Q1

Borel Contelli Lemma (I) $P(E_n \text{ i.o.}) = 0$ if $\sum_{n=1}^{\infty} P(E_n) < \infty$

Borel Contelli Lemma (II) $P(E_n \text{ i.o}) = 1$ if $\sum_{n=1}^{\infty} P(E_n) = \infty$ $\{E_n\}$ independent.

(a) when $\mathbb{P}(E_n) = \frac{1}{n}$ $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$ $\xrightarrow{BC(II)}$ $\mathbb{P}(E_n \text{ i.o.}) = 1$

$\{E_n\}$ independent

(b) $C = \frac{99}{100} < 1$

$\mathbb{P}(E_n) = \left(\frac{99}{100}\right)^n$

$\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \sum_{n=1}^{\infty} C^n = \sum_{n=0}^{\infty} C^n - 1$

$= \lim_{n\to\infty} \frac{C^{n+1} - 1}{C-1} - 1 = \lim_{n\to\infty} \frac{C(C^n - 1)}{C-1} = \frac{C}{1-C} < \infty$

By BC(I) $\mathbb{P}(E_n \text{ i.o}) = 0$

Q2.

$t_n(i) = \sum_{m=1}^{n} \mathbb{I}(X_m \in P_i)$

$\log R_n = \sum_{i=1}^{k} t_n(i) \cdot \log p_i = \sum_{i=1}^{k} \log p_i \left[\sum_{m=1}^{n} \mathbb{I}(X_m \in P_i)\right]$

$Y_m \triangleq \sum_{i=1}^{k} \log p_i \left[\mathbb{I}(X_m \in p_i)\right] \Rightarrow \log R_n = S_n = \sum_{m=1}^{n} Y_m.$

$\mathbb{E}|Y_m| = -\sum_{i=1}^{k} p_i \log p_i < \infty$

$\mathbb{E} Y_m = \sum_{i=1}^{k} p_i \log p_i$ $\{Y_m\}$ i.i.d.

By Strong Law of Large Number

$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{a.s.} 0$

$$S_n = \sum_{m=1}^{n} Y_m = \log R_n$$

$$\mathbb{E} S_n = \sum_{m=1}^{n} \mathbb{E} Y_m = -nh.$$

$$\Rightarrow \quad n^{-1} \log R_n \to -h \quad \text{as } n \to \infty.$$

Q3. (a)

$$x_i = \begin{cases} i & i^{-\alpha}/4 \\ -i & i^{-\alpha}/4 \\ 0 & 1 - i^{-\alpha}/2 \end{cases}$$

$$\mathbb{E} x_i = i \times i^{-\alpha}/4 + (-i)\, i^{-\alpha}/4 + 0 \cdot (1 - i^{-\alpha}/2) = 0$$

$$\mathbb{E} x_i^2 = i^2 \times i^{-\alpha}/4 + (i^2) \times i^{-\alpha}/4 + 0 \cdot (1 - i^{-\alpha}/2) = i^{2-\alpha}/2$$

$$Var(x_i) = \mathbb{E} x_i^2 - (\mathbb{E} x_i)^2 = \frac{1}{2} i^{2-\alpha}$$

$$Var(S_n) = \sum_{i=1}^{n} Var(x_i) = \sum_{i=1}^{n} \frac{1}{2} i^{2-\alpha} \sim \frac{n^{3-\alpha}}{2(3-\alpha)}$$

Let. $X_{n,m} = \frac{X_m}{n^{(3-\alpha)/2}}$ then. $\mathbb{E} X_{n,m} = 0$

$\sum_{m=1}^{n} \mathbb{E} X_{n,m}^2 = \sum_{m=1}^{n} \frac{m^{2-\alpha}}{2n^{3-\alpha}}$ since $\alpha \in (0,1)$ we have.

$\sum_{m=1}^{n} \mathbb{E} X_{n,m}^2 \longrightarrow \frac{1}{2(3-\alpha)} = \sigma^2 > 0$ $\quad \frac{3-\alpha}{2} > 1$

$n^{\frac{\alpha-1}{2}} < \varepsilon$. we have. $\sum_{m=1}^{n} \mathbb{E}(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) = 0$

then. $\lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{E}(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) = 0$

By Lindeberg-Feller Theorem.

$a_n(\alpha) = 0 \qquad b_n(\alpha) = \frac{n^{(3-\alpha)/2}}{\sqrt{2(3-\alpha)}}$ then.

$$\frac{S_n - a_n}{b_n} \Rightarrow N(0,1)$$

(b) $X_i = \begin{cases} 1 & \mathbb{P} = \frac{1}{i}. \\ 0 & \mathbb{P} = 1 - \frac{1}{i}. \end{cases}$

$\mathbb{E} X_i = \frac{1}{i}$ $\quad \mathbb{E} X_i^2 = \frac{1}{i}$ $\quad \mathrm{Var}(X_i) = \frac{1}{i} - \frac{1}{i^2}.$

$S_n = X_1 + X_2 + \cdots + X_n$

$\mathbb{E} S_n = \sum_{i=1}^{n} \frac{1}{i} \sim \log n,$

$\mathrm{Var}\, S_n = \sum_{i=1}^{n} \frac{1}{i} - \frac{1}{i^2} \sim \log n.$

Let $X_{n,m} = (X_m - \frac{1}{m}) / (\log n)^{1/2}.$

$\mathbb{E} X_{n,m} = 0$

$\sum_{m=1}^{n} \mathbb{E} X_{n,m}^2 = \sum_{m=1}^{n} \frac{\frac{1}{m}(1 - \frac{1}{m})}{\log n} \to 1$

$\exists \cdot n,$ s.t.

$\arg\max \left\{ \frac{|1 - \frac{1}{m}|}{(\log n)^{1/2}}, \frac{|\frac{1}{m}|}{(\log n)^{1/2}} \right\} < \varepsilon$

Then. $\sum_{m=1}^{n} \mathbb{E}(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) = 0$

$\Rightarrow \lim_{n \to \infty} \sum_{m=1}^{n} \mathbb{E}(|X_{n,m}|^2; |X_{n,m}| > \varepsilon) = 0$

By Lindeberg-Feller Theorem.

$a_n = \log n \quad b_n = (\log n)^{1/2}.$

then. $\frac{S_n - a_n}{b_n} \sim N(0,1)$

Q4 $\quad X_j = \begin{cases} j & \mathbb{P} = \frac{1}{2}\cdot j^{-\beta} \\ -j & \mathbb{P} = \frac{1}{2}\cdot j^{-\beta} \\ 0 & \mathbb{P} = 1 - j^{-\beta} \end{cases}$

(i)

$\mathbb{E}X_j = 0 \qquad \mathbb{E}|X_j| = j^{1-\beta} \qquad \mathbb{E}X_j^2 = j^2 \cdot j^{-\beta} = j^{2-\beta} \qquad Var(X_j) = j^{2-\beta}$

$A \triangleq 1$ then:

(1). $\sum_{n=1}^{\infty} Var(X_n \mathbb{I}(|X_n| \le 1)) = Var(X_1) = 1^{2-\beta}$. converges.

(2) $\sum_{n=1}^{\infty} \mathbb{E}(X_n \mathbb{I}(|X_n| \le 1)) = \mathbb{E}X_1 = 0$ converges.

(3) $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > 1) = \sum_{n=2}^{\infty} n^{-\beta}$ converges since $\beta > 1$

Thus, By Kolmogorov's three series theorem.

(1), (2), (3) $\Rightarrow \sum_n X_n < \infty$ a.s.

(ii) $\sum_{j=1}^{n} Var(X_j) = \sum_{j=1}^{n} j^{2-\beta} \sim n^{3-\beta}$

$\sum_{j=1}^{n} \mathbb{E}X_j = 0$

$\forall \varepsilon > 0$ as $\varepsilon > j \cdot n^{\frac{2}{3-\beta}}$.

Then. $\frac{1}{s_n^2} \sum_{j=1}^{n} \mathbb{E}(X_j^2 \mathbb{I}\{|X_j| > \varepsilon s_n\}) \to 0$

By Lindeberg - Feller CLT

$$\frac{S_n}{s_n} \to N(0,1) \Rightarrow \frac{S_n}{n^{(3-\beta)/2}} \to N(0,1)$$

(iii) $\mathbb{E}\exp\left(it \cdot \frac{\sum_{j=1}^{n} X_j}{n}\right) = \mathbb{E}\prod_{j=1}^{n} \exp\left(i\frac{t}{n} X_j\right) = \prod_{j=1}^{n} \mathbb{E}\exp\left(i\frac{t}{n} X_j\right)$

where.

$$\mathbb{E}e^{i\frac{t}{n}\cdot X_j} = e^{i\frac{t}{n}\cdot j}\cdot\frac{1}{2}\frac{1}{j} + e^{i\frac{t}{n}(-j)}\cdot\frac{1}{2j} + e^{i\frac{t}{n}\cdot 0}\cdot\left(1-\frac{1}{j}\right)$$

$$= e^{i\frac{t}{n}j}\cdot\frac{1}{2}\cdot\frac{1}{j} + e^{i\frac{t}{n}(-j)}\cdot\frac{1}{2}\cdot\frac{1}{j} + (1-\frac{1}{j})$$

$$= [\cos(\tfrac{t}{n}\cdot j) + i\sin(\tfrac{t}{n}j)]\cdot\frac{1}{2}\cdot\frac{1}{j}$$

$$+ [\cos(\tfrac{t}{n}(-j)) + i\sin(\tfrac{t}{n}\cdot j)]\cdot\frac{1}{2}\frac{1}{j}$$

$$+ (1-\frac{1}{j})$$

$$= \frac{1}{2j}\cdot(2\cos(\tfrac{t}{n}j)) + (1-\frac{1}{j})$$

$$= 1 + \frac{1}{j}\left(\cos(\tfrac{t}{n}j) - 1\right)$$

$$\Rightarrow \mathbb{E}\exp(itS_n/n) = \prod_{j=1}^{n}\mathbb{E}\exp(i\tfrac{t}{n}x_j)$$

$$= \prod_{j=1}^{n}\left(1+\frac{1}{n}\cdot(j/n)^{-1}[\cos(jt/n)-1]\right)$$

where $\sum_{j=1}^{n}\frac{1}{n}(j/n)^{-1}\{\cos(jt/n)-1\}$

suppose $x_i=\frac{j}{n}$ $\;=\; \frac{1}{n}\sum_{j=1}^{n} x^{-1}\cdot[\cos(xt)-1]$

then $0\le x\le 1$

$x^{-1}(1-\cos(tx))$ is bounded for $0\le x\le 1$

$\Rightarrow$ the Riemann sum

$\sum_{j=1}^{n}\frac{1}{n}(j/n)^{-1}\{\cos(jt/n)-1\}$

suppose $x_i=\frac{j}{n}$ $\;=\; \frac{1}{n}\sum_{j=1}^{n} x^{-1}\cdot[\cos(xt)-1]$

then $0\le x\le 1$

$$\to \int_0^1 x^{-1}[\cos(xt)-1]\,dx$$

$\Rightarrow$

$\mathbb{E}\exp(itS_n/n) \to \exp\{\int_0^1 x^{-1}[\cos(xt)-1]\,dx\}$ when $n\to\infty$.

RHS is continuous at 0

By Levy-Continuity Thm.

S_n/n converges in distribution to a distribution with characteristic function. $\exp(-\int_0^1 (1-\cos xt)/x\,dx)$

Q5

(a) $X_n \xrightarrow{\mathbb{P}} X \Rightarrow \forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon)$

$$|X_n - X_m| = |X_n - X + X - X_m| \le |X_n - X| + |X_m - X|$$

$$\{|X_n - X_m| > 2\varepsilon\} \subset \{|X_n - X| + |X_m - X| > 2\varepsilon\}$$

$$= \{ |X_n - X| > \delta \text{ and } |X_m - X| > \gamma \text{ where } \delta, \gamma > 0, \delta + \gamma = 2\varepsilon \}$$

$$\Rightarrow \mathbb{P}\{|X_n - X_m| > 2\varepsilon\} \subset \mathbb{P}(|X_n - X| > \delta) + \mathbb{P}(|X_m - X| > \gamma)$$

$$\to 0 \quad \text{when} \quad n.m \to \infty.$$

$\Rightarrow$ X_n is Cauchy convergent in probability

(b) Yes, it's true.

$\{X_n\}$ is a Cauchy sequence in $\mathbb{P}$

Choose a sequence $\{n_k\}$ such that

$\mathbb{P}\{|X_n - X_m| > 2^{-k}) < 2^{-k}$

for $n, m \geq n_k$

$Y_k \triangleq X_{n_k}$

$A_k \triangleq \{|Y_{k+1} - Y_k| > 2^{-k}\} \qquad \mathbb{P}(A_k) \le 2^{-k}$

$T = \sum_{k=1}^{\infty} \mathbb{I}\{A_k\} \qquad \mathbb{E}T \le 1 \qquad \mathbb{P}(T < \infty) = 1$

with probability 1, finitely many events A_k occurs.

This means that, for any w, $T(w) < \infty$

there exists a $k_0(w)$ $|Y_k(w) - Y_{k+1}(w)| < 2^{-k}$

for all $k \ge k_0(w)$, there has the inequality.

$|Y_k'(w) - Y_i'(w)| \le 2^{-k+1}$ for all $k \geq k_0(w)$.

$Y_n(\omega)$ is a numerical Cauchy sequence and hence.

$\exists \; X_1(\omega) \quad |Y_k(\omega) - X(\omega)| \to 0 \quad k \to \infty \quad Y_k \xrightarrow{a.s.} X$

and hence.

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \mathbb{P}(|X_n - X_{n_k}| \geq \tfrac{\varepsilon}{2}) + \mathbb{P}(|X_{n_k} - X| > \tfrac{\varepsilon}{2}) \to 0$$

$\Rightarrow X_n \xrightarrow{\mathbb{P}} X$

(c) Since $X_n \xrightarrow{\mathbb{P}} X$

the pair (X_i, X_j) and (Y_i, Y_j) have the same distribution for all i, j.

From (a) we have:

$$\forall \varepsilon > 0 \quad \mathbb{P}(|X_n - X_m| > \varepsilon) = \mathbb{P}(|Y_m - Y_n| > \varepsilon) \to 0$$

From (b) we have:

$\exists Y$ s.t. $Y_n \xrightarrow{\mathbb{P}} Y$

$$X_n - X_1 = \sum_{k=1}^{n-1} (X_{k+1} - X_k)$$
$$Y_n - Y_1 = \sum_{k=1}^{n-1} (Y_{k+1} - Y_k)$$

$\Rightarrow$ $\{X_n - X_1\}$ and $\{Y_n - Y_1\}$ have same distribution

$X_n - X_1 \xrightarrow{\mathbb{P}} X - X_1 \qquad Y_n - Y_1 \xrightarrow{\mathbb{P}} Y - Y_1$

$\Rightarrow \{X - X_1\}$ and $\{Y - Y_1\}$ have same distribution.

$\Rightarrow X, Y$ have same distribution.

Thus, $Y_n \xrightarrow{\mathbb{P}} Y$ $\quad Y$ has the same distribution as X

Q6. My research field is machine learning.

In machine learning, we train the model with traing data where training error is made; we evaluate the performance of the model with testing data where testing error is made.

An important question is: How much the testing error is close to the training data?

Since the traing data and testing data are sampled from the data pool randomly. We may as well solve the problem in the probability sense.

Let $X_1 \cdots X_n$ be n i.i.d. random variables with mean $\mu = \mathbb{E}X_i$

Let $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Given $\varepsilon > 0$, we are interested in estimating the tail probability

$$\mathbb{P}(\overline{X_n} \geq \mu + \varepsilon)$$

$$\mathbb{P}(\overline{X_n} \leq \mu - \varepsilon)$$

In machine learning, we can think $\overline{X_n}$ as training error observed on the training data.

The unknown mean μ is the test error which we want to minimize.

Therefore, in machine learning, these tail inequalities can be interpreted as: with high probability, the test error is close to training error.

Inequality often used includes:

1. Hoeffding Inequality
2. Bennett's Inequality
3. Bernstein Inequality

For example we can apply 1 Hoeffding Inequality to the function $\hat{f}$ (model) learned from the training data S_n since it is a random function that depends on S_n.

Actually, we can apply Hoeffding Inequality to each fixed function $f \in C$ (C is the concept class)

$$\mathbb{P}(err(f) \geq \hat{err}_{S_n}(f) + \varepsilon) \leq \exp(-2n\varepsilon^2)$$

Setting $N \exp(-2n\varepsilon^2) = \delta$ and solving for ε to get

$$\varepsilon = \sqrt{\frac{\ln(N/\delta)}{2n}}$$

$\Rightarrow$ with probability at least $1-\delta$, the following inequality holds for all $f \in C$

$$err(f) < \hat{err}_{S_n}(f) + \sqrt{\frac{\ln(N/\delta)}{2n}}$$

Such an result is called uniform convergence in the generalization analysis of machine learning.