

Lecture 2: Concentration without Independence

Tianpei Xie

Jan. 6th., 2023

Contents

1	Martingale-based Methods	2
1.1	Martingale	2
1.2	Concentration Inequalities for Martingale Difference Sequences	4
1.2.1	Bernstein Inequality for Martingale Difference Sequence	4
1.2.2	Azuma-Hoeffding Inequality	5
1.2.3	McDiarmid's Inequality	5
1.2.4	Applications	5
2	The Efron-Stein Inequality	5
2.1	Bounding Variance	5
2.2	Functions with Bounded Differences	8
2.3	Self-Bounding Functions	8
2.4	Lipschitz Functions of Gaussian Variables	10
2.5	A Proof of the EfronStein Inequality Based on Duality	10

1 Martingale-based Methods

1.1 Martingale

- **Definition (*Martingale*)** [Resnick, 2013]

Let $\{X_n, n \geq 0\}$ be a stochastic process on (Ω, \mathcal{F}) and $\{\mathcal{F}_n, n \geq 0\}$ be a **filtration**; that is, $\{\mathcal{F}_n, n \geq 0\}$ is an *increasing sub σ -fields* of \mathcal{F}

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}.$$

Then $\{(X_n, \mathcal{F}_n), n \geq 0\}$ is a **martingale (mg)** if

1. X_n is **adapted** in the sense that for each n , $X_n \in \mathcal{F}_n$; that is, X_n is \mathcal{F}_n -measurable.
2. $X_n \in L_1$; that is $\mathbb{E}[|X_n|] < \infty$ for $n \geq 0$.
3. For $0 \leq m < n$

$$\mathbb{E}[X_n | \mathcal{F}_m] = X_m, \quad \text{a.s.} \quad (1)$$

If the equality of (1) is replaced by \geq ; that is, things are getting better on the average:

$$\mathbb{E}[X_n | \mathcal{F}_m] \geq X_m, \quad \text{a.s.} \quad (2)$$

then $\{X_n\}$ is called a **sub-martingale (submg)** while if things are getting worse on the average

$$\mathbb{E}[X_n | \mathcal{F}_m] \leq X_m, \quad \text{a.s.} \quad (3)$$

$\{X_n\}$ is called a **super-martingale (supermg)**.

- **Remark** $\{X_n\}$ is **martingale** if it is *both* a **sub** and **supermartingale**. $\{X_n\}$ is a **super-martingale** if and only if $\{-X_n\}$ is a **submartingale**.
- **Remark** If $\{X_n\}$ is a **martingale**, then $\mathbb{E}[X_n]$ is *constant*. In the case of a **submartingale**, the mean *increases* and for a **supermartingale**, the mean *decreases*.
- **Proposition 1.1** [Resnick, 2013]
If $\{(X_n, \mathcal{F}_n), n \geq 0\}$ is a **(sub, super) martingale**, then

$$\{(X_n, \sigma(X_0, X_1, \dots, X_n)), n \geq 0\}$$

is also a **(sub, super) martingale**.

- **Definition (*Martingale Differences*)**. [Resnick, 2013]
 $\{(d_j, \mathcal{B}_j), j \geq 0\}$ is a **(sub, super) martingale difference sequence** or a **(sub, super) fair sequence** if

1. For $j \geq 0$, $\mathcal{B}_j \subset \mathcal{B}_{j+1}$.
2. For $j \geq 0$, $d_j \in L_1$, $d_j \in \mathcal{B}_j$; that is, d_j is *absolutely integrable* and \mathcal{B}_j -measurable.
3. For $j \geq 0$,

$$\begin{aligned} \mathbb{E}[d_{j+1} | \mathcal{B}_j] &= 0, & (\text{martingale difference / fair sequence}); \\ &\geq 0, & (\text{submartingale difference / subfair sequence}); \\ &\leq 0, & (\text{supermartingale difference / supfair sequence}) \end{aligned}$$

- **Proposition 1.2** (*Construction of Martingale From Martingale Difference*) [Resnick, 2013]
If $\{(d_j, \mathcal{B}_j), j \geq 0\}$ is a *(sub, super) martingale difference sequence*, and

$$X_n = \sum_{j=0}^n d_j,$$

then $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a *(sub, super) martingale*.

- **Proposition 1.3** (*Construction of Martingale Difference From Martingale*) [Resnick, 2013]
Suppose $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a *(sub, super) martingale*. Define

$$\begin{aligned} d_0 &:= X_0 - \mathbb{E}[X_0] \\ d_j &:= X_j - X_{j-1}, \quad j \geq 1. \end{aligned}$$

Then $\{(d_j, \mathcal{B}_j), j \geq 0\}$ is a *(sub, super) martingale difference sequence*.

- **Proposition 1.4** (*Orthogonality of Martingale Differences*). [Resnick, 2013]
If $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale* where X_n can be decomposed as

$$X_n = \sum_{j=0}^n d_j,$$

d_j is \mathcal{B}_j -measurable and $\mathbb{E}[d_j^2] < \infty$ for $j \geq 0$, then $\{d_j\}$ are *orthogonal*:

$$\mathbb{E}[d_i d_j] = 0 \quad i \neq j.$$

- **Example** (*Smoothing as Martingale*)

Suppose $X \in L_1$ and $\{\mathcal{B}_n, n \geq 0\}$ is an increasing family of sub σ -algebra of \mathcal{B} . Define for $n \geq 0$

$$X_n := \mathbb{E}[X | \mathcal{B}_n].$$

Then (X_n, \mathcal{B}_n) is a *martingale*. From this result, we see that $\{(d_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale difference sequence* when

$$d_n := \mathbb{E}[X | \mathcal{B}_n] - \mathbb{E}[X | \mathcal{B}_{n-1}], \quad n \geq 1. \quad (4)$$

Proof: See that

$$\begin{aligned} \mathbb{E}[X_{n+1} | \mathcal{B}_n] &= \mathbb{E}[\mathbb{E}[X | \mathcal{B}_{n+1}] | \mathcal{B}_n] \\ &= \mathbb{E}[X | \mathcal{B}_n] \quad (\text{Smoothing property of conditional expectation}) \\ &= X_n \quad \blacksquare \end{aligned}$$

- **Example** (*Sums of Independent Random Variables*)

Suppose that $\{Z_n, n \geq 0\}$ is an *independent* sequence of integrable random variables satisfying for $n \geq 0$, $\mathbb{E}[Z_n] = 0$. Set

$$\begin{aligned} X_0 &:= 0, \\ X_n &:= \sum_{i=1}^n Z_i, \quad n \geq 1 \\ \mathcal{B}_n &:= \sigma(Z_0, \dots, Z_n). \end{aligned}$$

Then $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale* since $\{(Z_n, \mathcal{B}_n), n \geq 0\}$ is a *martingale difference sequence*.

• **Example (*Likelihood Ratios*).**

Suppose $\{Y_n, n \geq 0\}$ are *independent identically distributed* random variables and suppose the *true density* of Y_n is f_0 (The word “*density*” can be understood with respect to some fixed reference measure μ .) Let f_1 be *some other probability density*. For simplicity suppose $f_0(y) > 0$, for all y . For $n \geq 0$, define the likelihood ratio

$$X_n := \frac{\prod_{i=0}^n f_1(Y_i)}{\prod_{i=0}^n f_0(Y_i)}$$

$$\mathcal{B}_n := \sigma(Y_0, \dots, Y_n)$$

Then (X_n, \mathcal{B}_n) is a *martingale*.

Proof: See that

$$\begin{aligned} \mathbb{E}[X_{n+1} | \mathcal{B}_n] &= \mathbb{E}\left[\left(\frac{\prod_{i=0}^n f_1(Y_i)}{\prod_{i=0}^n f_0(Y_i)}\right) \frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} \mid Y_0, \dots, Y_n\right] \\ &= X_n \mathbb{E}\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} \mid Y_0, \dots, Y_n\right] \\ &= X_n \mathbb{E}\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right] \quad (\text{by independence}) \\ &:= X_n \int \frac{f_1(y_{n+1})}{f_0(y_{n+1})} f_0(y_{n+1}) d\mu(y_{n+1}) = X_n. \quad \blacksquare \end{aligned}$$

1.2 Concentration Inequalities for Martingale Difference Sequences

1.2.1 Bernstein Inequality for Martingale Difference Sequence

• **Proposition 1.5 (*Bernstein Inequality, Martingale Difference Sequence Version*)**
[Wainwright, 2019]

Let $\{(D_k, \mathcal{B}_k), k \geq 1\}$ be a *martingale difference sequence*, and suppose that

$$\mathbb{E}[\exp(\lambda D_k) | \mathcal{B}_{k-1}] \leq \exp\left(\frac{\lambda^2 \nu_k^2}{2}\right)$$

almost surely for any $|\lambda| < 1/\alpha_k$. Then the following hold:

1. The sum $\sum_{k=1}^n D_k$ is *sub-exponential* with **parameters** $\left(\sqrt{\sum_{k=1}^n \nu_k^2}, \alpha_*\right)$ where $\alpha_* := \max_{k=1, \dots, n} \alpha_k$.
2. The sum satisfies *the concentration inequality*

$$\mathbb{P}\left\{\left|\sum_{k=1}^n D_k\right| \geq t\right\} \leq \begin{cases} 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^n \nu_k^2}\right) & \text{if } 0 \leq t \leq \frac{\sum_{k=1}^n \nu_k^2}{\alpha_*} \\ 2 \exp\left(-\frac{t}{\alpha_*}\right) & \text{if } t > \frac{\sum_{k=1}^n \nu_k^2}{\alpha_*}. \end{cases} \quad (5)$$

1.2.2 Azuma-Hoeffding Inequality

- **Corollary 1.6** (*Azuma-Hoeffding Inequality, Martingale Difference*) [Wainwright, 2019]
Let $\{(D_k, \mathcal{B}_k), k \geq 1\}$ be a **martingale difference sequence** for which there are constants $\{(a_k, b_k)\}_{k=1}^n$ such that $D_k \in [a_k, b_k]$ almost surely for all $k = 1, \dots, n$. Then, for all $t \geq 0$,

$$\mathbb{P} \left\{ \left| \sum_{k=1}^n D_k \right| \geq t \right\} \leq 2 \exp \left(- \frac{2t^2}{\sum_{k=1}^n (b_k - a_k)^2} \right) \quad (6)$$

1.2.3 McDiarmid's Inequality

- An important application of *Azuma-Hoeffding Inequality* concerns functions that satisfy a *bounded difference property*.

Definition (*Functions with Bounded Difference Property*)

Given vectors $x, x' \in \mathcal{X}^n$ and an index $k \in \{1, 2, \dots, n\}$, we define a new vector $x^{(-k)} \in \mathcal{X}^n$ via

$$x_j^{(-k)} = \begin{cases} x_j & j \neq k \\ x'_k & j = k \end{cases}$$

With this notation, we say that $f : \mathcal{X}^n \rightarrow \mathbb{R}$ satisfies ***the bounded difference inequality*** with parameters (L_1, \dots, L_n) if, for each index $k = 1, 2, \dots, n$,

$$\left| f(x) - f(x^{(-k)}) \right| \leq L_k, \quad \text{for all } x, x' \in \mathcal{X}^n. \quad (7)$$

- **Corollary 1.7** (*McDiarmid's Inequality / Bounded Differences Inequality*) [Wainwright, 2019]
Suppose that f satisfies ***the bounded difference property*** (7) with parameters (L_1, \dots, L_n) and that the random vector $X = (X_1, X_2, \dots, X_n)$ has **independent** components. Then

$$\mathbb{P} \{ |f(X) - \mathbb{E}[f(X)]| \geq t \} \leq 2 \exp \left(- \frac{2t^2}{\sum_{k=1}^n L_k^2} \right). \quad (8)$$

1.2.4 Applications

2 The Efron-Stein Inequality

2.1 Bounding Variance

- **Remark** (*Variance of Independence Random Variables*)

Let $X_n = \sum_{i=1}^n Z_i$ be the sum of **independent** real-valued random variables Z_1, \dots, Z_n . Then we have

$$\begin{aligned} \mathbb{E} \left[(X_n - \mathbb{E}[X_n])^2 \right] &= \sum_{i=1}^n \mathbb{E} \left[(Z_i - \mathbb{E}[Z_i])^2 \right] \\ &\Rightarrow \text{Var}(X_n) = \sum_{i=1}^n \text{Var}(Z_i). \end{aligned}$$

- **Remark (Variance of Smoothing Martingale Difference Sequence)**

Suppose $X \in L_1$ and $\{\mathcal{B}_n, n \geq 0\}$ is an increasing family of sub σ -algebra of \mathcal{B} formed by

$$\mathcal{B}_n := \sigma(Z_1, \dots, Z_n).$$

For $n \geq 1$, define

$$\begin{aligned} d_0 &:= \mathbb{E}[X] \\ d_n &:= \mathbb{E}[X|\mathcal{B}_n] - \mathbb{E}[X|\mathcal{B}_{n-1}] \\ &= \mathbb{E}[X|Z_1, \dots, Z_n] - \mathbb{E}[X|Z_1, \dots, Z_{n-1}]. \end{aligned}$$

From (4) we see that (d_n, \mathcal{B}_n) is a martingale difference sequence. By *orthogonality of martingale difference*, we see that

$$\mathbb{E}[d_i d_j] = 0 \quad i \neq j.$$

Therefore, based on the decomposition

$$X - EX = \sum_{i=1}^n d_i$$

we have

$$\begin{aligned} \text{Var}(X) &= \mathbb{E} \left[\left(\sum_{i=1}^n d_i \right)^2 \right] = \sum_{i=1}^n \mathbb{E}[d_i^2] + 2 \sum_{i>j} \mathbb{E}[d_i d_j] \\ &= \sum_{i=1}^n \mathbb{E}[d_i^2]. \end{aligned} \tag{9}$$

- **Remark (Variance of General Functions of Independent Random Variables)**

Then above formula (9) holds when $X = f(Z_1, \dots, Z_n)$ for general function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ with n independent random variables (Z_1, \dots, Z_n) . By *Fubini's theorem*,

$$\mathbb{E}[X|Z_1, \dots, Z_i] = \int_{\mathcal{Z}^{n-i}} f(Z_1, \dots, Z_i, z_{i+1}, \dots, z_n) d\mu_{i+1}(z_{i+1}) \dots d\mu_n(z_n)$$

where μ_j is the probability distribution of Z_j for $j \geq 1$. Define the conditional expectation of X given all random variables (Z_1, \dots, Z_n) **except for** Z_i as

$$\begin{aligned} \mathbb{E}_{(-i)}[X] &:= \mathbb{E}[X|Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n] \\ &= \int_{\mathcal{Z}} f(Z_1, \dots, Z_{i-1}, z_i, Z_{i+1}, \dots, Z_n) d\mu_i(z_i). \end{aligned}$$

Then, again by *Fubini's theorem (smoothing properties of conditional expectation)*,

$$\mathbb{E}[\mathbb{E}_{(-i)}[X]|Z_1, \dots, Z_i] = \mathbb{E}[X|Z_1, \dots, Z_{i-1}] \tag{10}$$

Denote $Z_{(-i)} := (Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n)$.

• **Proposition 2.1 (Efron-Stein Inequality)** [Boucheron et al., 2013]

Let Z_1, \dots, Z_n be **independent random variables** and let $X = f(Z)$ be a square-integrable function of $Z = (Z_1, \dots, Z_n)$. Then

$$\text{Var}(X) \leq \sum_{i=1}^n \mathbb{E} \left[(X - \mathbb{E}_{(-i)}[X])^2 \right] := \nu. \quad (11)$$

Moreover, if Z'_1, \dots, Z'_n are **independent** copies of Z_1, \dots, Z_n and if we define, for every $i = 1, \dots, n$,

$$X'_i := f(Z_1, \dots, Z_{i-1}, Z'_i, Z_{i+1}, \dots, Z_n),$$

then

$$\nu = \frac{1}{2} \sum_{i=1}^n \mathbb{E} \left[(X - X'_i)^2 \right] = \sum_{i=1}^n \mathbb{E} \left[(X - X'_i)_+^2 \right] = \sum_{i=1}^n \mathbb{E} \left[(X - X'_i)_-^2 \right]$$

where $x_+ = \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$ denote the **positive** and **negative** parts of a real number x . Also,

$$\nu = \inf_{X_i} \sum_{i=1}^n \mathbb{E} \left[(X - X_i)^2 \right],$$

where the infimum is taken over the class of all $Z_{(-i)}$ -measurable and square-integrable variables X_i , $i = 1, \dots, n$.

Proof: We begin with the proof of the first statement. Note that, using (10), we may write

$$\begin{aligned} d_i &:= \mathbb{E}[X | Z_1, \dots, Z_i] - \mathbb{E}[X | Z_1, \dots, Z_{i-1}] \\ &= \mathbb{E}[X | Z_1, \dots, Z_i] - \mathbb{E}[\mathbb{E}_{(-i)}[X] | Z_1, \dots, Z_i] \\ &= \mathbb{E}[X - \mathbb{E}_{(-i)}[X] | Z_1, \dots, Z_i]. \end{aligned}$$

By *Jensen's inequality* used conditionally,

$$d_i^2 \leq \mathbb{E} \left[(X - \mathbb{E}_{(-i)}[X])^2 | Z_1, \dots, Z_i \right]$$

Using (9) $\text{Var}(X) = \sum_{i=1}^n \mathbb{E} [d_i^2]$, we have

$$\text{Var}(X) \leq \sum_{i=1}^n \mathbb{E} \left[\mathbb{E} \left[(X - \mathbb{E}_{(-i)}[X])^2 | Z_1, \dots, Z_i \right] \right] = \sum_{i=1}^n \mathbb{E} \left[(X - \mathbb{E}_{(-i)}[X])^2 \right],$$

we obtain the desired inequality.

To prove the identities for ν , denote by $\text{Var}_{(-i)}$ the *conditional variance operator* conditioned on $Z_{(-i)} := (Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n)$. Then we may write ν as

$$\nu = \sum_{i=1}^n \mathbb{E} [\text{Var}_{(-i)}(X)].$$

Now note that one may simply use (conditionally) the elementary fact that if X and Y are *independent and identically distributed* real-valued random variables, then

$$\text{Var}(X) = \frac{1}{2} \mathbb{E} [(X - Y)^2] .$$

Since conditionally on $Z_{(-i)}$, X'_i is an independent copy of X , we may write

$$\text{Var}_{(i)}(X) = \frac{1}{2} \mathbb{E}_{(-i)} [(X - X'_i)^2] = \sum_{i=1}^n \mathbb{E}_{(-i)} [(X - X'_i)^2_+] = \sum_{i=1}^n \mathbb{E}_{(-i)} [(X - X'_i)^2_-] ,$$

where we used the fact that the conditional distributions of X and X'_i are *identical*.

The last identity is obtained by recalling that, for any real-valued random variable X ,

$$\text{Var}(X) = \inf_{a \in \mathbb{R}} \mathbb{E} [(X - a)^2] .$$

Using this fact conditionally, we have, for every $i = 1, \dots, n$,

$$\text{Var}_{(-i)}(X) = \inf_{X_i} \mathbb{E}_{(-i)} [(X - X_i)^2] .$$

Note that this infimum is achieved whenever $X_i = \mathbb{E}_{(-i)} [X]$. ■

2.2 Functions with Bounded Differences

- **Remark** Recall that a function $f : \mathcal{X}^n \rightarrow \mathbb{R}$ satisfies *the bounded difference inequality* with parameters (L_1, \dots, L_n) if, for each index $k = 1, 2, \dots, n$,

$$\left| f(x) - f(x^{(-k)}) \right| \leq L_k, \quad \text{for all } x, x' \in \mathcal{X}^n .$$

where

$$x_j^{(-k)} = \begin{cases} x_j & j \neq k \\ x'_k & j = k \end{cases}$$

- **Corollary 2.2** [Boucheron et al., 2013]
If f has the **bounded differences property** with parameters (L_1, \dots, L_n) , then

$$\text{Var}(f(X)) \leq \frac{1}{4} \sum_{i=1}^n L_i^2 .$$

2.3 Self-Bounding Functions

- Another simple property which is satisfied for many important examples is the so-called *self-bounding property*.

Definition (*Self-Bounding Property*)

A **nonnegative function** $f : \mathcal{X}^n \rightarrow [0, \infty)$ has the **self-bounding property** if there exist functions $f_i : \mathcal{X}^{n-1} \rightarrow \mathbb{R}$ such that for all $x_1, \dots, x_n \in \mathcal{X}$ and all $i = 1, \dots, n$,

$$0 \leq f(x_1, \dots, x_n) - f_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \leq 1 \quad (12)$$

and also

$$\sum_{i=1}^n (f(x_1, \dots, x_n) - f_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)) \leq f(x_1, \dots, x_n). \quad (13)$$

- **Remark** Clearly if f has the ***self-bounding property***,

$$\sum_{i=1}^n (f(x_1, \dots, x_n) - f_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n))^2 \leq f(x_1, \dots, x_n) \quad (14)$$

- **Corollary 2.3** [Boucheron et al., 2013]

If f has the ***self-bounding property***, then

$$\text{Var}(f(X)) \leq \mathbb{E}[f(X)].$$

- **Remark (*Relative Stability*)** [Boucheron et al., 2013]

A sequence of nonnegative random variables $(Z_n)_{n \in \mathbb{N}}$ is said to be **relatively stable** if

$$\frac{Z_n}{\mathbb{E}[Z_n]} \xrightarrow{\mathbb{P}} 1.$$

This property guarantees that ***the random fluctuations of Z_n around its expectation are of negligible size when compared to the expectation, and therefore most information about the size of Z_n is given by $\mathbb{E}[Z_n]$.***

Bounding the variance of Z_n by its expected value implies, in many cases, the relative stability of $(Z_n)_{n \in \mathbb{N}}$. If Z_n has the ***self-bounding property***, then, by *Chebyshev's inequality*, for all $\epsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{Z_n}{\mathbb{E}[Z_n]} - 1\right| > \epsilon\right\} \leq \frac{\text{Var}(Z_n)}{\epsilon^2(\mathbb{E}[Z_n])^2} \leq \frac{1}{\epsilon^2 \mathbb{E}[Z_n]}.$$

Thus, for relative stability, it suffices to have $\mathbb{E}[Z_n] \rightarrow \infty$.

- An important class of functions satisfying ***the self-bounding property*** consists of the so-called ***configuration functions***.

Definition (*Configuration Function*)

Assume that we have a property Π ***defined over the union of finite products*** of a set \mathcal{X} , that is, a sequence of sets

$$\Pi_1 \subset \mathcal{X}, \Pi_2 \subset \mathcal{X} \times \mathcal{X}, \dots, \Pi_n \subset \mathcal{X}^n.$$

We say that $(x_1, \dots, x_m) \in \mathcal{X}^m$ ***satisfies the property Π*** if $(x_1, \dots, x_m) \in \Pi_m$.

We assume that Π is **hereditary** in the sense that if (x_1, \dots, x_m) satisfies Π then so does ***any sub-sequence*** $\{x_{i_1}, \dots, x_{i_k}\}$ of (x_1, \dots, x_m) .

The function f that maps any vector $x = (x_1, \dots, x_n)$ to ***the size of a largest sub-sequence satisfying Π*** is **the configuration function** associated with property Π .

- **Corollary 2.4** [Boucheron et al., 2013]

Let f be a **configuration function**, and let $Z = f(X_1, \dots, X_n)$, where X_1, \dots, X_n are **independent** random variables. Then

$$\text{Var}(Z) \leq \mathbb{E}[Z].$$

- **Example (VC Dimension)**

Let \mathcal{H} be an arbitrary collection of subsets of \mathcal{X} , and let $x = (x_1, \dots, x_n)$ be a vector of n points of \mathcal{X} . Define the **trace** of \mathcal{H} on x by

$$\text{tr}(x) = \{A \cap \{x_1, \dots, x_n\} : A \in \mathcal{H}\}.$$

The shatter coefficient, (or *Vapnik-Chervonenkis growth function*) of \mathcal{H} in x is $\tau_{\mathcal{H}}(x) = |\text{tr}(x)|$, the size of the trace. $\tau_{\mathcal{H}}(x)$ is the number of different subsets of the n -point set $\{x_1, \dots, x_n\}$ generated by intersecting it with elements of \mathcal{H} . A subset $\{x_{i_1}, \dots, x_{i_k}\}$ of $\{x_1, \dots, x_n\}$ is said to be **shattered** if $2^k = T(x_{i_1}, \dots, x_{i_k})$.

The VC dimension $D(x)$ of \mathcal{H} (with respect to x) is the cardinality k of the largest shattered subset of x . From the definition it is obvious that $f(x) = D(x)$ is a **configuration function** (associated with the property of “**shatteredness**”) and therefore if X_1, \dots, X_n are *independent random variables*, then

$$\text{Var}(D(X)) \leq \mathbb{E}[D(X)].$$

2.4 Lipschitz Functions of Gaussian Variables

2.5 A Proof of the EfronStein Inequality Based on Duality

References

- Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- Sidney I Resnick. *A probability path*. Springer Science & Business Media, 2013.
- Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.