Lecture 2

$X_1, X_2, \ldots$ independent, identically distributed r.v. (iid)

Q. How much does

$Y_n = \dfrac{X_1 + \cdots + X_n}{n}$ deviate from

$\text{Ave}(X_i)$ in terms of n?

Simplifying assumption: $\text{Ave}(X_i) = 0$

(can arrange this by shifting: e.g. random() $\longrightarrow$ random() $- 1/2$)

Consider Y_n as r.v.

What is $\text{Ave}(Y_n)$?

$$= \text{Ave}\left(\tfrac{1}{n}(X_1 + \cdots + X_n)\right) = \tfrac{1}{n}\text{Ave}(X_1 + \cdots + X_n)$$

$$= \tfrac{1}{n}\left(\text{Ave}(X_1) + \cdots + \text{Ave}(X_n)\right)$$

For any i, since X_i are identically distributed $\longrightarrow$

$$= \tfrac{1}{n}\left(\text{Ave}(X_i) + \cdots + \text{Ave}(X_i)\right)$$

$$= \text{Ave}(X_i)$$

So Y_n average as each X_i, but expect it should "vary less"

How to measure variability?

Quantity called variance

$$\text{Var}(Z) = \text{Ave}\left(\left(Z - \text{Ave}(Z)\right)^2\right)$$

small when Z doesn't vary much

$$\text{Var}(X_i) = \text{Ave}\left((X_i - \text{Ave}(X_i))^2\right) = \text{Ave}(X_i^2)$$

$\to \mathrm{Var}(X_i) = \mathrm{Ave}((X_i - \underbrace{\mathrm{Ave}(X_i)}_{=0})^2) \quad \ldots$

(Standard deviation is square root of variance).

Now, compute $\mathrm{Var}(Y_n)$ in terms of $\mathrm{Var}(X_i)$ and n.

use $\mathrm{Ave}(Y_n) = \mathrm{Ave}(X_i) = 0$

$$\mathrm{Var}(Y_n) = \mathrm{Var}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \mathrm{Ave}\left(\left(\frac{X_1 + \ldots + X_n}{n}\right)^2\right)$$

$$= \frac{1}{n^2} \mathrm{Ave}\left((X_1 + \ldots + X_n)^2\right)$$

$$= \frac{1}{n^2} \mathrm{Ave}\left(X_1^2 + \ldots + X_n^2 + X_1X_2 + X_1X_3 + \ldots\right)$$

$$= \frac{1}{n^2}\left(\mathrm{Ave}(X_1^2) + \ldots + \mathrm{Ave}(X_n^2) + \mathrm{Ave}(X_1X_2) + \mathrm{Ave}(X_1X_3) + \ldots\right)$$

By independence of X_i, X_j, $i \neq j$

$$= \frac{1}{n^2}\left(\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)\right) + \mathrm{Ave}(X_1)\mathrm{Ave}(X_2) + \mathrm{Ave}(X_1)\mathrm{Ave}(X_3)$$

$$= \frac{1}{n^2}\left(n\,\mathrm{Var}(X_i) + 0\right)$$

$$= \frac{1}{n}\mathrm{Var}(X_i)$$

Q: How to rescale sums so that variance doesn't depend on n?

$$\mathrm{Var}\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\right) = \mathrm{Var}(X_i)$$

with this scaling, histograms should converge to fixed shape.

Central Limit Thm

$X_1, X_2, \ldots$ independent, identically dist r.v.

Assume $\mathrm{Ave}(X_i) = 0$, $\mathrm{Var}(X_i) = 1$

Assume [illegible]

(can shift/rescale any r.v. to get this)

Then

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \to \text{Standard Normal distribution (Bell curve)}$$

as $n \to \infty$.

(convergence as distributions)