Problem 1 : Jackknife : variance and bias:

(a) Variance bound

jackknife variance functional: $\hat{V} = \sum_{i=1}^{n} (z^{(i)} - \bar{z})^2$.

(a1)

$$\hat{V} = \sum_{i=1}^{n} (z^{(i)} - \bar{z})^2 = \sum_{i=1}^{n} \left((z^{(i)})^2 - 2 z^{(i)} \bar{z} + \bar{z}^2 \right)$$

$$= \sum_i (z^{(i)})^2 - 2 \cdot \left(\sum_i z^{(i)} \right) \bar{z} + n \bar{z}^2.$$

$$= \sum_i (z^{(i)})^2 - 2n\bar{z}^2 + n\bar{z}^2$$

$$= \sum_i (z^{(i)})^2 - n\bar{z}^2.$$

And,

$$\frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} (z^{(i)} - z^{(j)})^2 = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left((z^{(i)})^2 - 2 z^{(i)} z^{(j)} + (z^{(j)})^2 \right)$$

$$= \frac{1}{2n} \sum_{i=1}^{n} \left(n (z^{(i)})^2 - 2n z^{(i)} \bar{z} + \sum_j (z^{(j)})^2 \right)$$

$$= \frac{1}{2n} \left(n \cdot \sum_i (z^{(i)})^2 - 2n^2 \bar{z}^2 + n \cdot \sum_i (z^{(i)})^2 \right)$$

$$= \sum_i (z^{(i)})^2 - n\bar{z}^2.$$

Hence,

$$\boxed{\hat{V} = \sum_{i=1}^{n} (z^{(i)} - \bar{z})^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(z^{(i)} - z^{(j)})^2}{2n}}$$

(a2) Here,

$$z^{(i)} = f_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \quad i = 1, \ldots, n.$$

$$W = f_{n-1}(X_1, \ldots, X_{n-1}).$$

We want to show: $\operatorname{Var}(W) \le \mathbb{E}[\hat{V}]$.

Using i.i.d. symmetry,

$$\operatorname{Var}(W) = \operatorname{Var}(Z^{(1)}).$$

Using tensorization of the variance to $f_{n-1}(X_2, \dots, X_n)$,

$$\operatorname{Var}(Z^{(1)}) \le \sum_{k=2}^{n} \mathbb{E}\left[\operatorname{Var}\left(Z^{(1)} \mid X_2, \dots, X_{k-1}, X_{k+1}, \dots, X_n\right)\right]$$

Fix $k \in \{2, \dots, n\}$. Let,

$$F_k = \sigma(X_2, \dots, X_{k-1}, X_{k+1}, \dots, X_n).$$

Conditioning on F_k and by the symmetry of f_{n-1}, we can write

$$Z^{(1)} = f_{n-1}(X_2, \dots, X_n) = g(X_k),$$

for some random function g determined by the conditional variables.

Similarly, we have,

$$Z^{(k)} = f_{n-1}(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n) = g(X_1).$$

Now,

Under F_k, X_1 and X_k are i.i.d. and independent of F_k, so $g(X_1)$ and $g(X_k)$ will be i.i.d. Then,

$$\operatorname{Var}(Z^{(1)} \mid F_k) = \operatorname{Var}(g(X_k) \mid F_k) = \frac{1}{2}\mathbb{E}\left[(g(X_k) - g(X_1))^2 \mid F_k\right]$$

$$= \frac{1}{2}\mathbb{E}\left[(Z^{(1)} - Z^{(k)})^2 \mid F_k\right]$$

Taking expectations,

$$\mathbb{E}\left[\operatorname{Var}(Z^{(1)} \mid F_k)\right] = \frac{1}{2}\mathbb{E}\left[(Z^{(1)} - Z^{(k)})^2\right]$$

Putting back,

$$\boxed{\operatorname{Var}(Z^{(1)}) \le \frac{1}{2}\sum_{k=2}^{n}\mathbb{E}\left[(Z^{(1)} - Z^{(k)})^2\right]} \quad \cdots\cdots\cdots \quad 1$$

Because $(z^{(1)}, \dots, z^{(n)})$ is exchangable, for any $i \neq j$,

$$\mathbb{E}\left[(z^{(i)} - z^{(j)})^2\right] = \mathbb{E}\left[(z^{(1)} - z^{(2)})^2\right].$$

So,

$$\frac{1}{2} \sum_{k=2}^{n} \mathbb{E}\left[(z^{(1)} - z^{(k)})^2\right] = \frac{n-1}{2} \mathbb{E}\left[(z^{(1)} - z^{(2)})^2\right] \quad \text{.......} \; (2)$$

And, using (a1),

$$\hat{v} = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} (z^{(i)} - z^{(j)})^2.$$

Taking expectations, the $i=j$ terms vanish and there are $n(n-1)$ ordered pairs with $i \neq j$ each with same expectation as described above. So,

$$\mathbb{E}[\hat{v}] = \frac{1}{2n} \cdot n(n-1) \mathbb{E}\left[(z^{(1)} - z^{(2)})^2\right] = \frac{n-1}{2} \mathbb{E}\left[(z^{(1)} - z^{(2)})^2\right]$$

Thus, using (1) and (2), we get,

$$\mathrm{Var}(z^{(1)}) \leq \frac{n-1}{2} \mathbb{E}\left[(z^{(1)} - z^{(2)})^2\right] = \mathbb{E}[\hat{v}].$$

Since, $\mathrm{Var}(W) = \mathrm{Var}(z^{(1)})$,

$$\boxed{\mathrm{Var}(W) \leq \mathbb{E}[\hat{v}].}$$

(b) [Bonus] Exact bias for quadratic statistics

Here, the full sample estimator is,

$$z_n = f_n(X_1, \dots, X_n).$$

We assume the bias admits a first-order expansion:

$$\mathbb{E}[z_n] - \theta = \frac{c}{n} + O(n^{-2}) \quad \text{as } n \to \infty.$$

we further assume, $\mathbb{E}[z_n] - \theta = \frac{c}{n} + \frac{d}{n^2} + o(n^{-2})$ as $n \longrightarrow \infty$,

for some $d \in \mathbb{R}$.

Using the i.i.d. symmetry, $z^{(i)} = f_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ has the same distribution as $f_{n-1}(X_1, \dots, X_{n-1})$, so

$$\mathbb{E}[z^{(i)}] = \mathbb{E}[z_{n-1}] \quad \forall\ i.$$

Then,

$$\mathbb{E}[\bar{z}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[z^{(i)}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[z_{n-1}] = \mathbb{E}[z_{n-1}].$$

Now, taking expectation of the bias-corrected estimator,

$$\mathbb{E}[\tilde{z}_n] = n\,\mathbb{E}[z_n] - (n-1)\,\mathbb{E}[z_{n-1}].$$

Using the first order expansion of the bias for z_n and z_{n-1}:

$$\mathbb{E}[\tilde{z}_n] = n\left(\frac{c}{n} + \frac{d}{n^2} + o(n^{-2}) + \theta\right) - (n-1)\left(\frac{c}{n-1} + \frac{d}{(n-1)^2} + o(n^{-2}) + \theta\right)$$

$$\Rightarrow \mathbb{E}[\tilde{z}_n] - \theta = d\left(\frac{1}{n} - \frac{1}{n-1}\right) + o(n^{-2}).$$

$$= \frac{-d}{n(n-1)} + o(n^{-2}) = O(n^{-2}).$$

Hence,

$$\boxed{\mathbb{E}[\tilde{z}_n] - \theta = O(n^{-2}).}$$

i.e. for quadratic statistics with bias of $O(n^{-1})$, the jackknife can provide exact bias. //

Problem 2: Order Statistics: variance versus spacings:

(a) The Maximum

(a1) We will assume the X_i's are almost surely distinct to avoid ties.

We want to prove:

$$\mathrm{Var}(X_{(n)}) \leq \mathbb{E}\left[(X_{(n)} - X_{(n-1)})^2\right] = \mathbb{E}\left[\Delta_{n-1}^2\right].$$

We claim, for any $i \in \{1, \dots, n\}$,

$$\left(X_{(n)} - X_{(n)}^{(i)}\right)_+ \leq \left(X_{(n)} - X_{(n-1)}\right) \mathbb{1}\{X_i = X_{(n)}\}.$$

- <u>Case 1:</u> $X_i \neq X_{(n)}$.

  Then the maximum $X_{(n)}$ is attained by some index $j \neq i$. When we resample X_i, the value $X_j = X_{(n)}$ is unchanged. So,

  $$X_{(n)}^{(i)} \geq X_{(n)} \Rightarrow \left(X_{(n)} - X_{(n)}^{(i)}\right)_+ = 0.$$

  And,

  $\mathbb{1}\{X_i = X_{(n)}\} = 0$. So, the inequality holds.

- <u>Case 2:</u> $X_i = X_{(n)}$.

  Now, we are resampling a maximizer. When we replace X_i by X_i', the new $X_{(n)}^{(i)}$ is at least the max of the remaining $n-1$ original values i.e.,

  $$X_{(n)}^{(i)} \geq X_{(n-1)} \Rightarrow \left(X_{(n)} - X_{(n)}^{(i)}\right)_+ \leq \left(X_{(n)} - X_{(n-1)}\right).$$

  Because we have $\mathbb{1}\{X_i = X_{(n)}\} = 1$, the inequality holds.

Thus, for a given $i \in \{1, \dots, n\}$,

$$\boxed{\left(X_{(n)} - X_{(n)}^{(i)}\right)_+ \leq \left(X_{(n)} - X_{(n-1)}\right) \mathbb{1}\{X_i = X_{(n)}\}}$$

Now, summing over i,

$$\sum_{i=1}^{n}\left(X_{(n)}-X_{(n)}^{(i)}\right)_+^2 \leq \left(X_{(n)}-X_{(n-1)}\right)^2 \sum_{i=1}^{n} \mathbb{1}\{X_i = X_{(n)}\}$$

Because we assumed a.s. no ties, the indicator sum equals 1. So,

$$\sum_{i=1}^{n}\left(X_{(n)}-X_{(n)}^{(i)}\right)_+^2 \leq \left(X_{(n)}-X_{(n-1)}\right)^2 = \Delta_{n-1}^2 .$$

Taking expectations and using the one-sided ESS yields:

$$\mathrm{Var}\left(X_{(n)}\right) \leq \sum_{i=1}^{n}\mathbb{E}\left[\left(X_{(n)}-X_{(n)}^{(i)}\right)_+^2\right] \leq \mathbb{E}\left[\left(X_{(n)}-X_{(n-1)}\right)^2\right].$$

Hence,

$$\boxed{\mathrm{Var}\left(X_{(n)}\right) \leq \mathbb{E}\left[\left(X_{(n)}-X_{(n-1)}\right)^2\right] = \mathbb{E}\left[\Delta_{n-1}^2\right].}$$

(a2) Here, $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Uniform}[0,1]$.

for a random sample with CDF $F_X(x)$, the PDF of the order statistic is given by,

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!} f_X(x)\left[F_X(x)\right]^{k-1}\left[1-F_X(x)\right]^{n-k}$$

for $X_i \sim \mathrm{Uniform}[0,1]$, $F_X(x) = x$ and $f_X(x) = 1$ for $0 < x < 1$.

So,

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!} x^{k-1}(1-x)^{n-k}$$, which is a $\mathrm{Beta}(\alpha, \beta)$

distribution with $\alpha = k$ and $\beta = n-k+1$. Thus,

$$\boxed{X_{(k)} \sim \mathrm{Beta}(k, n-k+1)}$$ so $X_{(n)} \sim \mathrm{Beta}(n, 1)$.

Using the known variance of a Beta distribution,

$$\boxed{\mathrm{Var}(X_{(n)}) = \frac{n}{(n+1)^2(n+2)} \sim \frac{1}{n^2} \text{ for large } n \xrightarrow{n\to\infty} 0.}$$

Now, define the $n+1$ spacings in $[0,1]$:

$$D_0 = X_{(1)} - 0, \quad D_1 = X_{(2)} - X_{(1)}, \quad \dots, \quad D_{n-1} = X_{(n)} - X_{(n-1)}, \quad D_n = 1 - X_{(n)}.$$

Then $(D_0, \dots, D_n)$ has a Dirichlet$(1, \dots, 1)$ distribution on the simplex $\{d_k \geq 0, \sum d_k = 1\}$. We know, the marginal distributions of a K-dimensional Dirichlet distribution are univariate Beta distributions. So,

$$D_K \sim \mathrm{Beta}(1, n).$$

We have, $D_{n-1} = X_{(n)} - X_{(n-1)} = \Delta_{n-1}$. Thus,

$$\boxed{\Delta_{n-1} \sim \mathrm{Beta}(1,n).}$$

Now,

$$\mathbb{E}[(X_{(n)} - X_{(n-1)})^2] = \mathbb{E}[\Delta_{n-1}^2] = \mathrm{Var}(\Delta_{n-1}) + (\mathbb{E}[\Delta_{n-1}])^2.$$

Using known properties of Beta$(1,n)$:

$$\mathbb{E}[(X_{(n)} - X_{(n-1)})^2] = \frac{n}{(n+1)^2(n+2)} + \frac{1}{(n+1)^2}$$

$$= \frac{2n+2}{(n+1)^2(n+2)}$$

$$= \frac{2}{(n+1)(n+2)} \sim \frac{2}{n^2} \text{ for large } n.$$

Hence,

$$\boxed{\mathbb{E}[(X_{(n)} - X_{(n-1)})^2] = \frac{2}{(n+1)(n+2)} \xrightarrow{n\to\infty} 0.}$$

And, the ratio:

$$\frac{\mathbb{E}[\Delta_{n-1}^2]}{\mathrm{Var}(X_{(n)})} = \frac{2(n+1)}{n} = 2\left(1 + \frac{1}{n}\right) \xrightarrow{n\to\infty} 2. \quad \square$$

(a3) [Bonus] Here, $X_1, \dots, X_n \overset{iid}{\sim} Exp(1)$ with density $e^{-x}\mathbb{1}\{x \geq 0\}$.

We know, the minimum $X_{(1)}$ of n independent $Exp(1)$ r.v. is exponential with parameter n. Conditionally on $X_{(1)}$, $X_{(2)}$ is distributed as the sum of $X_{(1)}$ and $Exp(n-1)$, and so-on. So,

$$\begin{aligned} X_{(1)} &\sim Exp(n) \\ X_{(2)} \mid X_{(1)} &\sim X_{(1)} + Exp(n-1). \\ &\vdots \\ X_{(k)} \mid X_{(k-1)} &\sim X_{(k-1)} + Exp(n-k+1) \\ &\vdots \\ X_{(n)} \mid X_{(n-1)} &\sim X_{(n-1)} + Exp(1). \end{aligned}$$

So, we can write the marginal distribution of k^{th} order statistic as,

$$X_{(k)} = \sum_{i=1}^{k} Exp(n-i+1) \quad \Rightarrow X_{(n)} = \sum_{i=1}^{n} Exp(n-i+1)$$

We know, $Exp(\alpha) \sim \frac{Exp(1)}{\alpha}$. So,

$$X_{(n)} = \sum_{i=1}^{n} \frac{Exp(1)}{n-i+1} = Exp(1) \sum_{i=1}^{n} \frac{1}{n-i+1} = \sum_{j=1}^{n} \frac{Exp(1)}{j}.$$

Change of variable: $j = n-i+1$.

Then,

$$\boxed{Var(X_{(n)}) = \sum_{j=1}^{n} \frac{1}{j^2} \xrightarrow{n \to \infty} \frac{\pi^2}{6}.}$$

And,

$$X_{(n)} - X_{(n-1)} = Exp(1) \Rightarrow \mathbb{E}\left[(X_{(n)} - X_{(n-1)})^2\right] = Var(Exp(1)) + \left(\mathbb{E}[Exp(1)]\right)^2$$
$$= 2.$$

Hence, $\boxed{\mathbb{E}\left[(X_{(n)} - X_{(n-1)})^2\right] = 2.}$

And, the ratio: $\dfrac{\mathbb{E}[\Delta_{n-1}^2]}{Var(X_{(n)})} = 2\left(\sum_{j=1}^{n} \frac{1}{j^2}\right)^{-1} \xrightarrow{n \to \infty} \dfrac{12}{\pi^2}.$ $\square$

(b) [Bonus] General order statistics.

We generalize the previous result for any order statistic. Fix $k \in \{2, \dots, n-1\}$.

(b1) We want to prove that:

$$\mathrm{Var}(X_{(k)}) \le k\, \mathbb{E}[(X_{(k+1)} - X_{(k)})^2] = k\, \mathbb{E}[\Delta_k^2].$$

We claim, for any $i \in \{1, \dots, n\}$,

$$(X_{(k)}^{(i)} - X_{(k)})_+ \le (X_{(k+1)} - X_{(k)})\, \mathbb{1}\{X_i \le X_{(k)}\}.$$

- Case 1: $X_i \le X_{(k)}$. Here, $\mathbb{1}\{X_i \le X_{(k)}\} = 1$. We replace X_i by X_i'.
  - if $X_i' \le X_{(k)}$ then $X_{(k)}^{(i)} = X_{(k)}$, and the inequality holds.
  - if $X_i' > X_{(k)}$ then if $X_i' < X_{(k+1)}$, $X_{(k)}^{(i)} = X_i'$ and if $X_i > X_{(k+1)}$ $X_{(k)}^{(i)} = X_{(k+1)}$, so in both cases $(X_{(k)}^{(i)} - X_{(k)})_+ \le (X_{(k+1)} - X_{(k)})$ and the inequality holds.

- Case 2: $X_i > X_{(k)}$. Here, $\mathbb{1}\{X_i \le X_{(k)}\} = 0$. We replace X_i by X_i'.
  - if $X_i' \ge X_{(k)}$ then $X_{(k)}^{(i)} = X_{(k)}$, and the inequality holds.
  - if $X_i' < X_{(k)}$ then $X_{(k)}^{(i)} \le X_{(k)} \Rightarrow (X_{(k)}^{(i)} - X_{(k)})_+ = 0$ and the inequality holds.

Thus, for any given i,

$$\boxed{(X_{(k)}^{(i)} - X_{(k)})_+ \le (X_{(k+1)} - X_{(k)})\, \mathbb{1}\{X_i \le X_{(k)}\}.}$$

Summing over i,

$$\sum_{i=1}^{n} (X_{(k)}^{(i)} - X_{(k)})_+^2 \le (X_{(k+1)} - X_{(k)})^2 \sum_{i=1}^{n} \mathbb{1}\{X_i \le X_{(k)}\}.$$

Because we assumed the X_i's are a.s. distinct, the indicator sum $\le k$. So,

$$\sum_{i=1}^{n} (X_{(k)}^{(i)} - X_{(k)})_+^2 \le k (X_{(k+1)} - X_{(k)})^2.$$

Taking expectations and using the one-sided ESS on $-Z = -X_{(k)}$ gives,

$$\mathrm{Var}(X_{(k)}) \le \sum_{i=1}^{n} \mathbb{E}\left[\left(X_{(k)}^{(i)} - X_{(k)}\right)_+^2\right] \le k\, \mathbb{E}\left[\left(X_{(k+1)} - X_{(k)}\right)^2\right].$$

Hence,

$$\boxed{\mathrm{Var}(X_{(k)}) \le k\, \mathbb{E}\left[\left(X_{(k+1)} - X_{(k)}\right)^2\right] = k\, \mathbb{E}[\Delta_k^2].}$$

Now, we want to prove:

$$\mathrm{Var}(X_{(k)}) \le (n-k+1)\, \mathbb{E}\left[\left(X_{(k)} - X_{(k-1)}\right)^2\right] = (n-k+1)\, \mathbb{E}[\Delta_{k-1}^2].$$

We claim, for any $i \in \{1, \dots, n\}$,

$$\left(X_{(k)} - X_{(k)}^{(i)}\right)_+ \le \left(X_{(k)} - X_{(k-1)}\right) \mathbb{1}\{X_i \ge X_{(k)}\}.$$

- Case 1: $X_i \ge X_{(k)}$. Here, $\mathbb{1}\{X_i \ge X_{(k)}\} = 1$. Replace X_i by X_i'.
  - if $X_i' \ge X_{(k)}$ then $X_{(k)}^{(i)} = X_{(k)}$ and the inequality holds. ✓
  - if $X_i' < X_{(k)}$ then $X_{(k)}^{(i)} \ge X_{(k-1)} \Rightarrow X_{(k)} - X_{(k)}^{(i)} \le X_{(k)} - X_{(k-1)}$, and the inequality holds. ✓

- Case 2: $X_i < X_{(k)}$. Here, $\mathbb{1}\{X_i \ge X_{(k)}\} = 0$. Replace X_i by X_i'.
  - if $X_i' \le X_{(k)}$ then $X_{(k)}^{(i)} = X_{(k)}$ and the inequality holds. ✓
  - if $X_i' > X_{(k)}$ then $X_{(k)}^{(i)} > X_{(k)} \Rightarrow \left(X_{(k)} - X_{(k)}^{(i)}\right)_+ = 0$ and the inequality holds. ✓

Thus, for any given i,

$$\boxed{\left(X_{(k)} - X_{(k)}^{(i)}\right)_+ \le \left(X_{(k)} - X_{(k-1)}\right) \mathbb{1}\{X_i \ge X_{(k)}\}.}$$

Summing over i,

$$\sum_{i=1}^{n} \left(X_{(k)} - X_{(k)}^{(i)}\right)_+^2 \le \left(X_{(k)} - X_{(k-1)}\right)^2 \sum_{i=1}^{n} \mathbb{1}\{X_i \ge X_{(k)}\}.$$

Because we assumed the X_i's are a.s. distinct, the indicator sum $\le (n-k+1)$. So,

$$\sum_{i=1}^{n} \left(X_{(k)} - X_{(k)}^{(i)}\right)_+^2 \le (n-k+1)\left(X_{(k)} - X_{(k-1)}\right)^2.$$

Taking expectations and using the one-sided ESS on $Z = X_{(k)}$ gives,

$$\mathrm{Var}(X_{(k)}) \le \sum_{i=1}^{n} \mathbb{E}\left[\left(X_{(k)} - X_{(k)}^{(i)}\right)_+^2\right] \le (n-k+1)\,\mathbb{E}\left[\left(X_{(k)} - X_{(k-1)}\right)^2\right].$$

Hence,

$$\boxed{\mathrm{Var}(X_{(k)}) \le (n-k+1)\,\mathbb{E}\left[\left(X_{(k)} - X_{(k-1)}\right)^2\right] = (n-k+1)\,\mathbb{E}\left[\Delta_{k-1}^2\right].}$$

(b2) From (b1), we have

$$\mathrm{Var}(X_{(k)}) \le \min\left\{k\,\mathbb{E}[\Delta_k^2],\ (n-k+1)\,\mathbb{E}[\Delta_{k-1}^2]\right\}.$$

Then,

Choosing the first bound for $k \le \lfloor n/2 \rfloor$ and second for $k > \lfloor n/2 \rfloor$ yields:

$$\boxed{\mathrm{Var}(X_{(k)}) \le \begin{cases} k\,\mathbb{E}[\Delta_k^2], & 1 \le k \le \lfloor n/2 \rfloor, \\ (n-k+1)\,\mathbb{E}[\Delta_{k-1}^2], & \lfloor n/2 \rfloor < k \le n. \end{cases}}$$

Problem 3 : Rademacher processes : bounding the variance.

(a) A weak variance bound.

(a1) Here, $\mathbb{E}[\varepsilon_i] = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot(-1) = 0.$

and $\varepsilon_i^2 = 1$ with prob. 1 so $\mathbb{E}[\varepsilon_i^2] = 1$. Then,

$$\mathrm{Var}(\varepsilon_i) = \mathbb{E}[\varepsilon_i^2] - (\mathbb{E}[\varepsilon_i])^2$$
$$= 1 - 0 = 1.$$

Now, for fixed $t \in \mathbb{R}^n$,

$$\mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i t_i\right) = \sum_{i=1}^n \mathrm{Var}(\varepsilon_i t_i) \quad [\because \varepsilon_i\text{'s are independent.}]$$
$$= \sum_{i=1}^n t_i^2 \cdot \mathrm{Var}(\varepsilon_i)$$
$$= \sum_{i=1}^n t_i^2.$$

Taking the supremum over $t \in T$ yields:

$$\boxed{\sup_{t \in T} \mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i t_i\right) = \sup_{t\in T} \sum_{i=1}^n t_i^2 = \sigma^2.}$$

(a2) Here, we have,

$$Z = f(\varepsilon_1, \dots, \varepsilon_n) = \sup_{t\in T} \sum_{i=1}^n \varepsilon_i t_i \text{ , with independent inputs } \varepsilon_i.$$

For a given i, changing ε_i will atmost increase Z by $\sup_{t\in T} |t_i|$ or atmost decrease Z by $\sup_{t\in T} |t_i|$. So, the coordinate wise range of f is:

$$(D_i f)(\varepsilon_1, \dots, \varepsilon_n) = 2 \sup_{t \in T} |t_i|.$$

Now, using the bounded difference variance bound,

$$\begin{aligned}
\operatorname{Var}(Z) &\leq \frac{1}{4} \sum_{i=1}^{n} \left(2 \sup_{t \in T} |t_i|\right)^2 \\
&= \frac{1}{4} \sum_{i=1}^{n} 4 \cdot \sup_{t \in T} t_i^2 \\
&= \sum_{i=1}^{n} \sup_{t \in T} t_i^2 = \sigma_\infty^2 .
\end{aligned}$$

Hence,

$$\boxed{\operatorname{Var}(Z) \leq \sigma_\infty^2 .}$$

(b) [Bonus] A sharper variance bound.

(b1) We want to prove that $\operatorname{Var}(Z) \leq 2\sigma^2$.

Let $\varepsilon_1', \dots, \varepsilon_n'$ be an independent copy of $\varepsilon_1, \dots, \varepsilon_n$ and define

$$Z^{(i)} = \sup_{t \in T} \sum_{j=1}^{n} \varepsilon_j^{(i)} t_j \ , \quad \text{where} \quad \varepsilon^{(i)} = (\varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_i', \varepsilon_{i+1}, \dots, \varepsilon_n).$$

Let $t^* = t^*(\varepsilon) \in T$ be a measurable maximizer s.t. $Z = \sum_{j=1}^{n} \varepsilon_j t_j^*$.

Now, we claim: $(Z - Z^{(i)})_+ \leq |\varepsilon_i - \varepsilon_i'| \, |t_i^*|$.

Let $t^{**} \in T$ be a measurable maximizer s.t. $Z^{(i)} = \sum_{j=1}^{n} \varepsilon_j^{(i)} t_j^{**}$.

Then, we will have two cases:

- $\underline{t^* = t^{**}}$: $(Z - Z^{(i)})_+ = (\varepsilon_i t_i^* - \varepsilon_i' t_i^*)_+ \leq |\varepsilon_i - \varepsilon_i'| \, |t_i^*|$.

- $\underline{t^* \neq t^{**}}$: Here, $\sum_{j=1}^{n} \varepsilon_j^{(i)} t_j^{**} \geq \sum_{j=1}^{n} \varepsilon_j^{(i)} t_j^{*}$. So,

$$(Z - Z^{(i)})_+ = \left(\sum_{j=1}^{n} \varepsilon_j t_j^* - \sum_{j=1}^{n} \varepsilon_j^{(i)} t_j^{**} \right)_+ \leq \left(\sum_{j=1}^{n} \varepsilon_j t_j^* - \sum_{j=1}^{n} \varepsilon_j^{(i)} t_j^{*} \right)_+$$

$$= (\varepsilon_i t_i^* - \varepsilon_i' t_i^*)_+$$

$$\leq |\varepsilon_i - \varepsilon_i'| |t_i^*|.$$

So, we have,

$$(Z - Z^{(i)})_+ \leq |\varepsilon_i - \varepsilon_i'| |t_i^*| \Rightarrow (Z - Z^{(i)})_+^2 \leq (\varepsilon_i - \varepsilon_i')^2 (t_i^*)^2.$$

We know, ε_i and ε_i' are identical copies and independent, so

$$\mathbb{E}[(\varepsilon_i - \varepsilon_i')^2] = \mathbb{E}[\varepsilon_i^2 - 2\varepsilon_i \varepsilon_i' + (\varepsilon_i')^2]$$

$$= \mathbb{E}[\varepsilon_i^2] - 2\mathbb{E}[\varepsilon_i]\mathbb{E}[\varepsilon_i'] + \mathbb{E}[(\varepsilon_i')^2]$$

$$= 1 - 0 + 1 = 2,$$

because $\mathbb{E}[\varepsilon_i^2] = \mathbb{E}[(\varepsilon_i')^2] = 1$ and $\mathbb{E}[\varepsilon_i] = \mathbb{E}[\varepsilon_i'] = 0$.

Finally, using one-sided ESS on Z:

$$\mathrm{Var}(Z) \leq \sum_{i=1}^{n} \mathbb{E}[(Z - Z^{(i)})_+^2]$$

$$\leq \sum_{i=1}^{n} (t_i^*)^2 \, \mathbb{E}[(\varepsilon_i - \varepsilon_i')^2]$$

$$= 2 \sum_{i=1}^{n} (t_i^*)^2 \leq 2 \sup_{t \in T} \sum_{i=1}^{n} t_i^2 = 2\sigma^2.$$

Hence,

$$\boxed{\mathrm{Var}(Z) \leq 2\sigma^2.}$$

(b2) let $T = \{e_1, \dots, e_n\} \subset \mathbb{R}^n$ be the standard basis vectors of $\mathbb{R}^n$.

Then,

$$\sigma^2 = \sup_{t \in T} \sum_{i=1}^{n} t_i^2 = 1 \text{ and } \sigma_\infty^2 = \sum_{i=1}^{n} \sup_{t \in T} t_i^2 = n \Rightarrow \boxed{\frac{\sigma_\infty^2}{\sigma^2} = n.}$$

Problem 4 : Polynomial versus exponential moment method bounds

(a) Polynomial moments are at least as good.

We want to prove:

$$M(t) \leq C(t) \quad \text{for every } t > 0.$$

Fix $\lambda > 0$, then using power series,

$$\mathbb{E}[e^{\lambda Y}] = \mathbb{E}\left[\sum_{q=0}^{\infty} \frac{(\lambda Y)^q}{q!}\right].$$

Using monotone convergence, the sequence is bounded. So, using DCT:

$$\mathbb{E}[e^{\lambda Y}] = \sum_{q=0}^{\infty} \frac{\lambda^q}{q!} \mathbb{E}[Y^q]$$

$$= \sum_{q=0}^{\infty} \frac{(\lambda t)^q}{q!} \frac{\mathbb{E}[Y^q]}{t^q}.$$

Multiplying both sides by $e^{-\lambda t}$, we get,

$$e^{-\lambda t}\mathbb{E}[e^{\lambda Y}] = \sum_{q=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^q}{q!} \cdot \frac{\mathbb{E}[Y^q]}{t^q}.$$

Now, the PMF of a random variable $X \sim \text{Poisson}(\lambda t)$ is,

$$\mathbb{P}(X = q) = \frac{(\lambda t)^q e^{-\lambda t}}{q!} \Rightarrow \sum_{q=0}^{\infty} \frac{(\lambda t)^q e^{-\lambda t}}{q!} = 1.$$

And, Y is non-negative R.V. so $\mathbb{E}[Y^q] \geq 0$ for all $q \in \mathbb{Z}_+$. Then,

$$e^{-\lambda t}\mathbb{E}[e^{\lambda Y}] \geq \sum_{q=0}^{\infty} \frac{(\lambda t)^q e^{-\lambda t}}{q!} \cdot \inf_{q \in \mathbb{Z}_+} \frac{\mathbb{E}[Y^q]}{t^q}$$

$$\geq \inf_{q \in \mathbb{Z}_+} \frac{\mathbb{E}[Y^q]}{t^q} \cdot \sum_{q=0}^{\infty} \frac{(\lambda t)^q e^{-\lambda t}}{q!}$$

$$= \inf_{q \in \mathbb{Z}_+} \frac{\mathbb{E}[Y^q]}{t^q} = M(t).$$

Thus, for any $\lambda > 0$, $M(t) \leq e^{-\lambda t}\mathbb{E}[e^{\lambda Y}]$ and taking the infimum over $\lambda > 0$ gives:

$$\boxed{M(t) \leq \inf_{\lambda > 0} e^{-\lambda t}\mathbb{E}[e^{\lambda Y}] = C(t) \text{ for every } t > 0.}$$

(b) Chernoff is still useful.

let $Z = Y_1 + \cdots + Y_n$ where Y_i's are independent.

Then,

$$\mathbb{E}[e^{\lambda Z}] = \mathbb{E}[e^{\lambda \Sigma_i Y_i}] = \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda Y_i}\right]$$

$$= \prod_{i=1}^{n} \mathbb{E}[e^{\lambda Y_i}],$$

So the bound reduces to simple 1-dimensional calculations. However,

$\mathbb{E}[Z^q] = \mathbb{E}[(Y_1 + \cdots + Y_n)^q]$ expand into many mixed cross terms, making them messy or intractable to compute for large q.

Hence,

Chernoff bounds are still widely used, especially for sums of independent random variables.

Problem 5: Sub-Gaussian characterizations.

(a) Three sub-Gaussian properties.

Here, Z is a real-valued R.V. with $\mathbb{E}[Z]=0$ and the centered CGF is:

$$\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] \in (-\infty, \infty] \qquad (\lambda \in \mathbb{R}).$$

(a1) We want to prove:

$$\psi_Z(\lambda) \le \frac{\nu\lambda^2}{2} \;\; \forall \lambda \in \mathbb{R} \implies \mathbb{P}\{|Z| \ge t\} \le 2\exp\left(\frac{-t^2}{2\nu}\right) \;\; \forall\, t \ge 0.$$

Assume for all $\lambda \in \mathbb{R}$, $\psi_Z(\lambda) \le \frac{\nu\lambda^2}{2}$. Using Chernoff bound for Z:

$$\mathbb{P}\{Z \ge t\} \le \inf_{\lambda>0} \exp\left(-\lambda t + \psi_Z(\lambda)\right)$$

$$\le \inf_{\lambda>0} \exp\left(-\lambda t + \frac{\nu\lambda^2}{2}\right) \qquad [\because \text{using assumption.}]$$

$$\le \inf_{\lambda>0} \exp\left(f(\lambda)\right), \quad f(\lambda) = -\lambda t + \frac{\nu\lambda^2}{2}.$$

Here, $e>1$ i.e. the exponential function is stritctly monotic, so the minimum of $\exp(f(\lambda))$ will happen at the same value of λ where $f(\lambda)$ is minimized. So, setting $f'(\lambda)=0$ gives:

$$f'(\lambda) = -t + \nu\lambda = 0 \implies \lambda^* = \arg\min_{\lambda} f(\lambda) = \frac{t}{\nu} > 0.$$

Putting it back,

$$\mathbb{P}\{Z \ge t\} \le \exp\left(f(\lambda^*)\right) = \exp\left(\frac{-t^2}{2\nu}\right).$$

Thus, for all $t \ge 0$,

$$\boxed{\mathbb{P}\{Z \ge t\} \le \exp\left(\frac{-t^2}{2\nu}\right).}$$

Similarly, using Chernoff bound on $-Z$, we get

$$\mathbb{P}\{Z \le -t\} = \mathbb{P}\{-Z \ge t\} \le \inf_{\lambda > 0} \exp(-\lambda t + \psi_{-Z}(\lambda)).$$

Because the CGF bound works for all $\lambda \in \mathbb{R}$,

$$\psi_{-Z}(\lambda) = \log \mathbb{E}[e^{\lambda(-Z)}] = \psi_Z(-\lambda) \le \frac{\nu\lambda^2}{2}.$$

So,

$$\mathbb{P}\{Z \le -t\} \le \inf_{\lambda > 0} \exp\left(-\lambda t + \frac{\nu\lambda^2}{2}\right),$$ which is the same constraint optimization problem as before, so we deduce:

$$\boxed{\mathbb{P}\{Z \le -t\} \le \exp\left(\frac{-t^2}{2\nu}\right)} \quad \text{for all } t \ge 0.$$

Finally, using the union bound:

$$\mathbb{P}\{|Z| \ge t\} \le \mathbb{P}\{Z \ge t\} + \mathbb{P}\{Z \le -t\} = 2\exp\left(\frac{-t^2}{2\nu}\right).$$

Hence,

$$\boxed{\psi_Z(\lambda) \le \frac{\nu\lambda^2}{2} \ \forall \lambda \in \mathbb{R} \implies \mathbb{P}\{|Z| \ge t\} \le 2\exp\left(\frac{-t^2}{2\nu}\right) \ \forall t \ge 0.}$$

(a2) We assume, for all $t \ge 0$,

$$\mathbb{P}\{|Z| \ge t\} \le 2\exp(-t^2/2\nu).$$

We know, from Homework 1, problem 4(a2), if for all $t \ge 0$,

$$\mathbb{P}\{X \ge t\} \le C\exp(-ct^\alpha),$$

then $\|X\|_{L^p} = (\mathbb{E}[X^p])^{1/p} \le c' p^{1/\alpha} \ \forall p \ge 1$, where

- for $p \ge \alpha$, $c' = e\, C_0^{1/\alpha} (c\alpha)^{-1/\alpha} \max\{1, C/\alpha\}$
- for $p \le \alpha$, $c' = (C/c)^{1/\alpha}$.

So, in general we have, $\|X\|_{L_p} \leq c''(c^{-1}p)^{1/\alpha}$, c'' is another abs. const.

Here, for $\mathbb{P}\{|Z| \geq t\} \leq 2\exp(-t^2/2\nu)$,

$X = |Z|$, $\alpha = 2$, $c = \frac{1}{2\nu}$. So, using the result, we get,

$$(\mathbb{E}[|Z|^p])^{1/p} \leq c''(2\nu p)^{1/2} = c''\sqrt{2}\sqrt{\nu p}.$$

Hence, there exists a universal constant $K > 0$ such that,

$$\boxed{(\mathbb{E}[|Z|^p])^{1/p} \leq K\sqrt{\nu p}\ , \text{ for all } p \geq 1.}$$

(a3) [Bonus]. Assume, $\exists\, C > 0$ such that for all $p \geq 1$,

$$(\mathbb{E}[|Z|^p])^{1/p} \leq C\sqrt{\nu p}.$$

We then prove: for a universal constant $k > 0$,

$$\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] \leq \frac{k\nu\lambda^2}{2} \text{ for all } \lambda \in \mathbb{R}.$$

Expanding the MGF using power series:

$$\mathbb{E}[e^{\lambda Z}] = \sum_{K=0}^{\infty} \frac{\lambda^K \mathbb{E}[Z^K]}{K!}$$

$$= 1 + \sum_{K=1}^{\infty} \frac{\lambda^K \mathbb{E}[Z^K]}{K!} \qquad [\because \mathbb{E}[Z] = 0]$$

Using triangle inequality and noting $|a^n \cdot b| \leq |a|^n \cdot |b|$,

$$\mathbb{E}[e^{\lambda Z}] \leq 1 + \sum_{K=2}^{\infty} \frac{|\lambda|^K \cdot |\mathbb{E}[Z^K]|}{K!}$$

$$\leq 1 + \sum_{K=2}^{\infty} \frac{|\lambda|^K \mathbb{E}[|Z|^K]}{K!} \qquad [\because |\mathbb{E}Z^K| \leq \mathbb{E}|Z|^K].$$

$$\leq 1 + \sum_{K=2}^{\infty} \frac{|\lambda|^K (C\sqrt{\nu})^K K^{K/2}}{K!}. \qquad \text{[Using moment bound assumption]}$$

Using Stirling's bound, we know $k! \geq (k/e)^k$. So,

$$\mathbb{E}[e^{\lambda z}] \leq 1 + \sum_{k=2}^{\infty} \left(\frac{|\lambda| e C \sqrt{v}}{\sqrt{k}} \right)^k$$

$$= 1 + \sum_{k=2}^{\infty} \left(\frac{A}{\sqrt{k}} \right)^k, \quad A = |\lambda| e C \sqrt{v}.$$

Now, we split the sum at $k_0 = \lceil 4A^2 \rceil$. Then,

- For $k \geq 4A^2$: $\sqrt{k} \geq 2A \Rightarrow \frac{A}{\sqrt{k}} \leq \frac{1}{2}$. So,

$$\sum_{k=k_0}^{\infty} \left(\frac{A}{\sqrt{k}} \right)^k \leq \sum_{k=1}^{\infty} \left(\frac{1}{2} \right)^k = \frac{1/2}{1 - 1/2} = 1.$$

- For $2 \leq k < 4A^2$: let $f(x) = \left(\frac{A}{\sqrt{x}} \right)^x \Rightarrow \log f(x) = x \log A - \frac{x}{2} \log x$.

differentiating we get, $f'(x) = f(x) \left[\log A - \frac{1}{2} \log x - \frac{1}{2} \right] = 0$ gives $x^* = \arg\max_x f(x) = \frac{A^2}{e}$, and the maximum value is,

$$f(x^*) = \left(\frac{A}{\sqrt{A^2/e}} \right)^{A^2/e} = \exp\left(\frac{A^2}{2e} \right).$$

So,

$$\sum_{k=2}^{k_0 - 1} \left(\frac{A}{\sqrt{k}} \right)^k \leq \sum_{k=2}^{k_0 - 1} f(x^*) \leq 4A^2 \exp\left(\frac{A^2}{2e} \right).$$

Combining, we get,

$$\mathbb{E}[e^{\lambda z}] \leq 2 + 4A^2 \exp\left(\frac{A^2}{2e} \right) \leq \exp(c' A^2),$$ for sufficiently large constant c'. Substituting $A^2 = \lambda^2 e^2 C^2 v$ and taking log, we get,

$$\log \mathbb{E}[e^{\lambda z}] \leq e^2 c' C^2 v \lambda^2.$$

Define $K = 2e^2 c' C^2$, then $\log \mathbb{E}[e^{\lambda Z}] \leq \frac{K\nu\lambda^2}{2}$.

Hence, for a universal constant $K > 0$,

$$\boxed{\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] \leq \frac{K\nu\lambda^2}{2}, \ \forall \lambda \in \mathbb{R}.}$$

(b) Variance proxy.

We assume the CGF bound holds for some ν, i.e.,

$$\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] \leq \frac{\nu\lambda^2}{2} \quad \forall \lambda \in \mathbb{R}.$$

So, $\mathbb{E}[e^{\lambda Z}]$ is finite near 0. When $\mathbb{E}[e^{\lambda Z}] < \infty$ near 0, we know,

$$\psi_Z''(0) = \mathrm{Var}(Z).$$

Differentiating the CGF bound twice at $\lambda = 0$ gives,

$$\psi_Z''(0) \leq \nu \Rightarrow \mathrm{Var}(Z) \leq \nu. \quad \square$$

(c) [Bonus] Exponential-square integrability.

Here, we assume for all $t \geq 0$,

$$\mathbb{P}\{|Z| \geq t\} \leq 2\exp\left(\frac{-t^2}{2\nu}\right).$$

We want to show that, $\exists\, c > 0$ s.t. $\mathbb{E}\exp\left(\frac{Z^2}{c\nu}\right) \leq 2$.

For a non-decreasing $\Phi \geq 0$ with $\Phi(0) < \infty$, the tail integral identity for a non-negative random variable $|Z|$ is:

$$\mathbb{E}[\Phi(|Z|)] = \Phi(0) + \int_0^\infty \Phi'(t)\, \mathbb{P}\{|Z| \geq t\}\, dt.$$

Take $\Phi(t) = \exp\left(\frac{t^2}{cv}\right)$, then $\Phi(0) = 1$ and $\Phi'(t) = \frac{2t}{cv}\exp\left(\frac{t^2}{cv}\right)$.

Now, using the assumed tail bound gives,

$$\mathbb{E}\exp\left(\frac{Z^2}{cv}\right) \leq 1 + \int_0^\infty \frac{2t}{cv}\exp\left(\frac{t^2}{cv}\right) \cdot 2\exp\left(-\frac{t^2}{2v}\right) dt$$

$$= 1 + \frac{4}{cv}\int_0^\infty t\exp\left(t^2\left(\frac{1}{cv} - \frac{1}{2v}\right)\right) dt.$$

Choose $c > 2$ so that $\frac{1}{c} - \frac{1}{2} < 0$ and let $\alpha = \frac{1}{2v} - \frac{1}{cv} = \frac{c-2}{2cv} > 0$.

Then,

$$\mathbb{E}\exp\left(\frac{Z^2}{cv}\right) \leq 1 + \frac{4}{cv}\int_0^\infty t\exp(-\alpha t^2)\, dt$$

$$= 1 + \frac{4}{cv} \cdot \frac{1}{2\alpha}$$

$$= 1 + \frac{2}{cv} \cdot \frac{2cv}{c-2} = 1 + \frac{4}{c-2}.$$

Pick any $c \geq 6$, then $1 + \frac{4}{c-2} \leq 2$. Hence, we have

$$\boxed{\mathbb{E}\exp\left(\frac{Z^2}{cv}\right) \leq 2, \text{ for some constant } c \geq 6.}$$

Now,

We assume that $\mathbb{E}\exp\left(\frac{Z^2}{cv}\right) \leq 2$ for some $c > 0$. We then show:

$$\psi_Z(\lambda) = \log \mathbb{E}\left[e^{\lambda Z}\right] \leq \frac{c' v \lambda^2}{2} \text{ for all } \lambda \in \mathbb{R},\ c' > 0.$$

We will use Young's inequality: $|\lambda z| \leq \frac{a\lambda^2}{2} + \frac{z^2}{2a}$ for any $a > 0$.

Then,

$$\mathbb{E}[e^{\lambda z}] \leq \mathbb{E}[e^{|\lambda z|}] \leq \mathbb{E} \exp\left(\frac{a\lambda^2}{2} + \frac{z^2}{2a}\right)$$

$$= \exp\left(\frac{a\lambda^2}{2}\right) \cdot \mathbb{E} \exp\left(\frac{z^2}{2a}\right).$$

Choose $2a = cv \Rightarrow a = \frac{cv}{2}$ and use the given bound,

$$\mathbb{E}[e^{\lambda z}] \leq \exp\left(\frac{cv\lambda^2}{4}\right) \mathbb{E} \exp\left(\frac{z^2}{2a}\right) \leq 2 \exp\left(\frac{cv\lambda^2}{4}\right).$$

Taking logarithm,

$$\psi_z(\lambda) = \log \mathbb{E}[e^{\lambda z}] \leq \log 2 + \frac{cv\lambda^2}{4}.$$

For $|\lambda| \geq 1$, $\log 2 \leq \log 2 \cdot v\lambda^2$ so we can choose $c' = \frac{c}{2} + 2\log 2$ s.t.

$\psi_z(\lambda) \leq \frac{c'v\lambda^2}{2}$. For $|\lambda| < 1$, we can increase the constant infront of λ^2

to dominate $\log 2$ and choose c' s.t. $\psi_z(\lambda) \leq \frac{c'v\lambda^2}{2}$.

Hence, for some $c' > 0$,

$$\boxed{\psi_z(\lambda) = \log \mathbb{E}[e^{\lambda z}] \leq \frac{c'v\lambda^2}{2}.}$$