An upper bound on the error induced by saddlepoint approximations - Applications to information theory

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

An upper bound on the error induced by saddlepoint approximations - Applications to information theory

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

RESEARCH REPORT
N° 9329
February 2020
Project-Teams MARACAS and NEO

ISSN 0249-6399 ISRN INRIA/RR--9329--FR+ENG

An upper bound on the error induced by saddlepoint approximations - Applications to information theory

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

Project-Teams MARACAS and NEO

Research Report n° 9329 — version 1 — initial version February 2020 — revised version Mai 2020 — 56 pages

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, 2004 Route des Lucioles, 06902 Valbonne, France; and 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 in a special issue of Entropy (ISSN 1099-4300) in "Wireless Networks: Information Theoretic Perspectives".

Parts of this work have been published in the proceedings of the IEEE International Conference on Communications (ICC), Dublin, Ireland, Jun., 2020 [1].

RESEARCH CENTRE GRENOBLE – RHÔNE-ALPES

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

Abstract: This report introduces an upper bound on the absolute difference between: (a) the cumulative distribution function (CDF) of the sum of a finite number of independent and identically distributed random variables; and (b) a saddlepoint approximation of such CDF. This upper bound, which is particularly precise in the regime of large deviations is used to study the dependence testing (DT) bound and the meta converse (MC) bound on the decoding error probability (DEP) in point-to-point memoryless channels. Often, these bounds cannot be analytically calculated and thus, lower and upper bounds become particularly useful. Within this context, the main results include new upper bounds and lower bounds on the DT and MC bounds. A numerical analysis of these bounds is presented in the case of the binary symmetric channel, the additive white Gaussian noise channel, and the additive symmetric α-stable noise channel, in which the new bounds are observed to be tight.

Key-words: saddlepoint approximations, normal approximation, decoding error probability, memoryless channels

Résumé : Ce rapport propose une borne supérieure sur l'erreur induite par l'approximation du point de selle de la fonction de répartition de la somme des variables aléatoires identiquement distribuées. Cette borne est particulièrement précise sur la queue de la distribution. Ce résultat est appliqué pour étudier la borne "dependence testing (DT)" et celle du "meta converse (MC)" sur la probabilité d'erreur minimale de décodage d'un canal sans mémoire. Dans ce contexte, les résultats principaux sont les nouvelles bornes supérieures et inférieures sur les bornes DT et MC. Une analyse numérique de ces bornes est présentée pour les canaux binaires symétriques, les canaux avec un bruit blanc gaussien additif et les canaux avec un bruit impulsionnel additif. Les bornes obtenues par notre méthode sont meilleures que celles obtenues à l'aide du Théorème de Berry-Esseen.

Mots-clés : approximation du point de selle, approximation Gaussienne, probabilité d'erreur de décodage, canal sans mémoire

Contents

1 Introduction

This report focuses on approximating the cumulative distribution function (CDF) of sums of a finite number of real-valued independent and identically distributed (i.i.d.) random variables. More specifically, let $Y_1, Y_2, \ldots, Y_n$, with n an integer and $1 \leqslant n \leqslant \infty$, be real-valued random variables with probability distribution P_Y. Denote by F_Y the CDF associated with P_Y, and if it exists, denote by f_Y the corresponding probability density function (PDF). Let also

$$X_n = \sum_{t=1}^{n} Y_t \tag{1}$$

be a random variable with distribution P_{X_n}. Denote by F_{X_n} the CDF and if it exists, denote by f_{X_n} the PDF associated with P_{X_n}. The objective is to provide a positive function that approximates F_{X_n} and an upper bound on the resulting approximation error. In the following, a positive function $g : \mathbb{R} \to \mathbb{R}_+$ is said to approximate F_{X_n} with an *approximation error* that is upper bounded by a function $\epsilon : \mathbb{R} \to \mathbb{R}_+$, if for all $x \in \mathbb{R}$,

$$|F_{X_n}(x) - g(x)| \leqslant \epsilon(x). \tag{2}$$

The case in which $Y_1, Y_2, \ldots, Y_n$ in (1) are stable random variables with F_Y analytically expressible is trivial. This is essentially because the sum X_n follows the same distribution of a random variable $aY + b$, for some $(a, b) \in \mathbb{R}^2$ and Y a random variable whose CDF is F_Y. Examples of this case are random variables following the Gaussian distribution, the Cauchy distribution or the Levy distribution [2].
In general, the problem of calculating the CDF of X_n boils down to calculating $n-1$ convolutions. More specifically, it holds that

$$f_{X_n}(x) = \int_{-\infty}^{\infty} f_{X_{n-1}}(x - t)\, f_Y(t)\mathrm{d}t, \tag{3}$$

where $f_{X_1} = f_Y$. Even for discrete random variables and small values of n, the integral in (3) often requires excessive computation resources [3].
When the PDF of the random variable X_n cannot be conveniently obtained but only the r first moments are known, with $r \in \mathbb{N}$, an approximation of the PDF can be obtained by using an Edgeworth expansion. Nonetheless, the resulting relative error in the large deviation regime makes these approximations inaccurate [4].
When the cumulant generating function (CGF) associated with F_Y, denoted by $K_Y : \mathbb{R} \to \mathbb{R}$, is known, the PDF f_{X_n} can be obtained via the Laplace inversion lemma [3]. That is, given two reals $\alpha_- < 0$ and $\alpha_+ > 0$, if K_Y is analytic for all $z \in \{a+ib \in \mathbb{C} : (a, b) \in \mathbb{R}^2$ and $\alpha_- \leqslant a \leqslant \alpha_+\} \subset \mathbb{C}$, then,

$$f_{X_n}(x) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \exp\left(nK_Y(z) - zx\right) \mathrm{d}z, \tag{4}$$

with $i = \sqrt{-1}$ and $\gamma \in (\alpha_-, \alpha_+)$. Note that the domain of K_Y in (4) has been extended to the complex numbers and thus, it is often referred to as the complex CGF. With an abuse of notation, both the CGF and the complex CGF are identically denoted.
In the case in which n is sufficiently large, an approximation to the Bromwich integral in (4) can be obtained by choosing the countour to include the unique saddlepoint of the integrand as suggested in [5]. The intuition behind this lies on the following observations:
(i) the saddlepoint, denoted by z_0, is unique, real and $z_0 \in (\alpha_-, \alpha_+)$;
(ii) within a neighborhood around the saddlepoint of the form $|z - z_0| < \epsilon$, with $z \in \mathbb{C}$ and $\epsilon > 0$

sufficiently small, $\mathrm{Im}\left[nK_Y(z) - zx\right] = 0$ and $\mathrm{Re}\left[nK_Y(z) - zx\right]$ can be assumed constant; and (iii) outside such neighborhood, the integrand is negligible.
From (i), it follows that the derivative of $nK_Y(t) - tx$ with respect to t, with $t \in \mathbb{R}$, is equal to zero when it is evaluated at the saddlepoint z_0. More specifically, for all $t \in \mathbb{R}$,

$$\frac{\mathrm{d}}{\mathrm{d}t} K_Y(t) = \mathbb{E}_{P_Y}\left[Y \exp\left(tY - K_Y(t)\right)\right], \tag{5}$$

and thus,

$$\mathbb{E}_{P_Y}\left[Y \exp\left(z_0 Y - K_Y(z_0)\right)\right] = \frac{x}{n}, \tag{6}$$

which shows the dependence of z_0 on both x and n.
An expansion in Taylor series of the exponent $nK_Y(z) - zx$ in the neighborhood of z_0, leads to the following asymptotic expansion in powers of $\frac{1}{n}$ of the Bromwich integral in (4):

$$f_{X_n}(x) = \hat{f}_{X_n}(x)\left(1 + \frac{1}{n}\left(\frac{1}{8}\frac{K_Y^{(4)}(z_0)}{\left(K_Y^{(2)}(z_0)\right)^2} - \frac{5}{24}\frac{\left(K_Y^{(3)}(z_0)\right)^2}{\left(K_Y^{(2)}(z_0)\right)^3}\right) + O\left(\frac{1}{n^2}\right)\right), \tag{7}$$

where $\hat{f}_{X_n} : \mathbb{R} \to \mathbb{R}_+$ is

$$\hat{f}_{X_n}(x) = \sqrt{\frac{1}{2\pi n K_Y^{(2)}(z_0)}} \exp\left(nK_Y(z_0) - z_0 x\right), \tag{8}$$

and for all $k \in \mathbb{N}$ and $t \in \mathbb{R}$, the notation $K_Y^{(k)}(t)$ represents the k-th real derivative of the CGF K_Y evaluated at t. The first two derivatives $K_Y^{(1)}$ and $K_Y^{(2)}$ play a central role, and thus, it is worth to provide explicit expressions. That is,

$$K_Y^{(1)}(t) \triangleq \mathbb{E}_{P_Y}\left[Y \exp\left(tY - K_Y(t)\right)\right], \text{ and} \tag{9}$$

$$K_Y^{(2)}(t) \triangleq \mathbb{E}_{P_Y}\left[\left|Y - K_Y^{(1)}(t)\right|^2 \exp\left(tY - K_Y(t)\right)\right]. \tag{10}$$

The function $\hat{f}_{X_n}$ in (8) is referred to as the *saddlepoint approximation* of the PDF f_{X_n} and was first introduced in [5]. Nonetheless, $\hat{f}_{X_n}$ is not necessarily a PDF as often its integral on $\mathbb{R}$ is not equal to one. A particular exception is observed only in three cases [6]. First, when f_Y is the PDF of a Gaussian random variable, the saddlepoint approximation $\hat{f}_{X_n}$ is identical to f_{X_n}, for all $n > 0$. Second and third, when f_Y is the PDF associated with a Gamma distribution and an inverse normal distribution, respectively, the saddlepoint approximation $\hat{f}_{X_n}$ is exact up to a normalization constant for all $n > 0$.
An approximation to the CDF F_{X_n} can be obtained by integrating the PDF in (4), c.f., [7, 8] and [9]. In particular, the result reported in [7] leads to an asymptotic expansion of the CDF of X_n, for all $x \in \mathbb{R}$, of the form:

$$F_{X_n}(x) = \hat{F}_{X_n}(x) + O\left(n^{-1/2} \exp\left(nK_Y(z_0) - xz_0\right)\right), \tag{11}$$

where the function $\hat{F}_{X_n} : \mathbb{R} \to \mathbb{R}$ is the *saddlepoint approximation* of F_{X_n}. That is, for all $x \in \mathbb{R}$,

$$\hat{F}_{X_n}(x) = \mathbb{1}_{\{z_0 > 0\}} + (-1)^{\mathbb{1}_{\{z_0>0\}}} \exp\left(nK_Y(z_0) - z_0 x + \frac{1}{2} z_0^2 n K_Y^{(2)}(z_0)\right) Q\left(|z_0| \sqrt{n K_Y^{(2)}(z_0)}\right), \tag{12}$$

where the function $Q : \mathbb{R} \to [0, 1]$ is the complementary CDF of a Gaussian random variable with zero mean and unit variance. That is, for all $t \in \mathbb{R}$,

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x. \tag{13}$$

Finally, from the central limit theorem [4], for large values of n and for all $x \in \mathbb{R}$, a reasonable approximation to $F_{X_n}(x)$ is $1 - Q(x)$. In the following, this approximation is referred to as the *normal approximation* of F_{X_n}.

1.1 Contributions

The main contribution of this work is an upper bound on the error induced by the saddlepoint approximation $\hat{F}_{X_n}$ in (12) of the CDF F_{X_n} of the sum in (1) (Theorem 3 in Section 2.2). This result builds upon two observations. The first observation is that the CDF F_{X_n} can be written for all $x \in \mathbb{R}$ in the form,

$$\begin{aligned} &F_{X_n}(x) \\ &= \mathbb{1}_{\{z_0 \leqslant 0\}} \mathbb{E}_{P_{S_n}}\left[\exp(nK_Y(z_0) - z_0 S_n) \mathbb{1}_{\{S_n \leqslant x\}}\right] + \mathbb{1}_{\{z_0 > 0\}}\left(1 - \mathbb{E}_{P_{S_n}}\left[\exp(nK_Y(z_0) - z_0 S_n) \mathbb{1}_{\{S_n > x\}}\right]\right), \end{aligned} \tag{14}$$

where the random variable

$$S_n = \sum_{t=1}^{n} Y_t^{(z_0)} \tag{15}$$

has a probability distribution denoted by P_{S_n}, and the random variables $Y_1^{(z_0)}, Y_2^{(z_0)}, \ldots, Y_n^{(z_0)}$ are independent with probability distribution $P_{Y^{(z_0)}}$. The distribution $P_{Y^{(z_0)}}$ is an exponentially tilted distribution [10] with respect to the distribution P_Y at the saddlepoint z_0. More specifically, the Radon-Nikodym derivative of the distribution $P_{Y^{(z_0)}}$ with respect to the distribution P_Y satisfies for all $y \in \mathrm{supp} P_Y$,

$$\frac{\mathrm{d}P_{Y^{(z_0)}}}{\mathrm{d}P_Y}(y) = \exp\left(-\left(K_Y(z_0) - z_0 y\right)\right). \tag{16}$$

The second observation is that the saddlepoint approximation $\hat{F}_{X_n}$ in (12) can be written for all $x \in \mathbb{R}$ in the form,

$$\begin{aligned} &\hat{F}_{X_n}(x) \\ &= \mathbb{1}_{\{z_0 \leqslant 0\}} \mathbb{E}_{P_{Z_n}}\left[\exp(nK_Y(z_0) - z_0 Z_n) \mathbb{1}_{\{Z_n \leqslant x\}}\right] + \mathbb{1}_{\{z_0 > 0\}}\left(1 - \mathbb{E}_{P_{Z_n}}\left[\exp(nK_Y(z_0) - z_0 Z_n) \mathbb{1}_{\{Z_n > x\}}\right]\right), \end{aligned} \tag{17}$$

where Z_n is a Gaussian random variable with mean x, variance $nK_Y^{(2)}(z_0)$, and probability distribution P_{Z_n}. Note that the means of the random variable S_n in (14) and Z_n in (17) are equal to $nK_Y^{(1)}(z_0)$, whereas their variances are equal to $nK_Y^{(2)}(z_0)$. Note also that from (6), it holds that $x = nK_Y^{(1)}(z_0)$.

Using these observations, it holds that the absolute difference between F_{X_n} in (14) and $\hat{F}_{X_n}$ in (17) satisfies for all $x \in \mathbb{R}$,

$$\begin{aligned} &\left|F_{X_n}(x) - \hat{F}_{X_n}(x)\right| \\ &= \mathbb{1}_{\{z_0 \leqslant 0\}} \left|\mathbb{E}_{P_{S_n}}\left[\exp(nK_Y(z_0) - z_0 S_n) \mathbb{1}_{\{S_n \leqslant x\}}\right] - \mathbb{E}_{P_{Z_n}}\left[\exp(nK_Y(z_0) - z_0 Z_n) \mathbb{1}_{\{Z_n \leqslant x\}}\right]\right| \end{aligned}$$

$$+\mathbb{1}_{\{z_0>0\}}\left|\mathbb{E}_{P_{S_n}}\left[\exp\left(nK_Y(z_0)-z_0S_n\right)\mathbb{1}_{\{S_n>x\}}\right]-\mathbb{E}_{P_{Z_n}}\left[\exp\left(nK_Y(z_0)-z_0Z_n\right)\mathbb{1}_{\{Z_n>x\}}\right]\right|. \quad (18)$$

A step forward (Lemma 4 in Appendix A) is to note that when x is such that $z_0 \leqslant 0$, then,

$$\begin{aligned}&\left|\mathbb{E}_{P_{S_n}}\left[\exp\left(nK_Y(z_0)-z_0S_n\right)\mathbb{1}_{\{S_n\leqslant x\}}\right]-\mathbb{E}_{P_{Z_n}}\left[\exp\left(nK_Y(z_0)-z_0Z_n\right)\mathbb{1}_{\{Z_n\leqslant x\}}\right]\right|\\&\leqslant \exp\left(nK_Y(z_0)-z_0x\right)\min\left(1, 2\sup_{a\in\mathbb{R}}\left|F_{S_n}(a)-F_{Z_n}(a)\right|\right),\end{aligned} \quad (19)$$

and when x is such that $z_0 > 0$, it holds that

$$\begin{aligned}&\left|\mathbb{E}_{P_{S_n}}\left[\exp\left(nK_Y(z_0)-z_0S_n\right)\mathbb{1}_{\{S_n> x\}}\right]-\mathbb{E}_{P_{Z_n}}\left[\exp\left(nK_Y(z_0)-z_0Z_n\right)\mathbb{1}_{\{Z_n> x\}}\right]\right|\\&\leqslant \exp\left(nK_Y(z_0)-z_0x\right)\min\left(1, 2\sup_{a\in\mathbb{R}}\left|F_{S_n}(a)-F_{Z_n}(a)\right|\right),\end{aligned} \quad (20)$$

where F_{S_n} and F_{Z_n} are the CDFs of the random variables S_n and Z_n, respectively. The final result is obtained by observing that $\sup_{a\in\mathbb{R}}|F_{S_n}(a)-F_{Z_n}(a)|$ can be upper bounded using the Berry-Esseen Theorem (Theorem 1 in Section 2.1). This is essentially due to the fact that the random variable S_n is the sum of n independent random variables, i.e., (15), and Z_n is a Gaussian random variable, and both S_n and Z_n possess identical means and variances. Thus, the main result (Theorem 3 in Section 2.2) is that for all $x \in \mathbb{R}$,

$$\left|F_{X_n}(x)-\hat{F}_{X_n}(x)\right| \leqslant n^{-1/2}\exp\left(nK_Y(z_0)-z_0\,x\right)\frac{2\,c\,\xi_Y(z_0)}{\left(K_Y^{(2)}(z_0)\right)^{3/2}}, \quad (21)$$

where c can be chosen as $c = 0.476$ according to [11]; and $\xi_Y(z_0)$ is the third absolute central moment with respect to the distribution $P_{Y^{(z_0)}}$. Finally, note that (21) reflects the scaling law with respect to n suggested in (11).

1.2 Applications

In the realm of information theory, the normal approximation has played a central role in the calculation of bounds on the minimum decoding error probability (DEP) in point-to-point memoryless channels, c.f., [12, 13]. Thanks to the normal approximation, simple approximations for the dependence testing (DT) bound, the random coding union bound (RCU) bound, and the meta-converse (MC) bound have been obtained in [12, 14]. The success of these approximations stems from the fact that they are easy to calculate. Nonetheless, easy computation comes at the expense of loose upper and lower bounds, and thus, uncontrolled approximation errors.

On the other hand, saddlepoint techniques have been extensively used to approximate existing lower and upper bounds on the minimum DEP. See for instance, [15] and [16] in the case of the RCU bound and the MC bound. Nonetheless, the errors induced by saddlepoint approximations are often neglected due to the fact that calculating them involves a large number of optimizations and numerical integrations. Within this context, the main results of this report are used to provide new lower and upper bounds on the DT bound and the MC bound. Numerical analysis of these bounds are presented for the case of the binary symmetric channel (BSC), the additive white Gaussian noise (AWGN) channel, and the additive symmetric α-stable noise (SαS) channel, in which the new bounds are observed to be tight and obtained at low computational cost.

2 Sums of Independent and Identically Distributed Random Variables

In this section, upper bounds on the absolute error of approximating F_{X_n} by the *normal approximation* and the *saddlepoint approximation* are presented.

2.1 Error Induced by the Normal Approximation

Given a random variable Y, let the function $\xi_Y : \mathbb{R} \to \mathbb{R}$ be for all $t \in \mathbb{R}$:

$$\xi_Y(t) \triangleq \mathbb{E}_{P_Y}\left[\left|Y - K_Y^{(1)}(t)\right|^3 \exp\left(tY - K_Y(t)\right)\right]. \tag{22}$$

The following theorem, known as the Berry-Esseen theorem [4], introduces an upper bound on the approximation error induced by the normal approximation.

Theorem 1 (Berry-Esseen [4]) *Let Y_1, Y_2, ..., Y_n be i.i.d random variables with probability distribution P_Y. Let also Z_n be a Gaussian random variable with mean $n\,K_Y^{(1)}(0)$, variance $n\,K_Y^{(2)}(0)$ and CDF denoted by F_{Z_n}. Then, the CDF of the random variable $X_n = Y_1 + Y_2 + \ldots + Y_n$, denoted by F_{X_n}, satisfies*

$$\sup_{a\in\mathbb{R}} \left|F_{X_n}(a) - F_{Z_n}(a)\right| \leqslant \min\left(1, \frac{c\,\xi_Y(0)}{\sqrt{n(K_Y^{(2)}(0))^3}}\right), \tag{23}$$

where $c = 0.476$ and the functions $K_Y^{(1)}$, $K_Y^{(2)}$ and ξ_Y are defined in (9), (10), and (22).

The choice of $c = 0.476$ in Theorem 1 is justified in [11]. An immediate result from Theorem 1 consists in the following upper and lower bounds on $F_{X_n}(a)$, for all $a \in \mathbb{R}$,

$$F_{X_n}(a) \leqslant F_{Z_n}(a) + \min\left(1, \frac{c\,\xi_Y(0)}{\sqrt{n(K_Y^{(2)}(0))^3}}\right) \triangleq \bar{\Sigma}(a,n), \text{ and} \tag{24}$$

$$F_{X_n}(a) \geqslant F_{Z_n}(a) - \min\left(1, \frac{c\,\xi_Y(0)}{\sqrt{n(K_Y^{(2)}(0))^3}}\right) \triangleq \underline{\Sigma}(a,n). \tag{25}$$

The main drawback of Theorem 1 is that the upper bound on the approximation error does not depend on the exact value of a. More importantly, for some values of a and n, the upper bound on the approximation error might be particularly big, which leads to irrelevant results.

2.2 Error Induced by the Saddlepoint Approximation

The following theorem introduces an upper bound on the approximation error induced by approximating the CDF F_{X_n} of X_n in (1) by the function $\eta_Y : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ defined such that for all $(\theta, a, n) \in \mathbb{R}^2 \times \mathbb{N}$,

$$\eta_Y(\theta,a,n) \triangleq$$

$$\mathbb{1}_{\{\theta>0\}} + (-1)^{\mathbb{1}_{\{\theta>0\}}} \exp\left(\frac{1}{2}n\theta^2 K_Y^{(2)}(\theta) + nK_Y(\theta) - n\theta K_Y^{(1)}(\theta)\right) Q\left((-1)^{\mathbb{1}_{\{\theta\leqslant 0\}}} \frac{a + n\theta K_Y^{(2)}(\theta) - nK_Y^{(1)}(\theta)}{\sqrt{nK_Y^{(2)}(\theta)}}\right),$$

(26)

where the function $Q : \mathbb{R} \rightarrow [0,1]$ is the complementary CDF of the standard Gaussian distribution defined in (13). Note that $\eta_Y(\theta, n, a)$ is identical to $\hat{F}_{X_n}(a)$, when θ is chosen to satisfy the saddlepoint $K_Y^{(1)}(\theta) = \frac{a}{n}$. Note also that $\eta_Y(0, n, a)$ is the CDF of a Gaussian random variable with mean $nK_Y^{(1)}(0)$ and variance $nK_Y^{(2)}(0)$, which are the mean and the variance of X_n in (1), respectively.

Theorem 2 *Let Y_1, Y_2, ..., Y_n be i.i.d. random variables with probability distribution P_Y and CGF K_Y. Let also F_{X_n} be the CDF of the random variable $X_n = Y_1 + Y_2 + \ldots + Y_n$. Hence, for all $a \in \mathbb{R}$ and for all $\theta \in \Theta_Y$, it holds that*

$$|F_{X_n}(a) - \eta_Y(\theta, a, n)| \leqslant \exp(nK_Y(\theta) - \theta\, a) \min\left(1, \frac{2\, c\, \xi_Y(\theta)}{(K_Y^{(2)}(\theta))^{3/2}\sqrt{n}}\right), \tag{27}$$

where $c = 0.476$;

$$\Theta_Y \triangleq \{t \in \mathbb{R} : K_Y(t) < \infty\}; \tag{28}$$

and the functions $K_Y^{(2)}$, ξ_Y, and η_Y are defined in (10), (22), *and* (26), *respectively.*

Proof: The proof of Theorem 2 is presented in Appendix A. ■

The relevance of Theorem 2 is that given a pair $(a, n) \in \mathbb{R} \times \mathbb{N}$, the value $F_{X_n}(a)$ can be approximated by $\eta_Y(\theta, a, n)$ up to an approximation error that is not bigger than $\exp\big(nK_Y(\theta) -\theta\; a\big) \min\left(1, \frac{2\, c\, \xi_Y(\theta)}{(K_Y^{(2)}(\theta))^{3/2}\sqrt{n}}\right)$. This observation leads to the following upper and lower bounds on $F_{X_n}(a)$, for all $a \in \mathbb{R}$,

$$F_{X_n}(a) \leqslant \eta_Y(\theta, a, n) + \exp(nK_Y(\theta) - \theta\, a) \min\left(1, \frac{2\, c\, \xi_Y(\theta)}{(K_Y^{(2)}(\theta))^{3/2}\sqrt{n}}\right), \text{ and} \tag{29}$$

$$F_{X_n}(a) \geqslant \eta_Y(\theta, a, n) - \exp(nK_Y(\theta) - \theta\, a) \min\left(1, \frac{2\, c\, \xi_Y(\theta)}{(K_Y^{(2)}(\theta))^{3/2}\sqrt{n}}\right), \tag{30}$$

with $\theta \in \Theta_Y$.

The advantages of approximating F_{X_n} by using Theorem 2 instead of Theorem 1 are twofold. First, both the approximation η_Y and the corresponding approximation error depend on the exact value of a. In particular, the approximation can be optimized for each value of a via the parameter θ. Second, the parameter θ in (27) can be optimized to improve either the upper bound in (29) or the lower bound in (30) for some $a \in \mathbb{R}$. Nonetheless, such optimizations are not necessarily simple.

An alternative to the optimization on θ in (29) and (30) is to choose θ such that it minimizes $nK_Y(\theta) - \theta\, a$. This follows the intuition that, for some values of a and n, the term $\exp(nK_Y(\theta) - \theta\, a)$ is the one that influences the most the value of the right-hand side of (27). To build upon this idea, consider the following lemma.

Lemma 1 *Consider a random variable Y with probability distribution P_Y and CGF K_Y. Given $n \in \mathbb{N}$, let the function $h : \mathbb{R} \rightarrow \mathbb{R}$ be defined for all $a \in \mathbb{R}$ satisfying $\frac{a}{n} \in \text{int}\mathcal{C}_Y$, with $\text{int}\mathcal{C}_Y$ denoting the interior of the convex hull of* $\text{supp}\, P_{X_n}$*, as follows*

$$h(a) = \inf_{\theta \in \Theta_Y} nK_Y(\theta) - \theta\, a, \tag{31}$$

where Θ_Y is defined in (28)*. Then, the function h is concave and for all $a \in \mathbb{R}$,*

$$h(a) \leqslant h(n\mathbb{E}_{P_Y}[Y]) = 0. \tag{32}$$

Furthermore,

$$h(a) = nK_Y(\theta^\star) - \theta^\star a, \tag{33}$$

where $\theta^\star$ is the unique solution in θ to

$$nK_Y^{(1)}(\theta) = a, \tag{34}$$

with $K_Y^{(1)}$ is defined in (9)*.*

Proof: The proof of Lemma 1 is presented in Appendix B. ■

Given $(a, n) \in \mathbb{R} \times \mathbb{N}$, the value of $h(a)$ in (31) is the argument that minimizes the exponential term in (27). An interesting observation from Lemma 1 is that the maximum of h is zero and it is reached when $a = n\mathbb{E}_{P_Y}[Y] = \mathbb{E}_{P_{X_n}}[X_n]$. In this case, $\theta^\star = 0$, and thus, from (29) and (30), it holds that

$$\begin{aligned} F_{X_n}(a) &\leqslant \eta_Y(0, a, n) + \min\left(1, \frac{2\,c\,\xi_Y(0)}{(K_Y^{(2)}(0))^{3/2}\sqrt{n}}\right) \\ &= F_{Z_n}(a) + \min\left(1, \frac{2\,c\,\xi_Y(0)}{(K_Y^{(2)}(0))^{3/2}\sqrt{n}}\right), \text{ and} \end{aligned} \tag{35}$$

$$\begin{aligned} F_{X_n}(a) &\geqslant \eta_Y(0, a, n) - \min\left(1, \frac{2\,c\,\xi_Y(0)}{(K_Y^{(2)}(0))^{3/2}\sqrt{n}}\right) \\ &= F_{Z_n}(a) - \min\left(1, \frac{2\,c\,\xi_Y(0)}{(K_Y^{(2)}(0))^{3/2}\sqrt{n}}\right), \end{aligned} \tag{36}$$

where F_{Z_n} is the CDF defined in Theorem 1. Hence, the upper bound in (35) and the lower bound in (36) obtained from Theorem 2 are worse than those in (24) and (25) obtained from Theorem 1. In a nutshell, for values of a around the vicinity of $n\mathbb{E}_{P_Y}[Y] = \mathbb{E}_{P_{X_n}}[X_n]$, it is more interesting to use Theorem 1 instead of Theorem 2.

Alternatively, given that h is non-positive and concave, when $|a - n\mathbb{E}_{P_Y}[Y]| = \left|a - \mathbb{E}_{P_{X_n}}[X_n]\right| > \gamma$, with γ sufficiently large, it follows that

$$\exp\left(nK_Y(\theta^\star) - \theta^\star a\right) < \min\left(1, \frac{c\,\xi_Y(0)}{\sqrt{n(K_Y^{(2)}(0))^3}}\right), \tag{37}$$

with $\theta^\star$ defined in (34). Hence, in this case, the right-hand side of (27) is always smaller than the right-hand side of (23). That is, for such values of a and n, the upper and lower bounds in (29) and (30) are better than those in (24) and (25), respectively. The following theorem leverages this observation.

Theorem 3 *Let Y_1, Y_2, …, Y_n be i.i.d. random variables with probability distribution P_Y and CGF K_Y. Let also F_{X_n} be the CDF of the random variable $X_n = Y_1 + Y_2 + \ldots + Y_n$. Hence, for all $a \in \operatorname{int}\mathcal{C}_{X_n}$, with $\operatorname{int}\mathcal{C}_{X_n}$ the interior of the convex hull of $\operatorname{supp}P_{X_n}$, it holds that*

$$\left|F_{X_n}(a) - \hat{F}_{X_n}(a)\right| \leqslant \exp\left(nK_Y(\theta^\star) - \theta^\star a\right) \min\left(1, \frac{2\,c\,\xi_Y(\theta^\star)}{\sqrt{n}\left(K_Y^{(2)}(\theta^\star)\right)^{3/2}}\right), \tag{38}$$

where $\theta^\star$ *is defined in* (34), $c = 0.476$, *and the functions* $K_Y^{(2)}$, $\hat{F}_{X_n}$, *and* ξ_Y *are defined in* (10), (12), *and* (22), *respectively.*

Proof: The proof of Theorem 3 is presented in Appendix C. ■

An immediate result from Theorem 3 consists in the following upper and lower bounds on $F_X(a)$, for all $a \in \mathbb{R}$,

$$F_{X_n}(a) \leqslant \hat{F}_{X_n}(a) + \exp\left(nK_Y(\theta^\star) - \theta^\star a\right) \min\left(1, \frac{2\,c\,\xi_Y(\theta^\star)}{\left(K_Y^{(2)}(\theta^\star)\right)^{3/2}\sqrt{n}}\right) \triangleq \bar{\Omega}(a,n), \text{ and} \tag{39}$$

$$F_{X_n}(a) \geqslant \hat{F}_{X_n}(a) - \exp\left(nK_Y(\theta^\star) - \theta^\star a\right) \min\left(1, \frac{2\,c\,\xi_Y(\theta^\star)}{\left(K_Y^{(2)}(\theta^\star)\right)^{3/2}\sqrt{n}}\right) \triangleq \underline{\Omega}(a,n). \tag{40}$$

The following section presents two examples that highlight the observations mentioned above.

2.3 Examples

Example 1 (Discrete random variable) *Let the random variables* Y_1, Y_2, $\ldots$, Y_n *in* (1) *be i.i.d. Bernoulli random variables with parameter* $p = 0.2$ *and* $n = 100$. *In this case* $\mathbb{E}_{P_{X_n}}[X_n] = n\mathbb{E}_{P_Y}[Y] = 20$. *Figure 1 depicts the CDF* $F_{X_{100}}$ *of* X_{100} *in* (1); *the normal approximation* $F_{Z_{100}}$ *in* (23); *and the saddlepoint approximation* $\hat{F}_{X_{100}}$ *in* (12). *Therein, it is also depicted the upper and lower bounds due to the normal approximation* $\bar{\Sigma}$ *in* (24) *and* $\underline{\Sigma}$ *in* (25), *respectively; and the upper and lower bounds due to the saddlepoint approximation* $\bar{\Omega}$ *in* (39) *and* $\underline{\Omega}$ *in* (40), *respectively. These functions are plotted as a function of* a, *with* $a \in [5, 35]$. *Figure 2 and Figure 3 depict the same functions as a function of* a, *with* $a \in [0, 5]$ *and* $a \in [50, 60]$, *respectively.*

Example 2 (Continuous random variable) *Let the random variables* Y_1, Y_2, $\ldots$, Y_n *in* (1) *be i.i.d. chi-squared random variables with parameter* $k = 1$ *and* $n = 50$. *In this case* $\mathbb{E}_{P_{X_n}}[X_n] = n\mathbb{E}_{P_Y}[Y] = 50$. *Figure 4 depicts the CDF* $F_{X_{50}}$ *of* X_{50} *in* (1); *the normal approximation* $F_{Z_{50}}$ *in* (23); *and the saddlepoint approximation* $\hat{F}_{X_{50}}$ *in* (12). *Therein, it is also depicted the upper and lower bounds due to the normal approximation* $\bar{\Sigma}$ *in* (24) *and* $\underline{\Sigma}$ *in* (25), *respectively; and the upper and lower bounds due to the saddlepoint approximation* $\bar{\Omega}$ *in* (39) *and* $\underline{\Omega}$ *in* (40), *respectively. These functions are plotted as a function of* a, *with* $a \in [0, 100]$. *Figure 5 and Figure 6 depict the same functions as a function of* a, *with* $a \in [0, 25]$ *and* $a \in [100, 170]$, *respectively.*

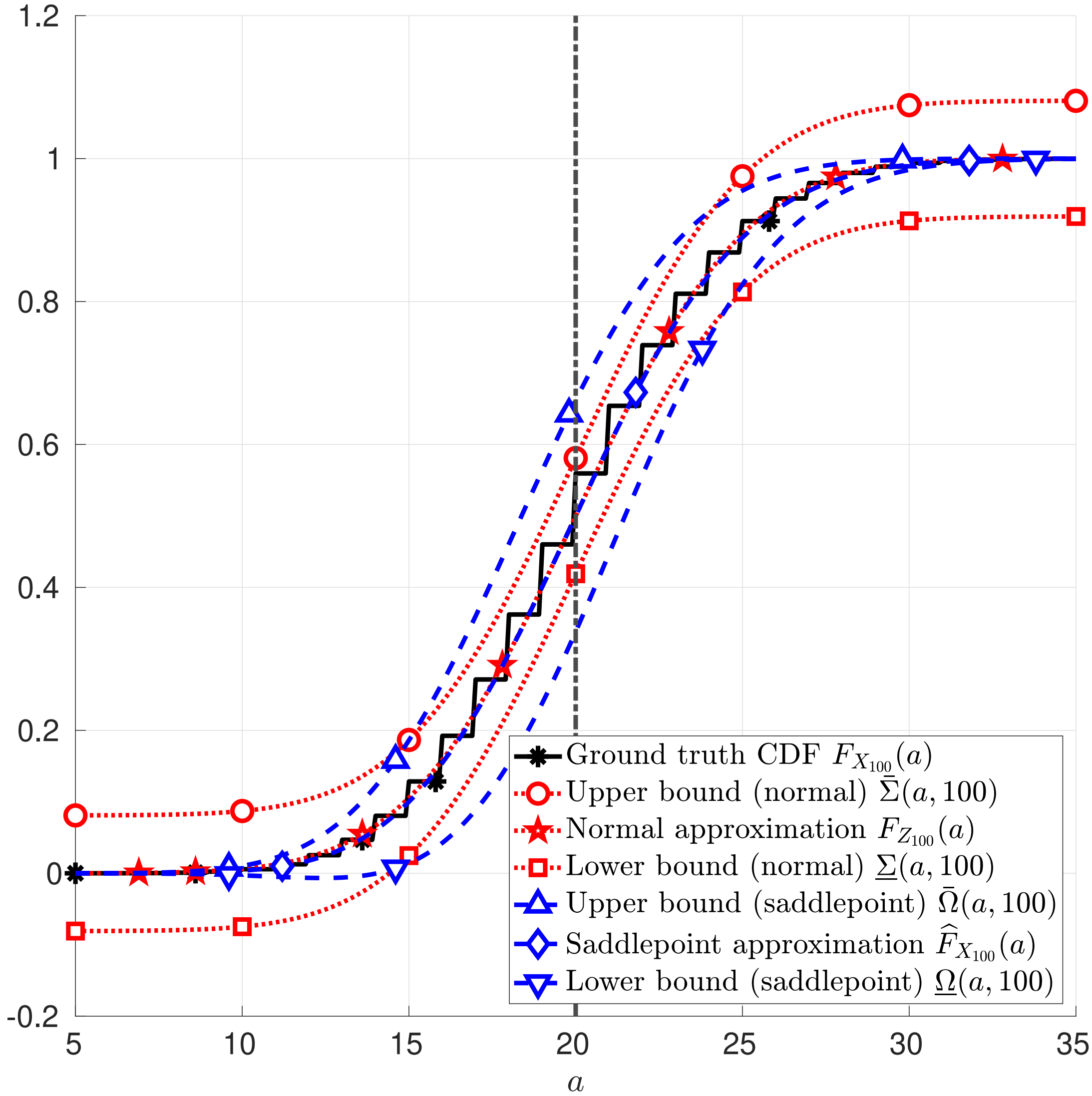

Figure 1: Sum of 100 Bernoulli random variables with parameter $p = 0.2$. Note that $\mathbb{E}[X_{100}] = 20$. The function $F_{X_{100}}(a)$ (asterisk markers $*$) in Example 1; the function $F_{Z_{100}}(a)$ (star markers $\star$) in (23); the function $\hat{F}_{X_{100}}(a)$ (diamond markers $\diamond$) in (12); the function $\bar{\Sigma}(a, 100)$ (circle marker $\circ$) in (24); the function $\underline{\Sigma}(a, 100)$ (square marker $\square$) in (25); the function $\bar{\Omega}(a, 100)$ (upward-pointing triangle marker $\triangle$) in (39); and the function $\underline{\Omega}(a, 100)$ (downward-pointing triangle marker $\triangledown$) in (40) as a function of a, with $a \in [5, 35]$.

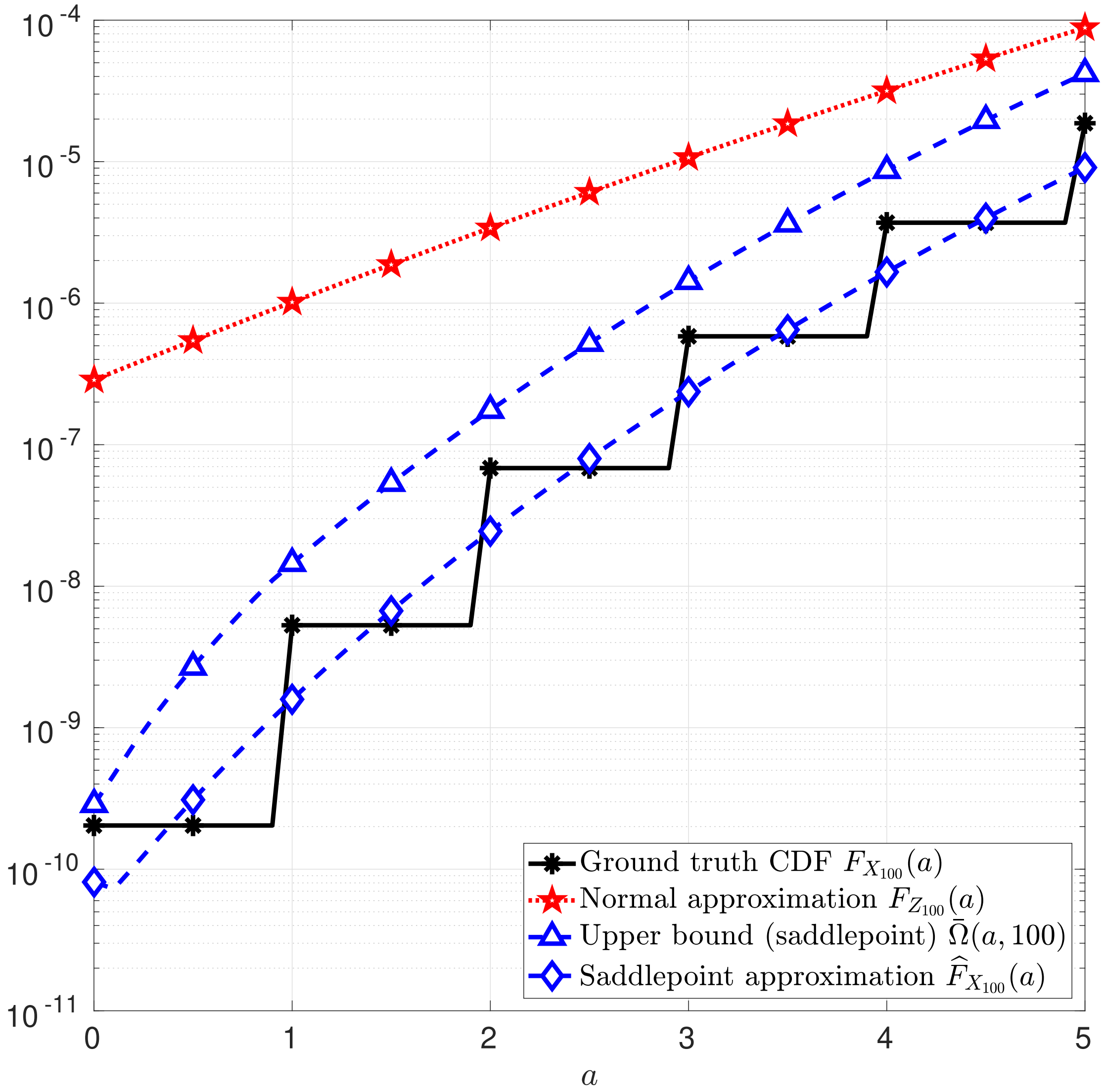

Figure 2: Sum of 100 Bernoulli random variables with parameter $p = 0.2$. Note that $\mathbb{E}[X_{100}] = 20$. The function $F_{X_{100}}(a)$ (asterisk markers $*$) in Example 1; the function $F_{Z_{100}}(a)$ (star markers $\star$) in (23); the function $\hat{F}_{X_{100}}(a)$ (diamond markers $\diamond$) in (12); $\bar{\Omega}(a, 100)$ (upward-pointing triangle marker $\triangle$) in (39) as a function of a, with $a \in [0, 5]$.

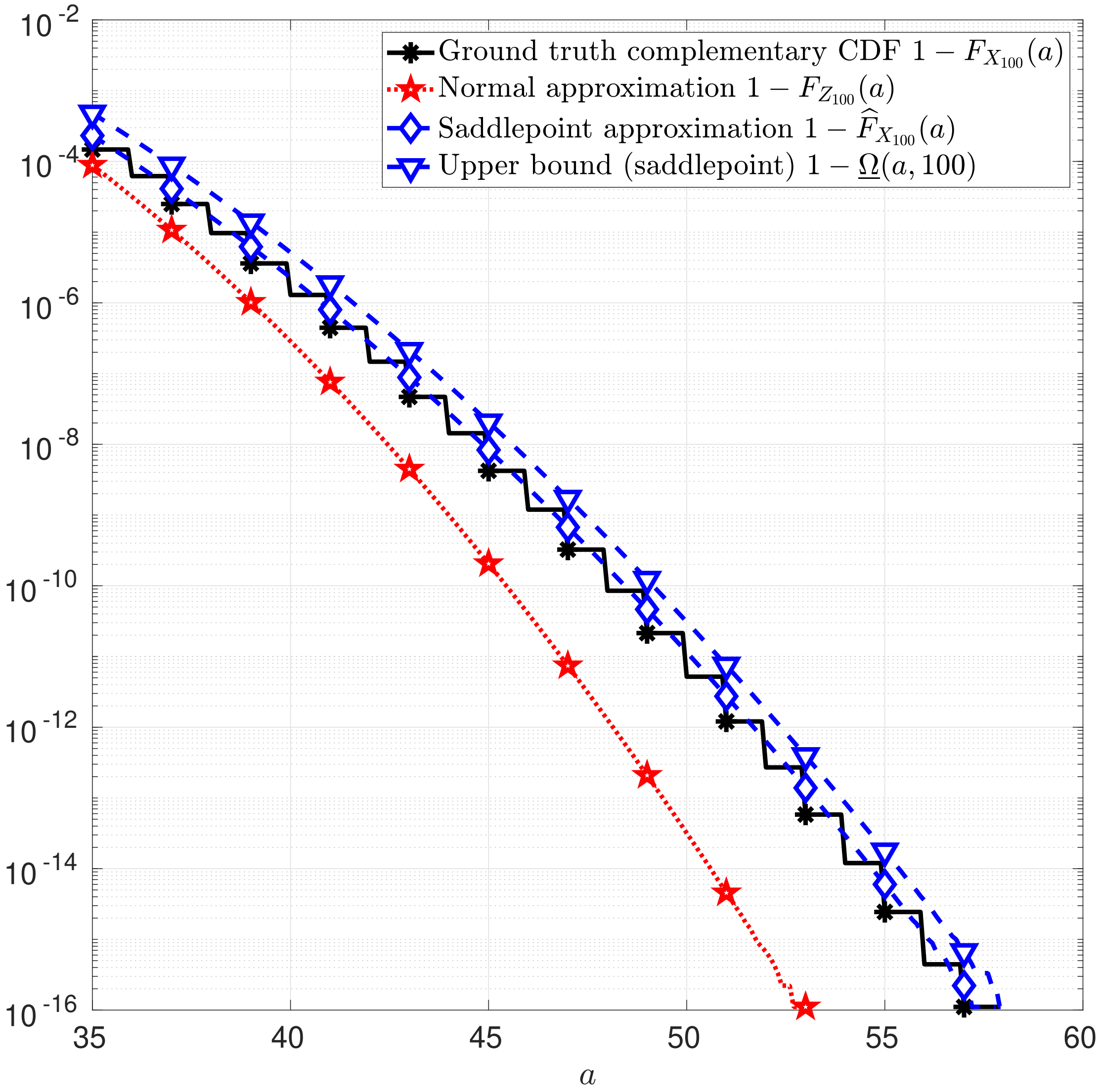

Figure 3: Sum of 100 Bernoulli random variables with parameter $p = 0.2$. Note that $\mathbb{E}[X_{100}] = 20$. The complementary CDF $1 - F_{X_{100}}(a)$ (asterisk markers $*$) in Example 1; the function $1 - F_{Z_{100}}(a)$ (star markers $\star$) in (23); the function $1 - \hat{F}_{X_{100}}(a)$ (diamond markers $\diamond$) in (12); The function $1 - \bar{\Omega}(a, 100)$ (downward-pointing triangle marker $\triangledown$) in (40) as a function of a, with $a \in [35, 60]$.

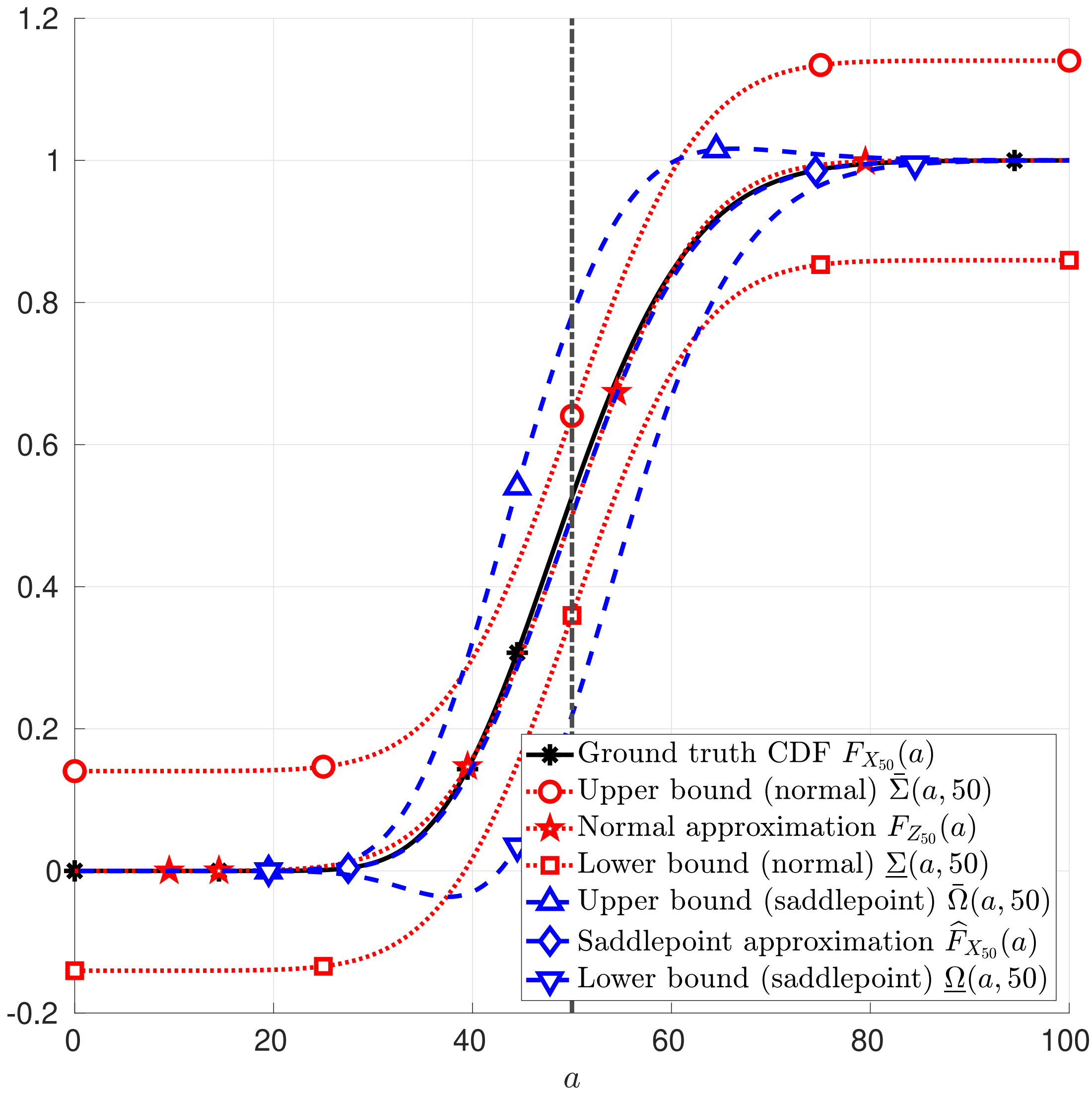

Figure 4: Sum of 50 Chi-squared random variables with parameter $k = 1$. Note that $\mathbb{E}[X_{50}] = 50$. The function $F_{X_{50}}(a)$ (asterisk markers $*$) in Example 2; the function $F_{Z_{50}}(a)$ (star markers $\star$) in (23); the function $\hat{F}_{X_{50}}(a)$ (diamond markers $\diamond$) in (12); the function $\bar{\Sigma}(a, 50)$ (circle marker $\circ$) in (24); the function $\underline{\Sigma}(a, 50)$ (square marker $\square$) in (25); $\bar{\Omega}(a, 50)$ (upward-pointing triangle marker $\triangle$) in (39); and the function $\underline{\Omega}(a, 50)$ (downward-pointing triangle marker $\triangledown$) in (40) as a function of a, with $a \in [0, 100]$.

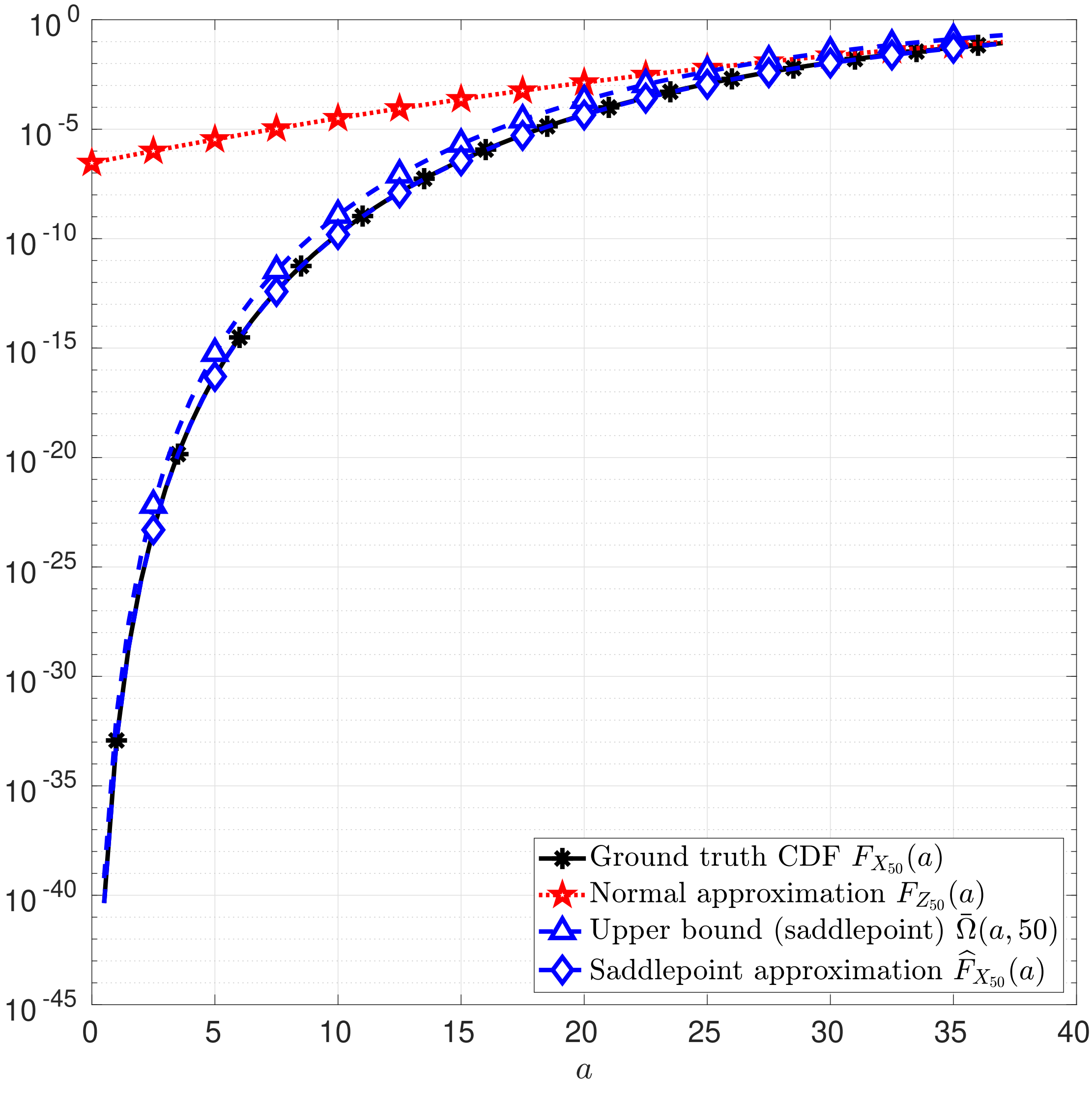

Figure 5: Sum of 50 Chi-squared random variables with parameter $k = 1$. Note that $\mathbb{E}[X_{50}] = 50$. The function $F_{X_{50}}(a)$ (asterisk markers $*$) in Example 2; the function $F_{Z_{50}}(a)$ (star markers $\star$) in (23); the function $\hat{F}_{X_{50}}(a)$ (diamond markers $\diamond$) in (12); $\bar{\Omega}(a, 50)$ (upward-pointing triangle marker $\triangle$) in (39) as a function of a, with $a \in [0, 40]$.

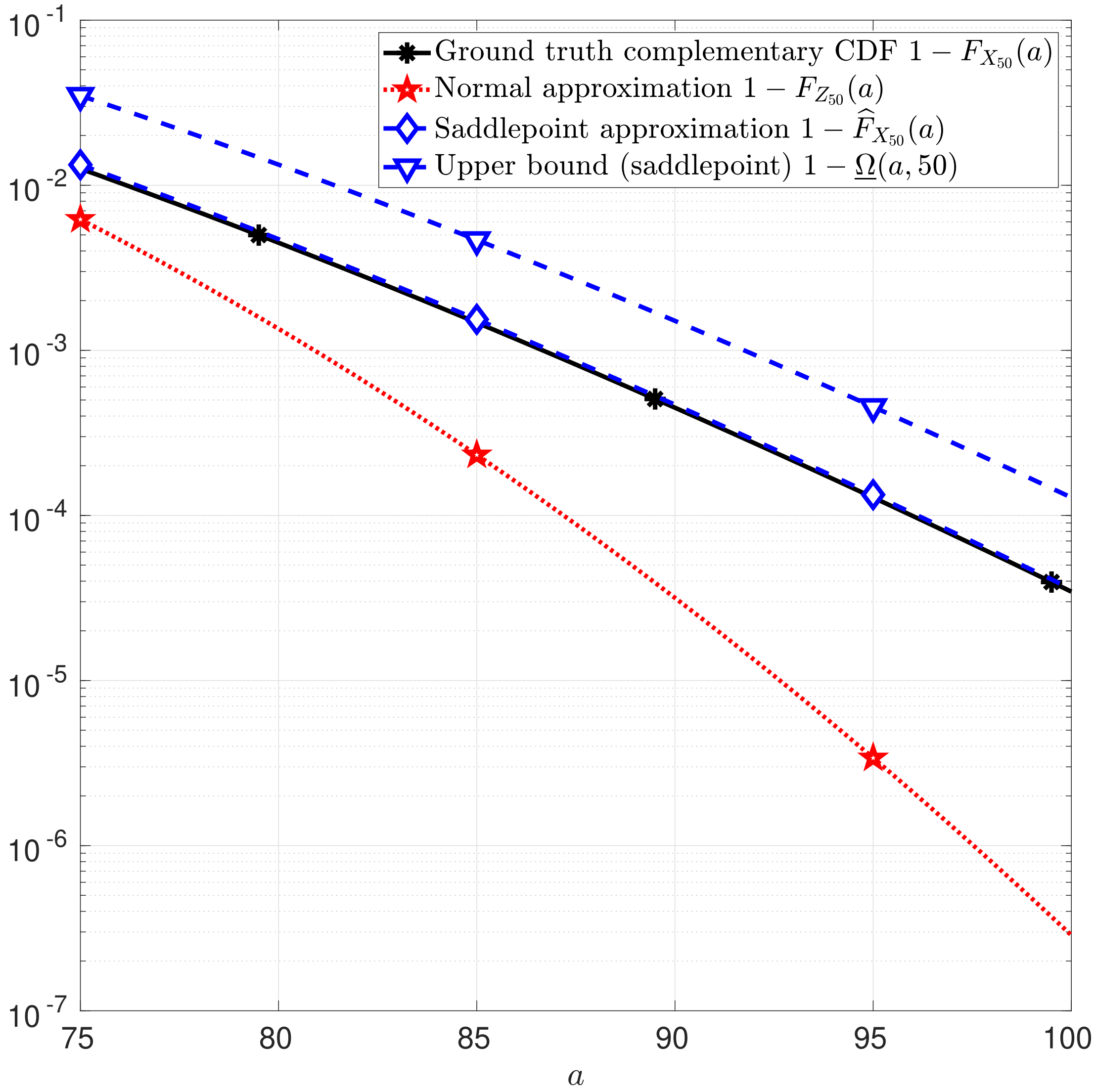

Figure 6: Sum of 50 Chi-squared random variables with parameter $k = 1$. Note that $\mathbb{E}[X_{50}] = 50$. The complementary CDF $1 - F_{X_{50}}(a)$ (asterisk markers $*$) in Example 2; the function $1 - F_{Z_{50}}(a)$ (star markers $\star$) in (23); the function $1 - \hat{F}_{X_{50}}(a)$ (diamond markers $\diamond$) in (12); $\underline{\Omega}(a, 50)$ (downward-pointing triangle marker $\triangledown$) in (40) as a function of a, with $a \in [70, 100]$.

3 Application to Information Theory: Channel Coding

This section focuses on the study of the DEP in point-to-point memoryless channels. The problem is formulated in Section 3.1. The main results presented in this section consist in lower and upper bounds on the DEP. The former, which are obtained building upon the existing DT bound [12], are presented in Section 3.2. The latter, which are obtained from the MC bound [12], are presented in Section 3.3.

3.1 System Model

Consider a point-to-point communication in which a transmitter aims at sending information to one receiver through a noisy memoryless channel. Such a channel can be modeled by a random transformation

$$(\mathcal{X}^n, \mathcal{Y}^n, P_{\boldsymbol{Y}|\boldsymbol{X}}), \tag{41}$$

where $n \in \mathbb{N}$ is the blocklength and $\mathcal{X}$ and $\mathcal{Y}$ are the channel input and channel output sets. Given the channel inputs $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, the outputs $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$ are observed at the receiver with probability

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{t=1}^{n} P_{Y|X}(y_t|x_t), \tag{42}$$

where, for all $x \in \mathcal{X}$, $P_{Y|X=x} \in \triangle(\mathcal{Y})$, with $\triangle(\mathcal{Y})$ the set of all possible probability distributions whose support is a subset of $\mathcal{Y}$. The objective of the communication is to transmit a message index i, which is a realization of a random variable W that is uniformly distributed over the set

$$\mathcal{W} \triangleq \{1, 2, \ldots, M\}, \tag{43}$$

with $1 < M < \infty$. To achieve this objective, the transmitter uses an (n, M, λ)-code, where $\lambda \in [0, 1]$.

Definition 1 **((n, M, λ)-code)** *Given a tuple $(M, n, \lambda) \in \mathbb{N}^2 \times [0,1]$, an (n, M, λ)-code for the random transformation in* (41) *is a system*

$$\left\{\Big(\boldsymbol{u}(1), \mathcal{D}(1)\Big), \Big(\boldsymbol{u}(2), \mathcal{D}(2)\Big), \ldots, \Big(\boldsymbol{u}(M), \mathcal{D}(M)\Big)\right\}, \tag{44}$$

where for all $(j, \ell) \in \mathcal{W}^2$, with $j \neq \ell$:

$$\boldsymbol{u}(j) = (u_1(j), u_2(j), \ldots, u_n(j)) \in \mathcal{X}^n, \tag{45a}$$

$$\mathcal{D}(j) \cap \mathcal{D}(\ell) = \emptyset, \tag{45b}$$

$$\bigcup_{j \in \mathcal{W}} \mathcal{D}(j) \subseteq \mathcal{Y}^n, \text{ and} \tag{45c}$$

$$\frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}(i)}}\left[\mathbb{1}_{\{\boldsymbol{Y} \notin \mathcal{D}(i)\}}\right] \leqslant \lambda. \tag{45d}$$

To transmit message index $i \in \mathcal{W}$, the transmitter uses the codeword $\boldsymbol{u}(i)$. For all $t \in \{1, 2, \ldots, n\}$, at channel use t, the transmitter inputs the symbol $u_t(i)$ into the channel. Assume that at the end of channel use t, the receiver observes the output y_t. After n channel uses, the receiver

uses the vector $\boldsymbol{y} = (y_1,y_2,\ldots, y_n)$ and determines that the symbol j was transmitted if $\boldsymbol{y} \in \mathcal{D}(j)$, with $j \in \mathcal{W}$.
Given the (n,M,λ)-code described by the system in (44), the DEP of the message index i is $\mathbb{E}_{P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}(i)}}\left[\mathbb{1}_{\{Y\notin\mathcal{D}(i)\}}\right]$. As a consequence, the average DEP is

$$\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}_{P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}(i)}}\left[\mathbb{1}_{\{Y\notin\mathcal{D}(i)\}}\right]. \tag{46}$$

Note that from (45d), the average DEP of such an (n, M, λ)-code is upper bounded by λ. Given a fixed pair $(n,M) \in \mathbb{N}^2$, the minimum λ for which an (n,M,λ)-code exists is defined hereunder.

Definition 2 *Given a pair $(n,M) \in \mathbb{N}^2$, the minimum average decoding error probability for the random transformation in* (41)*, denoted by $\lambda^*(n, M)$, is given by*

$$\lambda^*(n, M) = \min\left\{\lambda \in [0, 1] : \exists(n, M, \lambda)\text{-}code\right\}. \tag{47}$$

When λ is chosen accordingly with the reliability constraints, an (n, M, λ)-code is said to transmit at an information rate $R = \frac{\log_2(M)}{n}$ bits per channel use.
The remainder of this section introduces the DT bound and the MC bound. The DT bound is one of the tightest existing upper bounds on $\lambda^*(n, M)$ in (47), whereas the MC bound is one of the tightest lower bounds.

3.2 Dependence Testing Bound

This section describes an upper bound on $\lambda^*(n, M)$, for a fixed pair $(n, M) \in \mathbb{N}^2$. Given a probability distribution $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$, let the random variable $\iota(\boldsymbol{X};\boldsymbol{Y})$ satisfy

$$\iota(\boldsymbol{X};\boldsymbol{Y}) \triangleq \ln\left(\frac{\mathrm{d}P_{\boldsymbol{X}\boldsymbol{Y}}}{\mathrm{d}P_{\boldsymbol{X}}P_{\boldsymbol{Y}}}(\boldsymbol{X},\boldsymbol{Y})\right), \tag{48}$$

where the function $\frac{\mathrm{d}P_{\boldsymbol{X}\boldsymbol{Y}}}{\mathrm{d}P_{\boldsymbol{X}}P_{\boldsymbol{Y}}} : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}$ denotes the Radon-Nikodym derivative of the joint probability measure $P_{\boldsymbol{X}\boldsymbol{Y}}$ with respect to the product of probability measures $P_{\boldsymbol{X}}P_{\boldsymbol{Y}}$, with $P_{\boldsymbol{X}\boldsymbol{Y}} = P_{\boldsymbol{X}}P_{\boldsymbol{Y}|\boldsymbol{X}}$ and $P_{\boldsymbol{Y}}$ the corresponding marginal. Let the function $T : \mathbb{N}^2 \times \triangle(\mathcal{X}^n) \to \mathbb{R}_+$ be for all $(n,M) \in \mathbb{N}^2$ and for all probability distributions $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$,

$$T(n, M, P_{\boldsymbol{X}})=\mathbb{E}_{P_{\boldsymbol{X}}P_{\boldsymbol{Y}|\boldsymbol{X}}}\left[\mathbb{1}_{\left\{\iota(\boldsymbol{X};\boldsymbol{Y})\leqslant\ln\left(\frac{M-1}{2}\right)\right\}}\right] + \frac{M-1}{2}\mathbb{E}_{P_{\boldsymbol{X}}P_{\boldsymbol{Y}}}\left[\mathbb{1}_{\left\{\iota(\boldsymbol{X};\boldsymbol{Y})>\ln\left(\frac{M-1}{2}\right)\right\}}\right]. \tag{49}$$

Using this notation, the following lemma states the dependence testing bound.

Lemma 2 (Dependence testing bound [12]) *Given a pair $(n,M) \in \mathbb{N}^2$, the following holds for all $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$, with respect to the random transformation in* (41)*:*

$$\lambda^*(n, M) \leqslant T(n, M, P_{\boldsymbol{X}}), \tag{50}$$

with the function T defined in (49)*.*

Note that the input probability distribution $P_{\boldsymbol{X}}$ in Lemma 2 can be chosen among all possible probability distributions $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$ to minimize the right-hand side of (50), which improves the bound. Note also that with some lost of optimality, the optimization domain can be constrained to the set of probability distributions for which for all $\boldsymbol{x} \in \mathcal{X}^n$,

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{t=1}^{n} P_X(x_t), \tag{51}$$

with $P_X \in \triangle(\mathcal{X})$. Hence, subject to (42), the random variable $\iota(\boldsymbol{X};\boldsymbol{Y})$ in (48) can be written as the sum of i.i.d. random variables, i.e.,

$$\iota(\boldsymbol{X};\boldsymbol{Y}) = \sum_{t=1}^{n} \iota(X_t;Y_t). \tag{52}$$

This observation motivates the application of the results of Section 2 to provide upper and lower bounds on the function T in (49), for some given values $(n, M) \in \mathbb{N}^2$ and a given distribution $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$ for the random transformation in (41) subject to (42). These bounds become significantly relevant when the exact value of $T(n, M, P_{\boldsymbol{X}})$ cannot be calculated with respect to the random transformation in (41). In such a case, providing upper and lower bounds on $T(n, M, P_{\boldsymbol{X}})$ helps in approximating its exact value subject to an error sufficiently small such that the approximation is relevant.

3.2.1 Normal Approximation

This section describes the normal approximation of the function T in (49). That is, the random variable $\iota(\boldsymbol{X};\boldsymbol{Y})$ is assumed to satisfy (52) and to follow a Gaussian distribution. More specifically, for all $P_X \in \triangle(\mathcal{X})$, let

$$\mu(P_X) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\iota(X;Y)\right], \tag{53}$$

$$\sigma(P_X) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left(\iota(X;Y) - \mu(P_X)\right)^2\right], \text{ and} \tag{54}$$

$$\xi(P_X) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left|\iota(X;Y) - \mu(P_X)\right|^3\right], \tag{55}$$

be the first moment; the second central moment; and the third absolute central moment of the random variables $\iota(X_1;Y_1)$, $\iota(X_2;Y_2)$ $\ldots$ $\iota(X_n;Y_n)$. Using this notation consider the functions $D : \mathbb{N}^2 \times \triangle(\mathcal{X}) \to \mathbb{R}_+$ and $N : \mathbb{N}^2 \times \triangle(\mathcal{X}) \to \mathbb{R}_+$ such that for all $(n, M) \in \mathbb{N}^2$ and for all $P_X \in \triangle(\mathcal{X})$,

$$D(n, M, P_X) = \max\left(0, \alpha(n, M, P_X) - \frac{c\,\xi(P_X)}{\sigma(P_X)^{\frac{3}{2}}\sqrt{n}}\right), \text{ and} \tag{56}$$

$$N(n, M, P_X) = \min\left(1, \alpha(n, M, P_X) + \frac{3\,c\,\xi(P_X)}{\sigma(P_X)^{\frac{3}{2}}\sqrt{n}} + \frac{2\ln(2)}{\sigma(P_X)^{\frac{1}{2}}\sqrt{2n\pi}}\right), \tag{57}$$

where $c = 0.476$ and

$$\alpha(n, M, P_X) \triangleq Q\left(\frac{n\mu(P_X) - \ln\left(\frac{M-1}{2}\right)}{\sqrt{n\sigma(P_X)}}\right). \tag{58}$$

Using this notation, the following theorem introduces a lower bound and an upper bound on T in (49).

Theorem 4 *Given a pair $(n, M) \in \mathbb{N}^2$, for all input distributions $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$ subject to (51), the following holds with respect to the random transformation in (41) subject to (42),*

$$D(n, M, P_X) \leqslant T(n, M, P_{\boldsymbol{X}}) \leqslant N(n, M, P_X), \tag{59}$$

where the functions T, D and N are defined in (49), (56) and (57), respectively.

Proof: The proof of Theorem 4 is presented in [14]. Essentially, it consists in using Theorem 1 for upper and lower bounding the terms $\mathbb{E}_{P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}}\left[\mathbb{1}_{\left\{\iota(\boldsymbol{X};\boldsymbol{Y}) \leqslant \ln\left(\frac{M-1}{2}\right)\right\}}\right]$ in (49). The upper bound on $\mathbb{E}_{P_{\boldsymbol{X}} P_{\boldsymbol{Y}}}\left[\mathbb{1}_{\left\{\iota(\boldsymbol{X};\boldsymbol{Y}) > \ln\left(\frac{M-1}{2}\right)\right\}}\right]$ in (49) follows from Lemma 20 in [17]. ■

In [14], the function $\alpha(n, M, P_X)$ in (58) is often referred to as the *normal approximation* of $T(n, M, P_{\boldsymbol{X}})$, which is indeed a language abuse. In Section 2.1, a comment is given on the fact that the lower and upper bounds, i.e., the functions D in (56) and N in (57), are often too far from the normal approximation α in (58).

3.2.2 Saddlepoint Approximation

This section describes an approximation of the function T in (49) by using the saddlepoint approximation of the CDF of the random variable $\iota(\boldsymbol{X};\boldsymbol{Y})$, as suggested in Section 2.2. Given a distribution $P_X \in \triangle(\mathcal{X})$, the moment generating function of $\iota(X;Y)$ is

$$\varphi(P_X, \theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\exp\left(\theta\, \iota(X;Y)\right)\right], \tag{60}$$

with $\theta \in \mathbb{R}$. For all $P_X \in \triangle(\mathcal{X})$ and for all $\theta \in \mathbb{R}$, consider the following functions:

$$\mu(P_X, \theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\frac{\iota(X;Y)\exp\left(\theta\, \iota(X;Y)\right)}{\varphi(P_X, \theta)}\right], \tag{61}$$

$$V(P_X, \theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left(\iota(X;Y) - \mu(P_X, \theta)\right)^2 \frac{\exp\left(\theta \iota(X;Y)\right)}{\varphi(P_X, \theta)}\right], \text{ and} \tag{62}$$

$$\xi(P_X, \theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left|\iota(X;Y) - \mu(P_X, \theta)\right|^3 \frac{\exp\left(\theta \iota(X;Y)\right)}{\varphi(P_X, \theta)}\right]. \tag{63}$$

Using this notation, consider the functions $\beta_1 : \mathbb{N}^2 \times \mathbb{R} \times \triangle(\mathcal{X}) \to \mathbb{R}_+$ and $\beta_2 : \mathbb{N}^2 \times \mathbb{R} \times \triangle(\mathcal{X}) \to \mathbb{R}_+$:

$$\begin{aligned} &\beta_1(n, M, \theta, P_X) \\ &= \mathbb{1}_{\{\theta > 0\}} + (-1)^{\mathbb{1}_{\{\theta > 0\}}} \exp\left(n \ln\left(\varphi(P_X, \theta)\right) - \theta \ln\left(\frac{M-1}{2}\right) + \frac{1}{2}\theta^2 n V(P_X, \theta)\right) Q\left(\sqrt{n V(P_X, \theta)}|\theta|\right), \end{aligned} \tag{64}$$

and

$$\begin{aligned} &\beta_2(n, M, \theta, P_X) \\ &= \mathbb{1}_{\{\theta \leqslant -1\}} + (-1)^{\mathbb{1}_{\{\theta \leqslant -1\}}} \exp\left(n \ln\left(\varphi(P_X, \theta)\right) - (\theta+1) \ln\left(\frac{M-1}{2}\right) + \frac{1}{2}(\theta+1)^2 n V(P_X, \theta)\right) Q\left(\sqrt{n V(P_X, \theta)}|\theta+1|\right). \end{aligned} \tag{65}$$

Note that β_1 is the saddlepoint approximation of the CDF of the random variable $\iota(\boldsymbol{X};\boldsymbol{Y})$ in (52) when $\boldsymbol{X}$ and $\boldsymbol{Y}$ follow the distribution $P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}$. Note also that β_2 is the saddlepoint

approximation of the complementary CDF of the random variable $\iota(\boldsymbol{X};\boldsymbol{Y})$ in (52) when $\boldsymbol{X}$ and $\boldsymbol{Y}$ follow the distribution $P_{\boldsymbol{X}}P_{\boldsymbol{Y}}$.
Consider also the following functions:

$$G_1(n,M,\theta,P_X) = \beta_1(n,M,\theta,P_X) - \frac{2c\,\xi(P_X,\theta)}{V(P_X,\theta)^{3/2}\sqrt{n}}\exp\left(n\ln\left(\varphi(P_X,\theta)\right) - \theta\ln\left(\frac{M-1}{2}\right)\right), \quad (66)$$

$$G_2(n,M,\theta,P_X) = \beta_2(n,M,\theta,P_X) - \frac{2c\,\xi(P_X,\theta)}{V(P_X,\theta)^{3/2}\sqrt{n}}\exp\left(n\ln\left(\varphi(P_X,\theta)\right) - (\theta+1)\ln\left(\frac{M-1}{2}\right)\right), \quad (67)$$

$$G(n,M,\theta,P_X) = \max\left(0, G_1(n,M,\theta,P_X)\right) + \frac{M-1}{2}\max\left(0, G_2(n,M,\theta,P_X)\right), \quad \text{and} \quad (68)$$

$$S(n,M,\theta,P_X) = \min\left(1, \beta(n,M,\theta,P_X) + \frac{4c\,\xi(P_X,\theta)}{\left(V(P_X,\theta)\right)^{3/2}\sqrt{n}}\exp\left(n\ln\left(\varphi(P_X,\theta)\right) - \theta\ln\left(\frac{M-1}{2}\right)\right)\right). \quad (69)$$

The following theorem introduces new lower and upper bounds on T in (49).

Theorem 5 *Given a pair $(n,M)\in\mathbb{N}^2$, for all input distributions $P_{\boldsymbol{X}}\in\triangle(\mathcal{X}^n)$ subject to* (51)*, the following holds with respect to the random transformation in* (41) *subject to* (42)*,*

$$G(n,M,\theta,P_X) \leqslant T(n,M,P_{\boldsymbol{X}}) \leqslant S(n,M,\theta,P_X) \quad (70)$$

where θ is the unique solution in t to

$$n\mu(P_X,t) = \ln\left(\frac{M-1}{2}\right), \quad (71)$$

and the functions T, G, and S are defined in (49)*,* (68) *and* (69)*, with $c = 0.476$.*

Proof: The proof of Theorem 5 is provided in Appendix F. In a nutshell, the proof consists in using Theorem 3 for independently bounding the terms $\mathbb{E}_{P_{\boldsymbol{X}}P_{\boldsymbol{Y}|\boldsymbol{X}}}\left[\mathbb{1}_{\left\{\iota(\boldsymbol{X};\boldsymbol{Y})\leqslant\ln\left(\frac{M-1}{2}\right)\right\}}\right]$ and $\mathbb{E}_{P_{\boldsymbol{X}}P_{\boldsymbol{Y}}}\left[\mathbb{1}_{\left\{\iota(\boldsymbol{X};\boldsymbol{Y})\geqslant\ln\left(\frac{M-1}{2}\right)\right\}}\right]$ in (49). ■
In the following, the function

$$\beta(n,M,\theta,P_X) = \beta_1(n,M,\theta,P_X) + \frac{M-1}{2}\beta_2(n,M,\theta,P_X), \quad (72)$$

with β_1 in (64) and β_2 in (65), is referred to as the *saddlepoint approximation* of the function T in (49), which is indeed a language abuse.

3.2.3 Numerical Analysis

The normal approximation and the saddlepoint approximation of the DT bound as well as the corresponding upper bounds and lower bounds presented in Section 3.2.1 and in Section 3.2.2 are studied in the cases of the BSC, the AWGN channel, and the SαS channel. The latter is defined by the random transformation in (41) subject to (42) and for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$:

$$P_{Y|X}(y|x) = P_Z(y-x), \quad (73)$$

where P_Z is a probability distribution satisfying for all $t\in\mathbb{R}$,

$$\mathbb{E}_{P_Z}\left[\exp\left(itZ\right)\right] = \exp\left(-\left|\sigma t\right|^{\alpha}\right), \quad (74)$$

with $i = \sqrt{-1}$. The reals $\alpha \in (0, 2]$ and $\sigma \in \mathbb{R}_+$ in (74) are parameters of the SαS channel.
In the following figures, Figure 7 - Figure 9, the function T in (49), which is bounded by using Theorem 4 and Theorem 5, is studied. On the first hand, its normal approximation $\alpha\left(n, 2^{nR}, P_X\right)$ in (58) is plotted in black diamonds, whereas the corresponding lower and upper bounds, i.e., $D\left(n, 2^{nR}, P_X\right)$ in (56) and $N\left(n, 2^{nR}, P_X\right)$ in (57), are respectively plotted in red circles and blue squares. On the second hand, its saddlepoint approximation $\beta\left(n, 2^{nR}, \theta, P_X\right)$ in (72), is plotted in black stars whereas the corresponding upper and lower bounds, i.e., $S\left(n, 2^{nR}, \theta, P_X\right)$ in (69) and $G\left(n, 2^{nR}, \theta, P_X\right)$ in (68), are plotted in blue upward-pointing triangles and red downward-pointing triangles respectively. These functions are plotted only when their values are positive. The channel inputs are discrete $\mathcal{X} = \{-1, 1\}$, P_X is the uniform distribution, and θ is chosen to be the unique solution in t to the equality in (71).
Figure 7 concerns the case of a BSC with cross-over probability $\delta = 0.11$ and $R = 0.32$ bits per channel use. The function T in (49) can be calculated exactly and thus, it is plotted in magenta asterisks. Therein, it can be observed that both the saddlepoint approximation β and the function T overlap. These observations are in line with those reported in [15], in which the saddlepoint approximations of the RCU bound and the MC bound are both shown to be precise approximations. The new bounds provided in Theorem 5 show that the exact value of $T(n, M, P_X)$ is between $S(n, M, \theta, P_X)$ and $G(n, M, \theta, P_X)$. Hence, approximating T in (49) by the function $\alpha(n, M, P_X)$ in (58) might lead to erroneous conclusions. Indeed, when $n > 1000$ for instance, our lower bound $G(n, M, \theta, P_X)$ in (68) becomes bigger than the approximation $\alpha(n, M, P_X)$ in (58) and hence approximating T by α is too optimistic.
Figure 8 and Figure 9 concern the cases of a real-valued AWGN channel and a SαS channel, respectively. Moreover, the signal to noise ratio (SNR) is SNR $= 1$ for both channels. The information rate is $R = 0.425$ bits per channel use for the AWGN channel and $R = 0.38$ bits per channel use for the SαS channel, with $(\alpha, \sigma) = (1.4, 0.6)$. In both cases, the function T in (49) can not be computed explicitly and hence does not appear in Figure 8 and Figure 9. In addition, the lower bound $D(n, M, P_X)$ obtained from Theorem 4 is non-positive in these cases, and thus, does not appear on the figures.
Note that in Figure 7 - Figure 9, the upper bound $N(n, M, P_X)$ is several orders of magnitude far away from the normal approximation $\alpha(n, M, P_X)$. From this perspective, a proper analysis on the DT bound (Lemma 2) based on Theorem 4 does not lead to relevant conclusions.

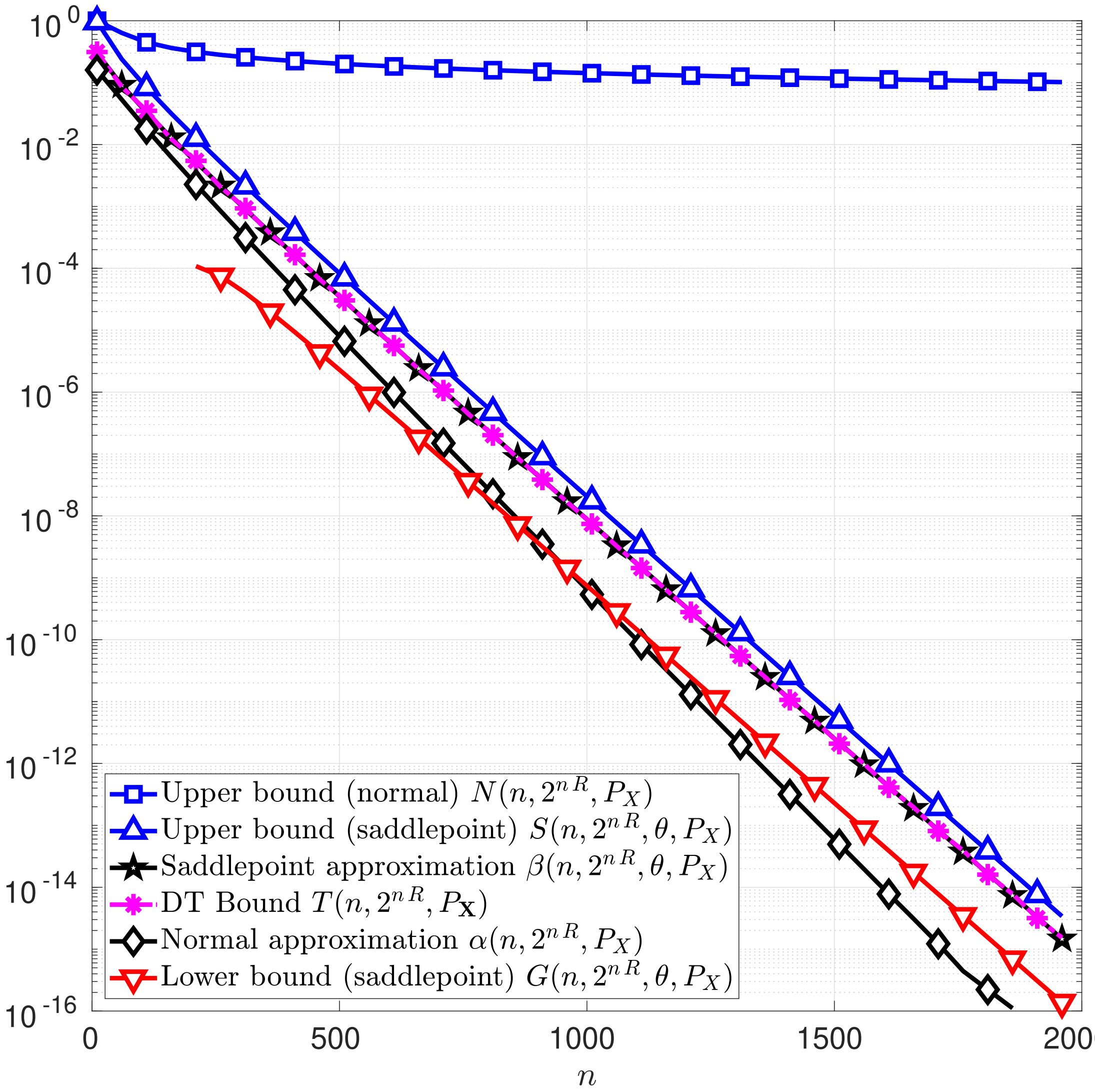

Figure 7: Normal and saddlepoint approximations of the function T in (49) as functions of the blocklength n for the case of a BSC with cross-over probability $\delta = 0.11$ at information rate $R = 0.32$ bits per channel use. The channel input distribution P_X is chosen to be the uniform distribution and θ chosen to be the unique solution in t to the equality in (71).

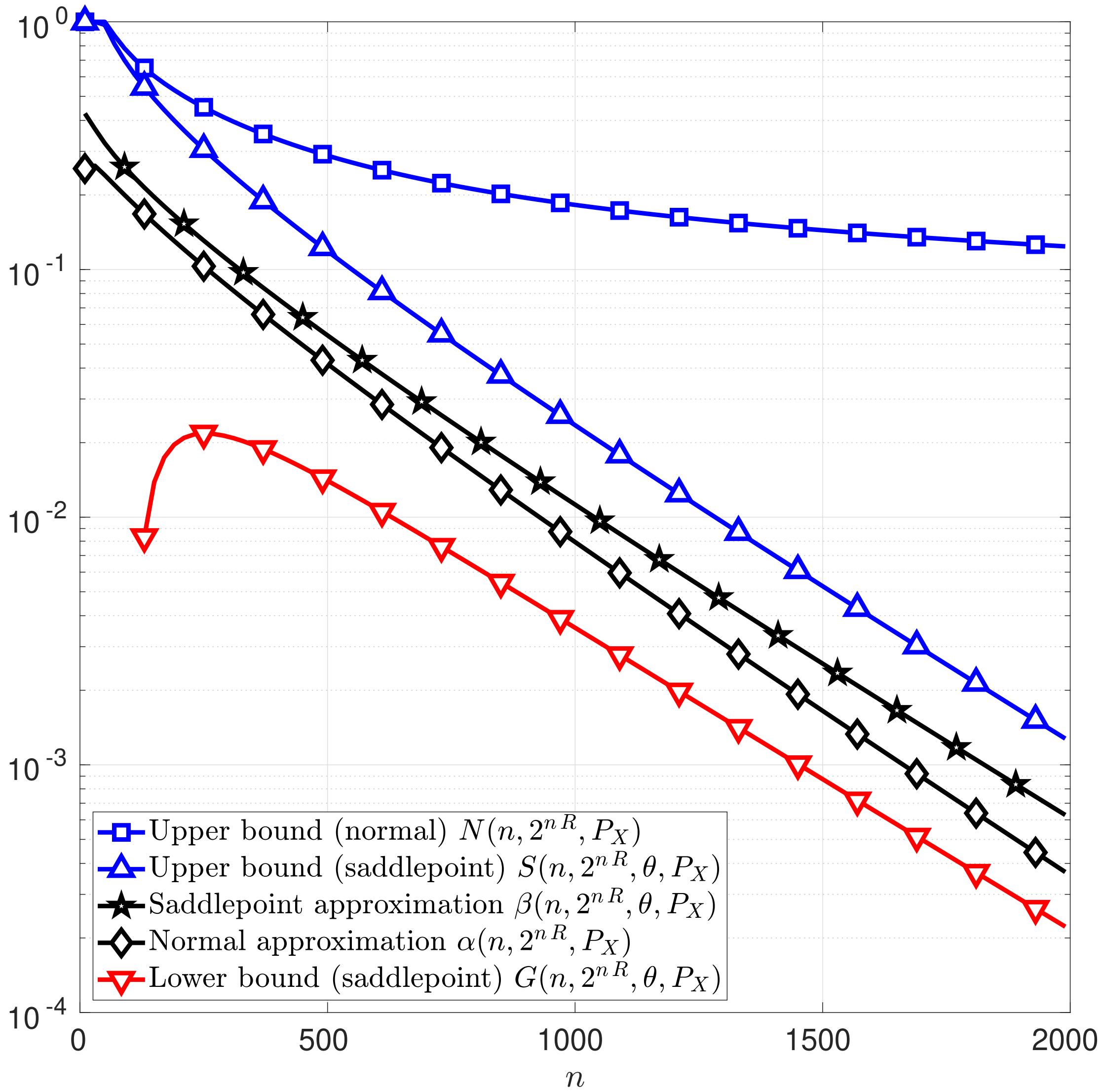

Figure 8: Normal and saddlepoint approximations of the function T in (49) as functions of the blocklength n for the case of a real-valued AWGN channel with discrete channel inputs, $\mathcal{X} = \{-1, 1\}$, and SNR $= 1$ at information rate $R = 0.425$ bits per channel use. The channel input distribution P_X is chosen to be the uniform distribution and θ chosen to be the unique solution in t to the equality in (71).

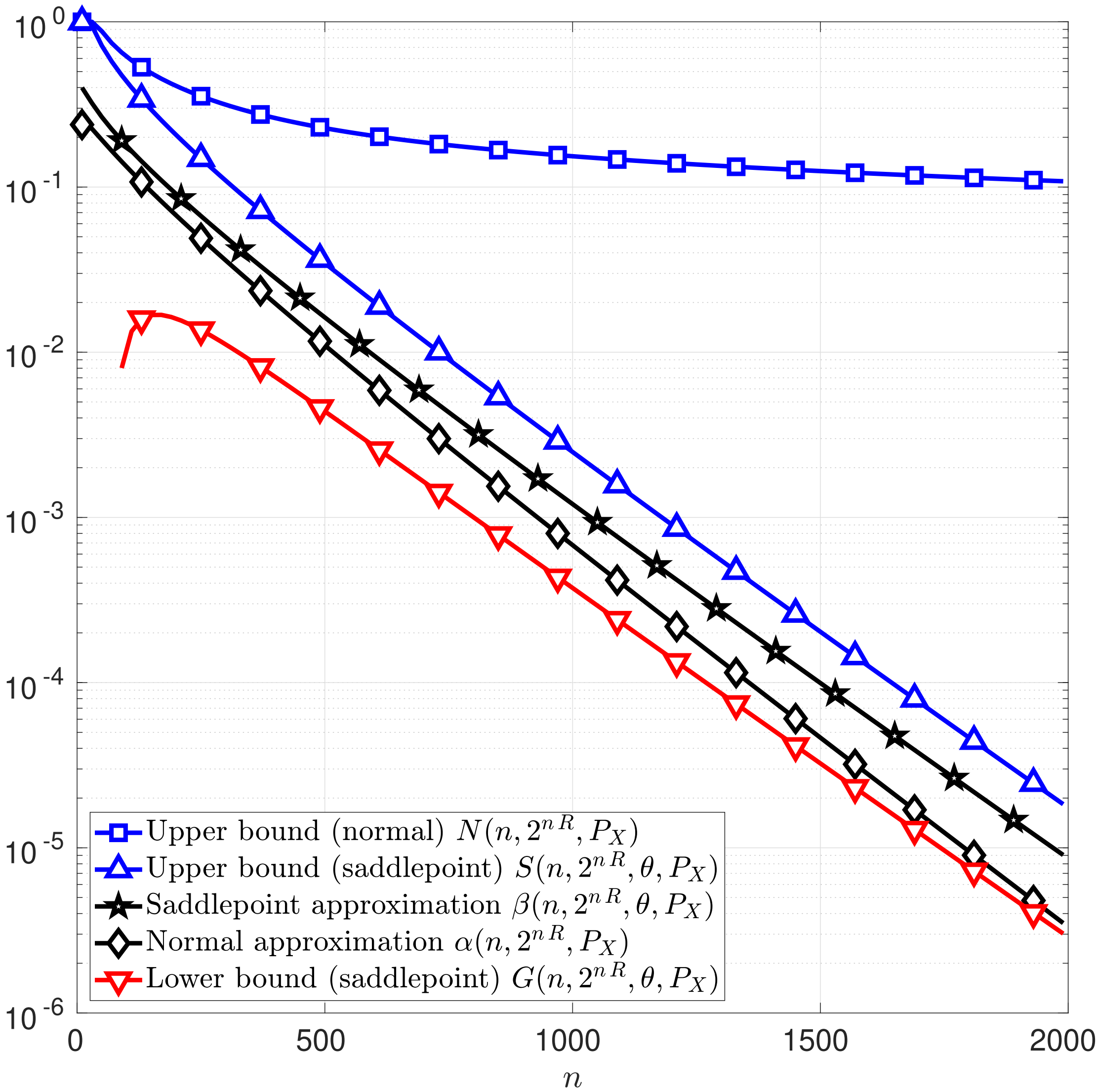

Figure 9: Normal and saddlepoint approximations of the function T in (49) as functions of the blocklength n for a real-valued SαS channel with discrete channel inputs, $\mathcal{X} = \{-1, 1\}$, $\alpha = 1.4$, and $\sigma = 0.6$ at information rate $R = 0.38$ bits per channel use. The channel input distribution P_X is chosen to be the uniform distribution and θ chosen to be the unique solution in t to the equality in (71).

3.3 Meta Converse Bound

This section describes a lower bound on $\lambda^*(n, M)$, for a fixed pair $(n, M) \in \mathbb{N}^2$. Given two probability distributions $P_{\boldsymbol{XY}} \in \triangle(\mathcal{X}^n \times \mathcal{Y}^n)$ and $Q_{\boldsymbol{Y}} \in \triangle(\mathcal{Y}^n)$, let the random variable $\tilde{\imath}(\boldsymbol{X}; \boldsymbol{Y}|Q_{\boldsymbol{Y}})$ satisfy

$$\tilde{\imath}(\boldsymbol{X}; \boldsymbol{Y}|Q_{\boldsymbol{Y}}) \triangleq \ln\left(\frac{\mathrm{d}P_{\boldsymbol{XY}}}{\mathrm{d}P_{\boldsymbol{X}}Q_{\boldsymbol{Y}}}(\boldsymbol{X}, \boldsymbol{Y})\right). \tag{75}$$

For all $(n,M,\gamma) \in \mathbb{N}^2 \times \mathbb{R}$ and for all probability distributions $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$ and $Q_{\boldsymbol{Y}} \in \triangle(\mathcal{Y}^n)$, let the function $C : \mathbb{N}^2 \times \triangle(\mathcal{X}^n) \times \triangle(\mathcal{Y}^n) \times \mathbb{R} \to \mathbb{R}_+$ be

$$C(n, M, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma) \triangleq \mathbb{E}_{P_{\boldsymbol{X}}P_{\boldsymbol{Y}|\boldsymbol{X}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}})\leqslant \ln(\gamma)\}}\right] + \gamma\left(\mathbb{E}_{P_{\boldsymbol{X}}Q_{\boldsymbol{Y}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}})>\ln(\gamma)\}}\right] - \frac{1}{M}\right). \tag{76}$$

Using this notation, the following lemma describes the MC bound.

Lemma 3 (MC Bound [12, 15]) *Given a pair* $(n,M) \in \mathbb{N}^2$*, the following holds for all* $Q_{\boldsymbol{Y}} \in \Delta(\mathcal{Y}^n)$*, with respect to the random transformation in* (41)*:*

$$\lambda^*(n, M) \geqslant \inf_{P_{\boldsymbol{X}} \in \Delta(\mathcal{X}^n)} \max_{\gamma \geqslant 0} C(n, M, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma), \tag{77}$$

where the function C *is defined in* (76)*.*

Note that the output probability distribution $Q_{\boldsymbol{Y}}$ in Lemma 3 can be chosen among all possible probability distributions $Q_{\boldsymbol{Y}} \in \triangle(\mathcal{Y}^n)$ to maximize the right-hand side of (76), which improves the bound. Note also that with some lost of optimality, the optimization domain can be constrained to the set of probability distributions for which for all $\boldsymbol{y} \in \mathcal{Y}^n$,

$$Q_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_{t=1}^{n} Q_Y(y_t), \tag{78}$$

with $Q_Y \in \triangle(\mathcal{Y})$. Hence, subject to (42), for all $\boldsymbol{x} \in \mathcal{X}^n$, the random variable $\tilde{\imath}(\boldsymbol{x}; \boldsymbol{Y}|Q_{\boldsymbol{Y}})$ in (76) can be written as the sum of the independent random variables, i.e.,

$$\tilde{\imath}(\boldsymbol{x}; \boldsymbol{Y}|Q_{\boldsymbol{Y}}) = \sum_{t=1}^{n} \tilde{\imath}(x_t; Y_t|Q_Y). \tag{79}$$

With some lost of generality, the focus is on a channel transformation of the form in 41 for which the following condition holds: The infimum in (77) is achieved by a product distribution, i.e., $P_{\boldsymbol{X}}$ is of the form in (51), when the probability distribution $Q_{\boldsymbol{Y}}$ satisfies (78). Note that this condition is met by memoryless channels such as the BSC, the AWGN and SαS channels with binary antipodal inputs, i.e. input alphabets are of the form $\mathcal{X} = \{a, -a\}$, with $a \in \mathbb{R}$. This follows from the fact that the random variable $\tilde{\imath}(\boldsymbol{x}; \boldsymbol{Y}|Q_{\boldsymbol{Y}})$ is invariant of the choice of $\boldsymbol{x} \in \mathcal{X}^n$ when the probability distribution $Q_{\boldsymbol{Y}}$ satisfies (78) and for all $y \in \mathcal{Y}$,

$$Q_Y(y) = \frac{P_{Y|X}(y|-a) + P_{Y|X}(y|a)}{2}. \tag{80}$$

Under these conditions, the random variable $\tilde{\imath}(\boldsymbol{X}; \boldsymbol{Y}|Q_{\boldsymbol{Y}})$ in (76) can be written as the sum of i.i.d. random variables, i.e.,

$$\tilde{\imath}(\boldsymbol{X}; \boldsymbol{Y}|Q_{\boldsymbol{Y}}) = \sum_{t=1}^{n} \tilde{\imath}(X_t; Y_t|Q_Y). \tag{81}$$

This observation motivates the application of the results of Section 2 to provide upper and lower bounds on the function C in (76), for some given values $(n, M) \in \mathbb{N}^2$ and given distributions $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$ and $Q_{\boldsymbol{Y}} \in \triangle(\mathcal{Y}^n)$. These bounds become significantly relevant when the exact value of $C(n, M, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma)$ cannot be calculated with respect to the random transformation in (41). In such a case, providing upper and lower bounds on $C(n, M, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma)$ helps in approximating its exact value subject to an error sufficiently small such that the approximation is relevant.

3.3.1 Normal Approximation

This section describes the normal approximation of the function C in (76), that is to say, the random variable $\tilde{\imath}(\boldsymbol{X}; \boldsymbol{Y}|Q_{\boldsymbol{Y}})$ is assumed to satisfy (81) and to follow a Gaussian distribution. That being said, for all $(P_X, Q_Y) \in \triangle(\mathcal{X}) \times \triangle(\mathcal{Y})$, let

$$\tilde{\mu}(P_X, Q_Y) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\tilde{\imath}(X; Y|Q_Y)\right], \tag{82}$$

$$\tilde{\sigma}(P_X, Q_Y) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left(\tilde{\imath}(X; Y|Q_Y) - \tilde{\mu}(P_X, Q_Y)\right)^2\right], \text{ and} \tag{83}$$

$$\tilde{\xi}(P_X, Q_Y) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left|\tilde{\imath}(X; Y|Q_Y) - \tilde{\mu}(P_X, Q_Y)\right|^3\right] \tag{84}$$

be the first, the second central, and the third absolute central moments, respectively, of the random variables $\tilde{\imath}(X_1; Y_1|Q_Y), \tilde{\imath}(X_2; Y_2|Q_Y), \ldots \tilde{\imath}(X_n; Y_n|Q_Y)$. Using this notation consider the functions $\tilde{D} : \mathbb{N}^2 \times \triangle(\mathcal{X}) \times \triangle(\mathcal{Y}) \times \mathbb{R}_+ \to \mathbb{R}_+$ and $\tilde{N} : \mathbb{N}^2 \times \triangle(\mathcal{X}) \times \triangle(\mathcal{Y}) \times \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $(n, M, \gamma) \in \mathbb{N}^2 \times \mathbb{R}$ and for all $P_X \in \triangle(\mathcal{X})$ and for all $Q_Y \in \triangle(\mathcal{Y})$,

$$\tilde{D}(n, M, P_X, Q_Y, \gamma) = \max\left(0, \tilde{\alpha}\left(n, M, P_X, Q_Y, \gamma\right) - \frac{c\,\tilde{\xi}(P_X, Q_Y)}{\tilde{\sigma}(P_X, Q_Y)^{\frac{3}{2}}\sqrt{n}}\right), \text{ and} \tag{85}$$

$$\tilde{N}(n, M, P_X, Q_Y, \gamma) = \min\left(1, \tilde{\alpha}\left(n, M, P_X, Q_Y, \gamma\right) + \frac{3\,c\,\tilde{\xi}(P_X, Q_Y)}{\tilde{\sigma}(P_X, Q_Y)^{\frac{3}{2}}\sqrt{n}} + \frac{2\ln(2)}{\tilde{\sigma}(P_X, Q_Y)^{\frac{1}{2}}\sqrt{2n\pi}}\right), \tag{86}$$

where $c = 0.476$ and

$$\tilde{\alpha}\left(n, M, P_X, Q_Y, \gamma\right) \triangleq Q\left(\frac{n\tilde{\mu}(P_X, Q_Y) - \ln(\gamma)}{\sqrt{n\tilde{\sigma}(P_X, Q_Y)}}\right) - \frac{\gamma}{M}. \tag{87}$$

Using this notation, the following theorem introduces a lower bound and an upper bound on C in (76).

Theorem 6 *Given a pair $(n, M) \in \mathbb{N}^2$, for all input distributions $P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)$ subject to* (51), *for all output distributions $Q_{\boldsymbol{Y}} \in \triangle(\mathcal{Y}^n)$ subject to* (78), *and for all $\gamma \geqslant 0$, the following holds with respect to the random transformation in* (41) *subject to* (42),

$$\tilde{D}(n, M, P_X, Q_Y, \gamma) \leqslant C(n, M, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma) \leqslant \tilde{N}(n, M, P_X, Q_Y, \gamma), \tag{88}$$

where the functions C, $\tilde{D}$, and $\tilde{N}$ are defined in (76), (85) *and* (86), *respectively.*

Proof: The proof of Theorem 6 is partially presented in [12]. Essentially, it consists in using Theorem 1 for upper and lower bounding the term $\mathbb{E}_{P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}}) \leqslant \ln(\gamma)\}}\right]$ in (76); and using Lemma 20 in [17] for upper bounding the term $\mathbb{E}_{P_{\boldsymbol{X}} Q_{\boldsymbol{Y}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}}) > \ln(\gamma)\}}\right]$ in (76). ■

The function $\tilde{\alpha}\left(n, M, P_X, Q_Y, \gamma\right)$ in (87) is often referred to as the *normal approximation* of $C(n, M, P_{\boldsymbol{X}})$, which is indeed a language abuse. In Section 2.1, a comment is given on the fact that the lower and upper bounds on the normal approximation, i.e., the functions $\tilde{D}$ in (85) and $\tilde{N}$ in (86), are often too far from the normal approximation $\tilde{\alpha}$ in (87).

3.3.2 Saddlepoint Approximation

This section describes an approximation of the function C in (76) by using the saddlepoint approximation of the CDF of the random variable $\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}})$, as suggested in Section 2.2. Given two distributions $P_X \in \triangle(\mathcal{X})$ and $Q_Y \in \triangle(\mathcal{Y})$, let the random variable $\tilde{\imath}(X;Y|Q_Y)$ satisfy

$$\tilde{\imath}(X;Y|Q_Y) \triangleq \ln\left(\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_X Q_Y}(X,Y)\right), \tag{89}$$

where $P_{Y|X}$ is in (42). The moment generating function of $\tilde{\imath}(X;Y|Q_Y)$ is

$$\tilde{\varphi}(P_X,Q_Y,\theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\exp\left(\theta\,\tilde{\imath}(X;Y|Q_Y)\right)\right], \tag{90}$$

with $\theta \in \mathbb{R}$. For all $P_X \in \triangle(\mathcal{X})$ and $Q_Y \in \triangle(\mathcal{Y})$, and for all $\theta \in \mathbb{R}$, consider the following functions:

$$\tilde{\mu}(P_X,Q_Y,\theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\frac{\tilde{\imath}(X;Y|Q_Y)\exp\left(\theta\,\tilde{\imath}(X;Y|Q_Y)\right)}{\tilde{\varphi}(P_X,Q_Y,\theta)}\right], \tag{91}$$

$$\tilde{V}(P_X,Q_Y,\theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left(\tilde{\imath}(X;Y|Q_Y)-\tilde{\mu}(P_X,Q_Y,\theta)\right)^2\frac{\exp\left(\theta\tilde{\imath}(X;Y|Q_Y)\right)}{\tilde{\varphi}(P_X,Q_Y,\theta)}\right], \text{ and} \tag{92}$$

$$\tilde{\xi}(P_X,Q_Y,\theta) \triangleq \mathbb{E}_{P_X P_{Y|X}}\left[\left|\tilde{\imath}(X;Y|Q_Y)-\tilde{\mu}(P_X,Q_Y,\theta)\right|^3\frac{\exp\left(\theta\tilde{\imath}(X;Y|Q_Y)\right)}{\tilde{\varphi}(P_X,Q_Y,\theta)}\right]. \tag{93}$$

Using this notation consider the functions $\tilde{\beta}_1 : \mathbb{N}\times\mathbb{R}^2\times\triangle(\mathcal{X})\times\triangle(\mathcal{Y}) \to \mathbb{R}_+$ and $\tilde{\beta}_2 : \mathbb{N}\times\mathbb{R}^2\times\triangle(\mathcal{X})\times\triangle(\mathcal{Y}) \to \mathbb{R}_+$:

$$\begin{aligned}&\tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y)\\&=\mathbb{1}_{\{\theta>0\}}+(-1)^{\mathbb{1}_{\{\theta>0\}}}\exp\!\left(n\ln\left(\tilde{\varphi}(P_X,Q_Y,\theta)\right)-\theta\ln(\gamma)+\frac{1}{2}\theta^2 n\tilde{V}(P_X,Q_Y,\theta)\right)Q\!\left(\sqrt{n\tilde{V}(P_X,Q_Y,\theta)}|\theta|\right),\end{aligned} \tag{94}$$

and

$$\begin{aligned}&\tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y)\\&=\mathbb{1}_{\{\theta\leqslant-1\}}+(-1)^{\mathbb{1}_{\{\theta\leqslant-1\}}}\exp\left(n\ln\left(\tilde{\varphi}(P_X,Q_Y,\theta)\right)-(\theta+1)\ln(\gamma)+\frac{1}{2}(\theta+1)^2 n\tilde{V}(P_X,Q_Y,\theta)\right)\\&\quad Q\!\left(\sqrt{n\tilde{V}(P_X,Q_Y,\theta)}|\theta+1|\right).\end{aligned} \tag{95}$$

Note that $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are the saddlepoint approximation of the CDF and the complementary CDF of the random variable $\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}})$ in (81) when $(\boldsymbol{X},\boldsymbol{Y})$ follows the distribution $P_{\boldsymbol{X}}P_{\boldsymbol{Y}|\boldsymbol{X}}$ and $P_{\boldsymbol{X}}Q_{\boldsymbol{Y}}$, respectively. Consider also the following functions:

$$\begin{aligned}&\tilde{G}_1(n,\gamma,\theta,P_X,Q_Y)\\&=\tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y)-\frac{2c\,\tilde{\xi}(P_X,Q_Y,\theta)}{\tilde{V}(P_X,Q_Y,\theta)^{3/2}\sqrt{n}}\exp\left(n\ln\left(\tilde{\varphi}(P_X,Q_Y,\theta)\right)-\theta\ln(\gamma)\right),\end{aligned} \tag{96}$$

$$\begin{aligned}&\tilde{G}_2(n,\gamma,\theta,P_X,Q_Y)\\&=\tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y)-\frac{2c\,\tilde{\xi}(P_X,Q_Y,\theta)}{\tilde{V}(P_X,Q_Y,\theta)^{3/2}\sqrt{n}}\exp\left(n\ln\left(\tilde{\varphi}(P_X,Q_Y,\theta)\right)-(\theta+1)\ln(\gamma)\right),\end{aligned} \tag{97}$$

$$\begin{aligned}&\tilde{G}(n,\gamma,\theta,P_X,Q_Y,M)\\&=\max\Big(0,\tilde{G}_1(n,\gamma,\theta,P_X,Q_Y)\Big)+\gamma\max\Big(0,\tilde{G}_2(n,\gamma,\theta,P_X,Q_Y)\Big)-\frac{\gamma}{M},\end{aligned}\tag{98}$$

$$\begin{aligned}&\tilde{S}(n,\gamma,\theta,P_X,Q_Y,M)\\&=\min\left(1,\tilde{\beta}\left(n,\gamma,\theta,P_X,Q_Y,M\right)+\frac{4c\,\tilde{\xi}(P_X,Q_Y,\theta)}{\Big(\tilde{V}(P_X,Q_Y,\theta)\Big)^{3/2}\sqrt{n}}\exp\left(n\ln\left(\tilde{\varphi}(P_X,Q_Y,\theta)\right)-\theta\ln\left(\gamma\right)\right)\right),\end{aligned}\tag{99}$$

and

$$\tilde{\beta}(n,\gamma,\theta,P_X,Q_Y,M)=\tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y)+\gamma\tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y)-\frac{\gamma}{M}.\tag{100}$$

The function $\tilde{\beta}(n,\gamma,\theta,P_X,Q_Y,M)$ in (100) is referred to as the *saddlepoint approximation* of the function C in (76), which is indeed a language abuse. The following theorem introduces a new lower bound and a new upper bound on C.

Theorem 7 *Given a pair $(n,M)\in\mathbb{N}^2$, for all input distributions $P_{\boldsymbol{X}}\in\triangle\left(\mathcal{X}^n\right)$ subject to* (51)*, for all output distributions $Q_{\boldsymbol{Y}}\in\triangle\left(\mathcal{Y}^n\right)$ subject to* (81) *such that for all $x\in\mathcal{X}$, $P_{Y|X=x}$ is absolutely continuous with respect to Q_Y, for all $\gamma\geqslant 0$, the following holds with respect to the random transformation in* (41) *subject to* (42)*,*

$$\tilde{G}(n,\gamma,\theta,P_X,Q_Y,M)\leqslant C(n,M,P_{\boldsymbol{X}},Q_{\boldsymbol{Y}},\gamma)\leqslant\tilde{S}(n,\gamma,\theta,P_X,Q_Y,M),\tag{101}$$

where θ is the unique solution in t to

$$n\mu(P_X,t)=\ln\left(\gamma\right),\tag{102}$$

and the functions C, $\tilde{G}$, and $\tilde{S}$ are defined in (76)*,* (98) *and* (99)*, with $c=0.476$.*

Proof: The proof of Theorem 7 is provided in Appendix G. ■

Note that in (101), the parameter γ can be optimized as in (77).

3.3.3 Numerical Analysis

The normal approximation and the saddlepoint approximation of the MC bound as well as the corresponding upper bounds and lower bounds presented in Section 3.3.1 and in Section 3.3.2 are studied in the cases of the BSC, the AWGN channel and the SαS channel. In the following figures, Figure 10 - Figure 12, the function C in (76), which is bounded by using Theorem 6 and Theorem 7, is studied. On the first hand, its normal approximation $\tilde{\alpha}\left(n,2^{nR},P_X,Q_Y,\gamma\right)$ in (87) is plotted in black diamonds, whereas the corresponding lower and upper bounds, i.e., $\tilde{D}\left(n,2^{nR},P_X,Q_Y,\gamma\right)$ in (85) and $\tilde{N}\left(n,2^{nR},P_X,Q_Y,\gamma\right)$ in (86), are respectively plotted in red circles and blue squares. On the second hand, its saddlepoint approximation $\tilde{\beta}\Big(n,\gamma,\theta,P_X,Q_Y,2^{nR}\Big)$ in (100), is plotted in black stars whereas the corresponding upper and lower bounds, i.e., $\tilde{S}\left(n,\gamma,\theta,P_X,Q_Y,2^{nR}\right)$ in (99) and $\tilde{G}\left(n,\gamma,\theta,P_X,Q_Y,2^{nR}\right)$ in (98), are plotted in blue upward-pointing triangles and red downward-pointing triangles respectively. These functions are plotted only when their values are positive. The channel inputs are discrete $\mathcal{X}=\{-1,1\}$, P_X is the uniform distribution, Q_Y is equal to the distribution P_Y, i.e. the marginal of $P_XP_{Y|X}$, γ is chosen to maximize the function C in (76), and θ is chosen to be the unique solution in t to the equality in (102).

Figure 10 concerns the case of a BSC with cross-over probability $\delta = 0.11$ and $R = 0.42$ bits per channel use. The function C in (76) can be calculated exactly and thus, it is plotted in magenta asterisks. Therein, it can be observed that both the saddlepoint approximation $\tilde{\beta}$ and the function C overlap. These observations are in line with those reported in [15], in which the saddlepoint approximations of the RCU bound and the MC bound are both shown to be precise approximations.
Figure 11 and Figure 12 concern the cases of a real-valued AWGN channel and a SαS channel, respectively. Moreover, the signal to noise ratio (SNR) is SNR $= 1$ for both channels. The information rate is $R = 0.425$ bits per channel use for the AWGN channel and $R = 0.38$ bits per channel use for the SαS channel, with $(\alpha, \sigma) = (1.4, 0.6)$. In both cases, the function C in (76) can not be computed explicitly and hence does not appear in Figure 11 and Figure 12. In addition, the lower bound $\tilde{D}\,(n, M, P_X, Q_Y, \gamma)$ obtained from Theorem 6 is non-positive in these cases, and thus, does not appear on the figures.
Note that in Figure 10 - Figure 12, the upper bound $\tilde{N}\left(n, 2^{nR}, P_X, Q_Y, \gamma\right)$ is several orders of magnitude far away from the normal approximation $\tilde{\alpha}\left(n, 2^{nR}, P_X, Q_Y, \gamma\right)$. From this perspective, a proper analysis on the MC bound (Lemma 3) based on Theorem 6 does not lead to relevant conclusions. These observations are in line with those reported in Figure 7 - Figure 9.

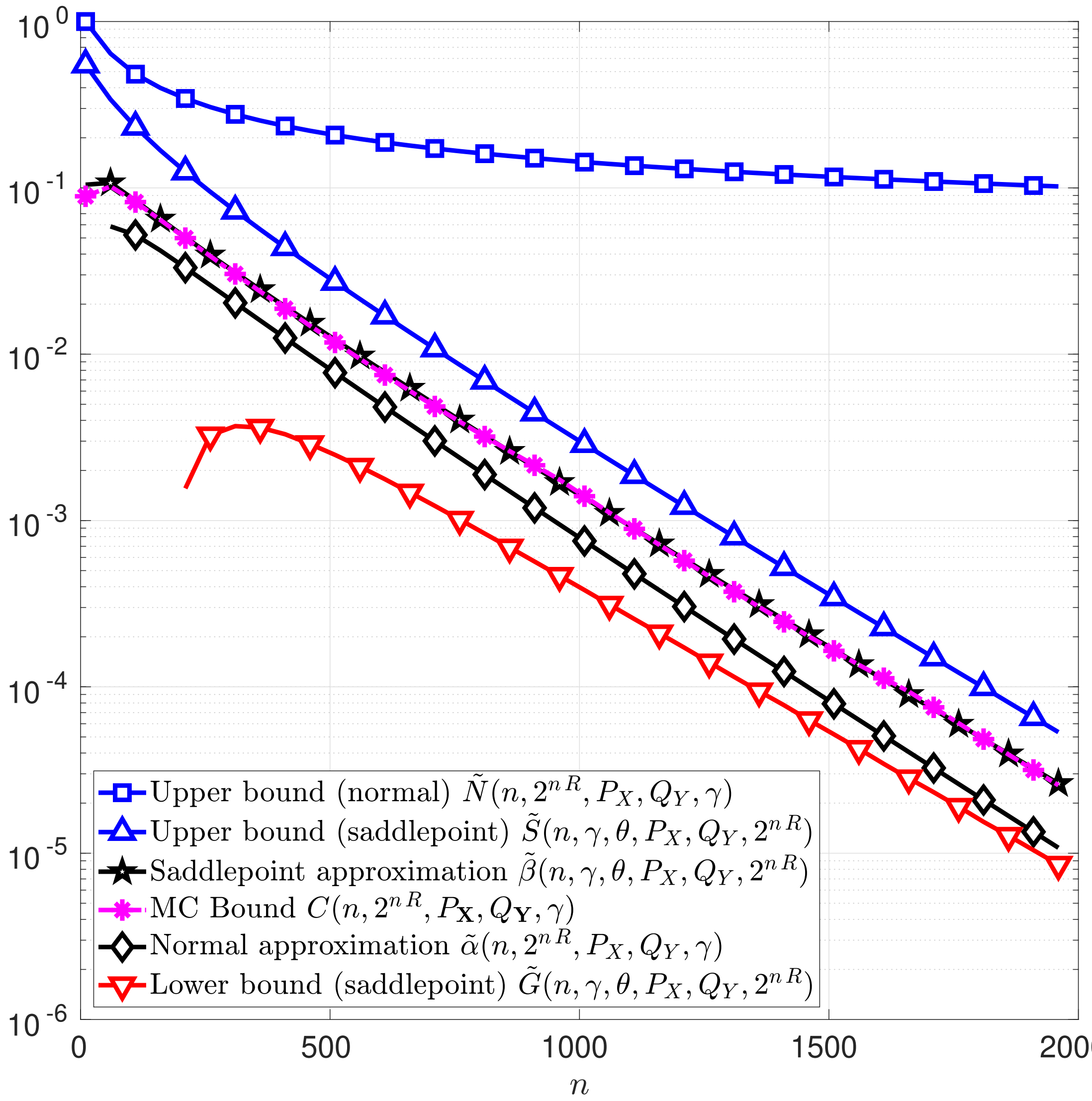

Figure 10: Normal and saddlepoint approximations to the function C in (76) as functions of the blocklength n for the case of a BSC with cross-over probability $\delta = 0.11$ at information rate $R = 0.42$ bits per channel use. The channel input distribution P_X is chosen to be the uniform distribution, the output distribution Q_Y chosen to be the channel output distribution P_Y, γ chosen to maximize C in (76), and θ chosen to be the unique solution in t of the equality in (102).

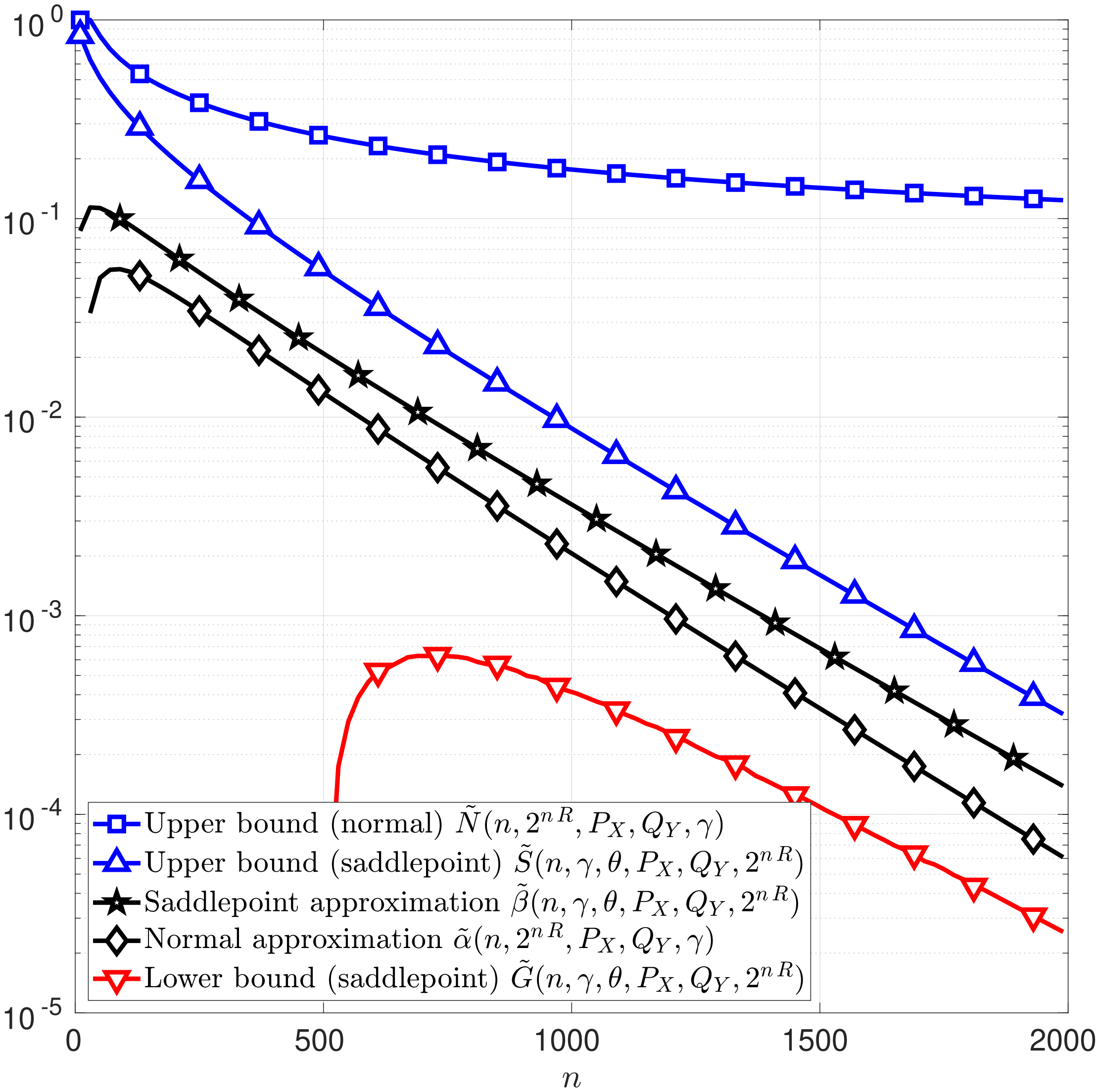

Figure 11: Normal and saddlepoint approximations to the function C in (76) as functions of the blocklength n for the case of a real-valued AWGN channel with discrete channel inputs, $\mathcal{X} = \{-1, 1\}$, and signal to noise ratio $\text{SNR} = 1$ at information rate $R = 0.425$ bits per channel use. The channel input distribution P_X is chosen to be the uniform distribution, the output distribution Q_Y chosen to be the channel output distribution P_Y, γ chosen to maximize C in (76), and θ chosen to be the unique solution in t of the equality in (102).

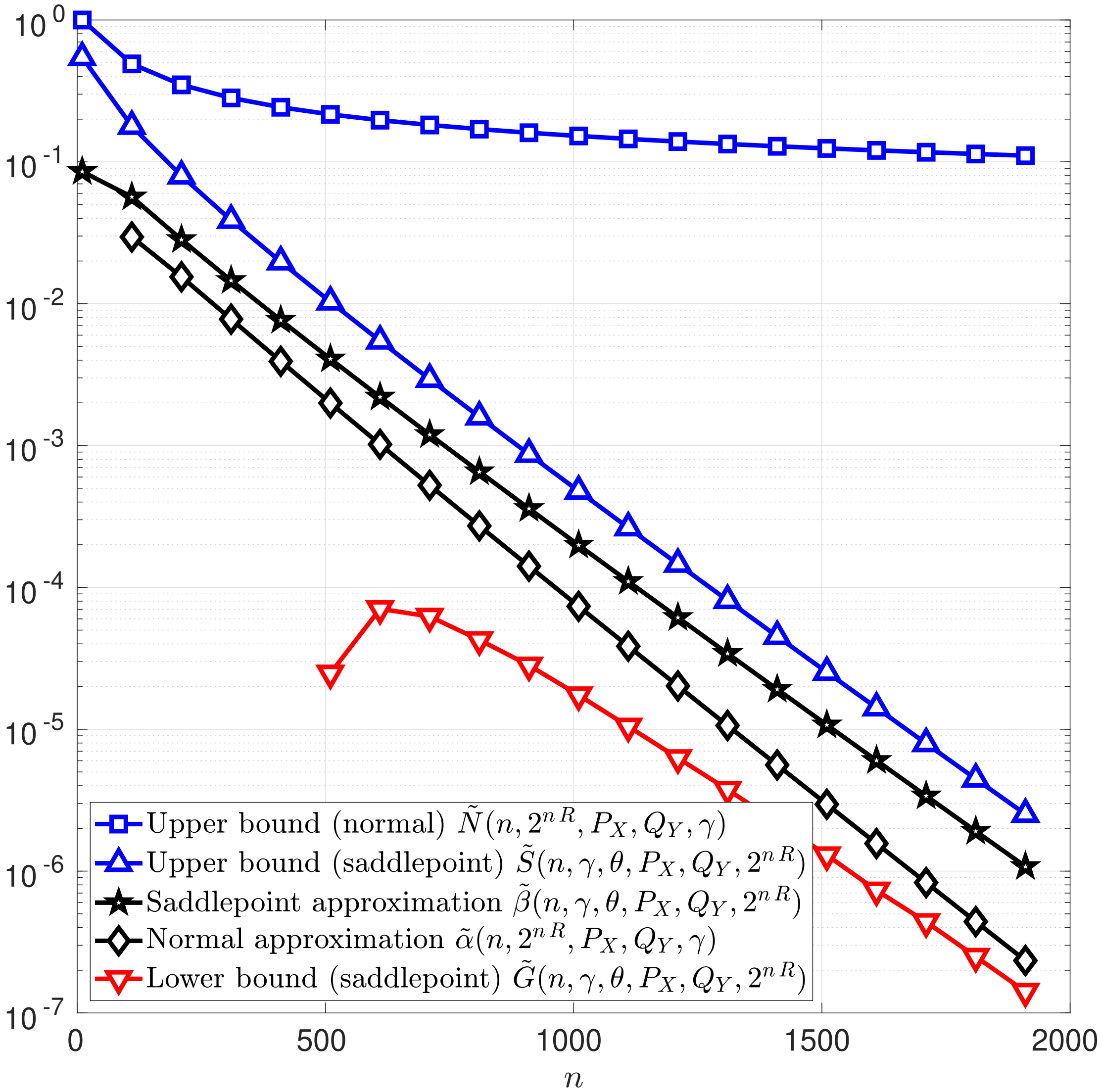

Figure 12: Normal and saddlepoint approximation to the function C in (76) as functions of the blocklength n for the case of a real-valued symmetric α-stable noise channel with discrete channel inputs, $\mathcal{X} = \{-1, 1\}$, a shape parameter $\alpha = 1.4$ and a dispersion parameter $\sigma = 0.6$ at information rate $R = 0.38$ bits per channel use. The channel input distribution P_X is chosen to be the uniform distribution, the output distribution Q_Y chosen to be the channel output distribution P_Y, γ chosen to maximize C in (76), and θ chosen to be the unique solution in t of the equality in (102).

4 Discussions and Further Work

One of the main results of this work is Theorem 3, which gives an upper bound on the error induced by the saddlepoint approximation of the CDF of a sum of i.i.d. random variables. This result paves the way to study channel coding problems at any finite blocklength and any constraint on the DEP. In particular, Theorem 3 is used to bound the DT and MC bounds in point-to-point memoryless channels. This leads to tighter bounds than those obtained from Berry-Esseen Theorem (Theorem 1), c.f., examples in Section 3.2.3 and Section 3.3.3, particularly for the small values of the DEP.

The bound on the approximation error presented in Theorem 2 uses a triangle inequality in the proof of Lemma 4, which is loose. This is essentially the reason why Theorem 2 is not reduced to the Berry-Esseen Theorem when the parameter θ is equal to zero. An interesting extension of this work is to tighten the inequality in Lemma 4 such that the Berry-Esseen Theorem can be obtained as a special case of Theorem 2, i.e., when $\theta = 0$. If such improvement on Theorem 2 is possible, Theorem 3 will be strongly improved and it would be more precise everywhere and in particular in the vicinity of the mean of the sum in (1).

Appendices

A Proof of Theorem 2

The proof of Theorem 2 relies on the notion of exponentially tilted distributions. Let φ_Y be the moment generating function of the distribution P_Y. Given $\theta \in \Theta_Y$, let $Y_1^{(\theta)}$, $Y_2^{(\theta)}$, ..., $Y_n^{(\theta)}$ be random variables whose joint probability distribution, denoted by $P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}$, satisfies for all $(y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$,

$$\frac{\mathrm{d}P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}{\mathrm{d}P_{Y_1 Y_2 \ldots Y_n}}(y_1, y_2, \ldots, y_n) = \frac{\exp\left(\theta \sum_{j=1}^n y_j\right)}{\left(\varphi_Y(\theta)\right)^n}. \tag{103}$$

That is, the distribution $P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}$ is an exponentially tilted distribution with respect to $P_{Y_1 Y_2 \ldots Y_n}$. Using this notation, for all $\mathcal{A} \subseteq \mathbb{R}$ and for all $\theta \in \Theta_Y$,

$$\begin{aligned}
P_{X_n}(\mathcal{A}) &= \mathbb{E}_{P_{X_n}}\left[\mathbb{1}_{\{X_n \in \mathcal{A}\}}\right] && \text{(104a)}\\
&= \mathbb{E}_{P_{Y_1 Y_2 \ldots Y_n}}\left[\mathbb{1}_{\left\{\sum_{j=1}^n Y_j \in \mathcal{A}\right\}}\right] && \text{(104b)}\\
&= \mathbb{E}_{P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}\left[\frac{\mathrm{d}P_{Y_1 Y_2 \ldots Y_n}}{\mathrm{d}P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}(Y_1^{(\theta)}, Y_2^{(\theta)}, \ldots, Y_n^{(\theta)}) \mathbb{1}_{\left\{\sum_{j=1}^n Y_j^{(\theta)} \in \mathcal{A}\right\}}\right] && \text{(104c)}\\
&= \mathbb{E}_{P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}\left[\left(\frac{\mathrm{d}P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}{\mathrm{d}P_{Y_1 Y_2 \ldots Y_n}}(Y_1^{(\theta)}, Y_2^{(\theta)}, \ldots, Y_n^{(\theta)})\right)^{-1} \mathbb{1}_{\left\{\sum_{j=1}^n Y_j^{(\theta)} \in \mathcal{A}\right\}}\right] && \text{(104d)}\\
&= \mathbb{E}_{P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}\left[\left(\frac{\exp\left(\theta \sum_{j=1}^n Y_j^{(\theta)}\right)}{\left(\varphi_Y(\theta)\right)^n}\right)^{-1} \mathbb{1}_{\left\{\sum_{j=1}^n Y_j^{(\theta)} \in \mathcal{A}\right\}}\right] && \text{(104e)}\\
&= \left(\varphi_Y(\theta)\right)^n \mathbb{E}_{P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}\left[\exp\left(-\theta \sum_{j=1}^n Y_j^{(\theta)}\right) \mathbb{1}_{\left\{\sum_{j=1}^n Y_j^{(\theta)} \in \mathcal{A}\right\}}\right] && \text{(104f)}
\end{aligned}$$

For the ease of the notation, consider the random variable

$$S_{n,\theta} = \sum_{j=1}^n Y_j^{(\theta)}, \tag{105}$$

whose probability distribution is denoted by $P_{S_{n,\theta}}$. Hence, plugging (105) in (104f) yields,

$$P_{X_n}(\mathcal{A}) = \left(\varphi_Y(\theta)\right)^n \mathbb{E}_{P_{S_{n,\theta}}}\left[\exp\left(-\theta S_{n,\theta}\right) \mathbb{1}_{\{S_{n,\theta} \in \mathcal{A}\}}\right]. \tag{106}$$

The proof continues by upper bounding the following absolute difference

$$\left| P_{X_n}(\mathcal{A}) - \left(\varphi_Y(\theta)\right)^n \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp\left(-\theta Z_{n,\theta}\right) \mathbb{1}_{\{Z_{n,\theta} \in \mathcal{A}\}}\right] \right|, \tag{107}$$

where $Z_{n,\theta}$ is a Gaussian random variable with the same mean and variance as $S_{n,\theta}$, and probability distribution denoted by $P_{Z_{n,\theta}}$. The relevance of the absolute difference in (107) is that it is equal to the error of calculating $P_{X_n}(\mathcal{A})$ under the assumption that the resulting random variable S_n follows a Gaussian distribution. The following lemma provides an upper bound on the absolute difference in (107) in terms of the Kolmogorov-Smirnov distance between the distributions $P_{S_{n,\theta}}$ and $P_{Z_{n,\theta}}$, denoted by

$$\Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right) \triangleq \sup_{x\in\mathbb{R}} \left|F_{S_{n,\theta}}(x) - F_{Z_{n,\theta}}(x)\right|, \tag{108}$$

where $F_{S_{n,\theta}}$ and $F_{Z_{n,\theta}}$ are the CDFs of the random variables $S_{n,\theta}$ and $Z_{n,\theta}$, respectively.

Lemma 4 *Given $\theta \in \Theta_Y$ and $a \in \mathbb{R}$ consider the following conditions:*
(i) $\theta \leqslant 0$ and $\mathcal{A} = (-\infty, a]$, and
(ii) $\theta > 0$ and $\mathcal{A} = (a, \infty)$.
If at least one of the above conditions is satisfied, then the absolute difference in (107) *satisfies,*

$$\left|P_{X_n}(\mathcal{A}) - (\varphi_Y(\theta))^n \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp(-\theta Z_{n,\theta}) \mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]\right| \leqslant \frac{(\varphi_Y(\theta))^n}{\exp(\theta a)} \min\left(1, 2\,\Delta(P_{S_{n,\theta}}, P_{Z_{n,\theta}})\right). \tag{109}$$

Proof: The proof of Lemma 4 is presented in Appendix D. ∎

The proof continues by providing an upper bound on $\Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right)$ in (109) leveraging the observation that $S_{n,\theta}$ is the sum of n independent and identically distributed random variables. This follows immediately from the assumptions of Theorem 2, nonetheless, for the sake of completeness, the following lemma provides a proof of this statement.

Lemma 5 *For all $\theta \in \Theta_Y$, $Y_1^{(\theta)}$, $Y_2^{(\theta)}$, ..., $Y_n^{(\theta)}$ are mutually independent and identically distributed random variables with probability distribution $P_{Y^{(\theta)}}$. Moreover, $P_{Y^{(\theta)}}$ is an exponential tilted distribution with respect to P_Y. That is, $P_{Y^{(\theta)}}$ satisfies for all $y \in \mathbb{R}$,*

$$\frac{\mathrm{d}P_{Y^{(\theta)}}}{\mathrm{d}P_Y}(y) = \frac{\exp(\theta y)}{\varphi_Y(\theta)}. \tag{110}$$

Proof: The proof of Lemma 5 is presented in Appendix E. ∎

Lemma 5 paves the way for obtaining an upper bound on $\Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right)$ in (109) via the Berry-Essen Theorem (Theorem 1). Let μ_θ, V_θ and ξ_θ be the mean, the variance and the third absolute central moment of the random variable $Y^{(\theta)}$, whose probability distribution is $P_{Y^{(\theta)}}$ in (110). More specifically:

$$\mu_\theta = \mathbb{E}_{P_{Y^{(\theta)}}}[Y^{(\theta)}] = \mathbb{E}_{P_Y}\left[\frac{Y\exp(\theta Y)}{\varphi_Y(\theta)}\right], \tag{111}$$

$$V_\theta = \mathbb{E}_{P_{Y^{(\theta)}}}[(Y^{(\theta)} - \mu_\theta)^2] = \mathbb{E}_{P_Y}\left[\frac{(Y-\mu_\theta)^2\exp(\theta Y)}{\varphi_Y(\theta)}\right], \text{and} \tag{112}$$

$$\xi_\theta = \mathbb{E}_{P_{Y^{(\theta)}}}[|Y^{(\theta)} - \mu_\theta|^3] = \mathbb{E}_{P_Y}\left[\frac{|Y-\mu_\theta|^3\exp(\theta Y)}{\varphi_Y(\theta)}\right]. \tag{113}$$

From Theorem 1, it follows that $\Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right)$ in (109) satisfies:

$$\Delta(P_{S_{n,\theta}}, P_{Z_{n,\theta}}) \leqslant \min\left(1, \frac{c\,\xi_\theta}{\sqrt{n(V_\theta)^3}}\right) \leqslant \frac{c\,\xi_\theta}{\sqrt{n(V_\theta)^3}}, \tag{114}$$

where $c = 0.476$. Plugging (114) in (109) yields,

$$\left|P_{X_n}(\mathcal{A}) - \frac{(\varphi_Y(\theta))^n}{\exp(\theta b)} \mathbb{E}_{P_{Z_{n,\theta}}}[\exp(-\theta Z_{n,\theta}) \mathbb{1}\left\{Z_{n,\theta} \in \mathcal{A}\right\}]\right| \leqslant \frac{(\varphi_Y(\theta))^n}{\exp(\theta a)} \min\left(1, 2\frac{c\,\xi_\theta}{\sqrt{n(V_\theta)^3}}\right), \tag{115}$$

under the assumption that at least one of the conditions of Lemma 4 is met.
The proof ends by obtaining a closed-form expression of the term $\mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp\left(-\theta Z_{n,\theta}\right)\mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]$ in (115) under the assumption that at least one of the conditions of Lemma 4 is met. First, assuming that condition (i) in Lemma 4 holds, it follows that:

$$
\begin{aligned}
&\mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp\left(-\theta Z_{n,\theta}\right)\mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]\\
&=\int_{-\infty}^{a}\exp\left(-\theta z\right)\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{(z-n\mu_\theta)^2}{2nV_\theta}\right)\mathrm{d}z &(116a)\\
&=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{z^2-2\,z\,n\mu_\theta+n^2\mu_\theta^2+2n\theta\,V_\theta\,z}{2nV_\theta}\right)\mathrm{d}z &(116b)\\
&=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{(z-n\mu_\theta+n\theta V_\theta)^2-n^2\theta^2V_\theta^2+2n\mu_\theta\,n\theta V_\theta}{2nV_\theta}\right)\mathrm{d}z &(116c)\\
&=\exp\left(-\theta n\mu_\theta+\frac{1}{2}nV_\theta\theta^2\right)\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{(z-n\mu_\theta+n\theta V_\theta)^2}{2nV_\theta}\right)\mathrm{d}z &(116d)\\
&=\exp\left(-\theta n\mu_\theta+\frac{1}{2}nV_\theta\theta^2\right)\int_{-\infty}^{\frac{a-n\mu_\theta+n\theta V_\theta}{\sqrt{nV_\theta}}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)\mathrm{d}t &(116e)\\
&=\exp\left(-\theta n\mu_\theta+\frac{1}{2}nV_\theta\theta^2\right)Q\left(-\frac{a-n\mu_\theta+n\theta V_\theta}{\sqrt{nV_\theta}}\right). &(116f)
\end{aligned}
$$

Second, assuming that condition (ii) in Lemma 4 holds, it follows that:

$$
\begin{aligned}
&\mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp\left(-\theta Z_{n,\theta}\right)\mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]\\
&=\int_{a}^{\infty}\exp\left(-\theta z\right)\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{(z-n\mu_\theta)^2}{2nV_\theta}\right)\mathrm{d}z &(117a)\\
&=\int_{a}^{\infty}\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{z^2-2\,z\,n\mu_\theta+n^2\mu_\theta^2+2n\theta\,V_\theta\,z}{2nV_\theta}\right)\mathrm{d}z &(117b)\\
&=\int_{a}^{\infty}\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{(z-n\mu_\theta+n\theta V_\theta)^2-n^2\theta^2V_\theta^2+2n\mu_\theta\,n\theta V_\theta}{2nV_\theta}\right)\mathrm{d}z &(117c)\\
&=\exp\left(-\theta n\mu_\theta+\frac{1}{2}nV_\theta\theta^2\right)\int_{a}^{\infty}\frac{1}{\sqrt{2\pi n V_\theta}}\exp\left(-\frac{(z-n\mu_\theta+n\theta V_\theta)^2}{2nV_\theta}\right)\mathrm{d}z &(117d)\\
&=\exp\left(-\theta n\mu_\theta+\frac{1}{2}nV_\theta\theta^2\right)\int_{\frac{a-n\mu_\theta+n\theta V_\theta}{\sqrt{nV_\theta}}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)\mathrm{d}t &(117e)\\
&=\exp\left(-\theta n\mu_\theta+\frac{1}{2}nV_\theta\theta^2\right)Q\left(\frac{a-n\mu_\theta+n\theta V_\theta}{\sqrt{nV_\theta}}\right), &(117f)
\end{aligned}
$$

where Q in (116f) and (117f) is the complementary cumulative distribution function of the standard Gaussian distribution defined in (13).
The expressions in (116f) and (117f) can be jointly written as follows:

$$
\mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp\left(-\theta Z_{n,\theta}\right)\mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]=\exp\left(-\theta n\mu_\theta+\frac{1}{2}nV_\theta\theta^2\right)Q\left((-1)^{\mathbb{1}_{\{\theta\leqslant 0\}}}\frac{a-n\mu_\theta+n\theta V_\theta}{\sqrt{nV_\theta}}\right), \quad (118)
$$

under the assumption that at least one of the conditions (i) or (ii) in Lemma 4 holds.
Finally, under the same assumption, plugging (118) in (115) yields

$$
\left|P_{X_n}(\mathcal{A})-\exp\left(n\ln\left(\varphi_Y(\theta)\right)-n\theta\mu_\theta+\frac{1}{2}n\theta^2V_\theta\right)Q\left((-1)^{\mathbb{1}_{\{\theta\leqslant 0\}}}\frac{a+n\theta V_\theta-n\mu_\theta}{\sqrt{nV_\theta}}\right)\right|
$$

$$\leqslant \exp\left(n\ln\left(\varphi_Y(\theta)\right) - \theta a\right) \min\left(1, \frac{2\,c\,\xi_\theta}{V_\theta^{3/2}\sqrt{n}}\right). \tag{119}$$

Under condition (i) in Lemma 4, the inequality in (119) can be written as follows:

$$\left| F_{X_n}(a) - \exp\left(n\ln\left(\varphi_Y(\theta)\right) - n\theta\mu_\theta + \frac{1}{2}n\theta^2 V_\theta\right) \cdot Q\left((-1)^{\mathbb{1}_{\{\theta\leqslant 0\}}}\frac{a + n\theta V_\theta - n\mu_\theta}{\sqrt{nV_\theta}}\right)\right|$$
$$\leqslant \exp\left(n\ln\left(\varphi_Y(\theta)\right) - \theta a\right) \min\left(1, \frac{2\,c\,\xi_\theta}{V_\theta^{3/2}\sqrt{n}}\right). \tag{120}$$

Alternatively, under condition (ii) in Lemma 4, it follows from (119) that

$$\left| 1 - F_{X_n}(a) - \exp\left(n\ln\left(\varphi_Y(\theta)\right) - n\theta\mu_\theta + \frac{1}{2}n\theta^2 V_\theta\right) \cdot Q\left((-1)^{\mathbb{1}_{\{\theta\leqslant 0\}}}\frac{a + n\theta V_\theta - n\mu_\theta}{\sqrt{nV_\theta}}\right)\right|$$
$$\leqslant \exp\left(n\ln\left(\varphi_Y(\theta)\right) - \theta a\right) \min\left(1, \frac{2\,c\,\xi_\theta}{V_\theta^{3/2}\sqrt{n}}\right). \tag{121}$$

Then, jointly writing (120) and (121), it follows that for all $a \in \mathbb{R}$ and for all $\theta \in \Theta_Y$,

$$\left| F_{X_n}(a) - \mathbb{1}_{\{\theta>0\}} - (-1)^{\mathbb{1}_{\{\theta>0\}}}\exp\left(n\ln\left(\varphi_Y(\theta)\right) - n\theta\mu_\theta + \frac{1}{2}n\theta^2 V_\theta\right) Q\left((-1)^{\mathbb{1}_{\{\theta\leqslant 0\}}}\frac{a + n\theta V_\theta - n\mu_\theta}{\sqrt{nV_\theta}}\right)\right|$$
$$\leqslant \exp\left(n\ln\left(\varphi_Y(\theta)\right) - \theta a\right) \min\left(1, \frac{2\,c\,\xi_\theta}{V_\theta^{3/2}\sqrt{n}}\right), \tag{122}$$

which can also be written as

$$\left| F_{X_n}(a) - \eta_Y\left(\theta, a, n\right)\right| \leqslant \exp\left(nK_Y(\theta) - \theta\, a\right) \min\left(1, \frac{2\,c\,\xi_Y(\theta)}{(K_Y^{(2)}(\theta))^{3/2}\sqrt{n}}\right). \tag{123}$$

This completes the proof.

B Proof of Lemma 1

Let $g : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ be for all $(\theta, a, n) \in \mathbb{R}^2 \times \mathbb{N}$,

$$g(\theta, a, n) = nK_Y(\theta) - \theta a = n\ln\left(\varphi_Y(\theta)\right) - \theta\, a. \tag{124}$$

First, note that for all $\theta \in \Theta_Y$ and for all $n \in \mathbb{N}$, the function g is a concave function of a. Hence, from the definition of the function h in (31), h is concave.
Second, note that $0 \in \Theta_Y$ given that $\phi_Y(0) = 1 < \infty$. Hence, from (31), it holds that for all $a \in \mathbb{R}$,

$$h(a) \leqslant nK_Y(0) = n\ln\left(\varphi_Y(0)\right) = n\ln\left(1\right) = 0. \tag{125a}$$

This shows that the function h in (31) is not positive.
Third, the next step of the proof consists of proving the equality in (33). For doing so, Let $\theta^\star : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ be for all $(a, n) \in \mathbb{R} \times \mathbb{N}$,

$$\theta^\star(a, n) = \arg\inf_{\theta\in\Theta_Y} g(\theta, a, n). \tag{126}$$

Note that the function g is a convex in θ. This follows by verifying that its second derivative with respect to θ is positive. That is,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}g(\theta,a,n) = \frac{n}{\varphi_Y(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi_Y(\theta) - a, \text{ and} \tag{127a}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}g(\theta,a,n) = \frac{n}{(\varphi_Y(\theta))^2}\left(\varphi_Y(\theta)\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\varphi_Y(\theta) - \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi_Y(\theta)\right)^2\right) \tag{127b}$$

$$= n\left(\frac{1}{\varphi_Y(\theta)}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\varphi_Y(\theta) - \left(\frac{1}{\varphi_Y(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi_Y(\theta)\right)^2\right) \tag{127c}$$

$$= n\left(\frac{1}{\varphi_Y(\theta)}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\mathbb{E}_{P_Y}[\exp(\theta Y)] - \left(\frac{1}{\varphi_Y(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{P_Y}[\exp(\theta Y)]\right)^2\right) \tag{127d}$$

$$= n\left(\frac{\mathbb{E}_{P_Y}[Y^2\exp(\theta Y)]}{\mathbb{E}_{P_Y}[\exp(\theta Y)]} - \left(\frac{\mathbb{E}_{P_Y}[Y\exp(\theta Y)]}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\right)^2\right) \tag{127e}$$

$$= n\left(\mathbb{E}_{P_Y}\left[\frac{Y^2\exp(\theta Y)}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\right] - \left(\mathbb{E}_{P_Y}\left[\frac{Y\exp(\theta Y)}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\right]\right)^2\right) \tag{127f}$$

$$= n\left(\mathbb{E}_{P_Y}\left[\frac{Y^2\exp(\theta Y)}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\right] - 2\mathbb{E}_{P_Y}\left[\frac{Y\exp(\theta Y)}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\right]K_Y^{(1)}(\theta) + \left(K_Y^{(1)}(\theta)\right)^2\right)$$

$$= n\mathbb{E}_{P_Y}\left[\frac{\left(Y - K_Y^{(1)}(\theta)\right)^2\exp(\theta Y)}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\right] > 0. \tag{127g}$$

Hence, if the first derivative of g with respect to θ (see (127a)) admits a zero in Θ_Y, then $\theta^\star(a,n)$ is the unique solution in θ to the following equality:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}g(\theta,a,n) = \frac{n}{\varphi_Y(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi_Y(\theta) - a = 0. \tag{128}$$

Equation (128) in θ can be rewritten as follows,

$$\frac{a}{n} = \frac{1}{\varphi_Y(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\varphi_Y(\theta) \tag{129a}$$

$$= \frac{1}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_{P_Y}[\exp(\theta Y)] \tag{129b}$$

$$= \frac{1}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\mathbb{E}_{P_Y}[Y\exp(\theta Y)] \tag{129c}$$

$$= \mathbb{E}_{P_Y}\left[\frac{Y\exp(\theta Y)}{\mathbb{E}_{P_Y}[\exp(\theta Y)]}\right] \tag{129d}$$

$$= K_Y^{(1)}(\theta). \tag{129e}$$

From (129d), it follows that $\frac{a}{n}$ is the mean of a random variable that follows an exponentially tilted distribution with respect to P_Y. Thus, there exists a solution in θ for (129d) if and only if $\frac{a}{n} \in \mathrm{int}\mathcal{C}_Y$. Hence, the equality in (33).
Finally, from (129d), $a = n\mathbb{E}_{P_Y}[Y]$ implies that $\theta^\star(a,n) = 0$. Hence, $h(n\mathbb{E}_{P_Y}[Y]) = 0$ from (33). This completes the proof for $h(n\mathbb{E}_{P_Y}[Y]) = 0$.

C Proof of Theorem 3

From Lemma 1, it holds that given $(a, n) \in \mathbb{R} \times \mathbb{N}$ such that $\frac{a}{n} \in \mathrm{int}\mathcal{C}_Y$,

$$nK_Y^{(1)}(\theta^\star) = a. \tag{130}$$

Then, plugging (130) in the expression of $\eta_Y(\theta^\star, a, n)$, with function η_Y defined in (26), the following holds

$$\begin{aligned}
&\eta_Y(\theta^\star, a, n) \\
=&\mathbb{1}_{\{\theta^\star>0\}} + (-1)^{\mathbb{1}_{\{\theta^\star>0\}}} \exp\left(\frac{1}{2}n(\theta^\star)^2 K_Y^{(2)}(\theta) + nK_Y(\theta^\star) - \theta^\star a\right) Q\left((-1)^{\mathbb{1}_{\{\theta^\star \leqslant 0\}}} \frac{a + n\theta^\star K_Y^{(2)}(\theta^\star) - a}{\sqrt{nK_Y^{(2)}(\theta^\star)}}\right) &\text{(131a)}\\
=&\mathbb{1}_{\{\theta^\star>0\}} + (-1)^{\mathbb{1}_{\{\theta^\star>0\}}} \exp\left(\frac{1}{2}n(\theta^\star)^2 K_Y^{(2)}(\theta) + nK_Y(\theta^\star) - \theta^\star a\right) Q\left((-1)^{\mathbb{1}_{\{\theta^\star \leqslant 0\}}} \theta^\star \sqrt{nK_Y^{(2)}(\theta^\star)}\right) &\text{(131b)}\\
=&\mathbb{1}_{\{\theta^\star>0\}} + (-1)^{\mathbb{1}_{\{\theta^\star>0\}}} \exp\left(\frac{1}{2}n(\theta^\star)^2 K_Y^{(2)}(\theta) + nK_Y(\theta^\star) - \theta^\star a\right) Q\left(|\theta^\star| \sqrt{nK_Y^{(2)}(\theta^\star)}\right) &\text{(131c)}\\
=& \hat{F}_{X_n}(a), &\text{(131d)}
\end{aligned}$$

where equality in (131d) follows (12). Finally, plugging (131d) in (27) yields

$$\left|F_{X_n}(a) - \hat{F}_{X_n}(a)\right| \leqslant \exp\left(nK_Y(\theta^\star) - \theta^\star a\right) \min\left(1, \frac{2\, c\, \xi_Y(\theta^\star)}{\left(K_Y^{(2)}(\theta^\star)\right)^{3/2} \sqrt{n}}\right). \tag{132}$$

This completes the proof by observing that $\frac{a}{n} \in \mathrm{int}\mathcal{C}_Y$ is equivalent to $a \in \mathrm{int}\mathcal{C}_{X_n}$.

D Proof of Lemma 4

The left-hand side of (109) satisfies

$$\begin{aligned}
&\left|P_{X_n}(\mathcal{A}) - (\varphi_Y(\theta))^n \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp(-\theta Z_{n,\theta}) \mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]\right| \\
=&(\varphi_Y(\theta))^n \left|\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp(-\theta S_{n,\theta}) \mathbb{1}_{\{S_{n,\theta}\in\mathcal{A}\}}\right] - \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp(-\theta Z_{n,\theta}) \mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]\right|. \qquad (133)
\end{aligned}$$

The focus is on obtaining explicit expressions for the terms $\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp(-\theta S_{n,\theta}) \mathbb{1}_{\{S_{n,\theta}\in\mathcal{A}\}}\right]$ and $\mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp(-\theta Z_{n,\theta}) \mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]$ in (133). First, consider the case in which the random variable $S_{n,\theta}$ is absolutely continuous and denote its probability density function by $f_{S_{n,\theta}}$ and its cumulative distribution function by $F_{S_{n,\theta}}$. Then,

$$\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp(-\theta S_{n,\theta}) \mathbb{1}_{\{S_{n,\theta}\in\mathcal{A}\}}\right] = \int_{\mathcal{A}} \exp(-\theta x) f_{S_{n,\theta}}(x) \mathrm{d}x. \tag{134}$$

Using integration by parts in (134), under the assumption (*i*) or (*ii*) in Lemma 4, the following holds:

$$\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp(-\theta S_{n,\theta}) \mathbb{1}_{\{S_{n,\theta}\in\mathcal{A}\}}\right] = (-1)^{\mathbb{1}_{\{\theta>0\}}} \exp(-\theta a) F_{S_{n,\theta}}(a) - \int_{\mathcal{A}} \theta \exp(-\theta x) F_{S_{n,\theta}}(x) \mathrm{d}x. \tag{135}$$

Second, consider the case in which the random variable $S_{n,\theta}$ is discrete and denote its probability mass function by $p_{S_{n,\theta}}$ and its cumulative distribution function by $F_{S_{n,\theta}}$. Let the support of $S_{n,\theta}$ be $\{s_0, s_1, \ldots, s_\ell\} \subset \mathbb{R}$, with $\ell \in \mathbb{N}$. Assume that condition (i) in Lemma 4 is satisfied. Then,

$$\mathcal{A} \cap \{s_0, s_1, \ldots, s_l\} = \{s_0, s_1, \ldots, s_u\}, \tag{136}$$

with $u \leqslant \ell$, and

$$\begin{aligned}
&\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp\left(-\theta S_{n,\theta}\right) \mathbb{1}_{\{S_{n,\theta}\in\mathcal{A}\}}\right] \\
&= \sum_{k=0}^{u} \exp\left(-\theta s_k\right) p_{S_{n,\theta}}(s_k) && \text{(137a)}\\
&= F_{S_{n,\theta}}(s_0)\exp\left(-\theta s_0\right) + \sum_{k=1}^{u}\left(F_{S_{n,\theta}}(s_k) - F_{S_{n,\theta}}(s_{k-1})\right)\exp\left(-\theta s_k\right) && \text{(137b)}\\
&= \sum_{k=0}^{u} F_{S_{n,\theta}}(s_k)\exp\left(-\theta s_k\right) - \sum_{k=1}^{u} F_{S_{n,\theta}}(s_{k-1})\exp\left(-\theta s_k\right) && \text{(137c)}\\
&= \sum_{k=0}^{u} F_{S_{n,\theta}}(s_k)\exp\left(-\theta s_k\right) - \sum_{k=0}^{u-1} F_{S_{n,\theta}}(s_k)\exp\left(-\theta s_{k+1}\right) && \text{(137d)}\\
&= F_{S_{n,\theta}}(s_u)\exp\left(-\theta s_u\right) - \sum_{k=0}^{u-1} F_{S_{n,\theta}}(s_k)\left(\exp\left(-\theta s_{k+1}\right) - \exp\left(-\theta s_k\right)\right) && \text{(137e)}\\
&= F_{S_{n,\theta}}(s_u)\exp\left(-\theta s_u\right) - \sum_{k=0}^{u-1}\int_{s_k}^{s_{k+1}} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(s_k)\mathrm{d}t && \text{(137f)}\\
&= F_{S_{n,\theta}}(s_u)\exp\left(-\theta s_u\right) - \int_{s_0}^{s_u} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(t)\mathrm{d}t && \text{(137g)}\\
&= F_{S_{n,\theta}}(a)\exp\left(-\theta a\right) - F_{S_{n,\theta}}(a)\exp\left(-\theta a\right) + F_{S_{n,\theta}}(s_u)\exp\left(-\theta s_u\right) - \int_{s_0}^{s_u} F_{S_{n,\theta}}(t)\theta\exp\left(-\theta t\right)\mathrm{d}t && \text{(137h)}\\
&= F_{S_{n,\theta}}(a)\exp\left(-\theta a\right) - F_{S_{n,\theta}}(s_u)\exp\left(-\theta a\right) + F_{S_{n,\theta}}(s_u)\exp\left(-\theta s_u\right) - \int_{s_0}^{s_u} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(t)\mathrm{d}t && \text{(137i)}\\
&= F_{S_{n,\theta}}(a)\exp\left(-\theta a\right) - F_{S_{n,\theta}}(s_u)\left(\exp\left(-\theta a\right) - \exp\left(-\theta s_u\right)\right) - \int_{s_0}^{s_u} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(t)\mathrm{d}t && \text{(137j)}\\
&= F_{S_{n,\theta}}(a)\exp\left(-\theta a\right) - \int_{s_u}^{a} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(s_u)\mathrm{d}t - \int_{s_0}^{s_u} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(t)\mathrm{d}t && \text{(137k)}\\
&= \exp\left(-\theta a\right) F_{S_{n,\theta}}(a) - \int_{s_0}^{a} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(t)\mathrm{d}t && \text{(137l)}\\
&= \exp\left(-\theta a\right) F_{S_{n,\theta}}(a) - \int_{-\infty}^{a} \theta\exp\left(-\theta t\right) F_{S_{n,\theta}}(t)\mathrm{d}t, && \text{(137m)}
\end{aligned}$$

which is an expression of the same form as the one in (135). Alternatively, assume that condition (ii) in Lemma 4 holds. Then,

$$\mathcal{A} \cap \{s_0, s_1, \ldots, s_l\} = \{s_u, s_{u+1}, \ldots, s_l\}, \tag{138}$$

with $u \leqslant \ell$, and

$$\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp\left(-\theta S_{n,\theta}\right) \mathbb{1}_{\{S_{n,\theta}\in\mathcal{A}\}}\right]$$

$$= \sum_{k=u}^{l} \exp\left(-\theta s_k\right) p_{S_{n,\theta}}(s_k) \tag{139a}$$

$$= \big(F_{S_{n,\theta}}(s_u) - F_{S_{n,\theta}}(a)\big) \exp\left(-\theta s_u\right) + \sum_{k=u+1}^{l} \big(F_{S_{n,\theta}}(s_k) - F_{S_{n,\theta}}(s_{k-1})\big) \exp\left(-\theta s_k\right) \tag{139b}$$

$$= -F_{S_{n,\theta}}(a) \exp\left(-\theta s_u\right) + \sum_{k=u}^{l} F_{S_{n,\theta}}(s_k) \exp\left(-\theta s_k\right) - \sum_{k=u+1}^{l} F_{S_{n,\theta}}(s_{k-1}) \exp\left(-\theta s_k\right) \tag{139c}$$

$$= -F_{S_{n,\theta}}(a) \exp\left(-\theta s_u\right) + \sum_{k=u}^{l} F_{S_{n,\theta}}(s_k) \exp\left(-\theta s_k\right) - \sum_{k=u}^{l-1} F_{S_{n,\theta}}(s_k) \exp\left(-\theta s_{k+1}\right) \tag{139d}$$

$$= F_{S_{n,\theta}}(s_l) \exp\left(-\theta s_l\right) - F_{S_{n,\theta}}(a) \exp\left(-\theta s_u\right) - \sum_{k=u}^{l-1} F_{S_{n,\theta}}(s_k) \left(\exp\left(-\theta s_{k+1}\right) - \exp\left(-\theta s_k\right)\right) \tag{139e}$$

$$= -F_{S_{n,\theta}}(a) \exp\left(-\theta s_u\right) - \int_{s_l}^{\infty} \theta \exp\left(-\theta s_t\right) F_{S_{n,\theta}}(s_l) \mathrm{d}t - \sum_{k=u}^{l-1} \int_{s_k}^{s_{k+1}} \theta \exp\left(-\theta t\right) F_{S_{n,\theta}}(s_k) \mathrm{d}t \tag{139f}$$

$$= -F_{S_{n,\theta}}(a) \exp\left(-\theta s_u\right) - \int_{s_u}^{\infty} \theta \exp\left(-\theta t\right) F_{S_{n,\theta}}(t) \mathrm{d}t \tag{139g}$$

$$= F_{S_{n,\theta}}(a) \exp\left(-\theta a\right) - F_{S_{n,\theta}}(a) \exp\left(-\theta a\right) - F_{S_{n,\theta}}(a) \exp\left(-\theta s_u\right) - \int_{s_u}^{\infty} \theta \exp\left(-\theta t\right) F_{S_{n,\theta}}(t) \mathrm{d}t \tag{139h}$$

$$= -F_{S_{n,\theta}}(a) \exp\left(-\theta a\right) - F_{S_{n,\theta}}(a) \left(\exp\left(-\theta s_u\right) - \exp\left(-\theta a\right)\right) - \int_{s_u}^{\infty} \theta \exp\left(-\theta t\right) F_{S_{n,\theta}}(t) \mathrm{d}t \tag{139i}$$

$$= -F_{S_{n,\theta}}(a) \exp\left(-\theta a\right) - \int_{a}^{s_u} \theta \exp\left(-\theta t\right) F_{S_{n,\theta}}(a) \mathrm{d}t - \int_{s_u}^{\infty} \theta \exp\left(-\theta t\right) F_{S_{n,\theta}}(t) \mathrm{d}t \tag{139j}$$

$$= -F_{S_{n,\theta}}(a) \exp\left(-\theta a\right) - \int_{a}^{\infty} \theta \exp\left(-\theta t\right) F_{S_{n,\theta}}(t) \mathrm{d}t, \tag{139k}$$

which is an expression of the same form as those in (135) and (137m).
Note that under the assumption that at least one of the conditions in Lemma 4 holds, the expressions in (135), (137m), and (139k) can be jointly written as follows:

$$\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp\left(-\theta S_{n,\theta}\right) \mathbb{1}_{\{S_{n,\theta} \in \mathcal{A}\}}\right] = (-1)^{\mathbb{1}_{\{\theta>0\}}} \exp\left(-\theta a\right) F_{S_{n,\theta}}(a) - \int_{\mathcal{A}} \theta \exp\left(-\theta x\right) F_{S_{n,\theta}}(x) \mathrm{d}x. \tag{140}$$

The expression in (140) does not involve particular assumptions on the random variable $S_{n,\theta}$ other than being discrete or absolutely continuous. Hence, the same expression holds with respect to the random variable $Z_{n,\theta}$ in (133). More specifically,

$$\mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp\left(-\theta Z_{n,\theta}\right) \mathbb{1}_{\{Z_{n,\theta} \in \mathcal{A}\}}\right] = (-1)^{\mathbb{1}_{\{\theta>0\}}} \exp\left(-\theta a\right) F_{Z_{n,\theta}}(a) - \int_{\mathcal{A}} \theta \exp\left(-\theta x\right) F_{Z_{n,\theta}}(x) \mathrm{d}x, \tag{141}$$

where $F_{Z_{n,\theta}}$ is the cumulative distribution function of the random variable $Z_{n,\theta}$.
The proof ends by plugging (140) and (141) in the right-hand side of (133). This yields,

$$\left| P_{X_n}(\mathcal{A}) - \left(\varphi_Y(\theta)\right)^n \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp\left(-\theta Z_{n,\theta}\right) \mathbb{1}_{\{Z_{n,\theta} \in \mathcal{A}\}}\right] \right|$$
$$= \left(\varphi_Y(\theta)\right)^n \left| (-1)^{\mathbb{1}_{\{\theta>0\}}} \exp\left(-\theta a\right) F_{S_{n,\theta}}(a) - \int_{\mathcal{A}} \theta \exp\left(-\theta x\right) F_{S_{n,\theta}}(x) \mathrm{d}x \right.$$

$$
\begin{aligned}
&-(-1)^{\mathbb{1}_{\{\theta>0\}}} \exp(-\theta a) F_{Z_{n,\theta}}(a) + \int_{\mathcal{A}} \theta \exp(-\theta x) F_{Z_{n,\theta}}(x)\mathrm{d}x\Big| && \text{(142a)}\\
&= (\varphi_Y(\theta))^n \left|(-1)^{\mathbb{1}_{\{\theta>0\}}} \exp(-a) \left(F_{S_{n,\theta}}(a) - F_{Z_{n,\theta}}(a)\right) - \int_{\mathcal{A}} \theta \exp(-\theta x) \left(F_{S_{n,\theta}}(x) - F_{Z_{n,\theta}}(x)\right) \mathrm{d}x\right| && \text{(142b)}\\
&\leq (\varphi_Y(\theta))^n \left(\left|\exp(-\theta a) \left(F_{S_{n,\theta}}(a) - F_{Z_{n,\theta}}(a)\right)\right| + \left|\int_{\mathcal{A}} \theta \exp(-\theta x) \left(F_{S_{n,\theta}}(x) - F_{Z_{n,\theta}}(x)\right) \mathrm{d}x\right|\right) && \text{(142c)}\\
&\leq (\varphi_Y(\theta))^n \left(\exp(-\theta a) \Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right) + \int_{\mathcal{A}} |\theta \exp(-\theta x)| \Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right) \mathrm{d}x\right) && \text{(142d)}\\
&= (\varphi_Y(\theta))^n \left(\exp(-\theta a) \Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right) + \Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right) \left|\int_{\mathcal{A}} \theta \exp(-\theta x) \mathrm{d}x\right|\right) && \text{(142e)}\\
&= (\varphi_Y(\theta))^n \left(\exp(-\theta a) \Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right) + \Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right) \exp(-\theta a)\right) && \text{(142f)}\\
&= 2\frac{(\varphi_Y(\theta))^n}{\exp(\theta a)} \Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right). && \text{(142g)}
\end{aligned}
$$

Finally, under the assumption that at least one of the conditions in Lemma 4 holds. Then,

$$
\begin{aligned}
&\left|P_{X_n}(\mathcal{A}) - (\varphi_Y(\theta))^n \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp(-\theta Z_{n,\theta}) \mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]\right|\\
&\leq (\varphi_Y(\theta))^n \max\left(\mathbb{E}_{P_{S_{n,\theta}}}\left[\exp(-\theta S_{n,\theta}) \mathbb{1}\left\{S_{n,\theta} \in \mathcal{A}\right\}\right], \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp(-\theta Z_{n,\theta}) \mathbb{1}\left\{Z_{n,\theta} \in \mathcal{A}\right\}\right]\right) && \text{(143a)}\\
&\leq (\varphi_Y(\theta))^n \exp(-\theta a) = \frac{(\varphi_Y(\theta))^n}{\exp(\theta a)}. && \text{(143b)}
\end{aligned}
$$

Under the same assumption, the expressions in (142g) and (143b) can be jointly written as follows:

$$
\left|P_{X_n}(\mathcal{A}) - (\varphi_Y(\theta))^n \mathbb{E}_{P_{Z_{n,\theta}}}\left[\exp(-\theta Z_{n,\theta}) \mathbb{1}_{\{Z_{n,\theta}\in\mathcal{A}\}}\right]\right| \leqslant \frac{(\varphi_Y(\theta))^n}{\exp(\theta a)} \min\left(2\Delta\left(P_{S_{n,\theta}}, P_{Z_{n,\theta}}\right), 1\right). \tag{144}
$$

This concludes the proof of Lemma 4.

E Proof of Lemma 5

In the case in which Y is discrete (p_Y, $p_{Y^{(\theta)}}$, $p_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}$ denote probability mass functions) or absolutely continuous random variables (p_Y, $p_{Y^{(\theta)}}$, $p_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}$ denote probability density functions), the following holds for all $(y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$,

$$
\frac{\mathrm{d}P_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}}{\mathrm{d}P_{Y_1 Y_2 \ldots Y_n}}(y_1, y_2, \ldots, y_n) = \frac{p_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}(y_1, y_2, \ldots, y_n)}{\prod_{j=1}^n p_Y(y_j)}, \tag{145}
$$

and for all $y \in \mathbb{R}$,

$$
\frac{\mathrm{d}P_{Y^{(\theta)}}}{\mathrm{d}P_Y}(y) = \frac{p_{Y^{(\theta)}}(y)}{p_Y(y)}. \tag{146}
$$

Equating the right-hand side of both (103) and (145), it yields for all $(y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$

$$
p_{Y_1^{(\theta)} Y_2^{(\theta)} \ldots Y_n^{(\theta)}}(y_1, y_2, \ldots, y_n) = \prod_{j=1}^n \frac{\exp(\theta y_j)}{\varphi_Y(\theta)} p_Y(y_j). \tag{147}
$$

Hence, $Y_1^{(\theta)}$, $Y_2^{(\theta)}$, ..., $Y_n^{(\theta)}$ are mutually independent and identically distributed. Moreover, for all $y \in \mathbb{R}$,

$$p_{Y^{(\theta)}}(y)=\frac{\exp(\theta y)}{\varphi_Y(\theta)}p_Y(y). \tag{148}$$

Finally, plugging (148) in (146) yields, for all $y \in \mathbb{R}$,

$$\frac{\mathrm{d}P_{Y^{(\theta)}}}{\mathrm{d}P_Y}(y)=\frac{\exp(\theta y)}{\varphi_Y(\theta)}, \tag{149}$$

which completes the proof.

F Proof of Theorem 5

Note that for a given distribution $P_{\boldsymbol{X}}$ subject (51) and for a random transformation in (41) subject to (42), the upper bound $T(n, M, P_{\boldsymbol{X}})$ in (49) can be written in the form of a weighted sum of the CDF and the complementary CDF of the random variables variables W_n and V_n that are sums of i.i.d random variables, respectively. That is

$$W_n=\sum_{t=1}^{n}\iota(X_t;Y_t), \text{ and} \tag{150}$$

$$V_n=\sum_{t=1}^{n}\iota(\bar{X}_t;Y_t), \tag{151}$$

where $(X_t, Y_t) \sim P_X P_{Y|X}$ and $(\bar{X}_t, Y_t) \sim P_{\bar{X}} P_Y$ with $P_X = P_{\bar{X}}$. More specifically, the function T in (49) can be rewritten in the form

$$T(n, M, P_{\boldsymbol{X}})=F_{W_n}\left(\ln\left(\frac{M-1}{2}\right)\right)+\frac{M-1}{2}\left(1-F_{V_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\right), \tag{152}$$

where F_{W_n} and F_{V_n} are the CDFs of W_n and V_n, respectively.
The next step consists in deriving the upper and lower bounds on $F_{W_n}\left(\ln\left(\frac{M-1}{2}\right)\right)$ and $1 - F_{V_n}\left(\ln\left(\frac{M-1}{2}\right)\right)$ by using the result of Theorem 3. That is,

$$F_{W_n}\left(\ln\left(\frac{M-1}{2}\right)\right) \leqslant \zeta_{\iota(X;Y)}\left(\theta, \ln\left(\frac{M-1}{2}\right), n\right)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\min\left(1, \frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \tag{153}$$

$$F_{W_n}\left(\ln\left(\frac{M-1}{2}\right)\right) \geqslant \zeta_{\iota(X;Y)}\left(\theta, \ln\left(\frac{M-1}{2}\right), n\right)-\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\min\left(1, \frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \tag{154}$$

$$1-F_{V_n}\left(\ln\left(\frac{M-1}{2}\right)\right) \leqslant 1-\zeta_{\iota(\bar{X};Y)}\left(\theta, \ln\left(\frac{M-1}{2}\right), n\right)+\exp\left(n\ln\left(\varphi_{\iota(\bar{X};Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\min\left(1, \frac{2\,c\,\xi_{\iota(\bar{X};Y)}(\theta)}{\left(V_{\iota(\bar{X};Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \tag{155}$$

and

$$
\begin{aligned}
&1 - F_{V_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\\
&\geqslant 1-\zeta_{\iota(\bar{X};Y)}\left(\theta, \ln\left(\frac{M-1}{2}\right), n\right) - \exp\left(n\ln\left(\varphi_{\iota(\bar{X};Y)}(\theta)\right) - \theta\ln\left(\frac{M-1}{2}\right)\right)\min\left(1, \frac{2\,c\,\xi_{\iota(\bar{X};Y)}(\theta)}{\left(V_{\iota(\bar{X};Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \quad (156)
\end{aligned}
$$

where θ and τ satisfy

$$
n\mu_{\iota(X;Y)}(\theta) = \ln\left(\frac{M-1}{2}\right) = n\mu_{\iota(\bar{X};Y)}(\tau), \quad (157)
$$

with for all $t \in \mathbb{R}$,

$$
\begin{aligned}
\varphi_{\iota(X;Y)}(t) &= \mathbb{E}_{P_X P_{Y|X}}\left[\exp(t\,\iota(X;Y))\right], && (158)\\
\varphi_{\iota(\bar{X};Y)}(t) &= \mathbb{E}_{P_{\bar{X}} P_Y}\left[\exp\left(t\,\iota(\bar{X};Y)\right)\right], && (159)\\
\mu_{\iota(X;Y)}(t) &= \mathbb{E}_{P_X P_{Y|X}}\left[\iota(X;Y)\frac{\exp(t\,\iota(X;Y))}{\varphi_{\iota(X;Y)}(t)}\right], && (160)\\
\mu_{\iota(\bar{X};Y)}(t) &= \mathbb{E}_{P_{\bar{X}} P_Y}\left[\iota(\bar{X};Y)\frac{\exp\left(t\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(\bar{X};Y)}(t)}\right], && (161)\\
V_{\iota(X;Y)}(t) &= \mathbb{E}_{P_X P_{Y|X}}\left[\left(\iota(X;Y) - \mu_{\iota(X;Y)}(t)\right)^2\frac{\exp(t\,\iota(X;Y))}{\varphi_{\iota(X;Y)}(t)}\right], && (162)\\
V_{\iota(\bar{X};Y)}(t) &= \mathbb{E}_{P_{\bar{X}} P_Y}\left[\left(\iota(\bar{X};Y) - \mu_{\iota(\bar{X};Y)}(t)\right)^2\frac{\exp\left(t\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(\bar{X};Y)}(t)}\right], && (163)\\
\xi_{\iota(X;Y)}(t) &= \mathbb{E}_{P_X P_{Y|X}}\left[\left|\iota(X;Y) - \mu_{\iota(X;Y)}(t)\right|^3\frac{\exp(t\,\iota(X;Y))}{\varphi_{\iota(X;Y)}(t)}\right], && (164)\\
\xi_{\iota(\bar{X};Y)}(t) &= \mathbb{E}_{P_{\bar{X}} P_Y}\left[\left|\iota(\bar{X};Y) - \mu_{\iota(\bar{X};Y)}(t)\right|^3\frac{\exp\left(t\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(\bar{X};Y)}(t)}\right], && (165)
\end{aligned}
$$

and for all $(t, a, n) \in \mathbb{R}^2 \times \mathbb{N}$

$$
\begin{aligned}
&\zeta_{\iota(X;Y)}(t,a,n)\\
&\triangleq \mathbb{1}_{\{t>0\}} + (-1)^{\mathbb{1}_{\{t>0\}}}\exp\left(\frac{1}{2}nt^2V_{\iota(X;Y)}(t) + n\ln\left(\varphi_{\iota(X;Y)}(t)\right) - ta\right)Q\left(|t|\sqrt{n\,V_{\iota(X;Y)}(t)}\right), \quad (166)\\
&\zeta_{\iota(\bar{X};Y)}(t,a,n)\\
&\triangleq \mathbb{1}_{\{t>0\}} + (-1)^{\mathbb{1}_{\{t>0\}}}\exp\left(\frac{1}{2}nt^2V_{\iota(\bar{X};Y)}(t) + n\ln\left(\varphi_{\iota(\bar{X};Y)}(t)\right) - ta\right)Q\left(|t|\sqrt{n\,V_{\iota(\bar{X};Y)}(t)}\right). \quad (167)
\end{aligned}
$$

The next step consists in simplifying the expressions in the right hand-side of (155) and (156) by studying the relation between $\varphi_{\iota(X;Y)}$ and $\varphi_{\iota(\bar{X};Y)}$, θ and τ, $V_{\iota(X;Y)}$ and $V_{\iota(\bar{X};Y)}$, $\xi_{\iota(X;Y)}$ and $\xi_{\iota(\bar{X};Y)}$.
First, from (158), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}} P_Y$ because $P_X P_{Y|X}$ is absolutely continuous with respect to $P_{\bar{X}} P_Y$, it holds that

$$
\varphi_{\iota(X;Y)}(t) = \mathbb{E}_{P_{\bar{X}} P_Y}\left[\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}} P_Y}\left(\bar{X};Y\right)\exp(t\,\iota(\bar{X};Y))\right] \quad (168)
$$

$$=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\exp\left((t+1)\,\iota(\bar{X};Y)\right)\right]. \tag{169}$$

Then, from (158) and (159), it holds that

$$\varphi_{\iota(X;Y)}(t)=\varphi_{\iota(\bar{X};Y)}(t+1). \tag{170}$$

This concludes the relation between $\varphi_{\iota(X;Y)}$ and $\varphi_{\iota(\bar{X};Y)}$.
Second, from (160), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}}P_Y$, it holds that

$$\mu_{\iota(X;Y)}(t)=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\iota(\bar{X};Y)\frac{\exp(t\,\iota(\bar{X};Y))}{\varphi_{\iota(X;Y)}(t)}\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}}P_Y}\left(\bar{X};Y\right)\right] \tag{171}$$

$$=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\iota(\bar{X};Y)\frac{\exp\left((t+1)\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(X;Y)}(t)}\right]. \tag{172}$$

Then, from (170) and (172), it holds that

$$\mu_{\iota(X;Y)}(t)=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\iota(\bar{X};Y)\frac{\exp\left((t+1)\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(\bar{X};Y)}(t+1)}\right]. \tag{173}$$

From (161) and (173), it holds that

$$\mu_{\iota(X;Y)}(t)=\mu_{\iota(\bar{X};Y)}(t+1). \tag{174}$$

This concludes the relation between $\mu_{\iota(X;Y)}$ and $\mu_{\iota(\bar{X};Y)}$.
Third, from (157) and (174), it holds that

$$\tau=\theta+1. \tag{175}$$

This concludes the relation between τ and θ.
Fourth, from (162), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}}P_Y$, it holds that

$$V_{\iota(X;Y)}(t)=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\left(\iota(\bar{X};Y)-\mu_{\iota(X;Y)}(t)\right)^2\frac{\exp(t\,\iota(\bar{X};Y))}{\varphi_{\iota(X;Y)}(t)}\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}}P_Y}\left(\bar{X};Y\right)\right] \tag{176}$$

$$=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\left(\iota(\bar{X};Y)-\mu_{\iota(X;Y)}(t)\right)^2\frac{\exp\left((t+1)\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(X;Y)}(t)}\right]. \tag{177}$$

From (170), (174) and (177), it holds that

$$V_{\iota(X;Y)}(t)=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\left(\iota(\bar{X};Y)-\mu_{\iota(\bar{X};Y)}(t+1)\right)^2\frac{\exp\left((t+1)\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(\bar{X};Y)}(t+1)}\right]. \tag{178}$$

From (163) and (178), it holds that

$$V_{\iota(X;Y)}(t)=V_{\iota(\bar{X};Y)}(t+1). \tag{179}$$

This concludes the relation between $V_{\iota(X;Y)}$ and $V_{\iota(\bar{X};Y)}$.
Fifth, from (164), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}}P_Y$, it holds that

$$\xi_{\iota(X;Y)}(t)=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\left|\iota(\bar{X};Y)-\mu_{\iota(X;Y)}(t)\right|^3\frac{\exp(t\,\iota(\bar{X};Y))}{\varphi_{\iota(X;Y)}(t)}\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}}P_Y}\left(\bar{X};Y\right)\right] \tag{180}$$

$$=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\left|\iota(\bar{X};Y)-\mu_{\iota(X;Y)}(t)\right|^3\frac{\exp\left((t+1)\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(X;Y)}(t)}\right]. \tag{181}$$

From (170), (174) and (181), it holds that

$$\xi_{\iota(X;Y)}(t)=\mathbb{E}_{P_{\bar{X}}P_Y}\left[\left|\iota(\bar{X};Y)-\mu_{\iota(\bar{X};Y)}(t+1)\right|^3\frac{\exp\left((t+1)\,\iota(\bar{X};Y)\right)}{\varphi_{\iota(\bar{X};Y)}(t+1)}\right]. \tag{182}$$

From (165) and (182), it holds that

$$\xi_{\iota(X;Y)}(t)=\xi_{\iota(\bar{X};Y)}(t+1). \tag{183}$$

This concludes the relation between $\xi_{\iota(X;Y)}$ and $\xi_{\iota(\bar{X};Y)}$.
Sixth, plugging (170), (174), and (179) into (166), for all $t\in\mathbb{R}$, it holds that

$$\begin{aligned}&\zeta_{\iota(\bar{X};Y)}(t,a,n)\\&\triangleq\mathbb{1}_{\{t>0\}}+(-1)^{\mathbb{1}_{\{t>0\}}}\exp\left(\frac{1}{2}nt^2V_{\iota(X;Y)}(t-1)+n\ln\left(\varphi_{\iota(X;Y)}(t-1)\right)-ta\right)Q\left(|t|\sqrt{nV_{\iota(X;Y)}(t-1)}\right).\end{aligned} \tag{184}$$

Then, from (65) and (184), it holds that

$$\zeta_{\iota(\bar{X};Y)}\left(t,\ln\left(\frac{M-1}{2}\right),n\right)=1-\beta_2(n,M,t-1,P_X). \tag{185}$$

Then, plugging (170), (174), (175), (179), (183), and (185) into the right hand-side of (155), it holds that

$$\begin{aligned}&1-F_{V_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\\&\leqslant\beta_2(n,M,\theta,P_X)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-(\theta+1)\ln\left(\frac{M-1}{2}\right)\right)\min\left(1,\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) &(186)\\&\leqslant\beta_2(n,M,\theta,P_X)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-(\theta+1)\ln\left(\frac{M-1}{2}\right)\right)\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}. &(187)\end{aligned}$$

Alternatively, plugging (170), (174), (175), (179), (183), and (185) into the right hand-side of (156), it holds that

$$\begin{aligned}&1-F_{V_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\\&\geqslant\beta_2(n,M,\theta,P_X)-\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-(\theta+1)\ln\left(\frac{M-1}{2}\right)\right)\min\left(1,\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) &(188)\\&\geqslant\beta_2(n,M,\theta,P_X)-\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-(\theta+1)\ln\left(\frac{M-1}{2}\right)\right)\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}} &(189)\\&=G_2(n,M,\theta,P_X), &(190)\end{aligned}$$

where the equality in (190) follows from (67). Observing that $1-F_{V_n}$ is a positive function, then from (189), it holds that

$$1-F_{V_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\geqslant \max\left(0, G_2(n,M,\theta,P_X)\right). \tag{191}$$

Seventh, from (64) and (166), it holds that

$$\zeta_{\iota(X;Y)}\left(t,\ln\left(\frac{M-1}{2}\right),n\right) = \beta_1(n,M,t,P_X). \tag{192}$$

Then, plugging (170), (174), (175), (179), (183), and (192) into the right hand-side of (153), it holds that

$$\begin{aligned}
&F_{W_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\\
&\leqslant \beta_1(n,M,\theta,P_X)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\min\left(1,\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) && (193)\\
&\leqslant \beta_1(n,M,\theta,P_X)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}. && (194)
\end{aligned}$$

Alternatively, plugging (170), (174), (175), (179), (183), and (185) into the right hand-side of (154), it holds that

$$\begin{aligned}
&F_{W_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\\
&\geqslant \beta_1(n,M,\theta,P_X)-\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\min\left(1,\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) && (195)\\
&\geqslant \beta_1(n,M,\theta,P_X)-\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\frac{2\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}} && (196)\\
&= G_1(n,M,\theta,P_X), && (197)
\end{aligned}$$

where the equality in (197) follows from (66). Observing that F_{W_n} is a positive function, then from (196), it holds that

$$F_{W_n}\left(\ln\left(\frac{M-1}{2}\right)\right)\geqslant \max\left(0, G_1(n,M,\theta,P_X)\right). \tag{198}$$

Finally, plugging (187) and (194) in (152), it holds that

$$\begin{aligned}
&T(n,M,P_{\boldsymbol{X}})\\
&\leqslant\beta_1(n,M,\theta,P_X)+\frac{M-1}{2}\beta_2(n,M,\theta,P_X)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\frac{4c\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}} && (199)\\
&= \beta(n,M,\theta,P_X)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\frac{4\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}, && (200)
\end{aligned}$$

where the equality in (200) follows from (72). Observing that $T(n,M,P_{\boldsymbol{X}})\leqslant 1$, from (200), it holds that

$$T(n,M,P_{\boldsymbol{X}})\leqslant\min\left(1,\beta(n,M,\theta,P_X)+\exp\left(n\ln\left(\varphi_{\iota(X;Y)}(\theta)\right)-\theta\ln\left(\frac{M-1}{2}\right)\right)\frac{4\,c\,\xi_{\iota(X;Y)}(\theta)}{\left(V_{\iota(X;Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) \tag{201}$$

$$=S(n,M,\theta,P_X), \tag{202}$$

where the equality in (197) follows from (69).
Alternatively, plugging (190) and (197) in (152), it holds that

$$T(n,M,P_{\boldsymbol{X}}) \geqslant \max\left(0, G_1(n,M,\theta,P_X)\right) + \frac{M-1}{2}\max\left(0, G_2(n,M,\theta,P_X)\right) \tag{203}$$

$$=G(n,M,\theta,P_X), \tag{204}$$

where the equality in (197) follows from (69). Combining (202) and (204) concludes the proof.

G Proof of Theorem 7

Note that for given distributions $P_{\boldsymbol{X}}$ subject (51), $Q_{\boldsymbol{Y}}$ subject to (81), and for a random transformation in (41) subject to (42), the lower bound $C(n{,}M{,}P_{\boldsymbol{X}}{,}Q_{\boldsymbol{Y}}{,}\gamma)$ in (76) can be written in the form of a weighted sum of the CDF and the complementary CDF of the random variables variables W_n and V_n that are sums of i.i.d random variables, respectively. That is

$$W_n = \sum_{t=1}^{n} \tilde{\imath}(X_t;Y_t|Q_Y), \tag{205}$$

$$V_n = \sum_{t=1}^{n} \tilde{\imath}(\bar{X}_t;Y_t|Q_Y), \tag{206}$$

where $(X_t,Y_t) \sim P_X P_{Y|X}$ and $(\bar{X}_t,Y_t) \sim P_{\bar{X}} Q_Y$ with $P_X = P_{\bar{X}}$. More specifically, the function C in (76) can be written in the form

$$C(n,M,P_{\boldsymbol{X}},Q_{\boldsymbol{Y}},\gamma) = F_{W_n}\left(\ln(\gamma)\right) + \gamma\left(1 - F_{V_n}\left(\ln(\gamma)\right)\right) - \frac{\gamma}{M}, \tag{207}$$

where F_{W_n} and F_{V_n} are the CDFs of the random variables W_n and V_n, respectively.
The next step consists in deriving the upper and lower bounds on $F_{W_n}\left(\ln(\gamma)\right)$ and $1-F_{V_n}\left(\ln(\gamma)\right)$ by using the result of Theorem 3. That is

$$F_{W_n}\left(\ln(\gamma)\right) \leqslant \zeta_{\tilde{\imath}(X;Y|Q_Y)}(\theta,\ln(\gamma),n) + \exp\left(n\ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right)\min\left(1, \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \tag{208}$$

$$F_{W_n}\left(\ln(\gamma)\right) \geqslant \zeta_{\tilde{\imath}(X;Y|Q_Y)}(\theta,\ln(\gamma),n) - \exp\left(n\ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right)\min\left(1, \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \tag{209}$$

$$1 - F_{V_n}\left(\ln(\gamma)\right) \leqslant 1 - \zeta_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta,\ln(\gamma),n) + \exp\left(n\ln\left(\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right)\min\left(1, \frac{2\,c\,\xi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \tag{210}$$

and

$$1 - F_{V_n}\left(\ln(\gamma)\right) \geqslant 1 - \zeta_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta,\ln(\gamma),n) - \exp\left(n\ln\left(\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right)\min\left(1, \frac{2\,c\,\xi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right), \tag{211}$$

where θ and τ satisfy

$$n\mu_{\tilde{\imath}(X;Y|Q_Y)}(\theta)=\ln(\gamma) = n\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}(\tau), \tag{212}$$

with for all $t \in \mathbb{R}$

$$\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_X P_{Y|X}}\left[\exp\left(t\,\tilde{\imath}(X;Y|Q_Y)\right)\right], \tag{213}$$

$$\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}} Q_Y}\left[\exp\left(t\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)\right], \tag{214}$$

$$\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_X P_{Y|X}}\left[\tilde{\imath}(X;Y|Q_Y)\frac{\exp(t\,\tilde{\imath}(X;Y|Q_Y))}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\right], \tag{215}$$

$$\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}} Q_Y}\left[\tilde{\imath}(\bar{X};Y|Q_Y)\frac{\exp\left(t\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)}\right], \tag{216}$$

$$V_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_X P_{Y|X}}\left[\left(\tilde{\imath}(X;Y|Q_Y)-\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)\right)^2\frac{\exp(t\,\tilde{\imath}(X;Y|Q_Y))}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\right], \tag{217}$$

$$V_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}} Q_Y}\left[\left(\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)\right)^2\frac{\exp\left(t\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)}\right], \tag{218}$$

$$\xi_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_X P_{Y|X}}\left[\left|\tilde{\imath}(X;Y|Q_Y)-\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)\right|^3\frac{\exp(t\,\tilde{\imath}(X;Y|Q_Y))}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\right], \tag{219}$$

$$\xi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}} Q_Y}\left[\left|\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)\right|^3\frac{\exp\left(t\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)}\right], \tag{220}$$

and for all $(t,a,n)\in\mathbb{R}^2\times\mathbb{N}$

$$\zeta_{\tilde{\imath}(X;Y|Q_Y)}(t,a,n) \triangleq \mathbb{1}_{\{t>0\}}+(-1)^{\mathbb{1}_{\{t>0\}}}\exp\left(\frac{1}{2}nt^2V_{\tilde{\imath}(X;Y|Q_Y)}(t)+n\ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)\right)-ta\right)Q\left(|t|\sqrt{nV_{\tilde{\imath}(X;Y|Q_Y)}(t)}\right), \tag{221}$$

$$\zeta_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t,a,n) \triangleq \mathbb{1}_{\{t>0\}}+(-1)^{\mathbb{1}_{\{t>0\}}}\exp\left(\frac{1}{2}nt^2V_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)+n\ln\left(\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)\right)-ta\right)Q\left(|t|\sqrt{nV_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t)}\right). \tag{222}$$

The next step consists in simplifying the expressions in the right hand-side of (210) and (211) by studying the relation between $\varphi_{\tilde{\imath}(X;Y|Q_Y)}$ and $\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}$, θ and τ, $V_{\tilde{\imath}(X;Y|Q_Y)}$ and $V_{\tilde{\imath}(\bar{X};Y|Q_Y)}$, $\xi_{\tilde{\imath}(X;Y|Q_Y)}$ and $\xi_{\tilde{\imath}(\bar{X};Y|Q_Y)}$ when the $P_{Y|X}$ is absolutely continuous with respect to Q_Y.
First, from (213), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}}Q_Y$ because $P_X P_{Y|X}$ is absolutely continuous with respect to $P_{\bar{X}}Q_Y$, it holds that

$$\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}} Q_Y}\left[\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}}Q_Y}\left(\bar{X};Y\right)\exp(t\,\tilde{\imath}(\bar{X};Y|Q_Y)\right] \tag{223}$$

$$=\mathbb{E}_{P_{\bar{X}} Q_Y}\left[\exp\left((t+1)\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)\right]. \tag{224}$$

Then, from (213) and (214), it holds that

$$\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)=\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1). \tag{225}$$

This concludes the relation between $\varphi_{\tilde{\imath}(X;Y|Q_Y)}$ and $\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}$.
Second, from (215), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}}Q_Y$, it holds that

$$\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}} Q_Y}\left[\tilde{\imath}(\bar{X};Y|Q_Y)\frac{\exp(t\,\tilde{\imath}(\bar{X};Y|Q_Y)}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}}Q_Y}\left(\bar{X};Y\right)\right] \tag{226}$$

$$=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\tilde{\imath}(\bar{X};Y|Q_Y)\frac{\exp\left((t+1)\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\right]. \quad (227)$$

Then, from (225) and (227), it holds that

$$\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\tilde{\imath}(\bar{X};Y|Q_Y)\frac{\exp\left((t+1)\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1)}\right]. \quad (228)$$

From (216) and (228), it holds that

$$\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1). \quad (229)$$

This concludes the relation between $\mu_{\tilde{\imath}(X;Y|Q_Y)}$ and $\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}$.
Third, from (212) and (229), it holds that

$$\tau=\theta+1. \quad (230)$$

This concludes the relation between τ and θ.
Fourth, from (217), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}}Q_Y$, it holds that

$$V_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\left(\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)\right)^2\frac{\exp(t\,\tilde{\imath}(\bar{X};Y|Q_Y)}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}}Q_Y}\left(\bar{X};Y\right)\right] \quad (231)$$

$$=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\left(\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)\right)^2\frac{\exp\left((t+1)\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\right]. \quad (232)$$

From (225), (229) and (232), it holds that

$$V_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\left(\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1)\right)^2\frac{\exp\left((t+1)\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1)}\right]. \quad (233)$$

From (218) and (233), it holds that

$$V_{\tilde{\imath}(X;Y|Q_Y)}(t)=V_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1). \quad (234)$$

This concludes the relation between $V_{\tilde{\imath}(X;Y|Q_Y)}$ and $V_{\tilde{\imath}(\bar{X};Y|Q_Y)}$.
Fifth, from (219), using the change of measure from $P_X P_{Y|X}$ to $P_{\bar{X}}Q_Y$, it holds that

$$\xi_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\left|\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)\right|^3\frac{\exp(t\,\tilde{\imath}(\bar{X};Y|Q_Y)}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\frac{\mathrm{d}P_X P_{Y|X}}{\mathrm{d}P_{\bar{X}}Q_Y}\left(\bar{X};Y\right)\right] \quad (235)$$

$$=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\left|\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(X;Y|Q_Y)}(t)\right|^3\frac{\exp\left((t+1)\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t)}\right]. \quad (236)$$

From (225), (229) and (236), it holds that

$$\xi_{\tilde{\imath}(X;Y|Q_Y)}(t)=\mathbb{E}_{P_{\bar{X}}Q_Y}\left[\left|\tilde{\imath}(\bar{X};Y|Q_Y)-\mu_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1)\right|^3\frac{\exp\left((t+1)\,\tilde{\imath}(\bar{X};Y|Q_Y)\right)}{\varphi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1)}\right]. \quad (237)$$

From (220) and (237), it holds that

$$\xi_{\tilde{\imath}(X;Y|Q_Y)}(t)=\xi_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t+1). \quad (238)$$

This concludes the relation between $\xi_{\tilde{\imath}(X;Y|Q_Y)}$ and $\xi_{\tilde{\imath}(\bar{X};Y|Q_Y)}$.
Sixth, plugging (225), (229), and (234) into (221), for all $t \in \mathbb{R}$, it holds that

$$\zeta_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t,a,n)$$
$$\triangleq \mathbb{1}_{\{t>0\}} + (-1)^{\mathbb{1}_{\{t>0\}}} \exp\left(\frac{1}{2} n t^2 V_{\tilde{\imath}(X;Y|Q_Y)}(t-1) + n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(t-1)\right) - ta\right) Q\left(|t| \sqrt{n V_{\tilde{\imath}(X;Y|Q_Y)}(t-1)}\right) \quad (239)$$

Then, from (95) and (239), it holds that

$$\zeta_{\tilde{\imath}(\bar{X};Y|Q_Y)}(t, \ln(\gamma), n) = 1 - \tilde{\beta}_2(n, \gamma, t-1, P_X, Q_Y). \quad (240)$$

Then, plugging (225), (229), (230), (234), (238), and (240) into the right hand-side of (210), it holds that

$$1 - F_{V_n}(\ln(\gamma))$$
$$\leqslant \tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y) + \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - (\theta+1)\ln(\gamma)\right) \min\left(1, \frac{2c\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) \quad (241)$$
$$\leqslant \tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y) + \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - (\theta+1)\ln(\gamma)\right) \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}. \quad (242)$$

Alternatively, plugging (225), (229), (230), (234), (238), and (240) into the right hand-side of (211), it holds that

$$1 - F_{V_n}(\ln(\gamma))$$
$$\geqslant \tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y) - \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - (\theta+1)\ln(\gamma)\right) \min\left(1, \frac{2c\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) \quad (243)$$
$$\geqslant \tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y) - \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - (\theta+1)\ln(\gamma)\right) \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}} \quad (244)$$
$$= \tilde{G}_2(n,\gamma,\theta,P_X,Q_Y), \quad (245)$$

where the equality in (245) follows from (97). Observing that $1 - F_{V_n}$ is a positive function, then from (244), it holds that

$$1 - F_{V_n}(\ln(\gamma)) \geqslant \max\left(0, \tilde{G}_2(n,\gamma,\theta,P_X,Q_Y)\right). \quad (246)$$

Seventh, from (94) and (221), it holds that

$$\zeta_{\tilde{\imath}(X;Y|Q_Y)}(t, \ln(\gamma), n) = \tilde{\beta}_1(n,\gamma,t,P_X,Q_Y). \quad (247)$$

Then, plugging (225), (229), (230), (234), (238), and (247) into the right hand-side of (208), it holds that

$$F_{W_n}(\ln(\gamma))$$
$$\leqslant \tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y) + \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right) \min\left(1, \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) \quad (248)$$
$$\leqslant \tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y) + \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right) \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}. \quad (249)$$

Alternatively, plugging (225), (229), (230), (234), (238), and (240) into the right hand-side of (209), it holds that

$$
\begin{aligned}
& F_{W_n}\left(\ln(\gamma)\right) \\
& \geqslant \tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y) - \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right) \min\left(1, \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) && (250)\\
& \geqslant \tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y) - \exp\left(n \ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right) \frac{2\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}} && (251)\\
& = \tilde{G}_1(n,\gamma,\theta,P_X,Q_Y), && (252)
\end{aligned}
$$

where the equality in (252) follows from (96). Observing that F_{W_n} is a positive function, then from (251), it holds that

$$
F_{W_n}\left(\ln(\gamma)\right) \geqslant \max\left(0, \tilde{G}_1(n,\gamma,\theta,P_X,Q_Y)\right). \tag{253}
$$

Finally, plugging (242) and (249) in (207), it holds that

$$
\begin{aligned}
& C(n,M,P_{\boldsymbol{X}},Q_{\boldsymbol{Y}},\gamma) \\
& \leqslant \tilde{\beta}_1(n,\gamma,\theta,P_X,Q_Y) + \gamma\tilde{\beta}_2(n,\gamma,\theta,P_X,Q_Y) + \exp\left(n\ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right)\frac{4c\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}} - \frac{\gamma}{M} && (254)\\
& = \tilde{\beta}(n,\gamma,\theta,P_X,Q_Y,M) + \exp\left(n\ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right)\frac{4c\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}, && (255)
\end{aligned}
$$

where the equality in (251) follows from (100). Observing that $C(n,M,P_{\boldsymbol{X}},Q_{\boldsymbol{Y}},\gamma) + \frac{\gamma}{M} \leqslant 1$, from (255), it holds that

$$
\begin{aligned}
& C(n,M,P_{\boldsymbol{X}},Q_{\boldsymbol{Y}},\gamma) \\
& \leqslant \min\left(1, \tilde{\beta}(n,\gamma,\theta,P_X,Q_Y) + \exp\left(n\ln\left(\varphi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right) - \theta\ln(\gamma)\right)\frac{4\,c\,\xi_{\tilde{\imath}(X;Y|Q_Y)}(\theta)}{\left(V_{\tilde{\imath}(X;Y|Q_Y)}(\theta)\right)^{3/2}\sqrt{n}}\right) && (256)\\
& = \tilde{S}(n,\gamma,\theta,P_X,Q_Y,M), && (257)
\end{aligned}
$$

where, (257) follows from (99).
Alternatively, plugging (245) and (252) in (207), it holds that

$$
\begin{aligned}
C(n,M,P_{\boldsymbol{X}},Q_{\boldsymbol{Y}},\gamma) & \geqslant \max\left(0, \tilde{G}_1(n,\gamma,\theta,P_X,Q_Y)\right) + \gamma\max\left(0, \tilde{G}_2(n,\gamma,\theta,P_X,Q_Y)\right) - \frac{\gamma}{M} && (258)\\
& = \tilde{G}(n,\gamma,\theta,P_X,Q_Y,M), && (259)
\end{aligned}
$$

where the equality in (259) follows from (98). Combining (257) and (259) concludes the proof.

References

[1] 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*, Dublin, Ireland, Jun. 2020.

[2] V. M. Zolotarev, *One-Dimensional Stable Distributions.* Providence, RI, USA: American Mathematical Society, 1986.

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

[4] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York, NY, USA: John Wiley and Sons, 1971, vol. 2.

[5] H. E. Daniels, "Saddlepoint approximations in statistics," *The Annals of Mathematical Statistics*, vol. 25, no. 4, pp. 631–650, Dec. 1954.

[6] ——, "Exact saddlepoint approximations," *Biometrika*, vol. 67, no. 1, pp. 59–63, 1980. [Online]. Available: https://doi.org/10.1093/biomet/67.1.59

[7] ——, "Tail probability approximations," *International Statistical Review / Revue Internationale de Statistique*, vol. 55, no. 1, pp. 37–48, Apr. 1987.

[8] N. Temme, "The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions," *SIAM J. Math. Anal.*, vol. 13, no. 2, pp. 239–253, 1982.

[9] R. Lugannani and S. Rice, "Saddle point approximation for the distribution of the sum of independent random variables," *Advances in Applied Probability*, vol. 12, no. 2, pp. 475–490, Jun. 1980.

[10] F. Esscher, "On the probability function in the collective theory of risk," *Skand. Aktuarie Tidskr.*, vol. 15, pp. 175–195, Mar. 1932.

[11] I. Shevtsova, "On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands," *ArXiv e-prints*, Nov. 2011.

[12] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[13] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for Gaussian multi-access channels via a central limit theorem for functions," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6719–6733, Dec. 2015.

[14] E. MolavianJazi, "A unified approach to Gaussian channels with finite blocklength," Ph.D. dissertation, University of Notre Dame, Jul. 2014.

[15] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. G. i Fàbregas, and A. Lancho, "Saddlepoint 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.

[16] 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.

[17] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, Nov. 2010.

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