Math Colloquium
Five Lectures on Probability Theory
Part 2: The Law of Large Numbers

Robert Niedzialomski, rniedzia@utm.edu

March 3rd, 2021

Probability - Intuition

Probability Theory = Mathematical framework for modeling/studying non-deterministic behavior where a source of randomness is introduced (this means that more than one outcome is possible)
The space of all possible outcomes is called the **sample space**. A set of outcomes is called an **event** and the source of randomness is called a **random variable**

Discrete Probability

A **discrete probability** space consists of a finite (or countable) set Ω of **outcomes** ω together with a set of non-negative real numbers p_ω assigned to each ω; p_ω is called the **probability of the outcome** ω. We require $\sum_{\omega\in\Omega} p_\omega = 1$.
An **event** is a set of outcomes, i.e., a subset $A \subset \Omega$. The probability of an event A is

$$P(A) = \sum_{\omega\in A} p_\omega.$$

A **random variable** is a function X mapping the set Ω to the set of real numbers. We write $X\colon \Omega \to \mathbb{R}$.
We note that the following Kolmogorov axioms of probability hold true:

- $P(\emptyset) = 0$
- if $A_1, A_2, \ldots$ are disjoint events, then $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

An Example of Rolling a Die Twice

Example (Rolling a Die Twice)

Suppose we roll a fair die twice and we want to model the probability of the sum of the numbers we roll. The sample space to $\Omega = \{(i,j)\colon i,j = 1,2,3,4,5,6\}$ with probability of each outcome $p_{ij} = 1/36$.

Let the random variable X represent the number after the first roll and let Y be the random variable that represents the number after the second roll. Hence

$$X(i,j) = i \quad \text{and} \quad Y(i,j) = j.$$

Our goal is to study the random variable $X + Y$. We compute

$$P(X' + Y' = 2) = P(\{(1,1)\}) = 1/36$$
$$P(X' + Y' = 3) = P(\{(1,2),(2,1)\} = 2 \cdot (1/36) = 1/18.$$

Example (Rolling a Die Twice Continued)

We continue our compuation of the probabilities for the random variable $X' + Y'$ and obtain

k	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

We have obtained a new probability space

$$\Omega_X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

with probability p_k of each outcome given in the table above. This probability is called the **distribution** of the random variable $X + Y$. The **expected value**, denoted by $E[X + Y]$ of this random variable is the weighted average (mean) of the locations k with weights p_k, i.e.,

$$E[X + Y] = \sum_{k=2}^{12} p_k \cdot k = 7.$$

Distribution and Expectation

Let $(\Omega, (p_\omega))$ be a discrete probability space and let X be a random variable on Ω. The **probability distribution** on X is the discrete probability

$$\Omega_X = \text{ the set of values of } X = \{X(\omega)\colon \omega \in \Omega\}$$

with probability of an outcome k given by

$$p_k = P(X = k) = P(\{\omega \in \Omega\colon X(\omega) = k\}).$$

The **expectation** $E[X]$ of X, called also **mean**, is given by

$$E[X] = \sum_{k \in \Omega_X} p_k \cdot k.$$

Remark: The formula for expectation makes sense for a probability defined on a real line without reference to a random variable.

Theorem

The expectation of a random variable X can be computed according to the formula

$$E[X] = \sum_{\omega \in \Omega} p_\omega X(\omega).$$

Proof.

We first notice that

$$p_k = P(\{\omega \in \Omega \colon X(\omega) = k\}) = \sum_{\{\omega \colon X(\omega) = k\}} p_\omega.$$

Therefore

$$\begin{aligned} E[X] &= \sum_{k \in \Omega_X} p_k \cdot k \\ &= \sum_{k \in \Omega_X} \sum_{\{\omega \colon X(\omega) = k\}} p_\omega X(\omega) = \sum_{\omega \in \Omega} p_\omega X(\omega). \end{aligned}$$

□

The theorem gives us the following properties of expectation:

- For two random variables X and Y we have

$$E[X+Y]=E[X]+E[Y].$$

- For a random variable X and a real number C we have

$$E[cX]=cE[X]$$

We say that a random variable X has **zero mean** if $E[X]=0$.

Bernoulli Distribution

Suppose we flip a biased coin with

$$\text{probability of heads } = p \quad \text{and} \quad \text{probability of tails } = q = 1 - p$$

The probability space is $\Omega = \{H, T\}$ with $p_H = p$ and $p_T = q$.

Let X be the random variable that assigns the value 0 to tails and value 1 to heads. This means that $X(T) = 0$ and $X(H) = 1$.

The probability distribution is $\Omega_X = \{0, 1\}$ with $p_1 = p$ and $p_0 = q$. This distribution is called the **Bernoulli distribution**. Its expectation is

$$E[X] = 0 \cdot p_0 + 1 \cdot p_1 = p.$$

Binomial Distribution

Suppose we flip a biased coin n times. **What is the probability of getting Heads k times?**

The sample space is $\Omega = \{(x_1, x_2, \ldots, x_n) \colon x_j = 0, 1\}$, where 0 represents tails and 1 represents heads. The probability of an outcome $(x_1, x_2, \ldots, x_n)$ is

$$p^{(\text{number of 1's})} \cdot q^{(\text{number of 0's})}.$$

Let S_n be the random variable that represents numbers of of Heads in n flips of the coin. We need to find $P(S_n = k)$. We see that

$$P(S_n = k) = \text{number of outcomes with } k \text{ ones} \cdot p^k q^{n-k}$$

Since the number of outcomes with k ones equals the number of k element subsets of an n element set with is *nchoosek*, we have

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}.$$

The distribution of the random variable S_n is

$$\text{probability space:} \quad \{0, 1, \ldots, n\}$$

$$\text{probability distribution:} \quad p_k = P(S_n = k) = \binom{n}{k} p^k q^{n-k}.$$

To find the expectation we need to compute

$$E[S_n] = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}.$$

This requires the use of the **Binomial Theorem** which says the following. For any real numbers a and b and any positive integer n we have

$$(a+b)^n = \sum_{k=1}^{n} \binom{n}{k} a^k b^{n-k}.$$

Let $X_1, X_2, \ldots, X_n$ be the random variables representing the 1st, 2nd,..., n-th flip of the coin. If we wanted to be precise, we would write

$$X_j(x_1, \ldots, x_n) = x_j$$

Each random variable X_j, where $j = 1, 2, \ldots, n$ has Bernoulli distribution. Hence

$$E[X_1] = E[X_2] = \ldots = E[X_n] = p.$$

Moreover, we see that

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Therefore

$$E[S_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = np.$$

What happens if we keep flipping the coin, record the number of Heads, and take the average by divide by the number of flips? In order words, we want to study

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \lim_{n\to\infty} \frac{S_n}{n}$$

Law of Large Numbers

Law of Averages: Suppose we repeat an experiment independently n times. Then

$$\frac{\#\text{ of successes in } n \text{ trials}}{n} \to P(\text{success})$$

Law of Large Numbers: Let the random variable X_i model the i-th trial of the experiment. This means that $P(X_i = 1) = P(\text{success}) = p$ and $P(X = 0) = P(\text{failure}) = q = 1 - p$. Then the random variables $X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.) with Bernoulli distribution, and

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to E[X_1] = P(\text{success}) = p$$

Theorem (Bernoulli, 1692)

It is the case that S_n/n converges to p as $n \to \infty$ in the sense that for any $\epsilon > 0$

$$P\left(p - \epsilon \leq \frac{S_n}{n} \leq p + \epsilon\right) \to 1 \quad when \quad n \to \infty.$$

Proof: Let $\epsilon > 0$. Then

$$P\left(\frac{S_n}{n} \geq p + \epsilon\right) = \sum_{k \geq n(p+\epsilon)} P(S_n = k) = \sum_{k=\lceil n(p+\epsilon)\rceil}^{n} \binom{n}{k} p^k q^{n-k}$$

Let $\lambda > 0$. Then $0 < \lambda[k - n(p+\epsilon)] = -\lambda n\epsilon + \lambda qk - \lambda p(n-k)$ and

$$P\left(\frac{S_n}{n} \geq p + \epsilon\right) \leq \sum_{k=\lceil n(p+\epsilon)\rceil}^{n} e^{\lambda[k-n(p+\epsilon)]} \binom{n}{k} p^k q^{n-k}$$

$$\leq e^{-\lambda n\epsilon} \sum_{k=0}^{m} \binom{n}{k} (pe^{\lambda q})^k (qe^{-\lambda p})^{n-k} = e^{-\lambda n\epsilon}(pe^{q\lambda} + qe^{-\lambda p})^n.$$

We will now use the inequality saying that

$$e^x \leq x + e^{x^2} \quad \text{where } x \text{ is any real number.}$$

Then

$$\begin{aligned} P\left(\frac{S_n}{n} \geq p + \epsilon\right) &\leq e^{-\lambda n\epsilon}(pe^{q\lambda} + qe^{-\lambda p})^n \\ &\leq e^{-\lambda n\epsilon}(p\lambda q + pe^{\lambda^2 q^2} - q\lambda p + qe^{\lambda^2 p^2})^n \\ &\leq e^{-\lambda n\epsilon}(pe^{\lambda^2} + qe^{\lambda^2})^n \\ &= e^{\lambda^2 n - \lambda n\epsilon}. \end{aligned}$$

The minimum of the function $\lambda \mapsto \lambda^2 n - \lambda n\epsilon = n\lambda(\lambda - \epsilon)$ occurs when $\lambda = \epsilon/2$. We get that

$$P\left(\frac{S_n}{n} \geq p + \epsilon\right) \leq e^{-\frac{1}{4}n\epsilon^2}.$$

This finishes the proof of the theorem.

Weak Law of Large Numbers

Theorem (Weak Law of Large Numbers)

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random variables with $E|X_1| = \mu < \infty$ and $E[|X_1|^2] < \infty$. Then the sequence

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

converges to μ in the following two ways:

- **in probability**: *this means that for any $\epsilon > 0$*

$$P\left(\left|\frac{S_n}{n} - \mu\right| \leq \epsilon\right) \to 1 \quad when \quad n \to \infty.$$

- **in the L^2 norm**: *this means that*

$$E\left[|\frac{S_n}{n} - \mu|^2\right] \to 0 \quad when \quad n \to \infty.$$

Strong Law of Large Numbers

Theorem (Strong Law of Large Numbers)

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random variables with finite mean $E[X_1] = \mu < \infty$. Then the sequence

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

converges to μ **almost surly**.

References

- Walsh, John B. *Knowing the odds.* Graduate Studies in Mathematics, 139. American Mathematical Society, Providence, RI, 2012
- Khoshnevisan Davar; Rassoul Agha Firac *Introduction to Probability*, lecture notes available online at: www.math.utah.edu/ davar/math5010/summer2012/Lectures.pdf
- Grimmett, Geoffrey R.; Stirzaker, David R. Probability and random processes. Third edition. Oxford University Press, New York, 2001