CS 3430: S24: Lecture 22
Reflections on Randomness: Part 02: Randomness of Numerical Sequences

Vladimir Kulyukin
Department of Computer Science
Utah State University

Introduction

In the previous lecture, we discussed a fundamental reason for us to study randomness, which is the nature of the universe.

Is the universe random? If the universe is random, it is unpredictable. If the universe is partially random, we can hope to predict the behavior of some natural phenomena.

When a scientist investigates a phenomenon, the scientist can start with the question: Is this phenomenon random? There can be three possible answers to this question: **yes**, **no**, **maybe**.

If a phenomenon is random, then the modern scientific method outlined by Francis Bacon in his 17-th century book "New Organon" (i.e., repeatable and replicable experiment) should be left alone, or the phenomenon should be studied with other methods, because randomness cannot be repeated or replicated. It can also be the case, of course, that our methods of measurements, i.e., our tools or observations, are not designed well.

Random Number Generators

In scientific computing, randomness is investigated with random number generators, or, to be more exact, with pseudorandom number generators (PRNGs), because we cannot really claim that numerical sequences we generate on deterministic digital computers are genuinely random.

How do we compare generators? We must have some tests of randomness. If the sequences produced by one generator G_1 are "more random" than the sequences output by another generator G_2, according to some randomness test T that we have agreed on, then we should, if our hardware allows it, use G_1.

In this lecture, we will study one such test – the χ^2 test (χ^2 is pronounced as *kai square*). The test's name is transliterated into English as *chi-square*, because the Greek letter χ (pronounced as *h*) is transliterated as *chi*. The symbol χ^2 takes its origin from the work by Karl Pearson, a 19-th century mathematician, who developed this test.

Random Variables

In probability theory and statistics, experiments are studied with the so-called *random variables*. Typically, random variables are referenced with capital letters, e.g., X, Y, Z.

The term *random variable* is a misnomer, however, because random variables are not variables, but *functions* that map some part of the observable universe to the universe of measurements. The latter is typically a subset of real numbers.

E.g., if our experiment consists of 10 flips of a fair coin, then we can have a random variable, say Y, to count the number of heads and the number of tails in each run of the experiment. If Y_1 is the actual (i.e., observed) number of heads and Y_2 is the actual number of tails, then in each experiment $0 \leq Y_1 \leq 10$ and $0 \leq Y_2 \leq 10$.

Random Variables

If in our coin flipping experiment we use the random variable Y to assign 1 to heads and 2 to tails, each run of the experiment generates a sequence of 10 numbers, e.g., 1222121121. Thus, the experiment can be viewed as a random number generator.

What are we investigating in this experiment? The same experiment can investigate multiple aspects of the same phonemenon. A straightforward answer is that we study the fairness of a specific coin. Our intuition tells us that Y_1 and Y_2 in each experiment should be close to each other or equal. If that is the case, in most runs, then we can conclude that the coin is *fair*. If not, then the coin is *biased*, e.g., there are always more heads than tails. Another, equally valid, answer is that we study the fairness of a specific coin flipper. I may be, knowingly or unknowingly, flipping a coin in such a way that heads are much more frequently observed than tails.

Observations and Categories

Suppose we have an experiment, e.g., 10 coin flips, 10 throws of a 6-faced die, 10 digits in the mantissa of π, pedestrians crossing a given street at a given time of day, etc., and we are reasonably sure that each trial within the experiment is independent of the other trials.

If we use a random variable to describe the outcome of an experiment, then we must come up with the values of the random variable. For coin flips, these values can be the set of 1 and 2, i.e., $\{1, 2\}$; for die throws, they can be $\{1, 2, 3, 4, 5, 6\}$; for π mantissa digits, they can be $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$; For the pedestrian crossings, they can be between 0 and some postive integer k. Each such value is a *category*. Thus, for coin flips, we have 2 categories; for die throws – 6, for π digits – 10, and for pedestrian crossings – k.

Note that we used natural numbers to label categories for the sake of convenience. This choice is frequent, but arbitrary. E.g., we might have just as well used the symbol H for an observed head and the symbol T for an observed tail. Then the outcome 1222121121 would look like $HTTTHTHHTH$.

Examples

Suppose I take a coin, flip it 10 times, and record each observation, then the outcome can be

$$HTHTHHHTHT$$

or, if we assign 1 to H and 2 to T,

$$1212111212.$$

Suppose I throw a 6-faced die 10 times, and record the outcome of each throw, i.e., the number of dots on a face, as a positive integer, then the outcome can be

$$1561123243.$$

Suppose I use the Chudnovsky algorithm to generate the first 10 digits of the π mantissa, then they are

$$1415926535.$$

Mathematical Expectation

We would like to develop some theory that allows us to make some theoretical predictions (aka mathematical expectations).

Let p_s be the probability of observing the category s as the outcome of an experiment. E.g., if we flip a fair coin, then

$$p_1 = p_2 = \frac{1}{2} = 0.5.$$

If we throw a fair 6-face die, then

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}.$$

If we generate the digits of the π mantissa, then

$$p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9 = \frac{1}{10}.$$

Mathematical Expectation

The general formula of the mathematical expectation of Y_i is

$$E(Y_i) = np_i,$$

where n is the number of individual trials in our experiment, e.g., coin flips, die throws, π digits, pedestrian crossings, etc. and p_s is the probability of seeing the category s at each observation.

Many tests of randomness, such as χ^2, assume that individual trials are *independent* of each other, i.e., the category of the current trial does not depend on the categories of any previous (or subsequent) trials. E.g., the current digit of the π mantissa does not depend on any previous digit.

What would be an example of *dependence*? If, when producing the next digit, we are not allowed to use any digits previously produced, our choice at each trial is dependent on the previously generated digits, and we will run out of digits after 10 trials.

Examples

For 10 coin flips, our expectations are:

$$E(Y_1) = E(Y_2) = 10\frac{1}{2} = 5.$$

In other words, we *expect* to see 5 heads and 5 tails. Said another way, we *predict* a priori, i.e., ahead of or prior to our edxperiment, that we will observe 5 heads and 5 tails.

For 10 die throws, our expectations are:

$$E(Y_1) = E(Y_2) = E(Y_1) = E(Y_3) =$$

$$E(Y_4) = E(Y_5) = E(Y_6) = 10\frac{1}{6} =$$

$$\frac{5}{3} = 1.66666...$$

Thus, we expect to see each face ≈ 1.67 times in 10 die throws. Note here that an expected count of a category is not always a whole number.

Examples

For the 10 π digits generated with Chudnovsky, our expectations are:

$$E(Y_0) = E(Y_1) = E(Y_2) = E(Y_3) = E(Y_4) =$$

$$E(Y_5) = E(Y_6) = E(Y_7) = E(Y_8) = E(Y_9) =$$

$$10\frac{1}{10} = 1.$$

Thus, we expect to see each digit once when we generate the first 10 digits of the π mantissa.

Expectation vs. Reality

The essence of the χ^2 test is the comparison of our expected counts for each category (i.e., the mathematical expectations) with the actually observed counts during an experiment. Thus, the first step in χ^2 is to compute the actual and expected counts for each category. E.g., from the first 10 digits of the π mantissa generated with Chudnovsky, i.e., 1415926535, we compile the following table.

s	Y_s	$E(Y_s)$
0	0	1
1	2	1
2	1	1
3	1	1
4	1	1
5	3	1
6	1	1
7	0	1
8	0	1
9	1	1

The $s = 1$ row tells us that we observed 2 occurences of 1 in 1415926535, but expected to see only 1 occurence.

χ^2 Test

We have an experiment with $k > 0$ categories, which means that we have $k - 1$ degrees of freedom (df). E.g., in coin flips, we have 2 categories and 1 degree of freedom; in 6-face die throws, we have 6 categories and 5 degrees of freedom; in the π mantissa generation, we have 10 categories and 9 degrees of freedom.

We perform the experiment and record the outcomes of individual trials and compute the expectations for each category. Then we compute the so called χ^2 statistic V for the outcome as

$$V = \sum_{s=1}^{k} \frac{(Y_s - np_s)^2}{np_s},$$

where n is the number of individual trials in each experiment, e.g., $n = 10$ in our coin flipping experiment. If we use 0 to label a category (and we do for the π digits on the previous slide), then s ranges from 0 to $k - 1$. Also, note that V is a positive real number.

Another Look at the χ^2 Statistic Formula

If we have k categories, then $Y_1 + Y_2 + \ldots + Y_k = n$ and $p_1 + p_2 + \ldots + p_k = 1$,

$$V = \sum_{s=1}^{k} \frac{(Y_s - np_s)^2}{np_s} = \sum_{s=1}^{k} \frac{{Y_s}^2 - 2Y_s np_s + n^2 {p_s}^2}{np_s} =$$

$$\frac{1}{n}\left(\sum_{s=1}^{k} \frac{{Y_s}^2}{p_s} - 2n \sum_{s=1}^{k} Y_s + n^2 \sum_{s=1}^{k} p_s \right) =$$

$$\frac{1}{n}\left(\sum_{k=1}^{k} \frac{{Y_s}^2}{p_s} - 2n^2 + n^2 \right) = \frac{1}{n} \sum_{s=1}^{k} \frac{{Y_s}^2}{p_s} - n.$$

A good mneumonic is

$$V = \sum_{s=1}^{k} \frac{(Observed - Expected)^2}{Expected}.$$

χ^2 Test

Once we have V, we go to the χ^2 table. To be more exact, we used to go to it, now we will likely use some software package to do the computation. An image below shows part of the χ^2 distribution table.

	$p = .01$	$p = .05$	$p = .25$	$p = .50$	$p = .75$	$p = .95$	$p = .99$
$\nu = 1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$\nu = 2$	0.02010	0.1026	0.5753	1.386	2.773	5.991	9.210
$\nu = 3$	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$\nu = 4$	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
$\nu = 5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$\nu = 6$	0.8720	1.635	3.455	5.348	7.841	12.59	16.81
$\nu = 7$	1.239	2.167	4.255	6.346	9.037	14.07	18.48
$\nu = 8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09
$\nu = 9$	2.088	3.325	5.899	8.343	11.39	16.92	21.67
$\nu = 10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$\nu = 11$	3.053	4.575	7.584	10.34	13.70	19.68	24.73
$\nu = 12$	3.571	5.226	8.438	11.34	14.84	21.03	26.22

The rows of this table are the degrees of freedom, i.e., $\nu = k - 1$, where $k > 1$ is the number of categories and the columns are specific probabilities, the so called p-values.

Example

	$p=.01$	$p=.05$	$p=.25$	$p=.50$	$p=.75$	$p=.95$	$p=.99$
$\nu=1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$\nu=2$	0.02010	0.1026	0.5753	1.386	2.773	5.991	9.210
$\nu=3$	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$\nu=4$	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
$\nu=5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$\nu=6$	0.8720	1.635	3.455	5.348	7.841	12.59	16.81
$\nu=7$	1.239	2.167	4.255	6.346	9.037	14.07	18.48
$\nu=8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09
$\nu=9$	2.088	3.325	5.899	8.343	11.39	16.92	21.67
$\nu=10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$\nu=11$	3.053	4.575	7.584	10.34	13.70	19.68	24.73
$\nu=12$	3.571	5.226	8.438	11.34	14.84	21.03	26.22

Let us say that $k=9$, then our degrees of freedom $\nu = k-1 = 8$. Suppose that $V = 20.09$. Since I deliberately took a real number that actually occurs in the table (so that we did not have to round), we find 20.09 in the cell of the row $\nu = 8$ and the column $p = 0.99$. In other words, the p-value corresponding to our V statistic is $p = 0.99$. What does it mean? It means that one would expect to see $V \leq 20.09$ as the χ^2 statistic with 8 degrees of freedom around 99% of the time and to see $V > 20.09$ to occur about 1% of the time.

Dr. Knuth's Criteria of Sufficient Randomness

If the answer at the end of the previous slide left you with a sense of uneasiness, you are not alone. Interpreting statistical results is difficult. By the way, there is a great book about this: Darrell Huff. *How to Lie with Statistics*. Norton, 1954.

Fortunately, there are some χ^2 interpretation criteria that have been emperically proved to work. E.g., below are the criteria given by Dr. Donald Knuth in Chapter 2 of the 2nd volume of his famous opus *The Art of Computer Programming*.

1. If the *p*-value of V is less than 1% or greater than 99%, i.e., $p < 0.01$ or $p > 0.99$ or $p \in [0, 0.01)$ or $p \in (0.99, 1.00]$, then the generated sequence is classified as *insufficiently random*.
2. If the *p*-value of V is in $[0.01, 0.05)$ or in $(0.95, 0.99]$, the sequence is *suspect*, i.e., highly likely not random.
3. If the *p*-value of V is in $[0.05, 0.1)$ or in $(0.9, 0.95]$, then the sequence is almost suspect, i.e., likely not random.

Given the above recommendations, a "safe" range for randomness appears to be $[0.1, 0.9]$. If we want to be more sure, we can narrow it to, say, $[0.2, 0.8]$ or even to $[0.3, 0.7]$.

Example

Let us go back to the π digit sequence 1415926535 generated with Chudnovsky and use the table on Slide 12 to compute its χ^2 statistic V.

$$V = \frac{(0-1)^2}{1} + \frac{(2-1)^2}{1} + \frac{(1-1)^2}{1} + \frac{(1-1)^2}{1} + \frac{(1-1)^2}{1} +$$

$$\frac{(3-1)^2}{1} + \frac{(1-1)^2}{1} + \frac{(0-1)^2}{1} + \frac{(0-1)^2}{1} + \frac{(1-1)^2}{1} = 8.$$

If we use the second version of the formula, then

$$V = \frac{1}{n}\sum_{s=0}^{9} \frac{Y_s}{p_s} - n =$$

$$\frac{1}{10}\left(\frac{0^2}{1/10} + \frac{2^2}{1/10} + \frac{1^2}{1/10} + \frac{1^2}{1/10} + \frac{1^2}{1/10} + \frac{3^2}{1/10} + \frac{1^2}{1/10} + \frac{1^2}{1/10}\right) - 10 =$$

$$\frac{1}{10}(10 \cdot 18) - 10 = 8.$$

χ^2 in Python

After we have the V value, we can use a χ^2 table to estimate the p-value of this V. We can take a shortcut and use `scipy.stats.chisquare()` that computes V and its p-value. It takes two arrays: an array of Y_s (the actually observed counts) and the array of the corresponding $E(Y_s)$ (the expected counts). So, we can use the values in the Y_s column in the table on Slide 12 to construct the first array and the values of the $E(Y_s)$ column to construct the second array.

```
import scipy
>>> scipy.stats.chisquare([0, 2, 1, 1, 1, 3, 1, 0, 0, 1],
                          [1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
Power_divergenceResult(statistic=8.0, pvalue=0.5341462169096916)
```

Since the p-value falls into the range recommended by Dr. Knuth, i.e, [0.1, 0.9], we consider this sequence as sufficiently random.

Another Example

Let us consider the sequence 1561123243, which is an outcome of the experiment of throwing a 6-face die 10 times. Let us compile the table for this sequence.

s	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Y_s	3	2	2	1	1	1
p_s	1/6	1/6	1/6	1/6	1/6	1/6

Then

$$V = \sum_{s=1}^{6} \frac{(Y_s - np_s)^2}{np_s} = \frac{(3-10/6)^2}{10/6} + \frac{(2-10/6)^2}{10/6} + \frac{(2-10/6)^2}{10/6} +$$

$$\frac{(1-10/6)^2}{10/6} + \frac{(1-10/6)^2}{10/6} + \frac{(1-10/6)^2}{10/6} \approx 2.0,$$

if we take $10/6 \approx 1.67$.

Another Example

We can use `scipy.stats.chisquare` to verify the χ^2 V statistic that we manually computed for the sequence 1561123243 generated by the experiment of throwing a 6-face die 10 times.

```
>>> scipy.stats.chisquare([3, 2, 2, 1, 1, 1], [10/6, 10/6, 10/6, 10/6, 10/6, 10/6])
Power_divergenceResult(statistic=2.0, pvalue=0.8491450360846096)
```

Thus, if we use Dr. Knuth's recommendations, the *p*-value $0.8491450360846096 \in [0.1, 0.9]$ gives us some assurance that 1561123243 is sufficiently random.

References

1. Knuth, D. E. *The Art of Computer Programming*, vol. 2. 2nd Ed. Reading, MA: Addison-Wesley, 1981.

2. F. Bacon. *New Organon*. Dodo Press.