Homework 4

Statistical Independence

Statistical independence refers to a situation in which two events or random variables do not influence each other in any way.

To understand this better, let us take as an example two statistical variables (X and Y) within a population

X/Y	**Thin**	**Fat**	**Normal**	**Maringal(Y)**
Low	5	10	15	30
Med	5	10	15	30
High	10	10	20	40
Marginal(X)	20	30	50	/

X: height
Y: weight

If we look at the first column, we find the height values fixed by weight. This conditioning is called **Conditional Distribution** and is represented as:

$$Y \mid X = x_i \qquad\qquad \sum_i Y \mid X = x_i = \text{Marginal of Y}$$

Various definitions of Statistical Independence:

- If through a conditional distribution the distributions of each section do not change then we are in a case of independence.
- If the distribution of each column and row has equal relative frequency, then the two variables are independent.
- n_{ij} is an element in (X, Y) distribution

X/Y	y_1	$\cdots$	y_i	$\cdots$	y_n	TOT
x_1						$\vdots$
$\vdots$						$\vdots$
x_i			n_{ij}			$n_{i\cdot}$
$\vdots$						$\vdots$
x_n						$\vdots$
TOT	$\cdots$	$\cdots$	$n_{\cdot j}$	$\cdots$	$\cdots$	$\cdots n$

> We can also see independence in the following way:
>
> $$\frac{n_{ij}}{n_{\cdot j}} = \frac{n_{i\cdot}}{n}$$
>
> Where:
>
> $$\frac{n_{ij}}{n_{\cdot j}} = f_{(X|Y=y_i)}$$
>
> $$\frac{n_{i\cdot}}{n} = f_{(X=x_i)}$$

We can therefore conclude that if this relationship is true:

$$f_{(X|Y=y_i)} = f_{(X=x_i)}$$

then X and Y are independent.

This also takes up the concept of **Mathematical Independence**:
two events A and B are independent if the joint probability of the two events is equal to the product of the probabilities of the individual events:

$$P(A \cap B) = P(A) \cdot P(B)$$

For random variables X and Y, their independence implies that the joint distribution $P(X, Y)$ is the product of the marginal distributions $P(X)$ and $P(Y)$.

Donsker Distribution

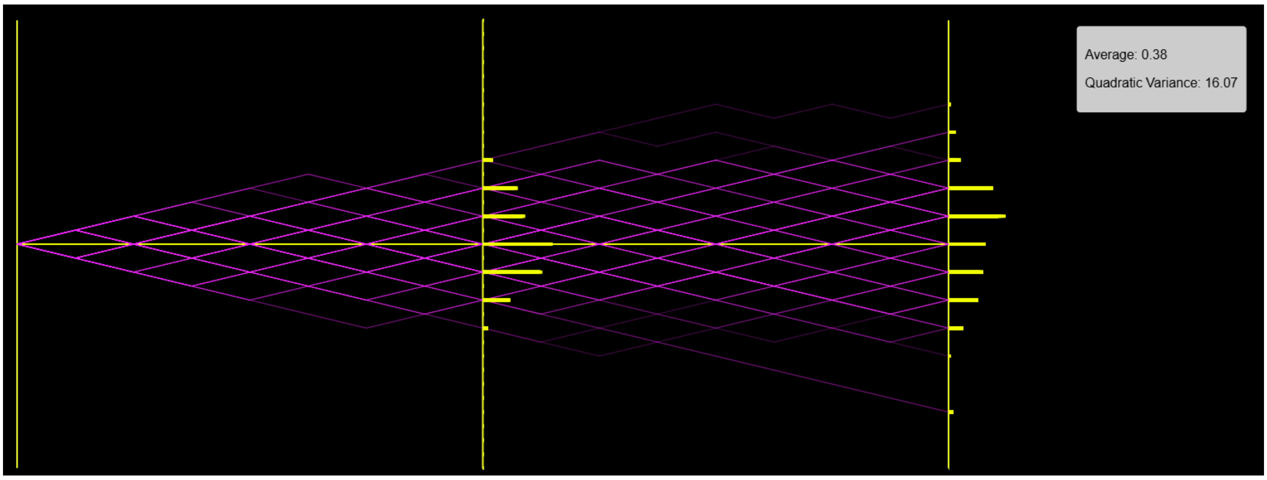
The **Donsker Distribution**, named after Monroe Donsker, is closely related to the **invariance principle** or **Donsker's theorem**. This theorem is fundamental in probability theory and states that **a properly normalized sum of random variables** (like those in a random walk) **converges to a Brownian motion** in the limit as the number of steps goes to infinity.

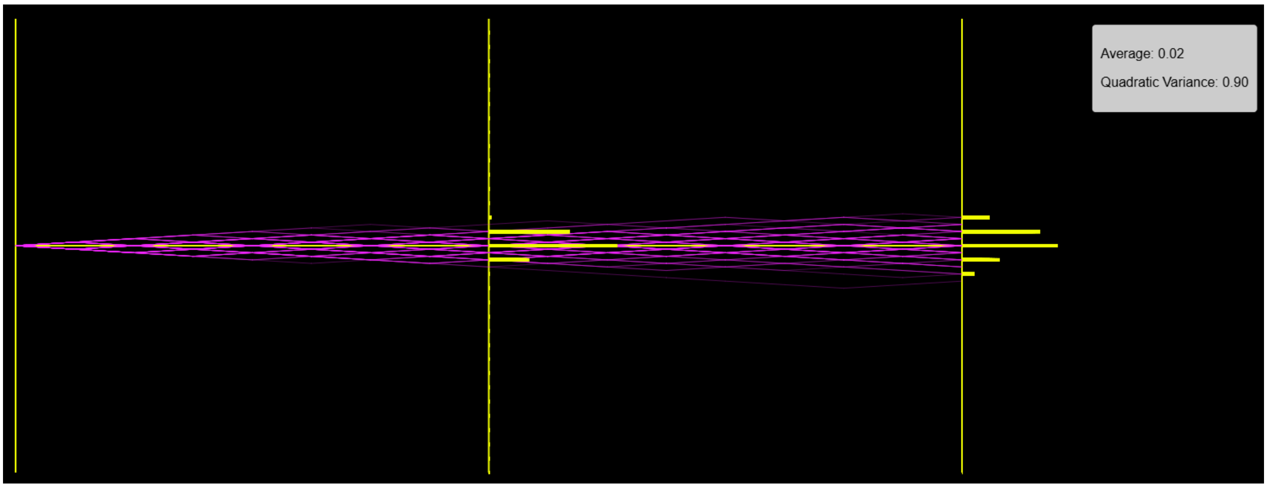
In a **Random Walk** Donsker's theorem provides a bridge between discrete random walks and continuous processes. It states that the path of a random walk, when rescaled appropriately, converges to a Brownian motion as the number of steps increases.

Statistical Graphs and Visualization: Donsker's theorem is valuable in visualizations because it allows analysts to interpret the **discrete paths of a random walk as approximations of a continuous Brownian motion**.

Graphical Differences

In this section we can see graphical differences between a classic random walk and one in which reduction is applied:





As we can see, the variance in the second case is smaller than the variance in the first graph. This is due to the scaling effect predicted by the Donsker's distribution.