# Probability and Statistics MA6.101
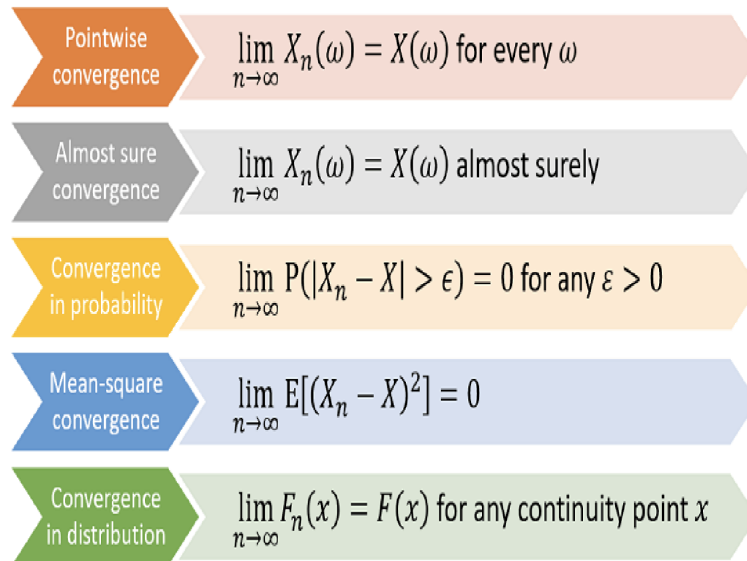
# Tutorial 8

Topics Covered: Convergence, Central Limit Theorem

## Summary



| Pointwise convergence | $\lim_{n \to \infty} X_n(\omega) = X(\omega)$ for every $\omega$ |
| Almost sure convergence | $\lim_{n \to \infty} X_n(\omega) = X(\omega)$ almost surely |
| Convergence in probability | $\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0$ for any $\varepsilon > 0$ |
| Mean-square convergence | $\lim_{n \to \infty} E[(X_n - X)^2] = 0$ |
| Convergence in distribution | $\lim_{n \to \infty} F_n(x) = F(x)$ for any continuity point $x$ |

`https://en.wikipedia.org/wiki/Convergence_of_random_variables`

Q1: Determine whether the following series' converges or diverges

(a) $\sum_{n=1}^{\infty} \frac{n!}{n^n}$

(b) $\sum_{n=1}^{\infty} \left(\frac{3n+1}{4n}\right)^n$

**Solution**

**(a) Ratio Test.**

For a series $\sum a_n$ with $a_n > 0$, define

$$L = \lim_{n \to \infty} \frac{a_{n+1}}{a_n}.$$

- If $L < 1$, the series converges absolutely.
- If $L > 1$ (or $L = \infty$), the series diverges.
- If $L = 1$, the test is inconclusive.

**Applying to** $a_n = \dfrac{n!}{n^n}$:

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!/(n+1)^{n+1}}{n!/n^n} = \frac{(n+1)n^n}{(n+1)^{n+1}} = \left(\frac{n}{n+1}\right)^n.$$

As $n \to \infty$,

$$L = \lim_{n\to\infty} \left(\frac{n}{n+1}\right)^n = \frac{1}{e} < 1.$$

Hence, the series $\sum_{n=1}^{\infty} \frac{n!}{n^n}$ **converges absolutely.**

**(b) Root Test.**

For a series $\sum a_n$, define

$$L = \lim_{n\to\infty} \sqrt[n]{|a_n|}.$$

- If $L < 1$, the series converges absolutely.
- If $L > 1$, the series diverges.
- If $L = 1$, the test is inconclusive.

**Applying to** $a_n = \left(\dfrac{3n+1}{4n}\right)^n$:

$$\sqrt[n]{|a_n|} = \frac{3n+1}{4n} \xrightarrow[n\to\infty]{} \frac{3}{4}.$$

Thus, $L = \frac{3}{4} < 1$, so the series **converges absolutely.**

Q2: Let $U$ be a random variable having a uniform distribution on the interval $[0, 1]$.

Now, define a sequence of random variables $\{X_n\}$ as follows:

$$X_1 = \mathbf{1}_{\{U\in[0,1]\}},$$
$$X_2 = \mathbf{1}_{\{U\in[0,1/2)\}}, \quad X_3 = \mathbf{1}_{\{U\in[1/2,1]\}},$$
$$X_4 = \mathbf{1}_{\{U\in[0,1/4)\}}, \quad X_5 = \mathbf{1}_{\{U\in[1/4,2/4)\}}, \quad X_6 = \mathbf{1}_{\{U\in[2/4,3/4)\}}, \quad X_7 = \mathbf{1}_{\{U\in[3/4,1]\}},$$
$$X_8 = \mathbf{1}_{\{U\in[0,1/8)\}}, \quad X_9 = \mathbf{1}_{\{U\in[1/8,2/8)\}}, \quad X_{10} = \mathbf{1}_{\{U\in[2/8,3/8)\}}, \quad \cdots$$
$$X_{16} = \mathbf{1}_{\{U\in[0,1/16)\}}, \quad X_{17} = \mathbf{1}_{\{U\in[1/16,2/16)\}}, \quad X_{18} = \mathbf{1}_{\{U\in[2/16,3/16)\}}, \quad \cdots$$

where $\mathbf{1}_{\{U\in[a,b)\}}$ is the indicator function of the event $\{U \in [a, b)\}$.

Find the probability limit (if it exists) of the sequence $\{X_n\}$.

**Solution:**

The sequence $\{X_n\}$ is constructed by partitioning $[0, 1]$ into finer and finer intervals. The pattern is:

- $X_1$ covers $[0, 1]$
- $X_2, X_3$ partition into halves: $[0, 1/2)$ and $[1/2, 1]$
- $X_4, X_5, X_6, X_7$ partition into quarters
- $X_8, \ldots, X_{15}$ partition into eighths

- And so on...

Any term $X_n$ of the sequence, being an indicator function, can take only two values:

- It can take value 1 with probability

$$\mathbb{P}(X_n = 1) = \mathbb{P}\left(U \in \left[\frac{j}{m}, \frac{j+1}{m}\right)\right) = \frac{1}{m}$$

  where $m$ is an integer satisfying

$$\frac{n}{2} < m \leq n$$

  and $j$ is an integer satisfying
$$n = m + j$$

- It can take value 0 with probability

$$\mathbb{P}(X_n = 0) = 1 - \mathbb{P}(X_n = 1) = 1 - \frac{1}{m}$$

By the previous inequality, $m$ goes to infinity as $n$ goes to infinity and

$$\lim_{n \to \infty} \mathbb{P}(X_n = 0) = \lim_{m \to \infty} \left(1 - \frac{1}{m}\right) = 1$$

Therefore, the probability that $X_n$ is equal to zero converges to 1 as $n$ goes to infinity. So, obviously, $\{X_n\}$ converges in probability to the constant random variable

$$X = 0$$

because, for any $\epsilon > 0$,

$$
\begin{aligned}
\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \epsilon) &= \lim_{n \to \infty} \mathbb{P}(|X_n - 0| > \epsilon) \\
&= \lim_{n \to \infty} \mathbb{P}(X_n > \epsilon) \quad \text{(because } X_n \text{ is positive)} \\
&= \lim_{n \to \infty} \mathbb{P}(X_n = 1) \quad \text{(because } X_n \text{ can take only value 0 or value 1)} \\
&= \lim_{n \to \infty} \frac{1}{m} = 0
\end{aligned}
$$

**Almost sure convergence:**

The sequence does **not** converge almost surely to 0 because for every $u \in [0, 1]$, the sequence $\{X_n(u)\}$ does not converge. It oscillates between 0 and 1 infinitely often.

Hence: $X_n \xrightarrow{\mathbb{P}} 0$ but $X_n \xrightarrow{a.s.} 0$.

Q3: If $X_n \xrightarrow{d} c$, where $c$ is a constant, then show that $X_n \xrightarrow{p} c$.

**Solution:** Since $X_n \xrightarrow{d} c$, we conclude that for any $\epsilon > 0$, we have

$$\lim_{n \to \infty} F_{X_n}(c - \epsilon) = 0, \qquad \lim_{n \to \infty} F_{X_n}\left(c + \frac{\epsilon}{2}\right) = 1.$$

We can write for any $\epsilon > 0$,

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \epsilon) &= \lim_{n\to\infty} \left[ \mathbb{P}(X_n \leq c - \epsilon) + \mathbb{P}(X_n \geq c + \epsilon) \right] \\
&= \lim_{n\to\infty} \mathbb{P}(X_n \leq c - \epsilon) + \lim_{n\to\infty} \mathbb{P}(X_n \geq c + \epsilon) \\
&= \lim_{n\to\infty} F_{X_n}(c - \epsilon) + \lim_{n\to\infty} \mathbb{P}(X_n \geq c + \epsilon) \\
&= 0 + \lim_{n\to\infty} \mathbb{P}(X_n \geq c + \epsilon) \quad (\text{since } \lim_{n\to\infty} F_{X_n}(c - \epsilon) = 0) \\
&\leq \lim_{n\to\infty} \mathbb{P}(X_n > c + \tfrac{\epsilon}{2}) \\
&= 1 - \lim_{n\to\infty} F_{X_n}\left(c + \tfrac{\epsilon}{2}\right) \\
&= 0 \quad (\text{since } \lim_{n\to\infty} F_{X_n}\left(c + \tfrac{\epsilon}{2}\right) = 1).
\end{aligned}
$$

Since $\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \epsilon) \geq 0$, we conclude that

$$
\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \epsilon) = 0, \quad \forall \epsilon > 0,
$$

which means $X_n \xrightarrow{p} c$.

Q4: Let $X_1, X_2, \cdots, X_n$ be i.i.d. with finite mean E(X) and variance Var(X). Then $S_n \to E(X)$ in m.s. Where $S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.
**Solution:** Here, we need to show that $\mathbb{E}\big(|S_n - E(X)|^2\big) \to 0$ as $n \to \infty$.

$$
\mathbb{E}\big(S_n\big) = \mathbb{E}\big(X\big)
$$

Hence,
$$
\mathbb{E}\big(|S_n - E(X)|^2\big) = \text{Var}(S_n) = \frac{\text{Var}(X)}{n} \to 0 \text{ as } n \to \infty.
$$

Q5: It is possible for a sequence of discrete random variables to converge in distribution to a continuous one. For example, if $Y_n$ is uniform on $\{1, \ldots, n\}$ and $X_n = Y_n/n$, then $X_n$ converges in distribution to a random variable which is uniform on $[0, 1]$.

**Solution:** For any $x \in [0, 1]$,

$$
F_{X_n}(x) = P(X_n \leq x) = P\left(Y_n \leq nx\right) = \frac{\lfloor nx \rfloor}{n}.
$$

For $x < 0$, $F_{X_n}(x) = 0$ and for $x > 1$, $F_{X_n}(x) = 1$.
Thus,
$$
\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \frac{\lfloor nx \rfloor}{n} = x = F_X(x),
$$

for $x \in [0, 1]$, 0 for $x < 0$ and 1 for $x > 1$, where $X$ is uniform on $[0, 1]$. Since $F_X$ is continuous on $[0, 1]$, we have $X_n \xrightarrow{d} X$.

Q6: Assume that a test has a mean score of 75 and a standard deviation of 10. Assume the distribution of scores is approximately normal.

  • What is the probability that a person chosen at random will make 100 or above on the test?

- In a group of 100 people, how many would you expect to score below 60?
- What is the probability that the mean of a group of 100 will score below 70?

**Solution:**

Let $X$ denote the test score. Given: $X \sim N(75, 10^2)$.

- **Probability of scoring 100 or above:**

$$\begin{aligned} \mathbb{P}(X \geq 100) &= \mathbb{P}\left(\frac{X - 75}{10} \geq \frac{100 - 75}{10}\right) \\ &= \mathbb{P}(Z \geq 2.5) \\ &= 1 - \Phi(2.5) \approx 1 - 0.9938 = 0.0062 \end{aligned}$$

So the probability is approximately $0.62\%$

- **Expected number scoring below 60 in a group of 100:**
First, find $\mathbb{P}(X < 60)$:

$$\begin{aligned} \mathbb{P}(X < 60) &= \mathbb{P}\left(\frac{X - 75}{10} < \frac{60 - 75}{10}\right) \\ &= \mathbb{P}(Z < -1.5) \\ &= \Phi(-1.5) \approx 0.0668 \end{aligned}$$

Expected number $= 100 \times 0.0668 \approx 6.68 \approx 7$ people.

- **Probability that the mean of 100 people is below 70:**
Let $\bar{X}$ be the sample mean of 100 people. Then:

$$\bar{X} \sim N\left(75, \frac{10^2}{100}\right) = N(75, 1)$$

$$\begin{aligned} \mathbb{P}(\bar{X} < 70) &= \mathbb{P}\left(\frac{\bar{X} - 75}{1} < \frac{70 - 75}{1}\right) \\ &= \mathbb{P}(Z < -5) \\ &\approx 2.87 \times 10^{-7} \end{aligned}$$

This probability is extremely small (essentially 0).

Q7: Find
$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i}.$$
where $X_i$ are i.i.d. and $X_i \sim U[0, 1]$.

**Solution:** By the Strong Law of Large Numbers (SLLN),

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}[X_1] = \frac{1}{2}$$

and
$$\frac{1}{n}\sum_{i=1}^{n}X_i^2 \xrightarrow{a.s.} \mathbb{E}[X_1^2] = \int_0^1 x^2\,dx = \frac{1}{3}.$$

Thus,
$$\frac{\sum_{i=1}^{n}X_i^2}{\sum_{i=1}^{n}X_i} = \frac{\frac{1}{n}\sum_{i=1}^{n}X_i^2}{\frac{1}{n}\sum_{i=1}^{n}X_i} \xrightarrow{a.s.} \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Q8:
- Find the probability of getting more than 55 heads after tossing a fair coin 100 times
- Find the probability of getting more than 220 heads after tossing the same coin for 400 times.
- Are the probabilities of these events the same?. If not, why so?
- For a 100 flips of the same fair coin, find the probability of getting 40 to 60 heads.

**Solution:**

Let $X$ be the number of heads in $n$ tosses of a fair coin. Then $X \sim \text{Binomial}(n, p = 0.5)$.

For large $n$, by the Central Limit Theorem:
$$\frac{X - np}{\sqrt{np(1-p)}} \approx N(0,1)$$

- **More than 55 heads in 100 tosses:**
  Here $n = 100$, $p = 0.5$, so $\mathbb{E}[X] = 50$ and $\text{Var}(X) = 25$, thus $\sigma = 5$.
  Using continuity correction:
  $$\mathbb{P}(X > 55) = \mathbb{P}(X \geq 56) \approx \mathbb{P}(X \geq 55.5)$$
  $$= \mathbb{P}\left(\frac{X-50}{5} \geq \frac{55.5-50}{5}\right)$$
  $$= \mathbb{P}(Z \geq 1.1)$$
  $$= 1 - \Phi(1.1) \approx 1 - 0.8643 = 0.1357$$

- **More than 220 heads in 400 tosses:**
  Here $n = 400$, so $\mathbb{E}[X] = 200$ and $\sigma = \sqrt{100} = 10$.

  $$\mathbb{P}(X > 220) \approx \mathbb{P}(X \geq 220.5)$$
  $$= \mathbb{P}\left(\frac{X-200}{10} \geq \frac{220.5-200}{10}\right)$$
  $$= \mathbb{P}(Z \geq 2.05)$$
  $$= 1 - \Phi(2.05) \approx 1 - 0.9798 = 0.0202$$

- **Are the probabilities the same?**
  No, the probabilities are different. In the first case, $\mathbb{P}(X > 55) \approx 0.1357$ while in the second case, $\mathbb{P}(X > 220) \approx 0.0202$.

Why is this so? Although both events represent getting more than 10% excess heads relative to the expected value, the standardized deviation from the mean is different. In the first case, we're 1.1 standard deviations above the mean, while in the second case, we're 2.05 standard deviations above the mean. As the sample size increases, higher deviations (in terms of standard deviations) become less likely.

- **Probability of 40 to 60 heads in 100 tosses:**

$$\mathbb{P}(40 \leq X \leq 60) \approx \mathbb{P}(39.5 < X < 60.5)$$
$$= \mathbb{P}\left(\frac{39.5 - 50}{5} < Z < \frac{60.5 - 50}{5}\right)$$
$$= \mathbb{P}(-2.1 < Z < 2.1)$$
$$= \Phi(2.1) - \Phi(-2.1)$$
$$= 2\Phi(2.1) - 1 \approx 2(0.9821) - 1 = 0.9642$$

Q9: A bank teller serves customers standing in the queue one by one. Suppose that the service time $X_i$ for customer $i$ has mean $\mathbb{E}[X_i] = 2$ minutes and $\mathrm{Var}(X_i) = 1$. Assume that service times for different customers are independent.

Let $Y$ be the total time the bank teller spends serving 50 customers. Find $\mathbb{P}(90 < Y < 110)$.

**Solution.**

Let
$$Y = X_1 + X_2 + \cdots + X_n,$$

where $n = 50$, $\mathbb{E}[X_i] = \mu = 2$, and $\mathrm{Var}(X_i) = \sigma^2 = 1$.

Then
$$\mathbb{E}[Y] = n\mu = 100, \quad \mathrm{Var}(Y) = n\sigma^2 = 50.$$

We can write

$$\mathbb{P}(90 < Y \leq 110) = \mathbb{P}\left(\frac{90 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - n\mu}{\sqrt{n}\sigma}\right).$$

Substituting the values:

$$\mathbb{P}(90 < Y \leq 110) = \mathbb{P}\left(\frac{90 - 100}{\sqrt{50}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - 100}{\sqrt{50}}\right) = \mathbb{P}\left(-\sqrt{2} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \sqrt{2}\right).$$

By the **Central Limit Theorem (CLT)**, the standardized variable

$$Z = \frac{Y - n\mu}{\sqrt{n}\sigma}$$

is approximately standard normal, $Z \sim N(0, 1)$.

Hence,
$$\mathbb{P}(90 < Y \leq 110) \approx \Phi(\sqrt{2}) - \Phi(-\sqrt{2}),$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

Using symmetry $\Phi(-x) = 1 - \Phi(x)$, we get:

$$\Phi(\sqrt{2}) - \Phi(-\sqrt{2}) = 2\Phi(\sqrt{2}) - 1 = 0.8427.$$

$$\boxed{\mathbb{P}(90 < Y < 110) \approx 0.8427.}$$