Convolution (Lecture 10)

Proposition

Let X and Y be independent, nonnegative integer-valued random variables. Then the distribution of $X+Y$ can be described as follows:

$$\mathbb{P}(X+Y=k)=\sum_{l=0}^{k}\mathbb{P}(X=l)\mathbb{P}(Y=k-l) \qquad (k\geq 0).$$

Furthermore, we have

$$\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y); \qquad D(X+Y)=\sqrt{D^2(X)+D^2(Y)}.$$

Convolution (Lecture 10)

Proposition

Let X and Y be independent, nonnegative integer-valued random variables. Then the distribution of $X+Y$ can be described as follows:

$$\mathbb{P}(X+Y=k)=\sum_{l=0}^{k}\mathbb{P}(X=l)\mathbb{P}(Y=k-l) \qquad (k\geq 0).$$

Furthermore, we have

$$\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y); \qquad D(X+Y)=\sqrt{D^2(X)+D^2(Y)}.$$

Proof. By splitting to pairwise disjoint subsets, and using the definition of independence:

$$\mathbb{P}(X+Y=k)=\sum_{l=0}^{k}\mathbb{P}(X=l,Y=k-l)=\sum_{l=0}^{k}\mathbb{P}(X=l)\mathbb{P}(Y=k-l).$$

Convolution: example

Proposition

Let X and Y be independent random variables with Poisson distribution, such that X has parameter λ, and Y has parameter μ. Then the random variable $X + Y$ also has Poisson distribution, with parameter $\lambda + \mu$; its expectation and variance are both equal to $\lambda + \mu$.

Convolution: example

Proposition

Let X and Y be independent random variables with Poisson distribution, such that X has parameter λ, and Y has parameter μ. Then the random variable $X+Y$ also has Poisson distribution, with parameter $\lambda+\mu$; its expectation and variance are both equal to $\lambda+\mu$.

Proof. Let $k \geq 0$ be fixed. Then by the definition of Poisson distribution:

$$\mathbb{P}(X+Y=k) = \sum_{l=0}^{k} \mathbb{P}(X=l)\mathbb{P}(Y=k-l) = \sum_{l=0}^{k} \frac{\lambda^l}{l!} e^{-\lambda} \cdot \frac{\mu^{k-l}}{(k-l)!} e^{-\mu} =$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^{k} \frac{k!}{l!(k-l)!} \lambda^l \mu^{k-l} =$$

$$= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{l=0}^{k} \binom{k}{l} \lambda^l \mu^{k-l} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},$$

where in the last step we used the binomial theorem.

Convolution of common distributions

- X, Y are indepdendent, and have Poisson distribution with parameters λ_1 and $\lambda_2 \Rightarrow X + Y$ has Poisson distribution with parameter $\lambda_1 + \lambda_2$;
- X, Y are independent and have binomial distribution, with order n_1, n_2, and the same parameter p

Convolution of common distributions

- X, Y are indepdendent, and have Poisson distribution with parameters λ_1 and $\lambda_2 \Rightarrow X + Y$ has Poisson distribution with parameter $\lambda_1 + \lambda_2$;
- X, Y are independent and have binomial distribution, with order n_1, n_2, and the same parameter $p \Rightarrow X + Y$ has binomial distribution with order $n_1 + n_2$ and parameter p;

Expected value and standard deviation of the mean

In statistics, it is a common question that if we repeat

- **the same measurement**
- sokszor, **independently**,
- then take the **average** of the results,
- then how does the avarage, as a random variable, behave.

Let $X_1, X_2, \ldots, X_n$ be **independent**, **identically distributed** random variables. What can we say about the **average**

$$\frac{X_1 + X_2 + \ldots + X_n}{n}:$$

what is its **expected value** and what is its **standard deviation**?

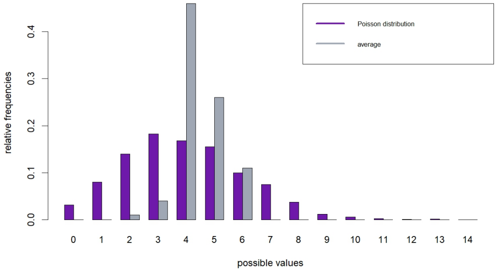
Identical distribution: $\mathbb{P}(X_j \in A) = \mathbb{P}(X_1 \in A)$ for every $j \geq 1$ and "appropriate" set $A \subseteq \mathbb{R}$; or: X_j and X_1 have the same distribution function for every $j \geq 1$.

identical distribution $\Rightarrow$ **equal expected value**, **equal standard deviation**

Independence of random variables

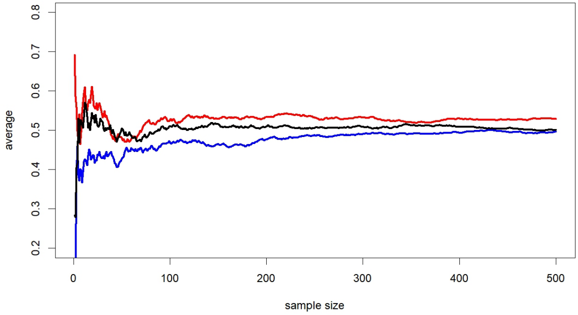
- **for two random variables**: random variables $X, Y : \Omega \to \mathbb{R}$ are **independent**, if
$$\mathbb{P}(X \leq t_1, Y \leq t_2) = \mathbb{P}(X \leq t_1) \cdot \mathbb{P}(Y \leq t_2)$$
holds for every real numbers $t_1, t_2 \in \mathbb{R}$.
- **for finitely many random variables**: random variables $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ are **independent**, if
$$\begin{aligned} \mathbb{P}(X_1 \leq t_1, &X_2 \leq t_2, \ldots, X_n \leq t_n) = \\ &= \mathbb{P}(X_1 \leq t_1) \cdot \mathbb{P}(X_2 \leq t_2) \ldots \mathbb{P}(X_n \leq t_n) \end{aligned}$$
holds for every real numbers $t_1, t_2, \ldots, t_n$.
- **for countably many random variables**: the random variables $X_1, X_2, X_3 \ldots$ are **independent**, if we get independent random variables with every choice of finitely many from $X_1, X_2, \ldots$.

Behavior of the average



Histogram of a sample of size 1000 from Poisson distribution with parameter $\lambda = 5$; and a sample of size 100, each being the average of ten independent Poisson distributed random variables with parameter $\lambda = 5 \rightarrow$ **the expected value does not change with averaging**, this is equal to 5 in both cases; **the variance decreases**

Convergence of the average



Average of a sample with uniform distribution on the interval $[0, 1]$, as a function of the sample size until $n = 500$, for three different samples

Expected value of the average

Proposition

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, for which $m = \mathbb{E}(X_1) < \infty$. Then we have

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \mathbb{E}(X_1) = m.$$

Expected value of the average

Proposition

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, for which $m = \mathbb{E}(X_1) < \infty$. Then we have

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \mathbb{E}(X_1) = m.$$

Proof.

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}\mathbb{E}(X_1 + \ldots + X_n) = \frac{1}{n} \cdot nm = m.$$

We used the linearity of expected value, and that this is determined by the distribution:

- $\mathbb{E}(cX) = c\mathbb{E}(X)$, ha $c \in \mathbb{R}$;
- $\mathbb{E}(Y + Z) = \mathbb{E}(Y) + \mathbb{E}(Z)$;
- if Y and Z has the same distribution (the same cumulative distribution function), then we have $\mathbb{E}(Y) = \mathbb{E}(Z)$.

Standard deviation of the average

Proposition

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, for which $\sigma = D(X_1) < \infty$. Then

$$D(\overline{X}) = D\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{D(X_1)}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}.$$

Standard deviation of the average

Proposition

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, for which $\sigma = D(X_1) < \infty$. Then

$$D(\overline{X}) = D\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{D(X_1)}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}.$$

Proof.

$$D(\overline{X}) = D\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{D(X_1 + \ldots + X_n)}{n} = \frac{\sqrt{n\sigma^2}}{n} = \frac{\sigma}{\sqrt{n}}.$$

We used the following properties of standard deviation:

- $D(cX) = |c|D(X)$, if $c \in \mathbb{R}$;
- $D^2(Y + Z) = D^2(Y) + D^2(Z)$, if Y and Z are independent;
- if Y and Z have the same distribution, then $D(Y) = D(Z)$

Inequalities

Suppose that a randomly chosen person is vegetarian with probability p – but p is **unknown**.

What is the minimal number of people to ask (supposing that everyone answers honestly), such that, **for every** p, the probability that the proportion of vegetarian people in the sample differs with **at most** 1% from p is **at least** 95%?

Inequalities

Suppose that a randomly chosen person is vegetarian with probability p – but p is **unknown**.

What is the minimal number of people to ask (supposing that everyone answers honestly), such that, **for every** p, the probability that the proportion of vegetarian people in the sample differs with **at most** 1% from p is **at least** 95%?

To answer such a question, we need the following elements:

- proportion is an average $\rightarrow$ how fast does the standard deviation decrease?
- if the **standard devation is small**, how can we prove that with large probability, the error is small $\rightarrow$ the methods here are the **Markov and Chebyshev inequalities**.

Markov inequality and its proof

Markov inequality. Let $t > 0$, and X be a **nonnegative random variable whose expected value exists**, that is, for which $X \geq 0$ holds for sure, and $\mathbb{E}(X)$ exists. Then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof. We define random variable Y as follows:

$$Y = \begin{cases} t, & \text{if } X \geq t; \\ 0, & \text{if } X < t. \end{cases}$$

Markov inequality and its proof

Markov inequality. Let $t > 0$, and X be a **nonnegative random variable whose expected value exists**, that is, for which $X \geq 0$ holds for sure, and $\mathbb{E}(X)$ exists. Then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof. We define random variable Y as follows:

$$Y = \begin{cases} t, & \text{if } X \geq t; \\ 0, & \text{if } X < t. \end{cases}$$

In both cases Y is at most X (by using $X \geq 0$):

$$Y \leq X \qquad \Rightarrow \qquad \mathbb{E}(Y) \leq \mathbb{E}(X).$$

On the other hand, Y has only two values, its expected value is as follows:

$$\mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + t \cdot \mathbb{P}(Y = t) = t \cdot \mathbb{P}(X \geq t) \leq \mathbb{E}(X).$$

Markov inequality and its proof

Markov inequality. Let $t > 0$, and X be a **nonnegative random variable whose expected value exists**, that is, for which $X \geq 0$ holds for sure, and $\mathbb{E}(X)$ exists. Then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof. We define random variable Y as follows:

$$Y = \begin{cases} t, & \text{if } X \geq t; \\ 0, & \text{if } X < t. \end{cases}$$

In both cases Y is at most X (by using $X \geq 0$):

$$Y \leq X \qquad \Rightarrow \qquad \mathbb{E}(Y) \leq \mathbb{E}(X).$$

On the other hand, Y has only two values, its expected value is as follows:

$$\mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + t \cdot \mathbb{P}(Y = t) = t \cdot \mathbb{P}(X \geq t) \leq \mathbb{E}(X).$$

We get the statement by dividing both sides of the last equality with positive number t.

Chebyshev inequality and its proof

Markov inequality. Let $t > 0$, and X be a **nonnegative random variable whose expected value exists**, that is, for which $X \geq 0$ holds for sure, and $\mathbb{E}(X)$ exists. Then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Chebyshev inequality. Let $t > 0$, and X be a random **random variable whose standard deviation exists**, that is, for which $D(X)$ exists. Then we have

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{D^2(X)}{t^2}.$$

Proof. Let $Z = (X - \mathbb{E}(X))^2$. This random variable is nonnegative and has finite expectation, hence we can apply the Markov inequality, for the positive number $t^2 > 0$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) = \mathbb{P}((X - \mathbb{E}(X))^2 \geq t^2) = \mathbb{P}(Z \geq t^2) \leq$$

$$\overset{Markov}{\leq} \frac{\mathbb{E}(Z)}{t^2} = \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{t^2} = \frac{D^2(X)}{t^2}$$

based on the definition of variance.

Application of Chebyshev inequality

What is the minimal number of people to ask (supposing that everyone answers honestly), such that, **for every** p, the probability that the proportion of vegetarian people in the sample differs with **at most** 1% from p is **at least** 95%?

n participants, everyone is vegetarian with probability p

X: number of vegetarian people among the participants

Necessary condition:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0.01\right) \geq 0.95$$

holds for every $0 \leq p \leq 1$ (we do not know p)

Application of Chebyshev inequality

n participants, everyone is vegetarian with probability p

X: number of vegetarian people among the participant

Since X has binomial distribution:

$$\mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot np = p; \qquad D\left(\frac{X}{n}\right) = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}.$$

Application of Chebyshev inequality

n participants, everyone is vegetarian with probability p

X: number of vegetarian people among the participant

Since X has binomial distribution:

$$\mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n} \cdot np = p; \qquad D\left(\frac{X}{n}\right) = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}.$$

Chebyshev inequality for the random variable X/n:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0.01\right) \leq \frac{D^2\left(\frac{X}{n}\right)}{0.01^2} = \frac{p(1-p)}{0.01^2 \cdot n} \leq \frac{1}{4 \cdot 0.01^2 \cdot n},$$

because $p(1-p) \leq 1/4$ always holds (e.g. by the arithmetic-geometric mean).

Application of Chebyshev inequality

n participants, everyone is vegetarian with probability p

X: number of vegetarian people among the participant

Chebyshev inequality implies that

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0.01\right) \leq \frac{1}{4 \cdot 0.01^2 \cdot n}$$

Necessary condition:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0.01\right) \geq 0.95,$$

that is, we have

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| > 0.01\right) \leq 0.05.$$

It follows that **the following is sufficient**:

$$\frac{1}{4 \cdot 0.01^2 \cdot n} \leq 0.05 \quad \Leftrightarrow \quad n \geq \frac{1}{4 \cdot 0.01^2 \cdot 0.05} = 50000.$$

Application of Chebyshev inequality

n participants, everyone is vegetarian with probability p

X: number of vegetarian people among the participant

Chebyshev inequality implies that

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \geq 0.01\right) \leq \frac{1}{4 \cdot 0.01^2 \cdot n}$$

Necessary condition:

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0.01\right) \geq 0.95,$$

that is, we have

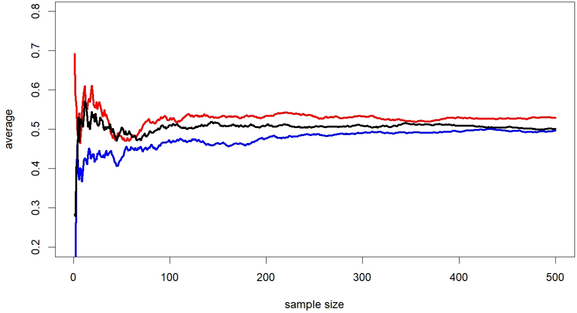
$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| > 0.01\right) \leq 0.05.$$

It follows that **the following is sufficient**:

$$\frac{1}{4 \cdot 0.01^2 \cdot n} \leq 0.05 \quad \Leftrightarrow \quad n \geq \frac{1}{4 \cdot 0.01^2 \cdot 0.05} = 50000.$$

If 0.01 would be replaced by 0.005 (its half), $n \geq 200000$ (four times 50000) would be the lower bound.

Convergence of the average



Average of a sample with uniform distribution on the interval $[0, 1]$, as a function of the sample size until $n = 500$, for three different samples

Types of convergence

Sequences of random variables might converge with respect to **different definitions**.

The sequence of random variables $Z_1, Z_2, \ldots,$ **converges in probability** to random variable Z if for every $\varepsilon > 0$ the following holds:

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \to 0$$

as $n \to \infty$.

Types of convergence

Sequences of random variables might converge with respect to **different definitions**.

The sequence of random variables $Z_1, Z_2, \ldots,$ **converges in probability** to random variable Z if for every $\varepsilon > 0$ the following holds:

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \to 0$$

as $n \to \infty$.

The sequence of random variables $Z_1, Z_2, \ldots,$ converges **with probability 1** to random variable Z if

$$\mathbb{P}(\omega \in \Omega : Z_n(\omega) \to Z(\omega) \ \text{ as } n \to \infty) = 1.$$

Types of convergence

Sequences of random variables might converge with respect to **different definitions**.

The sequence of random variables $Z_1, Z_2, \dots,$ **converges in probability** to random variable Z if for every $\varepsilon > 0$ the following holds:

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \to 0$$

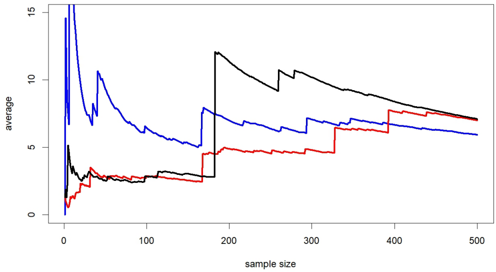
as $n \to \infty$.

The sequence of random variables $Z_1, Z_2, \dots,$ converges **with probability 1** to random variable Z if

$$\mathbb{P}(\omega \in \Omega : Z_n(\omega) \to Z(\omega) \;\text{ as } n \to \infty) = 1.$$

The sequence converges with probability 1 $\Rightarrow$ it converges in probability, but the other direction is not true.

Behavior of the average



The average as the function of sample size in a case when the expected value does not exist

Weak law of large numbers: proof

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with finite variance. Let $m = \mathbb{E}(X_1)$ and $\sigma = D(X_1)$.

As we have seen earlier:

$$\mathbb{E}(\overline{X}) = m; \qquad D^2(\overline{X}) = \frac{\sigma^2}{n}.$$

Weak law of large numbers: proof

Let $X_1, \dots, X_n$ be independent identically distributed random variables with finite variance. Let $m = \mathbb{E}(X_1)$ and $\sigma = D(X_1)$.

As we have seen earlier:

$$\mathbb{E}(\overline{X}) = m; \qquad D^2(\overline{X}) = \frac{\sigma^2}{n}.$$

Chebyshev inequality implies that for every $\varepsilon > 0$ we have

$$\mathbb{P}(|\overline{X} - m| > \varepsilon) \leq \frac{D^2(\overline{X})}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n} \to 0 \quad (n \to \infty).$$

Hence $\overline{X} \to m = \mathbb{E}(X_1)$ in probability.

Laws of large numbers

Theorem (Weak law of large numbers)

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables. Suppose that $D(X_1) < \infty$. Then for every $\varepsilon > 0$ we have

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}(X_1)| > \varepsilon) \to 0 \qquad (n \to \infty),$$

that is, $\overline{X}_n \to \mathbb{E}(X_1)$ in probability.

Laws of large numbers

Theorem (Weak law of large numbers)

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables. Suppose that $D(X_1) < \infty$. Then for every $\varepsilon > 0$ we have

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}(X_1)| > \varepsilon) \to 0 \qquad (n \to \infty),$$

that is, $\overline{X}_n \to \mathbb{E}(X_1)$ in probability.

Theorem (Strong law of large numbers)

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables. Suppose that $m = \mathbb{E}(X_1) < \infty$. Then we have

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}(X_1) = m$$

holds with probability 1 as $n \to \infty$.

In the second version, a stronger statement follows from a weaker condition.