Binary and Multi-Bit Coding for Stable Random Projections

Ping Li
Department of Statistics and Biostatistics
Department of Computer Science
Rutgers University
Piscataway, NJ 08854, USA
pingli@stat.rutgers.edu

Abstract
Motivated by the recent work on “one scan 1-bit compressed sensing”, which was based on α-stable random projections with small α, we develop efficient coding schemes for storing the measurements and recovering the $l_\alpha$ norms of signals. Coding at the bit level is crucial especially when the measurements are heavy-tailed (due to stable random projections). Interestingly, using merely 1-bit information does not result in significant loss of accuracy if the parameter is chosen appropriately. For example, when $\alpha = 0+, 1,$ and 2, the coefficients of the optimal estimation variances using full (i.e., infinite-bit) information are 1, 2, and 2, respectively. With the 1-bit scheme and appropriately chosen parameters, the corresponding variance coefficients are 1.544, $\pi^2/4$, and 3.066, respectively. Using 2 or more bits reduces the estimation variances and, importantly, stabilizes the performance so that the variances are not sensitive to parameters. The computational cost of the estimation procedure is minimal if we implement the method using look-up tables and a small number (e.g., 1, 2, or 3) of bits. Note that the method presented in this paper is applicable to the general scale distribution family, not just α-stable distributions.

1 Introduction

The research problem of interest is about efficient estimation of the scale parameter of the $\alpha$-stable distribution using binary (i.e., 1-bit) and multi-bit coding of the samples. That is, given $n$ i.i.d. samples,

$$y_j \sim S(\alpha, \Lambda_\alpha), \quad j = 1, 2, \ldots, n$$

from an $\alpha$-stable distribution $S(\alpha, \Lambda_\alpha)$, we hope to estimate the scale parameter $\Lambda_\alpha$ by using only 1-bit or multi-bit information of $|y_j|$. Here we adopt the parameterization \cite{11,10} such that, if $y \sim S(\alpha, \Lambda_\alpha)$, then the characteristic function is $E\left(e^{\sqrt{-1}t|y|}\right) = e^{-\Lambda_\alpha|t|^\alpha}$. When $\alpha = 2$, $S(2, \Lambda_2)$ is equivalent to a Gaussian distribution $N(0, \sigma^2 = 2\Lambda_2)$. When $\alpha = 1$, $S(1, 1)$ is the standard Cauchy distribution.

The use of $\alpha$-stable distributions was studied in the context of estimating frequency moments of data streams \cite{4,5}. In this paper, the development of binary (1-bit) and multi-bit coding schemes is motivated by the recent work on “one scan 1-bit compressed sensing”.

1.1 One Scan 1-Bit Compressed Sensing

The problem of estimating $\Lambda_\alpha$ using 1-bit or multi-bit information is motivated by the recent work \cite{6} on “one scan 1-bit compressed sensing”, which is based on $\alpha$-stable random projections with small $\alpha$. Here we
provide a brief description. Given \( n \) measurements \( y_j = \sum_{i=1}^{N} x_i s_{ij}, j = 1 \) to \( n \), where \( s_{ij} \sim S(\alpha, 1) \) i.i.d. and \( x_i, i = 1 \) to \( N \), is a sparse (and possibly dynamic/streaming) vector, the task is to recover \( x \) from only the signs of the measurements, i.e., \( \text{sign}(y_j) \). The recent work \( [6] \) provided a recipe to recover \( x \) from \( \text{sign}(y_j) \) by scanning the coordinates of the vector only once. The highly efficient procedure, however, requires the knowledge of “\( K \)”, which is the \( l_n \) norm \( \sum_{i=1}^{N} |x_i|^\alpha \) as \( \alpha \to 0+ \). In practice, this \( K \) will typically have to estimated and it will not make sense if we have to use too many measurements just for estimating \( K \). \( [6] \) showed that \( 12.3K \log N \) 1-bit measurements are needed using this one-scan 1-bit procedure. It was also shown in \( [6] \) that only a very small number (e.g., 5 or 10) of full measurements are needed to estimate \( K \) for the task of sparse recovery. In this paper, our work will elaborate that only 1 bit or a few bits per measurement can provide accurate estimates of \( K \) (as well as the general term \( \sum_{i=1}^{N} |x_i|^\alpha \) for \( 0 < \alpha \leq 2 \)).

Because the samples \( y_j \) are heavy-tailed, the storage requirement for each sample can be substantial. This consequently causes many issues in data retrieval, transmission and decoding speed. It is highly desirable if we just need to store 1 bit or a few bits for each \( |y_j| \). This is the motivation of our paper.

### 1.2 Sampling from \( \alpha \)-Stable Distributions

Although in general there is no closed-form expression for the density of \( S(\alpha, 1) \), we can sample from the distribution using a standard procedure provided by \( [1] \). That is, one can first sample an exponential \( w \sim \exp(1) \) and a uninform \( u \sim \text{unif}(-\pi/2, \pi/2) \), and then compute

\[
\begin{align*}
    s_\alpha &= \frac{\sin(\alpha u)}{\cos u} \left[ \frac{\cos(u - \alpha u)}{w} \right]^{(1-\alpha)/\alpha} \sim S(\alpha, 1) \\
\end{align*}
\]

In this paper, we will heavily use the distribution of \( |s_\alpha|^{\alpha} \), which is

\[
|s_\alpha|^{\alpha} = \frac{|\sin(\alpha u)|^{\alpha}}{\cos u} \left[ \frac{\cos(u - \alpha u)}{w} \right]^{(1-\alpha)}
\]

Clearly, \( 1/|s_\alpha|^{\alpha} \) converges to \( \exp(1) \) in distribution as formally established by \( [3] \).

### 2 Estimation of \( \Lambda_\alpha \) Using Full (Infinite-Bit) Information

In this section, given \( n \) i.i.d. samples \( y_j \sim S(\alpha, \Lambda_\alpha), j = 1 \) to \( n \), we review various estimators of the scale parameter \( \Lambda_\alpha \) using full information (i.e., infinite-bit). When \( \alpha = 2 \) (i.e., Gaussian), we can just use the arithmetic mean estimator, which is statistically optimal, i.e., the (asymptotic) variance reaches the reciprocal of the Fisher Information. The arithmetic mean estimator is

\[
\hat{\Lambda}_2 = \frac{1}{n} \sum_{j=1}^{n} |y_j|^2, \quad \text{Var} \left( \hat{\Lambda}_2 \right) = \frac{\Lambda_\alpha^2}{n}
\]

The papers \( [7, 5] \) discussed several estimators for \( \alpha < 2 \). The harmonic mean estimator is suitable for small \( \alpha \) and becomes optimal as \( \alpha \to 0+ \):

\[
\hat{\Lambda}_{\alpha, hm} = \frac{-2 \Gamma(-\alpha) \sin \left( \frac{\pi \alpha}{2} \right)}{\sum_{j=1}^{n} |y_j|^{\alpha}} \left( n - \left( \frac{-\pi \Gamma(-2\alpha) \sin \left( \pi \alpha \right)}{\Gamma(-\alpha) \sin \left( \frac{\pi \alpha}{2} \right)^2} - 1 \right) \right)
\]

\[
\text{Var} \left( \hat{\Lambda}_{\alpha, hm} \right) = \frac{\Lambda_\alpha^2}{n} \left( \frac{-\pi \Gamma(-2\alpha) \sin \left( \pi \alpha \right)}{\Gamma(-\alpha) \sin \left( \frac{\pi \alpha}{2} \right)^2} - 1 \right) + O \left( \frac{1}{n^2} \right)
\]
When $\alpha \to 0+$, the variance becomes
\[
\text{Var} \left( \hat{\Lambda}_{0+,hm} \right) = \frac{\Lambda_{0+}^2}{n} + O \left( \frac{1}{n^2} \right) \tag{7}
\]

The geometric mean estimator can be used for any $0 < \alpha \leq 2$:
\[
\hat{\Lambda}_{\alpha, gm} = \prod_{j=1}^{n} |y_j|^\alpha /n \\
\text{Var} \left( \hat{\Lambda}_{\alpha, gm} \right) = \frac{\Lambda_{\alpha}^2}{n} \left\{ \frac{\pi^2}{12} (\alpha^2 + 2) \right\} + O \left( \frac{1}{k^2} \right) \tag{8}
\]

The geometric mean estimator is not statistically optimal (for any $\alpha$) and the variance becomes relatively large when $\alpha$ approaches $0+$ or $2$. Interestingly, the following fractional power estimator is close to be optimal for the entire $0 < \alpha \leq 2$.
\[
\hat{\Lambda}_{\alpha, fp} = \left( \frac{1}{n} \sum_{j=1}^{n} |y_j|^\lambda^\alpha \right)^{1/\lambda^*} \\
\text{Var} \left( \hat{\Lambda}_{\alpha, fp} \right) = \frac{\Lambda_{\alpha}^2}{n} \left\{ \frac{\pi^2}{12} (\alpha^2 + 2) \right\} + O \left( \frac{1}{k^2} \right) \tag{9}
\]

where
\[
\lambda^* = \text{argmin}_{\frac{1}{2} < \lambda < \frac{1}{2}} \frac{1}{\lambda^2} \left( \frac{\frac{2}{\pi} \Gamma(1 - 2\lambda) \Gamma(2\lambda) \sin \left( \frac{\pi\lambda}{2} \right)}{\left[ \frac{2}{\pi} \Gamma(1 - \lambda) \Gamma(\lambda) \sin \left( \frac{\pi\lambda}{2} \right) \right]^2} - 1 \right) \tag{10}
\]

The fractional power estimator becomes the harmonic mean estimator when $\alpha \to 0+$ and it becomes the arithmetic mean estimator when $\alpha = 2$. Thus, it is optimal at these two $\alpha$ values.

When $\alpha = 1$, the optimal estimator requires solving a nonlinear MLE equation with the asymptotic variance being $\frac{\Lambda_1^2}{n} + O \left( \frac{1}{n^2} \right)$.

In summary, the optimal variances for $\alpha = 0+, 1$, and 2, are respectively
\[
\frac{\Lambda_{0+}^2}{n}, \quad \frac{\Lambda_1^2}{n}, \quad \text{and} \quad \frac{\Lambda_2^2}{n} \tag{12}
\]

Our goal is to develop 1-bit and multi-bit schemes to achieve variances which are close to be optimal.

### 3 1-Bit Coding and Estimation

Again, consider $n$ i.i.d. samples $y_j \sim S(\alpha, \Lambda_{\alpha})$, $j = 1$ to $n$. In this section, the task is to estimate $\Lambda_{\alpha}$ using just one bit information of each $|y_j|$. This means that we will need a (pre-determined) threshold in order to obtain this 1-bit information. To accomplish the goal, we consider a (positive) threshold $C$ (which can be a function of $\alpha$) and compare it with $z_j$ where
\[
z_j = |y_j|^\alpha, \quad j = 1, 2, ..., n \tag{13}
\]

In other word, we store a “0” if $z_j \leq C$ and a “1” if $z_j > C$. For convenience, we express $z_j$ in terms of $\Lambda_{\alpha}$ and $s_{\alpha} \sim S(\alpha, 1)$, i.e.,
\[
z_j = |y_j|^\alpha \sim \Lambda_{\alpha} |s_\alpha|^\alpha, \quad s_{\alpha} \sim S(\alpha, 1), \tag{14}
\]
Let \( f_\alpha \) and \( F_\alpha \) be the pdf and cdf of \(|s_\alpha|^\alpha\), respectively. Then we can define \( p_1 \) and \( p_2 \) as follows
\[
p_1 = \Pr (z_\alpha \leq C) = F_\alpha (C/\Lambda_\alpha), \quad p_2 = \Pr (z_\alpha > C) = 1 - p_1 = 1 - F_\alpha (C/\Lambda_\alpha)
\]
(15)
which are needed for computing the likelihood. Denote
\[
n_1 = \sum_{j=1}^{n} 1 \{ z_j \leq C \}, \quad n_2 = \sum_{j=1}^{n} 1 \{ z_j > C \}
\]
(16)
The log-likelihood of the \( n = n_1 + n_2 \) observations can be expressed as
\[
l = n_1 \log p_1 + n_2 \log p_2 = n_1 \log F_\alpha (C/\Lambda_\alpha) + n_2 \log [1 - F_\alpha (C/\Lambda_\alpha)]
\]
(17)
To seek the MLE (maximum likelihood estimator) of \( \Lambda_\alpha \), we need to compute the first derivative:
\[
l' = \frac{\partial l}{\partial \Lambda_\alpha} = n_1 f_\alpha (C/\Lambda_\alpha) \left( -\frac{C}{\Lambda_\alpha^2} \right) + n_2 f_\alpha (C/\Lambda_\alpha) \left( \frac{C}{\Lambda_\alpha^2} \right) - \frac{f_\alpha (C/\Lambda_\alpha)}{1 - F_\alpha (C/\Lambda_\alpha)} \left( -\frac{C}{\Lambda_\alpha^2} \right)
\]
(18)
Setting \( l' = 0 \) yields the MLE solution denoted by \( \hat{\Lambda}_\alpha \):
\[
F_\alpha^{-1} (n_1/n) = C/\Lambda_\alpha \implies \hat{\Lambda}_\alpha = C/F_\alpha^{-1} (n_1/n)
\]
(19)
To assess the estimation variance of \( \hat{\Lambda}_\alpha \), we resort to the classical statistical theory of Fisher Information, which says
\[
\text{Var} (\hat{\Lambda}_\alpha) = \frac{1}{-E (l'')} + O \left( \frac{1}{n^2} \right)
\]
(20)
Because \( E(n_1) = np_1, E(n_2) = np_2 \), we have
\[
E (l'') = n \left( \frac{C}{\Lambda_\alpha^2} \right) \left[ f_\alpha \left( -\frac{C}{\Lambda_\alpha^2} \right) - f_\alpha^2 (\frac{C}{\Lambda_\alpha^2}) - f_\alpha \left( -\frac{C}{\Lambda_\alpha^2} \right) - \frac{f_\alpha^2}{1 - F_\alpha} \right]
\]
\[
= -n \left( \frac{C}{\Lambda_\alpha^2} \right)^2 \left[ \frac{f_\alpha^2}{F_\alpha} + \frac{f_\alpha^2}{1 - F_\alpha} \right]
\]
\[
= -n \frac{C^2 f_\alpha^2}{\Lambda_\alpha^2 F_\alpha (1 - F_\alpha)}
\]
(21)
Note that, because \( E(l') = 0 \), we do not have to worry about the other part of the second directive (which is basically \( -\frac{2}{\Lambda_\alpha^2} l' \)).

To simplify the expressions, we introduce the notation \( \eta \) such that
\[
\eta = \frac{\Lambda_\alpha}{C}
\]
(22)
which allows us to write
\[
\frac{1}{-E (l'')} = \frac{\Lambda_\alpha^2}{n} \frac{\Lambda_\alpha^2 F_\alpha (1 - F_\alpha)}{f_\alpha^2} = \frac{\Lambda_\alpha^2}{n} V_\alpha (\eta), \quad \text{where} \quad V_\alpha (\eta) = \eta^2 \frac{F_\alpha (1 - F_\alpha)}{f_\alpha^2}
\]
(23)
We summarize the above results in Theorem[1].
Theorem 1  Given $n$ i.i.d. samples $y_j \sim S(\alpha, \Lambda)$, $j = 1$ to $n$, a threshold $C$, and $n_1 = \sum_{j=1}^{n} 1\{z_j \leq C\}$, the maximum likelihood estimator (MLE) of $\Lambda$ is

$$\hat{\Lambda}_\alpha = C / F_{\alpha}^{-1}(n_1/n)$$

(24)

The asymptotic variance of $\hat{\Lambda}_\alpha$ is

$$\text{Var} \left( \hat{\Lambda}_\alpha \right) = \frac{\Lambda_\alpha^2}{n} V_\alpha (\eta) + O \left( \frac{1}{n^2} \right)$$

(25)

where $\eta = \frac{\Lambda_\alpha}{C}$ and

$$V_\alpha (\eta) = \eta^2 \frac{\alpha_2 \alpha (1/\eta)(1 - \alpha_\alpha (1/\eta))}{\alpha_2 (1/\eta)}.$$ 

(26)

Here $\alpha_\alpha$ and $F_\alpha$ are the pdf and the cdf of $|S(\alpha, 1)|^\alpha$, respectively.

Next, we explicitly and separately compute $\hat{\Lambda}_\alpha$ and the variance for $\alpha = 0^+, 1$, and 2.

3.1  $\alpha \to 0^+$

As $\alpha \to 0^+$, we have $1/|s_\alpha|^\alpha \sim \exp(1)$. Hence we have

$$F_{0^+}(z) = e^{-1/z}, \quad f_{0^+}(z) = \frac{1}{z^2} e^{-1/z}, \quad F_{0^+}^{-1}(z) = \frac{1}{\log 1/z}$$

(27)

Thus,

$$\hat{\Lambda}_{0^+} = C / F_{0^+}^{-1}(n_1/n) = C \log n / n_1, \quad \text{Var} \left( \hat{\Lambda}_{0^+} \right) = \frac{\Lambda_{0^+}^2}{n} V_{0^+} (\eta) + O \left( \frac{1}{n^2} \right)$$

(28)

where

$$V_{0^+} (\eta) = \eta^2 \frac{\alpha_{0^+} (1/\eta)(1 - \alpha_\alpha (1/\eta))}{\alpha_2 (1/\eta)} = \frac{e^{-\eta} - e^{-2\eta}}{\eta^2 e^{-2\eta}} = \frac{e_\eta - 1}{\eta^2}$$

(29)

The minimum $V_{0^+} (\eta)$ is 1.544, attained at $\eta = 1.594$. (In this paper, we always keep 3 decimal places.)

3.2  $\alpha = 1$

By properties of Cauchy distribution, we know

$$F_1(z) = \frac{2}{\pi} \tan^{-1} z, \quad f_1(z) = \frac{2}{\pi} \frac{1}{1 + z^2}, \quad F_1^{-1}(z) = \tan \frac{\pi}{2} z$$

(30)

Thus,

$$\hat{\Lambda}_1 = C / \tan \frac{\eta}{2} n, \quad \text{Var} \left( \hat{\Lambda}_1 \right) = \frac{\Lambda_1^2}{n} V_1 (\eta) + O \left( \frac{1}{n^2} \right)$$

(31)

The minimum of $V_1(\eta)$ is $\frac{\pi^2}{4}$, attained at $\eta = 1$. Here, we provide a straightforward proof. For convenience, let $t = 1/\eta$. Then

$$V_1 (\eta) = \frac{1}{t^2} \frac{F_1(t)(1 - F_1(t))}{f_1^2(t)}$$
\[
\frac{\partial \log V_1(\eta)}{\partial t} = -\frac{2}{t} + \frac{f_1(t)}{F_1(t)} + \frac{-f_1(t)}{1 - F_1(t)} - \frac{2f_1'(t)}{f_1(t)}
\]

\[
= -\frac{2}{t} + \frac{4t}{1 + t^2} + \frac{1}{1 + t^2} - \frac{1}{1 - \frac{1}{\pi} \tan^{-1} t}
\]

\[
= \frac{1}{1 + t^2} \left[ t^2 - 1 + \frac{1}{\tan^{-1} t} - \frac{1}{\frac{1}{2} - \tan^{-1} t} \right]
\]

Setting \(\frac{\partial \log V_1(\eta)}{\partial t} = 0\), the solution is \(t = 1\). Hence the optimum is attained at \(\eta = 1\).

### 3.3 \(\alpha = 2\)

Because \(S(2, 1) \sim \sqrt{2} \times N(0, 1)\), i.e., \(|s_\alpha|^2 \sim 2 \chi^2_1\), we have

\[
F_2(z) = F_{\chi^2_1}(z/2), \quad f_2(z) = f_{\chi^2_1}(z/2)/2,
\]

(32)

where \(F_{\chi^2_1}\) and \(f_{\chi^2_1}\) are the cdf and pdf of a chi-square distribution with 1 degree of freedom, respectively.

The MLE is \(\hat{\Lambda}_2 = C \frac{f_2^{-1}(n_1/n)}{F_2^{-1}(n_1/n)}\) and the optimal variance of \(\hat{\Lambda}_2\) can be numerically shown to be \(\frac{\Lambda_2^2}{n} 3.066\), attained at \(\eta = \frac{\Lambda_2}{C} = 0.228\).

### 3.4 General \(0 < \alpha \leq 2\)

For general \(0 < \alpha \leq 2\), the cdf \(F_\alpha\) and pdf \(f_\alpha\) can be computed numerically. Figure 1 plots the \(V_\alpha(\eta)\) for \(0 < \eta < 2.5\) and \(\alpha\) from 0 to 2 spaced at 0.1. The lowest point on each curve corresponds to the optimal (smallest) \(V_\alpha(\eta)\). Figure 2 plots the optimal \(V_\alpha\) values (left panel) and optimal \(\eta\) values (right panel).

![Figure 1](image1.png)

Figure 1: The variance factor \(V_\alpha(\eta)\) defined in (26) for \(\alpha \in [0, 2]\), spaced at 0.1. The lowest point on each curve (for one particular \(\alpha\)) corresponds to the optimal variance at that \(\alpha\).

Figure 1 suggests that the 1-bit scheme performs very well at least for \(\alpha \leq 1\). The optimal variance coefficient \(V_\alpha\) is not much larger than the variance using full information. For example, when \(\alpha = 1\), the optimal variance coefficient using full information is 2 (i.e., see (12)), while the optimal variance coefficient of the 1-bit scheme is just \(\frac{\pi^2}{4} = 2.467\) which is only about 20% larger. Furthermore, we can see that, when \(\alpha \leq 1\), \(V_\alpha(\eta)\) is not very sensitive to \(\eta\) in a wide range of \(\eta\) values, which is practically very important.
Figure 2: The optimal variance values $V_\alpha(\eta)$ (left panel) and the corresponding optimal $\eta$ values (right panel). Each point on the curve corresponds to the lowest point of the curve for that $\alpha$ as shown in Figure 1.

### 3.5 Error Tail Bounds

While the variance analysis largely suffices for practical purposes, it is also interesting to examine the error tail bounds, which are in a sense more precise than variances only. The following two bounds can be proved:

**Right tail bound:**  
$$ \Pr \left( \hat{\Lambda}_\alpha \geq (1 + \epsilon)\Lambda_\alpha \right) \leq \exp \left( -n \frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}} \right), \quad \epsilon \geq 0 \tag{33} $$

**Left tail bound:**  
$$ \Pr \left( \hat{\Lambda}_\alpha \leq (1 - \epsilon)\Lambda_\alpha \right) \leq \exp \left( -n \frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}} \right), \quad 0 \leq \epsilon \leq 1 \tag{34} $$

where $G_R$ and $G_L$ will be specified later. Once we have the tail bounds, we can say more precisely about the sample complexity, i.e., the number of samples needed to reach certain desired accuracy. Suppose we would like to ensure

$$ \Pr \left( \hat{\Lambda}_\alpha \geq (1 + \epsilon)\Lambda_\alpha \right) + \Pr \left( \hat{\Lambda}_\alpha \leq (1 - \epsilon)\Lambda_\alpha \right) \leq \delta, \quad 0 \leq \delta \leq 1 \tag{35} $$

It suffices to let

$$ \exp \left( -n \frac{\epsilon^2}{G_{R,\alpha,C,\epsilon}} \right) + \exp \left( -n \frac{\epsilon^2}{G_{L,\alpha,C,\epsilon}} \right) \leq \delta \tag{36} $$

for which it suffices

$$ n \geq \frac{G_{\alpha,C,\epsilon}}{\epsilon^2} \log 2/\delta, \quad \text{where} \quad G_{\alpha,C,\epsilon} = \max \{G_{R,\alpha,C,\epsilon}, G_{L,\alpha,C,\epsilon}\} \tag{37} $$

Obviously, it will be even more precise to numerically compute $n$ from (36) instead of using the convenient sample complexity bound (37). On the other hand, as $\epsilon \to 0$, both $G_{R,\alpha,C,\epsilon}$ and $G_{L,\alpha,C,\epsilon}$ converge to $2 \times V_\alpha$ (Recall $V_\alpha$ is the variance coefficient), which is the reason why the variance analysis is largely sufficient.

Next we provide the exact expressions for $G_{R,\alpha,C,\epsilon}$ and $G_{L,\alpha,C,\epsilon}$. The proof is simple, based on the expression of the MLE estimator $\hat{\Lambda}_\alpha = C/F_\alpha^{-1}(n_1/n)$, the fact that $n_1 \sim Binomial(n, F_\alpha(1/\eta))$, and Chernoff’s original tail bounds [2] for the binomial distribution.
For the right tail bound, we have

\[
\Pr \left( \hat{\Lambda}_\alpha \geq (1 + \varepsilon)\Lambda_\alpha \right) \\
= \Pr \left( \frac{C}{F_\alpha^{-1}(n_1/n)} \geq (1 + \varepsilon)\Lambda_\alpha \right) \\
= \Pr \left( \frac{n_1}{n} \leq F_\alpha \left( \frac{C}{(1 + \varepsilon)\Lambda_\alpha} \right) \right) \\
= \Pr \left( \frac{n_1}{n} \leq F_\alpha \left( \frac{1}{(1 + \varepsilon)\eta} \right) \right) \\
\leq \left[ \frac{F_\alpha(1/\eta)}{F_\alpha(1/(1 + \varepsilon)\eta)} \right]^{nF_\alpha(1/\eta)} \left[ \frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1 + \varepsilon)\eta)} \right]^{n-nF_\alpha(1/\eta)} \\
= \exp \left( -\frac{\varepsilon^2}{G_{R,\alpha,C,\varepsilon}} \right)
\]

where

\[
\frac{\varepsilon^2}{G_{R,\alpha,C,\varepsilon}} = -F_\alpha(1/(1 + \varepsilon)\eta) \log \left[ \frac{F_\alpha(1/\eta)}{F_\alpha(1/(1 + \varepsilon)\eta)} \right] - (1 - F_\alpha(1/(1 + \varepsilon)\eta)) \log \left[ \frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1 + \varepsilon)\eta)} \right]
\]

(38)

Next, for the left tail bound, we have

\[
\Pr \left( \hat{\Lambda}_\alpha \leq (1 - \varepsilon)\Lambda_\alpha \right) \\
= \Pr \left( \frac{C}{F_\alpha^{-1}(n_1/n)} \leq (1 - \varepsilon)\Lambda_\alpha \right) \\
= \Pr \left( \frac{n_1}{n} \geq F_\alpha \left( \frac{C}{(1 - \varepsilon)\Lambda_\alpha} \right) \right) \\
= \Pr \left( \frac{n_1}{n} \geq F_\alpha \left( \frac{1}{(1 - \varepsilon)\eta} \right) \right) \\
\leq \left[ \frac{F_\alpha(1/\eta)}{F_\alpha(1/(1 - \varepsilon)\eta)} \right]^{nF_\alpha(1/\eta)} \left[ \frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1 - \varepsilon)\eta)} \right]^{n-nF_\alpha(1/\eta)} \\
= \exp \left( -\frac{\varepsilon^2}{G_{L,\alpha,C,\varepsilon}} \right)
\]

where

\[
\frac{\varepsilon^2}{G_{L,\alpha,C,\varepsilon}} = -F_\alpha(1/(1 - \varepsilon)\eta) \log \left[ \frac{F_\alpha(1/\eta)}{F_\alpha(1/(1 - \varepsilon)\eta)} \right] - (1 - F_\alpha(1/(1 - \varepsilon)\eta)) \log \left[ \frac{1 - F_\alpha(1/\eta)}{1 - F_\alpha(1/(1 - \varepsilon)\eta)} \right]
\]

(39)

Figure 3 provides the tail bound constants for $\alpha = 0^+$, i.e., $G_{R,0^+,C,\varepsilon}$ and $G_{L,0^+,C,\varepsilon}$ at selected $\eta$ values ranging from 1 to 2 (spaced at 0.1). Recall that when $\alpha = 0^+$, the optimal choice of $\eta$ is 1.594 and the optimal variance coefficient $V_{0^+}(\eta)$ is 1.544.
4 2-Bit Coding and Estimation

With the 2-bit scheme, we need to introduce 3 threshold values: \( C_1 \leq C_2 \leq C_3 \), and define

\[
p_1 = \Pr(z_j \leq C_1) = F_\alpha(C_1/\Lambda_\alpha) \quad (40)
\]
\[
p_2 = \Pr(C_1 < z_j \leq C_2) = F_\alpha(C_2/\Lambda_\alpha) - F_\alpha(C_1/\Lambda_\alpha) \quad (41)
\]
\[
p_3 = \Pr(C_2 < z_j \leq C_3) = F_\alpha(C_3/\Lambda_\alpha) - F_\alpha(C_2/\Lambda_\alpha) \quad (42)
\]
\[
p_4 = \Pr(z_j > C_3) = 1 - F_\alpha(C_3/\Lambda_\alpha) \quad (43)
\]

and

\[
n_1 = \sum_{j=1}^{n} 1\{z_j \leq C_1\}, \quad n_2 = \sum_{j=1}^{n} 1\{C_1 < z_j \leq C_2\} \quad (44)
\]
\[
n_3 = \sum_{j=1}^{n} 1\{C_2 < z_j \leq C_3\}, \quad n_4 = \sum_{j=1}^{n} 1\{z_j > C_3\} \quad (45)
\]

The log-likelihood of these \( n = n_1 + n_2 + n_3 + n_4 \) observations can be expressed as

\[
l = n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + n_4 \log p_4
\]
\[
= n_1 \log F_\alpha(C_1/\Lambda_\alpha) + n_2 \log [F_\alpha(C_2/\Lambda_\alpha) - F_\alpha(C_1/\Lambda_\alpha)]
\]
\[
+ n_3 \log [F_\alpha(C_3/\Lambda_\alpha) - F_\alpha(C_2/\Lambda_\alpha)] + n_4 \log [1 - F_\alpha(C_3/\Lambda_\alpha)] \quad (46)
\]

To seek the MLE of \( \Lambda_\alpha \), we need to compute the first derivative:

\[
\ell' = \frac{\partial l}{\partial \Lambda_\alpha} = n_1 \frac{f_\alpha(C_1/\Lambda_\alpha)}{F_\alpha(C_1/\Lambda_\alpha)} \left( \frac{C_1}{\Lambda_\alpha} \right) + n_2 \frac{f_\alpha(C_2/\Lambda_\alpha) \left( -\frac{C_2}{\Lambda_\alpha} \right)}{F_\alpha(C_2/\Lambda_\alpha) - F_\alpha(C_1/\Lambda_\alpha)}
\]
\[
+ n_3 \frac{f_\alpha(C_3/\Lambda_\alpha) \left( -\frac{C_3}{\Lambda_\alpha} \right)}{F_\alpha(C_3/\Lambda_\alpha) - F_\alpha(C_2/\Lambda_\alpha)} + n_4 \frac{-f_\alpha(C_3/\Lambda_\alpha) \left( -\frac{C_3}{\Lambda_\alpha} \right)}{1 - F_\alpha(C_3/\Lambda_\alpha) \left( -\frac{C_3}{\Lambda_\alpha} \right)} \quad (47)
\]
Setting \( l' = 0 \) yields the MLE solution:

\[
0 = n_1 \frac{C_1 f_\alpha (C_1/\Lambda_\alpha)}{F_\alpha (C_1/\Lambda_\alpha)} + n_2 \frac{C_2 f_\alpha (C_2/\Lambda_\alpha) - C_1 f_\alpha (C_1/\Lambda_\alpha)}{F_\alpha (C_2/\Lambda_\alpha) - F_\alpha (C_1/\Lambda_\alpha)} + n_3 \frac{C_3 f_\alpha (C_3/\Lambda_\alpha) - C_2 f_\alpha (C_2/\Lambda_\alpha)}{F_\alpha (C_3/\Lambda_\alpha) - F_\alpha (C_2/\Lambda_\alpha)} - n_4 \frac{-C_3 f_\alpha (C_3/\Lambda_\alpha)}{1 - F_\alpha (C_3/\Lambda_\alpha)}
\]

By noting that \( E(n_1) = np_1, E(n_2) = np_2, E(n_3) = np_3, \) and \( E(n_4) = np_4, \) we have

\[
\frac{\Lambda^2_\alpha}{n} E(l'') = C_i \left[ \frac{f'_\alpha (C_1/\Lambda_\alpha) - \frac{f_\alpha^2 (C_1/\Lambda_\alpha)}{F_\alpha (C_1/\Lambda_\alpha)}}{C_1} + \frac{f'_\alpha (C_2/\Lambda_\alpha) C_2 - \frac{f_\alpha^2 (C_1/\Lambda_\alpha)}{F_\alpha (C_1/\Lambda_\alpha)}}{C_1} - \frac{(f_\alpha (C_2/\Lambda_\alpha) C_2 - f_\alpha (C_1/\Lambda_\alpha) C_1)^2}{F_\alpha (C_2/\Lambda_\alpha) - F_\alpha (C_1/\Lambda_\alpha)} \right] \\
+ \frac{f'_\alpha (C_3/\Lambda_\alpha) C_3 - \frac{f_\alpha^2 (C_2/\Lambda_\alpha)}{F_\alpha (C_2/\Lambda_\alpha)}}{C_2} - \frac{(f_\alpha (C_3/\Lambda_\alpha) C_3 - f_\alpha (C_2/\Lambda_\alpha) C_2)^2}{F_\alpha (C_3/\Lambda_\alpha) - F_\alpha (C_2/\Lambda_\alpha)} \right]
\]

\[
= - C_1 \frac{f_\alpha^2 (C_1/\Lambda_\alpha)}{F_\alpha (C_1/\Lambda_\alpha)} + \frac{(f_\alpha (C_2/\Lambda_\alpha) C_2 - f_\alpha (C_1/\Lambda_\alpha) C_1)^2}{F_\alpha (C_2/\Lambda_\alpha) - F_\alpha (C_1/\Lambda_\alpha)} \\
- \frac{(f_\alpha (C_3/\Lambda_\alpha) C_3 - f_\alpha (C_2/\Lambda_\alpha) C_2)^2}{F_\alpha (C_3/\Lambda_\alpha) - F_\alpha (C_2/\Lambda_\alpha)} - C_3 \frac{f_\alpha^2 (C_3/\Lambda_\alpha)}{1 - F_\alpha (C_3/\Lambda_\alpha)}
\]

from which we can compute the asymptotic variance of the MLE.

We summarize the results for the 2-bit scheme in Theorem 2.

**Theorem 2** Given \( n \) i.i.d. samples \( y_j \sim S(\alpha, 1), \) \( j = 1 \) to \( n, \) three thresholds \( 0 < C_1 \leq C_2 \leq C_3, \)

\[
\eta_1 = \sum_{j=1}^n 1\{z_j \leq C_1\}, \quad \eta_2 = \sum_{j=1}^n 1\{C_1 < z_j \leq C_2\}, \quad \eta_3 = \sum_{j=1}^n 1\{C_2 < z_j \leq C_3\}, \quad \eta_4 = \sum_{j=1}^n 1\{z_j > C_3\}, \quad \text{and}
\]

\[
\eta_1 = \frac{\Lambda_\alpha}{C_1}, \quad \eta_2 = \frac{\Lambda_\alpha}{C_2}, \quad \eta_3 = \frac{\Lambda_\alpha}{C_3}
\]

the MLE, denoted by \( \hat{\Lambda}_\alpha \), is the solution to the following equation:

\[
0 = n_1 \frac{C_1 f_\alpha (1/\eta_1)}{F_\alpha (1/\eta_1)} + n_2 \frac{C_2 f_\alpha (1/\eta_2) - C_1 f_\alpha (1/\eta_1)}{F_\alpha (1/\eta_2) - F_\alpha (1/\eta_1)} + n_3 \frac{C_3 f_\alpha (1/\eta_3) - C_2 f_\alpha (1/\eta_2)}{F_\alpha (1/\eta_3) - F_\alpha (1/\eta_2)} - n_4 \frac{-C_3 f_\alpha (1/\eta_3)}{1 - F_\alpha (1/\eta_3)}
\]

The asymptotic variance of the MLE is

\[
Var \left( \hat{\Lambda}_\alpha \right) = \frac{\Lambda^2_\alpha}{n} Var \left( \eta_1, \eta_2, \eta_3 \right) + O \left( \frac{1}{\eta^2} \right)
\]

where

\[
Var \left( \eta_1, \eta_2, \eta_3 \right) = \frac{1}{\eta_1^2} \frac{f_\alpha^2 (1/\eta_1)}{F_\alpha (1/\eta_1)} + \frac{(f_\alpha (1/\eta_1)^2 - f_\alpha (1/\eta_1)^2)^2}{F_\alpha (1/\eta_1) - F_\alpha (1/\eta_1)} + \frac{(f_\alpha (1/\eta_2)^2 - f_\alpha (1/\eta_2)^2)^2}{F_\alpha (1/\eta_2) - F_\alpha (1/\eta_2)} + \frac{1}{\eta_3^2} \frac{f_\alpha^2 (1/\eta_3)}{1 - F_\alpha (1/\eta_3)}
\]

Note that, with a slight abuse of notation, we still use \( \hat{\Lambda}_\alpha \) to denote the MLE of the 2-bit scheme.
4.1 \( \alpha \to 0^+ \)

In this case, we can slightly simplify the expression of \( V_\alpha(\eta_1, \eta_2, \eta_3) \):

\[
V_\alpha(\eta_1, \eta_2, \eta_3) = \frac{1}{\eta_1^2 e^{-\eta_1} + \frac{\left[ \eta_2 e^{-\eta_2} - \eta_1 e^{-\eta_1} \right]^2}{e^{-\eta_2} - e^{-\eta_1}} + \frac{\left[ \eta_3 e^{-\eta_3} - \eta_2 e^{-\eta_2} \right]^2}{e^{-\eta_3} - e^{-\eta_2}} + \frac{\eta_3^2 e^{-2\eta_3}}{e^{-\eta_3} - 1}}
\]

\[
= \frac{1}{\eta_1^2 e^{-\eta_1} + \frac{\left[ \eta_2 e^{-\eta_2} - \eta_1 e^{-\eta_1} \right]^2}{e^{-\eta_2} - e^{-\eta_1}} + \frac{\left[ \eta_3 e^{-\eta_3} - \eta_2 e^{-\eta_2} \right]^2}{e^{-\eta_3} - e^{-\eta_2}} - \frac{\eta_3^2 e^{-\eta_3}}{e^{-\eta_3} - 1}}
\]

\[
= \frac{(\eta_1^2 - \eta_2)^2}{e^{\eta_1} - e^{\eta_2}} + \frac{(\eta_2 - \eta_3)^2}{e^{\eta_2} - e^{\eta_3}} + \frac{\eta_3^2}{e^{\eta_3} - 1} \tag{54}
\]

Numerically, the minimum of \( V_{0^+}(\eta_1, \eta_2, \eta_3) \) is 1.122, attained at

\[
\eta_1 = 3.365, \quad \eta_2 = 1.771, \quad \eta_3 = 0.754 \tag{55}
\]

The value 1.122 is substantially smaller than 1.544 which is the minimum variance coefficient of the 1-bit scheme. We can also see that with the 2-bit scheme, the variance is less sensitive to the choice of the thresholds, compared to the 1-bit scheme, as illustrated in Figure 4.

In practice, there are at least two simple strategies for selecting the parameters \( \eta_1 \geq \eta_2 \geq \eta_3 \):

- **Strategy 1**: First select \( \eta_3 \), then let \( \eta_2 = t\eta_3 \) and \( \eta_1 = t\eta_2 \), for some \( t > 1 \).
- **Strategy 2**: First select a “small” \( \eta_3 \) and a “large” \( \eta_1 \), then select a “reasonable” \( \eta_2 \) in between.

\[\text{Figure 4: Left panel (strategy 1): } V_{0^+}(\eta_1, \eta_2, \eta_3) \text{ for } \eta_2 = t\eta_3, \ \eta_1 = t\eta_2, \text{ at } t = 2, 3, 4. \text{ We let } \eta_3 \text{ vary from 0 to 2.5. Right panel (strategy 2): } V_{0^+}(\eta_1, \eta_2, \eta_3) \text{ for fixed } \eta_1 = 5, \ \eta_3 = 0.5, 0.75, \text{ or } 1, \ \text{and } \eta_2 \text{ varying between } \eta_3 \text{ and } \eta_1. \]

See the plots for the examples of the two strategies in Figure 4. We should re-iterate that for the task of estimating \( \Lambda_\alpha \) using only a few bits, we must choose parameters (thresholds) beforehand. While in general the optimal results are not attainable, the hope is that as long as \( \Lambda_\alpha \) falls in a “reasonable” range, we can select the parameters so that the estimation variance is not too far away from the theoretical optimal value.

4.2 \( \alpha = 1 \)

Again, by a numerical procedure, we know that the minimum of \( V_1(\eta_1, \eta_2, \eta_3) \) is 2.087, attained at

\[
\eta_1 = 1.927, \quad \eta_2 = 1.000, \quad \eta_3 = 0.519 \tag{56}
\]

\[\text{Figure 4: Left panel (strategy 1): } V_{0^+}(\eta_1, \eta_2, \eta_3) \text{ for } \eta_2 = t\eta_3, \ \eta_1 = t\eta_2, \text{ at } t = 2, 3, 4. \text{ We let } \eta_3 \text{ vary from 0 to 2.5. Right panel (strategy 2): } V_{0^+}(\eta_1, \eta_2, \eta_3) \text{ for fixed } \eta_1 = 5, \ \eta_3 = 0.5, 0.75, \text{ or } 1, \ \text{and } \eta_2 \text{ varying between } \eta_3 \text{ and } \eta_1. \]

See the plots for the examples of the two strategies in Figure 4. We should re-iterate that for the task of estimating \( \Lambda_\alpha \) using only a few bits, we must choose parameters (thresholds) beforehand. While in general the optimal results are not attainable, the hope is that as long as \( \Lambda_\alpha \) falls in a “reasonable” range, we can select the parameters so that the estimation variance is not too far away from the theoretical optimal value.
Note that the value 2.087 is very close to the optimal variance coefficient, i.e., 2, using full information. Figure 5 plots the examples of \( V_1(\eta_1, \eta_2, \eta_3) \) for both “strategy 1” (left panel) and “strategy 2” (right panel).

Figure 5: **Left panel** (strategy 1): \( V_1(\eta_1, \eta_2, \eta_3) \) for \( \eta_2 = t\eta_3, \eta_1 = t\eta_2, \) at \( t = 2, 3, 4 \). We let \( \eta_3 \) vary from 0 to 2. **Right panel** (strategy 2): \( V_1(\eta_1, \eta_2, \eta_3) \) for fixed \( \eta_1 = 3, \eta_3 = 0.25, 0.5, \) or 0.75, and \( \eta_2 \) varying between \( \eta_3 \) and \( \eta_1 \).

4.3 \( \alpha = 2 \)

Numerically, the minimum of \( V_2(\eta_1, \eta_2, \eta_3) \) is 2.236, attained at

\[
\eta_1 = 0.546, \quad \eta_2 = 0.195, \quad \eta_3 = 0.093
\]

(57)

Figure 6 presents the examples of \( V_2(\eta_1, \eta_2, \eta_3) \) for both strategies for choosing \( \eta_1, \eta_2, \) and \( \eta_3 \).

4.4 **Efficient Computational Procedure for the MLE Solutions**

With the 1-bit scheme, the computational cost for the MLE is negligible because we either have a closed-form solution (e.g., when \( \alpha = 0^+ \)) or the solution only requires a simple function inversion (which can be easily done by tabulating the results).

With the 2-bit scheme, however, the computational cost can be a major concern if we try to find the MLE solution numerically every time (at run time). The computationally efficient solution is (again) to tabulate the results, as long as the table is not too large. To see this, we can re-write the log-likelihood function

\[
l = \frac{n_1}{n} \log F_\alpha \left( 1/\eta_1 \right) + \frac{n_2}{n} \log \left[ F_\alpha \left( 1/\eta_2 \right) - F_\alpha \left( 1/\eta_1 \right) \right] + \frac{n_3}{n} \log \left[ F_\alpha \left( 1/\eta_3 \right) - F_\alpha \left( \eta_2 \right) \right] + \frac{n - (n_1 + n_2 + n_3)}{n} \log \left[ 1 - F_\alpha \left( \eta_3 \right) \right]
\]

(58)

This means, we only need to tabulate the results for the combination of \( n_1/n, n_2/n, n_3/n \) (which all vary between 0 and 1). Suppose we tabulate 100 values for each \( n_i/n \) (i.e., at an accuracy of 0.01), then the table size is only \( 100^3 = 10^6 \), which is very small. This provides a simple solution to the computational problem.
Multi-Bit (Multi-Partition) Coding and Estimation

Clearly, we can extend the method to more than 2 bits, but we must watch carefully for the computational and storage costs. Suppose we would like to fully exploit a 3-bit scheme. We will then need to use $2^3 - 1 = 7$ threshold values. In other words, we will partition the space into $2^3 = 8$ parts. Further suppose numerically solving for the MLE at run time is not acceptable and we would like to resort to the previously mentioned solution by tabulation. We encounter a memory problem because the table size will be too large.

For example, suppose we tabulate the results for 100 values of each $n_i/n$. The table size for the (full) 3-bit scheme becomes $100^7 = 10^{14}$, which is probably not (economically) realistic for most applications. On the other hand, most PCs these days would be able to handle a table of size $(100)^5 = 10^{10}$.

It is more flexible to consider schemes based on $(m + 1)$ partitions. For example, $m = 1$ for the 1-bit scheme, $m = 3$ for the 2-bit scheme, and $m = 7$ for the 3-bit scheme. We feel a $(5+1)$-partition scheme (i.e., $m = 5$) is particularly interesting, and we will focus on this case.

With the $(m + 1)$-partition scheme, the asymptotic variance of the MLE $\hat{\Lambda}_\alpha$ can be expressed as

$$Var\left(\hat{\Lambda}_\alpha\right) = \frac{\Lambda^2_\alpha}{n} V_\alpha(\eta_1, \ldots, \eta_m) + O\left(\frac{1}{n^2}\right)$$  \hspace{1cm} (59)$$

$$V_\alpha(\eta_1, \ldots, \eta_m) = \frac{1}{\eta_1^2} \frac{f_\alpha^2(1/\eta_1)}{F_\alpha(1/\eta_1)} + \sum_{s=1}^{m-1} \frac{f_\alpha(1/\eta_{s+1})/\eta_{s+1} - f_\alpha(1/\eta_s)/\eta_s}{F_\alpha(1/\eta_{s+1}) - F_\alpha(1/\eta_s)} + \frac{1}{\eta_m^2} \frac{f_\alpha^2(1/\eta_m)}{1 - F_\alpha(1/\eta_m)}$$  \hspace{1cm} (60)$$

Figure 6: Up-left panel (strategy 1): $V_2(\eta_1, \eta_2, \eta_3)$ for $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$. We let $\eta_3$ vary from 0 to 0.5. Other 3 panel (strategy 2): $V_2(\eta_1, \eta_2, \eta_3)$ for 3 values of $\eta_1$ (3, 2, 1, corresponding to each panel), $\eta_3 = 0.05, 0.1,$ or 0.2, and $\eta_2$ varying between $\eta_3$ and $\eta_1$. 

5 Multi-Bit (Multi-Partition) Coding and Estimation
5.1 $\alpha = 0+$ and $m = 5$

Numerically, the minimum of $V_{0+}(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ is 1.055, attained at

$$\eta_1 = 4.464, \quad \eta_2 = 2.871, \quad \eta_3 = 1.853, \quad \eta_4 = 1.099, \quad \eta_5 = 0.499$$  \tag{61}

Figure 7 (right panel) plots $V_{0+}(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying $\eta_5$ and $\eta_i = t\eta_{i+1}, i = 4,3,2,1$. For comparison, we also plot (in the left panel) $V_{0+}(\eta_1, \eta_2, \eta_3)$ for varying $\eta_3$, and $\eta_2 = t\eta_3, \eta_1 = t\eta_2$. We can see that with more partitions, the performance becomes significantly more robust.

5.2 $\alpha = 1$ and $m = 5$

Numerically, the minimum of $V_1(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ is 2.036, attained at

$$\eta_1 = 2.602, \quad \eta_2 = 1.498, \quad \eta_3 = 1.001, \quad \eta_4 = 0.668, \quad \eta_5 = 0.385$$  \tag{62}

Figure 8 (right panel) plots $V_1(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying $\eta_5$ and $\eta_i = t\eta_{i+1}, i = 4,3,2,1$. Again, for comparison, we also plot (in the left panel) $V_1(\eta_1, \eta_2, \eta_3)$ for varying $\eta_3$, and $\eta_2 = t\eta_3, \eta_1 = t\eta_2$. Clearly, using more partitions stabilizes the variances even when the parameters are chosen less appropriately.
5.3 $\alpha = 2$ and $m = 5$

Numerically, the minimum of $V_2(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ is 2.106, attained at

$$\eta_1 = 0.893, \quad \eta_2 = 0.339, \quad \eta_3 = 0.184, \quad \eta_4 = 0.111, \quad \eta_5 = 0.068$$

Figure (right panel) plots $V_2(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying $\eta_5$ and $\eta_i = t\eta_{i+1}$, $i = 4, 3, 2, 1$, as well as (in the left panel) $V_2(\eta_1, \eta_2, \eta_3)$ for varying $\eta_3$, and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$.

![Diagram](image-url)

Figure 9: **Left panel** (4-partition): $V_2(\eta_1, \eta_2, \eta_3)$ for varying $\eta_3$ and $\eta_2 = t\eta_3$, $\eta_1 = t\eta_2$, at $t = 2, 3, 4$.

**Right panel** (6-partition): $V_2(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$ for varying $\eta_5$ and $\eta_i = t\eta_{i+1}$, at $t = 2, 3, 4$.

6 Extension and Future Work

In this paper, we have focused on coding schemes for $\alpha$-stable random projections operated on individual data vectors. We feel an important line of future work would be the design of coding schemes for analyzing the relation of two or multiple data vectors, which will be very useful in the context of large-scale machine learning.

For example, [9] assumed sum-to-one (i.e., the $l_1$ norm = 1) nonnegative data vectors. After applying Cauchy stable random projections separately on two data vectors, the collision probability of the two signs of the projected data is essentially monotonic in the $\chi^2$ similarity (which is popular in computer vision). Now the open question is that, suppose we do not know the $l_1$ norms, how we should design coding schemes so that we can still evaluate the $\chi^2$ similarity (or other similarities) using Cauchy random projections.

[8] re-visited Gaussian random projections. By assuming unit $l_2$ norms of data vectors, [8] developed multi-bit coding schemes and estimators for the correlation between vectors. Can we, using just a few bits, still estimate the correlation if at the same time we must also estimate the $l_2$ norms?

7 Conclusion

Motivated by the recent work on “one scan 1-bit compressed sensing”, we have developed 1-bit and multi-bit coding schemes for estimating the scale parameter of $\alpha$-stable distributions. These simple coding schemes (even just 1-bit) perform well in that, if the parameters are chosen appropriately, their variances are actually not much larger than the variances using full (i.e., infinite-bit) information. In general, using more bits increases the computational cost or storage cost (e.g., the cost of tabulations), with the benefits of stabilizing the performance so that the estimation variances do not increase much even when the parameters are far from optimal. In practice, we expect the $(m + 1)$-partition scheme, combined with tabulation, for $m = 3, 4,$ or 5, should be overall preferable. Here $m = 3$ corresponds to the 2-bit scheme, $m = 1$ to the 1-bit scheme.
References

[1] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[2] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[3] N. Cressie. A note on the behaviour of the stable distributions for small index. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31(1):61–64, 1975.

[4] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of ACM*, 53(3):307–323, 2006.

[5] P. Li. Estimators and tail bounds for dimension reduction in $l_\alpha$ ($0 < \alpha \leq 2$) using stable random projections. In *SODA*, pages 10 – 19, San Francisco, CA, 2008.

[6] P. Li. One scan 1-bit compressed sensing. Technical report, [arXiv:1503.02346](https://arxiv.org/abs/1503.02346), 2015.

[7] P. Li and T. J. Hastie. A unified near-optimal estimator for dimension reduction in $l_\alpha$ ($0 < \alpha \leq 2$) using stable random projections. In *NIPS*, Vancouver, BC, Canada, 2007.

[8] P. Li, M. Mitzenmacher, and A. Shrivastava. Coding for random projections. In *ICML*, 2014.

[9] P. Li, G. Samorodnitsky, and J. Hopcroft. Sign cauchy projections and chi-square kernel. In *NIPS*, Lake Tahoe, NV, 2013.

[10] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, 1994.

[11] V. M. Zolotarev. *One-dimensional Stable Distributions*. American Mathematical Society, Providence, RI, 1986.