Random Variable Control

1 Maximum of i.i.d gaussian

Let $\xi_1, \ldots, \xi_k$ be k independent samples from $\mathcal{N}(0,1)$. Then

$$\mathbb{E}\left[\max\left\{\xi_1^2, \ldots, \xi_k^2\right\}\right] \leq 2\log(2k) \tag{.1}$$

2 Union bound for partial sums

2.1 Etemadi's inequality

Let $X_1, \ldots, X_n$ be independent random variables. For $i \in [n]$, let $Y_i = \sum_{j=1}^{i} X_j$ denote the partial sum up to i. Then for all $\alpha \geq 0$,

$$\Pr[\max_{i=1}^{n} |Y_i| > 3 \cdot \alpha] \leq 3 \cdot \max_{i=1}^{n} \Pr[|Y_i| > \alpha]. \tag{.2}$$

Proof Sketch. $\mathbb{P}[|Y_i| > \alpha]$ term을 얻기 위해서 $|Y_i - Y_n|$과 $|Y_i|$ 사이의 independence를 사용함. 그리고 partial sum의 maximum과 각 partial sum을 연결하기 위해서 i번째 partial sum이 처음으로 3α 보다 큰 event로 분해함. (Detail)

3 Random Singed Summation Bound

3.1 Khintchine ineqaulity

Let $\{\varepsilon_n\}_{n=1}^{N}$ be i.i.d. Rademacher random variables. Let $0 < p < \infty$ and let $x_1, \ldots, x_N \in \mathbb{C}$. Then

$$A_p \left(\sum_{n=1}^{N} |x_n|^2\right)^{1/2} \leq \left(\mathbb{E}\left|\sum_{n=1}^{N} \varepsilon_n x_n\right|^p\right)^{1/p} \leq B_p \left(\sum_{n=1}^{N} |x_n|^2\right)^{1/2} \tag{.3}$$

4 Summation Bound

4.1 Marcinkiewicz-Zygmund ineqaulity

If X_i, $i = 1, \ldots, n$ are independent random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[|X_i|^p], 1 < p < +\infty$, then

$$A_p \, \mathbb{E}\left[\left(\sum_{i=1}^{n} |X_i|^2\right)^{p/2}\right] \leq \mathbb{E}\left[\left|\sum_{i=1}^{n} X_i\right|^p\right] \leq B_p \, \mathbb{E}\left[\left(\sum_{i=1}^{n} |X_i|^2\right)^{p/2}\right] \tag{.4}$$

where A_p and B_p are positive constants, which depend only on p. for some constants A_p, B_p depending only on p.

4.2 Latala's inequality

If $p \geq 2$ and $X, X_1, \ldots, X_n$ are i.i.d. mean 0 random variables, then we have

$$\left\|\sum_{i=1}^{n} X_i\right\|_{L^p} \sim \sup\left\{\frac{p}{s}\left(\frac{n}{p}\right)^{1/s} \|X\|_{L^s} \,\middle|\, \max\left\{2, \frac{p}{n}\right\} \leq s \leq p\right\} \tag{.5}$$

5 Concentration ineqaulity

5.1 Bernstein's ineqaulity

Let $X_1, \dots X_n$ be independent random variables. Assume $\mathbb{E}[X_i] = 0, \mathbb{E}[X_i^2] = \sigma_i^2$, and $\Pr[|X_i| \le 1] = 1$ for each $i \in [n]$. Let $\sigma^2 := \sum_{i=1}^n \sigma_i^2$. Then for all $t \ge 0$,

$$\Pr\left[\sum_{i=1}^n X_i \ge t\right] \le \exp\left(\frac{-3t^2}{6\sigma^2 + 2t}\right) \tag{.6}$$

Proof Sketch. First bound the MGF of each X_i using taylor expansion. Then use Markov inequality for $\Pr\left[\sum_{i=1}^n X_i \ge t\right]$ with $\exp(\lambda \cdot)$ and minimize the upper bound with λ. Then use the following lemma to finish the proof.

Lemma. Let $v > -1$. Then $(1+v)\log(1+v) \ge v + \frac{3v^2}{2v+6}$

5.2 Matrix version Bernstein's inequality

Let $\boldsymbol{B}$ a fixed $q \times d$ matrix. Construct $q \times d$ matrix $\boldsymbol{R}$ such that

$$\mathbb{E}[\boldsymbol{R}] = \boldsymbol{B}, \quad \|\boldsymbol{R}\|_{\mathrm{op}} \le L \tag{.7}$$

Form the matrix sampling estimator

$$\bar{\boldsymbol{R}}_m = \frac{1}{m}\sum_{i=1}^m \boldsymbol{R}_i, \tag{.8}$$

where each $\boldsymbol{R}_i$ is an independent copy of $\boldsymbol{R}$. Then for every $t > 0$, the estimator satisfies

$$\mathbb{P}\left[\left\|\bar{\boldsymbol{R}}_m - \boldsymbol{B}\right\|_{\mathrm{op}} \ge t\right] \le (q+d)\cdot\exp\left(\frac{-mt^2}{m_2(\boldsymbol{R}) + 2Lt/3}\right), \tag{.9}$$

where $m_2(R)$ is the second moment $m_2(\boldsymbol{R}) = \max\left\{\|\mathbb{E}[\boldsymbol{R}^*\boldsymbol{R}]\|_{\mathrm{op}}, \|\mathbb{E}[\boldsymbol{R}\boldsymbol{R}^*]\|_{\mathrm{op}}\right\}$.

5.3 Hoeffding's inequality

Let $X_1, \dots, X_n$ be independent random variables such that $a_i \le X_i \le b_i$ a.s. Then for all $t > 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \ge t) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right), \tag{.10}$$

where $S_n = X_1 + \dots X_n$. Also, consider a set of r i.i.d. random variables $X_1 \dots . X_r$ such that $-\Delta \le X_i \le \Delta$ and $\mathbb{E}[X_i] = 0$ for each $i \in [r]$. Let $\sum_{i=1}^r X_i$. Then for any $\alpha \in (0, 1/2)$

$$\mathbb{P}[|M| > \alpha] \le 2\exp(-\frac{\alpha^2}{2r\Delta^2}) \tag{.11}$$

The proof uses the following:

Lemma. Let X be any real-valued random variable such that $a \le X \le b$ a.s. Then. for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \le \exp\left(\frac{\lambda^2(b-a)^2}{8}\right) \tag{.12}$$

Note: conditional expectation also works for the lemma.

5.4 Variance-only form

Consider a set of r independent random variables $X_1, \dots X_r$. Let $M = \sum_{i=1}^r X_i$. Then for $\alpha \in (0, 2\mathrm{Var}[M]/(\max_i |X_i - \mathbb{E}[X_i]|))$

$$\mathbb{P}[|M - \mathbb{E}[M]| > \alpha] \leq 2\exp\left(\frac{-\alpha^2}{4\sum_{i=1}^r \mathrm{Var}[X_i]}\right). \tag{.13}$$

5.5 Paley-Zygmund inequality

If $Z \geq 0$ is a random variable with finite variance, and if $0 \leq \theta \leq 1$, then

$$\mathbb{P}(Z > \theta\,\mathbb{E}[Z]) \geq (1-\theta)^2 \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]} \tag{.14}$$

5.6 Max of independent Gaussians

Let $X_1, X_2, \dots, X_n$ i.i.d. $\mathcal{N}(0,1)$, then

$$\mathbb{E}[\max(X_1, \dots, X_n)] = \sqrt{2\log(n)} + o(\sqrt{\log(n)}) \tag{.15}$$

5.7 DKW inequality

DKW inequality provides a bound on the worst-case distance of empirical CDF and the true CDF:

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \varepsilon\right) \leq Ce^{-2n\varepsilon^2} \quad \text{for every } \varepsilon > 0. \tag{.16}$$

For multivariate case, let $X_1, X_2, \dots, x_n$ be an i.i.d. sequence of k-dimensional vectors,

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}^k} |F_n(t) - F(t)| > \varepsilon\right) \leq (n+1)ke^{-2n\varepsilon^2} \tag{.17}$$

for every $\varepsilon, n, k > 0$.

Also see local DKW inequality.

Steinke version. Let $X_1, \dots, X_n$ be independent random variables with CDF $f(v) := \mathbb{P}[X_i \leq v]$ for all $i \in [n]$ and $v \in \mathbb{R}$. Let the empirical CDF be $F_x(v) := \frac{1}{n}\sum_{i=1}^n 1[X_i \leq v]$ for all $v \in \mathbb{R}$. Then, for all $\beta > 0$,

$$\mathbb{P}_X\left[\sup_{v\in\mathbb{R}} F_X(v) - f(v) \leq \sqrt{\frac{2\log(1/\beta)}{n}} + \frac{\log(1/\beta)}{2n}\right] \geq 1 - \beta. \tag{.18}$$

Lemma. For all $t, \lambda > 0$,

$$\mathbb{P}\left[\sup_{v\in\mathbb{R}} F_x(v)\log\left(1 + \frac{t}{f(v)}\right) > \frac{\lambda}{n}\right] \leq (1+t)^n e^{-\lambda} \leq e^{tn-\lambda}. \tag{.19}$$

Note: maximum이 bound되는 event 확률 구할 때는 martingale construction해서 optional stopping theorem적용하는 것도 좋음 ㅋ Lemma에서는 bionamial exponent에 놓아서 martingale 만듬.

5.8 Quadratic form

Definition (Subgaussian random variable). A centered random variable X is said to be v-subgaussian if its cumulant generating function is subquadratic:

$$\xi_X(t) \leq \frac{1}{2}vt^2 \quad \forall t \in \mathbb{R} \tag{.20}$$

5.9 Hanson-Wright tail bound

Let $\boldsymbol{x}$ be a random vector with independent centered v-subgaussian entries and let $\boldsymbol{A}$ be a square matrix. Then

$$\mathbb{P}\left(\left|\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}]\right| \geq t\right) \leq 2\exp\left(-\frac{c \cdot t^2}{v^2 \left\|\boldsymbol{A}\right\|_F^2 + v\left\|A\right\| t}\right), \tag{.21}$$

where $c > 0$ is a constant independent of $v, \boldsymbol{x}, t$ or $\boldsymbol{A}$.

5.10 Gaussian CCDF bound

$$1 - \Phi(w) \leq \min\left\{\frac{1}{2}, \frac{1}{w\sqrt{2\pi}}\right\} e^{-w^2/2}, \quad w > 0 \tag{.22}$$

5.11 McDiarmid's inequality

A function $f : \mathcal{X} \times \mathcal{X} \times \ldots \mathcal{X}_n \to \mathbb{R}$ satisfies the bounded differences property if substituting the value of the ith coordinate x_i changes the value of f by at most c_i. More formally, if there are constants $c_1, c_2, \ldots, c_n$ such that for all $i \in [n]$, and all $x_i \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, \ldots, x_n \in \mathcal{X}_3$,

$$\sup_{x_i' \in \mathcal{X}_i} |f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i \tag{.23}$$

Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfy the bounded differences property with bounds $c_1, c_2, \ldots, c_n$.
Consider independent random variables $X_1, X_2, \ldots, X_n$ where $X_i \in C_i$ for all i. Then, for any $\varepsilon > 0$,

$$\mathbb{P}(f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)] \geq \varepsilon) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right), \tag{.24}$$

$$\mathbb{P}(f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)] \leq -\varepsilon) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right) \tag{.25}$$

and as an immediate consequence,

$$\mathbb{P}(|f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots X_n)]| \geq \varepsilon) \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right) \tag{.26}$$

6 Decoupling lemma

6.1 Quadratic form

Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Let $X_1, \ldots X_n \in \mathbb{R}$ be independent mean-zero random variables. For $i, j \in [n]$, let $a_{i,j} \in \mathbb{R}$ be a constant. Then

$$\mathbb{E}\left[f\left(\sum_{i \neq j} a_{ij} X_i X_j\right)\right] \leq \mathbb{E}\left[f\left(4\sum_{i \neq j} a_{ij} X_i X_j'\right)\right], \tag{.27}$$

where $X_1', \ldots, X_n'$ are independent copies of $X_1, \ldots, X_n$.

Note: We can analyze $\boldsymbol{x}^* \boldsymbol{A}\boldsymbol{x}$ by $\boldsymbol{x}^* \boldsymbol{A}\boldsymbol{x}'$ with independent $\boldsymbol{x}'$.

7 Global variance control

7.1 Efron-Stein inequality

For $i \in [n]$ and tuple $Z = (Z_1, \ldots, Z_n)$, let $Z^{(i)}$ denote the tuple $(Z_1, \ldots, Z_{i-1}, \tilde{Z}_i, Z_{i+1}, \ldots, Z_n)$, where $\tilde{Z}_i$ is an independent copy of Z_i. For a scalar function $f(Z)$, the Efron-Stein inequality states that

$$\mathrm{Var}[f(Z)] = \mathbb{E}[(f(Z) - \mathbb{E}[f(Z)])^2] \le \frac{1}{2} \cdot \sum_{i \in [n]} \mathbb{E}\left[\left(f(Z) - f(Z^{(i)})\right)^2\right] \tag{.28}$$

$$\overset{\dagger}{=} \underbrace{\sum_{i \in [n]} \mathbb{E}\left[\left(f(Z) - E_i[f(Z^{(i)})]\right)^2\right]}_{\text{sum of conditional variance}} \tag{.29}$$

$\dagger$: Note that $\mathbb{E} = \mathbb{E}_{-i}\mathbb{E}_i, \mathbb{E}[f(Z^{(i)}) \mid Z] = \mathbb{E}_i[f(Z^{(i)})]$ and

$$\mathop{\mathbb{E}}_i[(Z - \mathop{\mathbb{E}}_i[Z])^2] = \frac{1}{2}\mathop{\mathbb{E}}_i[(Z - Z^{(i)})^2] \tag{.30}$$

8 Information

8.1 Fano's inequality

Let $X \in \{0,1\}^d$ be uniformly random and let $Y \in \mathbb{R}^d$ be a random variable that depends on X.
If $\mathbb{E}[\|X - Y\|_1] \le \alpha \cdot d$ for $\alpha \le \frac{1}{2}$, then

$$I(X;Y) \ge d \cdot D_{\mathrm{KL}}\left(\mathrm{Ber}(\alpha) || \mathrm{Ber}\left(\frac{1}{2}\right)\right). \tag{.31}$$

9 Do you like martingale?

9.1 Tail Distribution

Let X be a nonnegative cadlag submartingale. Then, for each $K, t > 0$,

$$K\mathbb{P}(X_t^* \ge K) \le \mathbb{E}[1_{\{X_t^* \ge K\}} X_t] \tag{.32}$$

10 Stochastic Dominance

10.1 Definition

Let $X, Y \in \mathbb{R}$ be random variables. We say X is *stochastically dominated* by Y if $\mathbb{P}[X > t] \le \mathbb{P}[Y > t]$ for all $t \in \mathbb{R}$. Equivalently, X is stochastically dominated by Y if there exists a coupling such that $\mathbb{P}[X \le Y] = 1$.

10.2 SD is preserved under sums/convolutions

Lemma. Suppose X_1 is stochastically dominate by Y_1. Suppose that, for all $x \in \mathbb{R}$, the conditional distribution $X_2 | X_1 = x$ is stochastically dominated by Y_2. Assume that Y_1 and Y_2 are independent. Then $X_1 + X_2$ is stochastically dominated by $Y_1 + Y_2$.