Law of large numbers and Central limit theorem

The law of large numbers

We will now turn our attention to one of the most central - and likewise most intuitive - laws in statistics that relates probability theory (so the theoretical, mathematical background of randomness and chance) to real-world phenomena. The law of large numbers gives us the statistical foundation to be able to relate real-world data to properties of theoretical distributions we have observed.

In that sense, the **law or large numbers** is our way of coping with random variation in random phenomenon, which we can eliminate from our measurements by repetition.

In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the **average of the results** obtained from a large number of trials should be close to the **expected value** and will tend to become closer to the expected value as more trials are performed.

As the number of trials/observations increases, the actual/observed probability approaches the theoretical/expected probability.

Example: Let's say that you flip a fair coin. The theoretical probability is 50% that you get heads. But if you flip a coin 10 times you are not guaranteed to get 5 heads. But is you flip a coin 10 times, and repeat that 1000 times, the observed proportion of heads will get closer and closer to 50%.

Gambler's Fallacy: A common misconception is that you expect a certain value to come. If you flip a coin 10 times, and the first 9 flips were tails, you would think that the next flip should be a heads (because the theoretical probability is 50/50) but that is not the case. Each flip of the coin is an independent event, and the outcome is unaffected by the previous event.

> In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the **average of the results** obtained from a large number of trials should be close to the **expected value** and will tend to become closer to the expected value as more trials are performed.

Proof of correctness: To provide statistical reasoning for why this rule - the *law of large numbers (LLN)* - holds, we consider the experiments as a sequence of random variables $X_1, X_2, ..., X_n$, where X_i is the specific result or outcome of the ith repetition of our experiment. Each of the random variables X_i has some expectation and standard deviation, which we denote as μ and σ. We confine ourselves to the situation where the experimental conditions and the subsequent experiments are identical, ie. we don't change the conditions of the experiment, and that the outcomes of some experiment doesn't influence the others, ie. we assume independence.

We call such a sequence an **independent and identically distributed (i.i.d) sequence**.

We can compute the average of this sequence through our standard method of averaging, i.e. we sum over the realisation of all random variables and divide by the number of repetitions:

$$\bar{X_n} = \frac{X_1 + X_2 + ... + X_n}{n}$$

$\bar{X}_n$ in itself is a random variable, since for a next n repetitions, we would assume a different value for $\bar{X}_n$. We can compute the expectation and variance of this new random variable.

Expectation

$$
\begin{aligned}
E[\bar{X}_n] &= E[\frac{1}{n}(X_1 + X_2 + ... + X_n)] \\
&= \frac{1}{n}E[X_1 + X_2 + ... + X_n] \\
&= \frac{1}{n}(E[X_1] + E[X_2] + ... + E[X_n]) \\
&= \frac{1}{n}(\mu + \mu + ... + \mu) \\
&= \frac{n \cdot \mu}{n} = \mu
\end{aligned}
$$

We see: When performing identical experiments a large number of times, the expected average value of the random experiment is equal to the expectation of each single random variable. Sometimes there will be values above our expectation, sometimes below - but if we average over it, we assume to obtain the same value again.

Variance

$$
\begin{aligned}
Var[\bar{X}_n] &= Var[\frac{1}{n}(X_1 + X_2 + ... + X_n)] \\
&= \frac{1}{n^2}Var[X_1 + X_2 + ... + X_n] \\
&= \frac{1}{n^2}(Var[X_1] + Var[X_2] + ... + Var[X_n]) \\
&= \frac{1}{n^2}(\sigma^2 + \sigma^2 + ... + \sigma^2) \\
&= \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}
$$

For the variance we have found an even more interesting property. The variance in the average of the realisation of a sequence of random variables, decreases the bigger the sequence, i.e. its standard deviation is less than that of a single realisation by a factor of $\sqrt{n}$. This property of the variance is what we use to proof the law of large numbers: The more often we repeat some experiment, the smaller the spread of the averaged results. \end{ocg}

Chebyshev's inequality

We have shown that increasing values of n in our random variable $\bar{X}_n$, which averages over the results of some random variable, result in a decreasing variance. If we were to plot the distribution of our new random variable for increasing values of n, we would observe a contraction of the probability mass around the expectation (less variance $\rightarrow$ less spread out probability mass). We will now learn about a tool, that will provide us with the final step to prove the law of large numbers formally.

Chebyshev's Inequality is a way to quantify the bound of the probability that any random variable Y is outside some symmetric interval of width $2a$ around its expectation. So outside the symmetric interval of: $[E[X] - a, E[X] + a]$. Chebyshev proved the following upper bound:

$$
P(|Y - E[Y]| > a) \leq \frac{1}{a^2} \cdot Var[Y]
$$

To explain Chebyshev's upper bound, it means the probability that the realisation of some random variable Y deviates more than a from its mean is **ALWAYS** lower than $\frac{1}{a^2} \cdot Var[Y]$. This is intuitive since we would expect the probability to get lower for bigger intervals (bigger values of a) and the probability to get higher for the higher variances.

The Chebyshev Inequality provides us with powerful tools to make statements about the probability mass of any random variable. If we want to say something about the probability that our random variable is within a few standard deviations we can employ Chebyshev's inequality.

Example of usage of Chebyshev's Inequality

Let's say at a factory we automatically fill up boxes with screws. The mean of the number of screws per box is 1000 and the variance is 25.

$\mu = 1000$ and $\sigma^2 = 25$ $\sigma = 5$. This variance is a problem since customers complain if they don't get 1000 screws but maybe get 992 screws and it is also a problem the other way around, since you are giving away free screws if the box has 1020 screws.

a) How many σ-units to the right of μ is 1009? So how many standard deviations away from the mean (to the right) is 1009?

There is no information given on the distribution of the data, so we cant expect a perfect normal distribution.

Useful formulas:

To convert a specific value of X into σ -units: $k = \frac{X-\mu}{\sigma}$ To convert a σ - unit to a specific value of X: $X = \mu + k\sigma$ Here k is the number of σ-units.

For problem a) we use the first formula.

$$k = \frac{1009 - 1000}{5} = 1.8$$

So this tells us that 1009 is 1.8 standard deviations away from the mean.

b) What X value is 2.6 σ-units to the left of μ?

Here we use the second formula:

$$X = 1000 - 2.6 \cdot 5 = 1000 - 13 = 987$$

This means that 2.6 standard deviation to the left of the mean lies the value: 987 screws.

Important: since the exercise asks for standard-deviation to the left, we insert a negative number. If they asked for standard-deviation to the right, we would add it, i.e.:

$$X = 1000 + 2.6 \cdot 5 = 1000 + 13 = 1013$$

c) Use Chebyshev's Inequality to find a bound on $P[994 < X < 1006]$

As we can see, the interval $[994 < X < 1006]$ is symmetric around the mean by 6. This means we can use Chebyshev's Inequality.

So first off we use the first formula to calculate how many standard deviations the two numbers are from the mean:

$$k = \frac{1006 - 1000}{5} = 1.2$$

Now we apply the second formula into Chebyshev's Inequality:

$$P(1000 - (1.2)(5) < X < 1000 + (1.2)(5)) \geq 1 - \frac{1}{(1.2)^2}$$

This is the same as just writing:

$$P(994 < X < 1006) \geq 1 - \frac{1}{(1.2)^2} = 0.3056$$

This means that the probability that X will lie between 994 and 1006 is AT LEAST 0.3056, note that you don't know exactly what the probability is, but Chebyshev's Inequality provides a bound.

Central Limit Theorem

The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution.

The central limit theory is a refinement of the law of large numbers. Again, we consider the average $\bar{X}_n$ over a series of independent, identically distributed (i.i.d) random variables $X_1, X_2, ..., X_n$ By the law of large numbers, we know that this average converges towards the true mean of the distribution $E[\bar{X}_n] = \mu$ for large n, which we were able to prove by applying the limit on Chebyshev's inequality.

The central limit theorem now extends this argument even further, stating that for a random variable modelling any kind of theoretical distribution, the average over n values, so $\bar{X}_n$ after normalisation converges towards the standard normal distribution.

An example of applying CLT to find a probability

Example: The average age at first marriage is 25 for women and 27.8 for men. If the standard deviation for women is 4 years, what is the probability that **a random selection of 32 women** have an average age at first marriage between 26 and 27?

The first thing we want to look at is the distribution. Since the problem, is a probability problem about an average that is derived from 32 women (so an average $\bar{X}$ derived from a random selection of women), then we can use the Central Limit Theorem. This tells us that $\bar{X}$ is normally distributed when the sample size is large (here we sampled 32 women, so we count that as a large n). So now we now we can use a Normal Distribution to calculate the problem.

We are only looking at a bell curve for the women, so we forget about the information given about the men.

So the μ for our normal distribution is 25.

IT IS IMPORTANT TO NOTE THAT we are not dealing with a normal X value, which represent the individual women's age at marriage. We are dealing with a sample of 32 women, so it not X we are dealing with, it is $\bar{X}$. It is not 1 randomly selected woman, it is 32 randomly selected women.

So, the mean for $\bar{X}$ is the same as μ. BUT the standard deviation is not just 4 we have to adjust it: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, so for out normal distribution we have:

$$\mu_{\bar{X}} = 25$$
$$\sigma_{\bar{X}} = \frac{4}{\sqrt{32}} = 0.7071068$$

So now we want to know the probability that the average lies between 26 and 27. (see drawing to the right, $z_1 = 26$ and $z_2 = 27$)

To calculate this we start off by calculating the z-scores of 26 and 27. recall that $z = \frac{X-\mu}{\sigma}$ and in this case $z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}}$

$$\text{z value for } 26 = \frac{26-25}{0.7071068} = 1.41 \text{z value for } 27 = \frac{27-25}{0.7071068} = 2.83$$

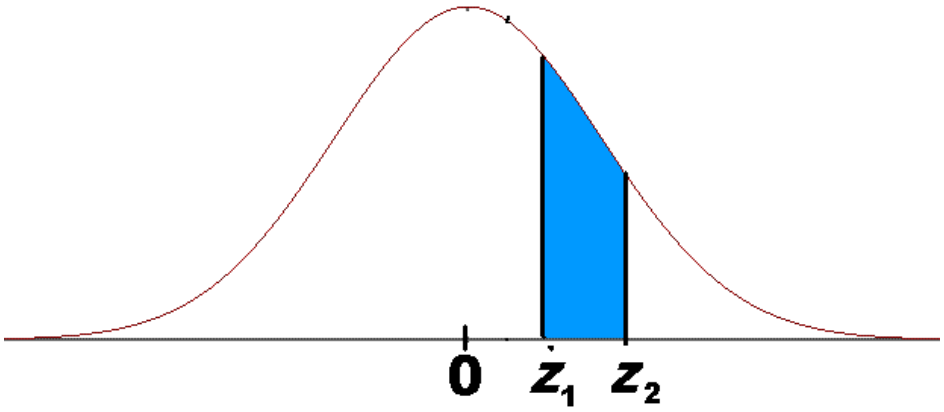

Figure 1: Normal distribution with marked area

You can either look it up at a z-table, or simply plug it into R:

```
pnorm(2.83) - pnorm(1.41)
[1] 0.07694244
```

You can also skip the z-score step and plug it into R like this:

```
pnorm(27, mean=25, sd=0.7071068) - pnorm(26, mean=25, sd=0.7071068)
[1] 0.07631074
```

So, $P(26 < \bar{X} < 27) = 0.07631 = 7.6\%$