The Law of Large Numbers

This is a key fundamental result, which is so intuitive that it is at the heart of the frequentist interpretation of probability, and is essential to the field of statistics.

**Theorem: The Law of Large Numbers (LLN)**

Let $X_1, X_2, \ldots$ be an infinite sequence of independent and identically distributed (i.i.d.) random variables, with mean $\mu_X = \mathbb{E}[X_i]$. Define the average

$$
\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i .
$$

Then, as $n \to \infty$, the average converges to a non-random real number. Specifically

$$
\lim_{{n \to \infty}} \bar{X}_n = \mu_x .
$$

In other words, if you repeat the same experiment in a independent way, the average of the outcomes is going to be very close to the mean !!!
Example

Take a regular coin, and start flipping it several times, taking note if it falls heads or tails (flip the coin in such a way that you cannot predict the outcome of each trial).

\[ X_i = \begin{cases} 1 & \text{if } i^{th} \text{ flip was ‘tails’} \\ 0 & \text{if } i^{th} \text{ flip was ‘heads’} \end{cases} \]

You see that the average is approaching very quickly the expected value, which is \( \frac{1}{2} \) in this case!!!
Example

Take a regular coin, and start flipping it several times, taking note if it falls heads or tails (flip the coin in such a way that you cannot predict the outcome of each trial).

\[
X_i = \begin{cases} 
1 & \text{if } i^{th} \text{ flip was 'tails'} \\
0 & \text{if } i^{th} \text{ flip was 'heads'}
\end{cases}
\]

You see that the average is approaching very quickly the expected value, which is \( \frac{1}{2} \) in this case!!!
Estimating Probabilities

This is one of the most common problems one needs to address. What is the probability of a certain event (e.g. what is the probability the cache of the router is full)?

Let $X$ be a random variable, to estimate $p = P(X \in A)$ where $A$ is an arbitrary set. Suppose we can generate $n$ independent and identically distributed samples from the distribution of $X$, denoted by $X_1, \ldots, X_n$. The Law of Large Numbers tells us that

$$\hat{p} = \frac{\text{number of } i \text{'s such that } X_i \in A}{n},$$

is a good estimator of $p$, in the sense that, as $n$ grows

$$\hat{p} \to p.$$ 

How large should we pick $n$ so that we get an accurate estimate?
Estimating Probabilities

Note that $Y = n\hat{p}$ is a binomial random variable $Y \sim \text{Bin}(n, p)$. As we have seen

$$
\mathbb{E}[Y] = np \quad \text{and} \quad \text{V}(Y) = np(1 - p) .
$$

This means that $\hat{p}$ is such that

$$
\mathbb{E}[\hat{p}] = p \quad \text{and} \quad \text{V}(\hat{p}) = \frac{p(1 - p)}{n} .
$$

Clearly the variability of $\hat{p}$ is decreasing fast, indicating that $\hat{p}$ should be getting closer and closer to the (unknown) constant $p$ with high probability.

So this gives us a simple way to use computers for the evaluation of probabilities - even in very complex scenarios !!!
Using Data to Estimate Probabilities

The Law of Large Numbers is what allows us to use data to estimate probabilities. Suppose you stand at the main entrance of the MetaForum, and count the number of female and male individuals you see entering the building for a period of 4 hours.

\[ x_i = \begin{cases} 
1 & \text{if student is a female} \\
0 & \text{if student is a male} 
\end{cases} \]

Say you counted 283 different individuals, and 27 were females.

Let \( Y \) be a random variable representing be the gender of a randomly chosen individual inside TU/e. Then it plausible that

\[ P(Y = 1) \approx \frac{27}{283} = 0.0954. \]

So provided you can make the assumption that the individuals entering the MetaForum are a representative independent and identically distributed sample of \( Y \) we can use the LLN to estimate it’s distribution…
Estimating Probability Mass Functions

Let $X$ be a discrete random variable, for which you want to determine the probability mass function. Suppose you can make $n$ i.i.d. observations of it. You can use this information to estimate the probability mass function of $X$ easily. Let's see an example:

(Bortkiewicz '1898) The number of fatalities resulting from being kicked by a horse was recorded for 10 corps of Prussian cavalry over a period of 20 years.

<table>
<thead>
<tr>
<th>Number of Deaths per Year</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>109</td>
</tr>
<tr>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

Let $X$ be a random variable representing the number of deaths within an year in a corp of Prussian cavalry.

$$P(X = x) \approx \begin{cases} 
109/203 & \text{if } x = 0 \\
65/203 & \text{if } x = 1 \\
22/203 & \text{if } x = 2 \\
3/203 & \text{if } x = 3 \\
1/203 & \text{if } x = 4 \\
0 & \text{otherwise}
\end{cases} = \begin{cases} 
0.527 & \text{if } x = 0 \\
0.320 & \text{if } x = 1 \\
0.108 & \text{if } x = 2 \\
0.015 & \text{if } x = 3 \\
0.005 & \text{if } x = 4 \\
0 & \text{otherwise}
\end{cases}$$