1. Rademacher Random Variables and Symmetrization

(a) Let $X$ be a Rademacher random variable, i.e., $P(X = \pm 1) = 1/2$. Show that

$$E[e^{\lambda X}] \leq e^{\lambda^2/2}.$$

**Solution**

$$E[e^{\lambda X}] = \frac{1}{2} [e^\lambda + e^{-\lambda}] = \frac{1}{2} \left[ \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} \right] = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \leq 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{2^k k!} = 1 + \sum_{k=1}^{\infty} \frac{(\lambda^2/2)^k}{k!} = e^{\lambda^2/2}.$$

(b) Let $X$ be a zero mean random variable supported on the interval $[a, b]$.

i. Assume that $X'$ is an independent copy of $X$, i.e., $X, X'$ are i.i.d. random variables. Using Jensen’s inequality, show that

$$E[e^{\lambda X}] \leq E_{X,X'}[e^{\lambda (X - X')}],$$

where $E_{X,X'}[\cdot]$ denotes expectation with respect to the joint distribution of $X, X'$.

ii. Assume that $\epsilon$ is a Rademacher random variable, independent of $X, X'$. Observe that the random variables $X - X'$ and $\epsilon (X - X')$ have the same distribution, which is symmetric. Use the tower property of conditional expectation and the inequality in 1a to show that

$$E[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/2}.$$

**Solution**

(i)

$$E_X[e^{\lambda X}] = E_X[e^{\lambda (X - E_{X'}[X'])}] \leq E_{X,X'}[e^{\lambda (X - X')}],$$

where $E[X'] = 0$ and Jensen’s inequality have been used.

(ii)

$$E_{X,X'}[e^{\lambda (X - X')}] = E_{X,X',\epsilon}[e^{\lambda \epsilon (X - X')}] = E_{X,X'}[E_{\epsilon}[e^{\lambda \epsilon (X - X')}]] \leq E_{X,X'}[e^{\lambda^2 (X - X')^2/2}].$$
Since \( X \in [a, b] \), we have that \( |X - X'| \leq (b-a) \) and therefore, \( E_{X,X'} \left[ e^{\frac{X^2 - X'^2}{2}} \right] \leq E_{X,X'} \left[ e^{\frac{(b-a)^2}{2}} \right] = e^{\frac{(b-a)^2}{2}}. \) Hence,

\[
E \left[ e^{\lambda X} \right] \leq e^{\frac{\lambda^2(b-a)^2}{2}}.
\]

2. Concentration of \( \chi^2 \)- Random Variables and Random Projections

A chi-squared random variable \( S \) with \( n \) degrees of freedom is a random variable of the form:

\[
S = \sum_{i=1}^{n} Z_i^2, \quad Z_i \sim \mathcal{N}(0, 1), \quad i = 1, 2, \ldots, n \quad \text{independent}
\]

and is denoted by \( S \sim \chi^2_n \). For this random variable, \( E[S] = n \), i.e., the mean value coincides with the degrees of freedom. Moreover, the following inequality holds:

\[
P \left( \left| \frac{1}{n} \sum_{i=1}^{n} Z_i^2 - 1 \right| > t \right) \leq 2e^{-nt^2/8}, \quad \forall t \in (0, 1).
\]

This is a concentration inequality for chi-squared random variables, because it demonstrates that as the number of degrees of freedom increases, the random variable \( S/n \) cannot be very far away from its mean value 1. This inequality plays an important role in analyzing random projections.

Suppose that we are given \( m \) data points \( \{x_1, x_2, \ldots, x_m\} \) in \( \mathbb{R}^d \). If \( d \) is large, it may be too expensive to store these vectors. This motivates the introduction of a mapping \( F : \mathbb{R}^d \rightarrow \mathbb{R}^n \), which preserves the “essential” information in the data points and allows for the storage of the projected vectors \( \{F(x_1), \ldots, F(x_m)\} \) instead of the initial set of vectors. Preserving the essential information of the data points corresponds, e.g., to the requirement that \( F \) satisfies

\[
(1-\delta) \|x_i - x_j\|_2^2 \leq \|F(x_i) - F(x_j)\|_2^2 \leq (1+\delta) \|x_i - x_j\|_2^2, \quad \forall i, j \in \{1, 2, \ldots, m\}
\]

for some \( \delta \in (0, 1) \). Here, \( \| \cdot \|_2 \) corresponds to the Euclidean norm either in \( \mathbb{R}^d \) or in \( \mathbb{R}^n \). Such a requirement can always be achieved if \( n \) is large enough, but the goal is to guarantee these inequalities for a “small” \( n \) relative to \( d \).

Define \( F : x \rightarrow Zx/\sqrt{n} \), where \( Z \in \mathbb{R}^{n \times d} \) is a matrix containing i.i.d. \( \mathcal{N}(0, 1) \) random variables. Verify that such an \( F \) satisfies (2) with high probability by proving the following steps:

(a) Let \( Z_i \in \mathbb{R}^d \) denote the \( i \)th row of \( Z \). Argue that \( \tilde{Z}_i = Z_i x/\|x\|_2 \) is a \( \mathcal{N}(0, 1) \) random variable.
Solution

For a fixed $x \neq 0$, we have that $\tilde{Z}_i = \sum_{j=1}^{d} Z_{ij} x_j / \|x\|_2$ ($Z = [Z_{ij}]$ for the matrix $Z$). Therefore, $\tilde{Z}_i$ is a linear combination of i.i.d. $\mathcal{N}(0,1)$ random variables and thus, $\tilde{Z}_i$ is also normal. Moreover, denoting the coefficients of this linear transformation as $a_i = \frac{x_i}{\|x\|_2}$, we observe that

$$E[\tilde{Z}_i] = \sum_{j=1}^{d} E[Z_{ij}] \frac{x_j}{\|x\|_2} = 0,$$

$$\text{Var}(\tilde{Z}_i) = \sum_{j=1}^{d} \text{Var}(Z_{ij}) \frac{x_j^2}{\|x\|_2^2} = \sum_{j=1}^{d} \frac{x_j^2}{\|x\|_2^2} = 1,$$

where in $\text{Var}(\tilde{Z}_i)$ the independence of $Z_{ij}$ has been used. Thus, $\tilde{Z}_i \sim \mathcal{N}(0,1)$.

(b) Conclude that $S = \sum_{i=1}^{n} \tilde{Z}_i^2$ is a chi-squared random variable with $n$ degrees of freedom.

Solution

$S = \sum_{i=1}^{n} \tilde{Z}_i^2$ is clearly a chi-squared random variable by the preliminary definitions in this problem. We only need to verify that $\tilde{Z}_1, \ldots, \tilde{Z}_n$ are independent. This is straightforward due to the fact that in each $\tilde{Z}_i$, a different set of $Z_{ij}$ participates.

(c) Use (1) to show that

$$P\left( \frac{\|F(x)\|_2^2}{\|x\|_2^2} \notin [(1 - \delta), (1 + \delta)] \right) \leq 2e^{-n\delta^2/n}, \text{ for any } 0 \neq x \in \mathbb{R}^d.$$
(d) Use the union bound to conclude that all the inequalities in (2) are satisfied with probability at least $1 - \epsilon$ for any $\epsilon \in (0, 1)$ if $n > \frac{16}{\delta^2} \log(m) + \frac{8}{\delta^2} \log \left( \frac{1}{\epsilon} \right)$.

Solution

Note that there are at most $\binom{m}{2}$ different pairs of data points $(x_i, x_j)$. Thus, by the union bound

$$P \left( \frac{\|F(x_i - x_j)\|^2}{\|x_i - x_j\|^2} \notin [(1 - \delta), (1 + \delta)] \right) \leq 2 \binom{m}{2} e^{-\frac{n\delta^2}{8}}.$$  

Requiring

$$2 \binom{m}{2} e^{-\frac{n\delta^2}{8}} \leq m^2 e^{-\frac{n\delta^2}{8}} \leq \epsilon$$

yields that $n > \frac{16}{\delta^2} \log(m) + \frac{8}{\delta^2} \log \left( \frac{1}{\epsilon} \right)$. I.e., $n > \frac{16}{\delta^2} \log(m) + \frac{8}{\delta^2} \log \left( \frac{1}{\epsilon} \right)$ is sufficient for all the inequalities given by (2) to hold with probability at least $1 - \epsilon$.

3. Convergence in Probability

(a) Let $X_1, X_2, \ldots$ be a sequence of random variables such that $E[X_n] \to \mu$ and $\text{Var}(X_n) \to 0$ as $n \to \infty$. Show that $X_n \to \mu$ as $n \to \infty$ in probability.

Solution

Consider the deterministic sequence $\{a_n = E[X_n]\}$. Then, for any $\epsilon > 0$, there exists $N_\epsilon > 0$ such that $|a_n - \mu| = |E[X_n] - \mu| < \frac{\epsilon}{2}, \forall n > N_\epsilon$. Therefore, $\forall n > N_\epsilon$:

$$P(|X_n - \mu| > \epsilon) = P(|X_n - E[X_n] + E[X_n] - \mu| > \epsilon)$$

$$\leq P(|X_n - E[X_n]| + |E[X_n] - \mu| > \epsilon)$$

$$= P \left( |X_n - E[X_n]| > \frac{\epsilon}{2} \right) \leq \frac{\text{Var}(X_n)}{(\epsilon^2/4)} \to 0 \text{ as } n \to \infty.$$  

Here, we use Chebyshev’s inequality in the last step. Hence, $\forall \epsilon > 0$, we have $\lim_{n \to \infty} P(|X_n - \mu| > \epsilon) = 0$, i.e., $X_n \overset{p}{\to} \mu$.

(b) Suppose that we distribute $n$ balls into $n$ boxes independently at random. Let $E_n = \mathbb{I}_1 + \mathbb{I}_2 + \cdots + \mathbb{I}_n$ be the number of empty boxes, where $\mathbb{I}_j$ is the emptiness indicator of the $j$th box. Using part 3a show that $\frac{1}{n} E_n \overset{p}{\to} 1/e$ as $n \to \infty$.

Solution

Note that $\mathbb{I}_j \sim \text{Ber} \left( (1 - \frac{1}{n})^n \right)$ random variables, because the probability of not occupying the $j$th box in every drop is $\left( 1 - \frac{1}{n} \right)$. This leads to $E[\mathbb{I}_j] = (1 - \frac{1}{n})^n$.
and \( E[E_n] = n \left( 1 - \frac{1}{n} \right)^n \). Moreover,

\[
E \left[ E_n^2 \right] = \sum_{j=1}^{n} E \left[ I_j^2 \right] + \sum_{i \neq j} E \left[ I_i I_j \right] \\
= \sum_{j=1}^{n} E \left[ I_j \right] + \sum_{i \neq j} E[I_i I_j] \\
= n \left( 1 - \frac{1}{n} \right)^n + \sum_{i \neq j} E[I_i I_j].
\]

We now observe that \( E[I_i I_j] = P(\text{boxes } i \text{ and } j \text{ are both empty}) = \left( 1 - \frac{2}{n} \right)^n \).

Combining, we see that

\[
\text{Var}(E_n) = E \left[ E_n^2 \right] - \left( E[E_n] \right)^2 \\
= n \left( 1 - \frac{1}{n} \right)^n + n(n-1) \left( 1 - \frac{2}{n} \right)^n - n^2 \left( 1 - \frac{1}{n} \right)^{2n}.
\]

Define the random variable \( S_n = \frac{E_n}{n} \). It is easy to see that \( E[S_n] = \frac{E[E_n]}{n} \to \frac{1}{e} \) as \( n \to \infty \), while

\[
\text{Var}(S_n) = \frac{1}{n} \left( 1 - \frac{1}{n} \right)^n + \left( 1 - \frac{1}{n} \right) \left( 1 - \frac{2}{n} \right)^n - \left( 1 - \frac{1}{n} \right)^{2n} \quad \to_{n \to \infty} 0 + e^{-2} - e^{-2} = 0.
\]

Therefore, by the previous part, \( S_n = \frac{E_n}{n} \overset{P}{\to} \frac{1}{e} \).

4. All Types of Convergence of Random Sequences

Consider a sequence of random variables \( X_1, X_2, \ldots \) with its generic term defined as follows:

\[
X_1 \sim \text{Ber}(1/2) \quad \text{and} \quad X_n = (X_{n-1} + 1) \mod 2.
\]

Does this sequence converge almost surely, in probability, in mean square and in distribution? Justify your answer for each type of convergence separately.

**Solution**

The sequence has only two outcomes (or sample paths) depending on \( X_1 \):

\[
X_1 = 1 : \text{Then } X_1 X_2 \ldots = 10101010 \ldots \\
X_1 = 0 : \text{Then } X_1 X_2 \ldots = 01010101 \ldots
\]

(a) \( \{X_n\} \) does not converge almost surely because the probability of every jump is \( \frac{1}{2} \).

(b) \( \{X_n\} \) does not converge in probability because the frequency of jumps is equal to \( \frac{1}{2} \).
(c) \( \{X_n\} \) does not converge to \( \frac{1}{2} \) (mean value of \( X_n \) for all \( n \)) in mean square sense since
\[
\lim_{n \to \infty} E \left[ \left( X_n - \frac{1}{2} \right)^2 \right] = E \left[ X_n^2 - X_n + \frac{1}{4} \right] = E[X_n^2] - E[X_n] + \frac{1}{4} = \frac{1}{2}
\]

(d) As \( n \to \infty \), the value of the \( X_n \) is determined by \( X_1 \) and therefore, \( X_n \sim \text{Ber} \left( \frac{1}{2} \right) \), \( \forall n \geq 1 \) (due to \( X_1 \)), hence \( X_n \overset{d}{\to} \text{Ber} \left( \frac{1}{2} \right) \).

5. Entropy Bounds

For a positive random variable \( X \), the following definition of entropy is introduced:
\[
H(X) = E[X \log X] - E[X] \log E[X].
\]
Consider the random variable \( e^{\lambda X} \) and the moment generating function \( m_X(\lambda) = E \left[ e^{\lambda X} \right] \). It is then easy to see that
\[
H \left( e^{\lambda X} \right) = \lambda m_X'(\lambda) - m_X(\lambda) \log m_X(\lambda).
\]
Assume that there is a constant \( \sigma^2 < \infty \) such that
\[
H \left( e^{\lambda X} \right) \leq \frac{\lambda^2 \sigma^2}{2} m_X(\lambda), \quad \forall \lambda \in \mathbb{R}_+.
\]
Show that
\[
E \left[ e^{\lambda(X - E[X])} \right] \leq e^{\frac{\lambda^2 \sigma^2}{2}}, \quad \forall \lambda \in \mathbb{R}_+.
\]

Solution

We need to show that \( \log m_X(\lambda) \leq \frac{\lambda^2 \sigma^2}{2} \). Note that \( H \left( e^{\lambda X} \right) \leq \frac{\lambda^2 \sigma^2}{2} m_X(\lambda) \) is equivalent to \( \lambda m_X'(\lambda) - m_X(\lambda) \log m_X(\lambda) \leq \frac{\lambda^2 \sigma^2}{2} m_X(\lambda) \) or \( \frac{m_X'(\lambda)}{\lambda m_X(\lambda)} - \frac{1}{\lambda^2} \log m_X(\lambda) \leq \frac{\sigma^2}{2} \), where we have divided by \( \lambda^2 m_X(\lambda) \). We now see that
\[
\frac{d}{d\lambda} \log m_X(\lambda) = \frac{m_X'(\lambda)}{\lambda m_X(\lambda)} - \frac{m_X(\lambda)}{\lambda^2}.
\]
Also,
\[
\lim_{\lambda \to 0} \frac{\log m_X(\lambda)}{\lambda} = \frac{m_X'(0)}{m_X(0)} = E[X].
\]
Therefore, for an arbitrary \( \lambda > 0 \),
\[
\int_0^\lambda \left[ \frac{d}{d\lambda} \log m_X(\lambda) \right] d\lambda = \frac{\log m_X(\lambda)}{\lambda} - E[X] \leq \int_0^\lambda \frac{\sigma^2}{2} d\lambda = \frac{\sigma^2 \lambda}{2}.
\]
Multiplying both sides by $\lambda$, we obtain
\[
\log E \left[ e^{\lambda(X-E[X])} \right] = \log E \left[ e^{\lambda X} \right] = \log E \left[ e^{\lambda X} \right] - \lambda E[X] \leq \frac{\sigma^2 \lambda^2}{2}.
\]

6. **Rademacher Complexity**

Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $X_i \in [a, b]$ for all $i \in \{1, 2, \ldots, n\}$. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, which is convex in each of its arguments and $L$-Lipschitz with respect to the Euclidean norm $\| \cdot \|_2$. Then the following inequality holds:
\[
E \left[ e^{\lambda(f(X_1, X_2, \ldots, X_n) - E[f(X_1, X_2, \ldots, X_n)])} \right] \leq e^{\lambda^2 L^2 (b-a)^2}, \ \forall \lambda \geq 0. \tag{3}
\]

Consider a set of vectors $\mathcal{C} \subset \mathbb{R}^n$. Then,
\[
\mathcal{R}(\mathcal{C}) = E \left[ \sup_{c \in \mathcal{C}} \sum_{i=1}^{n} \varepsilon_i c_i \right]
\]
corresponds to the **Rademacher complexity** of $\mathcal{C}$. Here, $\{\varepsilon_i\}$ are independent Rademacher random variables as these were defined in Problem 1 and $c = [c_1, c_2, \ldots, c_n]$ is a vector in $\mathcal{C}$. Let $\hat{R}(\mathcal{C}) = \sup_{c \in \mathcal{C}} \sum_{i=1}^{n} \varepsilon_i c_i$ denote the **empirical** version of $\mathcal{R}(\mathcal{C})$. Using (3), argue that
\[
P(\hat{R}(\mathcal{C}) \geq R(\mathcal{C}) + t) \leq e^{-\frac{t^2}{16(\text{diam}(\mathcal{C}))^2}},
\]
where $\text{diam}(\mathcal{C}) = \sup_{c \in \mathcal{C}} \|c\|_2$.

**Note**: A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be $L$-Lipschitz for some $L \geq 0$ with respect to the Euclidean norm $\| \cdot \|_2$ if
\[
|f(x) - f(y)| \leq L \|x - y\|_2, \ \forall x, y \in \mathbb{R}^n.
\]

Also, the pointwise supremum $f(x) = \sup_{i \in \mathcal{I}} f_i(x)$ of an arbitrary family of convex functions $\{f_i(x)\}_{i \in \mathcal{I}}$ is convex.

**Hint**: First, show that $\hat{L} = \text{diam}(\mathcal{C})$ is a valid upper bound to the Lipschitz constant of $f(\varepsilon_1, \ldots, \varepsilon_n) = \sup_{c \in \mathcal{C}} \sum_{i=1}^{n} \varepsilon_i c_i$ for $\varepsilon_i \in \{-1, +1\}$. Combine then (3) with the Chernoff bound to get the desired inequality by choosing $\lambda = \frac{t}{8(\text{diam}(\mathcal{C}))^2}$.

**Solution**

We note that the function $f(\varepsilon_1, \ldots, \varepsilon_n) = \sup_{c \in \mathcal{C}} \sum_{i=1}^{n} \varepsilon_i c_i$ is convex, since it is the supremum of a family of linear (i.e., convex) functions. Moreover, $f(\varepsilon_1, \ldots, \varepsilon_n)$ is
Lipschitz with $\bar{L} = \sup_{c \in C} \|c\|_2$ being a valid upper bound to the Lipschitz constant. To see this, consider first the linear form $g(\varepsilon; c) = \sum_{i=1}^n c_i \varepsilon_i$ for $c \in C$. Then,

$$|g(\varepsilon; c) - g(\varepsilon'; c)| = \left| \sum_{i=1}^n c_i (\varepsilon_i - \varepsilon'_i) \right| \leq \|c\|_2 \|\varepsilon - \varepsilon'\|_2 \leq \sup_{c \in C} \|c\|_2 \|\varepsilon - \varepsilon'\|_2.$$ 

In (a), the Cauchy-Schwarz inequality has been used. Therefore, $g(\varepsilon; c)$ is Lipschitz and $\bar{L}$ is a valid upper bound to the corresponding Lipschitz constant. For arbitrary $\varepsilon, \varepsilon'$, we can now see that

$$g(\varepsilon; c) \leq g(\varepsilon'; c) + \bar{L} \|\varepsilon - \varepsilon'\|_2.$$ 

Taking the supremum with respect to $c \in C$, first to the right hand side and then to the left, we obtain:

$$f(\varepsilon_1, \ldots, \varepsilon_n) - f(\varepsilon'_1, \ldots, \varepsilon'_n) \leq \bar{L} \|\varepsilon - \varepsilon'\|_2.$$ 

Moreover, interchanging $\varepsilon, \varepsilon'$, we obtain:

$$f(\varepsilon'_1, \ldots, \varepsilon'_n) - f(\varepsilon_1, \ldots, \varepsilon_n) \leq \bar{L} \|\varepsilon - \varepsilon'\|_2.$$ 

Combining, we have:

$$|f(\varepsilon_1, \ldots, \varepsilon_n) - f(\varepsilon'_1, \ldots, \varepsilon'_n)| \leq \bar{L} \|\varepsilon - \varepsilon'\|_2.$$ 

Therefore, $\bar{L}$ is an upper bound to the Lipschitz constant of $f$. Employing now the Chernoff bound and (3), we obtain:

$$P(\hat{R}(C) \geq R(C) + t) \leq \frac{E\left[e^{\lambda \hat{R}(C)}\right]}{e^{\lambda \hat{R}(C) + t}} = \frac{E\left[e^{\lambda(\hat{R}(C) - R(C))}\right]}{e^{\lambda t}} \leq e^{4\lambda^2 \bar{L}^2 - \lambda t}.$$ 

Minimizing $4\lambda^2 \bar{L}^2 - \lambda t$ with respect to $\lambda$ yields that

$$\lambda_* = \frac{t}{8\bar{L}^2} = \frac{t}{8(diam(C))^2}.$$ 

Plugging this value in the last inequality yields the desired result.

7. Convergence in Probability Again

Prove that $X_n \xrightarrow{p} X$ if and only if

$$\lim_{n \to \infty} E\left[\frac{|X_n - X|}{1 + |X_n - X|}\right] = 0.$$
Solution

Without loss of generality, take \( X = 0 \). We want to show that \( X_n \xrightarrow{p} 0 \) if and only if \( \lim_{n \to \infty} E \left[ \frac{|X_n|}{1 + |X_n|} \right] = 0 \).

(i) \( X_n \xrightarrow{p} 0 \implies \lim_{n \to \infty} E \left[ \frac{|X_n|}{1 + |X_n|} \right] = 0 \).

By \( X_n \xrightarrow{p} 0 \), we have that \( \forall \epsilon > 0 : \lim_{n \to \infty} P(|X_n| > \epsilon) = 0 \). Note that

\[
\frac{|X_n|}{1 + |X_n|} \leq \frac{|X_n|}{1 + |X_n|} \mathbb{I}(|X_n| > \epsilon) + \epsilon \mathbb{I}(|X_n| \leq \epsilon) \leq |X_n| \mathbb{I}(|X_n| > \epsilon) + \epsilon.
\]

Therefore,

\[
E \left[ \frac{|X_n|}{1 + |X_n|} \right] \leq E[|X_n| \mathbb{I}(|X_n| > \epsilon)] + \epsilon = P(|X_n| > \epsilon) + \epsilon.
\]

Taking the limit, we obtain \( \lim_{n \to \infty} E \left[ \frac{|X_n|}{1 + |X_n|} \right] \leq \epsilon \), and since \( \epsilon > 0 \) is arbitrary, we have that \( \lim_{n \to \infty} E \left[ \frac{|X_n|}{1 + |X_n|} \right] = 0 \).

(ii) \( \lim_{n \to \infty} E \left[ \frac{|X_n|}{1 + |X_n|} \right] = 0 \implies X_n \xrightarrow{p} 0 \).

Observe that the function \( f(x) = \frac{x}{x+1} \) is increasing. Therefore,

\[
\frac{\epsilon}{1 + \epsilon} \mathbb{I}(|X_n| > \epsilon) \leq \frac{|X_n|}{1 + |X_n|} \mathbb{I}(|X_n| > \epsilon) \leq \frac{|X_n|}{1 + |X_n|}.
\]

Taking expectations and then limits to both sides, we obtain:

\[
\frac{\epsilon}{1 + \epsilon} \lim_{n \to \infty} P(|X_n| > \epsilon) \leq \lim_{n \to \infty} E \left[ \frac{|X_n|}{|X_n| + 1} \right] = 0
\]

Since this holds for any \( \epsilon > 0 \), we have that \( \lim_{n \to \infty} P(|X_n| > \epsilon) = 0 \), \( \forall \epsilon > 0 \). therefore, \( X_n \xrightarrow{p} 0 \).

8. Almost Sure Convergence Again [Bonus Problem]

Let \( X_1, X_2, \ldots \) be a sequence of i.i.d. random variables with \( P(X_n > x) = e^{-x}, x \geq 0 \). Show that

\[
\frac{\max\{X_1, X_2, \ldots, X_n\}}{\log n} \overset{a.s.}{\to} 1 \text{ as } n \to \infty.
\]

Note: Use the Borel-Cantelli Lemma.
Solution

Let \( Y_n = \max\{X_1, \ldots, X_n\} \). For \( \epsilon > 0 \), we have

\[
P \left( \frac{\max(X_1, \ldots, X_n)}{\log n} < 1 - \epsilon \right) = P \left( \frac{Y_n}{\log n} < 1 - \epsilon \right)
\]
\[
= P \left( Y_n < (1 - \epsilon) \log n \right) = P \left( \bigcap_{i=1}^{n} \left\{ X_i < (1 - \epsilon) \log n \right\} \right)
\]
\[
= \left( 1 - e^{-\left(1-\epsilon\right) \log n} \right)^n
\]
\[
= \left( 1 - \frac{1}{n^{1-\epsilon}} \right)^n = \left( \left( 1 - \frac{1}{n^{1-\epsilon}} \right)^{n^{1-\epsilon}} \right)^{n^\epsilon}
\]

Using part (a) in the Borel-Cantelli lemma,

\[
\sum_{n=1}^{\infty} P \left( \frac{Y_n}{\log n} < 1 - \epsilon \right) = \sum_{n=1}^{\infty} \left[ \left( 1 - \frac{1}{n^{1-\epsilon}} \right)^{n^{1-\epsilon}} \right]^{n^\epsilon}
\]
\[
\leq \sum_{n=1}^{\infty} \left( e^{-1} \right)^{n^\epsilon} < \infty.
\]

Therefore, \( P \left( \frac{Y_n}{\log n} < 1 - \epsilon \text{ i.o.} \right) = 0 \). Using similar steps and focusing on the random variable \( X_n \), we have

\[
P \left( \frac{X_n}{\log n} > 1 + \epsilon \right) = e^{-\log n(1+\epsilon)} = \frac{1}{n^{1+\epsilon}}.
\]

Therefore,

\[
\sum_{n=1}^{\infty} P \left( \frac{X_n}{\log n} > 1 + \epsilon \right) = \sum_{n=1}^{\infty} \frac{1}{n^{1+\epsilon}} < \infty.
\]

By part (a) in the Borel-Cantelli lemma, \( P \left( \frac{X_n}{\log n} > 1 + \epsilon \text{ i.o.} \right) = 0 \). Also, \( \sum_{n=1}^{\infty} P \left( \frac{X_n}{\log n} > 1 - \epsilon \right) = \sum_{n=1}^{\infty} \frac{1}{n^{1-\epsilon}} = \infty \). By part (b) in the Borel-Cantelli lemma, \( P \left( \frac{X_n}{\log n} > 1 - \epsilon \text{ i.o.} \right) = 1 \). Combining the above results,

\[
\limsup_{n} \frac{X_n}{\log n} = 1 \text{ a.s.}
\]

and therefore, \( P \left( \frac{X_n}{\log n} > 1 + \epsilon \text{ i.o.} \right) = 0 \). Hence,

\[
\frac{Y_n}{\log n} = \max_{1 \leq i \leq n} \frac{X_i}{\log n} \text{ a.s.} \to 1.
\]