Unit 8 Limit Theorems and Classical Statistics

Inequalities, Convergence, and the Weak Law of Large Numbers

Weak Law of Large Numbers
Taking the average of more numbers converges to the actual mean
(Taking different sample means)

The definition of convergence here is a bit different as we are talking about a random variable converging to a number

The Markov inequality

- Use a bit of information about a distribution to learn something about probabilities of "extreme events"
- "If $X \geq 0$ and $\mathbf{E}[X]$ is small, then X is unlikely to be very large"

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$

$$E[X] = \int_0^\infty x f_X(x)\,dx \geq \int_a^\infty x f_X(x)\,dx$$

$$\geq \int_a^\infty a f_X(x)\,dx = a P(X \geq a)$$

The Markov inequality

- Use a bit of information about a distribution to learn something about probabilities of "extreme events"
- "If $X \geq 0$ and $\mathbf{E}[X]$ is small, then X is unlikely to be very large"

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$

$$Y = \begin{cases} 0, & \text{if } X < a \\ a, & \text{if } X \geq a \end{cases} \qquad a P(X \geq a) = E[Y] \leq E[X]$$

Assumes random variable is nonnegative

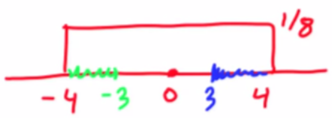
The Markov inequality

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}$

- **Example:** X is Exponential($\lambda = 1$): $\mathbf{P}(X \geq a) \leq \frac{1}{a}$

- **Example:** X is Uniform$[-4, 4]$: $\mathbf{P}(X \geq 3) \leq$ $P(|X| \geq 3) \leq \frac{E[|X|]}{3} = \frac{2}{3}$

$$= \frac{1}{2} P(|X| \geq 3) \leq \frac{1}{3}$$



The Chebyshev inequality

- Random variable X, with finite mean μ and variance σ^2
- "If the variance is small, then X is unlikely to be too far from the mean"

Chebyshev inequality: $\mathbf{P}\big(|X - \mu| \geq c\big) \leq \frac{\sigma^2}{c^2}$

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}$

$$P(|X-\mu| \geq c) = P\left(\underline{(X-\mu)^2} \geq c^2\right) \leq \frac{E[(X-\mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

c is positive here

The Chebyshev inequality

Chebyshev inequality: $\mathbf{P}\big(|X-\mu| \geq c\big) \leq \frac{\sigma^2}{c^2}$

$$\mathbf{P}\big(|X-\mu| \geq k\sigma\big) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} \qquad k=3 \qquad \leq \frac{1}{9}$$

- **Example:** X is Exponential($\lambda = 1$): $\mathbf{P}(X \geq a) \leq \frac{1}{a}$ (Markov)

$$P(X \geq a) = P(X-1 \geq a-1) \leq P(|X-1| \geq a-1) \leq \frac{1}{(a-1)^2} \sim \frac{1}{a^2}$$

Probability of being with k standard deviations is true no matter the distribution
In most cases, Chebyshev inequality is more telling than the Markov inequality because it also uses information about the variance

Exercise: Markov inequality

0/1 point (graded)

Let Z be a nonnegative random variable that satisfies $\mathbf{E}[Z^4] = 4$. Apply the Markov inequality to the random variable Z^4 to find the tightest possible (given the available information) upper bound on $\mathbf{P}(Z \geq 2)$.

$\mathbf{P}(Z \geq 2) \leq$ 1/8 ✖ **Answer:** 0.25

Solution:

We have

$$\mathbf{P}(Z \geq 2) = \mathbf{P}(Z^4 \geq 16) \leq \frac{\mathbf{E}[Z^4]}{16} = \frac{4}{16} = \frac{1}{4}.$$

Exercise: Chebyshev inequality

0/1 point (graded)

Let Z be normal with zero mean and variance equal to 4. For this case, the Chebyshev inequality yields:

$\mathbf{P}\big(|Z| \geq 4\big) \leq$ 1 ✖ **Answer:** 0.25

Solution:

We have

$$\mathbf{P}\big(|Z| \geq 4\big) \leq \frac{\mathsf{Var}\,(Z)}{4^2} = \frac{4}{4^2} = \frac{1}{4}.$$

Exercise: Chebyshev versus Markov

2/2 points (graded)

Let X be a random variable with zero mean and finite variance. The Markov inequality applied to $|X|$ yields

$$\mathbf{P}\big(\,|X| \geq a\big) \leq \frac{\mathbf{E}\big[\,|X|\,\big]}{a},$$

whereas the Chebyshev inequality yields

$$\mathbf{P}\big(\,|X| \geq a\big) \leq \frac{\mathbf{E}\,[X^2]}{a^2}.$$

a) Is it true that the Chebyshev inequality is stronger (i.e., the upper bound is smaller) than the Markov inequality, when a is very large?

Yes ✔ **Answer:** Yes

b) Is it true that the Chebyshev inequality is always stronger (i.e., the upper bound is smaller) than the Markov inequality?

No ✔ **Answer:** No

Solution:

a) Yes, because for very large a, the term $1/a^2$ will be much smaller than $1/a$.

b) No. For example, suppose that $a = 1$. It is certainly possible to have $\mathbf{E}\,[X^2] > \mathbf{E}\big[\,|X|\,\big]$, in which case the Markov inequality provides a stronger bound.

The Weak Law of Large Numbers (WLLN)

- $X_1, X_2, \ldots$ i.i.d.; finite mean μ and variance σ^2

  Sample mean: $M_n = \dfrac{X_1 + \cdots + X_n}{n}$ $\qquad \mu = E[X_i]$

- $\mathbf{E}[M_n] = \dfrac{E[X_1 + \cdots + X_n]}{n} = \dfrac{n\mu}{n} = \mu$

- $\text{Var}(M_n) = \dfrac{\text{Var}(X_1 + \cdots + X_n)}{n^2} = \dfrac{n\sigma^2}{n^2} = \dfrac{\sigma^2}{n}$

  $\mathbf{P}(|M_n - \mu| \geq \epsilon) \leq \dfrac{\text{var}(M_n)}{\varepsilon^2} = \dfrac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$ (fixed $\varepsilon > 0$)

$$\boxed{\textbf{WLLN: } \text{For } \epsilon > 0, \quad \mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) = \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0, \quad \text{as } n \to \infty}$$

Sample mean is a random variable as it is a sum of random variables
It is the simplest way of trying to estimate the true mean

Interpreting the WLLN

$M_n = (X_1 + \cdots + X_n)/n$

$$\boxed{\textbf{WLLN: } \text{For } \epsilon > 0, \quad \mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) = \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0, \quad \text{as } n \to \infty}$$

- One experiment
  - many measurements $X_i = \mu + W_i$
  - W_i: measurement noise; $\mathbf{E}[W_i] = 0$; independent W_i
  - **sample mean** M_n is unlikely to be far off from **true mean** μ
- Many independent repetitions of the same experiment
  - event A, with $p = \mathbf{P}(A)$ $\qquad E[X_i] = p$
  - X_i: indicator of event A $\qquad X_i = 1$, if A occurs; 0, o.w.
  - the sample mean M_n is the **empirical frequency** of event A

Far off refers to a distance of epsilon

Exercise: Sample mean bounds

1/2 points (graded)

By the argument in the last video, if the X_i are i.i.d. with mean μ and variance σ^2, and if $M_n = (X_1 + \cdots + X_n)/n$, then we have an inequality of the form

$$\mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) \leq \frac{a\sigma^2}{n},$$

for a suitable value of a.

a) If $\epsilon = 0.1$, then the value of a is: 10 ✖ **Answer:** 100

b) If we change $\epsilon = 0.1$ to $\epsilon = 0.1/k$, for $k \geq 1$ (i.e., if we are interested in k times higher accuracy), how should we change n so that the value of the upper bound does not change from the value calculated in part (a)?

n should

- ○ stay the same
- ○ increase by a factor of k
- ● increase by a factor of k^2
- ○ decrease by a factor of k
- ○ none of the above

✔

Solution:

a) Chebyshev's inequality yields

$$\mathbf{P}\big(\,|M_n - \mu| \geq \epsilon\big) \leq \frac{\sigma^2}{n\epsilon^2},$$

so that $a = 1/\epsilon^2 = 1/0.1^2 = 100.$

b) In order to keep the same upper bound, the term $n\epsilon^2$ in the denominator needs to stay constant. If we reduce ϵ by a factor of k, then ϵ^2 gets reduced by a factor of k^2. Thus, n will have to be increased by a factor of k^2.

The pollster's problem

- p: fraction of population that will vote "yes" in a referendum
- ith (randomly selected) person polled: $X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$ $E[X_i] = p$

  uniformly, independently

- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$
- Try $n = 10{,}000$
- $\mathbf{P}\big(|M_{10{,}000} - p| \geq 0.01\big) \leq \frac{\sigma^2}{n\varepsilon^2} = \frac{p(1-p)}{10^4 \cdot 10^{-4}} \leq \frac{1}{4}$ ← want $\leq 5\%$

$$\frac{1/4}{n\,10^{-4}} \leq \frac{5}{10^2} \iff n \geq \frac{10^6}{20} = 50{,}000$$ ← will suffice

Can't guarantee a error of no more than 1% but can say that the probability of getting a error above this is x%
Here we are saying 1/4 (25%), but that is really high
so we can work backwards and calculate how many people we need to sample to get the probability to less than 5%
But there are more accurate ways of doing this....

Exercise: Polling

2/4 points (graded)

We saw that if we want to have a probability of at least 95% that the poll results are within 1 percentage point of the truth, Chebyshev's inequality recommends a sample size of $n = 50,000$. This is very large compared to what is done in practice. Newspaper polls use smaller sample sizes for various reasons. For each of the following, decide whether it is a valid reason.

In the real world,

a) the accuracy requirements are looser.

Yes ✔ **Answer:** Yes

b) the Chebyshev bound is too conservative.

No ✖ **Answer:** Yes

c) the people sampled are all different, so their answers are not identically distributed.

No ✔ **Answer:** No

d) the people sampled do not have independent opinions.

Yes ✖ **Answer:** No

Solution:

a) Requiring the accuracy to be within one percentage point is too strict for most real world situations.

b) The Chebyshev bound is conservative as stated in the video.

c,d) No matter how opinions get formed, as long as we choose who to ask at random, independently and uniformly, the opinions reported will be i.i.d. random variables, so that the last two considerations do not apply.

Convergence "in probability"

WLLN: For any $\epsilon > 0$, $\mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) \to 0$, as $n \to \infty$

- Would like to say that "M_n converges to μ" $M_n \xrightarrow[n\to\infty]{i.P.} \mu$
- Need to define the word "converges"
- Sequence of random variables Y_n; not necessarily independent

Definition: A sequence Y_n **converges in probability** to a number a if:

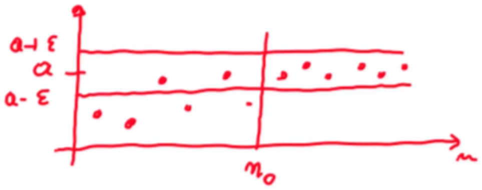
for any $\epsilon > 0$, $\lim\limits_{n\to\infty} \mathbf{P}\big(|Y_n - a| \geq \epsilon\big) = 0$

Understanding convergence "in probability"

- Ordinary convergence
  - Sequence a_n; number a

    $a_n \to a$

  "a_n eventually gets and stays (arbitrarily) close to a"

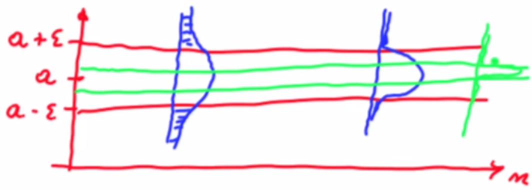


- For every $\epsilon > 0$, there exists n_0, such that for every $n \geq n_0$, we have $|a_n - a| \leq \epsilon$

- Convergence in probability
  - Sequence Y_n; number a

    $Y_n \to a$

- for any $\epsilon > 0$, $\mathbf{P}\big(|Y_n - a| \geq \epsilon\big) \to 0$



"(almost all) of the PMF/PDF of Y_n eventually gets concentrated (arbitrarily) close to a"

The narrower the band of epsilon the longer it will take to get most of the probability in that band

Some properties

- Suppose that $X_n \to a$, $Y_n \to b$, in probability

- If g is continuous, then $g(X_n) \to g(a)$ $\qquad X_n^2 \to a^2$

- $X_n + Y_n \to a + b$

- **But:** $\mathbf{E}[X_n]$ need not converge to a

Convergence of random variables does not imply convergence of expected values

Convergence in probability examples

$Y_n \xrightarrow[n\to\infty]{i.p.} 0.$

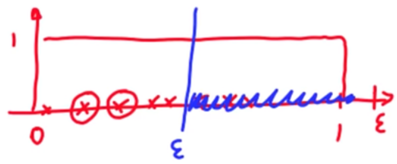
$\varepsilon > 0 \qquad P(|Y_n - 0| \geq \varepsilon) = 1/n \xrightarrow[n\to\infty]{} 0$

$\mathbf{E}[Y_n] = n^2 \cdot \frac{1}{n} = n \xrightarrow[n\to\infty]{} \infty$

- convergence in probability does **not** imply convergence of expectations

The expected value is very sensitive to the tail but the probability isn't affected when the probability mass of the tail is small

Convergence in probability examples



- X_i: i.i.d., uniform on $[0, 1]$
- $Y_n = \min\{X_1, \ldots, X_n\}$

$Y_{n+1} \leq Y_n$

$\mathbf{P}\big(|Y_n - 0| \geq \epsilon\big) = P(Y_n \geq \varepsilon)\,.$

$\varepsilon > 0 \qquad = P(X_1 \geq \varepsilon, \ldots, X_n \geq \varepsilon)$

$\varepsilon > 1 \qquad = P(X_1 \geq \varepsilon) \cdots P(X_n \geq \varepsilon)$

$\varepsilon \leq 1 \qquad = (1-\varepsilon)^n \xrightarrow[n\to\infty]{} 0$

$Y_n \xrightarrow[n\to\infty]{i.p.} 0$

Exercise: Convergence in probability

0/3 points (graded)

a) Suppose that X_n is an exponential random variable with parameter $\lambda = n$. Does the sequence $\{X_n\}$ converge in probability?

No ✖ **Answer:** Yes

b) Suppose that X_n is an exponential random variable with parameter $\lambda = 1/n$. Does the sequence $\{X_n\}$ converge in probability?

Yes ✖ **Answer:** No

c) Suppose that the random variables in the sequence $\{X_n\}$ are independent, and that the sequence converges to some number a, in probability. Let $\{Y_n\}$ be another sequence of random variables that are dependent, but where each Y_n has the same distribution (CDF) as X_n. Is it necessarily true that the sequence $\{Y_n\}$ converges to a in probability?

No ✖ **Answer:** Yes

Solution:

a) In the first case, for any $\epsilon > 0$, we have $\mathbf{P}(X_n \ge \epsilon) = e^{-n\epsilon}$, which converges to zero. Therefore, we have convergence in probability.

b) In the second case, for any $\epsilon > 0$, we have $\mathbf{P}(X_n \ge \epsilon) = e^{-\epsilon/n}$, which converges to one. Therefore, we do not have convergence in probability.

c) Dependence will not make a difference because the definition of convergence in probability involves probabilities of the form $\mathbf{P}\big(|Y_n - a| \ge \epsilon\big)$. These probabilities are completely determined by the marginal distributions of the random variables Y_n, and these marginal distributions are the same as for the sequence X_n.

Related topics

- Better bounds/approximations on tail probabilities $\le \sim \frac{1}{n}$
  - Markov and Chebyshev inequalities
  - Chernoff bound $P(|M_n - \mu| \ge a) \le e^{-n \underbrace{h(a)}_{>0}}$
  - Central limit theorem "$M_n \sim N(\mu, \sigma^2/n)$"
- Different types of convergence
  - Convergence in probability
  - Convergence "with probability 1" $P\left(\{w : Y_n(w) \xrightarrow[n\to\infty]{} Y(w)\}\right) = 1$
  - Strong law of large numbers $M_n \xrightarrow[n\to\infty]{wp1} \mu$
  - Convergence of a sequence of distributions (CDFs) to a limiting CDF

Strong law of large numbers means sample mean will converge to actual mean with probability 1

The Central Limit Theorem