

# Inequalities recap

## Inequalities recap

Name	Conditions	Formula	Use
Cauchy-Schwarz	$\mathbb{E}[X^2] < \infty, \mathbb{E}[Y^2] < \infty$	$ \mathbb{E}[XY]  \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$	
Jensen	$\mathbb{E}[X] < \infty, \phi$ convex	$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$	
Markov	$\mathbb{E}[X] < \infty, X \geq 0$	$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X]}{x}$	
Chebyshev	$\text{Var}(X) < \infty, x > 0$	$\mathbb{P}( X - \mathbb{E}[X]  \geq x) \leq \frac{\text{Var}(X)}{x^2}$	
Chernoff	$\mathbb{E}[e^{tX}] < \infty, t > 0, x > 0$	$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}}$	Tails

## Homework problems

### Problem 1

In a national survey, a random sample of people are chosen and asked whether they support a certain policy. Assume that everyone in the population is equally likely to be surveyed at each step, and that the sampling is with replacement (sampling without replacement is typically more realistic, but with replacement will be a good approximation if the sample size is small compared to the population size). Let  $n$  be the sample size, and let  $\hat{p}$  and  $p$  be the proportion of people who support the policy in the sample and in the entire population, respectively. Show that for every  $c > 0$ ,

$$\mathbb{P}(|\hat{p} - p| > c\sigma) \leq \frac{1}{4nc^2}$$

## Solution 1

Let  $X \sim \text{Bin}(n, p)$  describe the number of people, who support the policy. Then,

$$\hat{p} = \frac{X}{n}$$

Then,

$$\mathbb{E}[\hat{p}] = \frac{1}{n} \mathbb{E}[X] = \frac{1}{n} np = p$$

$$\text{Var}(\hat{p}) = \frac{1}{n^2} \text{Var}(X) = \frac{p(1-p)}{n}$$

Then we use Chebyshev inequality

$$\mathbb{P}(|\hat{p} - \mathbb{E}[\hat{p}]| > c) \leq \frac{\text{Var}(\hat{p})}{c^2}$$

$$\mathbb{P}(|\hat{p} - p| > c) \leq \frac{p(1-p)}{nc^2} \leq \frac{1}{4nc^2}$$

## Convergence of random variables

### Convergence of random variables

Consider infinite series of r.v.s  $X_1, X_2, \dots$

1.  $X_n$  converges to  $X$  **in probability** ( $X_n \xrightarrow{P} X$ ) if  $\forall \varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$$

2.  $X_n$  converges to  $X$  **almost surely** ( $X_n \xrightarrow{\text{a.s.}} X$ ) if

$$\mathbb{P}(X_n \rightarrow X) = 1$$

3.  $X_n$  converges to  $X$  **mean-square** ( $X_n \xrightarrow{\text{m.s.}} X$ ) if

$$\mathbb{E}[(X_n - X)^2] \rightarrow 0$$

4.  $X_n$  converges to  $X$  **in distribution** ( $X_n \xrightarrow{d} X$ ) if for any bounded continuous function  $\varphi$

$$\mathbb{E}[\varphi(X_n)] \rightarrow \mathbb{E}[\varphi(X)]$$

## Convergence in probability, detailed

- $X_n \xrightarrow{d} X$  is equivalent to  $F_{X_n}(x) \rightarrow F_X(x)$
- If r.v. takes only integer values, then  $X_n \xrightarrow{d} X$  is equivalent to  $\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k)$
- $X_n \xrightarrow{d} X$  is equivalent to convergence of characteristic functions, PGFs, MGFs

### Example 1

Let  $X_n \sim \text{Bin}(n, p_n)$  and  $np_n \rightarrow \lambda, n \rightarrow \infty$ . Show that  $X_n \xrightarrow{d} X \sim \text{Pois}(\lambda)$ .

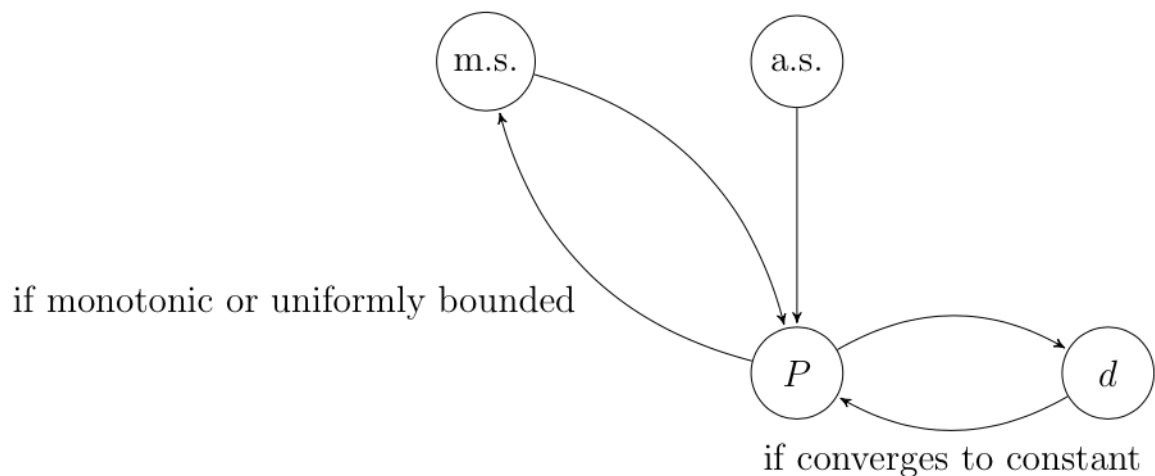
### Solution 1

We need to show  $\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k)$ .

$$\begin{aligned}
 \mathbb{P}(X_n = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right) \\
 &= \left[ \frac{\lambda^k}{k!} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \right] \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\rightarrow 1} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \\
 &= \mathbb{P}(X = k)
 \end{aligned}$$

## Relations between types of convergence

1.  $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$
2.  $X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{P} X$
3.  $X_n \xrightarrow{\text{m.s.}} X \Rightarrow X_n \xrightarrow{P} X$
4. If  $\{X_n\}$  is monotonic a.s. ( $X_{n+1} \geq X_n$  a.s.)  $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\text{m.s.}} X$
5. If  $X_n$  is uniformly bounded ( $|X_n| < a$  a.s.)  $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\text{m.s.}} X$
6.  $X_n \xrightarrow{d} C \Rightarrow X_n \xrightarrow{P} C$



## Limit theorems

### Limit theorems

We will discuss the two most famous theorems in probability:

- the law of large numbers (LLN)
- the central limit theorem (CLT)

Both tell us what happens to the **sample mean**  $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  as we obtain more and more data  $n \rightarrow \infty$ .

Limit theorems let us make approximations which are likely to work well when we have a large number of data points.

## Law of large numbers

Let  $X_1, X_2, \dots, X_n$  be i.i.d. r.v.s with  $\mathbb{E}[X_i] = m$  and  $\mathbb{V}\text{ar}(X) = \sigma^2$ . Then,

$$\mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n X_i - m \right)^2 \right] \rightarrow 0$$

So, LLN says that

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{m.s.}} m$$

We know that m.s. convergence implies  $P$ -convergence, so

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} m$$

## Law of large numbers

1. (Chebyshev form) The i.i.d. part can be relaxed to  $\text{cov}(X_i, X_j) \leq 0$
2. (Khinchin form) The existence of variance can be relaxed, but then we only have convergence in probability
3. Both parts can not be relaxed at the same time
4. Existence of expectation is essential

## Strong law of large numbers

Let  $X_1, X_2, \dots, X_n$  be i.i.d. r.v.s with  $\mathbb{E}[X_i] = m$ . Then,

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} m$$

## Example 2

Consider you have  $K_1$  money, that you want to multiply. You have two possibilities to do that:

- to put the money into the bank that guarantees  $b \cdot 100\%$  yearly interest
- to invest into stocks that will earn you  $X \cdot 100\%$  yearly interest, where  $X$  is an r.v. with expected value  $m_X > b > 0$

Denote the fraction of money put into the bank as  $u$  and the fraction of money invested into stocks  $v$ , such that  $0 \leq u + v \leq 1$ . This way, in a year, you will have

$$K_2 = K_1(1 + bu + Xv)$$

Obviously, the strategy maximizing the expected income is to set  $u = 0, v = 1$ . Show that

- $\mathbb{E}[K_t] \rightarrow \infty$
- If  $\mathbb{E}[\log(1 + X)] < 0$ , then  $\mathbb{E}[K_t] \xrightarrow{\text{a.s.}} 0$

## Solution 2.1

$$\mathbb{E}[K_{t+1}] = \mathbb{E}[K_t(1 + X_t)] = \mathbb{E}[K_1 \prod_{i=1}^t (1 + X_i)] = K_1(1 + m_x)^t \rightarrow \infty$$

## Solution 2.2

Consider change of variables

$$K_{t+1} = K_1 e^{tY_t}$$

where

$$Y_t = \frac{1}{t} \sum_{i=1}^t \log(1 + X_i)$$

Consider SLLN for  $\log(1 + X_i)$ , then

$$\frac{1}{t} \sum_{i=1}^t \log(1 + X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[\log(1 + X)] = -a < 0$$

Then,

$$K_{t+1} = K_1 e^{-at} \xrightarrow{\text{a.s.}} 0$$

## Central limit theorem

Let  $X_1, X_2, \dots, X_n$  be i.i.d. r.v.s,  $\mathbb{E}[X] = \mu$ ,  $\text{Var}(X) = \sigma^2$ . Then,

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

If additionally  $X_i$  are continuous, then

$$f_{Z_n}(x) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

It works for random vector as well. Let  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  be i.i.d. with  $\mathbb{E}[\mathbf{X}] = \mathbf{m}$  and covariance matrix  $\Sigma$ . Then,

$$\frac{\sum_{i=1}^n \mathbf{X}_i - n\mathbf{m}}{\sqrt{n}} \sim \mathcal{N}(0, \Sigma)$$

## Example 3

В выборах участвовали два кандидата Т и С. Т набрал  $N_T = 520000$  голосов, а С —  $N_C = 480000$ . Оказалось, что аппарат считающий бюллетени, был настроен неправильно, и случайно менял каждый голос на противоположный с вероятностью  $p = 0.45$ . Найти вероятность того, что число голосов за Т оказалось не меньше  $N_T$ , если изначально за Т и С было отдано поровну голосов:  $N_{0T} = N_{0C} = 500000$ .

Two parties  $L$  and  $D$  participate in election.  $L$  gets  $N_L = 520000$  votes,  $D$  gets  $N_D = 480000$  votes. Turns out that the voting machine was broken and randomly changed a vote with probability  $p = 0.45$ . Find probability that votes for  $L$  were at least  $N_L$  if the actual votes were  $N_{0L} = N_{0D} = 500000$ .

## Solution 3

Consider r.v.  $X_i \sim \text{Be}(p)$  - indicator of error  $D \rightarrow L$ , and  $Y_i \sim \text{Be}(p)$  - indicator of error  $L \rightarrow D$ . Then, the votes for  $L$  are

$$Z = \sum_{i=1}^{N_{0L}} X_i + \sum_{j=1}^{N_{0D}} (1 - Y_j)$$

By CLT,

$$\frac{Z - \mathbb{E}[Z]}{\sqrt{N_{OD} \text{Var}(X_i)}} \sim \mathcal{N}(0, 1)$$

$$\mathbb{E}[Z] = N_{0L}p + N_{0D}(1 - p) = N_{OD}$$

$$\text{Var}(X_i) = p(1 - p)$$

Then,

$$\mathbb{P}(Z \geq N_L) = \mathbb{P}\left(\frac{Z - N_{OD}}{\sqrt{2N_{OD}p(1-p)}} \geq \frac{N_L - N_{OD}}{\sqrt{2N_{OD}p(1-p)}}\right) \approx \mathbb{P}(V \geq a)$$



## Solution 3

$$\mathbb{P}(Z \geq N_L) = \mathbb{P}\left(\frac{Z - N_{OD}}{2N_{OD}p(1-p)} \geq \frac{N_L - N_{OD}}{2N_{OD}p(1-p)}\right) \approx \mathbb{P}(V \geq a)$$

where  $V \sim \mathcal{N}(0, 1)$  and

$$a = \frac{N_L - N_{OD}}{2N_{OD}p(1-p)} \approx 40$$

$$\mathbb{P}(Z \geq N_L) \approx 10^{-350}$$