Sampling statistics $\rightarrow$ Today (Wed)

Central Limit Theorem $\rightarrow$ Friday

Prob. - solving after this

Dice: $S = \{1, 2, 3, 4, 5, 6\}$ (if dice comes up as 4, you get 4Rs. So on for every)

M: money you make in n tosses averaged by n

↓ Sample mean

$\bar{M}_p$: toss 1000 times → 3.3

$\bar{M}_r$: toss 1000 times → 3.6

3.3

3.2

4.2

2.7

⋮

Sample mean is itself a random variable

Dice: $S = \{1, 2, 3, 4, 5, 6\}$

"average"

Expectation (true mean) $E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5$

Sample mean

1) $$\begin{array}{ccccccccc} [5, & 1, & 6, & 4, & 5, & 3, & 2, & 4 & \cdots] \\ X_1 & X_2 & X_3 & X_4 & \cdots & & & & X_n \end{array}$$

$$\bar{X} = \frac{X_1 + X_2 + X_3 \cdots + X_n}{n} = 3.8$$

2) $[1, 2, 6, 4, 1, 3, 4, 5 \cdots] \longrightarrow 2.9$

Eg: You get $\underset{(x)}{3}$ Rs if dice falls $\underset{\{x\}}{\{3\}}$

$$\underbrace{[3, 5, 1, 6, 4, 3, 2 \ldots\ldots]}_{1000 \text{ times}}$$

How much money do we "expect to make"

Some properties of the sample mean:

$X_i \leftarrow \{1, 2, 3, 4, 5, 6\}$

$$\bar{X} = \frac{X_1 + X_2 + \cdots X_n}{n} \longrightarrow \text{is a random variable}$$

$$E[\bar{X}] = E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]$$

$$= \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n}$$

$$= \frac{n(3.5)}{n}$$

$$= 3.5 = E[X] \rightarrow \mu$$

$E[X_1] = 3.5$

$E[X_2] = 3.5$

(X : R.V on $\{1, 2, 3, 4, 5, 6\}$

Expectation of the sample mean is the true expectation,

We saw in code the histogram of the sample mean

Observation (from histogram)

As 'n' increases, the "spread" of the sample mean decreases

$Var(\bar{X})$ reduces as "n" increases

Chrome File Edit View History Bookmarks Profiles Tab Window Help 100% Thu 16 Jun 8:01 AM

Variance :

$$Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

* conditions apply (independent)

$$= \frac{1}{n^2}\left(Var X_1 + Var X_2 + \cdots + Var X_n\right)$$

$$= \frac{1}{n^2} \; n \cdot Var X$$

$$= \frac{Var X}{n}$$

(n : num_tosses in code)

So far: We saw sample mean $\bar{X}$ (X: distribution from which we sample)

$$E[\bar{X}] = E[X]$$

$$Var[\bar{X}] = \frac{1}{n} Var[X]$$

What about sample variance

Chrome File Edit View History Bookmarks Profiles Tab Window Help — 100% Thu 16 Jun 8:01 AM

Variance :

$$E\left(\underline{X - E[X]}\right)^2 \rightarrow Var(X) \rightarrow \sigma^2$$

$$= E[X^2] - \left(E[X]\right)^2$$

$$E[X] = \sum_k k\,P[X=k] \rightarrow \mu$$

sample mean

$$E[\bar{X}] = \mu \text{ (or } E[X])$$

$$\bar{X} = \frac{X_1 + X_2 + \dots X_n}{n}$$

Sample variance ?

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - X)^2 + \dots (X_n - \bar{X})^2}{? \rightarrow (n-1)}$$

We want $E[S^2] = \sigma^2$

We have $\bar{X}$ as proxy for μ

$$E[\bar{X}] = \mu$$

We have S^2 as proxy for σ^2

$$E[S^2] \overset{?}{=} \sigma^2$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 \overset{1}{=} \sum_{i=1}^{n} \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)$$

$$\overset{2}{=} \sum_{i=1}^{n} x_i^2 - \underbrace{2\bar{x}\sum_{i=1}^{n} x_i} + \underbrace{\sum_{i=1}^{n} \bar{x}^2}$$

$$\overset{3}{=} \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$\overset{4}{=} \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$2\bar{x}\left(\sum_{i=1}^{n} x_i\right)$$

$$2\bar{x}\left(n\bar{x}\right)$$

$$\sum_{i=1}^{n} \bar{x}^2 = \underbrace{\bar{x}^2 + \bar{x}^2 + \dots + \bar{x}^2}_{n \text{ times}}$$

$$= n\bar{x}^2$$

$$E\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right] \overset{①}{=} E\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

$$\overset{②}{=} \sum_{i=1}^{n} E[X^2] - n\,E[\bar{X}^2]$$

$$\overset{③}{=} n\,E[X^2] - n\,E[\bar{X}^2]$$

$$\overset{④}{=} n\operatorname{Var} X + n(E[X])^2 - n\operatorname{Var}(\bar{X}) - n(E[\bar{X}])^2$$

$$\overset{⑤}{=} n\sigma^2 + n\cancel{\mu^2} - n\frac{\sigma^2}{n} - n\cancel{\mu^2}$$

$$\overset{⑥}{=} (n-1)\sigma^2$$

$EX_i = EX$

$EX_i^2 = EX^2$

$\operatorname{Var} Y = EY^2 - (E[Y])^2$

$EY^2 = \operatorname{Var} Y + (E[Y])^2$

Chrome File Edit View History Bookmarks Profiles Tab Window Help

100% Thu 16 Jun 8:01 AM

$$E\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right] = \sigma^2$$

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

Summary:

Distribution

X: $\mu = E[X] = \sum_k k\, P[X=k]$

$$\sigma^2 = \mathrm{Var}(X) = E\left(X - E[X]\right)^2$$

Sampling

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$E[\bar{X}] = \mu \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$E[S^2] = \sigma^2$$

$\rightarrow$ when in "expectation" you match what you want, we call it "unbiased"

$$X \longrightarrow \operatorname{Var} X$$

$$Y = 3X + b$$

$$\operatorname{Var}(Y) = 9\operatorname{Var}(X)$$

$EX^2 - (EX)^2$

$X_1 \,\&\, X_2$ $\{1, 2, 3, 4, 5, 6\}$

$\operatorname{Var} X$

$$(\operatorname{Var} X_1 = \operatorname{Var} X_2)$$

$$\operatorname{Var}\left(\frac{1}{2} X_1 + \frac{1}{2} X_2\right) = \frac{1}{4} \operatorname{Var} X_1 + \frac{1}{4} \operatorname{Var} X_2$$

$$= \frac{1}{4}\, 2 \operatorname{Var} X = \frac{\operatorname{Var} X}{2}$$