Selected Topics of Chapter 7: Limiting Distributions (and some Section 6.5)

Shih-Kang Chao
Department of Statistics
University of Missouri-Columbia

2019 Spring

Sequence of random variables

Statistical applications often involve data, $\{X_1, X_2, ..., X_n\}$ independent and identically distributed, i.e. they are independent and having the same cdf (and pdf)

- Body heights of high school students
- Housing prices
- Annual income

$\vdots$

Random sample

Definition (See Section 4.6 for details)

If random variables $X_1, X_2, ..., X_n$ are independent and identically distributed (i.i.d.), then $\{X_1, X_2, ..., X_n\}$ is called a random sample of size n

- To make sense of data, the first step is to compute the *sample average*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Is it a good estimate of the mean of X?

- For a moment, let's consider a general sequence of r.v.'s $Y_1, Y_2, \ldots Y_n$, and introduce some useful notions. Later, we will specialize the discussion on $Y_n = \bar{X}_n$

Converge in distribution

Definition (Definition 7.2.1)

$G_n(y) = P(Y_n \leq y)$ converges to $G(y)$ in distribution if

$$\lim_{n \to \infty} G_n(y) = G(y),$$

for all values y at which $G(y)$ is continuous, denoted by

$$Y_n \xrightarrow{d} Y$$

The distribution corresponding to the CDF $G(y)$ is called the limiting distribution of Y_n.

Smallest order statistics (Section 6.5)

An ideal situation: cdf $F(x)$ of X_i is known (all have same cdf).

$$X_{1:n} = \min\{X_1, X_2, ..., X_n\} \text{ : the smallest order statistics}$$

The CDF of $X_{1:n}$ is

$$\begin{aligned}
&P(X_{1:n} \leq x) \\
&= 1 - P(X_{1:n} > x) \\
&= 1 - P\big(\{X_1 > x\} \cap \{X_2 > x\} \cap ... \cap \{X_n > x\}\big) \\
&\qquad\qquad [\text{if the smallest of } X_i\text{'s is} > x, \text{ every single } X_i > x] \\
&= 1 - P(X_1 > x)P(X_2 > x)...P(X_n > x) \quad [\text{all } X_i \text{ independent}] \\
&= 1 - (1 - F(x))^n.
\end{aligned}$$

Hence, the CDF of $X_{1:n}$ is

$$1 - (1 - F(x))^n$$

Largest order statistics (Section 6.5)

$$X_{n:n} = \max\{X_1, X_2, ..., X_n\} : \text{the largest order statistics}$$

The CDF of $X_{n:n}$ is

$$\begin{aligned}
&P(X_{n:n} \leq x) \\
&= P\big(\{X_1 \leq x\} \cap \{X_2 \leq x\} \cap ... \cap \{X_n \leq x\}\big) \\
&\qquad\qquad [\text{if the largest of } X_i\text{'s is} \leq x, \text{ every single } X_i \leq x] \\
&= P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x) \\
&= F(x)^n.
\end{aligned}$$

Hence, the CDF of $X_{n:n}$ is

$$F(x)^n$$

Example of convergence in distribution

$F(x) = 1 - (1+x)^{-1}$ the Pareto(1,1) distribution.
$Y_n = nX_{1:n} = n * \min\{X_1, X_2, ..., X_n\}$.
If $y \le 0$, $P(Y_n \le y) = 0$ (why?). If $y > 0$,

$$\begin{aligned} P(Y_n \le y) &= P\Big(\min\{X_1, X_2, ..., X_n\} \le \frac{y}{n}\Big) \\ &= 1 - (1 - F(y/n))^n \\ &= 1 - \Big(1 + \frac{y}{n}\Big)^{-n} \end{aligned}$$

Calculus:

$$\Big(1 + \frac{y}{n}\Big)^{-n} \to e^{-y}, \quad \text{as } n \to \infty$$

Hence, as $n \to \infty$

$$P(Y_n \le y) \to \text{ the cdf of Exp(1)},$$

so, $Y_n \xrightarrow{d} \text{Exp}(1)$ as $n \to \infty$

Convergence in probability

Another notion of convergence is stronger than the convergence in distribution.

Definition (Definition 7.7.1)

Infinite sequence $Y_1, Y_2, \ldots$ is said to converge in probability to a random variable Y, if

$$\lim_{n\to\infty} P(|Y_n - Y| > \epsilon) = 0$$

Denote by $Y_n \xrightarrow{P} Y$

- Sometimes the limit Y can be a fixed constant c instead of a random variable

$\xrightarrow{P}$ is stronger than $\xrightarrow{d}$

Theorem (Theorem 7.7.1)

$$\text{If } Y_n \xrightarrow{P} Y, \text{ then } Y_n \xrightarrow{d} Y \tag{1}$$

In general, the converse of (1) is false. The converse of (1) is only true when $Y = c$ is a fixed constant, i.e. if $Y_n \xrightarrow{d} c$ then $Y_n \xrightarrow{P} c$.

Example of converging in probability

$X_1, X_2, ...X_n$ are independent and identically distributed. Suppose X_i's share the same mean μ and variance σ^2. However, they are unknown to us. The distribution of X_i is also unknown. Let

$$Y_n = \bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{sample average}$$

Distribution of Y_n is unknown, because we do not know the distribution of X_i's

$$\mathbb{E}[Y_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

$$Var\,[Y_n] = \frac{1}{n^2}\sum_{i=1}^{n} Var\,[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

Example of converging in probability

Recall that $\mathbb{E}[\bar{X}_n] = \mu$ and $Var\,[\bar{X}_n] = \dfrac{\sigma^2}{n}$, for any $\epsilon > 0$,

$$P(|\bar{X}_n - \mu| > \epsilon) = P\left(|\bar{X}_n - \mu| > \underbrace{\frac{\epsilon}{\cancel{\sqrt{Var\,(\bar{X}_n)}}}}_{\text{viewed as "}k\text{"}} \cancel{\sqrt{Var\,(\bar{X}_n)}}\right)$$

$$\le \frac{1}{k^2} \quad [\text{Chebychev's inequality (Sect. 2.4)}]$$

$$= \frac{\sigma^2}{n\epsilon^2}.$$

Hence, by squeezing,

$$0 \le \lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) \le \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0,$$

so

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0 \quad \text{that is, } \bar{X}_n \xrightarrow{P} \mu$$

Amazing things about the deduction above

- This is the (weak) law of large number (WLLN) (Theorem 7.6.2):

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu, \quad \text{as } n \to \infty$$

- in words, sample average is close to μ when huge data size n is available, so using $\bar{X}_n$ to estimate mean μ makes perfect sense!
- This holds even when the distribution of X is unknown to us

- Alternative view with normal distribution: see example 7.2.7

Estimation error

- Ok, now we know $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is close to μ when "huge" data size n is available
- But, say, I only have $n = 50$, this does not seem too large

Can we quantify the difference between $\bar{X}_n$ and μ for small n?

A solution is to use the square root of $Var\left[\bar{X}_n\right] = \dfrac{\sigma^2}{n}$ to measure the error, e.g.

$$\left|\bar{X}_n - \mu\right| \le q\sqrt{Var\left[\bar{X}_n\right]} = q\frac{\sigma}{\sqrt{n}} \tag{2}$$

with high probability if the constant $q > 0$ is large.

How large should q be?

- The answer depends on how much confidence one wants for (2) to be true
- Exact distribution of $\bar{X}_n$ is unknown since the distribution of X_i's are unknown

Central limit theorem (CLT)

Theorem (Theorem 7.3.1)

If $\{X_1, X_2, ..., X_n\}$ *are i.i.d. with mean* $\mathbb{E}[X] = \mu$ *and* $Var\,[X] = \sigma^2 < \infty$, *then*

$$Z_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \text{ as } n \to \infty. \tag{3}$$

- Although $n \to \infty$ in (3), Z usually approximates the distribution of Z_n well when n is mildly large, e.g. $n = 30$, and approximation is best when the distribution of X is symmetric
- The variance condition excludes Pareto distribution with $\kappa \leq 2$, so CLT does not hold for *all* distributions

Sketch of the proof of CLT

- Moment generating function of Z_n is (after Taylor expansion and some calculations)

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + \frac{d(n)}{n}\right)^n$$

where $\lim_{n\to\infty} d(n) \to 0$ is the remainder term

- By calculus, as $n \to \infty$,

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + \frac{d(n)}{n}\right)^n \to e^{\frac{t^2}{2}} = M_Z(t),$$

where $M_Z(t) = E[e^{tZ}]$ is the moment generating function of $Z \sim \mathcal{N}(0, 1)$

Now, using CLT to find a reasonable q in (2).

Divide both sides of (2) by $\frac{\sigma}{\sqrt{n}}$:

$$\left| \underbrace{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}_{\xrightarrow{d} Z \sim \mathcal{N}(0,1) \text{ by CLT}} \right| \leq q$$

so, we find q that satisfies

$$|Z| \leq q \quad \text{or} \quad -q < Z < q \tag{4}$$

with high probability $1 - \alpha$ (or small $\alpha > 0$)

$$
\begin{aligned}
1-\alpha &\overset{(4)}{=} P(-q<Z<q) \\
&= P(Z<q)-P(Z<-q) \\
&= P(Z<q)-(1-P(Z<q)) \\
&\qquad \text{symmetry of } Z\text{: } P(Z>-q)=P(Z<q) \\
&= 2P(Z<q)-1
\end{aligned}
$$

So, q should satisfy

$$
P(Z<q)=1-\frac{\alpha}{2} \quad \text{or} \quad q=z_{1-\frac{\alpha}{2}}
$$

γ	0.90	0.95	0.975	0.99	0.995	0.999
z_γ	1.282	1.645	1.960	2.326	2.576	3.090

Figure: The z_γ such that $P(Z<z_\gamma)=\gamma$ (bottom of Table 3).

Confidence interval

Plugging $q = z_{1-\frac{\alpha}{2}}$ in (2), we have

$$\left|\bar{X}_n - \mu\right| \leq z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}} \text{ with confidence } 1-\alpha$$

Breaking the absolute value yields the confidence interval for μ:

$$\mu \in \left[\bar{X}_n \pm z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}\right] \text{ with confidence } 1-\alpha \tag{5}$$

Say, $\alpha = 10\%$, $\gamma = 1 - \alpha/2 = 0.95$, and $z_\gamma = 1.645$. Suppose our data have size $n = 50$ with $\sigma = 2$

$$\mu \in \left[\bar{X}_n - 0.4653, \bar{X}_n + 0.4653\right]$$

with probability 90%.

$$\mu \in \left[\bar{X}_n \pm z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}\right] \text{ with confidence } 1-\alpha$$

The confidence interval is wider when

- confidence is high (α small and $\gamma = 1 - \alpha/2$ high)

γ	0.90	0.95	0.975	0.99	0.995	0.999
z_γ	1.282	1.645	1.960	2.326	2.576	3.090

- $\sigma^2 = Var(X)$ is large (more uncertainty in the data)

A caveat: σ is typically not known to practitioners, so one has to guess a σ before using (5)

We can deal with it by finding an estimator $\widehat{\sigma}^2$ of σ^2 that depends on the data $\{X_1, X_2, ..., X_n\}$ such that

$$\widehat{\sigma}^2 \xrightarrow{P} \sigma^2, \quad \text{as } n \to \infty$$

This can be achieved by setting

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

which is the sample variance

To determine q, can we have a modification of the CLT (3) such that

$$\frac{\bar{X}_n - \mu}{\widehat{\sigma}/\sqrt{n}} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \text{ as } n \to \infty? \tag{6}$$

The answer is YES!

Slutsky's Theorem

Theorem (Theorem 7.7.4)

If $W_n \xrightarrow{P} c$ *and* $V_n \xrightarrow{d} V$*, then*

1. $W_n + V_n \xrightarrow{d} c + V$
2. $W_n V_n \xrightarrow{d} cV$
3. $V_n / W_n \xrightarrow{d} V/c$*, if* $c \neq 0$

(6) is warranted by 2. of Slutsky's theorem by viewing

$$\frac{\bar{X}_n - \mu}{\widehat{\sigma}/\sqrt{n}} = \underbrace{\frac{\bar{X}_n - \mu}{\cancel{\sigma/\sqrt{n}}}}_{= V_n} \underbrace{\frac{\cancel{\sigma/\sqrt{n}}}{\widehat{\sigma}/\sqrt{n}}}_{= W_n} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \qquad (7)$$

and $V_n \xrightarrow{d} Z$ and $W_n \xrightarrow{P} 1$

Confidence interval with unknown σ

The confidence interval holds with probability $1 - \alpha$:

$$\left|\bar{X}_n - \mu\right| \leq z_{1-\alpha/2} * \frac{\widehat{\sigma}}{\sqrt{n}} \tag{8}$$

where $z_{1-\alpha/2}$ is determined as before

- This approach may require n to be larger than that needed for the case of known σ
- In practice, n may be small. A more refined confidence interval of using t distribution for q will be covered in the continuation of this course

Appendix: Chebychev's inequality

Theorem (Theorem 2.4.7)

If X is a random variable with mean μ and variance σ^2, then for any $k > 0$,

$$P\left(|X - \mu| > k\sigma\right) = P\left(|X - \mu| > k\sqrt{Var\,(X)}\right) \leq \frac{1}{k^2}$$

▸ Back to convergence in probability