

Non-Asymptotic Bounds

패턴인식 및 기계학습 여름학교
(Pattern Recognition and Machine Learning Summer School)
2017. 7. 12 (Wed.)

Yung-Kyun Noh (노영균)
Seoul National University

Many slides are borrowed and modified from Gábor Lugosi's concentration inequalities lecture slides!



<http://machinelearning.snu.ac.kr/PRMLSS2017/nonasymptotic.pdf>



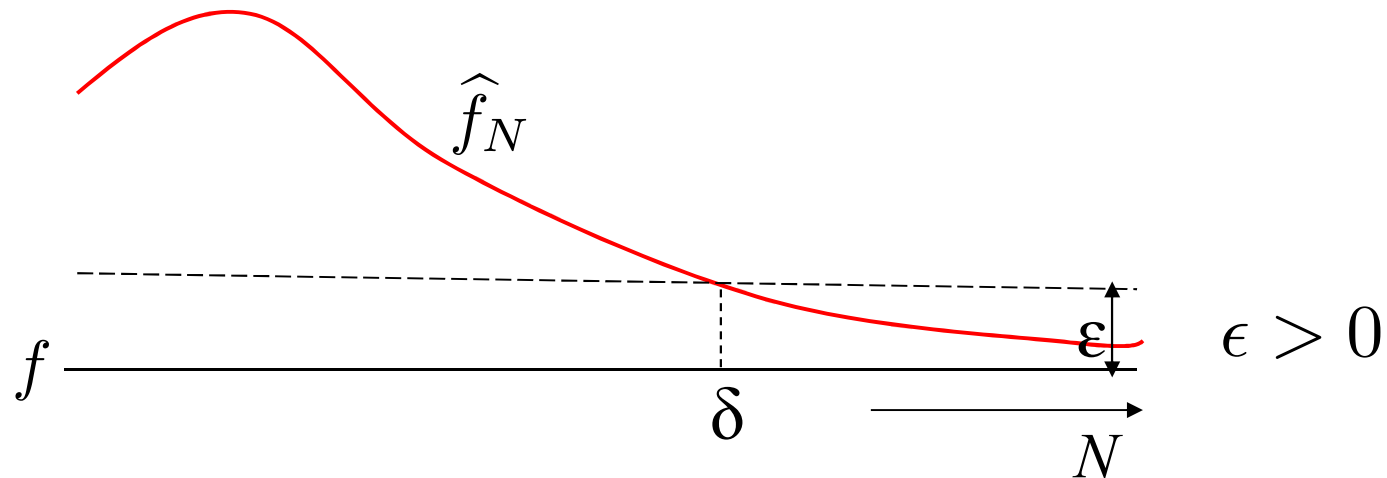
Contents

- Asymptotic Convergence
 - Converge in probability
 - Almost sure convergence
 - L2 convergence
 - Maximum likelihood
- Non-Asymptotic Bounds
 - Markov's inequality
 - Chernoff bound
 - Chebyshev's inequality
 - Efron-Stein inequality



Asymptotic / Non-asymptotic Bounds

- Limit, epsilon and delta $\lim_{N \rightarrow \infty} \hat{f}_N = f$



– For any ϵ , there exists δ such that

$$|\hat{f}_N - f| < \epsilon$$

Machine Learning and Consistency

- With increasing number of data
 - Expected loss

$$L = \mathbb{E}_P[l(y, f(\mathbf{x}))] \quad l(y, y'): \text{loss function}$$

- Estimated Error

$$\hat{L} = \sum_n l(y_n, f(\mathbf{x}_n)), \quad f(\mathbf{x}) \in \mathcal{H}$$

\mathcal{H} satisfies

$$\hat{L} \xrightarrow{N \rightarrow \infty} L$$

$$P\{\sup_{f \in \mathcal{H}} (L(f) - \hat{L}(f)) > \epsilon\} \rightarrow 0 \quad \text{for } \epsilon > 0$$

<Uniform convergence>

Interest in This Lecture

- We are interested in bounding random fluctuations of functions of many independent random variables.
- Let $f : \mathcal{X}^n \rightarrow \mathbb{R}$ and

$$Z = f(\mathbf{x}_1, \dots, \mathbf{x}_N)$$

with independent random variables $\mathbf{x}_1, \dots, \mathbf{x}_N$.



Interest in This Lecture

- The estimator will be close enough to the true value (or expectation of the estimator).

– **Asymptotically**

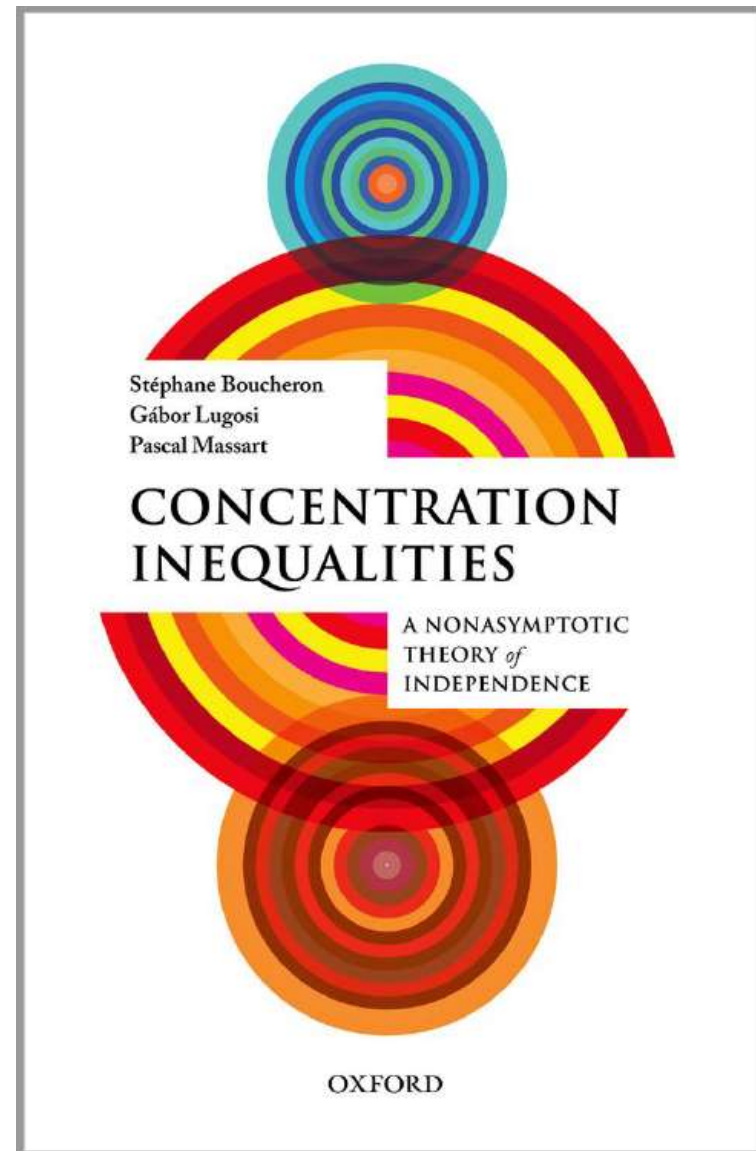
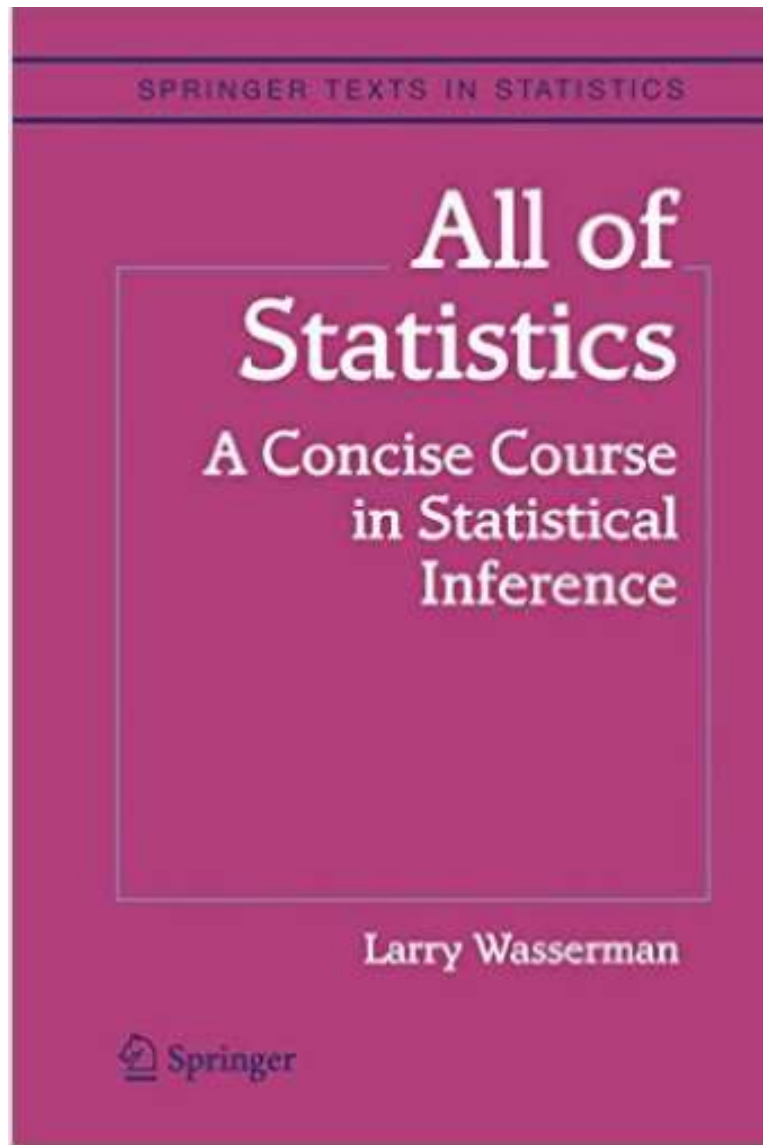
$$\hat{f}_N \xrightarrow[N \rightarrow \infty]{} f$$

– **Non-asymptotically (with fixed N)**

$$P\{Z > \mathbb{E}Z + t\} \text{ and } P\{Z < \mathbb{E}Z - t\}$$

for $t > 0$





ASYMPTOTIC CONVERGENCE



Convergence

- Convergence in probability: $Z_N \xrightarrow{P} Z$

$$P \{ |Z_N - Z| > \epsilon \} \rightarrow 0 \quad (N \rightarrow \infty, \epsilon > 0)$$

- Almost sure convergence: $Z_N \xrightarrow{as} Z$

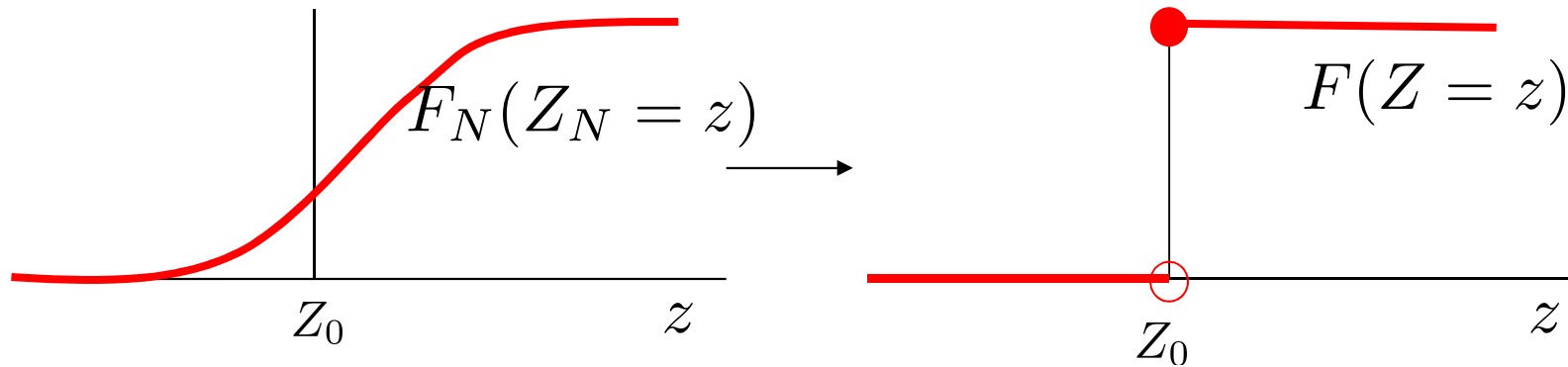
$$P \left(\lim_{N \rightarrow \infty} Z_N = Z \right) = 1$$

- L2-convergence: $Z_N \xrightarrow{qm} Z$

$$\mathbb{E}(Z_N - Z)^2 \rightarrow 0 \quad (N \rightarrow \infty)$$

Convergence

- Convergence in distribution $Z_N \rightsquigarrow Z$

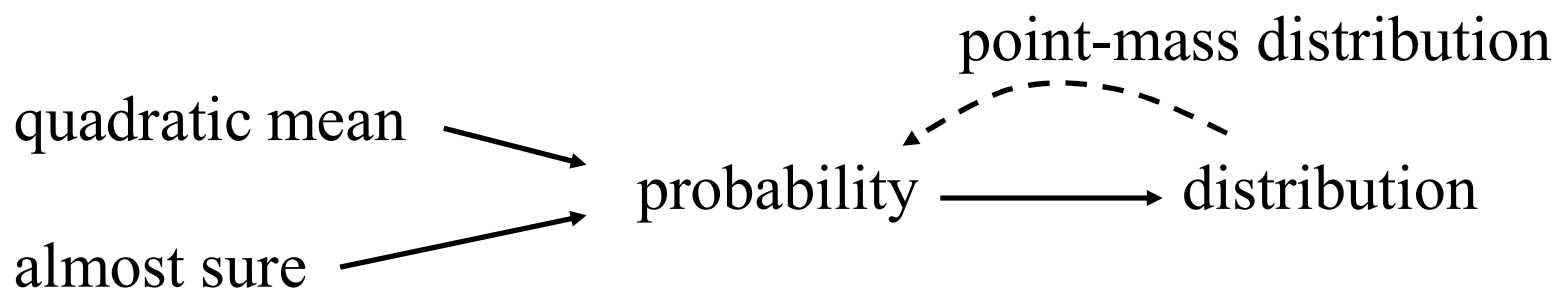


$$\lim_{N \rightarrow \infty} F_N(z) = F(z)$$

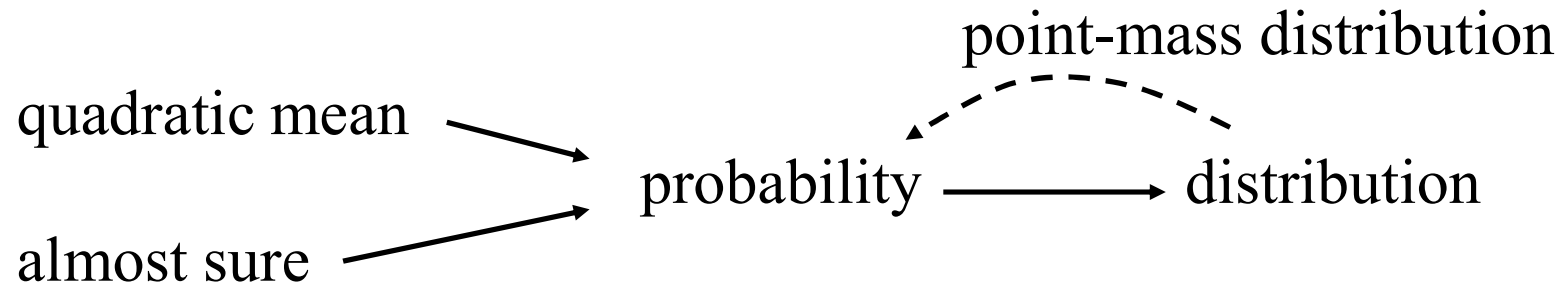
For this distribution, convergence in distribution implies

$$Z_N \xrightarrow{P} Z (= Z_0)$$

Relationship Between Types of Convergence



Relationship Between Types of Convergence



- Ex) $Z_N \rightsquigarrow Z$ & $P(Z = Z_0) = 1$
 $\longrightarrow Z_N \xrightarrow{P} Z (= Z_0)$

$$\begin{aligned}
 P(|Z_N - Z_0| > \epsilon) &= P(Z_N < Z_0 - \epsilon) + P(Z_N > Z_0 + \epsilon) \\
 &\leq P(Z_N \leq Z_0 - \epsilon) + P(Z_N > Z_0 + \epsilon) \\
 &= F_N(Z_0 - \epsilon) + 1 - F_N(Z_0 + \epsilon) \\
 &\rightarrow F(Z_0 - \epsilon) + 1 - F(Z_0 + \epsilon) \\
 &= 0 + 1 - 1 = 0
 \end{aligned}$$

Convergence in Probability

$$X(s) = s \quad s \sim \text{Unif}[0, 1]$$

A sequence X_1, X_2, \dots is defined as follows:

$$\begin{aligned} X_1(s) &= s + I_{[0,1]}(s) & X_2(s) &= s + I_{[0, \frac{1}{2}]}(s) \\ X_3(s) &= s + I_{[\frac{1}{2}, 1]}(s) & X_4(s) &= s + I_{[0, \frac{1}{3}]}(s) \\ X_5(s) &= s + I_{[\frac{1}{3}, \frac{2}{3}]}(s) & X_6(s) &= s + I_{[\frac{2}{3}, 1]}(s) \\ &\dots & & \end{aligned}$$



$$P(|X_N - X| \geq \epsilon) = 0 \quad (N \rightarrow \infty)$$

Convergence in Probability

$$X_1(s) = s + I_{[0,1]}(s)$$

$$X_2(s) = s + I_{[0, \frac{1}{2}]}(s)$$

$$X_3(s) = s + I_{[\frac{1}{2}, 1]}(s)$$

$$X_4(s) = s + I_{[0, \frac{1}{3}]}(s)$$

$$X_5(s) = s + I_{[\frac{1}{3}, \frac{2}{3}]}(s)$$

$$X_6(s) = s + I_{[\frac{2}{3}, 1]}(s)$$

...

- If $s = 3/8$,

$$X_1(s) = 11/8 \quad X_2(s) = 11/8 \quad X_3(s) = 3/8 \quad X_4(s) = 3/8$$

$$X_5(s) = 11/8 \quad X_6(s) = 3/8 \quad \dots$$

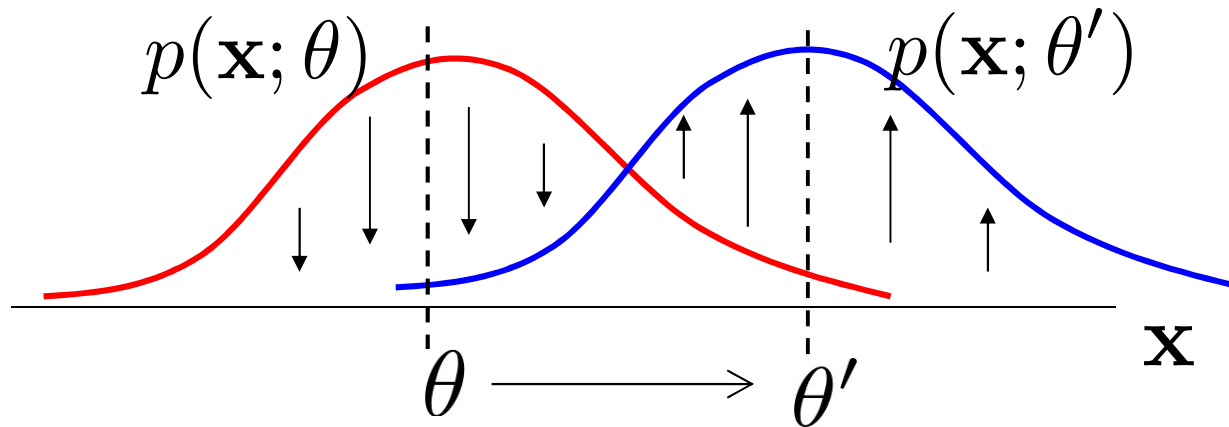
For every s , $X_N(s)$ alternates between s and $s+1$ infinitely often.

$$\lim_{N \rightarrow \infty} X_N \neq X \quad \text{Not a.s. convergence}$$



Convergence of Maximum Likelihood Estimation

- Sensitivity of probability density function with respect to the parameter θ : $\frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta}$ ($= \nabla_{\theta} \log p(\mathbf{x}; \theta)$)



We note

$$\begin{aligned} \mathbb{E}_{\mathbf{x} \sim p} [\nabla_{\theta} \log p(\mathbf{x}; \theta)] &= \int p(\mathbf{x}; \theta) \nabla_{\theta} \log p(\mathbf{x}; \theta) d\mathbf{x} \\ &= \int p(\mathbf{x}; \theta) \frac{\nabla_{\theta} p(\mathbf{x}; \theta)}{p(\mathbf{x}; \theta)} d\mathbf{x} = \nabla_{\theta} \int p(\mathbf{x}; \theta) d\mathbf{x} = \nabla_{\theta} 1 = 0 \end{aligned}$$

Convergence of MLE

- Fisher information

$$\begin{aligned} I(\theta) &= \mathbb{E}_{\mathbf{x} \sim p} \left[(\nabla_{\theta} \log p(\mathbf{x}; \theta))^2 \right] \\ &= \text{Var} (\nabla_{\theta} \log p(\mathbf{x}; \theta)) \\ &= - \int (\nabla_{\theta}^2 \log p(\mathbf{x}; \theta)) p(\mathbf{x}; \theta) d\mathbf{x} \end{aligned}$$

- Asymptotic Normality of MLE

$$\begin{aligned} \frac{(\hat{\theta}_N - \theta)}{\sqrt{1/(N \cdot I(\theta))}} &\rightsquigarrow \mathcal{N}(0, 1) && \left(\text{or } \frac{(\hat{\theta}_N - \theta)}{\sqrt{1/(N \cdot I(\hat{\theta}_N))}} \rightsquigarrow \mathcal{N}(0, 1) \right) \\ \text{se}(\hat{\theta}_N) &= \sqrt{1/(N \cdot I(\theta))} && \left(\text{or } \hat{\text{se}} = \sqrt{1/(N \cdot I(\hat{\theta}_N))} \right) \end{aligned}$$

Convergence of MLE

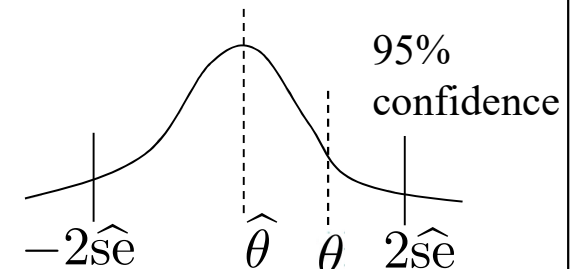
- **Ex)** $X_1, \dots, X_N \sim \text{Bernoulli}(\theta)$

$$p(x; \theta) = \theta^x (1 - \theta)^{1-x} \quad \hat{\theta}_N = \sum_{i=1}^N \frac{X_i}{N}$$

$$-\frac{\partial^2 \log p(x; \theta)}{\partial \theta^2} = \frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

$$I(\theta) = \mathbb{E}_{x \sim p} \left[-\frac{\partial^2 \log p(x; \theta)}{\partial \theta^2} \right] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$$

$$\widehat{\text{se}}(\hat{\theta}) = \frac{1}{\sqrt{N \cdot I(\hat{\theta})}} = \left(\frac{\hat{\theta}(1-\hat{\theta})}{N} \right)^{\frac{1}{2}}$$



NON-ASYMPTOTIC CONVERGENCE



Markov's Inequality

- If $Z \geq 0$

$$P(Z > t) \leq \frac{\mathbb{E}Z}{t}$$



Markov's Inequality

- If $Z \geq 0$

$$P(Z > t) \leq \frac{\mathbb{E}Z}{t}$$

$$\begin{aligned}\mathbb{E}Z &= \int_0^{\infty} ZP(Z)dZ \\ &= \int_0^t ZP(Z)dZ + \int_t^{\infty} ZP(Z)dZ \\ &\geq \int_t^{\infty} ZP(Z)dZ \geq \int_t^{\infty} tP(Z)dZ \\ &= t \int_t^{\infty} P(Z)dZ = tP(Z > t)\end{aligned}$$

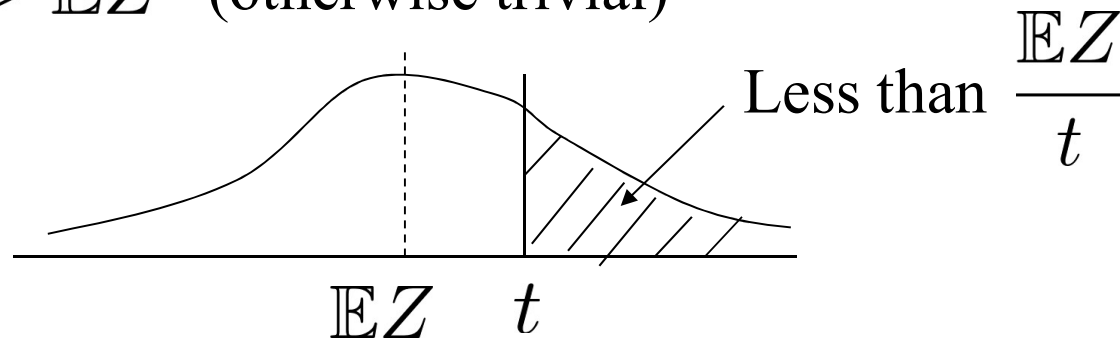


Markov's Inequality

- If $Z \geq 0$

$$P(Z > t) \leq \frac{\mathbb{E}Z}{t}$$

$t > \mathbb{E}Z$ (otherwise trivial)

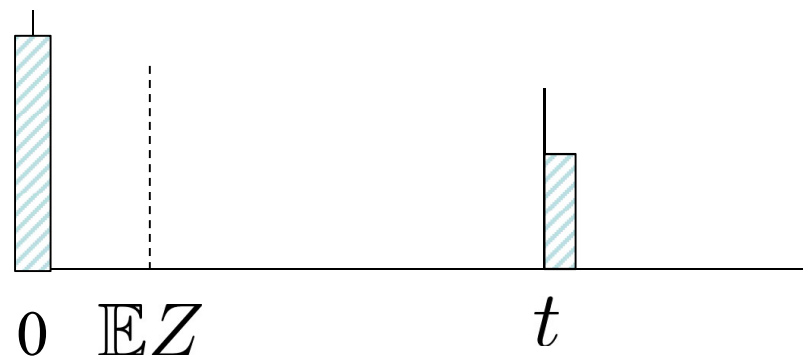


Markov's Inequality

- If $Z \geq 0$

$$P(Z > t) \leq \frac{\mathbb{E}Z}{t}$$

- The bound is tight when



Chebyshev's Inequality

$$\text{Var}(Z) = \mathbb{E}(Z - \mathbb{E}Z)^2$$

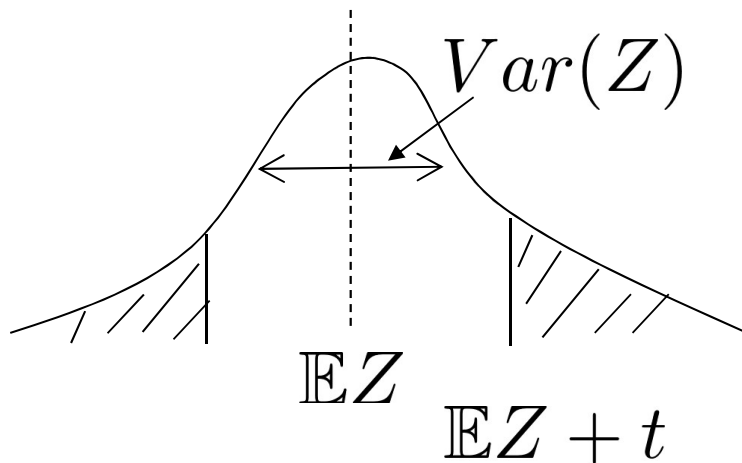
$$P(|Z - \mathbb{E}Z| > t) = P((Z - \mathbb{E}Z)^2 > t^2) \leq \frac{\text{Var}(Z)}{t^2}$$



Chebyshev's Inequality

$$\text{Var}(Z) = \mathbb{E}(Z - \mathbb{E}Z)^2$$

$$P(|Z - \mathbb{E}Z| > t) = P((Z - \mathbb{E}Z)^2 > t^2) \leq \frac{\text{Var}(Z)}{t^2}$$



Meaningless if
 $t^2 < \text{Var}(Z)$

Example of Concentration (Chebyshev)

- If we are interested in

$$Z = \frac{1}{N} \sum_{i=1}^N X_i$$

with independent R.V.s X_1, \dots, X_N

Example of Concentration (Chebyshev)

- If we are interested in

$$Z = \frac{1}{N} \sum_{i=1}^N X_i$$

with independent R.V.s X_1, \dots, X_N

$$\text{Var}(\sum X_i) = \sum_{i=1}^N \text{Var}(X_i) = N \text{Var}(X_1)$$

$$P \left(\left| \sum_{i=1}^N X_i - N \mathbb{E}X_1 \right| > t \right) \leq \frac{N \text{Var}(X_1)}{t^2}$$

Example of Concentration (Chebyshev)

- We are interested in

$$Z = \frac{1}{N} \sum_{i=1}^N X_i$$

with independent R.V.s X_1, \dots, X_N

- Equivalently,

$$P \left(\left| \frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}X_1 \right| > \frac{t}{\sqrt{N}} \right) \leq \frac{\text{Var}(X_1)}{t^2}$$

$$P \left(\left| \frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}X_1 \right| > t \right) \leq \frac{\text{Var}(X_1)}{Nt^2}$$

Chernoff Bounds

- Motivation:
 - Central limit theorem:

$$X \sim \mathcal{N}(0, \sigma^2)$$

$$\sqrt{N} \left(\frac{1}{N} \sum_{i=1}^N X_i - \mu \right) \rightsquigarrow X$$

Chernoff Bounds

- Motivation:
 - Central limit theorem:

$$\lim_{N \rightarrow \infty} P \left\{ \frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}X_i > \frac{t}{\sqrt{N}} \right\} = 1 - \Psi(t / \sqrt{\text{Var}(X_1)})$$

$$\Psi(\mathbf{x}) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mathbf{x}^2} d\mathbf{x}$$

Chernoff Bounds

- If we use

$$\int_a^\infty e^{-\frac{x^2}{2}} dx \leq \frac{e^{-\frac{a^2}{2}}}{a}$$

we obtain

$$\begin{aligned} \lim_{N \rightarrow \infty} P \left\{ \frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}X_i > \frac{t}{\sqrt{N}} \right\} &\leq \frac{\sqrt{\text{Var}(X_1)}}{\sqrt{2\pi t}} e^{-\frac{t^2}{2\text{Var}(X_1)}} \\ &\leq e^{-\frac{t^2}{2\text{Var}(X_1)}} \\ &\quad \left(t \text{ for } \frac{\sqrt{\text{Var}(X_1)}}{t} < 1 \right) \end{aligned}$$

Chernoff Bounds

- Chernoff Bounds

$$\begin{aligned} P \{ Z - \mathbb{E}Z > t \} &= P \left\{ e^{\lambda(Z - \mathbb{E}Z)} > e^{\lambda t} \right\} && \lambda > 0 \\ &\leq \frac{\mathbb{E}e^{\lambda(Z - \mathbb{E}Z)}}{e^{\lambda t}} \end{aligned}$$

Example of Concentration (Chernoff)

$$Z = \frac{1}{N} \sum X_i$$

$$\mathbb{E}e^{\lambda \sum X_i} = \mathbb{E} \prod_{i=1}^N e^{\lambda X_i} = \prod_{i=1}^N \mathbb{E}e^{\lambda X_i}$$

↑
Independence

From Hoeffding's inequality,

$$X_1, \dots, X_N \in [0, 1] \quad \mathbb{E}e^{\lambda(X_i - \mathbb{E}X_i)} \leq e^{\lambda^2/8}$$

$$\frac{\mathbb{E}e^{\lambda(X_i - \mathbb{E}X_i)}}{e^{\lambda t}} \leq \frac{e^{\lambda^2/8}}{e^{\lambda t}} = e^{\frac{\lambda^2}{8} - \lambda t}$$

$$\min_{\lambda} e^{\frac{\lambda^2}{8} - \lambda t} = e^{-2t^2}$$



Example of Concentration (Chernoff)

$$P \left\{ \left| \frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N X_i \right] \right| > t \right\} \leq 2e^{-2nt^2}$$



Martingale and Efron-Stein Inequality

$$\mathbb{E}_i[Z] = \mathbb{E}[Z | X_1, \dots, X_i]$$

$$\mathbb{E}_0[Z] = \mathbb{E}[Z]$$

$$Z = f(X_1, \dots, X_N)$$

Ex)

$$Z = X_1 + \dots + X_N$$

For $\mathbb{E}_i[Z]$

$$Z = \underbrace{X_1 + \dots + X_i}_{\text{fixed}} + \underbrace{X_{i+1} + \dots + X_N}_{\text{R.V.}}$$



Martingale and Efron-Stein Inequality

$$\begin{aligned}\mathbb{E}_i[Z] &= \int f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) \\ &\quad d\mathbf{x}_{i+1} \dots d\mathbf{x}_N \\ &= g_i(\mathbf{x}_1, \dots, \mathbf{x}_i) \\ &\quad \rightarrow \text{function of } \mathbf{x}_1, \dots, \mathbf{x}_i\end{aligned}$$

Martingale and Efron-Stein Inequality

- Exercise

$$j > i, \quad \mathbb{E}_i \mathbb{E}_j Z = \mathbb{E}_i Z$$

$$\mathbb{E}_i \mathbb{E}_j Z = \mathbb{E}_i \int \underline{f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_{j+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_j) d\mathbf{x}_{j+1} \dots d\mathbf{x}_N}$$

→ function of $\mathbf{x}_1, \dots, \mathbf{x}_j$

$$= \mathbb{E}_i g_j(\mathbf{x}_1, \dots, \mathbf{x}_j)$$

$$= \int g_j(\mathbf{x}_1, \dots, \mathbf{x}_j) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_N$$

$$= \int g_j(\mathbf{x}_1, \dots, \mathbf{x}_j) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_j | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_j$$

Martingale and Efron-Stein Inequality

$$= \iint f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_{j+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_j) d\mathbf{x}_{j+1} \dots d\mathbf{x}_N \\ p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_j | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_j$$

$$= \int f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_{j+1}, \mathbf{x}_j, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) \\ d\mathbf{x}_{i+1} \dots d\mathbf{x}_j d\mathbf{x}_{j+1} \dots d\mathbf{x}_N$$

$$= \mathbb{E}_i Z$$



Martingale and Efron-Stein Inequality

- Doob (Joseph Leo Doob) martingale representation

$$\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z$$

$$\begin{aligned} \sum_{i=1}^N \Delta_i &= \mathbb{E}_N Z - \mathbb{E}_{N-1} Z \\ &\quad + \mathbb{E}_{N-1} Z - \mathbb{E}_{N-2} Z + \dots \\ &\quad + \mathbb{E}_1 Z - \mathbb{E} Z \\ &= \mathbb{E}_N Z - \mathbb{E} Z \\ &= Z - \mathbb{E} Z \end{aligned}$$

Martingale and Efron-Stein Inequality

$$\begin{aligned} \text{Var}(Z) &= \mathbb{E} \left[(Z - \mathbb{E}Z)^2 \right] \\ &= \mathbb{E} \left[\left(\sum_{i=1}^N \Delta_i \right)^2 \right] \\ &= \sum_{i=1}^N \mathbb{E}[\Delta_i^2] + 2 \sum_{i,j;j>i} \mathbb{E}[\Delta_i \Delta_j] \end{aligned}$$

$$\mathbb{E}[\Delta_i \Delta_j]?$$



Martingale and Efron-Stein Inequality

$$\begin{aligned}
 \mathbb{E}[\Delta_i \Delta_j] &= \int \Delta_i \Delta_j p(\mathbf{x}_1, \dots, \mathbf{x}_N) d\mathbf{x}_1 \dots d\mathbf{x}_N \\
 &= \iint \Delta_i \Delta_j p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_N \\
 &\quad p(\mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_1 \dots d\mathbf{x}_i \\
 &= \int \underline{\mathbb{E}_i[\Delta_i \Delta_j]} p(\mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_1 \dots d\mathbf{x}_i
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{E}_i[\Delta_i \Delta_j] &= \\
 &\int \underbrace{g_i(\mathbf{x}_1, \dots, \mathbf{x}_i)}_{\text{function of } \mathbf{x}_1, \dots, \mathbf{x}_i} \underbrace{g_j(\mathbf{x}_1, \dots, \mathbf{x}_j)}_{\text{function of } \mathbf{x}_1, \dots, \mathbf{x}_j} p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_N \\
 &= g_i(\mathbf{x}_1, \dots, \mathbf{x}_i) \int g_j(\mathbf{x}_1, \dots, \mathbf{x}_j) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_N \\
 &= \Delta_i \mathbb{E}_i[\Delta_j]
 \end{aligned}$$

Martingale and Efron-Stein Inequality

$$j > i,$$

$$\mathbb{E}_i \Delta_j = \mathbb{E}_i (\mathbb{E}_j Z - \mathbb{E}_{j-1} Z)$$

$$= \mathbb{E}_i (\mathbb{E} f(\mathbf{x}_{j+1}, \dots, \mathbf{x}_N) - \mathbb{E} f(\mathbf{x}_j, \dots, \mathbf{x}_N))$$

$$= \int (g_j(\mathbf{x}_1, \dots, \mathbf{x}_j) - g_{j-1}(\mathbf{x}_1, \dots, \mathbf{x}_{j-1}))$$

$$p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_N$$

$$= \int g_j(\mathbf{x}_1, \dots, \mathbf{x}_j) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_j | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_j$$

$$- \int g_{j-1}(\mathbf{x}_1, \dots, \mathbf{x}_{j-1}) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_{j-1} | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_{j-1}$$



Martingale and Efron-Stein Inequality

$$\begin{aligned} &= \iint f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_{j+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_j) \\ &\quad p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_j | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_j d\mathbf{x}_{j+1} \dots d\mathbf{x}_N \\ &- \iint f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_j, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_{j-1}) \\ &\quad p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_{j-1} | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_{j-1} d\mathbf{x}_j \dots d\mathbf{x}_N \\ &= \int f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_N \\ &\quad - \int f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_{i+1}, \dots, \mathbf{x}_N | \mathbf{x}_1, \dots, \mathbf{x}_i) d\mathbf{x}_{i+1} \dots d\mathbf{x}_N \\ &= 0 \end{aligned}$$

Martingale and Efron-Stein Inequality

$$\text{Var}(Z) = \sum_{i=1}^N \mathbb{E}[\Delta_i^2]$$

$$\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z = \mathbb{E}_i [Z - \mathbb{E}^{(i)} Z]$$

$$\mathbb{E}^{(i)} Z = \int f(\mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_i | \mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_N) d\mathbf{x}_i$$

Show that

$$\mathbb{E}_i \mathbb{E}^{(i)} Z = \mathbb{E}_{i-1} Z$$

Martingale and Efron-Stein Inequality

$$\Delta_i = \mathbb{E}_i[Z - \mathbb{E}^{(i)} Z]$$

From Jensen's inequality
(Square of expectation vs.
Expectation of square),

$$\Delta_i^2 \leq \mathbb{E}_i \left[(Z - \mathbb{E}^{(i)} Z)^2 \right]$$

Note

$$\text{Var}(Z) = \sum_{i=1}^N \mathbb{E}[\Delta_i^2]$$

Martingale and Efron-Stein Inequality

Show that $\mathbb{E} [\mathbb{E}_i[a]] = \mathbb{E}[a]$.

$$\begin{aligned} \rightarrow \mathbb{E}[\Delta_i^2] &\leq \mathbb{E} \left[\mathbb{E}_i[(Z - \mathbb{E}^{(i)} Z)^2] \right] \\ &= \mathbb{E} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right] \end{aligned}$$

Martingale and Efron-Stein Inequality

$$\text{Var}^{(i)} Z \equiv \mathbb{E}^{(i)} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right]$$

Conditional variance operator conditioned on

$$X^{(i)} (= X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_N)$$

$$\begin{aligned} \mathbb{E}[\text{Var}^{(i)} Z] &= \mathbb{E} \left[\mathbb{E}^{(i)} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right] \right] \\ &= \mathbb{E} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right] \end{aligned}$$

(from $\mathbb{E} \left[\mathbb{E}^{(i)} [a] \right] = \mathbb{E}[a]$)

Also show this!!



Martingale and Efron-Stein Inequality

$$\begin{aligned}\mathbb{E}[\Delta_i^2] &= \mathbb{E}[(Z - \mathbb{E}^{(i)} Z)^2] \\ &= \mathbb{E}[Var^{(i)} Z]\end{aligned}$$

$$\begin{aligned}Var(Z) &\leq \sum_{i=1}^N \mathbb{E} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right] \\ &= \sum_{i=1}^N \mathbb{E} \left[Var^{(i)}(Z) \right]\end{aligned}$$

Martingale and Efron-Stein Inequality

- How can we use the inequality?

$$\textcircled{1} \quad \text{Var}(Z) \leq \frac{1}{2} \sum_{i=1}^N \mathbb{E} [(Z - Z'_i)^2]$$

Z'_i : independent copy of Z
conditionally on $X^{(i)}$

We use $\text{Var}^{(i)}(Z) = \frac{1}{2} \mathbb{E}^{(i)} [(Z - Z'_i)^2]$

$$\left(\frac{1}{2} \mathbb{E} [(X - Y)^2] = \mathbb{E} [(X - \mathbb{E}X)^2] \right)$$

Independent realizations $X, Y \sim P$

Martingale and Efron-Stein Inequality

- How can we use the inequality?

$$\textcircled{2} \quad \text{Var}(Z) \leq \sum_{i=1}^N \inf_{Z_i} \mathbb{E} [(Z - Z_i)^2]$$

We use

$$\text{Var}^{(i)}(Z) = \inf_{Z_i} \mathbb{E}^{(i)} [(Z - Z_i)^2]$$

Conditioned on $X^{(i)}$

Functions with Bounded Differences

$$f : \mathcal{X}^N \rightarrow \mathbb{R}$$

$$\sup_{\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{x}'_i \in \mathcal{X}} |f(\mathbf{x}_1, \dots, \mathbf{x}_N) - f(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}'_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_N)| \leq c_i$$

for some nonnegative constants c_1, \dots, c_N

(= If we change the i -th variable of f while keeping all the others fixed, the value of the function cannot change by more than c_i .)

$$\text{Var}(Z) \leq \frac{1}{4} \sum_{i=1}^N c_i^2$$



Functions with Bounded Differences

$$\text{Var}(Z) \leq \frac{1}{4} \sum_{i=1}^N c_i^2$$

Proof:

$$\text{Var}(Z) \leq \sum_{i=1}^N \inf_{Z_i} \mathbb{E} [(Z - Z_i)^2]$$

If we let
$$Z_i = \frac{1}{2} \left(\sup_{\mathbf{x}'_i \in \mathcal{X}} f(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}'_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_N) + \inf_{\mathbf{x}'_i \in \mathcal{X}} f(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}'_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_N) \right)$$

$$(Z - Z_i)^2 \leq \left(\frac{c_i}{2} \right)^2 = \frac{c_i^2}{4}$$

↑
Middle between max and min



Kernel Density Estimation

$$\phi_N(\mathbf{x}) = \frac{1}{Nh_N} \sum_{i=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_N}\right) \quad \int K = 1$$

$$Z(N) = f(\mathbf{x}_1, \dots, \mathbf{x}_N) = \int |\phi(\mathbf{x}) - \phi_N(\mathbf{x})| d\mathbf{x} \quad L_1 \text{ error}$$

$$\begin{aligned} & |f(\mathbf{x}_1, \dots, \mathbf{x}_N) - f(\mathbf{x}_1, \dots, \mathbf{x}'_i, \dots, \mathbf{x}_N)| \\ & \leq \frac{1}{Nh_N} \int \left| K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_N}\right) - K\left(\frac{\mathbf{x} - \mathbf{x}'_i}{h_N}\right) \right| d\mathbf{x} \leq \frac{2}{N} \end{aligned}$$

$$\text{Var}(Z(N)) \leq \frac{1}{N}$$

Kernel Density Estimation

$$\phi_N(\mathbf{x}) = \frac{1}{Nh_N} \sum_{i=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_N}\right) \quad \int K = 1$$

By Chebyshev's inequality

$$\begin{aligned} P\left\{\left|\frac{Z(N)}{\mathbb{E}Z(N)} - 1\right| \geq \epsilon\right\} &= P\left\{|Z(N) - \mathbb{E}Z(N)| \geq \epsilon\mathbb{E}Z(N)\right\} \\ &\leq \frac{\text{Var}(Z(N))}{\epsilon^2(\mathbb{E}Z(N))^2} \leq \frac{1}{N\epsilon^2(\mathbb{E}Z(N))^2} \rightarrow 0 \end{aligned}$$

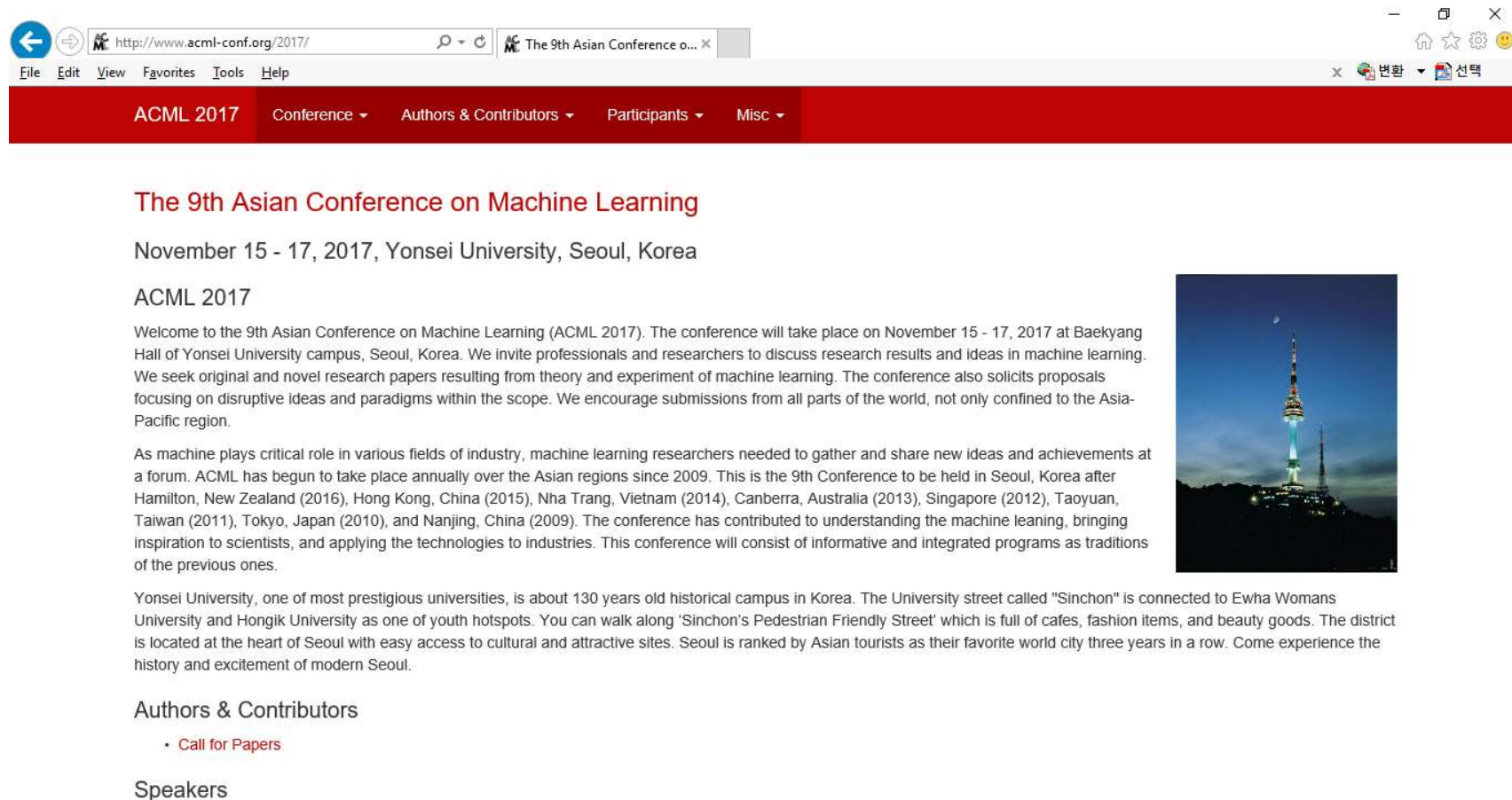
Bounded Difference

- Bounded difference extends to Rademacher average bounding and McDiarmid inequality.
- Foundations of learning theory



Asian Conference on Machine Learning (ACML) in Seoul

- <http://www.acml-conf.org/2017/>
- Nov. 15 - 17 (Wed. – Fri.), 2017



The screenshot shows a web browser window displaying the ACML 2017 website. The browser's address bar shows the URL <http://www.acml-conf.org/2017/>. The website has a red navigation bar with the following menu items: ACML 2017, Conference, Authors & Contributors, Participants, and Misc. The main content area features the title "The 9th Asian Conference on Machine Learning" in red, followed by the dates and location: "November 15 - 17, 2017, Yonsei University, Seoul, Korea". Below this is a section titled "ACML 2017" with a welcome message and details about the conference. A photograph of the N Seoul Tower at night is positioned to the right of the text. Further down, there are sections for "Authors & Contributors" (with a link for "Call for Papers") and "Speakers".

The 9th Asian Conference on Machine Learning

November 15 - 17, 2017, Yonsei University, Seoul, Korea

ACML 2017

Welcome to the 9th Asian Conference on Machine Learning (ACML 2017). The conference will take place on November 15 - 17, 2017 at Baekyang Hall of Yonsei University campus, Seoul, Korea. We invite professionals and researchers to discuss research results and ideas in machine learning. We seek original and novel research papers resulting from theory and experiment of machine learning. The conference also solicits proposals focusing on disruptive ideas and paradigms within the scope. We encourage submissions from all parts of the world, not only confined to the Asia-Pacific region.

As machine plays critical role in various fields of industry, machine learning researchers needed to gather and share new ideas and achievements at a forum. ACML has begun to take place annually over the Asian regions since 2009. This is the 9th Conference to be held in Seoul, Korea after Hamilton, New Zealand (2016), Hong Kong, China (2015), Nha Trang, Vietnam (2014), Canberra, Australia (2013), Singapore (2012), Taoyuan, Taiwan (2011), Tokyo, Japan (2010), and Nanjing, China (2009). The conference has contributed to understanding the machine learning, bringing inspiration to scientists, and applying the technologies to industries. This conference will consist of informative and integrated programs as traditions of the previous ones.

Yonsei University, one of most prestigious universities, is about 130 years old historical campus in Korea. The University street called "Sinchon" is connected to Ewha Womans University and Hongik University as one of youth hotspots. You can walk along 'Sinchon's Pedestrian Friendly Street' which is full of cafes, fashion items, and beauty goods. The district is located at the heart of Seoul with easy access to cultural and attractive sites. Seoul is ranked by Asian tourists as their favorite world city three years in a row. Come experience the history and excitement of modern Seoul.

Authors & Contributors

- [Call for Papers](#)

Speakers



THANK YOU

Yung-Kyun Noh
nohyung@snu.ac.kr

