

Exercise Walkthrough: Convergence of Normal Random Variables

Justin Lanfermann

25. June 2025

Overview

This document provides a detailed walkthrough for an exercise on the convergence of a sequence of normally distributed random variables. We are given a sequence of random variables $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ and we want to analyze the behavior of this sequence as the parameters μ_n and σ_n^2 converge.

The exercise is split into two parts:

1. **Part (i):** Proving convergence in distribution ($X_n \xrightarrow{d} X$) when the variance converges to a positive constant $\sigma^2 > 0$.
2. **Part (ii):** Proving convergence in probability ($X_n \xrightarrow{P} \mu$) when the variance converges to zero.

We will use key concepts from the lecture notes, such as characteristic functions, Lévy's Continuity Theorem, and Chebyshev's inequality, explaining each step and its justification.

1 Part (i): Convergence in Distribution

Problem Statement

Show that if $\mu_n \rightarrow \mu \in \mathbb{R}$ and $\sigma_n^2 \rightarrow \sigma^2 > 0$ as $n \rightarrow \infty$, then $X_n \xrightarrow{d} X$, where $X \sim \mathcal{N}(\mu, \sigma^2)$.

Explanation and Solution

Step 1: The Strategy

Our goal is to prove convergence in distribution. The hint points us towards using characteristic functions (CFs). This is a very powerful technique. According to **Lévy's Continuity Theorem** [2] (mentioned in **Def. 2.56 (iii)** of the script), a sequence of random variables X_n converges in distribution to X if and only if their characteristic functions $\phi_{X_n}(t)$ converge pointwise to the characteristic function $\phi_X(t)$ for every $t \in \mathbb{R}$.

So, our plan is:

1. Write down the characteristic function for $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$.
2. Write down the characteristic function for the target variable $X \sim \mathcal{N}(\mu, \sigma^2)$.
3. Show that the first CF converges to the second as $n \rightarrow \infty$.

Step 2: The Characteristic Functions

From **Example 2.48** in the script, we know the characteristic function [1] of a normal distribution $\mathcal{N}(m, s^2)$ is given by:

$$\phi(t) = \exp\left(imt - \frac{1}{2}s^2t^2\right)$$

Applying this formula to our variables:

- For $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$, the CF is:

$$\phi_{X_n}(t) = \exp\left(i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right)$$

- For the target variable $X \sim \mathcal{N}(\mu, \sigma^2)$, the CF is:

$$\phi_X(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$$

Step 3: Taking the Limit

Now we compute the limit of $\phi_{X_n}(t)$ as $n \rightarrow \infty$.

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \lim_{n \rightarrow \infty} \exp\left(i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right)$$

The exponential function $\exp(z)$ is continuous for all $z \in \mathbb{C}$. This is a crucial property, as it allows us to "push the limit inside" the function:

$$\lim_{n \rightarrow \infty} \exp(z_n) = \exp\left(\lim_{n \rightarrow \infty} z_n\right)$$

Applying this to our case, we get:

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \exp\left(\lim_{n \rightarrow \infty} \left(i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right)\right)$$

We are given that $\mu_n \rightarrow \mu$ and $\sigma_n^2 \rightarrow \sigma^2$. Since t is a fixed real number, we can evaluate the limit of the exponent:

$$\lim_{n \rightarrow \infty} \left(i\mu_n t - \frac{1}{2}\sigma_n^2 t^2\right) = i\left(\lim_{n \rightarrow \infty} \mu_n\right)t - \frac{1}{2}\left(\lim_{n \rightarrow \infty} \sigma_n^2\right)t^2 = i\mu t - \frac{1}{2}\sigma^2 t^2$$

Plugging this back in, we find:

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right) = \phi_X(t)$$

Step 4: Conclusion

We have shown that for every $t \in \mathbb{R}$, the characteristic function of X_n converges to the characteristic function of X . By Lévy's Continuity Theorem (**Def. 2.56**), this implies that the sequence of random variables X_n converges in distribution to X .

$$X_n \xrightarrow{d} X \sim \mathcal{N}(\mu, \sigma^2)$$

Check Your Understanding

Question: Why was the condition $\sigma^2 > 0$ important here? What might change if $\sigma^2 = 0$? (This is a great lead-in to the next part of the exercise!)

2 Part (ii): Convergence in Probability

Problem Statement

Show that if $\mu_n \rightarrow \mu \in \mathbb{R}$ and $\sigma_n^2 \rightarrow 0$ as $n \rightarrow \infty$, then $X_n \xrightarrow{P} \mu$.

Explanation and Solution

Here, the limit of the variance is zero. This means the normal distributions are getting "skinnier" and more concentrated. The hint suggests two possible paths. We'll explore both, as they showcase different but equally important concepts.

Method 1: Using Chebyshev's Inequality

This method is a direct proof using one of the fundamental probability inequalities.

The Strategy: According to **Def. 2.56 (i)**, convergence in probability means that for any $\epsilon > 0$, we must show:

$$\lim_{n \rightarrow \infty} P(|X_n - \mu| > \epsilon) = 0$$

We will use **Chebyshev's Inequality [3] (Thm. 2.40)**, which states for a random variable Y :

$$P(|Y - E[Y]| \geq \epsilon) \leq \frac{\text{Var}[Y]}{\epsilon^2}$$

Let's apply this to X_n . We have $E[X_n] = \mu_n$ and $\text{Var}[X_n] = \sigma_n^2$. Chebyshev's inequality gives us:

$$P(|X_n - \mu_n| \geq \epsilon) \leq \frac{\sigma_n^2}{\epsilon^2}$$

The challenge is that the inequality involves $|X_n - \mu_n|$, but we need to prove something about $|X_n - \mu|$. We can bridge this gap. For any $\epsilon > 0$, we can use the triangle inequality: $|X_n - \mu| = |X_n - \mu_n + \mu_n - \mu| \leq |X_n - \mu_n| + |\mu_n - \mu|$.

This implies that the event $\{|X_n - \mu| \geq \epsilon\}$ is a subset of the event $\{|X_n - \mu_n| + |\mu_n - \mu| \geq \epsilon\}$. Let's rewrite this as $\{|X_n - \mu_n| \geq \epsilon - |\mu_n - \mu|\}$.

Since $\mu_n \rightarrow \mu$, for any $\delta > 0$ (let's pick $\delta = \epsilon/2$), there exists an N such that for all $n > N$, we have $|\mu_n - \mu| < \epsilon/2$. For these large n , we have $\epsilon - |\mu_n - \mu| > \epsilon - \epsilon/2 = \epsilon/2$.

So, for $n > N$, the following holds:

$$P(|X_n - \mu| \geq \epsilon) \leq P(|X_n - \mu_n| \geq \epsilon - |\mu_n - \mu|) < P(|X_n - \mu_n| \geq \epsilon/2)$$

Now we can apply Chebyshev's inequality to the right-hand side, using $\epsilon/2$ as our new deviation:

$$P(|X_n - \mu_n| \geq \epsilon/2) \leq \frac{\text{Var}[X_n]}{(\epsilon/2)^2} = \frac{\sigma_n^2}{\epsilon^2/4} = \frac{4\sigma_n^2}{\epsilon^2}$$

So, for $n > N$, we have the bound:

$$0 \leq P(|X_n - \mu| \geq \epsilon) \leq \frac{4\sigma_n^2}{\epsilon^2}$$

As $n \rightarrow \infty$, we know $\sigma_n^2 \rightarrow 0$. Therefore, the right-hand side goes to 0. By the Squeeze Theorem, we conclude:

$$\lim_{n \rightarrow \infty} P(|X_n - \mu| > \epsilon) = 0$$

This proves that $X_n \xrightarrow{P} \mu$.

Method 2: Using Convergence in Distribution to a Constant

This method is more elegant and relies on theory we have already established.

The Strategy: We will use the fact that convergence in distribution to a constant implies convergence in probability (**Remark 2.57**) [4]. So the plan is:

1. Show that X_n converges in distribution to the constant value μ .
2. Invoke the theorem from the script to conclude convergence in probability.

From Part (i), we already know that if $\mu_n \rightarrow \mu$ and $\sigma_n^2 \rightarrow \sigma^2$, then $X_n \xrightarrow{d} X \sim \mathcal{N}(\mu, \sigma^2)$. In this part, we are given that $\sigma_n^2 \rightarrow 0$. So we can set $\sigma^2 = 0$.

This means X_n converges in distribution to a variable $X \sim \mathcal{N}(\mu, 0)$. A normal distribution with zero variance is not really "random"; it's a **degenerate distribution**, which is a point mass at its mean. That is, $P(X = \mu) = 1$. This is a constant random variable.

Let's confirm this using characteristic functions. The CF of the constant random variable $X = \mu$ is:

$$\phi_\mu(t) = E[e^{itX}] = E[e^{it\mu}] = e^{it\mu}$$

From our result in Part (i), the limit of the CFs of X_n is:

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \exp\left(i\mu t - \frac{1}{2}(0)t^2\right) = \exp(i\mu t) = \phi_\mu(t)$$

The CFs of X_n converge to the CF of the constant μ . Therefore, by Lévy's Continuity Theorem, $X_n \xrightarrow{d} \mu$.

Now we use the final piece of the puzzle. As stated in **Remark 2.57**, if a sequence converges in distribution to a constant c , it also converges in probability to that constant. Since $X_n \xrightarrow{d} \mu$, we can directly conclude:

$$X_n \xrightarrow{P} \mu$$

Summary and Takeaways

We have successfully proven both parts of the exercise.

- For Part (i), we saw that the convergence of the parameters (μ_n, σ_n^2) directly leads to the convergence of the characteristic functions, which in turn implies convergence in distribution. This highlights the power of CFs as a tool for proving distributional convergence.
- For Part (ii), we demonstrated two ways to prove convergence in probability. The first method, using Chebyshev's inequality, is a "from first principles" approach. The second, more streamlined method, shows how theoretical results (like the link between convergence modes) can make proofs much simpler by allowing us to reuse previous results.

This exercise is a foundational result for many statistical applications, where we often approximate the distribution of a statistic with a normal distribution, and understanding how these approximations behave is key.

Deeper Dive: Explaining the Concepts

[1] Characteristic Functions (CFs)

The characteristic function is like a unique "fingerprint" for a probability distribution. It's a function $\phi_X(t) = \mathbb{E}[e^{itX}]$ that always exists for any random variable X . Its main power comes from a few key properties (**Remark 2.47**):

- **Uniqueness:** If two random variables have the same CF, they have the same distribution.
- **Independence:** The CF of a sum of independent random variables is the product of their individual CFs ($\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$).
- **Convergence:** The convergence of CFs is linked to the convergence of distributions, as we saw in this exercise.

This makes it an incredibly powerful tool for proving theorems about distributions, especially for sums of random variables (like in the Central Limit Theorem).

[2] Lévy's Continuity Theorem

This is the formal theorem that provides the bridge we used. It states that a sequence of random variables X_n converges in distribution to a random variable X if and only if the sequence of their characteristic functions $\phi_{X_n}(t)$ converges pointwise to a function $\phi(t)$ for all $t \in \mathbb{R}$, where $\phi(t)$ is the characteristic function of X . Essentially, it guarantees that if the "fingerprints" converge, the distributions themselves converge. This is stated in **Def. 2.56** and **Remark 2.49**.

[3] Chebyshev's Inequality

Chebyshev's inequality (**Thm. 2.40**) is a beautifully simple and universal result. It provides an upper bound on the probability that a random variable deviates from its mean by more than a certain amount. The only information it needs is the variance of the variable. Its formula is $P(|X - \mu| \geq \epsilon) \leq \sigma^2/\epsilon^2$. Because it makes no assumptions about the shape of the distribution (it could be normal, uniform, Poisson, etc.), the bound it gives is often not very tight, but its generality makes it extremely useful for theoretical proofs, like the one for the Weak Law of Large Numbers (**Thm. 2.61**).

[4] Convergence Modes and Constants

As laid out in **Remark 2.57**, there's a hierarchy of convergence modes: almost sure convergence is the strongest, followed by convergence in probability, and then convergence in distribution.

$$\text{Almost Sure} \implies \text{In Probability} \implies \text{In Distribution}$$

The reverse implications are not generally true. However, there is a very important special case: when the limit is a constant. If a sequence of random variables X_n converges *in distribution* to a constant c , then it also converges *in probability* to c . This is a handy shortcut that we used in the second method for Part (ii). It makes sense intuitively: if the entire "shape" of the distribution collapses to a single point, then the probability of being anywhere else must be vanishing.