Probability Theory

Lecture 6

Per Sidén

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 $X_1, X_2, ...,$ also denoted by $\{X_n\}_{n=1}^{\infty}$.
- Example: what can we say about the sample mean $X_n = n^{-1} \sum_{i=1}^{n} Y_i$ 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

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

and therefore

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

Almost sure convergence

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

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

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

- Denoted by $X_n \overset{a.s.}{\to} X$.

Almost sure convergence

- $X_1, ...X_n$ and X are random variables on the same probability space.

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

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

- Denoted by $X_n \stackrel{a.s.}{\to} X$.

- For a given $\omega \in \Omega$, $X_n(\omega)$ $(n = 1, 2, ...)$ 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 $n \to \infty$ iff

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

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

Convergence in probability

Def X_n **converges in probability** to X as $n \to \infty$ iff

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

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

✏ Let $X_n \sim Beta(n, n)$ show that $X_n \xrightarrow{p} \frac{1}{2}$ as $n \to \infty$.

Convergence in probability

Def X_n **converges in probability** to X as $n \to \infty$ iff

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

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

✏ Let $X_n \sim Beta(n, n)$ show that $X_n \xrightarrow{p} \frac{1}{2}$ as $n \to \infty$.

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

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

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

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

Convergence in r-mean

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

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

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

Convergence in r-mean

Def X_n **converges in r-mean** to X as $n \to \infty$ iff

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

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

✏ Let X_n be a random variable with

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

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

Convergence in r-mean

Def X_n **converges in** r**-mean** to X as $n \to \infty$ iff

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

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

✏ Let X_n be a random variable with

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

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

Solution: we have

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

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

Convergence in distribution

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

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

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 $n \to \infty$ iff

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

at all continuity points of X.

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

✏ Suppose $X_n \sim Bin(n, \lambda/n)$. Show that $X_n \to Po(\lambda)$ as $n \to \infty$.

Convergence in distribution

Def X_n **converges in distribution** to X as $n \to \infty$ iff

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

at all continuity points of X.

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

✏ Suppose $X_n \sim Bin(n, \lambda/n)$. Show that $X_n \to Po(\lambda)$ as $n \to \infty$.

Solution: For fixed k we have

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

as $n \to \infty$

More on convergence

- Uniqueness: Theorem 6.2.1 tells us that if $X_n \to X$ and $X_n \to Y$, 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 $X_n \to X$ and $X_n \to Y$, 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 $X, X_1, X_2, \ldots$ 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 $X, X_1, X_2, \ldots$ 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 $X, X_1, X_2, \ldots$ be random variables, and suppose that

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

then

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

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

Convergence via transforms

- Let $X, X_1, X_2, \ldots$ 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 $X, X_1, X_2, \ldots$ be random variables, and suppose that

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

then

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

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

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

Law of large numbers - some preliminaries

- Let $X_1, .., X_n$ 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

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

Law of large numbers

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

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

- Proof: By Chebychev's inequality

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

- 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 $X_1, .., X_n$ be independent variables with mean μ and finite variance σ^2. Then

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

- Proof: By Chebychev's inequality

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

- 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 $X_1, .., X_n$ be independent variables with mean μ and finite variance σ^2. Then

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

- Proof: By Chebychev's inequality

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

- 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 $X_1, X_2, \ldots$ 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 $X_1, X_2, \ldots$ be iid random variables with finite expectation μ and variance σ^2. Then

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

Central limit theorem

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

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

- Proof by showing that

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

Central limit theorem

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

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

- Proof by showing that

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

- Application: empirical distribution function

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

then as $n \to \infty$

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

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

Convergence of sums of sequences of RVs

Th If $X_n \to X$ and $Y_n \to Y$, then $X_n + Y_n \to X + Y$.

- 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 $X_n \to X$ and $Y_n \to Y$, then $X_n + Y_n \to X + Y$.

- 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 $n \to \infty$

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

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

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

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

Convergence of sums of sequences of RVs

Th If $X_n \to X$ and $Y_n \to Y$, then $X_n + Y_n \to X + Y$.

- 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 $n \to \infty$

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

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

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

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

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

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

Convergence of functions of convergent RVs

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

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

Convergence of functions of convergent RVs

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

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

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

Convergence of functions of convergent RVs

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

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

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

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 $n \to \infty$.

Convergence of functions of convergent RVs

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

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

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

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 $n \to \infty$.

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

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