

APMA1690: Homework # 2 (Due by 11pm Sep 28)

1 Question 1

1.1 Review before Question 1: the Law of Large Numbers (LLN)

Theorem 1.1 (Etemadi's strong LLN, 1981, see [Etemadi \(1981\)](#) or Theorem 5.17 of [Klenke \(2020\)](#)). Let Z_1, Z_2, \dots be real-valued random variables defined on probability space (Ω, \mathbb{P}) . Suppose they are identically distributed and *pairwise independent* (i.e., Z_i and Z_j are independent for all i, j with $i \neq j$). If $\mathbb{E}Z_1$ exists, then $\lim_{n \rightarrow \infty} (\frac{1}{n} \sum_{i=1}^n Z_i) = \mathbb{E}Z_1$ with probability one, i.e.,

$$(1.1) \quad \mathbb{P} \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n Z_i(\omega) \right) = \mathbb{E}Z_1 \right\} = 1.$$

1.2 Setup of Question 1

Let (Ω, \mathbb{P}) be a probability space as follows

- The set Ω is defined by

$$\Omega = \underbrace{\{0, 1\} \times \{0, 1\} \times \cdots \times \{0, 1\} \times \cdots}_{\text{Cartesian product, infinitely many } \{0, 1\}},$$

where each element $\omega \in \Omega$ is an infinitely long vector of the form $\omega = \underbrace{(\omega_1, \omega_2, \dots, \omega_n, \dots)}_{\text{infinitely many entries}}$, and $\omega_i \in \{0, 1\}$ for any positive integer i .

- \mathbb{P} is a probability and satisfies the following: for any positive integer i , we have

$$\mathbb{P}\{\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega \mid \omega_i = 1\} = \mathbb{P}\{\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega \mid \omega_i = 0\} = \frac{1}{2};$$

furthermore, for any positive integer k and any vector

$$\mathbf{v} = (v_1, v_2, \dots, v_k) \in \underbrace{\{0, 1\} \times \cdots \times \{0, 1\}}_{\text{Cartesian product, } k \text{ times}},$$

we have $\mathbb{P}\{\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega \mid \omega_i = v_i \text{ for all } i = 1, 2, \dots, k\} = (1/2)^k$.

We define the following two subsets of Ω

$$\begin{aligned} \Omega_1 &= \left\{ \omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega \mid \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{i=1}^n \omega_i \right] = \frac{1}{2} \right\}, \\ \Omega_0 &= \left\{ \omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega \mid \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{i=1}^n \omega_i \right] \neq \frac{1}{2} \right\}; \end{aligned}$$

1.3 Question 1

- a (1 points) For the following two elements $\omega^{(1)}$ and $\omega^{(2)}$ in Ω , which of them belongs to Ω_1 ? Which of them belongs to Ω_0 ? Explain (rather than mathematically prove) your answer.

$$\begin{aligned}\omega^{(1)} &= (1, 0, 0, 1, 0, 0, 1, 0, 0, \dots), & (\text{repetition of the "1,0,0" pattern}), \\ \omega^{(2)} &= (1, 0, 1, 0, 1, 0, 1, 0, \dots), & (\text{repetition of the "1,0,1,0" pattern}).\end{aligned}$$

Both elements are members of Ω_1 . By the law of large numbers, the average should tend towards the expected value which is $1/2$ because having a 1 or a 0 is equally likely at each index.

- b (2 points) Prove the following using the LLN in a rigorous way

$$\mathbb{P}(\Omega_1) = 1 \quad \text{and} \quad \mathbb{P}(\Omega_0) = 0.$$

Let $\{X_i\}_{i=1}^{\infty}$ be an infinite sequence of iid random variables such that $X_1 \sim \text{Bernoulli}(\frac{1}{2})$. Then

$$\mathbb{E}X_1 = 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) = \mathbb{P}(X=1) = \frac{1}{2}$$

and

$$\begin{aligned}\mathbb{P}(\Omega_1) &= \mathbb{P}\left(\{\vec{\omega} \in \Omega \mid \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{i=0}^n X_i\right] = \frac{1}{2}\}\right) \\ &= \mathbb{P}\left(\{\vec{\omega} \in \Omega \mid \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{i=0}^n X_i\right] = \mathbb{E}X_1\}\right) \\ &\stackrel{LLN}{=} 1\end{aligned}$$

Then because Ω_1 and Ω_0 partition Ω , we know that

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega_1) + \mathbb{P}(\Omega_0) \implies 1 = 1 + \mathbb{P}(\Omega_0) \implies \mathbb{P}(\Omega_0) = 0 \quad \blacksquare$$

2 Question 2

2.1 Review before Question 2

We have the following two versions of the [Law of the Iterated Logarithm](#)

- (Heuristic/sloppy version.) Suppose X_1, \dots, X_n, \dots are iid random variables defined on the probability space (Ω, \mathbb{P}) . The mean is $\mu = \mathbb{E}X_1$ and the standard deviation is $\sigma = \sqrt{\text{Var}(X_1)}$. Then, we “approximately” (rather than “exactly”) have the following inequality when the sample size n is very large, where “log” is the [natural logarithm](#).

$$\left| \frac{1}{n} \sum_{i=1}^n X_i(\omega) - \mu \right| \leq \sigma \cdot \sqrt{\frac{2 \cdot \log(\log n)}{n}}.$$

- (Mathematical/rigorous version, see Theorem 22.11 of [Klenke \(2020\)](#); not required) Suppose X_1, \dots, X_n, \dots are iid random variables defined on the probability space (Ω, \mathbb{P}) . The mean is $\mu = \mathbb{E}X_1$ and the standard variance is $\sigma = \sqrt{\text{Var}(X_1)}$. Then, we have

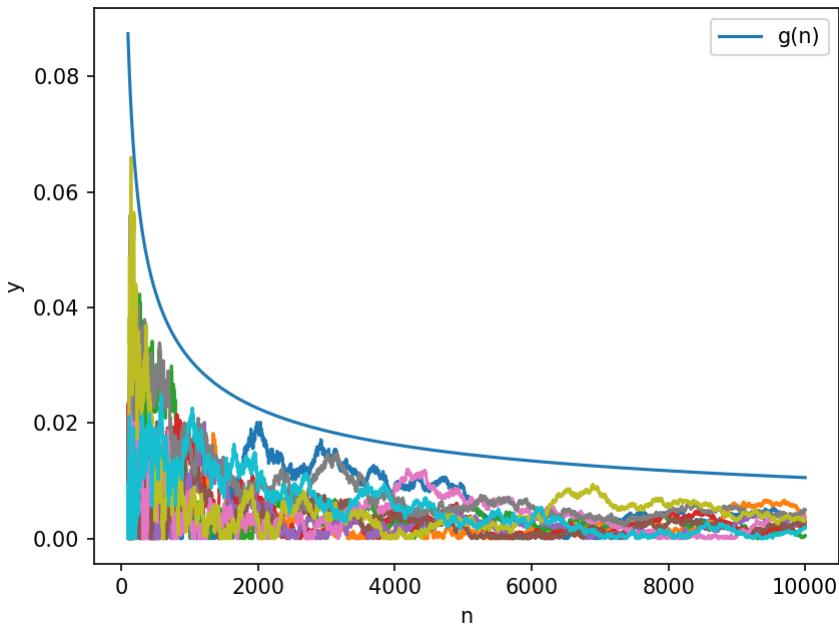
$$\mathbb{P} \left\{ \omega \in \Omega \mid \limsup_{n \rightarrow \infty} \left[\frac{\left| \frac{1}{n} \sum_{i=1}^n X_i(\omega) - \mu \right|}{\sigma \cdot \sqrt{\frac{2 \cdot \log(\log n)}{n}}} \right] = 1 \right\} = 1.$$

2.2 Question 2

Use any programming language of your choice to execute the following steps:

- Step 1: Generate 10,000 random numbers $x_1, x_2, \dots, x_{10000}$ from the distribution $\text{Bernoulli}(0.5)$.
- Step 2: For each n in the range 100 to 10,000, compute y_n using the formula $y_n = \left| \frac{1}{n} \sum_{i=1}^n x_i - 0.5 \right|$.
- Step 3: Plot the graph of “ y_n versus n ” (with n on the horizontal axis).
- Step 4: Repeat Steps 1-3 nine more times, and overlay all ten graphs on the same plot.
- Step 5: Define a function $g(n) = 0.5 \cdot \sqrt{\frac{2 \cdot \log(\log n)}{n}}$. Then, add the graph of “ $g(n)$ versus n ” to the plot from Step 4.

Display the picture obtained in Step 5 and provide the code used to generate it (2 points). Explain the picture from Step 5 (1 point).



```

import matplotlib.pyplot as plt
from scipy.stats import bernoulli
from math import sqrt, log

def avg(lst):
    return sum(lst)/len(lst)

len_rands = 10000
x = range(100, len_rands)

for plot in range(10):
    rands = bernoulli.rvs(0.5, size=len_rands)
    running_avg = 0
    y = []
    for n in x:
        running_avg = ((running_avg * (n-1)) + (rands[n] - 0.5))/n
        y.append(abs(running_avg))
    plt.plot(x, y)

g = list(map(lambda n: 0.5 * sqrt((2 * log(log(n)))/n), x))
plt.plot(x, g, label="g(n)")

plt.xlabel("n")
plt.ylabel("y")
plt.legend()
plt.show()

```

The plot shown above illustrates the law of the iterated logarithm. First, I generated a series of iid random variables $\{X_i\}_{i=1}^{10000} \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ with $\mu = \mathbb{E}X_1 = \frac{1}{2}$. Then for larger and larger n , I plotted the mean absolute error

$$e_n = |\bar{X} - \mu|$$

for ten different series of random variables to illustrate that they all share a similar shape but are random. Finally, the function $g(n)$ is the quantity $\sqrt{\text{Var}X_1} \cdot \sqrt{\frac{2 \log \log n}{n}}$. The final graph confirms the heuristic that the running error is less than the quantity $g(n)$.

3 Question 3

3.1 Review before Question 3

- (Empirical CDF.) Let X_1, X_2, \dots, X_n be random variables defined on the probability space (Ω, \mathbb{P}) . For each fixed ω , consider the indicator function $\mathbb{1}\{X_i(\omega) \leq t\}$, which is a function of t . The function $F_n(\omega, t)$, defined as follows, is called the empirical CDF based on the random variables X_1, X_2, \dots, X_n

$$F_n(\omega, t) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i(\omega) \leq t\} = \frac{\text{number of } \{X_i(\omega)\}_{i=1}^n \text{ such that } X_i(\omega) \leq t}{n}, \quad \text{for all } t \in \mathbb{R}.$$

- The following is the [Glivenko-Cantelli theorem](#).

Theorem 3.1 (Glivenko-Cantelli, 1933; see Theorem 5.23 of [Klenke \(2020\)](#)). *Let $F(t)$ be the CDF shared by iid random variables X_1, \dots, X_n, \dots , and $F_n(\omega, t)$ is the empirical CDF based on $\{X_i(\omega)\}_{i=1}^n$. Then, the empirical CDF converges to $F(t)$ uniformly with probability one, i.e.,¹,*

$$(3.1) \quad \mathbb{P} \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \left(\sup_{t \in \mathbb{R}} |F_n(\omega, t) - F(t)| \right) = 0 \right\} = 1.$$

3.2 You may directly apply the following results to solve Question 3

Please feel free to utilize the following two results without proving them.

- (i) (Dvoretzky-Kiefer-Wolfowitz inequality, 1956, see [Dvoretzky et al. \(1956\)](#))

Let $F(t)$ be the CDF shared by iid random variables X_1, \dots, X_n, \dots . The function $F_n(\omega, t)$ represents the empirical CDF based on $\{X_i(\omega)\}_{i=1}^n$. Then, there exists a positive constant C (not depending on F) such that

$$\mathbb{P} \left\{ \omega \in \Omega \mid \sup_{t \in \mathbb{R}} |F_n(\omega, t) - F(t)| > \epsilon \right\} \leq Ce^{-2n\epsilon^2},$$

for any $\epsilon > 0$ and positive integer n .

- (ii) (See Theorem 1.8 of [Shao \(2003\)](#)) Let $Z(\omega), Z_1(\omega), \dots, Z_n(\omega), \dots$ be random variables. If, for every $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ \omega \in \Omega \mid Z_n(\omega) \geq \epsilon \right\} < \infty,$$

then Z_n converges to 0 with probability one, i.e.,

$$\mathbb{P} \left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} Z_n(\omega) = 0 \right\} = 1.$$

¹The “sup” in Eq. (3.1) denotes the “supremum.” If you are not familiar with the “sup,” please feel free to just view it as [maximum](#), i.e., “max.”

3.3 Question 3

(2 points) Prove the Glivenko-Cantelli theorem (i.e., Theorem 3.1) using the two results presented above.

We let X_1, \dots, X_n be iid random variables sharing the CDF $F(t)$ which form the basis for the empirical CDF $F_n(\omega, t)$.

We note that the sup function fixes a value of t so the quantity $\sup_{t \in \mathbb{R}} |F_n(\omega, t) - F(t)|$ is a random variable. We call it $Z_n(\omega)$.

By the Dvoretzky-Kiefer-Wolfowitz inequality,

$$\mathbb{P}(Z_n > \varepsilon) \leq Ce^{-2n\varepsilon^2}$$

Thus, we calculate

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(Z_n \geq \varepsilon) &= \sum_{n=1}^{\infty} \mathbb{P}(Z_n = \varepsilon) + \sum_{n=1}^{\infty} \mathbb{P}(Z_n > \varepsilon) \\ &\leq \sum_{n=1}^{\infty} \mathbb{P}(Z_n = \varepsilon) + \sum_{n=1}^{\infty} Ce^{-2n\varepsilon^2} \end{aligned}$$

Because ε can be any positive real number, Z_n must be continuous or it can not hit every value of ε . Either way, $\mathbb{P}(Z_n = \varepsilon) = 0$ for any particular ε . Thus,

$$\sum_{n=1}^{\infty} \mathbb{P}(Z_n \geq \varepsilon) \leq \sum_{n=1}^{\infty} Ce^{-2n\varepsilon^2}$$

Looking at the RHS, we observe that the series $a_n = Ce^{-2\varepsilon^2}$ is positive (because C is positive) and monotonically decreasing so we can apply the integral convergence test (using the properties of the Gaussian Integral):

$$\int_1^{\infty} Ce^{-2x\varepsilon^2} dx \stackrel{a=2\varepsilon^2}{=} \int_1^{\infty} Ce^{-ax^2} dx < \int_0^{\infty} Ce^{-ax^2} = \frac{1}{2}\sqrt{\frac{\pi}{a}} < \infty$$

Then by Theorem 1.8 of Shao 2003,

$$\mathbb{P}\left(\left\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} Z_n(\omega) = 0\right\}\right) = \mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \left(\sup_{t \in \mathbb{R}} |F_n(\omega, t) - F(t)|\right) = 0\right\}\right) = 1$$

Which is precisely the Glivenko-Cantelli theorem. ■

4 Question 4

4.1 Review before Question 4

Definition 4.1. Suppose g is a real-valued function defined on \mathbb{R} .

- Let X be a discrete random variable whose CDF is $\sum_{k=0}^K p_k \cdot \mathbb{1}_{[x_k, \infty)}(x)$. If $\sum_{k=0}^K |g(x_k)| \cdot p_k < \infty$, then the following sum is called the mean/expected value of $g(X)$ and denoted as $\mathbb{E}[g(X)]$

$$\mathbb{E}[g(X)] \stackrel{\text{def}}{=} \sum_{k=0}^K g(x_k) \cdot p_k.$$

If $\sum_{k=0}^K |g(x_k)| \cdot p_k = \infty$, we say the expected value of $g(X)$ does not exist.

- Let X be a continuous random variable with PDF $p_X(x)$. If $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x) dx < \infty$, we call the following integral as the mean/expected value of $g(X)$ and denote it as $\mathbb{E}[g(X)]$

$$\mathbb{E}[g(X)] \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} g(x) \cdot p_X(x) dx.$$

If $\int_{-\infty}^{+\infty} |g(x)| \cdot p_X(x) dx = \infty$, we say the expected value of X does not exist.

4.2 Question 4

Let X be a continuous random variable whose PDF is the following²

$$(4.1) \quad p_X(x) = \frac{1}{\pi(1+x^2)}, \quad \text{for all } x \in \mathbb{R}.$$

a (1 point) Question: Does the expected value $\mathbb{E}X$ of X exist? Please prove your answer.

$$\int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} dx = \int_{-\infty}^0 \frac{-x}{\pi(1+x^2)} + \int_0^{\infty} \frac{x}{\pi(1+x^2)} dx = 2 \int_0^{\infty} \frac{x}{\pi(1+x^2)} dx$$

But

$$\begin{aligned} \frac{1}{\pi} \int_0^{\infty} \frac{2x}{1+x^2} dx &= \frac{1}{\pi} \int_0^{\infty} \frac{1}{u} du = \\ &\text{frac1}\pi [\ln |1+x^2|]_0^{\infty} = \infty - 0 = \infty \end{aligned}$$

So the integral does not converge.

Thus, the expected value

$$\mathbb{E}X = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx$$

does not exist. ■

²The PDF corresponds to a Cauchy distribution.

b Suppose we have done the following

- For each positive integer n , we generated random numbers $\xi_1, \xi_2, \dots, \xi_n$ from the distribution whose PDF is the one presented in Eq. (4.1) (i.e., the Cauchy distribution with parameters “ $(x_0 = 0, \gamma = 1)$ ”; see the Wikipedia page about [Cauchy distribution](#)).
- We computed the sample average $\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$.
- We plotted the “sample average $\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ vs. sample size n ” curves.
- We repeated this procedure 1000 times and got 1000 such curves. The 1000 curves are presented in Figure 1.

The R code for generating the figure is provided as follows.

```

1   m=1000
2   X=rcauchy(m)
3   X_bar=cumsum(X)/(1:m)
4   plot(1:m, X_bar, type = "l", xlab = "Sample size n", ylab = "Sample
average",
5       ylim = c(-50, 50), lwd=0.2,
6       main = "Sample average vs. sample size")
7   for (i in 1:999) {
8     X=rcauchy(m)
9     X_bar=cumsum(X)/(1:m)
10    lines(1:m, X_bar,
11          lwd=0.2)
12  }
13  abline(h=0, lty=2, col="red", lwd=2)
14

```

(1 point) Question: When $n \rightarrow \infty$, does $\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ converge to anything with probability one? Please heuristically (rather than mathematically/rigorously) explain your answer using both Figure 1 and the Law of Large Numbers.

The law of large numbers says that the sample average of a sequence of n independently and identically distributed random variables converges to the expected value of the corresponding CDF with probability 1 if the expected value exists. In the case of the Cauchy distribution, however, the expected value does not exist so the LLN does not apply. Experimental results (as in Fig. 1) suggest that there is no uniformity whatsoever in the convergence pattern of $\bar{\xi}_n$.

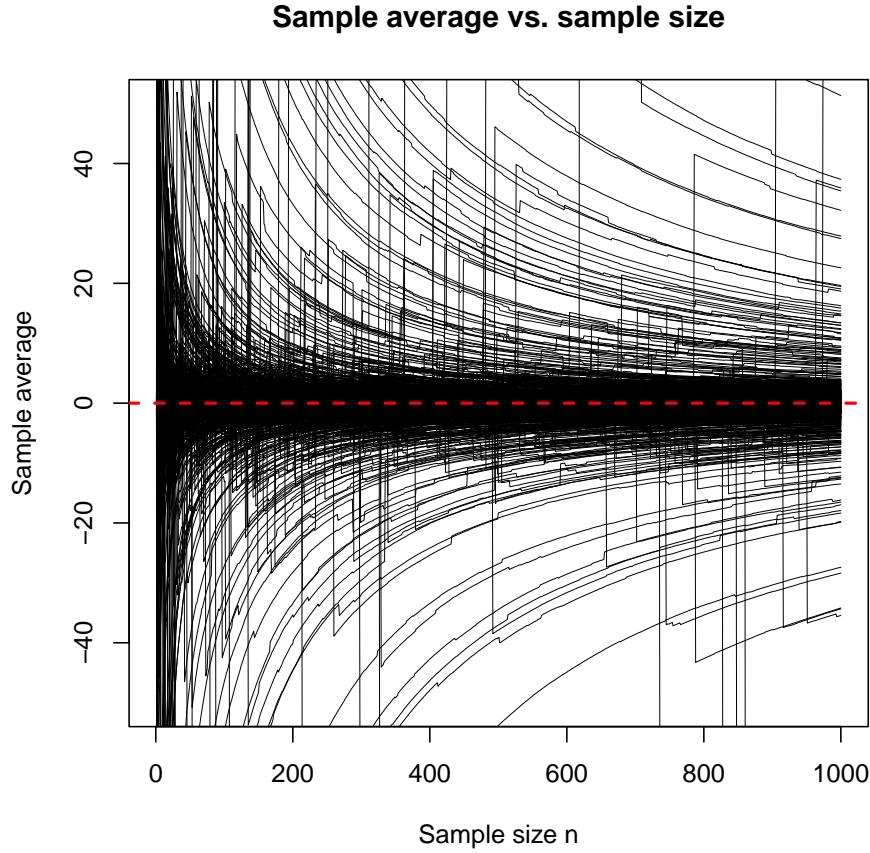


Figure 1: The 1000 “sample average $\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ vs. sample size n ” curves.

References

- A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.
- N. Etemadi. An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 55(1):119–122, 1981.
- A. Klenke. *Probability theory: a comprehensive course, 3rd Edition*. Springer Science & Business Media, 2020.
- J. Shao. *Mathematical statistics*. Springer Science & Business Media, 2003.