

PROBABILITY THEORY

LECTURE 6

Per Siden

**Division of Statistics
Dept. of Computer and Information Science
Linköping University**

OVERVIEW LECTURE 6

- ▶ Modes of convergence
 - ▶ almost surely
 - ▶ in probability
 - ▶ in r -mean
 - ▶ in distribution
- ▶ Law of large numbers
- ▶ Central limit theorem
- ▶ Convergence of sums, differences and products.

INTRODUCTION

- ▶ We are often interested in the **large sample**, or **asymptotic**, behavior of random variables.
- ▶ We are considering a **sequence** of random variables X_1, X_2, \dots , also denoted by $\{X_n\}_{n=1}^{\infty}$.
- ▶ Example: what can we say about the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n Y_i$ in large samples?
 - ▶ Does it converge to a single number? (**law of large numbers**)
 - ▶ How fast? (central limit theorem)
 - ▶ What is the distribution of the sample mean in large samples? (**central limit theorem**)
- ▶ The usual limit theorems from calculus will not do. Need to consider that X_n is a **random** variable.

MARKOV AND CHEBYSHEV'S INEQUALITIES

- ▶ **Markov's inequality.** For a positive random variable X and constant $a > 0$

$$\Pr(X \geq a) \leq \frac{E(X)}{a}$$

MARKOV AND CHEBYSHEV'S INEQUALITIES

- ▶ **Markov's inequality.** For a positive random variable X and constant $a > 0$

$$\Pr(X \geq a) \leq \frac{E(X)}{a}$$

- ▶ Proof: $a \cdot I_{X \geq a} \leq X$. Then $E(a \cdot I_{X \geq a}) = a \cdot \Pr(X \geq a) \leq E(X)$.

MARKOV AND CHEBYSHEV'S INEQUALITIES

- ▶ **Markov's inequality.** For a positive random variable X and constant $a > 0$

$$\Pr(X \geq a) \leq \frac{E(X)}{a}$$

- ▶ Proof: $a \cdot I_{X \geq a} \leq X$. Then $E(a \cdot I_{X \geq a}) = a \cdot \Pr(X \geq a) \leq E(X)$.
- ▶ **Chebyshev's inequality.** Let Y be a random variable with finite mean m and variance σ^2 . Then

$$\Pr(|Y - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

MARKOV AND CHEBYSHEV'S INEQUALITIES

- ▶ **Markov's inequality.** For a positive random variable X and constant $a > 0$

$$\Pr(X \geq a) \leq \frac{E(X)}{a}$$

- ▶ Proof: $a \cdot I_{X \geq a} \leq X$. Then $E(a \cdot I_{X \geq a}) = a \cdot \Pr(X \geq a) \leq E(X)$.
- ▶ **Chebyshev's inequality.** Let Y be a random variable with finite mean m and variance σ^2 . Then

$$\Pr(|Y - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

- ▶ Proof: Use Markov's inequality with $X = (Y - m)^2$ and $a = \varepsilon^2$, and that $E(X) = E(Y - m)^2 = \sigma^2$. We then have

$$\Pr\left((Y - m)^2 \geq \varepsilon^2\right) \leq \frac{\sigma^2}{\varepsilon^2}$$

and therefore

$$\Pr(|Y - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

ALMOST SURE CONVERGENCE

- ▶ X_1, \dots, X_n and X are random variables on the same probability space.

DEF X_n converges **almost surely** (a.s.) to X as $n \rightarrow \infty$ iff

$$P(\{\omega : X_n(\omega) \rightarrow X(\omega) \text{ as } n \rightarrow \infty\}) = 1.$$

- ▶ Denoted by $X_n \xrightarrow{\text{a.s.}} X$.

ALMOST SURE CONVERGENCE

- ▶ X_1, \dots, X_n and X are random variables on the same probability space.

DEF X_n converges **almost surely** (a.s.) to X as $n \rightarrow \infty$ iff

$$P(\{\omega : X_n(\omega) \rightarrow X(\omega) \text{ as } n \rightarrow \infty\}) = 1.$$

- ▶ Denoted by $X_n \xrightarrow{\text{a.s.}} X$.
- ▶ For a given $\omega \in \Omega$, $X_n(\omega)$ ($n = 1, 2, \dots$) and $X(\omega)$ are real numbers (not random variables).
- ▶ Almost sure convergence: check if the sequence of real numbers $X_n(\omega)$ converges to the real number $X(\omega)$ for all ω , except those ω that have probability zero.

CONVERGENCE IN PROBABILITY

DEF X_n converges in probability to X as $n \rightarrow \infty$ iff

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$


► Denoted by $X_n \xrightarrow{P} X$.

CONVERGENCE IN PROBABILITY

DEF X_n converges in probability to X as $n \rightarrow \infty$ iff

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

► Denoted by $X_n \xrightarrow{P} X$.


 Let $X_n \sim \text{Beta}(n, n)$ show that $X_n \xrightarrow{P} \frac{1}{2}$ as $n \rightarrow \infty$.

CONVERGENCE IN PROBABILITY

DEF X_n converges in probability to X as $n \rightarrow \infty$ iff

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

► Denoted by $X_n \xrightarrow{P} X$.

 Let $X_n \sim \text{Beta}(n, n)$ show that $X_n \xrightarrow{P} \frac{1}{2}$ as $n \rightarrow \infty$.

Solution: $E(X_n) = \frac{n}{n+n} = \frac{1}{2}$ and

$$\text{Var}(X_n) = \frac{n \cdot n}{(n+n)^2(n+n+1)} = \frac{1}{4(2n+1)}.$$

By Chebyshev's inequality, for all $\varepsilon > 0$

$$\text{Pr}(|X_n - 1/2| \geq \varepsilon) \leq \frac{1}{4(2n+1)\varepsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

CONVERGENCE IN R-MEAN

DEF X_n converges in r -mean to X as $n \rightarrow \infty$ iff

$$E |X_n - X|^r \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

- Denoted by $X_n \xrightarrow{r} X$.

CONVERGENCE IN R-MEAN

DEF X_n converges in r -mean to X as $n \rightarrow \infty$ iff

$$E |X_n - X|^r \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

► Denoted by $X_n \xrightarrow{r} X$.

 Let X_n be a random variable with

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{2n} \quad \text{and} \quad P(X_n = -1) = \frac{1}{2n}.$$

Show that $X_n \xrightarrow{r} 0$ as $n \rightarrow \infty$.

CONVERGENCE IN R-MEAN

DEF X_n converges in r -mean to X as $n \rightarrow \infty$ iff

$$E |X_n - X|^r \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

► Denoted by $X_n \xrightarrow{r} X$.

 Let X_n be a random variable with

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{2n} \quad \text{and} \quad P(X_n = -1) = \frac{1}{2n}.$$

Show that $X_n \xrightarrow{r} 0$ as $n \rightarrow \infty$.

Solution: we have

$$\begin{aligned} E |X_n - X|^r &= |0 - 0|^r \cdot \left(1 - \frac{1}{n}\right) + |1 - 0|^r \cdot \frac{1}{2n} + |-1 - 0|^r \cdot \frac{1}{2n} \\ &= \frac{1}{n} \rightarrow 0. \end{aligned}$$

as $n \rightarrow \infty$ for all $r > 0$.

CONVERGENCE IN DISTRIBUTION

DEF X_n converges in distribution to X as $n \rightarrow \infty$ iff

$$F_{X_n}(x) \rightarrow F(x) \quad \text{as } n \rightarrow \infty$$

at all continuity points of X .

- Denoted by $X_n \xrightarrow{d} X$.

CONVERGENCE IN DISTRIBUTION

DEF X_n converges in distribution to X as $n \rightarrow \infty$ iff

$$F_{X_n}(x) \rightarrow F(x) \quad \text{as } n \rightarrow \infty$$

at all continuity points of X .

► Denoted by $X_n \xrightarrow{d} X$.

📎 Suppose $X_n \sim \text{Bin}(n, \lambda/n)$. Show that $X_n \rightarrow \text{Po}(\lambda)$ as $n \rightarrow \infty$.


CONVERGENCE IN DISTRIBUTION

DEF X_n converges in distribution to X as $n \rightarrow \infty$ iff

$$F_{X_n}(x) \rightarrow F(x) \quad \text{as } n \rightarrow \infty$$

at all continuity points of X .

► Denoted by $X_n \xrightarrow{d} X$.

 Suppose $X_n \sim \text{Bin}(n, \lambda/n)$. Show that $X_n \rightarrow \text{Po}(\lambda)$ as $n \rightarrow \infty$.

Solution: For fixed k we have

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

as $n \rightarrow \infty$

MORE ON CONVERGENCE

- Uniqueness: Theorem 6.2.1 tells us that if $X_n \rightarrow X$ and $X_n \rightarrow Y$, then $X = Y$ almost surely ($X \stackrel{d}{=} Y$ for convergence in distribution).

MORE ON CONVERGENCE

- Uniqueness: Theorem 6.2.1 tells us that if $X_n \rightarrow X$ and $X_n \rightarrow Y$, then $X = Y$ almost surely ($X \stackrel{d}{=} Y$ for convergence in distribution).
- The different notions of convergence are related as follows:

$$\begin{array}{ccccc} X_n \xrightarrow{a.s.} X & \Rightarrow & X_n \xrightarrow{p} X & \Rightarrow & X_n \xrightarrow{d} X \\ & & \uparrow & & \\ & & X_n \xrightarrow{r} X & & \end{array}$$

- So $\xrightarrow{a.s.}$ is stronger than \xrightarrow{p} which is stronger than \xrightarrow{d} .

CONVERGENCE VIA TRANSFORMS

- ▶ Let X, X_1, X_2, \dots be random variables. What if the moment generating function of X_n converges to the moment generation function of X ? Does that mean that X_n converges to X ?

CONVERGENCE VIA TRANSFORMS

- ▶ Let X, X_1, X_2, \dots be random variables. What if the moment generating function of X_n converges to the moment generation function of X ? Does that mean that X_n converges to X ?

TH Let X, X_1, X_2, \dots be random variables, and suppose that

$$\varphi_{X_n}(t) \rightarrow \varphi_X(t) \text{ as } n \rightarrow \infty$$

then

$$X_n \xrightarrow{d} X \text{ as } n \rightarrow \infty.$$

TH The converse also holds. If $X_n \xrightarrow{d} X$, then $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$.

CONVERGENCE VIA TRANSFORMS

- ▶ Let X, X_1, X_2, \dots be random variables. What if the moment generating function of X_n converges to the moment generation function of X ? Does that mean that X_n converges to X ?

TH Let X, X_1, X_2, \dots be random variables, and suppose that

$$\varphi_{X_n}(t) \rightarrow \varphi_X(t) \text{ as } n \rightarrow \infty$$

then

$$X_n \xrightarrow{d} X \text{ as } n \rightarrow \infty.$$

TH The converse also holds. If $X_n \xrightarrow{d} X$, then $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$.

- ▶ Similar theorems hold for the generating function and moment generating function (Th 6.4.1-6.4.3).

LAW OF LARGE NUMBERS - SOME PRELIMINARIES

- ▶ Let X_1, \dots, X_n be independent variables with mean μ and variance σ^2 .
- ▶ Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ be the sample mean of n observations.
- ▶ We then have

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu$$

and

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

LAW OF LARGE NUMBERS

- ▶ **(Weak) law of large numbers.** Let X_1, \dots, X_n be independent variables with mean μ and finite variance σ^2 . Then

$$\bar{X}_n \xrightarrow{P} \mu.$$

LAW OF LARGE NUMBERS

- ▶ **(Weak) law of large numbers.** Let X_1, \dots, X_n be independent variables with mean μ and finite variance σ^2 . Then

$$\bar{X}_n \xrightarrow{P} \mu.$$

- ▶ Proof: By Chebychev's inequality

$$Pr(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

- ▶ This version of the law of large numbers requires a population variance which is finite. Theorem 6.5.1 gives a version where only the mean needs to be finite.

LAW OF LARGE NUMBERS

- ▶ **(Weak) law of large numbers.** Let X_1, \dots, X_n be independent variables with mean μ and finite variance σ^2 . Then

$$\bar{X}_n \xrightarrow{P} \mu.$$

- ▶ Proof: By Chebychev's inequality

$$Pr(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

- ▶ This version of the law of large numbers requires a population variance which is finite. Theorem 6.5.1 gives a version where only the mean needs to be finite.
- ▶ The strong law of large numbers proves that $\bar{X}_n \xrightarrow{a.s.} \mu$ if the mean is finite.

LAW OF LARGE NUMBERS

- ▶ **(Weak) law of large numbers.** Let X_1, \dots, X_n be independent variables with mean μ and finite variance σ^2 . Then

$$\bar{X}_n \xrightarrow{P} \mu.$$

- ▶ Proof: By Chebychev's inequality

$$Pr(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

- ▶ This version of the law of large numbers requires a population variance which is finite. Theorem 6.5.1 gives a version where only the mean needs to be finite.
- ▶ The strong law of large numbers proves that $\bar{X}_n \xrightarrow{a.s.} \mu$ if the mean is finite.
- ▶ The assumption of a finite mean is important. Example: if X_1, X_2, \dots are independent $C(0, 1)$, then $\bar{X}_n \stackrel{d}{=} X_1$ for all n . The law of large numbers does not hold.

CENTRAL LIMIT THEOREM

TH Let X_1, X_2, \dots be iid random variables with finite expectation μ and variance σ^2 . Then

$$\left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \right) \xrightarrow{d} N(0, 1) \text{ as } n \rightarrow \infty.$$

CENTRAL LIMIT THEOREM

TH Let X_1, X_2, \dots be iid random variables with finite expectation μ and variance σ^2 . Then

$$\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right) \xrightarrow{d} N(0, 1) \text{ as } n \rightarrow \infty.$$

► Proof by showing that

$$\varphi_{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}(t) \rightarrow \varphi_{N(0,1)}(t) = e^{-t^2/2}.$$

CENTRAL LIMIT THEOREM

TH Let X_1, X_2, \dots be iid random variables with finite expectation μ and variance σ^2 . Then

$$\left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \right) \xrightarrow{d} N(0, 1) \text{ as } n \rightarrow \infty.$$

► Proof by showing that

$$\varphi_{\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}}(t) \rightarrow \varphi_{N(0,1)}(t) = e^{-t^2/2}.$$

► Application: empirical distribution function

$$F_n(x) = \frac{\#\text{observations} \leq x}{n}$$

then as $n \rightarrow \infty$

$$F_n(x) \xrightarrow{p} F(x)$$

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} N(0, \sigma^2(x)), \quad \sigma^2(x) = F(x)[1 - F(x)].$$

CONVERGENCE OF SUMS OF SEQUENCES OF RVs

TH If $X_n \rightarrow X$ and $Y_n \rightarrow Y$, then $X_n + Y_n \rightarrow X + Y$.

- ▶ Holds for *a.s.*, *p* and *r*-convergence without assuming independence.
- ▶ The theorem also holds for *d*-convergence if we assume independence.

CONVERGENCE OF SUMS OF SEQUENCES OF RVs

TH If $X_n \rightarrow X$ and $Y_n \rightarrow Y$, then $X_n + Y_n \rightarrow X + Y$.

- ▶ Holds for *a.s.*, *p* and *r*-convergence without assuming independence.
- ▶ The theorem also holds for *d*-convergence if we assume independence.

TH If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, where a is a constant, then as $n \rightarrow \infty$

$$X_n + Y_n \xrightarrow{d} X + a$$

$$X_n - Y_n \xrightarrow{d} X - a$$

$$X_n \cdot Y_n \xrightarrow{d} X \cdot a$$

$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a} \text{ for } a \neq 0$$

CONVERGENCE OF SUMS OF SEQUENCES OF RVs

TH If $X_n \rightarrow X$ and $Y_n \rightarrow Y$, then $X_n + Y_n \rightarrow X + Y$.

- ▶ Holds for *a.s.*, *p* and *r*-convergence without assuming independence.
- ▶ The theorem also holds for *d*-convergence if we assume independence.

TH If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, where a is a constant, then as $n \rightarrow \infty$

$$X_n + Y_n \xrightarrow{d} X + a$$

$$X_n - Y_n \xrightarrow{d} X - a$$

$$X_n \cdot Y_n \xrightarrow{d} X \cdot a$$

$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a} \text{ for } a \neq 0$$

 Let X_1, X_2, \dots be independent $U(0, 1)$. Show that

$$\frac{X_1 + X_2 + \dots + X_n}{X_1^2 + X_2^2 + \dots + X_n^2} \xrightarrow{p} \frac{3}{2} \text{ as } n \rightarrow \infty.$$

CONVERGENCE OF FUNCTIONS OF CONVERGENT RVs


TH Let X_1, X_2, \dots be random variables such that $X_n \xrightarrow{p} a$ for some constant a . Let $g(\cdot)$ be a function which is continuous at a . Then

$$g(X_n) \xrightarrow{p} g(a).$$

CONVERGENCE OF FUNCTIONS OF CONVERGENT RVs

TH Let X_1, X_2, \dots be random variables such that $X_n \xrightarrow{P} a$ for some constant a . Let $g(\cdot)$ be a function which is continuous at a . Then


$$g(X_n) \xrightarrow{P} g(a).$$

 Let X_1, X_2, \dots be iid random variables with finite mean $\mu \geq 0$. Show that $\sqrt{\bar{X}_n} \xrightarrow{P} \sqrt{\mu}$ as $n \rightarrow \infty$.

CONVERGENCE OF FUNCTIONS OF CONVERGENT RVs

TH Let X_1, X_2, \dots be random variables such that $X_n \xrightarrow{P} a$ for some constant a . Let $g(\cdot)$ be a function which is continuous at a . Then

$$g(X_n) \xrightarrow{P} g(a).$$


 Let X_1, X_2, \dots be iid random variables with finite mean $\mu \geq 0$. Show that $\sqrt{\bar{X}_n} \xrightarrow{P} \sqrt{\mu}$ as $n \rightarrow \infty$.

Solution: from the law of large numbers we have $\bar{X}_n \xrightarrow{P} \mu$. Since $g(x) = \sqrt{x}$ is continuous at $x = \mu$ the above theorem proves that $\sqrt{\bar{X}_n} \xrightarrow{P} \sqrt{\mu}$ as $n \rightarrow \infty$.


CONVERGENCE OF FUNCTIONS OF CONVERGENT RVs

TH Let X_1, X_2, \dots be random variables such that $X_n \xrightarrow{P} a$ for some constant a . Let $g(\cdot)$ be a function which is continuous at a . Then

$$g(X_n) \xrightarrow{P} g(a).$$

 Let X_1, X_2, \dots be iid random variables with finite mean $\mu \geq 0$. Show that $\sqrt{\bar{X}_n} \xrightarrow{P} \sqrt{\mu}$ as $n \rightarrow \infty$.

Solution: from the law of large numbers we have $\bar{X}_n \xrightarrow{P} \mu$. Since $g(x) = \sqrt{x}$ is continuous at $x = \mu$ the above theorem proves that $\sqrt{\bar{X}_n} \xrightarrow{P} \sqrt{\mu}$ as $n \rightarrow \infty$.

 Let $Z_n \sim N(0, 1)$ and $V_n \sim \chi^2(n)$ be independent RVs. Show that

$$T_n = \frac{Z_n}{\sqrt{\frac{V_n}{n}}} \stackrel{d}{\sim} N(0, 1) \text{ as } n \rightarrow \infty.$$