

Contents

		1
1	Introduction	1
1.1	Basic Inequalities	1
1.2	Why bother?	2
1.2.1	Coin Tossing	2
1.2.2	Central Limit Theorem	2
2	Exponential Inequalities	4
2.1	Chevyshev vs Chernoff vs Hoeffding	4
2.2	Chernoff-Okamoto Inequalities	4
2.3	Hoeffding-Bernstein inequalities	5
3	Application to Estimation of Data Dimension	6
3.1	Proofs	8
4	Application to a Heuristic Algorithm for Travelling Salesman	11

1 Introduction

1.1 Basic Inequalities

Theorem 1.1.1 (Markov's inequality). For a random variable X with $\mathbf{P}\{X < 0\} = 0$ and $t > 0$, we have

$$\mathbf{P}\{X \geq t\} \leq \frac{\mathbf{E} X}{t}.$$

It follows that for a non-decreasing function φ which only takes non-negative values,

$$\mathbf{P}\{X \geq t\} = \mathbf{P}\{\varphi(X) \geq \varphi(t)\} \leq \frac{\mathbf{E} \varphi(X)}{\varphi(t)}.$$

Proof. In the first place, note that

$$\begin{aligned} X &= X \cdot \mathbf{1}_{X \geq t} + X \cdot \mathbf{1}_{X < t} \\ &\geq t \cdot \mathbf{1}_{X \geq t} + 0, \end{aligned}$$

and thus,

$$\mathbf{E} X \geq t \cdot \mathbf{E} \mathbf{1}_{X \geq t} = t \cdot \mathbf{P}\{X \geq t\}.$$

For the second statement, apply the same argument on the random variable $Y := \varphi(X)$ and the constant $s := \varphi(t)$. \square

Theorem 1.1.2 (Chebyshev's inequality). For $t > 0$ and a random variable X with mean $\mu = \mathbf{E} X$ and variance $\sigma^2 = \mathbf{Var} X$, then

$$\mathbf{P}\{|X - \mu| \geq t\} \leq \sigma^2 t^{-2}.$$

Proof. Applying Markov's inequality with $\varphi : x \mapsto x^2$ we obtain,

$$\mathbf{P}\{|X - \mu| \geq t\} = \mathbf{P}\{|X - \mu|^2 \geq t^2\} \leq \frac{\mathbf{E} [(X - \mu)^2]}{t^2} = \sigma^2 t^{-2}.$$

\square

Theorem 1.1.3 (Jensen's inequality). For any real valued random variable X and convex function φ

$$\varphi(\mathbf{E} X) \leq \mathbf{E} \varphi(X)$$

1.2 Why bother?

The concentration inequalities are used to obtain information on how a random variable is distributed at some specific places of its domain. In the most common scenarios, these inequalities will be used to quantify how concentrated a random variable at its tails, for example,

$$\mathbf{P}\{|X - \mu| \geq t\} < f(t) \ll 1.$$

A concentration inequality is specially useful when this probability cannot be calculated at a low computational cost or estimated with high precision. The following will illustrate a case where using concentration inequalities achieves the best results.

1.2.1 Coin Tossing

A coin tossing game is fair if the chances of winning are equal to the chances of losing. We can verify from a sample of N games that the game is not rigged if the number of heads in the sample is not very distant from the average $N/2$. However, there's a chance that one may classify the coin as rigged, even when the coin is fair. By the *Law of Large Numbers*, we know that the larger the sample, the less likely it is to obtain a false positive. But let's ask ourselves how fast this probability converges to 0.

Let $S_N \sim \text{Bi}(N, 1/2)$ denote the number of heads in a fair coin tossing game. Then,

$$\mu = \mathbf{E} S_N = \frac{N}{2}, \quad \sigma^2 = \mathbf{Var} S_N = \frac{N}{4}.$$

For a fixed $\varepsilon > 0$, we may classify a coin tossing game as rigged if, after N trials, the ratio of heads vs tails in the sample is greater than $[1 + \varepsilon : 1 - \varepsilon]$, or similarly,

$$S_N \geq \mu + \frac{\varepsilon}{2}N = \frac{1 + \varepsilon}{2}N.$$

Using the Chebyshev inequality 1.1.2, we assert that

$$\mathbf{P}\left\{S_N \geq \mu + \frac{\varepsilon}{2}N\right\} \leq \mathbf{P}\left\{|S_N - \mu| \geq \frac{\varepsilon}{2}N\right\} \leq \sigma^2 \frac{4}{\varepsilon^2 N^2} = \frac{1}{\varepsilon^2 N}.$$

Therefore, the probability of bad events tends to 0 at least linearly with the number of games.

1.2.2 Central Limit Theorem

The proof of the following theorems can be found in (ref)

Theorem 1.2.1. Let X_i be a i.i.d. sample. Let $S_N = \sum_{i=1}^N X_i$, with mean $\mu = \mathbf{E} S_N$ and variance $\sigma^2 = \mathbf{Var} S_N$. If

$$Z_N = \frac{S_N - N \cdot \mathbf{E} X_i}{\sqrt{N \cdot \mathbf{Var} X_i}} = \frac{S_N - \mu}{\sqrt{N} \sigma},$$

1 Introduction

66 then,

$$67 \quad Z_N \rightarrow Z \sim \mathcal{N}(0, 1), \text{ in distribution.}$$

68

□

69 **Theorem 1.2.2** (Tails of the Normal Distribution). Let $Z \sim \mathcal{N}(0, 1)$, for $t > 0$ we have

$$70 \quad \left(\frac{1}{t} - \frac{1}{t^3}\right) \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) \leq \mathbf{P}\{Z \geq t\} \leq \frac{1}{t} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right).$$

71

□

72 With that in mind, we might naively assume that better bounds can be obtained by
73 using the previous theorem. For a large enough N we can say that for the coin tossing,

$$74 \quad Z_N = \frac{S_N - N/2}{\sqrt{N/4}}$$

75

$$76 \quad \Rightarrow \mathbf{P}\left\{S_N \geq \frac{1+\varepsilon}{2}N\right\} = \mathbf{P}\left\{Z_N \geq \varepsilon\sqrt{N}\right\} \sim \mathbf{P}\left\{Z \geq \varepsilon\sqrt{N}\right\}.$$

77 However, this raises the question of whether we can draw the following conclusion from
78 Theorem 1.2.2:

$$79 \quad \mathbf{P}\left\{S_N \geq \frac{1+\varepsilon}{2}N\right\} \leq \frac{1}{\varepsilon\sqrt{N}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\varepsilon^2 \cdot N}{2}\right).$$

80 Unfortunately, the answer is no. The following theorem will show why.

81 **Theorem 1.2.3** (Convergence Rate for Central Limit Theorem). For Z_N, Z in Theo-
82 rem 1.2.1, we have:

$$83 \quad |\mathbf{P}\{Z_N \geq t\} - \mathbf{P}\{Z \geq t\}| \in O\left(\frac{1}{\sqrt{N}}\right).$$

84

□

85 Since the approximation error is greater than the bound, the previous results cannot
86 be taken into account.

87 In the context of coin tossing, this may not matter at all because the linear bound
88 obtained using Chebyshev's inequality indicates that the probability of wrongly classi-
89 fying a fair coin as a rigged coin converges at least linearly to zero. Even the Central
90 Limit Theorem shows in a less precise way this convergence. However, for some specific
91 problems in statistics, these basic tools are not precise enough to solve them. In the fol-
92 lowing chapters, we will show some examples where better crafted strategies are needed
93 in order to get bounds to the tails of the random variables.

2 Exponential Inequalities

94

Even if we are satisfied with the linear convergence rate provided by Chebyshev's inequality, there is a simple way to improve this bound with the following result.

95

96

Theorem 2.0.1 (Hoeffding's inequality). Let X_1, \dots, X_N be independent random variables, where $X_i \sim \text{Be}(p_i)$. Then, for every $t > 0$, we have

97

98

$$P \left\{ \sum (X_i - \mathbf{E} X_i) \geq t \right\} \leq \exp \left(\frac{-2t^2}{N} \right)$$

99

Proof. TODO

□ 100

Another exponential bound we can derive with a similar idea is the following

101

Theorem 2.0.2 (Chernoff's inequality). Let X_1, \dots, X_N be independent random variables, where for every $i = 1, \dots, N$, $X_i \in [a_i, b_i]$. Define $S_N = \sum X_i$ and let $\mu = \mathbf{E} S_N$. Then, for every $t > 0$, we have

102

103

104

$$P \{ S_N \geq \mu + t \} \leq \exp \left(\frac{-2t^2}{\sum (a_i - b_i)^2} \right)$$

105

2.1 Chevyshev vs Chernoff vs Hoeffding

106

2.2 Chernoff-Okamoto Inequalities

107

Applying Markov's Inequality to $Y = e^{uX}$, we can assert that

108

$$\mathbf{P}\{X \geq \lambda + t\} \leq e^{-u(\lambda+t)} \mathbf{E} e^{uX} = e^{-u(\lambda+t)} (1 - p + pe^u)^n.$$

109

The right hand equation is minimized when,

110

$$e^u = \frac{\lambda + t}{(n - \lambda - t)} \cdot \frac{1 - p}{p}.$$

111

Therefore, for $0 \leq t \leq n - \lambda$,

112

$$\mathbf{P}\{X \geq \lambda + t\} \leq \left(\frac{\lambda}{\lambda + t} \right)^{\lambda+t} \left(\frac{n - \lambda}{n - \lambda - t} \right)^{n-\lambda-t} \quad (2.1)$$

113

2 Exponential Inequalities

114 **Theorem 2.2.1.** Let X be random variable with the binomial distribution $\text{Bi}(n, p)$ with
115 $\lambda := np = \mathbf{E} X$, then for $t \geq 0$,

116
$$\mathbf{P}\{X \geq \lambda + t\} \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right) \quad (2.2)$$

117
118
$$\mathbf{P}\{X \leq \lambda - t\} \leq \exp\left(-\frac{t^2}{2\lambda}\right) \quad (2.3)$$

119 **Used in:** Theorem 3.0.2

120 *Proof.* (TODO I've already written the proof on paper) □

121 2.3 Hoeffding-Bernstein inequalities

122 **Theorem 2.3.1.** Let $\|f\|_\infty < c$, $\mathbf{E} f(X_1, \dots, X_m) = 0$ and $\sigma^2 = \mathbf{E} f^2(X_1, \dots, X_m)$.
123 Then for any $t > 0$,

124
$$\mathbf{P}\{U_m^n(f, P) > t\} \leq \exp\left(\frac{\frac{n}{m}t^2}{2\sigma^2 + \frac{2}{3}ct}\right) \quad (2.1)$$

125 **Used in:** Theorem 3.0.4

126 *Proof.* Proposition 2.3. (a) Arcones and Giné (1993) □

3 Application to Estimation of Data Dimension

The article [Díaz et al. \(2019\)](#) explains how we can estimate the dimension d of a manifold M embedded on a Euclidean space of dimension m , say \mathbb{R}^m . First, we are going to introduce the method they used, and then, we will show how does the exponential inequalities can be used to prove two important results in the paper. The procedure starts with an example on a uniformly distributed sample on a d -sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$, but will be later generalized for samples of any distribution on any manifold.

In the first place, let Z_1, \dots, Z_k be a i.i.d. sample uniformly distributed on \mathbb{S}^{d-1} . Then, we have the following formula for the variance of the angles between $Z_i, Z_j, i \neq j$:

$$\beta_d := \mathbf{Var}(\arccos \langle Z_i, Z_j \rangle) = \begin{cases} \frac{\pi^2}{4} - 2 \sum_{j=1}^k (2j-1)^{-2}, & \text{if } d = 2k+1 \text{ is odd,} \\ \frac{\pi^2}{12} - 2 \sum_{j=1}^k (2j)^{-2}, & \text{if } d = 2k+2 \text{ is even.} \end{cases} \quad (3.1)$$

The previous formula for the angle variance is proven in [Díaz et al. \(2019\)](#). In order to give more insight on how we will be choosing an estimator \hat{d} of the dimension of the sphere, consider the following theorem.

Theorem 3.0.1 (Bounds for β_d). For every $d > 1$,

$$\frac{1}{d} \leq \beta_d \leq \frac{1}{d-1}.$$

□

Knowing that for every $d > 1$, β_d is in the interval $[\frac{1}{d}, \frac{1}{d-1}]$, one can guess the dimension of the sphere by estimating β_d , and then, taking d from the lower bound of the interval where our estimator is. Since β_d is the variance of the angles in our sphere, our best choice for an estimator is the angle's sample variance,

$$U_k = \binom{k}{2}^{-1} \sum_{i < j \leq k} \left(\arccos \langle Z_i, Z_j \rangle - \frac{\pi^2}{2} \right)^2. \quad (3.2)$$

3 Application to Estimation of Data Dimension

150 In Proposition 1. of [Díaz et al. \(2019\)](#) the authors prove that it's the Minimum Vari-
151 ance Unbiased Estimator for β_d on the unit sphere.

152
153 Furthermore, the authors also prove that this result can be generalized for any mani-
154 fold with samples of any distribution. Let X_1, \dots, X_n be a i.i.d. sample from a random
155 distribution P on a manifold $M \subset \mathbb{R}^m$, and let $p \in M$ a point. For $C > 0 \in \mathbb{R}$, let
156 $k = \lceil C \ln(n) \rceil$ and define $R(n) = L_{k+1}(p)$ as the distance between p and its $(k+1)$ -
157 nearest neighbor. W.L.O.G. assume that $p = 0 \in M$ and that X_1, \dots, X_k are the
158 k -nearest neighbors of p . Additionally, for the following theorem to be true, we require
159 that at any neighborhood of p , the probability in that neighborhood is greater than 0.

160
161 The following theorem uses a special inequality from Chernoff-Okamoto, and it's cru-
162 cial in the idea behind this generalization.

163 **Theorem 3.0.2** (Bound k -neighbors). For any sufficiently large $C > 0$, we have that,
164 there exists n_0 such that, with probability 1, for every $n \geq n_0$,

$$165 \quad R(n) \leq f_{p,P,C}(n) = O(\sqrt[d]{\ln(n)/n}), \quad (3.3)$$

166 where the function $f_{p,P,C}$ is a deterministic function which depends on p , P and C .

167 . □

168 The following theorem, although it does not require concentration inequalities, is
169 important for connecting the idea of the previous theorem to the main frame. Let
170 $\pi : \mathbb{R}^m \rightarrow T_p M$ be the orthogonal projection on the Tangent Space of M at p . Also,
171 define $W_i := \pi(X_i)$ and then normalize,

$$172 \quad Z_i := \frac{X_i}{\|X_i\|}, \quad \widehat{W}_i := \frac{W_i}{\|W_i\|}. \quad (3.4)$$

173 **Theorem 3.0.3** (Projection Distance Bounds). For any $i < j \leq n$,

$$174 \quad (i) \quad \|X_i - \pi(X_i)\| = O(\|\pi(X_i)\|^2) \quad (3.5)$$

$$175 \quad (ii) \quad \|Z_i - \widehat{W}_i\| = O(\|\pi(X_i)\|) \quad (3.6)$$

176 (iii) The inner products (cosine of angles) can be bounded as it follows:

$$177 \quad |\langle Z_i, Z_j \rangle - \langle \widehat{W}_i, \widehat{W}_j \rangle| \leq Kr, \quad (3.7)$$

178 for a constant $K \in \mathbb{R}$, whenever $r \geq \max(\|\pi(X_i)\|, \|\pi(X_j)\|)$.

179 . □

180 What follows is that if we know W_1, \dots, W_k are behaved similar to a uniformly dis-
181 tributed sample on the sphere \mathbb{S}^d , then, Z_1, \dots, Z_k (the normalized k -nearest neighbors
182 of p) also behave like they are uniformly distributed on \mathbb{S}^d . The following theorem is
183 made by combining the ideas of the previous theorems.

Theorem 3.0.4 (Projection's Angle Variance Statistic). For $k = O(\ln n)$, let

$$V_{k,n} = \binom{k}{2}^{-1} \sum_{i < j \leq k} \left(\arccos \langle \widehat{W}_i, \widehat{W}_j \rangle - \frac{\pi^2}{2} \right)^2, \quad (3.8)$$

and let $U_{k,n} = U_k$ from equation 3.2. The following statements hold

- (i) $k|U_{k,n} - V_{k,n}| \xrightarrow{n \rightarrow \infty} 0$, in probability. (3.9)
- (ii) $\mathbf{E} |U_{k,n} - V_{k,n}| \xrightarrow{n \rightarrow \infty} 0$.

□

This last theorem is as far as this document is planned to cover. However, the last result in the paper provides the main statement. It says that if we estimate β_d as we did with $U_{k,n}$ from 3.0.4, and then, extract \widehat{d} from the interval where $U_{k,n}$ is located, it follows that,

Theorem 3.0.5 (Consistency). When $n \rightarrow \infty$,

$$\mathbf{P}\{\widehat{d} \neq d\} \rightarrow 0.$$

3.1 Proofs

Proof Theorem 3.0.1: The even and the odd cases must be distinguished:

- (1): When $d = 2k + 2$ is even: In the first place, remember that,

$$\lim_{k \rightarrow \infty} \sum_{j=1}^k j^{-2} = \frac{\pi^2}{6}.$$

It follows that

$$\begin{aligned} \beta_d &= \frac{\pi^2}{12} - 2 \sum_{j=1}^k (2j)^{-2} = \frac{\pi^2}{12} - \frac{1}{2} \sum_{j=1}^k j^{-2} \\ &= \frac{1}{2} \sum_{j=k+1}^{\infty} j^{-2}. \end{aligned}$$

Since $(j^{-2})_{j \in \mathbb{N}}$ is a monotonically decreasing sequence, it follows that (**TODO:** Improve array syntax)

$$\begin{aligned} \frac{1}{d} &= \frac{1}{2k+2} = \frac{1}{2} \int_{k+1}^{\infty} x^{-2} dx \\ &\leq \beta_d \leq \frac{1}{2} \int_{k+1/2}^{\infty} x^{-2} dx \\ &= \frac{1}{2k+1} = \frac{1}{d-1}. \end{aligned}$$

3 Application to Estimation of Data Dimension

(2): When $d = 2k + 3$ is odd: On the other hand, note that

$$\begin{aligned} \lim_{k \rightarrow \infty} \sum_{j=1}^k (2j-1)^{-2} &= \lim_{k \rightarrow \infty} \sum_{j=1}^{2k-1} j^{-2} - \sum_{j=1}^{k-1} (2j)^{-2} \\ &= \lim_{k \rightarrow \infty} \sum_{j=1}^{2k-1} j^{-2} - \frac{1}{4} \sum_{j=1}^{k-1} j^{-2} \\ &= \frac{\pi^2}{6} - \frac{\pi^2}{24} = \frac{\pi^2}{8} \end{aligned}$$

Then,

$$\begin{aligned} \beta_d &= \frac{\pi^2}{4} - 2 \sum_{j=1}^k (2j-1)^{-2} \\ &= 2 \sum_{j=k+1}^{\infty} (2j-1)^{-2}. \end{aligned}$$

Using a similar argument we conclude that (**TODO**: Improve array syntax)

$$\begin{aligned} \frac{1}{d} &= \frac{1}{2k+1} = 2 \int_{k+1}^{\infty} (2x-1)^{-2} dx \\ &\leq \beta_d \leq 2 \int_{k+1/2}^{\infty} (2x-1)^{-2} dx \\ &= \frac{1}{2k+2} = \frac{1}{d-1}. \end{aligned}$$

□

Proof Theorem 3.0.2: The volume of a d -sphere of radius r is equal to:

$$v_d r^d = \frac{\pi^{d/2}}{\Gamma(\frac{n}{2} + 1)} r^d.$$

Where v_d is the volume of the unit d -sphere. For the assumptions we made on P and M around $p = 0$, we can say that for any $r > 0$, there's a percent (greater than 0) of the sample that is within a range r from p . This proportion is subordinated only by the volume of a d -sphere of radius r and a constant $\alpha := \alpha(P)$ that depends on the distribution P :

$$\rho = \mathbf{P}\{X \in M : |X| < r\} \geq \alpha v_d r^d > 0.$$

We can now define a binomial process based on how many neighbors does p has within a range r . Let $N = N_r \sim \text{Bi}(n, \rho)$ be the number of neighbors, using **Theorem 2.2.1** with $\lambda = n\rho$ and $t = \frac{\lambda}{2}$ we obtain,

$$\mathbf{P}\{N \leq \lambda - t\} = P\{2N \leq \lambda\} \leq \exp(-\lambda/8).$$

3 Application to Estimation of Data Dimension

Since $n(\alpha v_d r^d) \leq n\rho = \lambda$, it follows that, by choosing $r(n)$ such that 224

$$r(n) = \left(\frac{C}{\alpha v_d} \cdot \frac{\ln n}{n} \right)^{1/d} = O(\sqrt[d]{\ln(n)/n}), \quad (\star) \quad 225$$

and thus, 226

$$C \ln n = n(\alpha v_d r(n)^d) \leq \lambda, \quad 227$$

we obtain: 228

$$P\{2N \leq C \ln n\} \leq P\{2N \leq \lambda\}, \quad 229$$

and, 230

$$\exp(-\lambda/8) \leq \exp\left(\frac{-C \ln n}{8}\right) = n^{-C/8}. \quad 231$$

Therefore, 232

$$P\{2N \leq C \ln n\} \leq n^{-C/8}. \quad 233$$

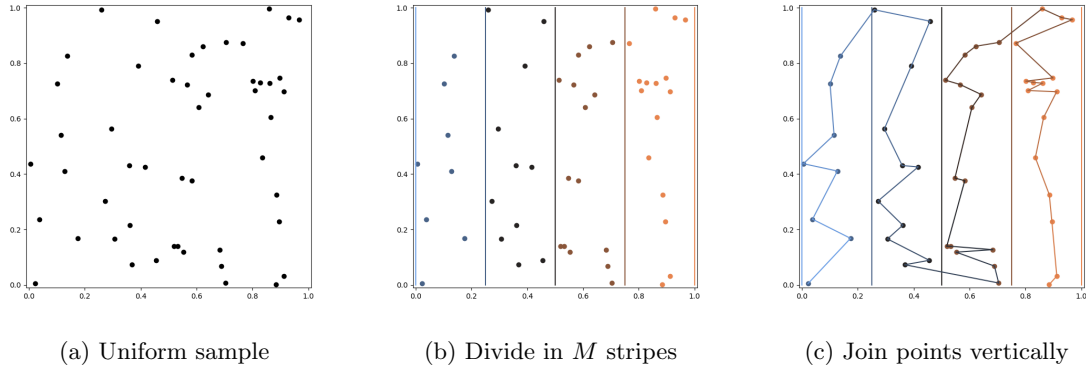
Finally, with this last expression we proved that if $k = \frac{C}{2} \ln n$, then the k -neighbors of p 234
are contained in the ball of radius $r(n)$ with a probability that converges exponentially 235
to 1. □ 236

4 Application to a Heuristic Algorithm for Travelling Salesman

In this section we are going to present an application of the Azuma-Hoeffding inequality to prove the convergence to the mean of a linear approximation algorithm for the *Travelling Salesman Problem*.

The Algorithm

Let X_1, \dots, X_N be a sample of N uniformly distributed points in a compact square $[0, L] \times [0, L]$. The algorithm divides this square in M stripes of width L/M each. Then, it connects each of the points in each of the stripes vertically and connects the top-most of one stripe with the top-most of the next one (or viceversa as the image below shows).



In the reference [Gzyl et al. \(1990\)](#) the authors assert that by choosing a number of stripes $M^* = \lfloor 0.58N^{1/2} \rfloor$, one can achieve the best result in comparison to the real TSP solution. If t_N is the TSP solution distance for our sample and d_N is the algorithm's answer with the optimal M^* , then the error is asymptotically:

$$\frac{d_N - t_N}{t_N} \approx 0.23.$$

The result that we are going to prove is that d_N converges with an exponential rate to its mean. To prove our point, we are going to modify the algorithm's trajectory as it follows. Let e_N be trajectory distance that for any empty stripe in the plane we sum the length of its diagonal $\sqrt{L^2 + L^2/M^2}$ and then it skips the empty stripe. When there are

4 Application to a Heuristic Algorithm for Travelling Salesman

no empty stripes $e_N = d_N$ and the probability that any given stripe is empty converges exponentially to 0:

$$\begin{aligned} (1 - 1/M)^N &= (1 - 0.58^{-1}N^{-1/2})^N \\ &= \left((1 - 1/M)^M\right)^{0.58^{-1}N^{1/2}} \\ &\sim \exp(-0.58^{-1}N^{1/2}). \end{aligned}$$

Let $\mathcal{A}_i := \sigma\{X_1, \dots, X_i\}$ be the sigma algebra corresponding to revealing the first i points, $\mathcal{A}_0 = \{\emptyset, [0, L]^2\}$. The expected value of the trajectory e_N given that we only know the positions of the first i points in the sample is $\mathbf{E}(e_N|\mathcal{A}_i)$. Define

$$Z_i = \mathbf{E}(e_N|\mathcal{A}_i) - \mathbf{E}(e_N|\mathcal{A}_{i-1}),$$

As the difference of this expectations when we reveal 1 more point. Note that since

$$\mathbf{E}(Z_i|\mathcal{A}_i) = \mathbf{E}(e_N|\mathcal{A}_i, \mathcal{A}_i) - \mathbf{E}(e_N|\mathcal{A}_{i-1}, \mathcal{A}_i) = \mathbf{E}(e_N|\mathcal{A}_i) - \mathbf{E}(e_N|\mathcal{A}_i) = 0,$$

The Z'_i s form a vertex exposure martingale sequence.

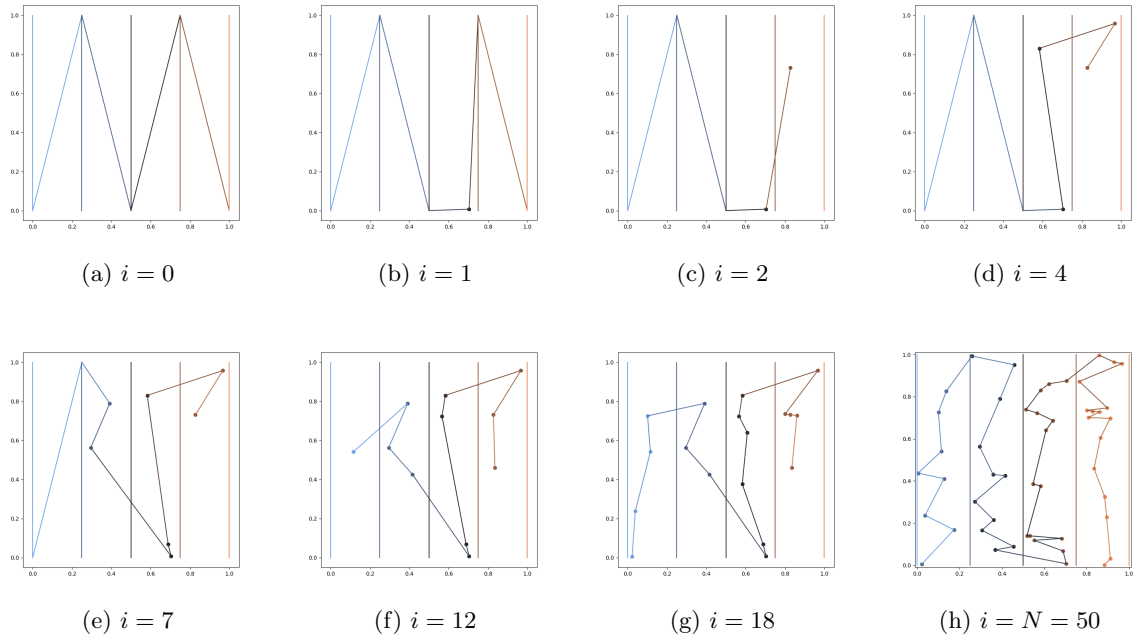


Figure 4.1: Evolution of the Martingale

Define $e_N^{[i]}$ as the distance of the trajectory when we remove the i -th point from the sample. Intuitively from the figure above and the triangle inequality, we can obtain

$$e_N^{[i]} \leq e_N \leq e_N + 2L/M,$$

4 Application to a Heuristic Algorithm for Travelling Salesman

269 meaning that revealing one point cannot increase more than 2 widths the distance of
270 the trajectory. Thus,

$$271 \quad \|Z_i\|_\infty = \sup_{X_1, \dots, X_N} \|\mathbf{E}(e_N|\mathcal{A}_i) - \mathbf{E}(e_N|\mathcal{A}_{i-1})\| \leq 2L/M.$$

272 On the other hand, by telescopic sums we obtain that

$$273 \quad e_N - Ee_N = \mathbf{E}(e_N|\mathcal{A}_N) - \mathbf{E}(e_N|\mathcal{A}_0) = \sum_{i=1}^N Z_i.$$

274 Therefore, by the Azuma-Hoeffding inequality,

$$275 \quad \mathbf{P}\{|e_N - Ee_N| > t\} \leq 2 \exp\left(\frac{-t^2}{2} \sum_{i=1}^N \|Z_i\|_\infty^2\right).$$

276 Finally,

$$277 \quad \sum_{i=1}^N \|Z_i\|_\infty^2 \leq \frac{4NL^2}{M^2},$$

278 which implies that

$$279 \quad \mathbf{P}\{|e_N - Ee_N| > t\} \leq 2 \exp\left(\frac{-t^2}{2} \sum_{i=1}^N \frac{4NL^2}{M^2}\right) \sim e^{-t^2 KN},$$

280 for some $K \in \mathbb{R}^+$.

Bibliography

281

- Miguel A Arcones and Evarist Giné. Limit theorems for u-processes. *The Annals of Probability*, pages 1494–1542, 1993. 282
283
- Mateo Díaz, Adolfo J Quiroz, and Mauricio Velasco. Local angles and dimension estimation from data on manifolds. *Journal of Multivariate Analysis*, 173:229–247, 2019. 284
285
- H Gzyl, R Jiménez, and AJ Quiroz. The physicist’s approach to the travelling salesman problem—ii. *Mathematical and Computer Modelling*, 13(7):45–48, 1990. 286
287