Theoretical Stats and Machine Learning (Homework 2)

Student Name: Vu Huyen Trang Pham

UT EID: vp22673

Question 1. Jackknife: variance and bias

a. Variance bound

a1. From the jackknife variance functional: $\widehat{V} = \sum_{i=1}^{n} \left(Z^{(i)} - \bar{Z}\right)^2, \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z^{(i)}$, we prove the identity:

$$\widehat{V} = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(Z^{(i)} - Z^{(j)}\right)^2$$

First, expanding the definition of $\widehat{V}$, we have:

$$\begin{aligned}\widehat{V} = \sum_{i=1}^{n}\left(Z^{(i)} - \bar{Z}\right)^2 &= \sum_{i=1}^{n}(Z^{(i)})^2 - 2\bar{Z}\sum_{i=1}^{n} Z^{(i)} + \sum_{i=1}^{n}\bar{Z}^2 \\ &= \sum_{i=1}^{n}(Z^{(i)})^2 - 2n\bar{Z}^2 + n\bar{Z}^2 = \sum_{i=1}^{n}(Z^{(i)})^2 - n\bar{Z}^2 \qquad \text{(because } \sum_{i=1}^{n} Z^{(i)} = n\bar{Z}\text{)}\end{aligned} \tag{1}$$

Next, expand the double sum, we obtain:

$$\begin{aligned}S &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(Z^{(i)} - Z^{(j)}\right)^2 \\ &= \sum_{i,j}\left((Z^{(i)})^2 + (Z^{(j)})^2 - 2Z^{(i)}Z^{(j)}\right)\end{aligned}$$

Evaluate each term separately, we have:

$$\begin{cases} \sum_{i,j}(Z^{(i)})^2 = n\sum_{i=1}^{n}(Z^{(i)})^2 \\ \sum_{i,j}(Z^{(j)})^2 = n\sum_{j=1}^{n}(Z^{(j)})^2 \\ \sum_{i,j} Z^{(i)}Z^{(j)} = \left(\sum_{i=1}^{n} Z^{(i)}\right)^2 \end{cases}$$

Therefore, we have:

$$\begin{aligned}
S &= 2n\sum_{i=1}^{n}(Z^{(i)})^2 - 2\left(\sum_{i=1}^{n} Z^{(i)}\right)^2 \\
&= 2n\sum_{i=1}^{n}(Z^{(i)})^2 - 2n^2\bar{Z}^2 \qquad \left(\text{because } \sum_{i=1}^{n} Z^{(i)} = n\bar{Z}\right) \qquad (2) \\
\iff \frac{S}{2n} &= \sum_{i=1}^{n}(Z^{(i)})^2 - n\bar{Z}^2
\end{aligned}$$

From 1 and 2, we conclude:

$$\boxed{\widehat{V} = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(Z^{(i)} - Z^{(j)}\right)^2}$$

a2. Let $W := f_{n-1}(X_1, \ldots, X_{n-1})$; $Z^{(i)} := f_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$; $\bar{Z} := \frac{1}{n}\sum_{i=1}^{n} Z^{(i)}$. Because $(X_1, \ldots, X_{n-1}) \overset{d}{=} (X_2, \ldots, X_n)$, we apply the same statistic f_{n-1} to obtain:

$$W \overset{d}{=} Z^{(1)} \quad \Longrightarrow \quad \mathrm{Var}(W) = \mathrm{Var}(Z^{(1)})$$

Therefore, instead of finding the variance bound of W, we find the variance bound of $Z^{(1)}$ instead. Now, consider $Z^{(1)} = f_{n-1}(X_2, \ldots, X_n)$ as a function of the independent inputs $X_2, \ldots, X_n$. By variance tensorization (Efron-Stein), we have:

$$\mathrm{Var}(Z^{(1)}) \le \sum_{k=2}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z^{(1)} \,\middle|\, X_{-k}\right)\right], \qquad X_{-k} := (X_2, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n) \qquad (3)$$

We then $k \in \{2, \ldots, n\}$ and condition on X_{-k}. Then all arguments of f_{n-1} are fixed except the single variable X_k, so there exists a (random) function $g = g_{k,X_{-k}}$ such that:

$$Z^{(1)} = g(X_k) \qquad \text{(under the conditioning on } X_{-k}\text{)}$$

Moreover, we have: $Z^{(k)} = f_{n-1}(X_1, X_2, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n) = g(X_1)$ because under the same conditioning the only unfixed argument is X_1, and f_{n-1} is symmetric so the position of the unfixed argument does not matter.

We note that under the conditioning on X_{-k}, the variables X_1 and X_k are i.i.d. and independent of X_{-k}. Therefore, $g(X_k)$ and $g(X_1)$ are iid and independent of X_{-k}, which implies:

$$\mathrm{Var}\Big(Z^{(1)}\,\Big|\,X_{-k}\Big) = \mathrm{Var}(g(X_k)\,|\,X_{-k}) = \frac{1}{2}\mathbb{E}\Big[\big(g(X_k)-g(X_1)\big)^2\,\Big|\,X_{-k}\Big] = \frac{1}{2}\mathbb{E}\Big[\big(Z^{(1)}-Z^{(k)}\big)^2\,\Big|\,X_{-k}\Big]$$

Replace $\mathrm{Var}\big(Z^{(1)}\,|\,X_{-k}\big)$ obtain above to 3, we have:

$$\begin{aligned}
&\mathrm{Var}(Z^{(1)}) \le \sum_{k=2}^{n}\mathbb{E}\left[\frac{1}{2}\mathbb{E}\Big[(Z^{(1)}-Z^{(k)})^2 \mid X_{-k}\Big]\right] \\
\iff &\mathrm{Var}(Z^{(1)}) \le \frac{1}{2}\sum_{k=2}^{n}\mathbb{E}\left[\mathbb{E}\Big[(Z^{(1)}-Z^{(k)})^2 \mid X_{-k}\Big]\right] \\
\implies &\mathrm{Var}(Z^{(1)}) \le \frac{1}{2}\sum_{k=2}^{n}\mathbb{E}\Big[(Z^{(1)}-Z^{(k)})^2\Big] \qquad \text{(tower property)}
\end{aligned} \tag{4}$$

Because the sample is i.i.d. and f_{n-1} is symmetric, the vector $(Z^{(1)},\ldots,Z^{(n)})$ is exchangeable. Therefore $\mathbb{E}[(Z^{(i)}-Z^{(j)})^2]$ is the same for all $i \neq j$, which implies that:

$$\begin{aligned}
&\sum_{k=2}^{n}\mathbb{E}\Big[\big(Z^{(1)}-Z^{(k)}\big)^2\Big] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\Big[\big(Z^{(i)}-Z^{(j)}\big)^2\Big] \\
\implies &\mathrm{Var}(Z^{(1)}) \le \frac{1}{2}\sum_{k=2}^{n}\mathbb{E}\Big[(Z^{(1)}-Z^{(k)})^2\Big] = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\Big[\big(Z^{(i)}-Z^{(j)}\big)^2\Big]
\end{aligned} \tag{5}$$

From part (a1), we have:

$$\begin{aligned}
\widehat{V} = \sum_{i=1}^{n}(Z^{(i)}-\bar{Z})^2 &= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}(Z^{(i)}-Z^{(j)})^2 \\
\implies \mathbb{E}[\widehat{V}] &= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\Big[\big(Z^{(i)}-Z^{(j)}\big)^2\Big]
\end{aligned}$$

Combine with 4 and 5, we conclude:

$$\boxed{\mathrm{Var}(W) = \mathrm{Var}(Z^{(1)}) \le \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\Big[\big(Z^{(i)}-Z^{(j)}\big)^2\Big] = \mathbb{E}[\widehat{V}]}$$

b. [Bonus] Exact bias for quadratic statistics

Let $Z_n = f_n(X_1, \dots, X_n)$ and assume $\mathbb{E}Z_n - \theta = \frac{c}{n} + O(n^{-2}) \quad (n \to \infty)$.

For $i = 1, \dots, n$, we define the leave-one-out statistic:

$$Z^{(i)} := f_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), \qquad \bar{Z} := \frac{1}{n}\sum_{i=1}^{n} Z^{(i)}$$

By exchangeability of $(X_1, \dots, X_n)$, for each i we have $Z^{(i)} \xrightarrow{d} f_{n-1}(X_1, \dots, X_{n-1}) =: Z_{n-1}$. Therefore:

$$\mathbb{E}\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}Z^{(i)} = \mathbb{E}Z_{n-1}$$

The jackknife bias-corrected estimator is $\tilde{Z}_n := nZ_n - (n-1)\bar{Z}$.

Take expectations and use the previous identity, we obtain:

$$\mathbb{E}\tilde{Z}_n = n\,\mathbb{E}Z_n - (n-1)\,\mathbb{E}Z_{n-1} \tag{6}$$

We now write the bias expansions in a form that isolates the $O(n^{-2})$ term: there exist sequences (r_n) and (r_{n-1}) with $r_n = O(n^{-2})$ and $r_{n-1} = O((n-1)^{-2}) = O(n^{-2})$ such that:

$$\mathbb{E}Z_n = \theta + \frac{c}{n} + r_n, \qquad \mathbb{E}Z_{n-1} = \theta + \frac{c}{n-1} + r_{n-1}$$

Plug in 6 to compute the expectation of the bias-corrected estimator, we have:

$$\begin{aligned}
\mathbb{E}\tilde{Z}_n &= n\Big(\theta + \frac{c}{n} + r_n\Big) - (n-1)\Big(\theta + \frac{c}{n-1} + r_{n-1}\Big) \\
&= \big(n\theta + c + nr_n\big) - \big((n-1)\theta + c + (n-1)r_{n-1}\big) \\
&= \theta + \underbrace{nr_n - (n-1)r_{n-1}}_{(\star)}
\end{aligned}$$

Our task now is to bound $(\star)$. Because $r_n = O(n^{-2})$, we have $nr_n = O(n^{-1})$. Similarly, $(n-1)r_{n-1} = O(n^{-1})$, which implies that $(\star) = nr_n - (n-1)r_{n-1} = O(n^{-1}) - O(n^{-1})$.

To see this difference, we use the sharper representation $r_n = \frac{a}{n^2} + O(n^{-3})$ for some constant a

(which is equivalent to $r_n = O(n^{-2})$ together with existence of a second-order expansion). We have:

$$nr_n = \frac{a}{n} + O(n^{-2}), \qquad (n-1)r_{n-1} = \frac{a}{n-1} + O(n^{-2})$$

Therefore $(\star) = \left(\frac{a}{n} - \frac{a}{n-1}\right) + O(n^{-2})$.

Because $\frac{1}{n} - \frac{1}{n-1} = \frac{(n-1)-n}{n(n-1)} = -\frac{1}{n(n-1)} = O(n^{-2})$, we have $(\star) = O(n^{-2})$. This leads us to the final conclusion: $\boxed{\mathbb{E}\tilde{Z}_n - \theta = O(n^{-2})}$.

Question 2: Order statistics: variance versus spacings

a. The maximum

Let $Z := X_{(n)}$ and for each $i \in \{1, \dots, n\}$ let $Z^{(i)} := X_{(n)}^{(i)}$ be the maximum of the resampled vector $(X_1^{(i)}, \dots, X_n^{(i)})$ where $X_j^{(i)} = X_j$ for $j \neq i$ and $X_i^{(i)} = X_i'$.

Recall the one-sided ESS inequality:

$$\mathrm{Var}(Z) \;\leq\; \sum_{i=1}^{n} \mathbb{E}\Big[(Z - Z^{(i)})_+^2\Big], \qquad (a)_+ := \max\{a, 0\}.$$

a1. First of all, we prove the pointwise bound:

$$\big(X_{(n)} - X_{(n)}^{(i)}\big)_+ \;\leq\; \big(X_{(n)} - X_{(n-1)}\big)\,\mathbf{1}\{X_i = X_{(n)}\}. \tag{7}$$

Fix an outcome of the experiment.

Case 1: $X_i \neq X_{(n)}$. Then the (original) maximum $X_{(n)}$ is attained by some index $j \neq i$, so X_j is unchanged in the resampled vector. Therefore, $X_{(n)}^{(i)} \geq X_j = X_{(n)}$. We then have:

$$\big(X_{(n)} - X_{(n)}^{(i)}\big)_+ = 0 \leq \big(X_{(n)} - X_{(n-1)}\big)\,\mathbf{1}\{X_i = X_{(n)}\} = 0$$

Case 2: $X_i = X_{(n)}$. Then, after replacing X_i by X_i', all the other $n-1$ values remain the same, so in particular the second-largest order statistic $X_{(n-1)}$ remains present in the resampled vector. This means $X_{(n)}^{(i)} \geq X_{(n-1)}$.

Consequently, we have:

$$X_{(n)} - X_{(n)}^{(i)} \leq X_{(n)} - X_{(n-1)} =: \Delta_{n-1}$$

and because $(\cdot)_+$ is nonnegative and bounded above by the argument when the argument is positive, we obtain:

$$\big(X_{(n)} - X^{(i)}_{(n)}\big)_+ \le \Delta_{n-1} = \big(X_{(n)} - X_{(n-1)}\big)\,\mathbf{1}\{X_i = X_{(n)}\}$$

This proves (7).

Now, squaring (7), we obtain:

$$\begin{gathered}\big(X_{(n)} - X^{(i)}_{(n)}\big)_+^2 \le \Delta_{n-1}^2\,\mathbf{1}\{X_i = X_{(n)}\}\\ \implies \sum_{i=1}^{n} \mathbb{E}\Big[\big(X_{(n)} - X^{(i)}_{(n)}\big)_+^2\Big] \le \mathbb{E}\left[\Delta_{n-1}^2 \sum_{i=1}^{n} \mathbf{1}\{X_i = X_{(n)}\}\right]\end{gathered}$$

If the maximum is a.s. unique (e.g. the X_i are a.s. distinct), then $\sum_{i=1}^{n} \mathbf{1}\{X_i = X_{(n)}\} = 1$ almost surely, hence the right-hand side equals $\mathbb{E}[\Delta_{n-1}^2]$. Applying one-sided ESS, we obtain:

$$\mathrm{Var}(X_{(n)}) \le \sum_{i=1}^{n} \mathbb{E}\Big[\big(X_{(n)} - X^{(i)}_{(n)}\big)_+^2\Big] \le \mathbb{E}[\Delta_{n-1}^2] = \mathbb{E}\big[(X_{(n)} - X_{(n-1)})^2\big]$$

We prove that the bound still holds even if ties occur, when $X_{(n)} = X_{(n-1)}$.

On the event $\{X_{(n)} = X_{(n-1)}\}$, the maximum is attained by at least two indices. Hence, for any i, after replacing X_i by X_i', there remains at least one observation equal to $X_{(n)}$ in the resampled vector. As a result, $X^{(i)}_{(n)} \ge X_{(n)}$ and therefore $(X_{(n)} - X^{(i)}_{(n)})_+ = 0$.

Because $\Delta_{n-1} = 0$ on this event, it follows that:

$$\sum_{i=1}^{n} \mathbb{E}\Big[(X_{(n)} - X^{(i)}_{(n)})_+^2 \mathbf{1}\{X_{(n)} = X_{(n-1)}\}\Big] = \mathbb{E}\big[\Delta_{n-1}^2 \mathbf{1}\{X_{(n)} = X_{(n-1)}\}\big]\,.$$

Combining this with the argument on the event $\{X_{(n)} > X_{(n-1)}\}$, we conclude that $\mathrm{Var}(X_{(n)}) \le \mathbb{E}[\Delta_{n-1}^2]$ holds without requiring uniqueness of the maximum.

a2. Let $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Unif}(0,1)$ and denote $X_{(1)} \le \dots \le X_{(n)}$.

For the maximum $X_{(n)}$, we have for $0 \le x \le 1$:

$$\mathbb{P}(X_{(n)} \le x) = \mathbb{P}(X_1 \le x, \dots, X_n \le x) = x^n,$$

so $X_{(n)}$ has density $f(x) = nx^{n-1}$ on $(0, 1)$. Therefore, we have:

$$\begin{cases} \mathbb{E}[X_{(n)}] = \int_0^1 x f_{X_{(n)}}(x)\,dx = \int_0^1 x \cdot nx^{n-1}\,dx = n\int_0^1 x^n\,dx = \dfrac{n}{n+1}, \\ \mathbb{E}[X_{(n)}^2] = \int_0^1 x^2 f_{X_{(n)}}(x)\,dx = \int_0^1 x^2 \cdot nx^{n-1}\,dx = n\int_0^1 x^{n+1}\,dx = \dfrac{n}{n+2} \end{cases}$$

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

Now we compute $\mathbb{E}[(X_{(n)} - X_{(n-1)})^2] = \mathbb{E}[\Delta_{n-1}^2]$.

Step 1: Joint density of $(X_{(n-1)}, X_{(n)})$. For i.i.d. samples with CDF F and density f, the joint density of $(X_{(k)}, X_{(l)})$ $(1 \le k < l \le n)$ is written as follows:

$$f_{X_{(k)},X_{(l)}}(x, y) = \frac{n!}{(k-1)!(l-k-1)!(n-l)!}\,[F(x)]^{k-1}\,[F(y)-F(x)]^{l-k-1}\,[1-F(y)]^{n-l}\,f(x)f(y), \quad x < y.$$

In our case $F(t) = t$ and $f(t) = 1$ on $(0, 1)$, and we take $k = n-1$, $l = n$. Then $(l-k-1) = 0$ and $(n-l) = 0$, therefore for $0 < x < y < 1$, we have:

$$f_{X_{(n-1)},X_{(n)}}(x, y) = \frac{n!}{(n-2)!\,0!\,0!}\,x^{n-2} = n(n-1)\,x^{n-2}, \qquad 0 < x < y < 1.$$

Step 2: Marginal density of $\Delta_{n-1} = X_{(n)} - X_{(n-1)}$. Let $\Delta := y - x$. For a fixed $\Delta \in (0, 1)$, the constraint $0 < x < y < 1$ becomes $0 < x < 1 - \Delta$ with $y = x + \Delta$. Therefore the density of Δ is obtained by integrating out x:

$$f_\Delta(\Delta) = \int_0^{1-\Delta} f_{X_{(n-1)},X_{(n)}}\big(x, x+\Delta\big)\,dx = \int_0^{1-\Delta} n(n-1)\,x^{n-2}\,dx = n(1-\Delta)^{n-1}, \qquad 0 < \Delta < 1.$$

This is exactly the $\mathrm{Beta}(1, n)$ density.

Step 3: Compute $\mathbb{E}[\Delta_{n-1}^2]$. Because $\Delta_{n-1} \sim \mathrm{Beta}(1, n)$, we use the known formulas for the mean and variance of a $\mathrm{Beta}(a, b)$ random variable:

$$\mathbb{E}[Y] = \frac{a}{a+b}, \qquad \mathrm{Var}(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

With $a = 1$ and $b = n$, we obtain:

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

Therefore:

$$\begin{aligned}\mathbb{E}[\Delta_{n-1}^2] &= \mathrm{Var}(\Delta_{n-1}) + \big(\mathbb{E}[\Delta_{n-1}]\big)^2 = \frac{n}{(n+1)^2(n+2)} + \frac{1}{(n+1)^2} \\ &= \frac{n+(n+2)}{(n+1)^2(n+2)} = \frac{2}{(n+1)(n+2)}\end{aligned}$$

This implies $\mathbb{E}[\Delta_{n-1}^2] \sim \frac{2}{n^2} \to 0$ as $n \to \infty$.

Ratio with $\mathrm{Var}(X_{(n)})$ *and limit.*

We have:

$$\frac{\mathbb{E}[\Delta_{n-1}^2]}{\mathrm{Var}(X_{(n)})} = \frac{\frac{2}{(n+1)(n+2)}}{\frac{n}{(n+1)^2(n+2)}} = \frac{2(n+1)}{n} \longrightarrow 2 \qquad \text{as } n \to \infty.$$

a3. [Bonus] Assume $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathrm{Exp}(1)$ with density $e^{-x}\mathbf{1}\{x \geq 0\}$. We compute explicitly $\mathrm{Var}(X_{(n)})$ and $\mathbb{E}[(X_{(n)} - X_{(n-1)})^2]$, then the limits and ratio.

Step 1: Spacings representation for exponential order statistics. Define $X_{(0)} := 0$ and spacings $D_k := X_{(k)} - X_{(k-1)}$ for $k = 1, \dots, n$. Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathrm{Exp}(1)$, with order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$, and define $X_{(0)} := 0$, $D_k := X_{(k)} - X_{(k-1)}$. Then $D_k \sim \mathrm{Exp}(n-k+1)$ and $D_1, \dots, D_n$ are independent. We now prove that:

$$D_k \;\sim\; \mathrm{Exp}(n-k+1) \qquad \text{and} \qquad D_1, \dots, D_n \text{ are independent.}$$

Indeed, first, we have:

$$\mathbb{P}(D_1 > t) = \mathbb{P}(X_{(1)} > t) = \prod_{i=1}^{n} \mathbb{P}(X_i > t) = (e^{-t})^n = e^{-nt}$$

so $D_1 \sim \mathrm{Exp}(n)$.

Let $T := X_{(1)}$ and let I be the index attaining the minimum (assume that it is unique). For each

$j \neq I$, define the residual $Y_j := X_j - T$. By the memoryless property of Exp(1), we have:

$$(Y_j \mid X_j > T) \ \sim \ \text{Exp}(1)$$

and the residuals $\{Y_j : j \neq I\}$ are conditionally independent given (I, T). Therefore:

$$D_2 = X_{(2)} - X_{(1)} = \min_{j \neq I} Y_j \ \sim \ \text{Exp}(n-1)$$

Because for all i and t, we have $\mathbb{P}(D_2 > s \mid I = i,\, T = t) = e^{-(n-1)s}$. which does not depend on (i, t), we have:

$$\mathbb{P}(D_2 > s \mid D_1 = t) = \mathbb{P}(D_2 > s \mid T = t) = \sum_{i=1}^{n} \mathbb{P}(D_2 > s \mid I = i,\, T = t)\, \mathbb{P}(I = i \mid T = t) = e^{-(n-1)s}.$$

Therefore, the conditional law of D_2 given $D_1 = t$ does not depend on t, so D_2 is independent of D_1. Repeat the same argument after the second order statistic: after time $X_{(2)}$, there remain $n-2$ residual waiting times, which (by memorylessness) are i.i.d. Exp(1) and independent of (D_1, D_2). Therefore:

$$D_3 = \min(n-2 \text{ i.i.d. Exp}(1) \text{ residuals}) \sim \text{Exp}(n-2)$$

and D_3 is independent of (D_1, D_2). Continuing inductively for all k, we obtain:

$$D_k \sim \text{Exp}(n-k+1) \quad \text{and} \quad D_k \perp (D_1, \ldots, D_{k-1})$$

so $D_1, \ldots, D_n$ are mutually independent.

Apply the above result:

$$D_k \ \sim \ \text{Exp}(n-k+1) \qquad \text{and} \qquad D_1, \ldots, D_n \text{ are independent.}$$

Therefore $X_{(n)} \ = \ \sum_{k=1}^{n} D_k$.

Step 2: Compute $\mathrm{Var}(X_{(n)})$. If $Y \sim \mathrm{Exp}(\lambda)$ (rate λ), then $\mathrm{Var}(Y) = 1/\lambda^2$. We have:

$$\mathrm{Var}(X_{(n)}) \;=\; \mathrm{Var}\left(\sum_{k=1}^{n} D_k\right) \;=\; \sum_{k=1}^{n} \mathrm{Var}(D_k) \;=\; \sum_{k=1}^{n} \frac{1}{(n-k+1)^2} \;=\; \sum_{j=1}^{n} \frac{1}{j^2}$$

Step 3: Compute $\mathbb{E}[(X_{(n)} - X_{(n-1)})^2]$. We have $X_{(n)} - X_{(n-1)} \;=\; D_n$. Moreover, $D_n \sim \mathrm{Exp}(1)$ (because $n - n + 1 = 1$). For $Y \sim \mathrm{Exp}(\lambda)$, we have:

$$\mathbb{E}[Y^2] \;=\; \mathrm{Var}(Y) + (\mathbb{E}Y)^2 \;=\; \frac{1}{\lambda^2} + \frac{1}{\lambda^2} \;=\; \frac{2}{\lambda^2}$$

Therefore, $\mathbb{E}[(X_{(n)} - X_{(n-1)})^2] \;=\; \mathbb{E}[D_n^2] \;=\; 2.$

Step 4: Limits and ratio. We have

$$\mathrm{Var}(X_{(n)}) \;=\; \sum_{j=1}^{n} \frac{1}{j^2} \;\xrightarrow[n\to\infty]{}\; \sum_{j=1}^{\infty} \frac{1}{j^2} \;=\; \frac{\pi^2}{6}$$

and $\mathbb{E}[(X_{(n)} - X_{(n-1)})^2] \;=\; 2$ for all n. Therefore, the ratio of RHS to LHS equals

$$\frac{\mathbb{E}[(X_{(n)} - X_{(n-1)})^2]}{\mathrm{Var}(X_{(n)})} \;=\; \frac{2}{\sum_{j=1}^{n} j^{-2}} \;\xrightarrow[n\to\infty]{}\; \frac{2}{\pi^2/6} \;=\; \frac{12}{\pi^2}$$

So for Exp(1), the ESS bound $\mathrm{Var}(X_{(n)}) \le \mathbb{E}[\Delta_{n-1}^2]$ is valid and asymptotically within a constant factor $12/\pi^2$.

b. [Bonus] General order statistics.

b1. Fix $k \in \{2, \ldots, n-1\}$. We prove the two inequalities

$$\mathrm{Var}(X_{(k)}) \;\le\; k\,\mathbb{E}[(X_{(k+1)} - X_{(k)})^2] \qquad \text{and} \qquad \mathrm{Var}(X_{(k)}) \;\le\; (n-k+1)\,\mathbb{E}[(X_{(k)} - X_{(k-1)})^2]$$

Throughout, we may assume $X_1, \ldots, X_n$ are a.s. distinct (ties can be handled; when a spacing is 0 the bounds are automatic).

Part 1: Upper spacing bound $\mathrm{Var}(X_{(k)}) \le k\,\mathbb{E}[\Delta_k^2]$. Set $Z := X_{(k)}$. We will show the *pointwise* inequality

$$\big(X_{(k)}^{(i)} - X_{(k)}\big)_+ \;\le\; \big(X_{(k+1)} - X_{(k)}\big)\,\mathbf{1}\{X_i \le X_{(k)}\}. \tag{8}$$

Why (8) *is true.* Fix a realization of $(X_1, \ldots, X_n)$ and consider resampling only coordinate i.

- If $X_i > X_{(k)}$, then the value X_i lies *above* the kth order statistic. Removing it and replacing it by X_i' cannot *increase* the kth order statistic, because we are potentially taking away a large element, not a small one. Hence $X_{(k)}^{(i)} \le X_{(k)}$, so the positive part is 0:

$$\big(X_{(k)}^{(i)} - X_{(k)}\big)_+ \;=\; 0 \qquad \text{when } X_i > X_{(k)}.$$

  This matches the RHS of (8) because $\mathbf{1}\{X_i \le X_{(k)}\} = 0$.

- If $X_i \le X_{(k)}$, Then X_i is among the k smallest sample points. Removing it may cause the kth order statistic to move either upward or downward, depending on the value of X_i'. However, if it increases (i.e. if $X_{(k)}^{(i)} > X_{(k)}$), the increase is limited.

  Indeed, if $X_{(k)}^{(i)} > X_{(k)}$ after deleting one of the k smallest points, among the remaining $n-1$ original values $\{X_j : j \neq i\}$ the kth smallest value is at most the original $(k+1)$st order statistic $X_{(k+1)}$. Replacing the deleted value by X_i' cannot make the new kth order statistic exceed the larger of the kth smallest among the unchanged values, and the inserted value X_i'. Therefore, whenever $X_{(k)}^{(i)} > X_{(k)}$ we must have $X_{(k)}^{(i)} \le X_{(k+1)}$. Therefore, for $X_{(k)}^{(i)} > X_{(k)}$:

$$0 < X_{(k)}^{(i)} - X_{(k)} \le X_{(k+1)} - X_{(k)}$$

  If instead $X_{(k)}^{(i)} \le X_{(k)}$, then the positive part is again zero, which implies that in all cases:

$$\big(X_{(k)}^{(i)} - X_{(k)}\big)_+ \;\le\; X_{(k+1)} - X_{(k)}.$$

  Because $\mathbf{1}\{X_i \le X_{(k)}\} = 1$ in this case, (8) follows.

  Combining the two cases proves the desired pointwise inequality.

Apply one-sided ESS for $-X_{(k)}$, we obtain:

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

By (8), we have:

$$\big(X^{(i)}_{(k)} - X_{(k)}\big)_+^2 \;\le\; \big(X_{(k+1)} - X_{(k)}\big)^2\, \mathbf{1}\{X_i \le X_{(k)}\}.$$

Summing over i and taking expectation, we obtain:

$$\begin{aligned}\mathrm{Var}(X_{(k)}) &\le \sum_{i=1}^{n} \mathbb{E}\Big[\big(X_{(k+1)} - X_{(k)}\big)^2\, \mathbf{1}\{X_i \le X_{(k)}\}\Big] \\ &= \mathbb{E}\left[\big(X_{(k+1)} - X_{(k)}\big)^2 \sum_{i=1}^{n} \mathbf{1}\{X_i \le X_{(k)}\}\right]\end{aligned}$$

Moreover, we have $\sum_{i=1}^n \mathbf{1}\{X_i \le X_{(k)}\} \;=\; k$. Therefore:

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

Part 2: We prove the lower spacing bound $\mathrm{Var}(X_{(k)}) \le (n-k+1)\mathbb{E}[\Delta_{k-1}^2]$. Again set $Z := X_{(k)}$, but now use the original one-sided ESS:

$$\mathrm{Var}(X_{(k)}) \;\le\; \sum_{i=1}^{n} \mathbb{E}\Big[\big(X_{(k)} - X^{(i)}_{(k)}\big)_+^2\Big].$$

We show the pointwise inequality

$$\big(X_{(k)} - X^{(i)}_{(k)}\big)_+ \;\le\; \big(X_{(k)} - X_{(k-1)}\big)\, \mathbf{1}\{X_i \ge X_{(k)}\}. \tag{9}$$

The reasoning is symmetric to before:

- If $X_i < X_{(k)}$, then removing a small element cannot decrease the kth order statistic; hence $X^{(i)}_{(k)} \ge X_{(k)}$ and the positive part is 0, matching RHS since $\mathbf{1}\{X_i \ge X_{(k)}\} = 0$.

- If $X_i \ge X_{(k)}$, then X_i lies among the $n-k+1$ largest points. Removing it can only push $X_{(k)}$ downward, but it cannot drop below the previous order statistic $X_{(k-1)}$ of the original sample. Therefore the decrease is at most $X_{(k)} - X_{(k-1)}$, which implies (9).

Squaring (9), summing over i, and take the expectation, we obtain:

$$\mathrm{Var}(X_{(k)}) \le \mathbb{E}\left[\left(X_{(k)} - X_{(k-1)}\right)^2 \sum_{i=1}^{n} \mathbf{1}\{X_i \ge X_{(k)}\}\right]$$

There are $n-k+1$ points satisfy $X_i \ge X_{(k)}$, which means that the sum equals $n-k+1$. Therefore:

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

b2. From *b1*, we have two valid upper bounds:

$$\mathrm{Var}(X_{(k)}) \;\le\; k\,\mathbb{E}[\Delta_k^2] \qquad \text{and} \qquad \mathrm{Var}(X_{(k)}) \;\le\; (n-k+1)\,\mathbb{E}[\Delta_{k-1}^2].$$

Therefore, for every k, $\mathrm{Var}(X_{(k)}) \;\le\; \min\Big\{\, k\,\mathbb{E}[\Delta_k^2],\; (n-k+1)\,\mathbb{E}[\Delta_{k-1}^2] \,\Big\}$.

We split at $\lfloor n/2 \rfloor$ to choose the bound with the smaller counting coefficient: if $k \le n/2$ then $k \le n-k+1$, so we use the first bound, while if $k > n/2$ then $n-k+1 \le k$, so we use the second:

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

However, I think the expectations of the spacings may differ, it does not guarantee which bound is numerically smaller.

Question 3: Rademacher processes: bounding the variance

a. A weak variance bound

a1. We have:

$$\mathrm{Var}\left(\sum_{i=1}^{n} \varepsilon_i t_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^{n} \varepsilon_i t_i\right)^2\right] - \left(\mathbb{E}\left[\sum_{i=1}^{n} \varepsilon_i t_i\right]\right)^2$$

Because $\mathbb{E}[\varepsilon_i] = 1.\frac{1}{2} + (-1).\frac{1}{2} = 0$, we have $\mathbb{E}[\sum_{i=1}^{n} \varepsilon_i t_i] = \sum_{i=1}^{n} t_i \mathbb{E}[\varepsilon_i] = 0$. Therefore:

$$\mathrm{Var}\left(\sum_{i=1}^{n} \varepsilon_i t_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^{n} \varepsilon_i t_i\right)^2\right]$$

Expand the square, we obtain:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\varepsilon_i t_i\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n} t_i^2\varepsilon_i^2 + 2\sum_{1\le i<j\le n} t_i t_j\,\varepsilon_i\varepsilon_j\right]$$

Using $\mathbb{E}[\varepsilon_i^2] = 1^2.\frac{1}{2} + (-1)^2.\frac{1}{2} = 1$, we have:

$$\mathbb{E}\left[\sum_{i=1}^{n} t_i^2\varepsilon_i^2\right] = \sum_{i=1}^{n} t_i^2\mathbb{E}[\varepsilon_i] = \sum_{i=1}^{n} t_i^2$$

For $i \neq j$, independence and $\mathbb{E}[\varepsilon_i] = 0$ imply $\mathbb{E}[\varepsilon_i\varepsilon_j] = \mathbb{E}[\varepsilon_i]\mathbb{E}[\varepsilon_j] = 0$, so all cross terms vanish. Therefore:

$$\boxed{\mathrm{Var}\left(\sum_{i=1}^{n}\varepsilon_i t_i\right) = \sum_{i=1}^{n} t_i^2}$$

Consequently:

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

a2. For each i, let $Z^{(i)}$ denote the value of Z obtained after replacing ε_i by $-\varepsilon_i$ while keeping all other coordinates fixed.

For any fixed $t \in T$, we have:

$$\sum_{j=1}^{n}\varepsilon_j t_j - \left(\sum_{j\neq i}\varepsilon_j t_j - \varepsilon_i t_i\right) = 2\varepsilon_i t_i$$

Therefore:

$$\left|\sum_{j=1}^{n}\varepsilon_j t_j - \left(\sum_{j\neq i}\varepsilon_j t_j - \varepsilon_i t_i\right)\right| = 2|t_i|$$

Using the inequality:

$$\left|\sup_{t\in T} f(t) - \sup_{t\in T} g(t)\right| \le \sup_{t\in T}|f(t) - g(t)| \tag{10}$$

Indeed, (10) can be proved as follows. For every $t \in T$, we have:

$$f(t) \le g(t) + |f(t) - g(t)|$$

For all $t \in T$, we have:

$$f(t) \leq g(t) + |f(t) - g(t)| \leq \sup_{t \in T} g(t) + \sup_{s \in T} |f(s) - g(s)|$$

Because the above result is true for any $f(t)$, then it is true for $\sup_{t \in T} f(t)$, which implies that:

$$\sup_{t \in T} f(t) \leq \sup_{t \in T} g(t) + \sup_{s \in T} |f(s) - g(s)|$$

Therefore:

$$\sup_{t \in T} f(t) - \sup_{t \in T} g(t) \leq \sup_{t \in T} |f(t) - g(t)|$$

By exchanging the roles of f and g, we also obtain

$$\sup_{t \in T} g(t) - \sup_{t \in T} f(t) \leq \sup_{t \in T} |f(t) - g(t)|.$$

Combining the two inequalities yields (10).

Apply 10, we obtain:

$$|Z - Z^{(i)}| = \left| \sup_{t \in T} \sum_{j=1}^{n} \varepsilon_j t_j - \sup_{t \in T} \left(\sum_{j \neq i} \varepsilon_j t_j - \varepsilon_i t_i \right) \right| \leq 2 \sup_{t \in T} |t_i|$$

Thus flipping ε_i changes Z by at most $c_i = 2 \sup_{t \in T} |t_i|$.

By the bounded differences variance bound, we obtain:

$$\mathrm{Var}(Z) \leq \frac{1}{4} \sum_{i=1}^{n} c_i^2 = \frac{1}{4} \sum_{i=1}^{n} \left(2 \sup_{t \in T} |t_i| \right)^2 = \sum_{i=1}^{n} \sup_{t \in T} t_i^2 = \sigma_\infty^2.$$

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

b. [Bonus] A sharper variance bound.

b1. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher random variables and let $\varepsilon'_1, \ldots, \varepsilon'_n$ be an independent copy. For each $i \in \{1, \ldots, n\}$ define

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

Fix a deterministic tie-breaking rule on the finite set T and let $t^* = t^*(\varepsilon) \in T$ be the (measurable) maximizer so that

$$Z = \sum_{j=1}^{n} \varepsilon_j \, t_j^*.$$

First, because $Z^{(i)}$ is a supremum, we evaluate it at t^* to obtain:

$$Z^{(i)} \geq \sum_{j=1}^{n} \varepsilon_j^{(i)} \, t_j^* = \sum_{j \neq i} \varepsilon_j t_j^* + \varepsilon_i' t_i^*$$

Subtracting from $Z = \sum_{j \neq i} \varepsilon_j t_j^* + \varepsilon_i t_i^*$, we have:

$$Z - Z^{(i)} \leq (\varepsilon_i - \varepsilon_i') \, t_i^*$$

Take the positive part, we obtain:

$$\begin{aligned} &(Z - Z^{(i)})_+ \leq \left|(\varepsilon_i - \varepsilon_i') \, t_i^*\right| = |\varepsilon_i - \varepsilon_i'| \, |t_i^*| \\ \Longrightarrow \; &(Z - Z^{(i)})_+^2 \leq (\varepsilon_i - \varepsilon_i')^2 \, (t_i^*)^2 \end{aligned} \tag{11}$$

Now, apply the tower property with respect to ε:

$$\mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2 \, (t_i^*)^2\big] = \mathbb{E}\big[\mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2 \, (t_i^*)^2 \,\big|\, \varepsilon\big]\big] \, .$$

Because $(t_i^*)^2$ depends on ε only, when conditioning on ε it is treated as a constant and can be pulled out:

$$\mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2 \, (t_i^*)^2 \,\big|\, \varepsilon\big] = (t_i^*)^2 \, \mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2 \,\big|\, \varepsilon\big] \tag{12}$$

Combining the previous two equations, we get:

$$\mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2\,(t_i^*)^2\big] = \mathbb{E}\big[(t_i^*)^2\,\mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2 \,\big|\, \varepsilon\big]\big] \tag{13}$$

Next compute the conditional expectation. Because ε_i' is independent of ε and $\mathbb{E}[\varepsilon_i'] = 0$, we have

$$\mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2 \,\big|\, \varepsilon\big] = \mathbb{E}\big[\varepsilon_i^2 + (\varepsilon_i')^2 - 2\varepsilon_i\varepsilon_i' \,\big|\, \varepsilon\big] = \varepsilon_i^2 + \mathbb{E}[(\varepsilon_i')^2] - 2\varepsilon_i\,\mathbb{E}[\varepsilon_i'] = 1 + 1 - 0 = 2$$

Plug this into (13), we have:

$$\mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2\,(t_i^*)^2\big] = 2\,\mathbb{E}[(t_i^*)^2] \tag{14}$$

Therefore: $\mathbb{E}[(Z - Z^{(i)})_+^2] \leq \mathbb{E}\big[(\varepsilon_i - \varepsilon_i')^2\,(t_i^*)^2\big] = 2\,\mathbb{E}[(t_i^*)^2]$

By the one-sided Efron–Stein inequality, we obtain:

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

Finally, because $t^*(\varepsilon) \in T$ for every of ε, we have:

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

and therefore $\boxed{\mathrm{Var}(Z) \leq 2\,\mathbb{E}[\sigma^2] = 2\sigma^2}$.

b2. Take $T = \{e_1, \ldots, e_n\}$, the standard basis of $\mathbb{R}^n$. Then $\sigma^2 = \sup_{t\in T} \|t\|_2^2 = 1$ and $\sigma_\infty^2 = \sum_{i=1}^{n} \sup_{t\in T} t_i^2 = \sum_{i=1}^{n} 1 = n$, so $\boxed{\sigma_\infty^2/\sigma^2 = n}$.

Question 4: Polynomial versus exponential moment method bounds

a. Polynomial moments are at least as good

Let $Y \geq 0$ and fix $t > 0$. We show that $M(t) \leq C(t)$.

Fix any $\lambda > 0$. Apply the power series expansion $e^{\lambda y} = \sum_{q=0}^{\infty} \frac{\lambda^q y^q}{q!}$ for $y \geq 0$, and utilize the fact that all terms are nonnegative, we exchange expectation and summation to obtain:

$$\begin{aligned}
\mathbb{E}\big[e^{\lambda Y}\big] &= \sum_{q=0}^{\infty} \frac{\lambda^q}{q!}\,\mathbb{E}\big[Y^q\big] \\
\iff e^{-\lambda t}\mathbb{E}\big[e^{\lambda Y}\big] &= \sum_{q=0}^{\infty}\Big(e^{-\lambda t}\frac{(\lambda t)^q}{q!}\Big)\,\frac{\mathbb{E}[Y^q]}{t^q} = \sum_{q=0}^{\infty} w_q(\lambda,t)\,\frac{\mathbb{E}[Y^q]}{t^q},
\end{aligned}$$

where $w_q(\lambda,t) := e^{-\lambda t}\frac{(\lambda t)^q}{q!}$.

We observe that $w_q(\lambda,t) \geq 0$ and $\sum_{q=0}^{\infty} w_q(\lambda,t) = e^{-\lambda t}\sum_{q=0}^{\infty}\frac{(\lambda t)^q}{q!} = e^{-\lambda t}e^{\lambda t} = 1$, so $(w_q(\lambda,t))_{q\geq 0}$ is Poisson(λt)) distribution. Therefore, the right-hand side is a convex combination of the values $a_q := \mathbb{E}[Y^q]/t^q$, and therefore it is at least their infimum:

$$e^{-\lambda t}\mathbb{E}\big[e^{\lambda Y}\big] = \sum_{q=0}^{\infty} w_q(\lambda,t)a_q \;\geq\; \inf_{q\in\mathbb{Z}_+} a_q = \inf_{q\in\mathbb{Z}_+}\frac{\mathbb{E}[Y^q]}{t^q} = M(t)$$

Because this holds for every $\lambda > 0$, we take $\inf_{\lambda>0}$ on $e^{-\lambda t}\mathbb{E}[e^{\lambda Y}]$ to obtain:

$$C(t) = \inf_{\lambda>0} e^{-\lambda t}\mathbb{E}\big[e^{\lambda Y}\big] \;\geq\; M(t).$$

We conclude that $\boxed{M(t) \leq C(t)}$ for every $t > 0$.

b. Chernoff is still useful

Although optimizing over polynomial moments can be numerically at least as effective, it is often difficult to compute or control $\mathbb{E}[Y^q]$ for many values of q, especially for sums, due to the rapidly increasing number of terms arising from exponentiation. In contrast, exponential moments factorize conveniently: if $Y = \sum_{i=1}^{n} X_i$ with independent summands, then $\mathbb{E}\,e^{\lambda Y} = \prod_{i=1}^{n}\mathbb{E}\,e^{\lambda X_i}$. Consequently, it becomes easier to derive clean, closed-form bounds by analyzing each $\mathbb{E}\,e^{\lambda X_i}$ separately and then optimizing over λ. This makes Chernoff bounds widely practical and modular for concentration results involving sums.

Question 5. Sub-Gaussian Characterization

a. Three sub-Gaussian properties

a1. Prove $(i) \Rightarrow (ii)$.

Assume that for all $\lambda \in \mathbb{R}$,

$$\psi_Z(\lambda) = \log \mathbb{E} e^{\lambda Z} \leq \frac{v\lambda^2}{2}$$

Upper tail. By the Chernoff bound (Laplace transform method), for any $\lambda \geq 0$ and $t \geq 0$, we have:

$$\mathbb{P}(Z \geq t) \leq \exp(-\lambda t + \psi_Z(\lambda)) \leq \exp\left(-\lambda t + \frac{v\lambda^2}{2}\right)$$

Optimizing over $\lambda \geq 0$ to find $\sup_{\lambda \geq 0}(\lambda t - \frac{v\lambda^2}{2})$, we obtain $\lambda^* = t/v$. Therefore:

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

Lower tail. We have $\mathbb{P}(Z \leq -t) = \mathbb{P}(-Z \geq t)$ and $\psi_{-Z}(\lambda) = \log \mathbb{E} e^{\lambda(-Z)} = \log \mathbb{E} e^{-\lambda Z} = \psi_Z(-\lambda)$. Applying Laplace transform method to $-Z$, we have:

$$\mathbb{P}(-Z \geq t) \leq \exp(-\lambda t + \psi_Z(-\lambda)) \leq \exp\left(-\lambda t + \frac{v\lambda^2}{2}\right)$$

Employ the same argument as above, we obtain $\mathbb{P}(-Z \geq t) \leq \exp\left(-\frac{t^2}{2v}\right)$

By the union bound, we can conclude:

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

a2. Prove $(ii) \Rightarrow (iii)$. Assume (ii): for all $t \geq 0$,

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

Fix $p \geq 1$. By the tail-integral identity (HW1 Q4(a1)), we have:

$$\mathbb{E}|Z|^p = p\int_0^\infty t^{p-1}\mathbb{P}(|Z| \geq t)\, dt \leq 2p\int_0^\infty t^{p-1}\exp\left(-\frac{t^2}{2v}\right)\, dt$$

To compute the integral, we use the change of variables $u = t^2/(2v)$ (so $t = \sqrt{2vu}$ and $dt = \sqrt{\frac{v}{2u}}\, du$).

Therefore:

$$\begin{aligned}
\mathbb{E}|Z|^p &\le 2p \int_0^\infty (2vu)^{(p-1)/2} e^{-u} \sqrt{\frac{v}{2u}}\, du \\
&= 2p\,(2v)^{(p-1)/2} \sqrt{\frac{v}{2}} \int_0^\infty u^{(p-1)/2-1/2} e^{-u}\, du \\
&= C\,p\,v^{p/2} \int_0^\infty u^{p/2-1} e^{-u}\, du = C\,p\,v^{p/2}\Gamma(p/2),
\end{aligned}$$

for a universal constant $C > 0$ (absorbing powers of 2 into C). Using the fact that $\Gamma(s) \le s^s$ for $s \ge 1$ and $p \ge 2$:

$$\begin{aligned}
\mathbb{E}|Z|^p &\le C\,p\,v^{p/2} \left(\frac{p}{2}\right)^{p/2} \\
\iff (\mathbb{E}|Z|^p)^{1/p} &\le C\,v^{1/2}\,p^{1/2}\,p^{1/p} \le C'\sqrt{vp}
\end{aligned}$$

because $p^{1/p} \le 2$ for all $p \ge 1$ (absorbed into C').

For $1 \le p \le 2$, apply monotonicity of L_p norms, we have:

$$(\mathbb{E}|Z|^p)^{1/p} \le (\mathbb{E}|Z|^2)^{1/2} \le C'\sqrt{v} \le C'\sqrt{vp}.$$

Therefore, there exists a universal constant $C' > 0$ such that for all $p \ge 1$:

$$\boxed{(\mathbb{E}|Z|^p)^{1/p} \le C'\sqrt{vp}}$$

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

$$\big(\mathbb{E}|Z|^p\big)^{1/p} \le C\sqrt{vp}$$

Fix $\lambda \in \mathbb{R}$ and use the Maclaurin series for $e^{\lambda Z}$, we obtain:

$$\mathbb{E}e^{\lambda Z} = \sum_{k=0}^{\infty} \frac{\lambda^k \mathbb{E}[Z^k]}{k!} = 1 + \sum_{k=1}^{\infty} \frac{\lambda^k \mathbb{E}[Z^k]}{k!}$$

For $k \ge 1$, we have $|\mathbb{E}[Z^k]| \le \mathbb{E}|Z|^k \le (C\sqrt{vk})^k$ by Jensen inequality and by (iii) with $p = k$.

Therefore:

$$\mathbb{E}e^{\lambda Z} \le 1 + \sum_{k=1}^{\infty} \frac{\left(|\lambda|\, C\sqrt{v}\,\sqrt{k}\right)^k}{k!} \tag{15}$$

Now apply Stirling's lower bound (valid for $k \ge 1$), we have: $k! \ge \left(\frac{k}{e}\right)^k$.

Therefore, for $k \ge 1$,

$$\frac{\left(|\lambda|\, C\sqrt{v}\,\sqrt{k}\right)^k}{k!} \le \left(\frac{e|\lambda|C\sqrt{v}\sqrt{k}}{k}\right)^k = \left(\frac{A}{\sqrt{k}}\right)^k, \qquad A := eC|\lambda|\sqrt{v}.$$

Plug this into (15), we obtain:

$$\mathbb{E}e^{\lambda Z} \le 1 + \sum_{k=1}^{\infty} \left(\frac{A}{\sqrt{k}}\right)^k \tag{16}$$

Case 1: $A < 1$**.** For all $k \ge 1$, we have $\sqrt{k} \ge 1$, which implies that:

$$\left(\frac{A}{\sqrt{k}}\right)^k \le A^k$$

Therefore:

$$\mathbb{E}e^{\lambda Z} \le= 1 + \sum_{k=1}^{\infty} A.A^{k-1} = 1 + \frac{A}{1-A} \qquad \text{(geometric sum for } A < 1\text{)}$$

If $A < 1/2$ then $1 + \frac{A}{1-A} \le 2 \le \exp(4A^2)$ (because $e^{4A^2} \ge 1$). If $1/2 \le A < 1$ then $1 + \frac{A}{1-A} \le \frac{1}{1-A} \le 2 \le \exp(4A^2)$ as well. Thus, in all cases $A < 1$, we obtain:

$$\mathbb{E}e^{\lambda Z} \le \exp(C_0 A^2) \tag{17}$$

for some absolute constant $C_0 > 0$.

Case 2: $A \ge 1$**.** Let $k_0 := \lfloor A^2 \rfloor \ge 1$ and split the sum in (16), we obtain:

$$\mathbb{E}e^{\lambda Z} \le 1 + \sum_{k=1}^{k_0} \left(\frac{A}{\sqrt{k}}\right)^k + \sum_{k=k_0+1}^{\infty} \left(\frac{A}{\sqrt{k}}\right)^k =: 1 + S_{\text{small}} + S_{\text{tail}}$$

(i) Small $k \le k_0$. Because $\sqrt{k} \ge 1$ for $k \ge 1$, we have $(A/\sqrt{k})^k \le A^k$, which implies:

$$S_{\text{small}} \le \sum_{k=1}^{k_0} A^k \le k_0 A^{k_0}$$

Because $k_0 \le A^2$ and $A^{k_0} \le A^{A^2}$, we get:

$$k_0 A^{k_0} \le A^2 \cdot A^{A^2} = \exp\big(2\log A + A^2 \log A\big) \le \exp(C_1 A^2),$$

using $\log A \le A$ for $A \ge 1$ and absorbing constants into C_1.

(ii) Tail $k \ge k_0 + 1$. Split further into the “middle” block and the geometric tail:

$$S_{\text{tail}} = \sum_{k=k_0+1}^{\lfloor 4A^2 \rfloor} \left(\frac{A}{\sqrt{k}}\right)^k + \sum_{k \ge \lceil 4A^2 \rceil} \left(\frac{A}{\sqrt{k}}\right)^k =: S_{\text{mid}} + S_{\text{geo}}$$

For $k \ge k_0 + 1 \ge A^2$, we have $A/\sqrt{k} \le 1$, which means each term is ≤ 1 and therefore:

$$S_{\text{mid}} \le \big(\lfloor 4A^2 \rfloor - k_0\big) \cdot 1 \le 4A^2$$

For $k \ge 4A^2$, we have $\sqrt{k} \ge 2A$, so $A/\sqrt{k} \le 1/2$ and thus

$$S_{\text{geo}} \le \sum_{k \ge \lceil 4A^2 \rceil} 2^{-k} = \frac{2^{-\lfloor 4A^2 \rfloor}}{\frac{1}{2}} \le 2$$

Combining, $S_{\text{tail}} \le 4A^2 + 2 \le \exp(C_2 A^2)$ for an absolute constant $C_2 > 0$ (because $e^{C_2 A^2}$ dominates any polynomial in A when $A \ge 1$).

(iii) Combine in Case 2. We have shown $S_{\text{small}} \le \exp(C_1 A^2)$ and $S_{\text{tail}} \le \exp(C_2 A^2)$, and also $1 \le \exp(C_2 A^2)$. Therefore, for $C_3 = \max(C_1, C_2)$ and $C_4 A^2 \ge C_3 A^2 + \log 3$, we have:

$$\mathbb{E}e^{\lambda Z} \le 1 + S_{\text{small}} + S_{\text{tail}} \le 3\exp(C_3 A^2) \le \exp(C_4 A^2)$$

From (17) and Case 2, we can conclude that there exists an absolute constant $K > 0$ such that for all $\lambda \in \mathbb{R}$:

$$\mathbb{E}e^{\lambda Z} \leq \exp\bigl(KA^2\bigr) = \exp\bigl(Ke^2C^2v\lambda^2\bigr)$$

Take logs at both sides of the above result, we obtain:

$$\boxed{\psi_Z(\lambda) = \log \mathbb{E}e^{\lambda Z} \leq Ke^2C^2v\lambda^2 = \frac{(cv)\lambda^2}{2}, \qquad c := 2Ke^2C^2}$$

which is (i).

b. **Variance proxy**. Assume $\psi_Z(\lambda) = \log \mathbb{E}e^{\lambda Z} \leq \frac{\nu\lambda^2}{2}$, $\forall \lambda \in \mathbb{R}$ and $\mathbb{E}Z = 0$. From the assumption,

$$\mathbb{E}e^{\lambda Z} \leq \exp\left(\frac{\nu\lambda^2}{2}\right) < \infty \quad \forall \lambda \in \mathbb{R},$$

so the moment generating function is finite in a neighborhood of 0, hence it is twice differentiable around 0 and differentiation under the expectation is justified.

$$\begin{aligned}
&\psi_Z(\lambda) = \log \mathbb{E}e^{\lambda Z} \\
\implies &\psi_Z'(\lambda) = \frac{\frac{d}{d\lambda}\mathbb{E}e^{\lambda Z}}{\mathbb{E}e^{\lambda Z}} = \frac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}e^{\lambda Z}} \\
\implies &\psi_Z''(\lambda) = \frac{\mathbb{E}[Z^2e^{\lambda Z}]}{\mathbb{E}e^{\lambda Z}} - \left(\frac{\mathbb{E}[Ze^{\lambda Z}]}{\mathbb{E}e^{\lambda Z}}\right)^2
\end{aligned}$$

Because $\psi_Z(\lambda) \leq \frac{\nu\lambda^2}{2}$ for all λ, we consider λ in a neighborhood of 0 (when $\lambda \to 0$).

First, we have: $\psi_Z''(0) = \mathbb{E}Z^2 - (\mathbb{E}Z)^2 = \mathrm{Var}(Z)$.

Second, because $\psi_Z(0) = 0$, $\psi_Z'(0) = 0$, we write the Taylor expansion at 0 of $\psi_Z(\lambda)$ to obtain: $\psi_Z(\lambda) = \frac{\psi_Z''(0)}{2}\lambda^2 + o(\lambda^2)$.

Therefore, we have:

$$\begin{aligned}
&\frac{\psi_Z''(0)}{2}\lambda^2 + o(\lambda^2) \leq \frac{\nu}{2}\lambda^2 \\
\implies &\frac{\psi_Z''(0)}{2} + \frac{o(\lambda^2)}{\lambda^2} \leq \frac{\nu}{2} \qquad \text{(because } \lambda^2 > 0 \text{ when } \lambda \neq 0)
\end{aligned}$$

We have $\lim_{\lambda\to 0}\frac{o(\lambda^2)}{\lambda^2} = 0$. Therefore, we obtain:

$$\frac{\psi_Z''(0)}{2} \le \frac{\nu}{2} \Longrightarrow \psi_Z''(0) \le \nu$$

Therefore $\boxed{\mathrm{Var}(Z) \le \nu}$.

c. [Bonus] Exponential-square integrability.

Prove (ii) $\Rightarrow \mathbb{E}\exp(Z^2/(cv)) \le 2$ *for a universal* $c > 0$. Assume the tail bound (ii): for all $t \ge 0$:

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

We now prove the tail-integral identity:

$$\begin{aligned}
\mathbb{E}[f(X)] &= \int f(X(\omega))\, d\mathbb{P}(\omega) \qquad \text{(definition of expectation)}\\
&= \int f(x)\, dP_X(x) \qquad \text{(rewrite using the distribution of } X)\\
&= \int \Big(f(0) + \int_0^x f'(t)\, dt\Big)\, dP_X(x) \qquad \text{(fundamental theorem of calculus)}\\
&= f(0) + \int \left(\int_0^x f'(t)\, dt\right) dP_X(x) \qquad \text{(separate the constant term)}\\
&= f(0) + \int_0^\infty f'(t) \left(\int \mathbf{1}_{\{t\le x\}}\, dP_X(x)\right) dt \qquad \text{(Tonelli's theorem)}\\
&= f(0) + \int_0^\infty f'(t)\, P_X([t,\infty))\, dt\\
&= f(0) + \int_0^\infty f'(t)\, \mathbb{P}(X \ge t)\, dt.
\end{aligned}$$

Let $f(t) := \exp\big(t^2/(cv)\big)$ for $t \ge 0$. Because f is increasing and absolutely continuous, the tail-integral identity, we obtain:

$$\mathbb{E}f(|Z|) = f(0) + \int_0^\infty f'(t)\, \mathbb{P}(|Z| \ge t)\, dt = 1 + \int_0^\infty \frac{2t}{cv}\exp\left(\frac{t^2}{cv}\right)\mathbb{P}(|Z| \ge t)\, dt.$$

Plugging (ii), we have:

$$\mathbb{E}\exp\left(\frac{Z^2}{cv}\right) \le 1 + \int_0^\infty \frac{2t}{cv}\exp\left(\frac{t^2}{cv}\right)\cdot 2\exp\left(-\frac{t^2}{2v}\right)\,dt$$
$$= 1 + \frac{4}{cv}\int_0^\infty t\exp\left(-\frac{t^2}{v}\left(\frac{1}{2}-\frac{1}{c}\right)\right)\,dt.$$

Choose any $c > 2$. Using $\int_0^\infty te^{-at^2}\,dt = \frac{1}{2a}$ with $a = \frac{1}{v}\left(\frac{1}{2}-\frac{1}{c}\right)$, we obtain

$$\mathbb{E}\exp\left(\frac{Z^2}{cv}\right) \le 1 + \frac{4}{c}\cdot\frac{1}{2(\frac{1}{2}-\frac{1}{c})} = 1 + \frac{4}{c-2}.$$

In particular, taking $c = 6$ yields $\mathbb{E}\exp(Z^2/(6v)) \le 2$.

Conversely, prove that if $\mathbb{E}\exp(Z^2/(cv)) \le 2$ then Z is sub-Gaussian in the sense of (i') (up to constants).

Assume that for some universal constant $c > 0$, $\mathbb{E}\exp\left(\frac{Z^2}{c\nu}\right) \le 2$.

Fix $\lambda \in \mathbb{R}$ and let $a > 0$. By Young's inequality, we have:

$$\lambda Z \le |\lambda Z| \le \frac{a\lambda^2}{2} + \frac{Z^2}{2a}$$
$$\implies \mathbb{E}e^{\lambda Z} \le \exp\left(\frac{a\lambda^2}{2}\right)\,\mathbb{E}\exp\left(\frac{Z^2}{2a}\right)$$

Choose $a = \frac{c\nu}{2}$, so that $\frac{Z^2}{2a} = \frac{Z^2}{c\nu}$. Then:

$$\mathbb{E}e^{\lambda Z} \le \exp\left(\frac{c\nu}{4}\lambda^2\right)\,\mathbb{E}\exp\left(\frac{Z^2}{c\nu}\right) \le 2\exp\left(\frac{c\nu}{4}\lambda^2\right)$$
$$\implies \psi_Z(\lambda) = \log\mathbb{E}e^{\lambda Z} \le \log 2 + \frac{c\nu}{4}\lambda^2$$

Because $\psi_Z(0) = 0$ and ψ_Z is convex, for $|\lambda| \le 1$, we have:

$$\psi_Z(\lambda) \le |\lambda|\,\psi_Z(1) \le |\lambda|\left(\log 2 + \frac{c\nu}{4}\right) \le \left(\log 2 + \frac{c\nu}{4}\right)\lambda^2$$

using $|\lambda| \le \lambda^2$ for $|\lambda| \le 1$. For $|\lambda| \ge 1$ we simply use

$$\log 2 \le (\log 2)\lambda^2$$

so that for all $\lambda \in \mathbb{R}$, we obtain:

$$\psi_Z(\lambda) \leq \left(\log 2 + \frac{c\nu}{4}\right) \lambda^2 = \frac{\tilde{c}\,\nu\,\lambda^2}{2}$$

where $\tilde{c} := \frac{2\log 2}{\nu} + \frac{c}{2}$ is a universal constant, which implies that there exists a universal constant $c' > 0$ such that:

$$\boxed{\psi_Z(\lambda) \leq \frac{(c'\nu)\lambda^2}{2} \qquad \forall \lambda \in \mathbb{R}}$$

which is exactly the sub-Gaussian CGF bound up to universal constants.