Collected Papers 2021 - A Special Issue on Statistics and Probability

Mikko I. Malinen

5th February, 2022

Chapter 1

Means Of Several Random Variables

1.1 Introduction

We consider a line segment of length 1, where we put several points at random locations. We examine what are the mean positions of the points starting from the leftmost point and proceeding to the middle points and to the rightmost point. This setting gives also the mean values of several identically distributed random variables starting from random variable with the lowest value and proceeding to higher value random variables.

1.2 Two Points On a Line Segment

Let's consider first two randomly located points on a line segment of length 1, see Figure 1.1. We will derive the mean locations of the points. Probability

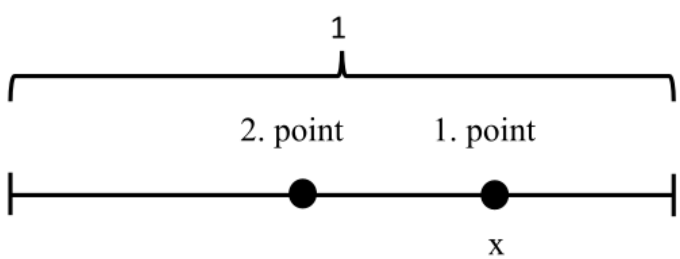

Figure 1.1: Two random points on a line segment of length 1.

that the second point is on the left side of the first point $p_1 = x$. The mean location for the second point, if it is on the left side of the first point $x_1 = x/2$. Probability that the second point if it is on the right side of the first point is $p_2 = 1 - x$. The mean location for the second point, if it is on the right side of the first point is $x_2 = x + (1 - x)/2$. Now, the mean position from the left of the leftmost point is

$$x_{1,mean} = p_1 \cdot x_1 + p_2 \cdot x_2 \tag{1.1}$$
$$= x \cdot \frac{x}{2} + (1 - x) \cdot x \tag{1.2}$$
$$= \frac{x^2}{2} + x - x^2 \tag{1.3}$$
$$= -\frac{x^2}{2} + x. \tag{1.4}$$

We now introduce the following lemma:

Lemma 1. The mean $\bar{f}(s)$ of a function $f(s)$ on an interval is

$$\bar{f}(s) = \frac{1}{b - a} \int_a^b f(s) ds,$$

where a is the starting point of the interval and b is the endpoint of the interval. We now use the Lemma 1 in calculating the mean position of the leftmost point by calculating the mean when x goes from 0 to 1:

$$\bar{f}(x) = \frac{1}{1 - 0} \int_0^1 -\frac{x^2}{2} + x dx \tag{1.5}$$
$$= /_0^1 - \frac{1}{6} x^3 + \frac{1}{2} x^2 \tag{1.6}$$
$$= -\frac{1}{6} + \frac{1}{2} \tag{1.7}$$
$$= \frac{1}{3}. \tag{1.8}$$

This is our result. The mean of the leftmost point is 1/3. By symmetry, the mean of the rightmost point is 2/3.

1.3 Three Points On a Line Segment

In the case of three points, the mean of the middle point is 0.5 because the points have to be symmetrically on the line segment. Moreover, because of symmetry, there has to be one point on the left side of the middle and one point on the right side of the middle. The mean of the leftmost point is in the middle of the left end and 0.5, that is, at 0.25. The rightmost point respectively at point 0.75. To summarize, the means of the points are in 0.25, 0.5, and 0,75.

1.4 n Points On a Line Segment

If all other points are kept fixed and one point can be moved, the mean of it is at the mean of neighboring points or, if there is no neighbor on one side, at the mean of end and neighboring point. When all points are moved like this one point at a time, and this process is repeated several times, they will settle uniformly on the line segment. Thus the mean locations of n points will be

$$\frac{1}{n+1}, \frac{2}{n+1}, ..., \frac{n}{n+1}. \tag{1.9}$$

1.5 Analogy to Random Variables

There is a correspondence between the points, which we have been talking so far, and uniformly and identically distributed random variables. Let there be as many random variables as there are points. Then the random variable having the minimum value corresponds to the leftmost point and the random variable having the second smallest value corresponds to the second point from the left. The means of the values of the random variables have equal spacing as have the points. But what if the distribution of the random variables is something different from uniform distribution? Then the mean values of the random variables are such that there is equal amount of probability mass between two neighboring random variables. To get those values, proceed similarly as in sampling from arbitrary distribution. Take n values from an uniform distribution, with equal spacing, and transform those to the values from the arbitrary distribution as you would do when sampling from that distribution.

1.6 Summary

When placing n random point on a line segment, the means of the point locations will be set uniformly, with equal spacing, to the line segment.

When sampling values for uniformly and identically distributed random variables, the spacing between two variables neighboring by value is equal in every such pair.

When sampling values for arbitrarily and identically distributed random variables, the spacing by probability mass between two variables neigboring by value is equal in every such pair.