Math Colloquium
Five Lectures on Probability Theory
Part 3: Central Limit Theorem

Robert Niedzialomski, rniedzia@utm.edu

March 10th, 2021

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 $X\colon \Omega \to \mathbb{R}$.
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

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

Expectation and Variance

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

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

One can show that

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

The **variance** of a random variable X is defined by

$$\mathrm{Var}[X] = E[(X - EX)^2] = E[X^2] - E[X]^2.$$

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

For two random variables X and Y define the **correlation** of X and Y by

$$\mathrm{Cor}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Theorem

We have the following properties of the variance:

- $Var[b] = 0$
- $Var[aX] = a^2 Var[X]$
- $Var[X + Y] = Var[X] + 2Cor(X, Y) + Var[Y]$

Indeed,

$$\begin{aligned}
\mathrm{Var}[X + Y] &= E[((X + Y) - E(X + Y))^2] \\
&= E[((X - E[X]) + (Y - E[Y]))^2] \\
&= E[(X - EX)^2] + 2E[(X - E[X])(Y - E[Y])] + E[(Y - EY)^2] \\
&= \mathrm{Var}[X] + 2\mathrm{Cor}(X, Y) + \mathrm{Var}[Y]
\end{aligned}$$

Independence

Let Ω be a sample space with probability measure P. This means that to an event $A \subset \Omega$ we have assigned the probability $P(A)$ of A happening. Suppose we have two events A and B with $P(B) > 0$. The **conditional probability** of A given B is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Intuitively, two events are **independent** if one of them happening has no influence on the other happening. If we denote the events by A and B we need

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

Therefore

$$\frac{P(A \cap B)}{P(B)} = P(A) \quad \text{and} \quad \frac{P(B \cap A)}{P(A)} = P(B)$$

We obtain that $P(A \cap B) = P(A)P(B)$.

Two events A and B are called **independent** if $P(A \cap B) = P(A) \cdot P(B)$.

Let X and Y be two random variables defined on a discrete probability space. We say that X and Y are **independent** if for any two sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ the events $(X \in A)$ and $(X \in B)$ are independent, i.e.,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Lemma

Two random variables X and Y defined on some discrete probability space are independent if and only if for any two real numbers i and j subsets $A, B \subset \mathbb{R}$

$$P(X = i, Y = j) = P(X = i)P(Y = j)$$

Indeed, we have

$$P(X \in A, Y \in B) = \sum_{i \in A, j \in B} P(X = i, Y = j) = \sum_{i \in A, j \in B} P(X = i)P(Y = j)$$
$$= \sum_{i \in A} P(X = i) \sum_{j \in B} P(Y = j) = P(X \in A)P(Y \in B).$$

Lemma

If two random variables X and Y defined on some discrete probability space are independent, then $E[XY] = E[X]E[Y]$.

Indeed, when X and Y are independent, we have

$$E[XY] = \sum_{i\in\Omega_X, j\in\Omega_y} P(X=i, Y=j)ij = \sum_{i\in\Omega_X, j\in\Omega_y} P(X=i)P(Y=j)ij$$

$$= \sum_{i\in\Omega_X} P(X=i)i \cdot \sum_{j\in\Omega_Y} P(Y=j)j = E[X]E[Y]$$

Lemma

If two random variables X and Y defined on some discrete probability space are independent, then $Var(X+Y) = Var(X) + Var(Y)$.

We just need to show that the correlation of X and Y is zero. We compute

$$\begin{aligned}\mathrm{Cor}(X,Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - E[Y]X - E[X]Y + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

Since X and Y are independent, we get that $\mathrm{Cor}(X,Y) = 0$.

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 -1 to tails and value 1 to heads. This means that $X(T) = -1$ and $X(H) = 1$.

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

$$E[X] = -1 \cdot p_{-1} + 1 \cdot p_1 = p - q.$$

The variance is

$$\mathrm{Var}[X] = E[X^2] - (p - q)^2 = 1 - (p - q)^2 = 4pq.$$

In particular then $p = q = \frac{1}{2}$, then $E[X] = 0$ and $\mathrm{Var}[X] = 1$.

Binomial Distribution

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

Today: Suppose we flip a coin. When we get tails we jump to the left and when we get heads we jump to the right. **What is our position after n flips of the coin?**
The sample space is $\Omega = \{(x_1, x_2, \ldots, x_n) \colon x_j = -1, 1\}$, where -1 represents tails and 1 represents heads.

Let S_n be the random variable that represents our position after n flips of the coin. Suppose $n = 2m$ is an even number. Then the possible values of S_{2m} are $-2m, -2m-2, \ldots, -2, 0, 2, \ldots, 2m-2, 2m$. Moreover, for $k = -m, -m-1, \ldots, m-1, m$ we have

$$P(S_{2m} = 2k) = \binom{2m}{m-k} q^{m-k} p^{m+k}.$$

Indeed, let i denotes the number of -1's in an outcome that produces $S_{2m} = 2k$. Then the number of 1's in that outcome is $2m - i$. Therefore, we need

$$-i + (2m - i) = 2k.$$

We solve for i and get that $i = m - k$. Hence

$$P(S_{2m} = 2k) = \binom{2m}{m-k} q^{m-k} p^{m+k}.$$

Therefore the probability distribution of the random variable S_{2m} is

$$\text{probability space:} \quad \{-2m, -2m-2, \ldots, -2, 0, 2, \ldots, 2m-2, 2m\}$$

$$\text{probability distribution:} \quad p_{2k} = P(S_{2m} = 2k) = \binom{2m}{m-k} q^{m-k} p^{m+k}.$$

What are the expectation and the variance of the random variable S_n?

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

$$E[X_1] = E[X_2] = \ldots = E[X_n] = p - q,$$
$$\mathrm{Var}[X_1] = \mathrm{Var}[X_2] = \ldots = \mathrm{Var}[X_n] = 4pq$$

Moreover, we see that

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

Therefore

$$E[S_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = n(p - q).$$

and since the random variables $X_1, X_2, \ldots$ are independent, we have

$$\mathrm{Var}[S_n] = \mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n] = 4npq.$$

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 bounded mean and variance. Denote $E[X_1] = \mu$. Then the sequence

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

converges to μ

- *in probability. This means that for any $\epsilon > 0$*

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \to 0 \quad when \quad n \to \infty.$$

- *in the L^2 norm. This means that*

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

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random variables with means 0 and variance 1. This means that $E[X_1] = E[X_2] = \ldots = 0$ and $\mathrm{Var}[X_1] = \mathrm{Var}[X_2] = \ldots = 1$.

Recall the following **Chebyshev inequality**. For any $\epsilon > 0$ and any random variable X we have

$$P\left(|X - EX| > \epsilon\right) \leq \frac{\mathrm{Var}[X]}{\epsilon^2}$$

Lemma

(Improvement of Weak Law Of Large Numbers) Let $\delta > 0$ and let $S_n = X_1 + \ldots X_n$. Then

$$P\left(\left|\frac{S_n}{n^{1/2+\delta}}\right| > \epsilon\right) \to 0 \quad when \quad n \to \infty.$$

Proof of Lemma

First we see that

$$E[S_n] = 0 \quad \text{and} \quad \mathrm{Var}[S_n] = n.$$

Therefore by Chebyshev's inequality

$$P\left(\left|\frac{S_n}{n^{1/2+\delta}}\right| > \epsilon\right) = P\left(|S_n| > n^{1/2+\delta}\epsilon\right)$$
$$\leq \frac{\mathrm{Var}[S_n]}{(n^{1/2+\delta}\epsilon)^2} = \frac{1}{\epsilon^2 n^{2\delta}} \to 0.$$

Remark: What about the behavior of $\frac{S_n}{\sqrt{n}}$ when $n \to \infty$?

We see that

$$E\left[\frac{S_n}{\sqrt{n}}\right] = 0 \quad \text{and} \quad \mathrm{Var}\left[\frac{S_n}{\sqrt{n}}\right] = 1.$$

Central Limit Theorem

Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random variables with means 0 and variance 1. Then the random variable $S_n/\sqrt{n}$ converges in distribution to a random variable with the **standard normal distribution***. This means that for any number t we have that*

$$P\left(\frac{S_n}{\sqrt{n}} \leq t\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}}\, dx$$

Equivalently, for any $a < b$

$$P\left(a \leq \frac{S_n}{\sqrt{n}} \leq b\right) \to \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{x^2}{2}}\, dx$$

De Moivre-Laplace Theorem

Theorem

The central limit theorem holds for the Bernoulli distribution taking values -1 and 1 with probability $1/2$.

Sketch of Proof: Recall that

$$P(S_{2n} = 2k) = \binom{2n}{n-k}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n+k)!(n-k)!}\left(\frac{1}{2}\right)^{2n}.$$

We will take advantage of the following **Stirling's Formula**:

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n}$$

Therefore

$$\frac{(2n)!}{(n+k)!(n-k)!}\left(\frac{1}{2}\right)^{2n} \sim \frac{1}{\sqrt{2\pi}} \cdot \frac{(2n)^{2n+1/2}}{(n+k)^{n+k+1/2}(n-k)^{n-k+1/2}} \cdot \left(\frac{1}{2}\right)^{2n}$$

Hence

$$P(S_{2n} = 2k) \sim \frac{1}{\sqrt{\pi}} \cdot \frac{n^{2n} \cdot \sqrt{n}}{(n^2 - k^2)^n (n+k)^k (n-k)^{-k} \sqrt{n^2 - k^2}}$$
$$= \frac{1}{\sqrt{\pi}} \cdot \frac{(n^2)^n}{(n^2 - k^2)^n} \cdot \frac{n^k}{(n+k)^k} \cdot \frac{n^{-k}}{(n-k)^{-k}} \cdot \sqrt{\frac{n}{n^2 - k^2}}$$
$$= \frac{1}{\sqrt{\pi}} \left(1 - \frac{k^2}{n^2}\right)^{-n} \left(1 + \frac{k}{n}\right)^{-k} \left(1 - \frac{k}{n}\right)^{k} \sqrt{\frac{n}{n^2 - k^2}}$$

We are interested in k of order $k = r\sqrt{n}$. For such k the above becomes

$$\frac{1}{\sqrt{\pi}} \left(1 - \frac{r^2}{n}\right)^{-n} \left(1 + \frac{r}{\sqrt{n}}\right)^{-r\sqrt{n}} \left(1 - \frac{r}{\sqrt{n}}\right)^{r\sqrt{n}} \sqrt{\frac{1}{n - r^2}}$$

Therefore for n large the above is approximated by

$$P(S_{2n} = 2k) = \frac{(2n)!}{(n+k)!(n-k)!} \sim \frac{1}{\sqrt{\pi}} e^{r^2} e^{-r^2} e^{-r^2} \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{2}{n}} e^{-k^2/n}$$

Therefore

$$\lim_{n\to\infty} P\left(a \le \frac{S_{2n}}{\sqrt{2n}} \le b\right) = \lim_{n\to\infty} \sum_{a\le \frac{2k}{\sqrt{2n}} \le b} \frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{2}{n}} e^{-k^2/n}$$
$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx,$$

since the sum is the Riemann sum of the function $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, where the sub-intervals have length $\sqrt{2/n}$;

for example plug in $k = 1$ to $\frac{2k}{\sqrt{2n}}$ to get $\sqrt{2/n}$ and plug in $\frac{2k}{\sqrt{2n}}$ for x in the power $-x^2/2$ to get that

$$-\frac{1}{2}\left(\frac{2k}{\sqrt{2n}}\right)^2 = -\frac{k}{n^2}.$$

References

- Lawler, Gregory F. Random walk and the heat equation. Student Mathematical Library, 55. American Mathematical Society, Providence, RI, 2010