\boldsymbol{Y}_2\ldots\boldsymbol{Y}_n}}(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_2 \ldots \boldsymbol{y}_n) = \prod_{j=1}^{n} \frac{\mathrm{d}P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}{\mathrm{d}P_{\boldsymbol{Y}}}(\boldsymbol{y}_j) \tag{19}$$

$$= \exp\left(\sum_{j=1}^{n}\left(\boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{y}_j - K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\right). \tag{20}$$

Using this notation, for all $\mathcal{A} \subseteq \mathbb{R}^k$ and for all $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$, with $\Theta_{\boldsymbol{Y}}$ defined in (17), it holds that

$$P_{\boldsymbol{X}_n}(\mathcal{A}) = \mathbb{E}_{P_{\boldsymbol{X}_n}}[\mathbb{1}_{\{\boldsymbol{X}_n \in \mathcal{A}\}}] \tag{21a}$$

$$= \mathbb{E}_{P_{\boldsymbol{Y}_1\boldsymbol{Y}_2\ldots\boldsymbol{Y}_n}}\left[\mathbb{1}_{\left\{\sum_{j=1}^n \boldsymbol{Y}_j \in \mathcal{A}\right\}}\right] \tag{21b}$$

$$= \mathbb{E}_{P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}\left[\mathbb{1}_{\left\{\sum_{j=1}^n \boldsymbol{Y}_j^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}} \frac{\mathrm{d}P_{\boldsymbol{Y}_1\boldsymbol{Y}_2\ldots\boldsymbol{Y}_n}}{\mathrm{d}P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}\left(\boldsymbol{Y}_1^{(\boldsymbol{\theta})}, \boldsymbol{Y}_2^{(\boldsymbol{\theta})}, \ldots, \boldsymbol{Y}_n^{(\boldsymbol{\theta})}\right)\right] \tag{21c}$$

$$= \mathbb{E}_{P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}\left[\mathbb{1}_{\left\{\sum_{j=1}^n \boldsymbol{Y}_j^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}}\left(\frac{\mathrm{d}P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}{\mathrm{d}P_{\boldsymbol{Y}_1\boldsymbol{Y}_2\ldots\boldsymbol{Y}_n}}\left(\boldsymbol{Y}_1^{(\boldsymbol{\theta})}, \boldsymbol{Y}_2^{(\boldsymbol{\theta})}, \ldots, \boldsymbol{Y}_n^{(\boldsymbol{\theta})}\right)\right)^{-1}\right] \tag{21d}$$

$$= \mathbb{E}_{P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}\left[\mathbb{1}_{\left\{\sum_{j=1}^n \boldsymbol{Y}_j^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}} \exp\left(\sum_{j=1}^{n}\left(K_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{Y}_j^{(\boldsymbol{\theta})}\right)\right)\right] \tag{21e}$$

$$= \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right) \mathbb{E}_{P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})}\boldsymbol{Y}_2^{(\boldsymbol{\theta})}\ldots\boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}\left[\mathbb{1}_{\left\{\sum_{j=1}^n \boldsymbol{Y}_j^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \sum_{j=1}^{n} \boldsymbol{Y}_j^{(\boldsymbol{\theta})}\right)\right]. \tag{21f}$$

To ease the notation, consider the random vector

$$\boldsymbol{S}_n^{(\boldsymbol{\theta})} = \sum_{j=1}^{n} \boldsymbol{Y}_j^{(\boldsymbol{\theta})}, \tag{22}$$

which induces the probability measure $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$ on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$. Hence, plugging (22) in (21f) yields,

$$P_{\boldsymbol{X}_n}(\mathcal{A}) = \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right) \mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}}\left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{S}_n^{(\boldsymbol{\theta})}\right) \mathbb{1}_{\left\{\boldsymbol{S}_n^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}}\right]. \tag{23}$$

The equality in (23) is central as it expresses the probability $P_{\boldsymbol{X}_n}(\mathcal{A})$ in terms of another measure $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$, which is the sum of n independent random vectors. From this observation, it follows that an approximation on $P_{\boldsymbol{X}_n}(\mathcal{A})$ can be obtained by arbitrarily replacing $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$ by its Gaussian approximation, i.e., a probability measure $P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}$ induced on the measurable space $\left(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k)\right)$ by a Gaussian random vector $\boldsymbol{Z}_n^{(\boldsymbol{\theta})}$ with the same mean vector and covariance matrix as the random vector $\boldsymbol{S}_n^{(\boldsymbol{\theta})}$ in (22). Denote by $\boldsymbol{\mu}_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}} \in \mathbb{R}^k$ and $\underline{\boldsymbol{v}}_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}} \in \mathbb{R}^{k\times k}$ the mean vector and the covariance matrix of the random vector $\boldsymbol{Z}_n^{(\boldsymbol{\theta})}$, respectively. Hence,

$$\boldsymbol{\mu}_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}} \triangleq \mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}}\left[\boldsymbol{S}_n^{(\boldsymbol{\theta})}\right] \tag{24a}$$

$$= n\mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}\left[\boldsymbol{Y}^{(\boldsymbol{\theta})}\right] \tag{24b}$$

$$= n\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\boldsymbol{Y}\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y} - K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\right] \tag{24c}$$

$$= nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta}), \tag{24d}$$

where $K_{\boldsymbol{Y}}^{(1)}$ is the gradient vector of the CGF $K_{\boldsymbol{Y}}$ defined in (2). Alternatively,

$$\underline{\boldsymbol{v}}_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}} \triangleq \mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}}\left[\left(\boldsymbol{S}_n^{(\boldsymbol{\theta})} - nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\left(\boldsymbol{S}_n^{(\boldsymbol{\theta})} - nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\right] \tag{25a}$$

$$= n\mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}\left[\left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\right] \tag{25b}$$

$$= n\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\left(\boldsymbol{Y} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\left(\boldsymbol{Y} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y} - K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\right] \tag{25c}$$

$$= nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta}), \tag{25d}$$

where $K_{\boldsymbol{Y}}^{(2)}$ is the Hessian matrix of the CGF $K_{\boldsymbol{Y}}$ defined in (3). The equality (25b) is a consequence of the random vector $S_n^{(\boldsymbol{\theta})}$ in (22) being a sum of independent random vectors. Hence, the central idea for providing an approximation to $P_{\boldsymbol{X}_n}$ is to approximate the RHS of (23) by the function $\eta_{\boldsymbol{Y}} : \mathbb{R}^k\times \mathscr{B}(\mathbb{R}^k) \times \mathbb{N} \to \mathbb{R}$, which is such that

$$\eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n) \triangleq \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\mathbb{E}_{P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}}\left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Z}_n^{(\boldsymbol{\theta})}\right)\mathbb{1}_{\left\{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}\in\mathcal{A}\right\}}\right]. \tag{26}$$

Note that $\eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)$ can also be expressed as follows:

$$\eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)$$

$$= \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\int_{\mathbb{R}^k}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{z}\right)\mathbb{1}_{\{\boldsymbol{z}\in\mathcal{A}\}}\mathrm{d}P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}(\boldsymbol{z}) \tag{27a}$$

$$= \int_{\mathcal{A}}\frac{\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{z}\right)}{\sqrt{(2\pi)^k\det\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)}}\exp\left(-\frac{\left(\boldsymbol{z} - nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{-1}\left(\boldsymbol{z} - nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)}{2}\right)\mathrm{d}\nu_k(\boldsymbol{z}) \tag{27b}$$

$$= \frac{\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)}{\sqrt{(2\pi)^k\det\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)}}$$

$$\int_{\mathcal{A}} \exp\left(-\frac{\left(\boldsymbol{z}-nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{-1}\left(\boldsymbol{z}-nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)}{2}\right)\exp\left(-\frac{2\boldsymbol{\theta}^{\mathsf{T}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{z}}{2}\right)\mathrm{d}\nu_k(\boldsymbol{z}) \tag{27c}$$

$$= \frac{\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})+\frac{\boldsymbol{\theta}^{\mathsf{T}}nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\boldsymbol{\theta}}{2}-\boldsymbol{\theta}^{\mathsf{T}}nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)}{\sqrt{(2\pi)^k\det\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)}}$$

$$\int_{\mathcal{A}} \exp\left(-\frac{\left(\boldsymbol{z}-nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})+\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\boldsymbol{\theta}\right)^{\mathsf{T}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{-1}\left(\boldsymbol{z}-nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})+\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\boldsymbol{\theta}\right)}{2}\right)\mathrm{d}\nu_k(\boldsymbol{z}) \tag{27d}$$

$$= \exp\left(n\left(K_{\boldsymbol{Y}}(\boldsymbol{\theta})\frac{\boldsymbol{\theta}^{\mathsf{T}}K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\boldsymbol{\theta}}{2}-\boldsymbol{\theta}^{\mathsf{T}}K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\right)\frac{1}{\sqrt{(2\pi)^k\det\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)}}$$

$$\int_{\mathcal{A}} \exp\left(-\frac{\left(\boldsymbol{z}-nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})+\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\boldsymbol{\theta}\right)^{\mathsf{T}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{-1}\left(\boldsymbol{z}-nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})+\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\boldsymbol{\theta}\right)}{2}\right)\mathrm{d}\nu_k(\boldsymbol{z}) \tag{27e}$$

$$= \exp\left(n\left(K_{\boldsymbol{Y}}(\boldsymbol{\theta})+\frac{\boldsymbol{\theta}^{\mathsf{T}}K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\boldsymbol{\theta}}{2}-\boldsymbol{\theta}^{\mathsf{T}}K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\right)P_{\boldsymbol{H}_n^{(\boldsymbol{\theta})}}(\mathcal{A}) \tag{27f}$$

where the probability measure $P_{\boldsymbol{H}_n^{(\boldsymbol{\theta})}}$ is the probability measure induced by a Gaussian random vector $\boldsymbol{H}_n^{(\boldsymbol{\theta})}$ with mean vector $n\left(K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})-K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\,\boldsymbol{\theta}\right)$ and covariance matrix $nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})$ on the measurable space $\left(\mathbb{R}^k,\mathscr{B}(\mathbb{R}^k)\right)$.

The rest of this section follows by upper bounding the error induced by replacing $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$ by its Gaussian approximation $P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}$ in (23). That is, establishing an upper-bound on

$$\left|P_{\boldsymbol{X}_n}(\mathcal{A})-\eta_{\boldsymbol{Y}}(\boldsymbol{\theta},\mathcal{A},n)\right|, \tag{28}$$

which is the purpose of the following lemma.

Lemma 2 *Given $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_k)\in\Theta_{\boldsymbol{Y}}$, with $\Theta_{\boldsymbol{Y}}$ in (17), and a convex set $\mathcal{A}\in\mathscr{B}(\mathbb{R}^k)$, it holds that*

$$\left|P_{\boldsymbol{X}_n}(\mathcal{A})-\eta_{\boldsymbol{Y}}(\boldsymbol{\theta},\mathcal{A},n)\right|\leqslant\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})-\boldsymbol{\theta}^{T}\boldsymbol{a}\left(\mathcal{A},\boldsymbol{\theta}\right)\right)\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}},P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right), \tag{29}$$

where

$$\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}},P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right)\triangleq\sup_{\mathcal{B}\in\mathcal{C}_k}\left|P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}(\mathcal{B})-P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}(\mathcal{B})\right|; \tag{30}$$

the collection $\mathcal{C}_k$ contains all convex sets in $\mathscr{B}(\mathbb{R}^k)$; and the vector $\boldsymbol{a}\left(\mathcal{A},\boldsymbol{\theta}\right)=(a_1(\mathcal{A},\boldsymbol{\theta}),\ a_2(\mathcal{A},\boldsymbol{\theta}),$

..., $a_k(\mathcal{A}, \boldsymbol{\theta}))$ *is such that for all* $i \in \{1, 2, \ldots, k\}$,

$$a_i(\mathcal{A}, \boldsymbol{\theta}) \triangleq \begin{cases} 0 & \text{if } \theta_i = 0 \\ \inf_{(b_1, b_2, \ldots, b_k) \in \mathcal{A}} b_i & \text{if } \theta_i > 0 \\ \sup_{(b_1, b_2, \ldots, b_k) \in \mathcal{A}} b_i & \text{if } \theta_i < 0. \end{cases} \tag{31}$$

Proof: The proof of Lemma 2 is presented in Appendix A. ∎

Note that the term $\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right)$ in (30) can be upper bounded by using Theorem 1. For doing so, consider the function $\xi_{\boldsymbol{Y}} : \mathbb{R}^k \to \mathbb{R}$, such that for all $\boldsymbol{t} \in \mathbb{R}^k$,

$$\xi_{\boldsymbol{Y}}(\boldsymbol{t}) \triangleq \mathbb{E}_{P_{\boldsymbol{Y}}}\left[\left(\left(\boldsymbol{Y} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{t})\right)^{\mathsf{T}} \left(K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{t})\right)^{-1} \left(\boldsymbol{Y} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{t})\right)\right)^{3/2} \exp(\boldsymbol{t}^{\mathsf{T}}\boldsymbol{Y} - K_{\boldsymbol{Y}}(\boldsymbol{t}))\right]. \tag{32}$$

Using this notation, the following theorem introduces an upper bound on the error induced by the approximation of the probability $P_{\boldsymbol{X}_n}(\mathcal{A})$ by $\eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)$, where $\mathcal{A} \subseteq \mathbb{R}^k$ is a convex Borel measurable set and $\boldsymbol{\theta} \in \mathbb{R}^k$ is a fixed parameter.

Theorem 3 *For all* $\mathcal{A} \in \mathcal{C}_k$, *with* $\mathcal{C}_k$ *the collection of all convex sets in* $\mathscr{B}(\mathbb{R}^k)$, *and for all* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta_{\boldsymbol{Y}}$, *with* $\Theta_{\boldsymbol{Y}}$ *in* (17), *it holds that*

$$|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)| \leqslant \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{a}(\mathcal{A}, \boldsymbol{\theta})\right) \min\left(1, \frac{c(k)\xi_{\boldsymbol{Y}}(\boldsymbol{\theta})}{\sqrt{n}}\right), \tag{33}$$

where the functions c *and* $\eta_{\boldsymbol{Y}}$ *are respectively defined in* (14) *and* (27f)*; the vector* $\boldsymbol{a}(\mathcal{A}, \boldsymbol{\theta}) \triangleq (a_1(\mathcal{A}, \boldsymbol{\theta}), a_2(\mathcal{A}, \boldsymbol{\theta}), \ldots, a_k(\mathcal{A}, \boldsymbol{\theta}))$ *is defined in* (31)*; and the function* $\xi_{\boldsymbol{Y}}$ *is defined in* (32).

Proof: The proof of Theorem 3 is presented in Appendix B. ∎

4.2 Approximation of the CDF

The CDF $F_{\boldsymbol{X}_n}$ can be written in the form of the probability of a convex set in $\mathscr{B}(\mathbb{R}^k)$. That is, for all $\boldsymbol{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, let the set $\mathcal{A}_{\boldsymbol{x}}$ be such that

$$\mathcal{A}_{\boldsymbol{x}} = \left\{(t_1, t_2, \ldots, t_k) \in \mathbb{R}^k : \forall i \in \{1, 2, \ldots, k\}, t_i \leq x_i\right\}. \tag{34}$$

Then, for all $\boldsymbol{x} \in \mathbb{R}^k$, it holds that

$$F_{\boldsymbol{X}_n}(\boldsymbol{x}) = P_{\boldsymbol{X}_n}(\mathcal{A}_{\boldsymbol{x}}). \tag{35}$$

This observation allows to use Theorem 3 to approximate the CDF $F_{\boldsymbol{X}_n}$ of the random vector $\boldsymbol{X}_n$ in (4). Explicitly, for all $\boldsymbol{x} \in \mathbb{R}^k$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{Y}}$, with $\boldsymbol{\Theta}_{\boldsymbol{Y}}$ in (17), it holds that

$$|F_{\boldsymbol{X}_n}(\boldsymbol{x}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}_{\boldsymbol{x}}, n)| \leqslant \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}(\mathcal{A}_{\boldsymbol{x}}, \boldsymbol{\theta})\right) \min\left(1, \frac{c(k)\, \xi_{\boldsymbol{Y}}(\boldsymbol{\theta})}{\sqrt{n}}\right). \tag{36}$$

The approximation of the CDF $F_{\boldsymbol{X}_n}$ in (36) can be enhanced by choosing the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{Y}}$ that minimizes the right-hand side (RHS) of (36). From this standpoint, the parameter $\boldsymbol{\theta}$ must be searched within a subset of $\boldsymbol{\Theta}_{\boldsymbol{Y}}$ in which

$$\boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}(\mathcal{A}_{\boldsymbol{x}}, \boldsymbol{\theta}) = \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{x} < +\infty. \tag{37}$$

More specifically, given $\boldsymbol{A}_{\boldsymbol{x}}$ in (34), it follows from (31) that the minimization must be restricted to the set

$$\boldsymbol{\Theta}_{\boldsymbol{Y}}^{-} \triangleq \{(t_1, t_2, \ldots, t_k) \in \boldsymbol{\Theta}_{\boldsymbol{Y}}, \forall i \in \{1, 2, \ldots, k\},\, t_i \leq 0\}. \tag{38}$$

An arbitrary choice of $\boldsymbol{\theta}$ is the one that minimizes the exponential term $\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}(\mathcal{A}_{\boldsymbol{x}}, \boldsymbol{\theta})\right)$, which depends on $\boldsymbol{x}$. Denote such a choice by $\boldsymbol{\theta}(\boldsymbol{x})$, which is defined in terms of the following quantity:

$$\boldsymbol{\tau}(\boldsymbol{x}) \triangleq \arg\min_{\boldsymbol{t} \in \mathrm{clo}\boldsymbol{\Theta}_{\boldsymbol{Y}}^{-}} \left(nK_{\boldsymbol{Y}}(\boldsymbol{t}) - \boldsymbol{t}^{\mathsf{T}}\, \boldsymbol{x}\right), \tag{39}$$

where $\mathrm{clo}\boldsymbol{\Theta}_{\boldsymbol{Y}}^{-}$ is the closure of $\boldsymbol{\Theta}_{\boldsymbol{Y}}^{-}$. The uniqueness of $\boldsymbol{\tau}(\boldsymbol{x})$ in (39), for a given $\boldsymbol{x}$, follows from the fact that the set $\boldsymbol{\Theta}_{\boldsymbol{Y}}^{-}$ in (38) is convex and the function $nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}(\mathcal{A}_{\boldsymbol{x}}, \boldsymbol{\theta})$ is strictly convex with respect to $\boldsymbol{\theta}$. More specifically, the difference between a strictly convex function, i.e., $nK_{\boldsymbol{Y}}(\boldsymbol{\theta})$ and a linear function, i.e., $\boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}(\mathcal{A}_{\boldsymbol{x}}, \boldsymbol{\theta})$ is strictly convex. The former is strictly convex due to the fact that the covariance matrix $K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{0})$ is a positive definite matrix, c.f., [3, Section 1.2] and [15, Theorem 7.1].
Given $\boldsymbol{\tau}(\boldsymbol{x})$ in (39), the choice of the vector $\boldsymbol{\theta}$ to reduce the RHS of (36) is

$$\boldsymbol{\theta}(\boldsymbol{x}) = \begin{cases} \boldsymbol{\tau}(\boldsymbol{x}) & \text{if } \boldsymbol{\tau}(\boldsymbol{x}) \in \boldsymbol{\Theta}_{\boldsymbol{Y}}^{-} \\ \boldsymbol{\tau}(\boldsymbol{x}) + \boldsymbol{\epsilon} & \text{otherwise,} \end{cases} \tag{40}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^k$ is chosen such that two conditions are simultaneously met: First, $||\boldsymbol{\epsilon}|| < r$, with $r > 0$ arbitrary small; and second, $\boldsymbol{\theta}(\boldsymbol{x}) \in \boldsymbol{\Theta}_{\boldsymbol{Y}}^{-}$.

The following lemma, presents some of the properties of $\boldsymbol{\theta}(\boldsymbol{x})$.

Lemma 4 *For all $\boldsymbol{x} \in \mathbb{R}^k$, $\boldsymbol{\theta}(\boldsymbol{x})$ in (40) satisfies*

$$(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}_n})^{\mathsf{T}} \boldsymbol{\theta}(\boldsymbol{x}) \geq 0, \tag{41}$$

and

$$\boldsymbol{\theta}(\boldsymbol{\mu}_{\boldsymbol{X}_n}) = \boldsymbol{0}, \tag{42}$$

where,

$$\boldsymbol{\mu}_{\boldsymbol{X}_n} = \left(\mu_{\boldsymbol{X}_{n,1}}, \mu_{\boldsymbol{X}_{n,2}}, \ldots, \mu_{\boldsymbol{X}_{n,k}}\right)^{\mathsf{T}} \tag{43}$$

is the mean of the random vector $\boldsymbol{X}_n$ in (4).

Proof: The proof of Lemma 4 is presented in Appendix C. ■

Let the set $\mathcal{E}_{\boldsymbol{X}_n}$ be defined by

$$\mathcal{E}_{\boldsymbol{X}_n} \triangleq \left\{(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k : \forall i \in \{1, 2, \ldots, k\}, x_i > \mu_{X_{n,i}}\right\}, \tag{44}$$

where for all $i \in \{1, 2, \ldots, k\}$, $\mu_{X_{n,i}}$ is defined in (43). From (44), it holds that for all $\boldsymbol{x} \in \mathcal{E}_{\boldsymbol{X}_n}$, the vector $\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}_n}$ is such that all components are strictly positive. Similarly, from (38), it follows that $\boldsymbol{\theta}(\boldsymbol{x})$ is a vector whose components are all nonpositive. Hence, from (41), it follows that,

$$\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{0}, \tag{45}$$

which leads to the Gaussian approximation of the CDF $F_{\boldsymbol{X}_n}$ at the point $\boldsymbol{x}$. Hence, for all $\boldsymbol{x} \in \mathcal{E}_{\boldsymbol{X}_n}$ the choice of $\boldsymbol{\theta}$ in (40) can still be improved. In this case, the objective is to focus on $1 - F_{\boldsymbol{X}_n}(\boldsymbol{x})$ and write it as a sum of probability measures of convex sets with respect to $P_{\boldsymbol{X}_n}$. The following lemma provides such a result.

Lemma 5 *For all $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)^{\mathsf{T}} \in \mathbb{R}^k$, with $k \in \mathbb{N}$, it holds that*

$$1 - F_{\boldsymbol{X}_n}(\boldsymbol{x}) = \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} P_{\boldsymbol{X}_n}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})\right), \tag{46}$$

where the sets $\mathscr{S}(k)$ and $\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})$, with $\mathcal{J} \in \mathscr{S}(k)$, are respectively,

$$\mathscr{S}(k) = \{\mathcal{L} \subseteq \{1, 2, \ldots, k\} : \mathcal{L} \neq \emptyset\}; \text{ and} \tag{47}$$

$$\mathcal{B}_{\boldsymbol{x}}(\mathcal{J}) = \left\{\boldsymbol{t} = (t_1, t_2, \ldots, t_k) \in \mathbb{R}^k : \forall i \in \mathcal{J}, t_i > x_i\right\}. \tag{48}$$

Proof: The proof of Lemma 5 is presented in Appendix D ■

For all $\mathcal{J} \in \mathscr{S}(k)$, the probability $P_{\boldsymbol{X}_n}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})\right)$ in (46) can be approximated by using Theorem 3. More specifically, for all $\mathcal{J} \in \mathscr{S}(k)$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{Y}}$,

$$\left|P_{\boldsymbol{X}_n}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})\right) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), n)\right| \leqslant \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), \boldsymbol{\theta}\right)\right) \min\left(1, \frac{c(k)\, \xi_{\boldsymbol{Y}}(\boldsymbol{\theta})}{\sqrt{n}}\right). \tag{49}$$

Similar to the previous discussion on the minimization of the RHS of (36), the minimization of the RHS of (49) is focused only on the term $\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), \boldsymbol{\theta}\right)\right)$, and thus, the choice of $\boldsymbol{\theta}$ must be constrained to a subset of $\Theta_{\boldsymbol{Y}}$ in which $\boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), \boldsymbol{\theta}\right)$ is finite. That is, for all $\boldsymbol{x} \in \mathbb{R}^k$, the choice of $\boldsymbol{\theta}$ must satisfy that

$$\boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), \boldsymbol{\theta}\right) = \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{x} < +\infty. \tag{50}$$

More specifically, given a set $\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})$, it follows from (31), that the the choice of $\boldsymbol{\theta}$ must be restricted to the set

$$\boldsymbol{\Theta}_{\boldsymbol{Y}}^{\mathcal{J}} \triangleq \{(\theta_1, \theta_2, \ldots, \theta_k) \in \boldsymbol{\Theta}_{\boldsymbol{Y}} : \forall i \in \{1, 2, \ldots, k\},\, \theta_i \geq 0 \text{ if } i \in \mathcal{J}, \text{ and } \theta_i = 0 \text{ otherwise}\}. \tag{51}$$

Denote such a choice by $\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x})$, which is defined in terms of the following quantity:

$$\boldsymbol{\tau}_{\mathcal{J}}(\boldsymbol{x}) = \underset{\boldsymbol{t} \in \mathrm{clo}\boldsymbol{\Theta}_{\boldsymbol{Y}}^{\mathcal{J}}}{\arg\min} \left(nK_{\boldsymbol{Y}}(\boldsymbol{t}) - \boldsymbol{t}^{\mathsf{T}}\, \boldsymbol{x}\right) \tag{52}$$

with $\mathrm{clo}\boldsymbol{\Theta}_{\boldsymbol{Y}}^{\mathcal{J}}$ the closure of $\boldsymbol{\Theta}_{\boldsymbol{Y}}^{\mathcal{J}}$. The uniqueness of $\boldsymbol{\tau}_{\mathcal{J}}(\boldsymbol{x})$ in (52), for a given $\boldsymbol{x}$, follows from the fact that the set $\boldsymbol{\Theta}_{\boldsymbol{Y}}^{\mathcal{J}}$ in (51) is convex and the function $nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), \boldsymbol{\theta})$ is strictly convex with respect to $\boldsymbol{\theta}$, as discussed above.

Given $\boldsymbol{\tau}_{\mathcal{J}}(\boldsymbol{x})$ in (52), the choice of the vector $\boldsymbol{\theta}$ to reduce the RHS of (49) is

$$\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x}) = \begin{cases} \boldsymbol{\tau}_{\mathcal{J}}(\boldsymbol{x}) & \text{if } \boldsymbol{\tau}_{\mathcal{J}}(\boldsymbol{x}) \in \boldsymbol{\Theta}_{\boldsymbol{Y}}^{\mathcal{J}} \\ \boldsymbol{\tau}_{\mathcal{J}}(\boldsymbol{x}) + \boldsymbol{\epsilon} & \text{otherwise,} \end{cases} \tag{53}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^k$ is chosen such that two conditions are simultaneously met: First, $||\boldsymbol{\epsilon}|| < r$, with $r > 0$ arbitrary small; and second, $\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x}) \in \boldsymbol{\Theta}_{\boldsymbol{Y}}^{\mathcal{J}}$.

Finally, for all $\mathcal{J} \in \mathscr{S}(k)$, it holds from (49) that the probability $P_{\boldsymbol{X}_n}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})\right)$ can be approximated by $\eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), n)$. Using these approximations in (46), the approximation error can be upper bounded as follows:

$$\left|1 - F_{\boldsymbol{X}_n}(\boldsymbol{x}) - \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x}), \mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), n)\right|$$

$$= \left| \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} P_{\boldsymbol{X}_n}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})\right) - \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x}), \mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), n) \right| \tag{54}$$

$$\leq \sum_{\mathcal{J} \in \mathscr{S}(k)} \left| P_{\boldsymbol{X}_n}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})\right) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x}), \mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), n) \right| \tag{55}$$

$$\leq \sum_{\mathcal{J} \in \mathscr{S}(k)} \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}_{\mathcal{J}}^{\mathsf{T}}(\boldsymbol{x})\, \boldsymbol{x} \right) \min\left(1, \frac{c(k)\, \xi_{\boldsymbol{Y}}\left(\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x})\right)}{\sqrt{n}}\right), \tag{56}$$

where the inequality in (55) follows from the triangular inequality; and the inequality in (56) follows from (49).

In order to ease the notation, let the functions $\zeta_{\boldsymbol{Y}} : \mathbb{N} \times \mathbb{R}^k \to \mathbb{R}$ and $\delta_{\boldsymbol{Y}} : \mathbb{N} \times \mathbb{R}^k \to \mathbb{R}$ be such that for all $(n, \boldsymbol{x}) \in \mathbb{N} \times \mathbb{R}^k$,

$$\zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) \triangleq \begin{cases} \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}(\boldsymbol{x}), \mathcal{A}_{\boldsymbol{x}}, n) & \text{if } \boldsymbol{x} \notin \mathcal{E}_{\boldsymbol{X}_n} \\ 1 + \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{|\mathcal{J}|} \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x}), \mathcal{B}_{\boldsymbol{x}}(\mathcal{J}), n) & \text{if } \boldsymbol{x} \in \mathcal{E}_{\boldsymbol{X}_n}, \end{cases} \tag{57}$$

and

$$\delta_{\boldsymbol{Y}}(n, \boldsymbol{x}) \triangleq \begin{cases} \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}(\boldsymbol{x})) - \boldsymbol{\theta}(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{x}\right) \min\left(1, \frac{c(k)\, \xi_{\boldsymbol{Y}}(\boldsymbol{\theta}(\boldsymbol{x}))}{\sqrt{n}}\right) & \text{if } \boldsymbol{x} \notin \mathcal{E}_{\boldsymbol{X}_n} \\ \sum_{\mathcal{J} \in \mathscr{S}(k)} \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x})) - \boldsymbol{\theta}_{\mathcal{J}}^{\mathsf{T}}(\boldsymbol{x})\, \boldsymbol{x}\right) \min\left(1, \frac{c(k)\, \xi_{\boldsymbol{Y}}(\boldsymbol{\theta}_{\mathcal{J}}(\boldsymbol{x}))}{\sqrt{n}}\right) & \text{if } \boldsymbol{x} \in \mathcal{E}_{\boldsymbol{X}_n}. \end{cases} \tag{58}$$

Using this notation, the following theorem summarizes the discussion above.

Theorem 6 For all $\boldsymbol{x} \in \mathbb{R}^k$, it holds that

$$|F_{\boldsymbol{X}_n}(\boldsymbol{x}) - \zeta_{\boldsymbol{Y}}(n, \boldsymbol{x})| \leqslant \delta_{\boldsymbol{Y}}(n, \boldsymbol{x}), \tag{59}$$

where the functions $\zeta_{\boldsymbol{Y}}$ and $\delta_{\boldsymbol{Y}}$ are respectively defined in (57) and (58).

An immediate result from Theorem 6 is the following upper and lower bounds on $F_{\boldsymbol{X}_n}(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathbb{R}^k$,

$$\underline{\Omega}(n, \boldsymbol{x}) \leqslant F_{\boldsymbol{X}_n}(\boldsymbol{x}) \leqslant \bar{\Omega}(n, \boldsymbol{x}), \tag{60}$$

where,

$$\bar{\Omega}(n, \boldsymbol{x}) \triangleq \zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) + \delta_{\boldsymbol{Y}}(n, \boldsymbol{x}), \text{ and} \tag{61}$$

$$\underline{\Omega}(n, \boldsymbol{x}) \triangleq \zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) - \delta_{\boldsymbol{Y}}(n, \boldsymbol{x}). \tag{62}$$

Finally, the following claim underlines the relation between the saddlepoint approximation $\hat{F}_{\boldsymbol{X}_n}$ in (10) and the function $\zeta_{\boldsymbol{Y}}$ in (57).

Claim 7 *For all $\boldsymbol{x} \in \mathcal{D}$, with $\mathcal{D}$ in (7), it holds that*

$$\zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) = \hat{F}_{\boldsymbol{X}_n}(\boldsymbol{x}), \tag{63}$$

where the function $\hat{F}_{\boldsymbol{X}_n}$ and $\zeta_{\boldsymbol{Y}}$ are respectively defined in (10) and (57).

Proof: From (7), note that for all $\boldsymbol{x} \in \mathcal{D}$ the solution in $\boldsymbol{t}$ to

$$nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{t}) = \boldsymbol{x} \tag{64}$$

denoted by $\boldsymbol{\tau}_0$ exists and the components of $\boldsymbol{\tau}_0$ are all strictly negative. Thus, $\boldsymbol{\tau}_0 \in \Theta_{\boldsymbol{Y}}^-$, with $\Theta_{\boldsymbol{Y}}^-$ in (38).
Note that the vector $\boldsymbol{\tau}_0$ is also a solution to (39) for all $\boldsymbol{x} \in \mathcal{D}$. This follows from the fact that the CGF $K_{\boldsymbol{Y}}$ is strictly convex and $K_{\boldsymbol{Y}}^{(1)}$ is the gradient vector of the CGF $K_{\boldsymbol{Y}}$. Thus, the vector $\boldsymbol{\theta}(\boldsymbol{x})$ in (40) satisfies

$$\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{\tau}_0. \tag{65}$$

Then, for all $\boldsymbol{x} \in \mathcal{D}$, from (65), all the components of $\boldsymbol{\theta}(\boldsymbol{x})$ are strictly negative and thus, $\boldsymbol{x} \notin \mathcal{E}_{\boldsymbol{X}_n}$. Then, plugging (65) in (57) yields

$$\begin{aligned}
\zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) =& \eta_{\boldsymbol{Y}}(\boldsymbol{\tau}_0, \mathcal{A}_{\boldsymbol{x}}, n), & (66)\\
=& \exp\left(n\left(K_{\boldsymbol{Y}}(\boldsymbol{\tau}_0) - \boldsymbol{\tau}_0^{\mathsf{T}} K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\tau}_0) + \frac{\boldsymbol{\tau}_0^{\mathsf{T}}\, K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\, \boldsymbol{\tau}_0}{2}\right)\right) P_{\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}}(\mathcal{A}_{\boldsymbol{x}}) & (67)\\
=& \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\tau}_0) - n\boldsymbol{\tau}_0^{\mathsf{T}} K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\tau}_0) + \frac{n\boldsymbol{\tau}_0^{\mathsf{T}}\, K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\, \boldsymbol{\tau}_0}{2}\right) P_{\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}}(\mathcal{A}_{\boldsymbol{x}}) & (68)\\
=& \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\tau}_0) - \boldsymbol{\tau}_0^{\mathsf{T}} \boldsymbol{x} + \frac{n\boldsymbol{\tau}_0^{\mathsf{T}}\, K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\, \boldsymbol{\tau}_0}{2}\right) P_{\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}}(\mathcal{A}_{\boldsymbol{x}}) & (69)\\
=& \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\tau}_0) - \boldsymbol{\tau}_0^{\mathsf{T}} \boldsymbol{x} + \frac{n\boldsymbol{\tau}_0^{\mathsf{T}}\, K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\, \boldsymbol{\tau}_0}{2}\right) F_{\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}}(\boldsymbol{x}), & (70)
\end{aligned}$$

where $P_{\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}}$ is the probability measure induced by a Gaussian random vector $\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}$ with mean vector $n\Big(K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\tau}_0) - K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_0\Big)$ and covariance matrix $nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)$ on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$; and $F_{\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}}$ is the CDF of the random vector $\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}$. The equality in (67) follows from (27f). The equality in (69) follows from $\boldsymbol{\tau}_0$ being the solution in $\boldsymbol{t}$ to $nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{t}) = \boldsymbol{x}$.
Let the random vector $\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)}$ be such that

$$\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)} = \boldsymbol{H}_n^{(\boldsymbol{\tau}_0)} - n\left(K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\tau}_0) - K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_0\right), \tag{71}$$

which induces the probability measure $P_{\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)}}$ on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$ and the corresponding CDF is denoted by $F_{\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)}}$. Then, the mean vector and the covariance matrix of $\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)}$ are respectively $(0, 0, \ldots, 0)$ and $nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)$. Thus, for all $\boldsymbol{x} \in \mathbb{R}^k$, it holds that

$$\begin{aligned}
F_{\boldsymbol{H}_n^{(\boldsymbol{\tau}_0)}}(\boldsymbol{x}) =& F_{\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)}}\left(\boldsymbol{x} - n\left(K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\tau}_0) - K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_0\right)\right) & (72)\\
=& F_{\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_0\right), & (73)
\end{aligned}$$

where the equality in (73) follows from $\boldsymbol{\tau}_0$ being the solution in $\boldsymbol{t}$ to $nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{t}) = \boldsymbol{x}$.
Plugging (73) in (70) yields

$$\zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) = \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\tau}_0) - \boldsymbol{\tau}_0^{\mathsf{T}} \boldsymbol{x} + \frac{n\boldsymbol{\tau}_0^{\mathsf{T}}\, K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\, \boldsymbol{\tau}_0}{2}\right) F_{\boldsymbol{Z}_n^{(\boldsymbol{\tau}_0)}}\left(nK_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_0\right) \tag{74}$$

$$=\hat{F}_{\boldsymbol{X}_n}(\boldsymbol{x}), \tag{75}$$

where the equality in (75) follows from (10). This concludes the proof. ∎

5 Examples

Consider the case in which the independent random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_n$ in (4), with n fixed, are such that for all $i \in \{1, 2, \ldots, n\}$,

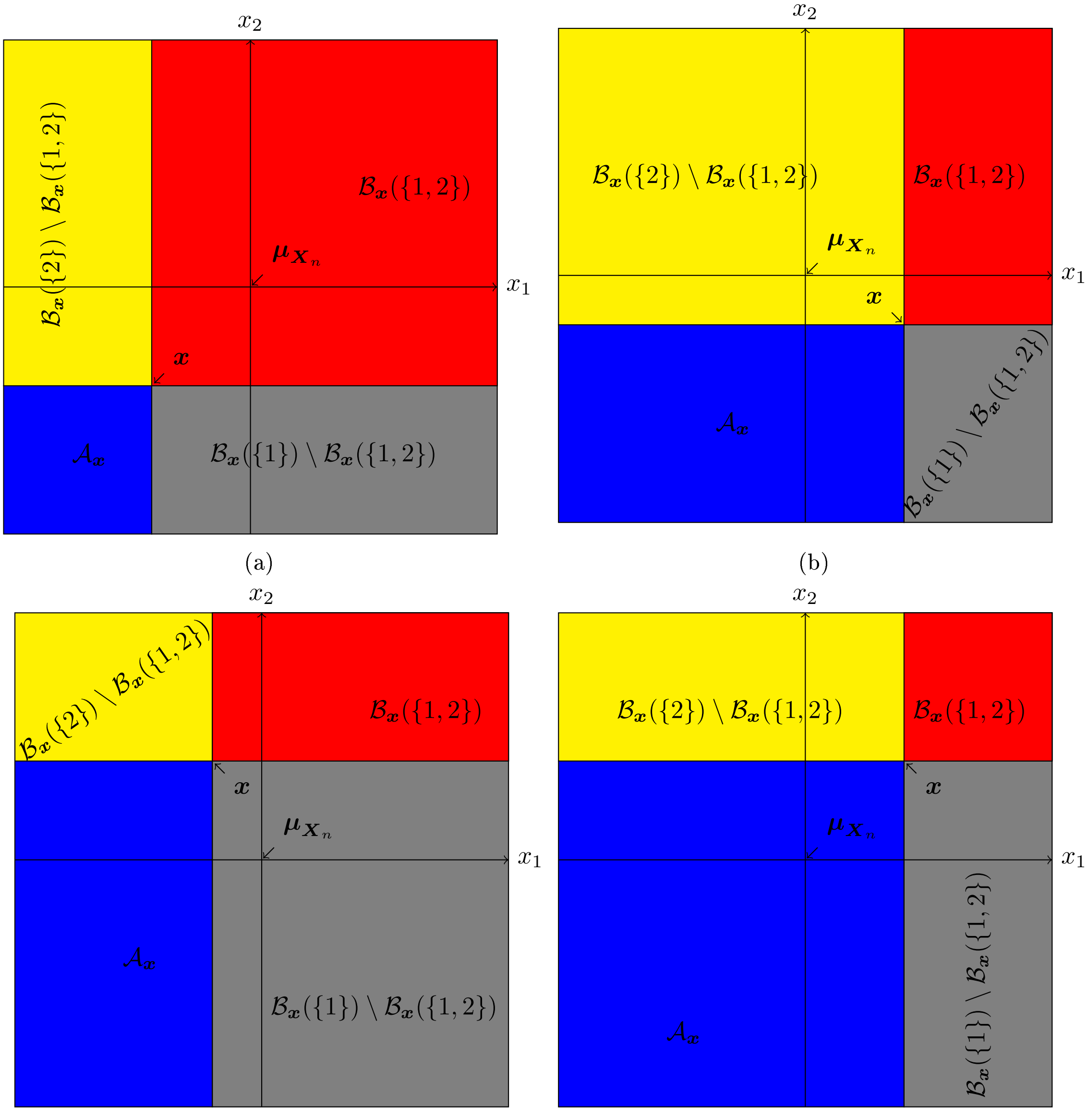
$$\boldsymbol{Y}_i \triangleq \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \tag{76}$$

where $\rho \in [0, 1)$ is the Pearson correlation coefficient between the components of $\boldsymbol{Y}_i$; and both B_1 and B_2 are independent Bernoulli random variables with parameter $p = 0.25$. The mean of $\boldsymbol{X}_n$ in (4) is $\boldsymbol{\mu}_{\boldsymbol{X}_n} = np\left(1, \rho + \sqrt{1-\rho^2}\right)^{\mathsf{T}}$.

Given a vector $\boldsymbol{x} \in \mathbb{R}^2$, Figure 1 depicts the set $\mathcal{A}_{\boldsymbol{x}}$ in (34) (blue rectangle); the set $\mathcal{B}_{\boldsymbol{x}}(\{1\}) \setminus \mathcal{B}_{\boldsymbol{x}}(\{1,2\})$ (grey rectangle), the set $\mathcal{B}_{\boldsymbol{x}}(\{2\}) \setminus \mathcal{B}_{\boldsymbol{x}}(\{1,2\})$ (yellow rectangle); and the set $\mathcal{B}_{\boldsymbol{x}}(\{1,2\})$ (red rectangle) in (48). Four cases are distinguished with respect to the given vector $\boldsymbol{x} \in \mathbb{R}^2$ and the mean vector $\boldsymbol{\mu}_{\boldsymbol{X}_n}$ in (43). In Sub-figure 1a, the mean vector $\boldsymbol{\mu}_{\boldsymbol{X}_n}$ belongs to the set $\mathcal{B}_{\boldsymbol{x}}(\{1,2\})$. In Sub-figure 1b, the mean vector $\boldsymbol{\mu}_{\boldsymbol{X}_n}$ belongs to the set $\mathcal{B}_{\boldsymbol{x}}(\{2\})$. In Sub-figure 1c, the mean vector $\boldsymbol{\mu}_{\boldsymbol{X}_n}$ belongs to the set $\mathcal{B}_{\boldsymbol{x}}(\{1\})$. In these three cases, the approximation of the CDF $F_{\boldsymbol{X}_n}$ is done using the set $\mathcal{A}_{\boldsymbol{x}}$, i.e., using the equality in (35). In Sub-figure 1d, the mean vector $\boldsymbol{\mu}_{\boldsymbol{X}_n}$ belongs to the set $\mathcal{A}_{\boldsymbol{x}}$. In this case, the approximation of the CDF $F_{\boldsymbol{X}_n}$ is done using the sets $\mathcal{B}_{\boldsymbol{x}}(\{1\})$, $\mathcal{B}_{\boldsymbol{x}}(\{2\})$, and $\mathcal{B}_{\boldsymbol{x}}(\{1,2\})$. That is, using the equality in (46).

Figures 2-4 depict the CDF $F_{\boldsymbol{X}_n}$ of $\boldsymbol{X}_n$ in (4); the Gaussian approximation $F_{\boldsymbol{Z}_n}$ in (16); the saddlepoint approximation $\zeta_{\boldsymbol{Y}}$ in (57); and the saddlepoint upper and lower bounds $\bar{\Omega}$ in (61) and $\underline{\Omega}$ in (62); through the line $a\boldsymbol{d} + \boldsymbol{\mu}_{\boldsymbol{X}_n}$. The plots on the left and the center in Figures 2-4 are respectively for fixed vectors $\boldsymbol{d} = (1, 1)^{\mathsf{T}}$ and $\boldsymbol{d} = (1, -1)^{\mathsf{T}}$, as a function of a. The plots on the right in Figures 2-4 are in function of ρ for a fixed point in the line $a\boldsymbol{d} + \boldsymbol{\mu}_{\boldsymbol{X}_n}$, with $a \in \{-6, -12, -24\}$ and $\boldsymbol{d} = (1, 1)^{\mathsf{T}}$, *i.e.*, the tail of the distribution in the direction of the vector $\boldsymbol{d} = (1, 1)^{\mathsf{T}}$. Note that Gaussian and saddlepoint approximations are particularly precise near to the mean $\boldsymbol{\mu}_{\boldsymbol{X}_n}$. That is, when $a = 0$. Nonetheless, away from the mean, i.e., $a < -4$ when $n = 25$, or $a < -10$ when $n \in \{50, 100\}$, the Gaussian approximation induces a large approximation error. Note that this is in sharp contrast with the saddlepoint approximation.

For the value of $n = 50$, Figure 3, the lower bound $\underline{\Omega}$ is negative, except when $a > 5$. Alternatively, the Gaussian upper and lower bounds $\bar{\Sigma}$ in (16a) and $\underline{\Sigma}$ in (16b) are trivial. That is, the lower bound is negative and the upper bound is bigger than one, which highlights the lack of formal mathematical arguments to evaluate the Gaussian approximation. For instance, note that when $a < -10$, the Gaussian approximation is bigger than the upper bound due to the saddle point approximation. In particular, note that Figure 3 (Right) highlights the fact that the same observation holds for all values of ρ.



(a)

(b)

(c)

(d)

Figure 1: Example of the set $\mathcal{A}_{\boldsymbol{x}}$ in (34) (blue rectangle); the set $\mathcal{B}_{\boldsymbol{x}}(\{1\}) \setminus \mathcal{B}_{\boldsymbol{x}}(\{1,2\})$ (grey rectangle), the set $\mathcal{B}_{\boldsymbol{x}}(\{2\}) \setminus \mathcal{B}_{\boldsymbol{x}}(\{1,2\})$ (yellow rectangle); and the set $\mathcal{B}_{\boldsymbol{x}}(\{1,2\})$ (red rectangle) in (48).

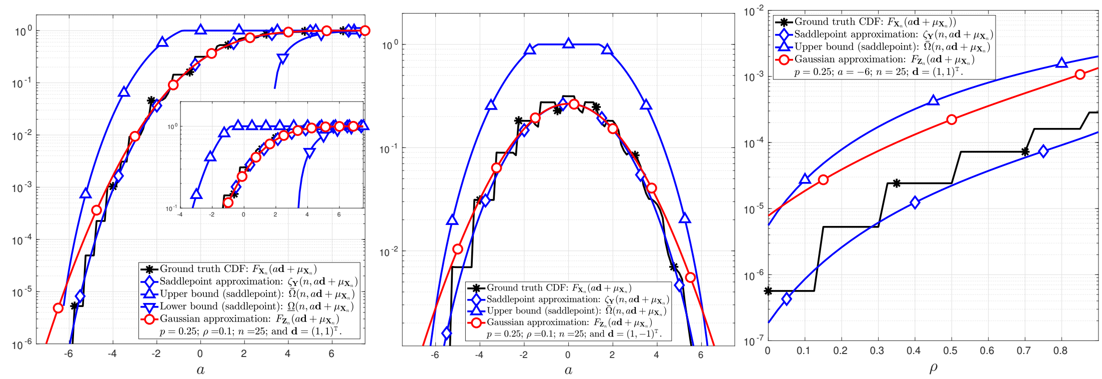


Figure 2: Sum of the independent random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_n$, with $n = 25$, such that for all $i \in \{1, 2, \ldots, n\}$, $\boldsymbol{Y}_i$ satisfies (76).

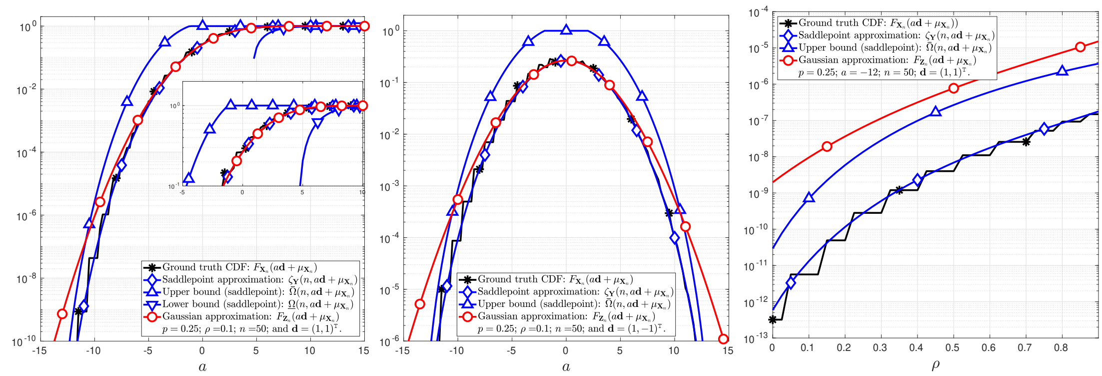


Figure 3: Sum of the independent random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_n$, with $n = 50$, such that for all $i \in \{1, 2, \ldots, n\}$, $\boldsymbol{Y}_i$ satisfies (76).

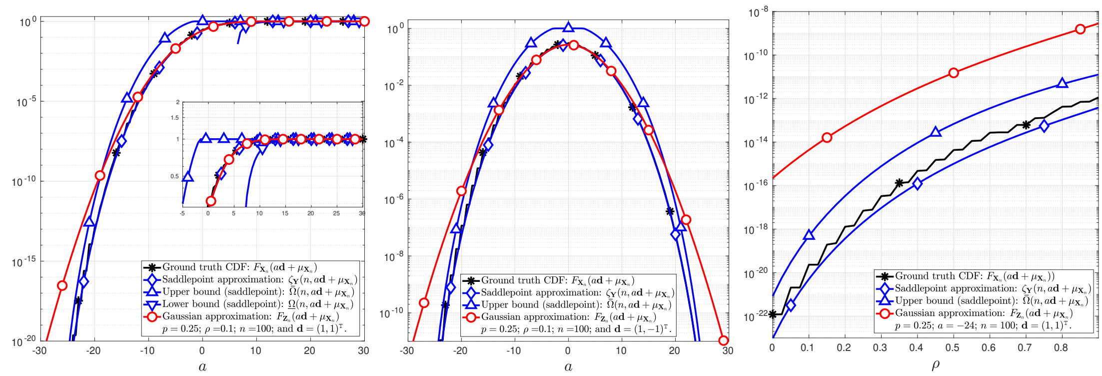


Figure 4: Sum of the independent random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_n$, with $n = 100$, such that for all $i \in \{1, 2, \ldots, n\}$, $\boldsymbol{Y}_i$ satisfies (76).

6 Final Remarks and Discussion

A final remark is the fact that for all $\boldsymbol{x} \in \mathcal{D}$, the saddlepoint approximation $\hat{F}_{\boldsymbol{X}_n}(\boldsymbol{x})$ in (10) is identical to $\zeta_{\boldsymbol{Y}}(n, \boldsymbol{x})$ in (57). That is, the saddlepoint approximation $\hat{F}_{\boldsymbol{X}_n}$ can be obtained from Theorem 6 in the special case in which the vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_n$ in (4) are either lattice or absolutely continuous random vectors. Additionally, it is worth to highlight that Theorem 6 holds for all random variables whose CGF exists. Under this condition, the Multivariate Berry-Essen theorem in [14, Theorem 1.1], Theorem 1, is a special case of Theorem 3 for the choice $\boldsymbol{\theta} = \mathbf{0}$.

The advantages of approximating the probability of a convex set in $\mathscr{B}(\mathbb{R}^k)$ by using Theorem 3 instead of Theorem 1 are twofold. First, the proposed upper bound on the approximation error depends on the set to be approximated. Second, both the approximation and the upper bound on the approximation error are parametrized by $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$, with $\Theta_{\boldsymbol{Y}}$ in (17). Thus, the vector $\boldsymbol{\theta}$ in (33) can be tuned to minimize the upper bound on the approximation error. Nonetheless, such optimization is not trivial. In this work, a non necessarily optimal choice has been made for obtaining an approximation of the CDF $F_{\boldsymbol{X}_n}$ in Theorem 6. That being said, the possibility to obtain tighter upper bounds on the approximation error on the measure $P_{\boldsymbol{X}_n}$ from Theorem 3; and on the approximation error on the CDF $F_{\boldsymbol{X}_n}$ from Theorem 6 is not negligible.
In the single dimension case, i.e., $k = 1$, Theorem 3 leads to the same approximation on the measure $P_{\boldsymbol{X}_n}$ in [16, Theorem 2]. Nonetheless, the upper bound provided in [16, Theorem 2] on the approximation error is better than the one provided by Theorem 3.

A Proof of Lemma 2

The proof relies on noticing that:

$$|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)| = \left| P_{\boldsymbol{X}_n}(\mathcal{A}) - \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right) \mathbb{E}_{P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}} \left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{Z}_n^{(\boldsymbol{\theta})} \right) \mathbb{1}_{\left\{ \boldsymbol{Z}_n^{(\boldsymbol{\theta})} \in \mathcal{A} \right\}} \right] \right|, \quad (77)$$

and thus, plugging (23) in the RHS of (77) yields

$$\begin{aligned} &|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)| \\ &= \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right) \left| \mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}} \left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{S}_n^{(\boldsymbol{\theta})} \right) \mathbb{1}_{\left\{ \boldsymbol{S}_n^{(\boldsymbol{\theta})} \in \mathcal{A} \right\}} \right] - \mathbb{E}_{P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}} \left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{Z}_n^{(\boldsymbol{\theta})} \right) \mathbb{1}_{\left\{ \boldsymbol{Z}_n^{(\boldsymbol{\theta})} \in \mathcal{A} \right\}} \right] \right|. \end{aligned} \quad (78)$$

In the following, explicit expressions for the terms

$$\mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}} \left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{S}_n^{(\boldsymbol{\theta})} \right) \mathbb{1}_{\left\{ \boldsymbol{S}_n^{(\boldsymbol{\theta})} \in \mathcal{A} \right\}} \right], \text{ and} \quad (79)$$

$$\mathbb{E}_{P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}} \left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{Z}_n^{(\boldsymbol{\theta})} \right) \mathbb{1}_{\left\{ \boldsymbol{Z}_n^{(\boldsymbol{\theta})} \in \mathcal{A} \right\}} \right], \quad (80)$$

in (78) are obtained.

A.1 Explicit Expression of (79)

From (79), the following holds

$$\mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}} \left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{S}_n^{(\boldsymbol{\theta})} \right) \mathbb{1}_{\left\{ \boldsymbol{S}_n^{(\boldsymbol{\theta})} \in \mathcal{A} \right\}} \right] = \int_{\mathbb{R}^k} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{x} \right) \mathbb{1}_{\{\boldsymbol{x} \in \mathcal{A}\}} \mathrm{d}P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}(\boldsymbol{x}). \quad (81)$$

The next step consists in writing the function $\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{x}\right)$ in the right hand-side of (81) as a Lebesgue integral. For doing so, consider the set $\mathcal{K} = \{1, 2, \ldots, k\}$ and let the set-valued functions $\mathcal{I}^- : \Theta_{\boldsymbol{Y}} \to 2^{\mathcal{K}}$, $\mathcal{I} : \Theta_{\boldsymbol{Y}} \to 2^{\mathcal{K}}$, and $\mathcal{I}^+ : \Theta_{\boldsymbol{Y}} \to 2^{\mathcal{K}}$ be respectively defined such that for all $\boldsymbol{u} = (u_1, u_2, \ldots, u_k) \in \Theta_{\boldsymbol{Y}}$

$$\mathcal{I}^-(\boldsymbol{u}) = \{i \in \{1, 2, \ldots, k\} : u_i < 0\}, \quad (82)$$

$$\mathcal{I}(\boldsymbol{u}) = \{i \in \{1, 2, \ldots, k\} : u_i = 0\}, \text{ and} \quad (83)$$

$$\mathcal{I}^+(\boldsymbol{u}) = \{i \in \{1, 2, \ldots, k\} : u_i > 0\}. \quad (84)$$

Then, for all $\boldsymbol{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, the following holds

$$\begin{aligned} &\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{x}\right) \\ &= \exp\left(-\theta_1 x_1 - \theta_2 x_2 - \ldots - \theta_k x_k\right) && (85) \\ &= \exp(-\theta_1 x_1) \exp(-\theta_2 x_2) \ldots \exp(-\theta_k x_k) && (86) \\ &= \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \exp\left(-\theta_i\, x_i\right) \right) \left(\prod_{j \in \mathcal{I}(\boldsymbol{\theta})} \exp\left(-\theta_j\, x_j\right) \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \exp\left(-\theta_s\, x_s\right) \right) && (87) \\ &= \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \int_{-\infty}^{x_i} -\theta_i \exp\left(-\theta_i\, t_i\right) \mathrm{d}t_i \right) \left(\prod_{j \in \mathcal{I}(\boldsymbol{\theta})} \int_0^1 \mathrm{d}t_j \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \int_{x_s}^{\infty} \theta_s \exp\left(-\theta_s\, t_s\right) \mathrm{d}t_s \right) && (88) \end{aligned}$$

$$= (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \int_{-\infty}^{x_i} \theta_i \exp\left(-\theta_i t_i\right) \mathrm{d}t_i \right) \left(\prod_{j \in \mathcal{I}(\boldsymbol{\theta})} \int_0^1 \mathrm{d}t_j \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \int_{x_s}^{\infty} \theta_s \exp\left(-\theta_s t_s\right) \mathrm{d}t_s \right). \quad (89)$$

To ease the notation, for all $u \in \mathbb{R}$ and for all $a \in \mathbb{R}$, let the set $\mathcal{B}_{u,a}$ be:

$$\mathcal{B}_{u,a} \triangleq \begin{cases} [0,1] & \text{if } u = 0 \\ (-\infty, a] & \text{if } u < 0 \\ [a, \infty) & \text{if } u > 0. \end{cases} \quad (90)$$

Then, using the notation in (90), the equality in (89) can be written as follow

$$\begin{aligned} &\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{x}\right) \\ &= (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \int_{\mathcal{B}_{\theta_i, x_i}} \theta_i \exp\left(-\theta_i t_i\right) \mathrm{d}t_i \right) \left(\prod_{j \in \mathcal{I}(\boldsymbol{\theta})} \int_{\mathcal{B}_{\theta_j, x_j}} \mathrm{d}t_j \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \int_{\mathcal{B}_{\theta_s, x_s}} \theta_s \exp\left(-\theta_s t_s\right) \mathrm{d}t_s \right) \quad (91) \\ &= (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \theta_i \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \theta_s \right) \int_{\mathcal{B}_{\theta_1, x_1}} \int_{\mathcal{B}_{\theta_2, x_2}} \cdots \int_{\mathcal{B}_{\theta_k, x_k}} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{t}\right) \mathrm{d}t_k \ldots \mathrm{d}t_2 \mathrm{d}t_1, \quad (92) \end{aligned}$$

where the equality in (92) follows from the linearity of the integration. To ease the notation, consider the set defined as

$$\mathcal{B}_{\boldsymbol{\theta}, \boldsymbol{x}} = \mathcal{B}_{\theta_1, x_1} \times \mathcal{B}_{\theta_2, x_2} \times \ldots \times \mathcal{B}_{\theta_k, x_k}, \quad (93)$$

where for all $i \in \{1, 2, \ldots, k\}$, the set $\mathcal{B}_{\theta_i, x_i}$ is defined in (90). Then, plugging (93) in (89) yields

$$\begin{aligned} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{x}\right) &= (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \theta_i \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \theta_s \right) \int_{\mathcal{B}_{\boldsymbol{\theta}, \boldsymbol{x}}} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{t}\right) \mathrm{d}\nu_k\left(\boldsymbol{t}\right) \quad (94) \\ &= (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \theta_i \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \theta_s \right) \int_{\mathbb{R}^k} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{t}\right) \mathbb{1}_{\{\boldsymbol{t} \in \mathcal{B}_{\boldsymbol{\theta}, \boldsymbol{x}}\}} \mathrm{d}\nu_k\left(\boldsymbol{t}\right), \quad (95) \end{aligned}$$

where ν_k is the Lebesgue measure on the measurable space $\left(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k)\right)$.
Then, plugging (95) in (81) yields

$$\begin{aligned} &\mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}} \left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{S}_n^{(\boldsymbol{\theta})}\right) \mathbb{1}_{\left\{\boldsymbol{S}_n^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}} \right] \\ &= \int_{\mathbb{R}^k} (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \theta_i \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \theta_s \right) \int_{\mathbb{R}^k} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{t}\right) \mathbb{1}_{\{\boldsymbol{t} \in \mathcal{B}_{\boldsymbol{\theta}, \boldsymbol{x}}\}} \mathrm{d}\nu_k\left(\boldsymbol{t}\right) \mathbb{1}_{\{\boldsymbol{x} \in \mathcal{A}\}} \mathrm{d}P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}(\boldsymbol{x}) \quad (96) \\ &= (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \theta_i \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \theta_s \right) \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{t}\right) \mathbb{1}_{\{\boldsymbol{t} \in \mathcal{B}_{\boldsymbol{\theta}, \boldsymbol{x}}\}} \mathbb{1}_{\{\boldsymbol{x} \in \mathcal{A}\}} \mathrm{d}\nu_k\left(\boldsymbol{t}\right) \mathrm{d}P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}(\boldsymbol{x}) \quad (97) \\ &= (-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|} \left(\prod_{i \in \mathcal{I}^-(\boldsymbol{\theta})} \theta_i \right) \left(\prod_{s \in \mathcal{I}^+(\boldsymbol{\theta})} \theta_s \right) \int_{\mathbb{R}^{k \times 2}} \exp\left(-\boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{t}\right) \mathbb{1}_{\{\boldsymbol{t} \in \mathcal{B}_{\boldsymbol{\theta}, \boldsymbol{x}}\}} \mathbb{1}_{\{\boldsymbol{x} \in \mathcal{A}\}} \mathrm{d}\nu_k \cdot P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\boldsymbol{t}, \boldsymbol{x}\right), \quad (98) \end{aligned}$$

where the Lebesgue integral in (98) is with respect to the product measure $\nu_k \cdot P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$ on the measurable space $\left(\mathbb{R}^{k\times 2}, \mathscr{B}(\mathbb{R}^{k\times 2})\right)$. Note that the integral in (98) is absolutely integrable. Thus, using the Fubini's Theorem [17], the right hand-side of (98) can be written as follows

$$\begin{aligned}&\mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}}\left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{S}_n^{(\boldsymbol{\theta})}\right)\mathbb{1}_{\left\{\boldsymbol{S}_n^{(\boldsymbol{\theta})}\in\mathcal{A}\right\}}\right]\\ =&(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{t}\right)\mathbb{1}_{\{\boldsymbol{t}\in\mathcal{B}_{\boldsymbol{\theta},\boldsymbol{x}}\}}\mathbb{1}_{\{\boldsymbol{x}\in\mathcal{A}\}}\mathrm{d}P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}(\boldsymbol{x})\,\mathrm{d}\nu_k(\boldsymbol{t}).\end{aligned} \tag{99}$$

From (93), the indicator $\mathbb{1}_{\{\boldsymbol{t}\in\mathcal{B}_{\boldsymbol{\theta},\boldsymbol{x}}\}}$ in (99) can be written as follows

$$\mathbb{1}_{\{\boldsymbol{t}\in\mathcal{B}_{\boldsymbol{\theta},\boldsymbol{x}}\}} = \left(\prod_{j\in\mathcal{I}(\boldsymbol{\theta})}\mathbb{1}_{\{t_j\in[0,1]\}}\right)\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\mathbb{1}_{\{t_i\leqslant x_i\}}\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\mathbb{1}_{\{t_s\geqslant x_s\}}\right). \tag{100}$$

To ease the notation, let the set

$$\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}} = \bar{\mathcal{B}}_{\theta_1,t_1}\times\bar{\mathcal{B}}_{\theta_2,t_2}\times\ldots\times\bar{\mathcal{B}}_{\theta_k,t_k}, \tag{101}$$

where, for all $i\in\{1,2,\ldots,k\}$, the set $\bar{\mathcal{B}}_{\theta_i,t_i}$ is defined by:

$$\bar{\mathcal{B}}_{\theta_i,t_i} = \begin{cases}\mathbb{R} & \text{if } \theta_i = 0\\ (-\infty, t_i] & \text{if } \theta_i > 0\\ [t_i,\infty) & \text{if } \theta_i < 0.\end{cases} \tag{102}$$

Then, plugging (101) in (100) yields

$$\mathbb{1}_{\{\boldsymbol{t}\in\mathcal{B}_{\boldsymbol{\theta},\boldsymbol{x}}\}} = \left(\prod_{j\in\mathcal{I}(\boldsymbol{\theta})}\mathbb{1}_{\{t_j\in[0,1]\}}\right)\mathbb{1}_{\left\{\boldsymbol{x}\in\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right\}}. \tag{103}$$

Hence, plugging (103) in (99) yields

$$\begin{aligned}&\mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}}\left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{S}_n^{(\boldsymbol{\theta})}\right)\mathbb{1}_{\left\{\boldsymbol{S}_n^{(\boldsymbol{\theta})}\in\mathcal{A}\right\}}\right]\\ =&(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{t}\right)\left(\prod_{j\in\mathcal{I}(\boldsymbol{\theta})}\mathbb{1}_{\{t_j\in[0,1]\}}\right)\\ &\mathbb{1}_{\left\{\boldsymbol{x}\in\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right\}}\mathbb{1}_{\{\boldsymbol{x}\in\mathcal{A}\}}\mathrm{d}P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}(\boldsymbol{x})\,\mathrm{d}\nu_k(\boldsymbol{t}) &(104)\\ =&(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathbb{R}^k}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{t}\right)\left(\prod_{j\in\mathcal{I}(\boldsymbol{\theta})}\mathbb{1}_{\{t_j\in[0,1]\}}\right)\\ &\int_{\mathbb{R}^k}\mathbb{1}_{\left\{\boldsymbol{x}\in\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right\}}\mathbb{1}_{\{\boldsymbol{x}\in\mathcal{A}\}}\mathrm{d}P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}(\boldsymbol{x})\,\mathrm{d}\nu_k(\boldsymbol{t}) &(105)\\ =&(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathbb{R}^k}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{t}\right)\left(\prod_{j\in\mathcal{I}(\boldsymbol{\theta})}\mathbb{1}_{\{t_j\in[0,1]\}}\right)P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)\mathrm{d}\nu_k(\boldsymbol{t}). &(106)\end{aligned}$$

Note that the support of the integrand in (106) is a subset of $\mathbb{R}^k$. Hence, the objective of the following lines is to characterize a subset of $\mathbb{R}^k$, denoted by $\mathcal{D}(\mathcal{A},\boldsymbol{\theta})$, that contains the support of the integrand in (106). The integrand is different from zero if $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)$ and $\prod_{j\in\mathcal{I}(\boldsymbol{\theta})}\mathbb{1}_{\{t_j\in[0,1]\}}$ are simultaneously strictly positive. On the first hand, given a vector $\boldsymbol{t}=(t_1,\, t_2,\, \ldots,\, t_k)\in\mathbb{R}^k$, the product $\prod_{j\in\mathcal{I}(\boldsymbol{\theta})}\mathbb{1}_{\{t_j\in[0,1]\}}$ is strictly positive if and only if for all $i\in\{1,2,\ldots,k\}$, it holds that

$$t_i\in[0,1] \text{ if } \theta_i=0. \tag{107}$$

On the other hand, given $\boldsymbol{t}=(t_1,\, t_2,\, \ldots,\, t_k)\in\mathbb{R}^k$, a necessary condition for $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)$ to be strictly positive is that the set $\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}$ is not empty. Now a necessary condition for the non-emptiness of $\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}$ is that the set $\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}$ contains at least one element $\boldsymbol{e}=(e_1,e_2,\ldots,e_k)$ such that for all $i\in\{1,2,\ldots,k\}$,

$$\inf_{(b_1,b_2,\ldots,b_k)\in\mathcal{A}} b_i \leqslant e_i \leqslant \sup_{(b_1,b_2,\ldots,b_k)\in\mathcal{A}} b_i. \tag{108}$$

The inequalities in (108) impose some conditions on the given vector $\boldsymbol{t}$. More specifically, from the definition of the set $\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}$ in (101), the vector $\boldsymbol{t}=(t_1,t_2,\ldots,t_k)$ must satisfy for all $i\in\{1,2,\ldots,k\}$,

$$\begin{cases} t_i \geqslant \inf\limits_{(b_1,b_2,\ldots,b_k)\in\mathcal{A}} b_i & \text{if } \theta_i>0; \text{ and}\\ t_i \leqslant \sup\limits_{(b_1,b_2,\ldots,b_k)\in\mathcal{A}} b_i & \text{if } \theta_i<0. \end{cases} \tag{109}$$

Hence, from (107) and (109) the set $\mathcal{D}(\mathcal{A},\boldsymbol{\theta})$ can be defined as follows

$$\begin{aligned}\mathcal{D}(\mathcal{A},\boldsymbol{\theta})\triangleq\Bigg\{&(t_1,t_2,\ldots,t_k)\in\mathbb{R}^k:\forall i\in\{1,2,\ldots,k\},\, t_i\in[0,1] \text{ if } \theta_i=0,\\ & t_i\geqslant \inf_{(b_1,b_2,\ldots,b_k)\in\mathcal{A}} b_i \text{ if } \theta_i>0, \text{ and } t_i\leqslant \sup_{(b_1,b_2,\ldots,b_k)\in\mathcal{A}} b_i \text{ if } \theta_i<0\Bigg\}.\end{aligned} \tag{110}$$

Then, the equality in (106) can be written as follows

$$\begin{aligned}&\mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}}\left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{S}_n^{(\boldsymbol{\theta})}\right)\mathbb{1}_{\left\{\boldsymbol{S}_n^{(\boldsymbol{\theta})}\in\mathcal{A}\right\}}\right]\\ &=(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{t}\right)P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)\mathrm{d}\nu_k\left(\boldsymbol{t}\right).\end{aligned} \tag{111}$$

A.2 Explicit Expression of (80)

Following similar steps as in Subsection A.1, the following holds with the random vector $\boldsymbol{Z}_n^{(\boldsymbol{\theta})}$:

$$\begin{aligned}&\mathbb{E}_{P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}}\left[\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{Z}_n^{(\boldsymbol{\theta})}\right)\mathbb{1}_{\left\{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}\in\mathcal{A}\right\}}\right]\\ &=(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{t}\right)P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)\mathrm{d}\nu_k\left(\boldsymbol{t}\right).\end{aligned} \tag{112}$$

A.3 Upper Bound on (78)

The proof ends by plugging (111) and (112) in the right hand-side of (78). This yields,

$$\begin{aligned}
&|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)| \\
=&\exp(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\Bigg|(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{t}\right)P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)\mathrm{d}\nu_k(\boldsymbol{t}) \\
&-(-1)^{|\mathcal{I}^-(\boldsymbol{\theta})|}\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{t}\right)P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)\mathrm{d}\nu_k(\boldsymbol{t})\Bigg| \qquad (113) \\
=&\exp(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\Bigg|\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right) \\
&\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{t}\right)\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)-P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)\right)\mathrm{d}\nu_k(\boldsymbol{t})\Bigg|. \qquad (114)
\end{aligned}$$

The set $\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\cap\mathcal{A}$ in (114) is convex Borel measurable, given that $\mathcal{A}$ is convex Borel measurable from the assumptions of the lemma. From (30), it holds that

$$\left|P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)-P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\left(\mathcal{A}\cap\bar{\mathcal{B}}_{\boldsymbol{\theta},\boldsymbol{t}}\right)\right|\leqslant\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}},P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right). \qquad (115)$$

Then, plugging (115) in (114) yields

$$\begin{aligned}
&|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)| \\
\leqslant&\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}-\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{t}\right)\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}},P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right)\nu_k(\mathrm{d}\boldsymbol{t}) \qquad (116)\\
=&\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}},P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right)\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}-\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{t}\right)\nu_k(\mathrm{d}\boldsymbol{t}) \qquad (117)
\end{aligned}$$

The expression $\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{t}\right)\mathrm{d}\nu_k(\boldsymbol{t})$ in (117) using the notation in (90) and (93) can be written in the form

$$\begin{aligned}
&\int_{\mathcal{D}(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{t}\right)\mathrm{d}\nu_k(\boldsymbol{t}) \\
=&\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}\int_{-\infty}^{a_i(\mathcal{A},\boldsymbol{\theta})}\exp\left(-\theta_i\,t_i\right)\mathrm{d}t_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\int_{a_s(\mathcal{A},\boldsymbol{\theta})}^{\infty}\exp\left(-\theta_s\,t_s\right)\mathrm{d}t_s\right) \qquad (118)\\
=&\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}-\frac{\exp\left(-\theta_i\,a_i\left(\mathcal{A},\boldsymbol{\theta}\right)\right)}{\theta_i}\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\frac{\exp\left(-\theta_s\,a_s\left(\mathcal{A},\boldsymbol{\theta}\right)\right)}{\theta_s}\right) \qquad (119)\\
=&\frac{\exp\left(-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{a}\left(\mathcal{A},\boldsymbol{\theta}\right)\right)}{\left(\prod_{i\in\mathcal{I}^-(\boldsymbol{\theta})}-\theta_i\right)\left(\prod_{s\in\mathcal{I}^+(\boldsymbol{\theta})}\theta_s\right)}, \qquad (120)
\end{aligned}$$

where the vector $\boldsymbol{a}\left(\mathcal{A},\boldsymbol{\theta}\right)=\left(a_1(\mathcal{A},\boldsymbol{\theta}),a_2(\mathcal{A},\boldsymbol{\theta}),\ldots,a_k(\mathcal{A},\boldsymbol{\theta})\right)$ is defined in (31). Hence, plugging (120) in (117) yields

$$|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)|\leqslant\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta})-\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{a}\left(\mathcal{A},\boldsymbol{\theta}\right)\right)\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}},P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right), \qquad (121)$$

which completes the proof.

B Proof of Theorem 3

From Lemma 2, it holds that

$$|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)| \leqslant \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}\,(\mathcal{A}, \boldsymbol{\theta})\right) \Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right). \tag{122}$$

Hence, the objective is to provide an upper bound on $\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right)$. An upper bound would be obtained immediately from Theorem 1, except for the fact that the vectors $\boldsymbol{Y}_1^{(\boldsymbol{\theta})}$, $\boldsymbol{Y}_2^{(\boldsymbol{\theta})}$, ..., $\boldsymbol{Y}_n^{(\boldsymbol{\theta})}$ in (22) do not have means $(0, 0, \ldots, 0)$ and variances $\frac{1}{n}\mathrm{diag}\,(1, 1, \ldots, 1)$, as required by Theorem 1. Denote by $\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \in \mathbb{R}^k$ and $\underline{\boldsymbol{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \in \mathbb{R}^{k\times k}$, respectively, the mean vector and the covariance matrix of these random vectors, for some $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$, with $\Theta_{\boldsymbol{Y}}$ in (17). Then, the following holds,

$$\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \triangleq \mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}[\boldsymbol{Y}^{(\boldsymbol{\theta})}] \tag{123a}$$

$$= \mathbb{E}_{P_{\boldsymbol{Y}}}\left[\boldsymbol{Y}\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{Y} - K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\right] \tag{123b}$$

$$= K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta}), \tag{123c}$$

where $K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})$ is the gradient vector of the CGF $K_{\boldsymbol{Y}}$ defined in (2). Alternatively,

$$\underline{\boldsymbol{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \triangleq \mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}\left[\left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\right] \tag{123d}$$

$$= \mathbb{E}_{P_{\boldsymbol{Y}}}\left[\left(\boldsymbol{Y} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\left(\boldsymbol{Y} - K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{Y} - K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\right] \tag{123e}$$

$$= K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta}), \tag{123f}$$

where $K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})$ is the Hessian matrix of the CGF $K_{\boldsymbol{Y}}$ defined in (3). Let the Cholesky decomposition of the matrix $\underline{\boldsymbol{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}$ be

$$\underline{\boldsymbol{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} = \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{\mathsf{T}}, \tag{124}$$

where $\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}$ is a real lower triangular matrix. Note that the matrix $\underline{\boldsymbol{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}$ is nonsingular. This follows from the assumption that the covariance matrix $K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{0})$ is positive definite, which implies that the CGF $K_{\boldsymbol{Y}}$ is strictly convex and thus, its Hessian matrix $K_{\boldsymbol{Y}}^{(2)}$ is positive definite.
Let the random vector $\boldsymbol{R}_n^{(\boldsymbol{\theta})}$ be such that

$$\boldsymbol{R}_n^{(\boldsymbol{\theta})} \triangleq \frac{1}{\sqrt{n}}\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1}\left(\boldsymbol{S}_n^{(\boldsymbol{\theta})} - n\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right) \tag{125}$$

which induces the probability measure $P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}$ on the measurable space $\left(\mathbb{R}^k, \mathscr{B}\left(\mathbb{R}^k\right)\right)$. Plugging (22) in (125) yields:

$$\boldsymbol{R}_n^{(\boldsymbol{\theta})} = \frac{1}{\sqrt{n}}\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{Y}_j^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right) \tag{126}$$

$$= \sum_{j=1}^{n}\frac{1}{\sqrt{n}}\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1}\left(\boldsymbol{Y}_j^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right) \tag{127}$$

$$=\sum_{j=1}^{n} \boldsymbol{U}_j^{(\boldsymbol{\theta})}, \tag{128}$$

where

$$\boldsymbol{U}_j^{(\boldsymbol{\theta})} \triangleq \frac{1}{\sqrt{n}} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \left(\boldsymbol{Y}_j^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \tag{129}$$

is a random vector that induces the probability measure $P_{\boldsymbol{U}^{(\boldsymbol{\theta})}}$ on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$. Thus, the random vector $\boldsymbol{R}_n^{(\boldsymbol{\theta})}$ in (125) is the sum of n IID random vectors $\boldsymbol{U}_1^{(\boldsymbol{\theta})}$, $\boldsymbol{U}_2^{(\boldsymbol{\theta})}$, ..., $\boldsymbol{U}_n^{(\boldsymbol{\theta})}$ such that each of them induces the probability measure $P_{\boldsymbol{U}^{(\boldsymbol{\theta})}}$ on the measurable space $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$, which satisfies,

$$\begin{aligned}
\mathbb{E}_{P_{\boldsymbol{U}^{(\boldsymbol{\theta})}}} \left[\boldsymbol{U}^{(\boldsymbol{\theta})} \right] &= \frac{1}{\sqrt{n}} \mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}} \left[\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \right] && (130)\\
&= \frac{1}{\sqrt{n}} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}} \left[\left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \right] && (131)\\
&= 0, && (132)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_{P_{\boldsymbol{U}^{(\boldsymbol{\theta})}}} \left[\boldsymbol{U}^{(\boldsymbol{\theta})} \boldsymbol{U}^{(\boldsymbol{\theta})\mathsf{T}} \right] &= \frac{1}{n} \mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}} \left[\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \left(\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \right)^{\mathsf{T}} \right] && (133)\\
&= \frac{1}{n} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}} \left[\left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \left(\boldsymbol{Y}^{(\boldsymbol{\theta})} - \boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right)^{\mathsf{T}} \right] \left(\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \right)^{\mathsf{T}} && (134)\\
&= \frac{1}{n} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \underline{\boldsymbol{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \left(\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \right)^{\mathsf{T}} && (135)\\
&= \frac{1}{n} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{\mathsf{T}} \left(\underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \right)^{\mathsf{T}} && (136)\\
&= \frac{1}{n} \mathrm{diag}\left(1, 1, \ldots, 1\right). && (137)
\end{aligned}$$

Similarly, let the random vector $\boldsymbol{W}_n^{(\boldsymbol{\theta})}$ be such that

$$\boldsymbol{W}_n^{(\boldsymbol{\theta})} \triangleq \frac{1}{\sqrt{n}} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \left(\boldsymbol{Z}_n^{(\boldsymbol{\theta})} - n\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right), \tag{138}$$

which induce the probability measure $P_{\boldsymbol{W}_n^{(\boldsymbol{\theta})}}$ on the measurable space $\left(\mathbb{R}^k, \mathscr{B}\left(\mathbb{R}^k\right)\right)$. The mean vector and the covariance matrix of the random vector $\boldsymbol{Z}_n^{(\boldsymbol{\theta})}$ are identical to those of the random vector $\boldsymbol{S}_n^{(\boldsymbol{\theta})}$. See for instance (24) and (25). Then, from (125) and (138), it holds that the mean vector and the covariance matrix of the random vector $\boldsymbol{W}_n^{(\boldsymbol{\theta})}$ are identical to those of the random vector $\boldsymbol{R}_n^{(\boldsymbol{\theta})}$.
The rest of the proof follows by noticing that for all $\mathcal{B} \in \mathscr{B}\left(\mathbb{R}^k\right)$, the set $\mathcal{S}\left(\mathcal{B}\right)$ defined by

$$\begin{aligned}
\mathcal{S}\left(\mathcal{B}\right) &\triangleq \left\{ \boldsymbol{y} \in \mathbb{R}^k : \exists \boldsymbol{x} \in \mathcal{B},\ \boldsymbol{y} = \frac{1}{\sqrt{n}} \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}^{-1} \left(\boldsymbol{x} - n\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \right\} && (139)\\
&= \left\{ \boldsymbol{y} \in \mathbb{R}^k : \left(\sqrt{n}\, \underline{\boldsymbol{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \boldsymbol{y} + n\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}} \right) \in \mathcal{B} \right\}. && (140)
\end{aligned}$$

allows writing that

$$P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}\left(\mathcal{B}\right) = \mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}} \left[\mathbb{1}_{\left\{ \boldsymbol{S}_n^{(\boldsymbol{\theta})} \in \mathcal{B} \right\}} \right] \tag{141}$$

$$
\begin{aligned}
&= \mathbb{E}_{P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}}\left[\mathbb{1}_{\left\{\left(\sqrt{n}\,\boldsymbol{\underline{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\boldsymbol{R}_n^{(\boldsymbol{\theta})}+n\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)\in\mathcal{B}\right\}}\right] && (142)\\
&= P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}\left(\left\{\boldsymbol{y}\in\mathbb{R}^k : \left(\sqrt{n}\,\boldsymbol{\underline{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\boldsymbol{y}+n\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)\in\mathcal{B}\right\}\right) && (143)\\
&= P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}\left(\mathcal{S}\left(\mathcal{B}\right)\right). && (144)
\end{aligned}
$$

Similarly, from (138) and (139), it holds that

$$
P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\left(\mathcal{B}\right) = P_{\boldsymbol{W}_n^{(\boldsymbol{\theta})}}\left(\mathcal{S}\left(\mathcal{B}\right)\right). \tag{145}
$$

This implies that,

$$
\begin{aligned}
\Delta\Big(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\Big) &= \sup_{\mathcal{B}\in\mathcal{C}_k}\left|P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}\left(\mathcal{S}\left(\mathcal{B}\right)\right) - P_{\boldsymbol{W}_n^{(\boldsymbol{\theta})}}\left(\mathcal{S}\left(\mathcal{B}\right)\right)\right|, && (146a)\\
&\leq \sup_{\mathcal{S}\in\mathcal{C}_k}\left|P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}\left(\mathcal{S}\right) - P_{\boldsymbol{W}_n^{(\boldsymbol{\theta})}}\left(\mathcal{S}\right)\right| && (146b)\\
&= \Delta\left(P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{W}_n^{(\boldsymbol{\theta})}}\right), && (146c)
\end{aligned}
$$

where the inequality in (146b) is a consequence of the fact that the collection $\mathcal{C}_k$ of all convex sets in $\mathscr{B}(\mathbb{R}^k)$ is stable under linear transformations. Then, from the multivariate Berry-Essen Theorem (Theorem 1), it holds that

$$
\begin{aligned}
&\Delta\left(P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{W}_n^{(\boldsymbol{\theta})}}\right)\\
&\leq \min\left(1, c(k)\, n\, \mathbb{E}_{P_{\boldsymbol{U}^{(\boldsymbol{\theta})}}}\left[||\boldsymbol{U}^{(\boldsymbol{\theta})}||^3\right]\right) && (147a)\\
&= \min\left(1, c(k)\, n\, \mathbb{E}_{P_{\boldsymbol{U}^{(\boldsymbol{\theta})}}}\left[\left(\boldsymbol{U}^{(\boldsymbol{\theta})\mathsf{T}}\boldsymbol{U}^{(\boldsymbol{\theta})}\right)^{\frac{3}{2}}\right]\right) && (147b)\\
&= \min\left(1, \frac{c(k)}{\sqrt{n}}\mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}\left[\left(\left((\boldsymbol{\underline{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})^{-1}\left(\boldsymbol{Y}^{(\boldsymbol{\theta})}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)\right)^{\mathsf{T}}(\boldsymbol{\underline{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})^{-1}\left(\boldsymbol{Y}^{(\boldsymbol{\theta})}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)\right)^{\frac{3}{2}}\right]\right) && (147c)\\
&= \min\left(1, \frac{c(k)}{\sqrt{n}}\mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}\left[\left(\left(\boldsymbol{Y}^{(\boldsymbol{\theta})}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)^{\mathsf{T}}\left((\boldsymbol{\underline{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})^{-1}\right)^{\mathsf{T}}(\boldsymbol{\underline{l}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})^{-1}\left(\boldsymbol{Y}^{(\boldsymbol{\theta})}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)\right)^{\frac{3}{2}}\right]\right) && (147d)\\
&= \min\left(1, \frac{c(k)}{\sqrt{n}}\mathbb{E}_{P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}\left[\left(\left(\boldsymbol{Y}^{(\boldsymbol{\theta})}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)^{\mathsf{T}}(\boldsymbol{\underline{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})^{-1}\left(\boldsymbol{Y}^{(\boldsymbol{\theta})}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}\right)\right)^{\frac{3}{2}}\right]\right) && (147e)\\
&= \min\left(1, \frac{c(k)}{\sqrt{n}}\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\left((\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})^{\mathsf{T}}\,(\boldsymbol{\underline{v}}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})^{-1}\,(\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}^{(\boldsymbol{\theta})}})\right)^{3/2}\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\,\boldsymbol{Y}-K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\right]\right), && (147f)
\end{aligned}
$$

where c is the function defined in (14).
Finally, plugging (123c) and (123f) in (147f) yields

$$
\begin{aligned}
&\Delta\left(P_{\boldsymbol{R}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{W}_n^{(\boldsymbol{\theta})}}\right)\\
&\leq \min\left(1, \frac{c(k)}{\sqrt{n}}\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\left(\left(\boldsymbol{Y}-K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)^{\mathsf{T}}\left(K_{\boldsymbol{Y}}^{(2)}(\boldsymbol{\theta})\right)^{-1}\left(\boldsymbol{Y}-K_{\boldsymbol{Y}}^{(1)}(\boldsymbol{\theta})\right)\right)^{3/2}\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}-K_{\boldsymbol{Y}}(\boldsymbol{\theta})\right)\right]\right) && (148a)\\
&= \min\left(1, \frac{c(k)\xi_{\boldsymbol{Y}}\left(\boldsymbol{\theta}\right)}{\sqrt{n}}\right), && (148b)
\end{aligned}
$$

where the function $\xi_{\boldsymbol{Y}}$ is defined in (32). Plugging (148b) in (146c) yields

$$\Delta\left(P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}, P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}\right) \leq \min\left(1, \frac{c(k)\xi_{\boldsymbol{Y}}(\boldsymbol{\theta})}{\sqrt{n}}\right). \tag{149}$$

Finally, plugging (149) in (122) yields

$$|P_{\boldsymbol{X}_n}(\mathcal{A}) - \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}, \mathcal{A}, n)| \leqslant \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{a}(\mathcal{A}, \boldsymbol{\theta})\right)\min\left(1, \frac{c(k)\xi_{\boldsymbol{Y}}(\boldsymbol{\theta})}{\sqrt{n}}\right), \tag{150}$$

which completes the proof.

C Proof of Lemma 4

Note that for all $\boldsymbol{x} \in \mathbb{R}^k$, it holds from (40) and from the fact that $\boldsymbol{0} \in \boldsymbol{\Theta}_{\boldsymbol{Y}}^-$, that

$$nK_{\boldsymbol{Y}}(\boldsymbol{\theta}(\boldsymbol{x})) - \boldsymbol{\theta}^{\mathsf{T}}(\boldsymbol{x})\,\boldsymbol{x} \leq \left(nK_{\boldsymbol{Y}}(\boldsymbol{0}) - \boldsymbol{0}^{\mathsf{T}}\boldsymbol{x}\right) = 0. \tag{151}$$

Moreover, for all $\boldsymbol{x} \in \mathbb{R}^k$ and for all $\boldsymbol{\theta} \in \mathrm{clo}\boldsymbol{\Theta}_{\boldsymbol{Y}}^-$, it holds that

$$\begin{aligned}
nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x} &= n\log\left(\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}\right)\right]\right) - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x} &(152)\\
&\geq n\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\log\left(\exp\left(\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}\right)\right)\right] - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x} &(153)\\
&= n\mathbb{E}_{P_{\boldsymbol{Y}}}\left[\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}\right] - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x} &(154)\\
&= \boldsymbol{\theta}^{\mathsf{T}} n\mathbb{E}_{P_{\boldsymbol{Y}}}[\boldsymbol{Y}] - \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x} &(155)\\
&= \boldsymbol{\theta}^{\mathsf{T}}\left(n\mathbb{E}_{P_{\boldsymbol{Y}}}[\boldsymbol{Y}] - \boldsymbol{x}\right) &(156)\\
&= \boldsymbol{\theta}^{\mathsf{T}}\left(\mathbb{E}_{P_{\boldsymbol{X}_n}}[\boldsymbol{X}_n] - \boldsymbol{x}\right) &(157)\\
&= \boldsymbol{\theta}^{\mathsf{T}}\left(\boldsymbol{\mu}_{\boldsymbol{X}_n} - \boldsymbol{x}\right), &(158)
\end{aligned}$$

where the inequality in (153) follows from Jensen's inequality [18, Section 2.6]; the equality in (157) follows from (4); and the equality in (158) follows from (43).
From (151) and (158), it holds that for all $\boldsymbol{x} \in \mathbb{R}^k$, $\boldsymbol{\theta}(\boldsymbol{x})$ in (40) satisfies

$$\boldsymbol{\theta}^{\mathsf{T}}(\boldsymbol{x})\left(\boldsymbol{\mu}_{\boldsymbol{X}_n} - \boldsymbol{x}\right) \leq 0, \tag{159}$$

which implies that $(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}_n})^{\mathsf{T}}\boldsymbol{\theta}(\boldsymbol{x}) \geq 0$ and proves the inequality in (41).
From (151) and (158), it holds that for $\boldsymbol{x} = \boldsymbol{\mu}_{\boldsymbol{X}_n}$,

$$nK_{\boldsymbol{Y}}\left(\boldsymbol{\tau}(\boldsymbol{\mu}_{\boldsymbol{X}_n})\right) - \boldsymbol{\tau}^{\mathsf{T}}(\boldsymbol{\mu}_{\boldsymbol{X}_n})\,\boldsymbol{\mu}_{\boldsymbol{X}_n} = \left(nK_{\boldsymbol{Y}}(\boldsymbol{0}) - \boldsymbol{0}^{\mathsf{T}}\boldsymbol{\mu}_{\boldsymbol{X}_n}\right) = 0. \tag{160}$$

Thus, the uniqueness of $\boldsymbol{\tau}(\boldsymbol{\mu}_{\boldsymbol{X}_n})$ implies from (160) that

$$\boldsymbol{\tau}(\boldsymbol{\mu}_{\boldsymbol{X}_n}) = \boldsymbol{0}. \tag{161}$$

Finally, note that $\boldsymbol{0} \in \boldsymbol{\Theta}_{\boldsymbol{Y}}^-$ and thus, from (40), it holds that

$$\boldsymbol{\theta}(\boldsymbol{\mu}_{\boldsymbol{X}_n}) = \boldsymbol{\tau}(\boldsymbol{\mu}_{\boldsymbol{X}_n}) = \boldsymbol{0}, \tag{162}$$

which concludes the proof.

D Proof of Lemma 5

For all $\boldsymbol{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, it holds that

$$1 - F_{\boldsymbol{X}_n}(\boldsymbol{x}) = 1 - \mathbb{E}_{\boldsymbol{X}_n}\left[\prod_{t=1}^{k} \mathbb{1}_{\{X_{n,t} \leq x_t\}}\right] \tag{163}$$

$$= \mathbb{E}_{\boldsymbol{X}_n}\left[1 - \prod_{t=1}^{k} \mathbb{1}_{\{X_{n,t} \leq x_t\}}\right] \tag{164}$$

$$= \mathbb{E}_{\boldsymbol{X}_n}\left[\max\left\{\mathbb{1}_{\{X_{n,t} > x_t\}} : t \in \{1, 2, \ldots, k\}\right\}\right]. \tag{165}$$

The proof continues by using a property of the max function provided by the following lemma.

Lemma 8 *For all $k \in \mathbb{N}$, and for all $(a_1,\, a_2,\, \ldots,\, a_k) \in \{0,1\}^k$, it holds that*

$$\max\{a_1, a_2, \ldots, a_k\} = \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} a_j, \tag{166}$$

where the collection $\mathscr{S}(k)$ is defined in (47)*.*

Proof: The proof is made by recurrence. For $k = 1$, the results is trivial. For $k = 2$, it holds that

$$\max\{a_1, a_2\} = a_1 + a_2 - a_1 a_2. \tag{167}$$

Assume that for all $k \in \mathbb{N}$ and for all $(a_1,\, a_2,\, \ldots,\, a_k) \in \{0,1\}^k$, it holds that

$$\max\{a_1, a_2, \ldots, a_k\} = \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} a_j, \tag{168}$$

where the collection $\mathscr{S}(k)$ is defined in (47). Let a_{k+1} a real be such that $a_{k+1} \in \{0,1\}$. Then, it holds that

$$\max\{a_1, a_2, \ldots, a_k, a_{k+1}\} = \max\{a_{k+1}, \max\{a_1, a_2, \ldots, a_k\}\}. \tag{169}$$

Note that $\max\{a_1, a_2, \ldots, a_k\} \in \{0,1\}$. Then, plugging (167) in (169) yields

$$\max\{a_1, a_2, \ldots, a_k, a_{k+1}\} = a_{k+1} + \max\{a_1, a_2, \ldots, a_k\} - a_{k+1} \max\{a_1, a_2, \ldots, a_k\}. \tag{170}$$

Then, plugging (168) in (170) yields

$$\max\{a_1, a_2, \ldots, a_k, a_{k+1}\}$$

$$= a_{k+1} + \left(\sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} a_j\right) - a_{k+1}\left(\sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} a_j\right) \tag{171}$$

$$= a_{k+1} + \left(\sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} a_j\right) + \left(\sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{2+|\mathcal{J}|} a_{k+1} \prod_{j \in \mathcal{J}} a_j\right) \tag{172}$$

$$= a_{k+1} + \left(\sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} a_j\right) + \left(\sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{2+|\mathcal{J}|} \prod_{j \in \mathcal{J} \cup \{k+1\}} a_j\right). \tag{173}$$

Note for all $k \in \mathbb{N}$, using the definition in (47), it holds that

$$\mathscr{S}(k+1) = \big\{\{k+1\}\big\} \cup \mathscr{S}(k) \cup \{\{k+1\} \cup \mathcal{J} : \mathcal{J} \in \mathscr{S}(k)\}. \tag{174}$$

Then, plugging (174) in (173) yields

$$\max\{a_1, a_2, \ldots, a_k, a_{k+1}\} = \sum_{\mathcal{J} \in \mathscr{S}(k+1)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} a_j, \tag{175}$$

which concludes the proof by recurrence.

■

Using Lemma 8 in (165), it follows that

$$1 - F_{\boldsymbol{X}_n}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{X}_n}\left[\sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \prod_{j \in \mathcal{J}} \mathbb{1}_{\{X_{n,j} > x_j\}}\right] \tag{176}$$

$$= \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} \mathbb{E}_{\boldsymbol{X}_n}\left[\prod_{j \in \mathcal{J}} \mathbb{1}_{\{X_{n,j} > x_j\}}\right] \tag{177}$$

$$= \sum_{\mathcal{J} \in \mathscr{S}(k)} (-1)^{1+|\mathcal{J}|} P_{\boldsymbol{X}_n}\left(\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})\right), \tag{178}$$

where the set $\mathcal{B}_{\boldsymbol{x}}(\mathcal{J})$ is defined in (48). The equality in (177) follows from the linearity of the expectation. This concludes the proof.

References

[1] J. E. Kolassa, "Multivariate saddlepoint tail probability approximations," *The Annals of Statistics*, vol. 31, no. 1, pp. 274–286, Feb. 2003.

[2] J. Kolassa and J. Li, "Multivariate saddlepoint approximations in tail probability and conditional inference," *Bernoulli*, vol. 16, no. 4, pp. 1191–1207, Nov. 2010.

[3] J. L. Jensen, *Saddlepoint Approximations*. New York, NY, USA: Clarendon press - Oxford, 1995.

[4] R. Butler, *Saddlepoint approximations with applications*. Cambridge, NY, USA: Cambridge University Press, 2007.

[5] D. Anade, J.-M. Gorce, P. Mary, and S. M. Perlaza, "On the saddlepoint approximation of the dependence testing bound in memoryless channels," in *Proc. of the IEEE International Conference on Communications (ICC)*, Dublin, Ireland, Jun. 2020.

[6] A. Martinez, J. Scarlett, M. Dalai, and A. G. i Fàbregas, "A complex-integration approach to the saddlepoint approximation for random-coding bounds," in *Proc. of the 11th International Symposium on Wireless Communications Systems (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 618–621.

[7] T. Erseghe, "Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6854–6883, Dec. 2016.

[8] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. G. i Fàbregas, and A. Lancho, "Saddle-point approximations of lower and upper bounds to the error probability in channel coding," in *Proc. of the 52nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, Mar. 2018, pp. 1–6.

[9] G. Vazquez-Vilar, A. G. i Fàbregas, T. Koch, and A. Lancho, "Saddlepoint approximation of the error probability of binary hypothesis testing," in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, Vail, CO, USA, 2018, pp. 2306–2310.

[10] J. Font-Segura, A. Martinez, and A. G. i Fàbregas, "Saddlepoint approximation of the cost-constrained random coding error probability," in *Proc. of the IEEE Information Theory Workshop (ITW)*, Guangzhou, China, Jan. 2018, pp. 1–5.

[11] J. Font-Segura, A. Martinez, and A. G. i Fàbregas, "Asymptotics of the random coding error probability for constant-composition codes," in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, Paris, France, Jul 2019, pp. 2947–2951.

[12] A. Lancho, J. Ostman, G. Durisi, T. Koch, and G. Vazquez-Vilar, "Saddlepoint approximations for short-packet wireless communications," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4831–4846, Jul. 2020.

[13] R. J. Beerends, *Fourier and Laplace transforms.* Cambridge, NY, USA: Cambridge University Press, 2003.

[14] M. Raič, "A multivariate BerryâĂŞEsseen theorem with explicit constants," *Bernoulli*, vol. 25, no. 4A, pp. 2824–2853, Nov. 2019.

[15] O. E. Nielsen, *Information and exponential families : in statistical theory.* Chichester U.K., NY, USA: John Wiley & Sons, 2014.

[16] D. Anade, J.-M. Gorce, P. Mary, and S. M. Perlaza, "An upper bound on the error induced by saddlepoint approximations - Applications to information theory," *Entropy*, vol. 22, no. 6, p. 690, Jun. 2020.

[17] R. Ash, *Probability and measure theory.* San Diego, CA, USA: Harcourt/Academic Press, 2000.

[18] T. M. Cover, *Elements of information theory.* Hoboken, N.J, USA: Wiley-Interscience, 2006.

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399