

The following experiment is conducted. We randomly select tree in a forest, measure its height and record it as  $x_1$ . Then we again select random tree in the same forest, independently on the first choice, measure its height and record it as  $x_2$ , and so on.

1. Can we consider  $(x_1, \dots, x_n)$  as an i.i.d. (independent and identically distributed) sample from some random variable? I.e. can we assume that these values are obtained as independent realizations of some random variable  $X$ ? 2. Now we sort values  $(x_1, \dots, x_n)$  in ascending order and thus obtain new sequence of numbers denoted by  $(y_1, \dots, y_n)$ . Can we say that  $(y_1, \dots, y_n)$  is an i.i.d. sample from some random variable?

Hint: it can be helpful to solve the following problem before answering question 2. Let  $X_1$  and  $X_2$  be two independent Bernoulli random variables with  $p=1/2$  (i.e. they take values 0 and 1 with equal probabilities.) Consider new random variables  $Y_1 = \min(X_1, X_2)$  and  $Y_2 = \max(X_1, X_2)$ . Is it true that they  $Y_1$  and  $Y_2$  are independent random variables? Use definition of independent random variables.

1. Yes, one can consider  $(x_1, \dots, x_n)$  as an i.i.d. sample from some random variable  $X$ . Let us prove it using the definition of the arbitrary i.i.d sample. Firstly, each tree is selected randomly and independently, meaning that the selection of the next tree is not influenced in any way by the previous tree. Hence, the measurements are independent by definition. Secondly, one can assume that the heights of the trees sample follow the same distribution from the population. Hence, using the law of large numbers, one can assume that the measurements are identically distributed when the sample size is big enough. ■
2. No, such sample is not i.i.d. by definition. Sorting any sample would mean that there's a clear dependency. If we have sorted the heights and obtained the sample  $(y_1, \dots, y_n)$  in the ascending order, there's a dependency such that for each  $i$  it follows that  $y_i < y_{i+1}$ . Likewise, the distribution of the sample will not be the same as the distribution of the population. Let us now prove that these events are not independent using the Bernoulli random variables. From the experiment description it follows that if:

$$x_1 = 0, x_2 = 0 \implies y_1 = 0, y_2 = 0$$

$$x_1 = 0, x_2 = 1 \implies y_1 = 0, y_2 = 1$$

$$x_1 = 1, x_2 = 0 \implies y_1 = 0, y_2 = 1$$

$$x_1 = 1, x_2 = 1 \implies y_1 = 1, y_2 = 1$$

$\implies$  there's a clear dependence of outcome based on the prior (obtained) information

$\therefore$  this sample is not i.i.d. ■