Probability Part II - Repeated Independent Trials

Yan Tao

Advanced 1 - Fall 2023

1 Independent Random Variables

This is a direct continuation of last week's worksheet, so all relevant definitions and examples should be found there.

In this worksheet, we'll continue to study variances. When trying to measure the expected value of a random variable, it makes intuitive sense to measure the random variable again and again, and take the average values. For this approach to work, the variance in our measurements should be low. Before we show this, however, we'll have to introduce some necessary definitions.

Definition 1 *Let $(\Omega, \mathbb{P})$ be a finite probability space. Two events A and B are **independent** if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

(that is, if the probability of A and B equals the product of the probabilities of A and B)

Intuitively, two events are independent precisely when one of them doesn't affect the other one. While this doesn't always match the mathematical calculation (so it is no substitute for it), it is good to keep in mind.

Problem 1 *Consider an experiment where we roll a fair six-sided dice once. For each pair of events below, decide intuitively, and then mathematically, whether they are independent.*

- $A = \{\textit{odd number}\}, B = \{\textit{number divisible by } 3\}$

  Solution: These are independent.

- $A = \{6\}, B = \{\textit{not } 6\}$

  Solution: These are not independent.

- $A = \{1, 4, 6\}, B = \{2, 6\}$

  Solution: These are independent (not intuitively, probably).

- $A = \{\textit{even number}\}, B = \{\textit{prime number}\}$

  Solution: These are not independent.

- $A = \emptyset, B$ *is any event*

  Solution: These are independent.

Recall that a random variable is a function $X : \Omega \to \mathbb{R}$. Since each random variable has corresponding events $\{X = x\}$, we can determine whether those events are independent.

Definition 2 *Let $(\Omega, \mathbb{P})$ be a finite probability space. Two random variables X and Y on it are **independent** if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all $x, y \in \mathbb{R}$.*

Problem 2 *Consider an experiment where we flip one fair coin. Let X be the number of heads, and Y the number of tails. Show whether X and Y are independent.*

Solution: They are not independent - for instance,

$$0 = \mathbb{P}(\{X = 0\} \cap \{Y = 0\}) \neq \mathbb{P}(\{X = 0\})\mathbb{P}(\{Y = 0\}) = \frac{1}{4}$$

Problem 3 *Consider an experiment where we flip two fair coins. Let X be the number of heads on the first flip, and Y the number of heads on the second flip. Show whether X and Y are independent.*

Solution: They are intuitively independent. To show this mathematically, consider

$$\{X = 0\} = \{TH, TT\} \text{ and } \{X = 1\} = \{HH, HT\}$$
$$\{Y = 0\} = \{HT, TT\} \text{ and } \{X = 1\} = \{HH, TH\}$$

(The events $\{X = x\}$, $\{Y = y\}$ for $x, y \neq 0, 1$ are empty.) We can check the required independences.

Problem 4 *Consider an experiment where we flip two fair coins. Let X be the number of heads on the first flip, and Y always equal 1. Show whether X and Y are independent.*

Solution: They are independent. $\{Y = 1\} = \Omega$ (and all other Y-events are empty), so we can check the required independences.

Problem 5 *Consider an experiment where we roll a fair die twice. Let X_1 be the number on the first roll, and X_2 be the number on the second roll. Explain (intuitively) why X_1 and X_2 are independent. (Bonus) Prove it.*

Solution: The first roll does not affect the second roll.

Problem 6 *In general, it seems that the first and second (and third, etc.) trials of an experiment are independent. But can you think of any real-life examples where this might not be the case?*

Solution: A chemistry experiment where the beakers weren't clean, a physics experiment where a spring wore out, and many other examples.

2 Repeated Trials and Sample Mean

From now on, we'll assume successive trials of the same experiment to be independent. We'll also need a stronger property for our random variables.

Definition 3 *Let $(\Omega, \mathbb{P})$ be a finite probability space. Two random variables X_1 and X_2 on it are **identically distributed** if they have the same PMF; that is, if $\mathbb{P}(\{X_1 = x\}) = \mathbb{P}(\{X_2 = x\})$ for all real numbers x.*

If $X_1, ..., X_n$ are independent and identically distributed, we will say they are *iid* for short. We now give some examples that suggest that repeated trials of the same experiment will not only be independent, but also identically distributed (so we'll assume this too).

Problem 7 *Consider an experiment where we roll a fair die n times, and let X_1 be the first roll, X_2 the second roll, and so on. Show that $X_1, ..., X_n$ are iid.*

Solution: Independence follows from Problem 5. By definition each X_i is identically distributed to one roll of the die, so they are identically distributed to each other.

Problem 8 *Consider the same setup as above but with an unfair die. Assuming we are rolling the same unfair die each time, show that $X_1, ..., X_n$ are still iid.*

Solution: The logic from Problem 5 still shows independence, and the logic from Problem 6 still shows identical distribution.

Last week, we estimated the expected value of a fair six-sided die roll by rolling a fair six-sided die repeatedly. What we actually measured was the average, or *mean* of our results. We'll now show that this mean is a good way to measure the expected value.

Definition 4 *Let $X_1, ..., X_n$ be iid random variables. The **sample mean** $\overline{X}$ is the random variable*

$$\overline{X} := \frac{X_1 + ... + X_n}{n}$$

Problem 9 *Using a computer dice roller (preferably), roll 200 fair six-sided dice, in groups of 10. Write down the 20 sample means (one from each group of 10).*

Problem 10 *Based on this data, what do you think is the expected value of the sample mean?*

Let's prove (or disprove) this conjecture.

Problem 11 *Let $X_1, ..., X_n$ be iid random variables. Show that $\mathbb{E}[\overline{X}] = \mathbb{E}[X_1] = ... = \mathbb{E}[X_n]$ (Hint: Use Problem 12 from last week.)*

Solution: Because $X_1, ..., X_n$ are iid, they have the same expected value, and the first equality follows by linearity of expectation.

3 Variance of the Sample Mean

In order to use the sample mean as a good measurement of the expected value, its variance (which measures how far off it will typically be) should be close to zero. As we did last time, we will do another experiment to show the variance.

Problem 12 *Using a computer dice roller, roll 400 fair six-sided dice, in groups of 20. Write down the 20 sample means (one from each group of 20). If you want, you can also repeat this in groups of 30 as well.*

Problem 13 *Based on this data (and the data from Problem 9), hypothesize a relationship between the number of trials n and the variance of the sample mean. (Hint: It may help to draw some sort of graph.)*

In order to prove (or disprove) your conjecture, we'll need some useful facts about variance.

Problem 14 *(a) Show that for any (real) constant c,* $Var(cX) = c^2 Var(X)$

Solution: By linearity of expectation

$$Var(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = \mathbb{E}[(cX - c\mathbb{E}[X])^2] = c^2\mathbb{E}[(X - \mathbb{E}[X])^2]$$

(b) (Bonus) Show that if X *and* Y *are independent, then* $Var(X+Y) = Var(X) + Var(Y)$.

Problem 15 *Let $X_1, ..., X_n$ be iid random variables. Write down a formula for $Var(\overline{X})$, in terms of $Var(X_1)$ and the number of trials n.*

Solution: By Problem 14,

$$Var(\overline{X}) = Var\left(\frac{X_1 + ... + X_n}{n}\right) = \frac{1}{n^2}Var(X_1 + ... + X_n) = \frac{n}{n^2}Var(X_1) = \frac{Var(X_1)}{n}$$

Problem 16 *(Law of Large Numbers) Let $X_1, ..., X_n$ be iid random variables. Show that as n becomes larger, $Var(\overline{X})$ becomes smaller and approaches zero.*

Solution: This is intuitively clear from the formula from Problem 15.

Problem 17 *(Bonus) Last week, we discussed Buffon's needle experiment as a way to calculate π. Now, let's figure out how many times we should drop the needle in order to get an accurate measurement. As a reminder, we determined that when we drop a needle of length 1 onto a sequence of parallel lines distance 1 apart, it touches a line with probability $2/\pi$.*

- *Let X be the random variable that counts the number of times the needle intersects a line. Find $Var(X)$.*
- *Suppose we want to approximate π to the nearest hundredth digit. What is the acceptable margin of error for $\overline{X}$, which measures $2/\pi$?*
- *Typically when approximating, our acceptable margin of error should be at least 2* ***standard deviations****. The standard deviation $\sigma(X)$ is the square root of the variance $Var(X)$. How many trials n do we need for $\sigma(\overline{X})$ to be small enough to make the necessary approximation? Do you think this is a particularly good way to calculate π?*