Probability & Statistics II

Contents

5.2 The Law of Large Numbers (LLN)

The theme in this section is the following: "if we add lots of random variables then the "errors" average out."

Before we state and prove the LLN, let us recall the following property of the variance which plays a very important role in the proof.

Lemma 18. *If $X_1, X_2, ..., X_n$ is a sequence of independent random variables with $E(X_i) = \mu_j$, $\text{Var}(X_j) = \sigma_j^2$ then*

$$\text{Var}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \sigma_j^2.$$

Proof. See Coursework for the proof. □

Theorem 19 (Law of Large Numbers)**.** *Suppose that $X_1, X_2, \ldots$ is a sequence of independent random variables with mean μ and variance σ^2. Let*

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then for any number $\varepsilon > 0$

$$P(|Y_n - \mu| \le \varepsilon) \to 1 \qquad \text{as } n \to \infty.$$

Proof. We have

$$E(Y_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu,$$

and

$$\text{Var}(Y_n) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{\sigma^2}{n},$$

where we use two properties of variance: $\text{Var}(cZ) = c^2\text{Var}(Z)$ and Lemma 18.

Hence by Chebyshev's inequality we have

$$P(|Y_n - \mu| > \varepsilon) \le \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Since $\frac{\sigma^2}{n\varepsilon^2}$ tends to zero as $n \to \infty$, so does $P(|Y_n - \mu| > \varepsilon)$. Hence

$$P(|Y_n - \mu| \le \varepsilon) = 1 - P(|Y_n - \mu| > \varepsilon) \to 1.$$

□

Remark. This is also called the *weak LLN*. It basically says that for some specified "large" n, the average Y_n of the $(X_1, \ldots, X_n)$ is likely to be close to the mean μ. In fact, we can repeat the arguments of the above proof, to also proved a useful estimate:

$$P(|Y_n - \mu| \le \varepsilon) \ge 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

5.3 Central Limit Theorem

We have seen that the average value $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ converges to the mean (when the X_k are independent identically distributed random variables with finite variance). However, different random variables converge at different rates to the mean (that is some converge more quickly than others). The Central Limit Theorem gives a much more precise description of the behaviour of the Y_n. We basically define a "scaled version" of Y_n which has zero mean and variance 1.

Theorem 20 (Central Limit Theorem)**.** *Suppose that $X_1, X_2, X_3, \ldots$ are independent identically distributed random variables with mean μ and variance σ^2. Let*

$$Z_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}}$$

Then Z_n converges, as $n \to \infty$, to a normal random variable with parameters $(0,1)$ in the sense that, for any s, t, such that $s < t$, we have

$$P(s \leq Z_n \leq t) \to \int_s^t \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \Phi(t) - \Phi(s),$$

where $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ is the cumulative distribution function of a standard Normal random variable.

Proof. Can be proved using moment generating functions and such a proof can be found in standard textbooks of probability (this is beyond the scope of this course). □

Statistical remarks

(1) The Central Limit Theorem (CLT) only tells you about what happens as $n \to \infty$.

(2) However, in Statistics, this is commonly (and very conveniently) used for finite but large values of n.

Suppose that $X_1, X_2, \ldots$ are independent identically distributed random variables with mean μ and variance σ^2. For "large" n, we define their sum by

$$S_n = \sum_{k=1}^{n} X_k.$$

According to the CLT,

the distribution of the random variable $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ *is approximately standard normal.*

Using this we can also characterise the distribution of the average Y_n as n increases, which is a very useful result in statistical applications. Namely, we get that

the distribution of the average $Y_n := \frac{S_n}{n}$ *is approximately* $N(\mu, \frac{\sigma^2}{n})$.

This result justifies also the extensive use of Normal distributions in real-life applications to model data resulting from many different independent factors (roughly independent) or when the distribution of data is unknown.

Finally, we can see from the above statements that

the distribution of the random variable S_n is approximately $N(n\mu, n\sigma^2)$.

Example. *Suppose that W_i is the amount (in pounds) that gambler i wins at a visit in a casino, for $i = 1, 2, \ldots$. These amounts are considered to be independent random variables with mean β pounds and variance $4\beta^2$.*

(i). *What is the approximate distribution of the total profit of 100 gamblers?*

(ii). *What is the approximate probability that the total profit of 100 gamblers is negative (i.e. casino wins money), if their individual average profit is $-£5$ for each player (negative profit translates to a loss)?*

Answer. (i). We know that the total profit of 100 customers is given by $T_{100} = W_1 + W_2 + \cdots + W_{100}$ where $W_1, W_2, \ldots$ are their individual profits and we know that they are independent random variables with mean β and variance $4\beta^2$. Hence $E(W_i) = \beta$ and $\mathrm{Var}(W_i) = 4\beta^2$. Therefore $E(T_{100}) = 100\beta$ and $\mathrm{Var}(T_{100}) = 400\beta^2$.

Hence by the approximate Central Limit Theorem

$$T_{100} \approx N\left(100\beta, 400\beta^2\right).$$

(ii). In this case, we have $\beta = -5$. The approximate distribution (from part (i)) is therefore

$$T_{100} \approx N\left(-500, 10000\right).$$

Supposing that $Z \sim N(0,1)$, this implies that

$$P(T_{100} < 0) \approx P\Big(Z < \frac{500}{100}\Big) = P(Z < 5) = 0.9987$$

Therefore, we do not expect that the casino will lose any money with a 99.87% chance. □