SDS 384 11: Theoretical Statistics

Lecture 7a: Efron Stein inequality

Purnamrita Sarkar

Department of Statistics and Data Science

The University of Texas at Austin

Efron Stein inequality

- Consider n independent random variables in some metric space $\mathcal{X}$.
- Consider a function $g : \mathcal{X}^n \to \mathbb{R}$
- Let $Z := g(X_1, \ldots, X_n)$
- We are interested in computing $\text{var}(g(X_1, \ldots, X_n))$
- Define $E_i(Z) = E[Z|X_{1:i-1}, X_{i+1:n}]$

An upper bound

Theorem

$$var(Z) \leq \sum_{i=1}^{n} E\left[Z - E_i[Z]\right]^2$$

Proof.

- For two arbitrary bounded random variables X, Y, we have:
  $E[XY] = E[E[XY|Y]] = E[YE[X|Y]]$

- Let $V := Z - E[Z]$

- Let $V_i := E[Z|X_{1:i}] - E[Z|X_{1:i-1}]$

- Clearly $V = \sum_i V_i$

$\square$

Proof continued

$$\text{var}(Z) = E\left[\sum_i V_i\right]^2 \tag{1}$$

$$= \sum_i E[V_i^2] + 2\sum_{i<j} E[V_i V_j] = \sum_i E[V_i^2] \tag{2}$$

- Why is the last step true? For $i > j$

$$E[V_i V_j] = E[E[V_i V_j | X_1, \ldots, X_j]]$$
$$= E[V_j E[V_i | X_1, \ldots, X_j]] = 0$$

Proof cont.

- Note that for three independent random variables X, Y, Z

$$E[g(X, Y, Z)|X] = E[[g(X, Y, Z)|X, Z]|X, Y]$$

$$LHS = \int_{y,z} g(x, y, z) f(y, z|x) dy dz = \int_z (\int_y g(x, y, z) f(y|x, z) dy) f(z|x) dz$$

$$= \int_z E[g(X, Y, Z)|X, Z] f(z|x) dz$$

$$\stackrel{independence}{=} \int_z E[g(X, Y, Z)|X, Z] f(z|x, y) dz$$

$$= E[E[g(X, Y, Z)|X, Z]|X, Y]$$

Proof cont.

-
$$V_i^2 = (E[Z|X_{1:i}] - E[Z|X_{1:i-1}])^2$$
$$= (E[Z|X_{1:i}] - E[Z|X_{1:i-1}])^2$$
$$= (E[E[Z|X_{1:n}]|X_{1:i}] - E[E[Z|X_{1:i-1}, X_{i+1:n}]|X_{1:i}])^2$$
$$= (E[E[Z|X_{1:n}] - E[Z|X_{1:i-1}, X_{i+1:n}]|X_{1:i}])^2$$
$$= (E[Z - E_i Z|X_{1:i}])^2$$
$$\leq E[(Z - E_i Z)^2|X_{1:i}]$$
$$E[V_i^2] \leq E[(Z - E_i Z)^2]$$

The Efron Stein inequality

Theorem

Let $X_1', \ldots X_n'$ denote an independent copy of $X_1, \ldots, X_n$. Let $Z_i' = g(X_{1:i-1}, X_i', X_{i+1:n})$. We have:

$$var(Z) \leq \frac{1}{2} \sum_i E[(Z - Z_i')^2].$$

Proof.

- If X, Y are iid, $var(X) = \frac{E[X - Y]^2}{2}$

- Conditioned on $X_{1:i-1}, X_{i:n}$, Z and Z_i' are independent and so

$$E_i[Z - E_i[Z]]^2 = \frac{E_i[Z - Z_i']^2}{2}$$

$$var(Z) \leq \sum_{i=1}^{n} E\left[Z - E_i[Z]\right]^2 = \frac{E[E_i[Z - Z_i']^2]}{2}$$

Remarks

- For $g(X_1, \ldots, X_n) = \sum_i X_i$ we have an equality.

- So in some sense, sums of independent random variables are the least concentrated functions

- Consider a function with the Bounded Difference property, i.e.

$$\sup_{x_{1:n}, x_i' \in \mathcal{X}} |g(x_1, \ldots, x_n) - g(x_{1:i-1} x_i' x_{i+1:n})| \leq c_i$$

- We have:

$$\mathrm{var}(g(X)) \leq \frac{1}{2} \sum_i c_i^2$$

Example: longest common subsequence

Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be two sequences of coin flips. Z is the length of the longest common subsequence.

$$Z = \max\{k : X_{i_1} = Y_{j_1}, \ldots, X_{i_k} = Y_{j_k}\}$$

where $1 \leq i_1 < i_2 \ldots$ and $1 \leq j_1 < j_2 \ldots$.

- It is well known that $E[Z]/n \to \mu$ where $\mu \in [0.757, 0.837]$.
- If you change one bit of X, it can change Z by at most one, so,

$$\text{var}(Z) \leq n/2$$

- So Z concentrates around its mean.

Uniform deviation

For $X_1, \ldots, X_n$ iid random variables, let $\hat{P}_n(A) = \frac{1}{n} 1(X_i \in A)$ and $P_n(A) = P(X_i \in A)$. We are interested in te quantity $Z := \sup_A |\hat{P}_n(A) - P_n(A)|$

- If we change one X_i, Z changes by $1/n$ at most.
- So $\mathrm{var}(Z) \leq \dfrac{1}{2n}$ by the Efron Stein inequality.
- Can we do better?

Uniform deviation

For $X_1, \ldots, X_n$ iid random variables, let

$$Z = \sup_{f \in \mathcal{F}} \sum_j f(X_j).$$

For simplicity, assume $Ef[X_i] = 0$. We will show that the E/S inequality gives a much tighter upper bound that the one we just derived.

- $\text{var}(Z) \leq \dfrac{1}{2} \sum_i E[(Z - Z_i')^2]$

- Say f^* achieves the supremum for Z and f_* achieves the supremum for Z_i

$$f_*(X_i) - f_*(X_i') \leq Z - Z_i \leq f^*(X_i) - f^*(X_i')$$
$$(Z - Z_i)^2 \leq \max((f_*(X_i) - f_*(X_i'))^2, (f^*(X_i) - f^*(X_i'))^2)$$
$$\leq \sup_{f \in \mathcal{F}} (f(X_i) - f(X_i'))^2$$

Uniform deviation

$$\text{var}(Z) \leq \frac{1}{2} \sum_i E \left[\sup_{f \in \mathcal{F}} (f(X_i) - f(X_i'))^2 \right]$$

$$\overset{(i)}{\leq} \sum_i E \left[\sup_{f \in \mathcal{F}} (f(X_i)^2 + f(X_i')^2) \right]$$

$$\leq 2 \sum_i E \sup_{f \in \mathcal{F}} f(X_i)^2$$

- (i) uses $|2ab| \leq a^2 + b^2$
- If $f(X_i) \in [-1, 1]$ we get $\text{var}(Z) \leq 2n$
- But if the maximum variance of $f(X_i)$ is small we have a significant improvement.

11

Minimum of empirical loss

Consider a function class $\mathcal{F}$ of binary valued functions on some space $\mathcal{X}$. Given an iid sample $(X_i, Y_i) \in \mathcal{X} \times \{0, 1\}$, for each $f \in \mathcal{F}$ we define the empirical loss:

$$L_n(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(X_i), Y_i) \qquad \text{where } \ell(y, y') = 1(y \neq y')$$

Define the empirical loss as $\hat{L} = \inf_{f \in \mathcal{F}} L_n(f)$.

- Naive application of Efron Stein shows $\text{var}(\hat{L}) \leq 2/n$
- Is this enough?

Minimum of empirical loss

- Let $Z = n\hat{L}$

- Let $Z_i = \min\limits_{f \in \mathcal{F}} \left(\sum\limits_{j \neq i} \ell(f(X_j), Y_j) + \ell(f(X_i'), Y_i') \right)$

- $\text{var}(Z) \leq \dfrac{1}{2} \sum\limits_i E[Z - Z_i']^2 = \sum\limits_i E[Z - Z_i']^2 1(Z_i' > Z)$

- Note that
  $$0 \geq (Z - Z_i')1(Z_i' > Z) \geq (\ell(f^*(X_i), Y_i) - \ell(f^*(X_i'), Y_i'))1(Z_i' > Z)$$

- So $(Z - Z_i')^2 1(Z_i' > Z) \leq (\ell(f^*(X_i), Y_i) - \ell(f^*(X_i'), Y_i'))^2 1(Z_i' > Z) \leq \ell(f^*(X_i'), Y_i')1(\ell(f^*(X_i), Y_i) = 0)$

- So, $E \sum\limits_i (Z - Z_i')^2 1(Z_i' > Z) \leq E \sum\limits_{\ell(f^*(X_i), Y_i) = 0} E_{X_i', Y_i'} \ell(f^*(X_i'), Y_i') \leq$
  $nEL(f^*)$

- Often you can show that $EL(f^*) = E\hat{L} + O(n^{-1/2})$

- So $\text{var}(\hat{L}) \leq \dfrac{E\hat{L}}{n} + o(1)$

Self bounding functions

Definition

A non-negative function $g : \mathcal{X}^n] \to \mathcal{R}$ has the self bounding property if there exist functions $g_i : \mathcal{X}^{n-1} \to \mathcal{R}$ such that for all $x_1, \ldots, x_n \in \mathcal{X}$ and $i \in [n]$,

- $0 \leq g(x_1, \ldots, x_n) - g_i(x_{1:i-1}, x_{i+1:n}) \leq 1$
- $\sum_i \left(g(x_1, \ldots, x_n) - g_i(x_{1:i-1}, x_{i+1:n}) \right) \leq g(x_1, \ldots, x_n)$

- Clearly, $\sum_i (g(x_{1:n}) - g_i(x_{1:i-1}, x_{i+1:n}))^2 \leq g(x_1, \ldots, x_n) =: Z$

- Now Theorem 1 gives:

$$\text{var}(Z) \leq \sum_i E[(Z - E_i[Z])^2] \leq \sum_i E[(Z - g_i(x_{1:i-1}, x_{i+1:n}))^2] \leq E[g(x_{1:n})]$$

- So $\text{var}(Z) \leq E[Z]$

Concentration of self bounding functions

Theorem

Consider $Z := g(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are independent random variables. For all $t \geq 0$,

$$P\left(Z \geq E[Z] + t\right) \leq \exp\left(-\frac{t^2}{2(EZ + t/3)}\right)$$

$$P\left(Z \leq E[Z] - t\right) \leq \exp\left(-\frac{t^2}{2EZ}\right)$$

Relative Stability

- A sequence of non-negative random variables $\{Z_n\}$ are said to be relatively stable if $Z_n/E[Z_n] \xrightarrow{P} 0$

- If Z_n also satisfies the self bounding property,

$$P\left(|\frac{Z_n}{E[Z_n]} - 1| \geq \epsilon\right) \leq \frac{\text{var}(Z_n)}{\epsilon^2 E[Z_n]^2} \leq \frac{1}{\epsilon^2 E[Z_n]}$$

- So as long as $E[Z_n] \rightarrow \infty$, Z_n satisfies the relative stability condition

Example: empirical processes

Consider a function class $\mathcal{F}$ of functions in $[0, 1]$. $Z := \sup_{f \in \mathcal{F}} \sum_i f(X_i)$. We show that Z is self bounding.

- Let $Z_i := \sup_{f \mathcal{F}} \sum_{j \neq i} f(X_i)$

- Let f^* maximize Z and f_i maximize Z_i

- We have $0 \leq f_i(X_i) \leq f^*(X_i) \leq 1$

- So $\sum_i (Z - Z_i) \leq \sum_i f^*(X_i) = Z$

- Hence $\text{var}(Z) \leq E[Z]$, while a naive application of E-S will give us $\text{var}(Z) \leq n/2$

Rademacher averages

Consider a function class $\mathcal{F}$ of functions in $[-1, 1]$. Let $\{\epsilon_i\}_1^n$ denote n independent Rademacher variables independent of $X_1, \ldots, X_n$. The conditional Rademacher average is defined as

$$Z := E\left[\sup_{f \in \mathcal{F}} \sum_i \epsilon_i f(X_i) \,|X_{1:n}\right]$$

Z has the self bounding property and so $\mathrm{var}(Z) \leq E[Z]$.

- Define $Z_i := E\left[\sup_{f \in \mathcal{F}} \sum_{j \neq i} \sigma_j f(X_j) \,|X_{1:n}\right]$

Rademacher avg cont.

- Let f^* maximize Z and f_i maximize Z_i. Note that:

$$Z - Z_i \leq E[\epsilon_i f^*(X_i)|X_{1:n}] \leq 1$$

- On the other hand,

$$Z - Z_i \geq E[\epsilon_i f_i(X_i)|X_{1:n}] = 0$$

- The last step is true because ?

- So $\sum_i (Z - Z_i) \leq Z$

- Hence Z has the self-bounding property and has $var(Z) \leq E[Z]$