An Upper Bound on the Error Induced by Saddlepoint Approximations: Applications to Wireless Communications

Samir M. Perlaza

Joint work with: Dadja Anade, Jean-Marie Gorce, and Philippe Mary

Workshop on **"Performance Guarantees in Wireless Networks"**

March 6, 2023

Motivation

Source $\mathcal{W} = \{1, 2, \ldots, M\}$

$i \in \mathcal{W}$

Transmitter

$\boldsymbol{u}(i) = (u_1(i), u_2(i), \ldots, u_n(i))^{\mathsf{T}} \in \mathcal{X}^n$

Channel: $P_{\boldsymbol{Y}|\boldsymbol{X}}$

$\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^{\mathsf{T}} \in \mathcal{Y}^n$

Receiver

$\tilde{i} \in \mathcal{W}$

- **Memoryless Channel**: $(\mathcal{X}^n, \mathcal{Y}^n, P_{\boldsymbol{Y}|\boldsymbol{X}})$
  $\forall\ \boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ and $\forall\ \boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$,

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

- **An (n,M,λ)-code**:

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

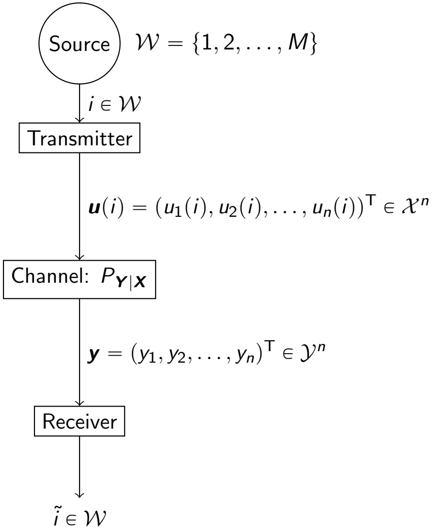
where for all $(j, \ell) \in \mathcal{W}^2$, with $j \neq \ell$:

$$\mathcal{D}(j) \cap \mathcal{D}(\ell) = \varnothing,$$

$$\bigcup_{j \in \mathcal{W}} \mathcal{D}(j) = \mathcal{Y}^n, \text{ and}$$

$$\underbrace{\frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}(i)}} \left[\mathbb{1}_{\{Y \notin \mathcal{D}(i)\}} \right]}_{\textbf{Average Decoding Error Probability}} \leqslant \lambda.$$

Motivation



- **An** (n,M,λ)**-code**:

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

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

$$\mathcal{D}(j) \cap \mathcal{D}(\ell) = \varnothing,$$

$$\bigcup_{j \in \mathcal{W}} \mathcal{D}(j) = \mathcal{Y}^n, \text{ and}$$

$$\underbrace{\frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{u}(i)}} \left[\mathbb{1}_{\{\boldsymbol{Y} \notin \mathcal{D}(i)\}} \right]}_{\textbf{Average Decoding Error Probability}} \leqslant \lambda.$$

- **Minimum** Average Decoding Error Probability

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

Motivation

Lemma 1 (DT bound [**Polyanskiy2010**], MC Bound [**FontSegura2018**])

Given a pair $(n, M) \in \mathbb{N}^2$*, the following holds for all probability measures* $Q_{\boldsymbol{Y}}$ *on the measurable space* $(\mathcal{Y}^n, \mathscr{B}(\mathcal{Y}^n))$*:*

$$\inf_{P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)} \max_{\gamma \geqslant 0} \left(T(n, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma) - \frac{\gamma}{M} \right) \leqslant \lambda^*(n, M) \leqslant \inf_{P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)} T\left(n, P_{\boldsymbol{X}}, P_{\boldsymbol{Y}}, \frac{M-1}{2}\right),$$

where

$$T(n, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma) \triangleq \mathbb{E}_{P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X}; \boldsymbol{Y}|Q_{\boldsymbol{Y}}) \leqslant \ln(\gamma)\}}\right] + \gamma \mathbb{E}_{P_{\boldsymbol{X}} Q_{\boldsymbol{Y}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X}; \boldsymbol{Y}|Q_{\boldsymbol{Y}}) > \ln(\gamma)\}}\right], \text{ and}$$

$$\tilde{\imath}(\boldsymbol{x}; \boldsymbol{y}|Q_{\boldsymbol{Y}}) \triangleq \ln\left(\frac{\mathrm{d}P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}}{\mathrm{d}Q_{\boldsymbol{Y}}}(\boldsymbol{y})\right).$$

Memoryless and Stationary Assumptions: Sum of IID Random Variables

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

Objective

Summary of the state of the art

- Bounds on decoding error probabilities are difficult to evaluate
- CDF of sums of IID random vectors
- Bounds provided by Berry-Esseen theorem are too loose.
- Saddlepoint approximations are good but **unknown bounds on the error**.

Objective

- Provide an appoximation to the CDF of sums of random vectors
- Characterize the approximation error

Approximation of CDFs of Sums of Random Vectors

- Consider n **independent** random vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$
- For all $i \in \{1, 2, \ldots, n\}$, $\boldsymbol{Y}_i \sim P_{\boldsymbol{Y}} \in \Delta\left(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k)\right)$.
- The **cumulant generating function** associated to the measure $P_{\boldsymbol{Y}}$ is $K_{\boldsymbol{Y}}$
- $$\boldsymbol{X}_n \triangleq \sum_{t=1}^{n} \boldsymbol{Y}_t \qquad \sim P_{\boldsymbol{X}_n}.$$
- The **cumulative distribution function** (CDF) of $\boldsymbol{X}_n$ is $F_{\boldsymbol{X}_n}$.
- The **approximation** of $F_{\boldsymbol{X}_n}$ is denoted by $\hat{F}_{\boldsymbol{X}_n}$.

Problem:

How to determine an upper bound on the absolute difference $\left| F_{\boldsymbol{X}_n} - \hat{F}_{\boldsymbol{X}_n} \right|$?

Preliminary Results - Esscher (exponential) Tilting

For all $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$, with

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

let $\boldsymbol{Y}_1^{(\boldsymbol{\theta})}$, $\boldsymbol{Y}_2^{(\boldsymbol{\theta})}$, ..., $\boldsymbol{Y}_n^{(\boldsymbol{\theta})}$ be independent random vectors with probability measure $P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}$ that satisfies for all $\boldsymbol{y} \in \mathbb{R}^k$,

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

Definition 2 (Esscher Tilting)

The probability measure $P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}$ is **Esscher tilted** with respect to $P_{\boldsymbol{Y}}$.

Preliminary Results - Esscher (exponential) Tilting

Given $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$, let the probability measure $P_{\boldsymbol{Y}(\boldsymbol{\theta})}$ be the solution to:

$$\min_{P \in \Delta(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))} \int \boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{y} \, \mathrm{d}P(\boldsymbol{y}) + D\left(P \| P_{\boldsymbol{Y}}\right).$$

Then, for all $\boldsymbol{t} \in \mathbb{R}^k$,

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

Samir M. Perlaza, Gaetan Bisson, Iñaki Esnaola, Alain Jean-Marie, Stefano Rini, **"Empirical Risk Minimization with Relative Entropy Regularizations"**. Research Report, INRIA, No. RR-9454, Sophia Antipolis, France, Feb., 2022.

Samir M. Perlaza, Iñaki Esnaola and H. Vincent Poor. "**Sensitivity of the Gibbs Algorithm to Data Aggregation in Supervised Machine Learning**". Research Report, INRIA, No. RR-9474, Sophia Antipolis, France, Jun., 2022.

Preliminary Results - Change of Measure

For all $\mathcal{A} \in \mathscr{B}(\mathbb{R}^k)$ and for all $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$,

$$
\begin{aligned}
P_{\boldsymbol{X}_n}(\mathcal{A}) =& \mathbb{E}_{P_{\boldsymbol{Y}_1 \boldsymbol{Y}_2 \ldots \boldsymbol{Y}_n}}\left[\mathbb{1}_{\left\{\sum_{j=1}^{n} \boldsymbol{Y}_j \in \mathcal{A}\right\}}\right]\\
=& \mathbb{E}_{P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})} \boldsymbol{Y}_2^{(\boldsymbol{\theta})} \ldots \boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}\left[\mathbb{1}_{\left\{\sum_{j=1}^{n} \boldsymbol{Y}_j^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}}\left(\prod_{j=1}^{n} \frac{\mathrm{d}P_{\boldsymbol{Y}^{(\boldsymbol{\theta})}}}{\mathrm{d}P_{\boldsymbol{Y}}}(\boldsymbol{Y}_j^{(\boldsymbol{\theta})})\right)^{-1}\right]\\
=& \mathbb{E}_{P_{\boldsymbol{Y}_1^{(\boldsymbol{\theta})} \boldsymbol{Y}_2^{(\boldsymbol{\theta})} \ldots \boldsymbol{Y}_n^{(\boldsymbol{\theta})}}}\left[\mathbb{1}_{\left\{\sum_{j=1}^{n} \boldsymbol{Y}_j^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}} \exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}} \sum_{j=1}^{n} \boldsymbol{Y}_j^{(\boldsymbol{\theta})}\right)\right]\\
=& \mathbb{E}_{P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}}\left[\exp\left(nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{S}_n^{(\boldsymbol{\theta})}\right)\mathbb{1}_{\left\{\boldsymbol{S}_n^{(\boldsymbol{\theta})} \in \mathcal{A}\right\}}\right],
\end{aligned}
$$

where

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

with probability measure $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$.

Preliminary Results - Gaussian Approximations

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

Warning:

Finding the measure $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$ is as difficult as finding $P_{\boldsymbol{X}_n}$.

Solution:

Approximate $\boldsymbol{S}_n^{(\boldsymbol{\theta})}$ by a **Gaussian random vector:**

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

where $P_{\boldsymbol{Z}_n^{(\boldsymbol{\theta})}}$ is the Gaussian approximation of $P_{\boldsymbol{S}_n^{(\boldsymbol{\theta})}}$.

$\eta_{\boldsymbol{Y}}(\boldsymbol{\theta},\mathcal{A},n)$ is the **exponentially-tilted Gaussian Approximation** of $P_{\boldsymbol{X}_n}(\mathcal{A})$.

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

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

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

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

The functions $K_{\boldsymbol{Y}}^{(1)}$ and $K_{\boldsymbol{Y}}^{(2)}$ are respectively the **gradient** and the **Hessian** of the CGF $K_{\boldsymbol{Y}}$.

Preliminary Results - Approximation Error

Theorem 3

For all convex sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^k)$*, and for all* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta_{\boldsymbol{Y}}$*,*

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

where the vector $\boldsymbol{a}\left(\mathcal{A}, \boldsymbol{\theta}\right) = \left(a_1(\mathcal{A}, \boldsymbol{\theta}),\ a_2(\mathcal{A}, \boldsymbol{\theta}),\ \ldots,\ a_k(\mathcal{A}, \boldsymbol{\theta})\right)$ *is such that for all* $i \in \{1, 2, \ldots, k\}$*,*

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

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

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

Dadja Anade, Jean-Marie Gorce, Philippe Mary, and Samir M. Perlaza, "Saddlepoint Approximation of Cumulative Distribution Functions of Sums of Random Vectors", in Proc. of the IEEE International Symposium on Information Theory (ISIT), 2021

Main Results (Approximation of the CDF)

For all $\boldsymbol{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, $F_{\boldsymbol{X}_n}(\boldsymbol{x}) = P_{\boldsymbol{X}_n}(\mathcal{A}_{\boldsymbol{x}})$, where
$\mathcal{A}_{\boldsymbol{x}} = \left\{(t_1, t_2, \ldots, t_k) \in \mathbb{R}^k : \forall i \in \{1, 2, \ldots, k\}, t_i \leqslant x_i\right\}$.

Observation

For all $\boldsymbol{x} \in \mathbb{R}^k$ and for all $\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}$, it holds that

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

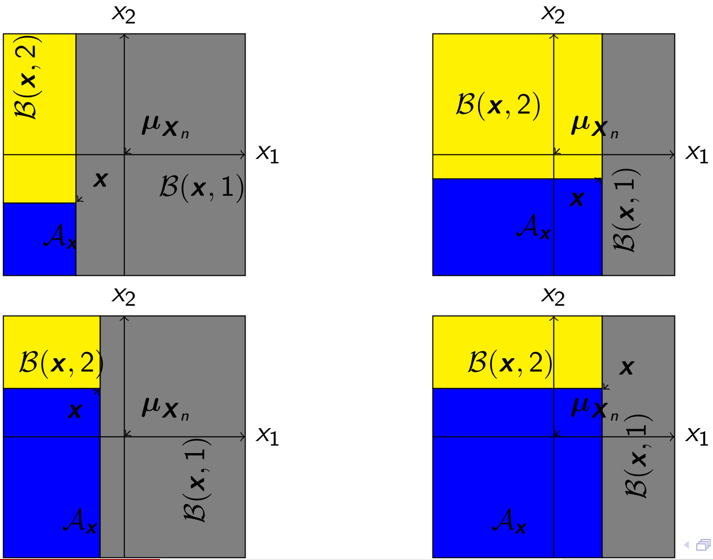
Minimize the exponential part of the upper bound (**chosen choice** denoted by $\boldsymbol{\theta}(\boldsymbol{x})$):

$$\min_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{Y}}} nK_{\boldsymbol{Y}}(\boldsymbol{\theta}) - \boldsymbol{\theta}^{\mathsf{T}}\, \boldsymbol{a}(\mathcal{A}_{\boldsymbol{x}}, \boldsymbol{\theta})$$

Warning

When $\boldsymbol{x} \in \mathcal{E}_{\boldsymbol{X}_n}$, then $\boldsymbol{\theta}(\boldsymbol{x}) = 0$ (Gaussian Approximation), where
$\mathcal{E}_{\boldsymbol{X}_n} \triangleq \left\{(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k : \forall i \in \{1, 2, \ldots, k\}, x_i > \mu_{X_{n,i}}\right\}$ and
$\boldsymbol{\mu}_{\boldsymbol{X}_n} = \left(\mu_{X_{n,1}}, \mu_{X_{n,2}}, \ldots, \mu_{X_{n,k}}\right)^{\mathsf{T}}$ is the mean of $\boldsymbol{X}_n$.

Main Results (Intuition behind $\mathcal{E}_{\boldsymbol{x}_n}$)



Main Results

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

$$\zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) \triangleq \begin{cases} \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}(\boldsymbol{x}), \mathcal{A}_{\boldsymbol{x}}, n) & \text{if } \boldsymbol{x} \notin \mathcal{E}_{\boldsymbol{X}_n} \\ 1 - \sum_{i=1}^{k} \eta_{\boldsymbol{Y}}(\boldsymbol{\theta}_i(\boldsymbol{x}), \mathcal{B}(\boldsymbol{x}, i), n) & \text{if } \boldsymbol{x} \in \mathcal{E}_{\boldsymbol{X}_n}, \end{cases}$$

and

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

with $\mathcal{B}(\boldsymbol{x}, i) = \left\{\boldsymbol{t} = (t_1, t_2, \ldots, t_k) \in \mathbb{R}^k : \forall j \in \{1, 2, \ldots, k\},\ t_j \leqslant x_j \text{ if } j < i, \text{ and } t_i > x_i\right\}$.

Theorem 4

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

$$\left|F_{\boldsymbol{X}_n}(\boldsymbol{x}) - \zeta_{\boldsymbol{Y}}(n, \boldsymbol{x})\right| \leqslant \delta_{\boldsymbol{Y}}(n, \boldsymbol{x}).$$

Main Results (Summary)

Upper and Lower Bounds on $F_{\boldsymbol{X}_n}$

For all $\boldsymbol{x} \in \mathbb{R}^k$,

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

where,

$$\bar{\Omega}(n, \boldsymbol{x}) \triangleq \zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) + \delta_{\boldsymbol{Y}}(n, \boldsymbol{x}), \text{ and}$$
$$\underline{\Omega}(n, \boldsymbol{x}) \triangleq \zeta_{\boldsymbol{Y}}(n, \boldsymbol{x}) - \delta_{\boldsymbol{Y}}(n, \boldsymbol{x}).$$

Connections to Saddlepoint Approximations

For all $\boldsymbol{x} \in \mathcal{D} \triangleq \left\{\boldsymbol{u} \in \mathbb{R}^k : \exists \boldsymbol{t} \in]-\infty, 0[^k,\ nK_{\boldsymbol{Y}}^{(1)}(\boldsymbol{t}) = \boldsymbol{u}\right\}$, it holds that $\zeta_{\boldsymbol{Y}}(n, \boldsymbol{x})$ is the **saddlepoint approximation** to $F_{\boldsymbol{X}_n}(\boldsymbol{x})$.

Approximation Error (Scalar)

Corollary 5

Consider the interval $\mathcal{A} = (e, b)$*, with* $e < b$*. Then, for all* $\theta \in \Theta_Y$*,*

$$|P_{X_n}(\mathcal{A}) - \eta_Y(\theta, \mathcal{A}, n)| \leqslant \exp\left(nK_Y(\theta) - \theta\, a\left(\mathcal{A}, \theta\right)\right) \min\left(1, \frac{c\, \xi_Y(\theta)}{\sqrt{n}}\right),$$

where c *is a constant and*

$$a\left(\mathcal{A}, \theta\right) \triangleq \begin{cases} 0 & \text{if } \theta = 0 \\ e & \text{if } \theta > 0 \\ b & \text{if } \theta < 0. \end{cases}$$

For all $x \in \mathbb{R}$, $F_{X_n}(x) = P_{X_n}\left(\mathcal{A}_x\right)$, where $\mathcal{A}_x = (-\infty, x]$:

- if $\theta \leqslant 0$, then $\theta\, a\left(\mathcal{A}_x, \theta\right) = \theta\, x$ and $\theta\, a\left(\mathcal{A}_x^{\mathsf{c}}, \theta\right) = -\infty$
- if $\theta > 0$, then $\theta\, a\left(\mathcal{A}_x, \theta\right) = -\infty$ and $\theta\, a\left(\mathcal{A}_x^{\mathsf{c}}, \theta\right) = \theta\, x$

How to choose θ ? Let $\theta(x) \in \arg\min_{\theta \in \Theta_Y} nK_Y(\theta) - \theta x$. Then,

$$(x - \mu_{X_n})\theta(x) \geqslant 0.$$

Approximation Error (Scalar)

Let the function $\zeta_Y : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ be such that for all $(n, x) \in \mathbb{N} \times \mathbb{R}$,

$$\zeta_Y(n,x) \triangleq \begin{cases} \eta_Y(\theta(x), \mathcal{A}_x, n) & \text{if } x \leqslant \mu_{X_n} \\ 1 - \eta_Y(\theta(x), \mathcal{A}_x^{\mathsf{c}}, n) & \text{else.} \end{cases}$$

Theorem 6

For all $x \in \mathbb{R}$, it holds that

$$|F_{X_n}(x) - \zeta_Y(n,x)| \leqslant \exp\left(nK_Y(\theta(x)) - \theta(x)\, x\right) \min\left(1, \frac{c(1)\, \xi_Y(\theta(x))}{\sqrt{n}}\right).$$

Upper and Lower Bounds on F_{X_n}

For all $x \in \mathbb{R}$,

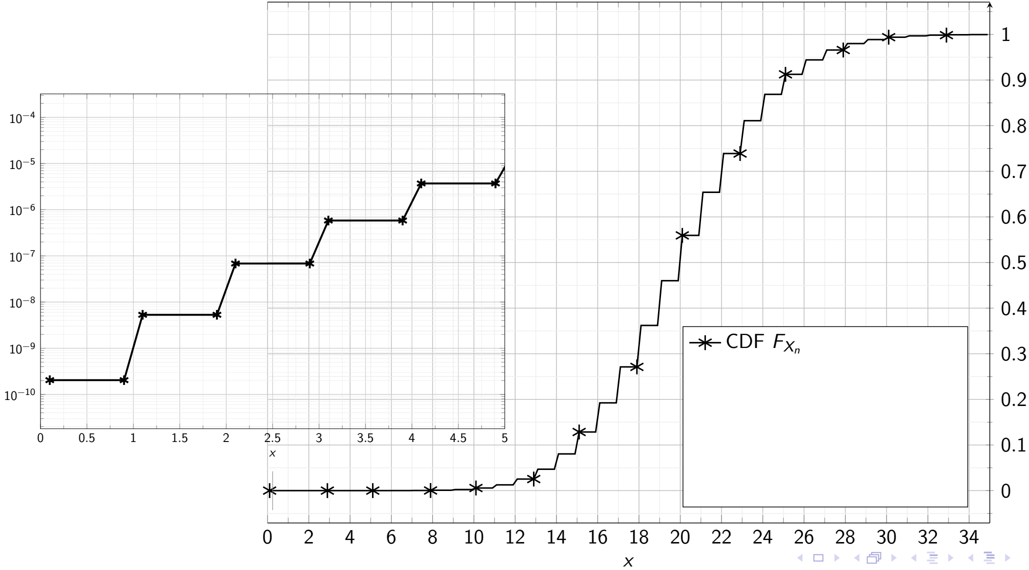
$$\underline{\Omega}(n,x) \leqslant F_{X_n}(x) \leqslant \bar{\Omega}(n,x),$$

where

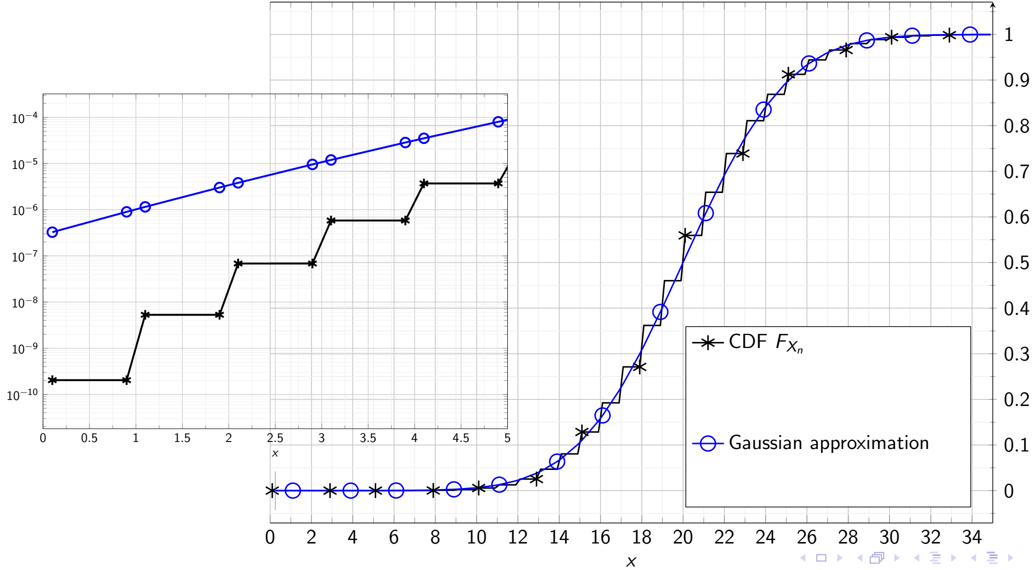
$$\bar{\Omega}(n,x) \triangleq \zeta_Y(n,x) + \exp\left(nK_Y(\theta(x)) - \theta(x)\, x\right) \min\left(1, \frac{c(1)\, \xi_Y(\theta(x))}{\sqrt{n}}\right),$$

$$\underline{\Omega}(n,x) \triangleq \zeta_Y(n,x) - \exp\left(nK_Y(\theta(x)) - \theta(x)\, x\right) \min\left(1, \frac{c(1)\, \xi_Y(\theta(x))}{\sqrt{n}}\right).$$

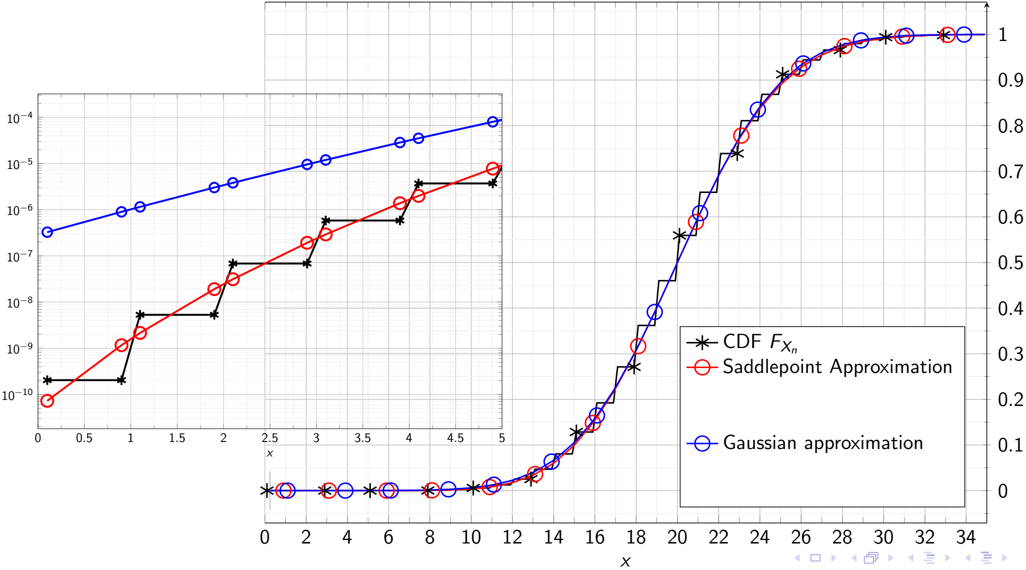
Examples: Sum of 100 Bernoulli random variables with $p = 0.2$.



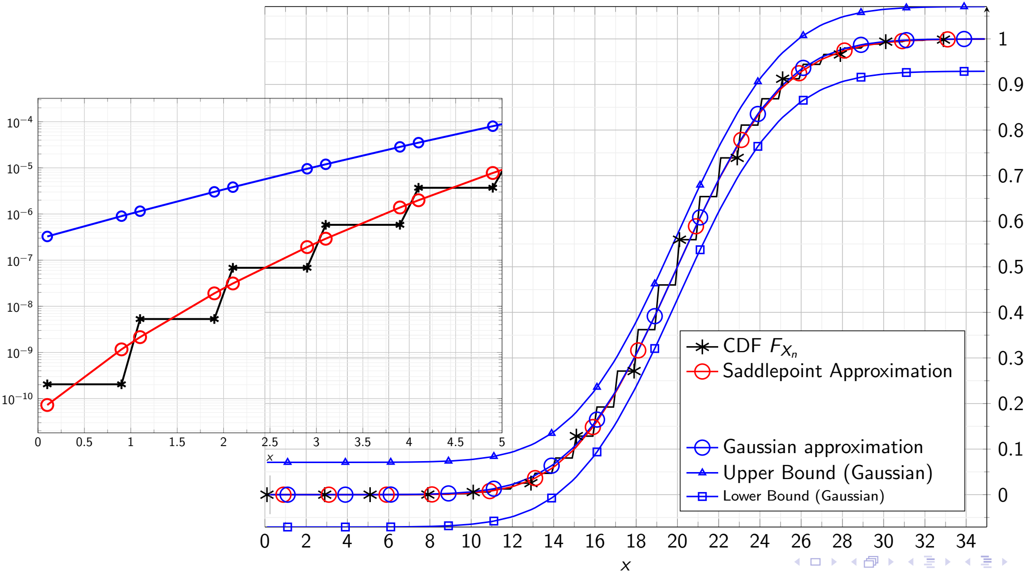
Examples: Sum of 100 Bernoulli random variables with $p = 0.2$.



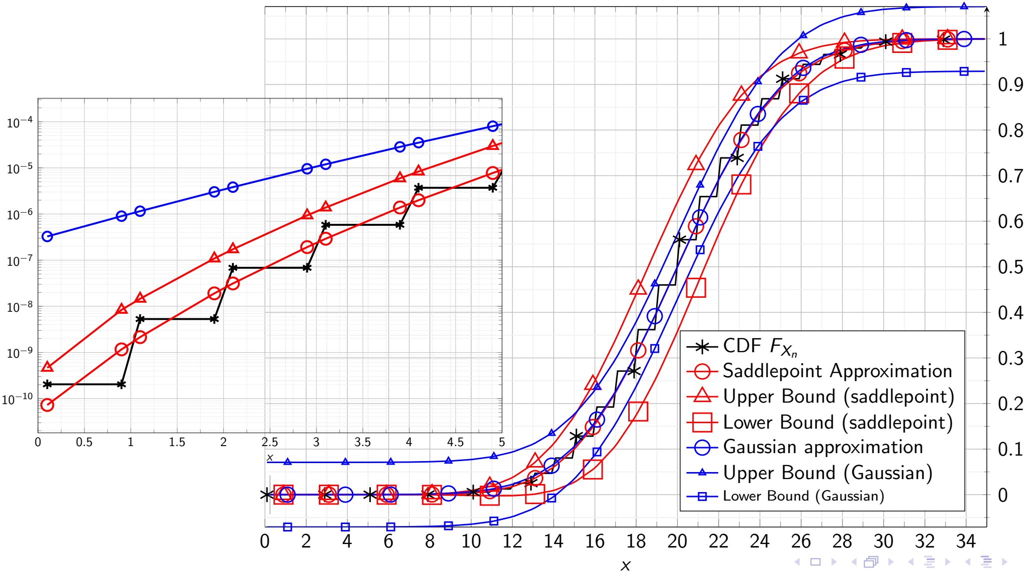
Examples: Sum of 100 Bernoulli random variables with $p = 0.2$.



Examples: Sum of 100 Bernoulli random variables with $p = 0.2$.



Examples: Sum of 100 Bernoulli random variables with $p = 0.2$.

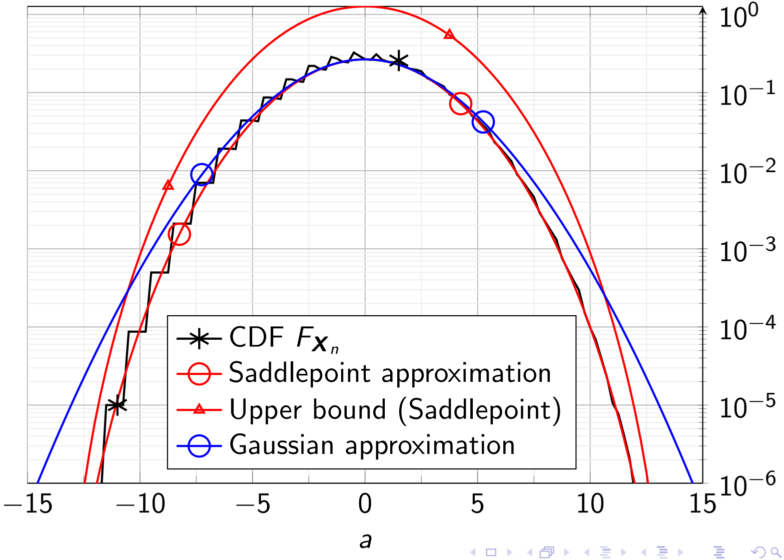


Examples: Random Vectors

- For all $i \in \{1, 2, \ldots, n\}$,

$$\boldsymbol{Y}_i \triangleq \begin{pmatrix} 1 & 0 \\ 0.1 & \sqrt{0.99} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix},$$

- B_1 and B_2 are independent **Bernoulli random variables** with parameter $p = 0.25$.
- Evaluation of CDF $F_{\boldsymbol{X}_n}$ at $\boldsymbol{x} = \boldsymbol{\mu}_{\boldsymbol{X}_n} + a\boldsymbol{d}$, with $\boldsymbol{d} = (1, -1)^{\mathsf{T}}$



Contribution Summary on Approximations of CDF

- upper and lower bounds for the CDF of random vectors/variables using Esscher tilting and Gaussian approximation.
- these bounds include the bounds for Gaussian approximation and saddlepoint approximation for specific values of theta.

Dadja Anade, Jean-Marie Gorce, Philippe. Mary, and Samir M. Perlaza, "An Upper Bound on the Error Induced by Saddlepoint Approximations - Applications to Information Theory", Entropy, vol. 22, num. 6, pp. 690.

Dadja Anade, Jean-Marie Gorce, Philippe. Mary, and Samir M. Perlaza, "Saddlepoint Approximation of Cumulative Distribution Functions of Sums of Random Vectors", in Proc. of the IEEE International Symposium on Information Theory (ISIT), 2021.

Point to Point Channels: Bounds

Lemma 7 (DT bound [**Polyanskiy2010**], MC Bound [**FontSegura2018**])

Given a pair $(n,M) \in \mathbb{N}^2$, the following holds for all probability measures $Q_{\boldsymbol{Y}}$ on the measurable space $(\mathcal{Y}^n, \mathscr{B}(\mathcal{Y}^n))$:

$$\inf_{P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)} \max_{\gamma \geqslant 0} \left(T(n, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma) - \frac{\gamma}{M} \right) \leqslant \lambda^*(n, M) \leqslant \inf_{P_{\boldsymbol{X}} \in \triangle(\mathcal{X}^n)} T\left(n, P_{\boldsymbol{X}}, P_{\boldsymbol{Y}}, \frac{M-1}{2}\right),$$

where

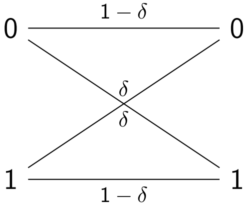
$$T(n, P_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}, \gamma) \triangleq \mathbb{E}_{P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}}) \leqslant \ln(\gamma)\}}\right] + \gamma \mathbb{E}_{P_{\boldsymbol{X}} Q_{\boldsymbol{Y}}}\left[\mathbb{1}_{\{\tilde{\imath}(\boldsymbol{X};\boldsymbol{Y}|Q_{\boldsymbol{Y}}) > \ln(\gamma)\}}\right], \text{ and}$$

$$\tilde{\imath}(\boldsymbol{x}; \boldsymbol{y}|Q_{\boldsymbol{Y}}) \triangleq \ln\left(\frac{\mathrm{d}P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}}{\mathrm{d}Q_{\boldsymbol{Y}}}(\boldsymbol{y})\right).$$

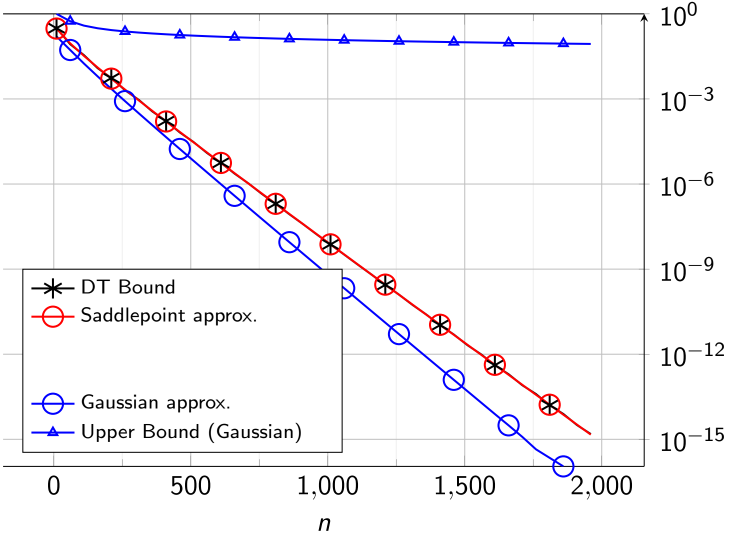
Memoryless and Stationary Assumptions: Sum of IID Random Vectors

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

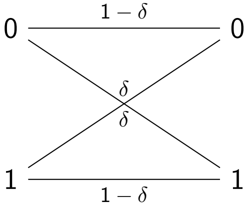
Binary Symmetric Channel: DT Bound



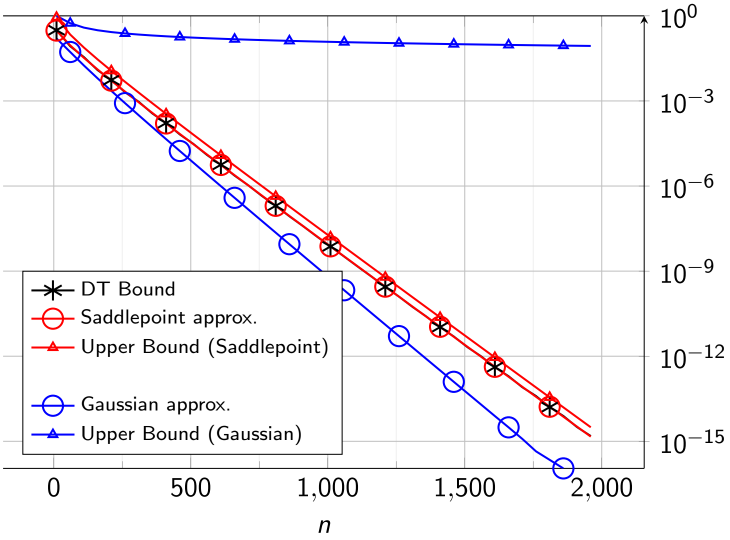
- Cross-over Probability $\delta = 0.11$,
- Information rate $R = 0.32$.



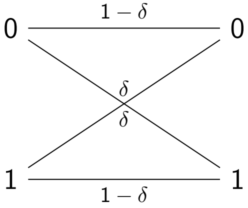
Binary Symmetric Channel: DT Bound



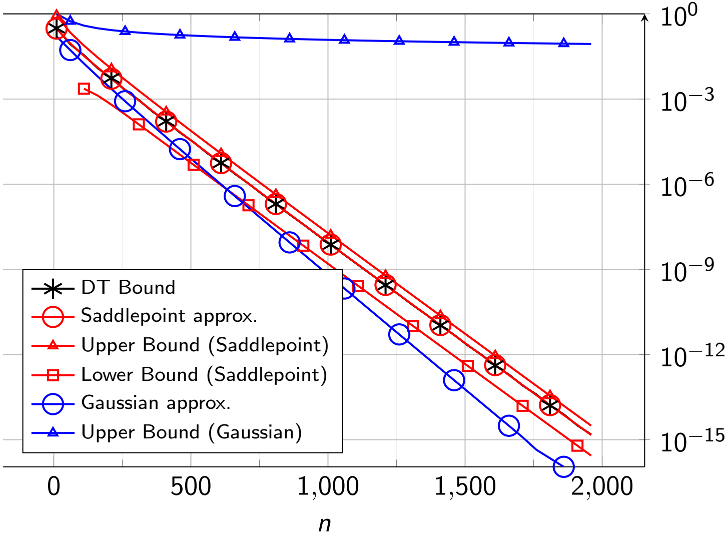
- Cross-over Probability $\delta = 0.11$,
- Information rate $R = 0.32$.



Binary Symmetric Channel: DT Bound



- Cross-over Probability $\delta = 0.11$,
- Information rate $R = 0.32$.



symmetric α-stable noise channel: MC Bound

For all $i \in \{1, 2, \ldots, n\}$,

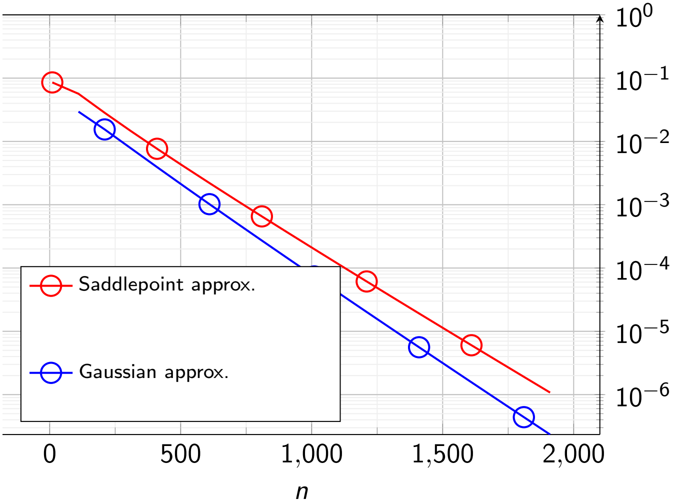
$$Y_i = X_i + Z_i$$

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

$$\mathbb{E}_{P_{Z_i}}\left[\exp\left(jtZ_i\right)\right] = \exp\left(-\left|\sigma t\right|^{\alpha}\right),$$

with $j = \sqrt{-1}$,

- Shape parameter: $\alpha = 1.4$,
- Dispersion parameter: $\sigma = 0.6$,
- Inputs: $\mathcal{X} = \{-1, 1\}$,
- Information rate: $R = 0.38$.



Laurent Clavier, Troels Pedersen, Ignacio Rodriguez, Mads Lauridsen, Malcolm Egan, "Experimental Evidence for Heavy Tailed Interference in the IoT", statistical models ; IoT ; Interference ; subexponential distributions ; heavy tails, Mar. 2020.

For all $i \in \{1, 2, \ldots, n\}$,

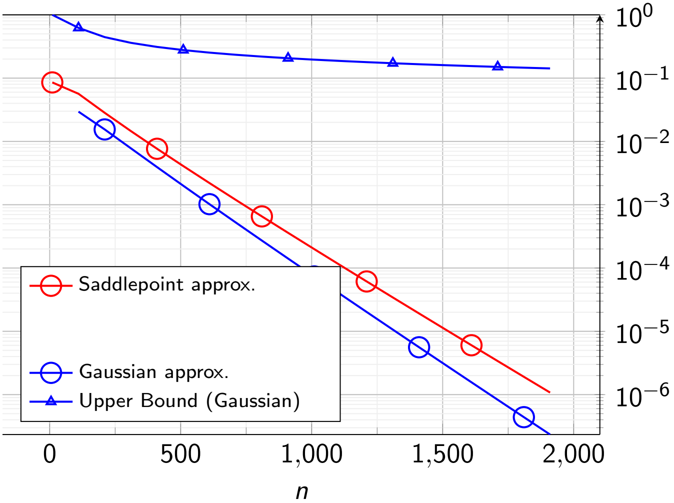
$$Y_i = X_i + Z_i$$

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

$$\mathbb{E}_{P_{Z_i}}\left[\exp\left(jtZ_i\right)\right] = \exp\left(-\left|\sigma t\right|^{\alpha}\right),$$

with $j = \sqrt{-1}$,

- Shape parameter: $\alpha = 1.4$,
- Dispersion parameter: $\sigma = 0.6$,
- Inputs: $\mathcal{X} = \{-1, 1\}$,
- Information rate: $R = 0.38$.



Laurent Clavier, Troels Pedersen, Ignacio Rodriguez, Mads Lauridsen, Malcolm Egan, "Experimental Evidence for Heavy Tailed Interference in the IoT", statistical models ; IoT ; Interference ; subexponential distributions ; heavy tails, Mar. 2020.

symmetric α-stable noise channel: MC Bound

For all $i \in \{1, 2, \ldots, n\}$,

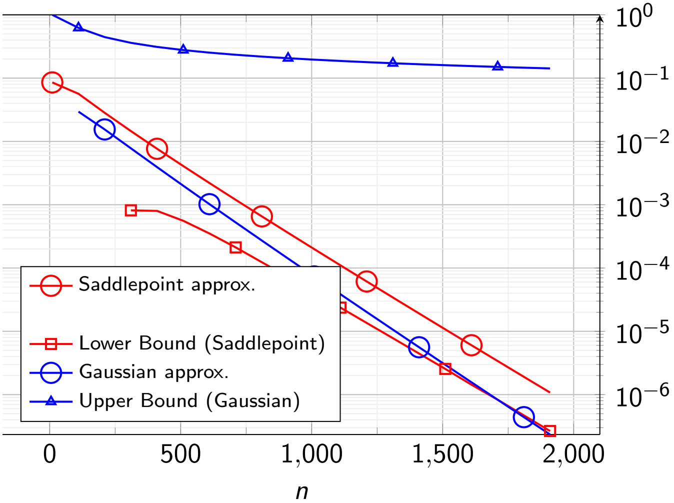
$$Y_i = X_i + Z_i$$

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

$$\mathbb{E}_{P_{Z_i}}\left[\exp\left(jtZ_i\right)\right] = \exp\left(-\left|\sigma t\right|^{\alpha}\right),$$

with $j = \sqrt{-1}$,

- Shape parameter: $\alpha = 1.4$,
- Dispersion parameter: $\sigma = 0.6$,
- Inputs: $\mathcal{X} = \{-1, 1\}$,
- Information rate: $R = 0.38$.



Laurent Clavier, Troels Pedersen, Ignacio Rodriguez, Mads Lauridsen, Malcolm Egan, "Experimental Evidence for Heavy Tailed Interference in the IoT", statistical models ; IoT ; Interference ; subexponential distributions ; heavy tails, Mar. 2020.

symmetric α-stable noise channel: MC Bound

For all $i \in \{1, 2, \ldots, n\}$,

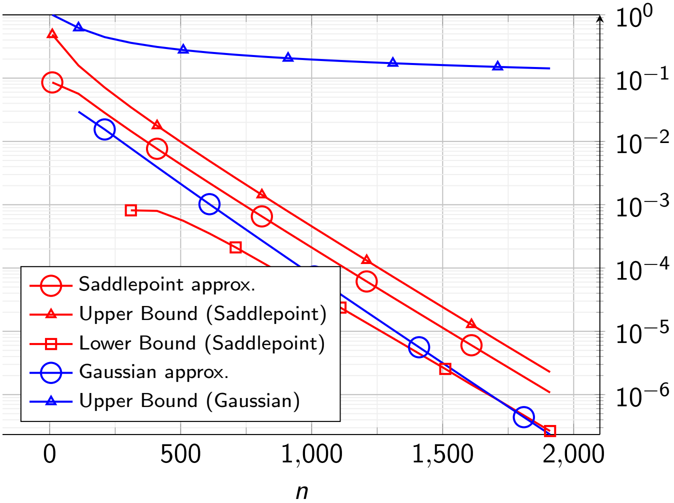
$$Y_i = X_i + Z_i$$

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

$$\mathbb{E}_{P_{Z_i}}\left[\exp\left(jtZ_i\right)\right] = \exp\left(-\left|\sigma t\right|^{\alpha}\right),$$

with $j = \sqrt{-1}$,

- Shape parameter: $\alpha = 1.4$,
- Dispersion parameter: $\sigma = 0.6$,
- Inputs: $\mathcal{X} = \{-1, 1\}$,
- Information rate: $R = 0.38$.



Laurent Clavier, Troels Pedersen, Ignacio Rodriguez, Mads Lauridsen, Malcolm Egan, "Experimental Evidence for Heavy Tailed Interference in the IoT", statistical models ; IoT ; Interference ; subexponential distributions ; heavy tails, Mar. 2020.

Conclusions and Further work

Contributions

- **Upper bounds on the error** of exponentially-tilted Gaussian approximations:
  - Choice of $\theta = 0$: Gaussian approximation
  - Choice of $\theta = \theta(x)$: Saddlepoint approximation
- **Applications:** Approximations of decoding error probability:
  - Point to point symmetric α-stable noise channels
  - Gaussian Multiple Access Channels

Thanks

This work was partially funded by the French National Agency for Research (ANR) under grant ANR-16-CE25-0001.