Class 22: Limit theorems

Tuesday, December 3, 2019   11:37 AM

- HW 8 due Fri, Dec 6

Today: (and basically the field of statistics)
What to do with iid samples.

Let $X_1, X_2, \dots$ be iid random variables.
independent, identically distributed

Recall:

① $\mathbb{E}[X_1 + X_2 + \dots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n]$

↳ (true without independence)

② $Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$

↳ does require independence

- Even without independence:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} Cov(X_i, X_j)$$

Similar to: $\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j$

since $Cov(X, X) = Var(X)$

Proof (n=2 case):

$$Var(X_1 + X_2) = \mathbb{E}[(X_1 + X_2)^2] - (\mathbb{E}[X_1 + X_2])^2$$

$$= \mathbb{E}[X_1^2 + X_2^2 + 2X_1X_2] - \left((\mathbb{E}[X_1])^2 + (\mathbb{E}[X_2])^2 + 2\mathbb{E}[X_1]\mathbb{E}[X_2]\right)$$

$$= \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + 2\mathbb{E}[X_1X_2] - (\dots)$$

$$= \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2$$

$$+ \mathbb{E}[X_2^2] - (\mathbb{E}[X_2])^2$$

$$= \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2$$
$$+ \mathbb{E}[X_2^2] - (\mathbb{E}[X_2])^2$$
$$+ 2\left(\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]\right)$$
$$= Var(X_1) + Var(X_2) + 2\,Cov(X_1, X_2)$$
$$= Cov(X_1, X_1) + Cov(X_2, X_2) + Cov(X_1, X_2) + Cov(X_2, X_1)$$
$$= \sum_{i=1}^{2}\sum_{j=1}^{2} Cov(X_i, X_j)$$

- **Corollary**: If $Cov(X_i, X_j) = 0$ (Pairwise uncorrelated) for all $i \neq j$, then
$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n).$$

- **Simple case**: $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$.

Fact: If a r.v. X satisfies $Var(X) = 0$, then it must be constant. That is, there exists $c \in \mathbb{R}$ such that $\mathbb{P}(X = c) = 1$.

Proof:
If $0 = Var(X) = \mathbb{E}[\underbrace{(X - \mathbb{E}[X])^2}_{\geq 0}]$ then $\underbrace{X - \mathbb{E}[X]}_{c \in \mathbb{R}} = 0$.

Law of Large Numbers (LLN)

Let $X_1, X_2, \ldots$ be iid with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$.

Define the sample average $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ for each $n = 1, 2, \ldots$

Facts:

Facts:

① $\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n} m = m.$

"unbiased estimator of m"

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

$Var(cX) = c^2 Var(X)$

$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$

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

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

This decreases to 0 as $n \to \infty$.

So $\bar{X}_n$ is "roughly constant" as $n \to \infty$.

Theorem (weak LLN):

Let $X_1, X_2, \dots$ be iid with mean $m \in \mathbb{R}$ and finite variance.

Then $Var(\bar{X}_n) \to 0$ as $n \to \infty$, and

$$\lim_{n\to\infty} \mathbb{P}\left(\underbrace{|\bar{X}_n - m|}_{\text{sample error}} > \underbrace{a}_{\text{error threshold}}\right) = 0, \quad \text{for any } a > 0.$$

Ex: Flip n coins. Mark $X_i = \begin{cases} 1 & \text{if } i^{th} \text{ flip is heads} \\ 0 & \text{if } i^{th} \text{ flip is tails.} \end{cases}$

Then X_i are iid $Ber(1/2)$.

So $m = \mathbb{E}[X_i] = 1/2$, $\sigma^2 = Var(X_i) = \frac{1}{4}$.

Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ = fraction of n flips which show heads.

The LLN says basically that $\bar{X}_n \simeq \frac{1}{2}$ if n large.

The LLN says basically that $\bar{X}_n \approx \frac{1}{2}$ if n lage.

Or: $\mathbb{P}\left(|\bar{X}_n - \tfrac{1}{2}| > a \right) \to 0$ as $n \to \infty$, for $a > 0$.

Thm (Markov's inequality):

If X is a non-negative r.v. and $a > 0$,

then $\mathbb{P}(X > a) \leq \dfrac{\mathbb{E}[X]}{a}$.

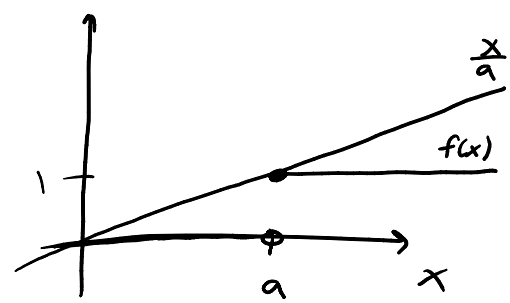
Proof:

Define a function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \begin{cases} 1 & \text{if } x > a \\ 0 & \text{if } x \leq a \end{cases}$.

(indicator function)

Then

$$\mathbb{E}[f(X)] = 1 \cdot \mathbb{P}(X > a) + 0 \cdot \mathbb{P}(X \leq a)$$
$$= \mathbb{P}(X > a).$$



But also $f(x) \leq \frac{x}{a}$ for all $x \geq 0$.

Thus

$$\mathbb{P}(X > a) = \mathbb{E}[f(X)] \leq \mathbb{E}\left[\frac{X}{a}\right] = \frac{\mathbb{E}[X]}{a}.$$

Proof of weak LLN:

We already saw that $Var(\bar{X}_n) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$.

Let $a > 0$. Then

$$\mathbb{P}(|\bar{X}_n - m| > a) = \mathbb{P}(|\bar{X}_n - m|^2 > a^2)$$

$$\leq \frac{\mathbb{E}[|\bar{X}_n - m|^2]}{a^2}$$

by Markov's inequality with $X = |\bar{X}_n - m|^2$ and $a \to a^2$

$$= \mathbb{E}[(\bar{X}_n - \mathbb{E}[\bar{X}_n])^2]$$

$$a^2$$

and $a \to a^2$

$$= \frac{\mathbb{E}[(\bar{X}_n - \mathbb{E}[\bar{X}_n])^2]}{a^2} \qquad \text{since } \mathbb{E}[\bar{X}_n] = m$$

$$= \frac{\mathrm{Var}(\bar{X}_n)}{a^2} \qquad \text{(sometimes called Chebyshev's inequality)}$$

$$= \frac{\sigma^2}{n a^2} \to 0 \quad \text{as } n \to \infty.$$

<u>Strong LLN</u>:

$$\mathbb{P}\left(\lim_{n \to \infty} \bar{X}_n = m \right) = 1.$$

LLN says roughly that $\bar{X}_n \approx m$.

The <u>central limit theorem</u> (CLT) tells us what the <u>errors</u> $(\bar{X}_n - m)$ look like.

<u>Theorem (CLT)</u>:

Let $X_1, X_2, \ldots$ be iid with mean $m \in \mathbb{R}$ and finite variance $\sigma^2 > 0$.

Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $Z_n = \sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right)$.

Then, for any $x \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathbb{P}(Z_n \le x) = \mathbb{P}(Z \le x) \quad \text{where } Z \sim N(0,1).$$

standard normal

Equivalently, note $\mathbb{P}(Z \le x) = \Phi(x) =$ CDF of standard normal

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, dy.$$

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

① Rearrange $\bar{Z}_n = \sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right)$ to get

$$\bar{X}_n = m + \frac{\sigma}{\sqrt{n}} \bar{Z}_n.$$

(Recall LLN $\Rightarrow$ $\bar{X}_n \simeq m$.)

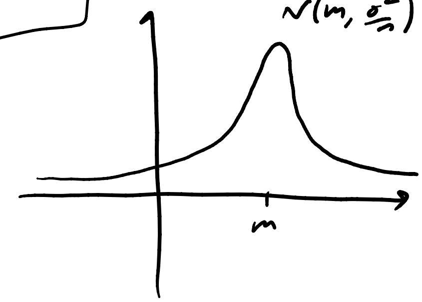
CLT says $\bar{Z}_n \simeq N(0,1)$.

Or in other words, since $m + \frac{\sigma}{\sqrt{n}} N(0,1) = m + N(0, \frac{\sigma^2}{n})$

$$= N(m, \frac{\sigma^2}{n}),$$

we have $\bar{X}_n \simeq N(m, \frac{\sigma^2}{n})$.

PDF of $N(m, \frac{\sigma^2}{n})$



② This works for ANY iid r.v.'s X_i (with $\sigma^2 < \infty$).

(universality)

③ Rewrite

$$\bar{Z}_n = \sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right) = \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - m}{\sigma}\right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right).$$

CLT says:

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sigma} \le x\right) = \mathbb{P}(Z \le x), \quad Z \sim N(0,1).$$

④ Note

$$E[\ldots] = E\left[\frac{1}{n}\sum (X_i - m)\right] = \frac{1}{n}\sum E[X_i] - m$$

④ Note

$$E[\bar{Z}_n] = \mathbb{E}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right)\right] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\mathbb{E}[X_i] - m}{\sigma} = 0$$

since $E[X_i] = m$.

and

$$\begin{aligned} Var(\bar{Z}_n) &= Var\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right)\right) \\ &= \frac{1}{n} Var\left(\sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right)\right) \\ &= \frac{1}{n} \sum_{i=1}^{n} Var\left(\frac{X_i - m}{\sigma}\right) && \leftarrow \text{ since iid} \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma^2} Var(X_i) && \leftarrow \text{ using } Var(aX+b) = a^2 Var(X) \text{ with } a = \tfrac{1}{\sigma},\ b = -\tfrac{m}{\sigma}. \\ &= \frac{1}{n} \sum_{i=1}^{n} 1 \\ &= 1. \end{aligned}$$

Sometimes call $\bar{Z}_n$ the "rescaled"/"normalized" error, since it has mean 0 and variance 1.

⑤ More generally: For $Z \sim N(0,1)$

$$\lim_{n \to \infty} \mathbb{P}(a \le \bar{Z}_n \le b) = \mathbb{P}(a \le Z \le b).$$

In particular,

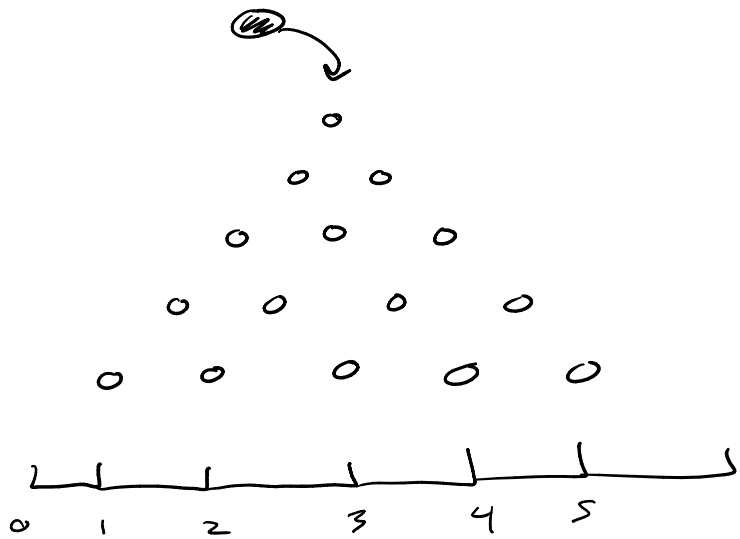
$$\lim_{n \to \infty} \mathbb{P}(|\bar{Z}_n| \le a) = \mathbb{P}(|Z| \le a).$$

Ex: Flip n coins, so $X_i \sim Ber(\frac{1}{2})$.

Then $\sum_{i=1}^{n} X_i \sim Bin(n, \frac{1}{2})$.

Galton Board: https://galtonboard.com/Content/Videos/galton-board-large_1_mobile_1.mp4

Galton Board: https://galtonboard.com/Content/Videos/galton-board-large_1_mobile_1.mp4



success = move right
failure = move left

CLT says $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sigma} \approx N(0,1)$.

Here $m = E[X_i] = \frac{1}{2}$, $\sigma^2 = Var(X_i) = \frac{1}{4}$, $\sigma = \frac{1}{2}$

So CLT says $\hat{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \frac{1}{2}}{1/2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (2X_i - 1)$

$$= \frac{1}{\sqrt{n}} \left(2\left(\sum_{i=1}^{n} X_i\right) - n \right).$$

Thus $\frac{\sqrt{n}\,\hat{Z}_n + n}{2} = \sum_{i=1}^{n} X_i \sim Bin(n, \frac{1}{2})$

(CLT) $\{$

$$\frac{\sqrt{n}\,N(0,1) + n}{2} \approx \frac{N(0,n) + n}{2}$$

$$= N\left(\frac{n}{2}, \frac{n}{4}\right).$$

i.e. $B(n, \frac{1}{2}) \approx N\left(\frac{n}{2}, \frac{n}{4}\right)$.