Probability Theory – Exercise 8

Junwoo Yang

December 13, 2020

Problem 1

Use the Central Limit Theorem to estimate the probability that the number of Heads in 1000 independent tosses differs from 500 by less than 2%.

Proof. In this case, X_n is a sequence of independent Bernoulli random variables, with each X_n taking the values 1 and 0, each with probability $\frac{1}{2}$. Then, $\mathbb{E}(X_n) = \frac{1}{2}$, $\mathrm{Var}(X_n) = \frac{1}{4}$. By de Moivre-Laplace Theorem,

$$P(Z_n \leq t) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}x^2}\,\mathrm{d}x \quad \text{as } n \to \infty \quad \text{where } Z_n = \frac{S_n - n/2}{\sqrt{n}/2}.$$

Thus, we approximate the probability as follows.

$$\begin{aligned} P(|S_{1000} - 500| < 10) &= P(\frac{|S_{1000} - 500|}{\sqrt{1000}/2} < \frac{10}{\sqrt{1000}/2}) = P(-\frac{10}{\sqrt{250}} < Z_{1000} < \frac{10}{\sqrt{250}}) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-10/\sqrt{250}}^{10/\sqrt{250}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x \approx 0.4729. \end{aligned}$$

□

Problem 2

How many tosses of a coin are required to have the probability at least 0.99 that the average number of Heads differs from 0.5 by less than 1%?

Proof. Similarly, under the condition on n, we approximate the probability as follows.

$$\begin{aligned} P(|\frac{S_n}{n} - 0.5| < 0.005) &= P(\frac{|S_n - 0.5n|}{0.5\sqrt{n}} < \frac{0.005n}{0.5\sqrt{n}}) = P(|Z_n| < 0.01\sqrt{n}) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-0.01\sqrt{n}}^{0.01\sqrt{n}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x \geq 0.99. \end{aligned}$$

Hence, at least 66,349 tosses of a coin are required. □

Problem 3

Assume that $\frac{1}{c_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}(|X_k - \mu_k|^{2+\delta}) \to 0$ for some $\delta > 0$. Show that Lindeberg's condition (8.4) holds.

Proof. For fixed $\varepsilon > 0$ and δ that satisfies given assumption,

$$\begin{aligned}
\frac{1}{c_n^2}\sum_{k=1}^{n}\int_{\{x:|x-\mu_k|\geq\varepsilon c_n\}}(x-\mu_k)^2\,\mathrm{d}P_{X_k}(x) &\leq \frac{1}{c_n^2}\sum_{k=1}^{n}\int_{\{x:|x-\mu_k|\geq\varepsilon c_n\}}\frac{|x-\mu_k|^{2+\delta}}{(\varepsilon c_n)^{\delta}}\,\mathrm{d}P_{X_k}(x)\\
&= \frac{1}{\varepsilon^{\delta}c_n^{2+\delta}}\sum_{k=1}^{n}\int_{\{x:|x-\mu_k|\geq\varepsilon c_n\}}|x-\mu_k|^{2+\delta}\,\mathrm{d}P_{X_k}(x)\\
&\leq \frac{1}{\varepsilon^{\delta}c_n^{2+\delta}}\sum_{k=1}^{n}\int_{\mathbb{R}}|x-\mu_k|^{2+\delta}\,\mathrm{d}P_{X_k}(x)\\
&= \frac{1}{\varepsilon^{\delta}c_n^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}(|X_k-\mu_k|^{2+\delta}).
\end{aligned}$$

Since the last term converges to zero by the assumption, Lindeberg's condition holds. □

Problem 4

Suppose that $X_1, \cdots, X_n$ are independent random variables. Show that for any $a \geq 0$,

$$P(\max_{1\leq k\leq n}|S_k| \geq 3a) \leq 3\max_{1\leq k\leq n}P(|S_k| \geq a)$$

where $S_k = X_1 + \cdots + X_k$.

Proof. We describe first instance that $|S_k|$ exceeds $3a$. Namely, we write

$$A_k = \{\omega : |S_1(\omega)| < 3a, \cdots, |S_{k-1}(\omega)| < 3a, |S_k(\omega)| \geq 3a\}.$$

Then A_k are clearly pairwise disjoint and

$$\bigcup_{k=1}^{n} A_k = \{\omega : \max_{1\leq k\leq n}|S_k| \geq 3a\}.$$

$$\begin{aligned}
P(\max_{1\leq k\leq n}|S_k| \geq 3a) &= P(\{\max_{1\leq k\leq n}|S_k| \geq 3a\}\cap\{|S_n| \geq a\}) + P(\{\max_{1\leq k\leq n}|S_k| \geq 3a\}\cap\{|S_n| < a\})\\
&\leq P(|S_n| \geq a) + P(\bigcup_{k=1}^{n} A_k \cap \{|S_n - S_k| > 2a\})\\
&= P(|S_n| \geq a) + \sum_{k=1}^{n} P(A_k \cap \{|S_n - S_k| > 2a\})\\
&= P(|S_n| \geq a) + \sum_{k=1}^{n} P(A_k)P(|S_n - S_k| > 2a) \qquad (1)\\
&\leq P(|S_n| \geq a) + \max_{1\leq k\leq n}P(|S_n - S_k| > 2a)\sum_{k=1}^{n} P(A_k)\\
&\leq P(|S_n| \geq a) + \max_{1\leq k\leq n}P(|S_n - S_k| > 2a)
\end{aligned}$$

$$\begin{aligned}
&\le P(|S_n| \ge a) + \max_{1\le k\le n} \left(P(|S_n| \ge a) + P(|S_k| \ge a)\right) \\
&= 2P(|S_n| \ge a) + \max_{1\le k\le n} P(|S_k| \ge a) \\
&\le 3 \max_{1\le k\le n} P(|S_k| \ge a)
\end{aligned}$$

where (1) holds since X_k are independent, so S_k and $S_n - S_k$ are independent. This is Etemadi's inequality. $\square$

Problem 5

Suppose that $\{X_n\}$ are independent random variables and $\mathbb{E}(X_n) = 0$. Show that if $\sum_n \mathrm{Var}(X_n) < \infty$, then $\sum_n X_n$ converges with probability 1.

Proof. Consider partial sum $S_N = \sum_{n=1}^N X_n$. To show that $\sum_{n=1}^\infty X_n = \lim_{N\to\infty} S_N$ converges with probability 1, it is sufficient to prove that

$$\limsup_{N\to\infty} S_N - \liminf_{N\to\infty} S_N = 0$$

with probability 1. For any $m \in \mathbb{N}$,

$$\limsup_{N\to\infty} S_N - \liminf_{N\to\infty} S_N = \limsup_{N\to\infty}(S_N - S_m) - \liminf_{N\to\infty}(S_N - S_m) \le 2\max_{k\in\mathbb{N}} \left|\sum_{i=1}^k X_{m+i}\right|.$$

Thus, for any $\varepsilon > 0$,

$$\begin{aligned}
P(\limsup_{N\to\infty} S_N - \liminf_{N\to\infty} S_N \ge \varepsilon) &\le P(2\max_{k\in\mathbb{N}} \left|\sum_{i=1}^k X_{m+i}\right| \ge \varepsilon) = P(\max_{k\in\mathbb{N}} \left|\sum_{i=1}^k X_{m+i}\right| \ge \frac{\varepsilon}{2}) \\
&= P(\lim_{n\to\infty} \max_{1\le k\le n} \left|\sum_{i=1}^k X_{m+i}\right| \ge \frac{\varepsilon}{2}) = \lim_{n\to\infty} P(\max_{1\le k\le n} \left|\sum_{i=1}^k X_{m+i}\right| \ge \frac{\varepsilon}{2}) \\
&\le \lim_{n\to\infty} \frac{4}{\varepsilon^2} \mathrm{Var}(\sum_{i=1}^n X_{m+i}) = \frac{4}{\varepsilon^2} \lim_{n\to\infty} \sum_{i=1}^n \mathrm{Var}(X_{m+i}).
\end{aligned}$$

While the second inequality is due to Kolmogorov's inequality. Since $\sum_{n=1}^\infty \mathrm{Var}(X_n) < \infty$, it follows that the last term tends to 0 as m goes to infinity, for every arbitrary $\varepsilon > 0$. $\square$

Problem 6

Let $\{X_n\}$ be i.i.d. random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}(|X_1|^p) < 1$ for some $1 < p < 2$. Show that

$$\lim_{n\to\infty} \frac{S_n}{n^{1/p}} = 0 \quad \text{a.s.} \tag{1}$$

Proof of Problem 6

It is Marcinkiewicz–Zygmund theorem. To prove this theorem we need the following lemmas.

Lemma 1. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of identically distributed random variables. If* $E(|X_1|^p) < \infty$ *where* $1 < p < 2$*, then*

$$\frac{\sum_{i=1}^{n} |X_i| \mathbb{I}_{\{|X_i| > n^{1/p}\}}}{n^{1/p}} \to 0 \quad a.s. \tag{2}$$

Proof of Lemma 1. Let $U_n = |X_n|^p \mathbb{I}_{\{|X_n| > n^{1/p}\}}$, $n \geq 1$. Note that condition $E(|X_1|^p) < \infty$ is equivalent to the relation

$$\sum_{n=1}^{\infty} P(|X_1| > n^{1/p}) < \infty. \tag{3}$$

Thus, we have

$$\sum_{n=1}^{\infty} P(U_n \neq 0) = \sum_{n=1}^{\infty} P(|X_n| > n^{1/p}) = \sum_{n=1}^{\infty} P(|X_1| > n^{1/p}) < \infty.$$

Therefore, by Borel-Cantelli lemma,

$$U_n \to 0 \quad \text{a.s.} \tag{4}$$

Moreover

$$\frac{\sum_{i=1}^{\infty} |X_i| \mathbb{I}_{\{|X_i| > n^{1/p}\}}}{n^{1/p}} \leq \frac{\sum_{i=1}^{\infty} |X_i|^p \mathbb{I}_{\{|X_i| > n^{1/p}\}}}{n} \leq \frac{\sum_{i=1}^{\infty} |X_i|^p \mathbb{I}_{\{|X_i| > i^{1/p}\}}}{n}. \tag{5}$$

By (4) the right-hand side of (5) converges to zero almost sure and relation (2) follows. □

Lemma 2. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of identically distributed random variables. If* $E(|X_1|^p) < \infty$ *where* $1 < p < 2$*, then*

$$\frac{\sum_{i=1}^{n} |X_i| \mathbb{I}_{\{|X_i| > n^{1/p}\}}}{n^{1/p}} \to 0 \quad (n \to \infty). \tag{6}$$

Lemma 3. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of identically distributed random variables. If* $E(|X_1|^p) < \infty$ *where* $1 < p < 2$*, then*

$$\sum_{n=1}^{\infty} \frac{1}{2^{\frac{2n}{p}}} \sum_{k=1}^{2^n} \mathbb{E}(|X_k|^2 \mathbb{I}_{\{|X_k| \leq 2^{\frac{n}{p}}\}}) < \infty. \tag{7}$$

Proof of Problem 6. Let

$$X_i^{(n)} = X_i \mathbb{I}_{\{|X_i| \leq n^{1/p}\}}, \quad i \geq 1,\, n \geq 1,$$

$$S_j^{(n)} = \sum_{i=1}^{j} X_i^{(n)}, \quad j \geq 1,\, n \geq 1.$$

Step 1. Let us prove that

$$\frac{S_n - S_n^{(n)}}{n^{1/p}} \to 0 \quad \text{a.s.} \tag{8}$$

We have

$$\frac{|S_n - S_n^{(n)}|}{n^{1/p}} = \frac{|\sum_{i=1}^{n} X_i \mathbb{I}_{\{|X_i| > n^{1/p}\}}|}{n^{1/p}} \leq \frac{\sum_{i=1}^{n} |X_i| \mathbb{I}_{\{|X_i| > n^{1/p}\}}}{n^{1/p}}.$$

Application of Lemma 1 yields to (8).

Step 2. Let us prove that

$$\frac{\mathbb{E}(S_n^{(n)})}{n^{1/p}} \to 0 \quad (n \to \infty). \tag{9}$$

We have $\square$