e.g.: Roll fair dice 600 times independently, X be the # of sixes that appear.

(a): What type of r.v. is X? Dist of X?

support of X: $\{0, 1, 2, \cdots, 600\}$

$\Downarrow$

X is a discrete r.v.

Repeated indep experiment $\Rightarrow$ binomial dist

$$X \sim B(600, \tfrac{1}{6})$$

(b): Find $\mathbb{P}(X \leq 100)$

$$\mathbb{P}(X \leq 100) = \sum_{k=0}^{100} \underbrace{\mathbb{P}(X=k)}_{\text{pmf of binomial}} = \sum_{k=0}^{100} \binom{600}{k} \left(\tfrac{1}{6}\right)^k \left(\tfrac{5}{6}\right)^{600-k}$$

(c): Use a limit thm (LLN, CLT) to show why the prob in (b) can be approximated by value $\frac{1}{2}$.

We shall use CLT.

SLLN: If $X_1, \cdots, X_n, \cdots$ i.i.d. r.v., $\mathbb{E}|X_1| < \infty$,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} \mathbb{E}X_1 \quad (n \to \infty)$$

Since $X \sim B(600, \frac{1}{6})$, let X_i be an indicator of whether the i-th roll gives six.

$$X = X_1 + \cdots + X_{\boxed{600}}$$ (X is i.i.d. sum of large enough n=600 r.v.)

$$\begin{cases} \mathbb{E}X = 600 \times \frac{1}{6} = 100 \\ Var(X) = 600 \times \frac{1}{6} \times \frac{5}{6} = \frac{500}{6} \end{cases}$$

Apply CLT:

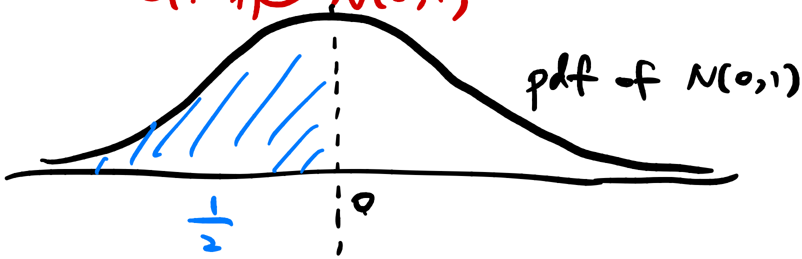
$$\frac{X - \mathbb{E}X}{\sqrt{Var(X)}} \overset{approx}{\sim} N(0,1)$$

$$\frac{X - 100}{\sqrt{\frac{500}{6}}} \overset{approx}{\sim} N(0,1)$$

$$\mathbb{P}(X \le 100) = \mathbb{P}\left(\frac{X-100}{\sqrt{\frac{500}{6}}} \le \underbrace{\frac{100-100}{\sqrt{\frac{500}{6}}}}_{=0}\right)$$

$$= \mathbb{P}\left(\boxed{\frac{X-100}{\sqrt{\frac{500}{6}}}} \le 0\right) \approx \frac{1}{2}$$

CLT: approx $N(0,1)$



We have a dist F, we have r.v. $X \sim F$,

want to estimate $\underset{\overset{\shortparallel}{f}}{\mathbb{E}\, h(X)}$ $\Rightarrow$ Monte Carlo

①: Sampling,

sample n i.i.d. random variable

$X_1, \cdots, X_n \sim F$.

②: Average

$$\hat{f} = \frac{h(X_1) + h(X_2) + \cdots + h(X_n)}{n}$$

(sample mean)

$\hat{f}$ is an estimator for f.

By SLLN, $\hat{f} \xrightarrow{a.s.} \mathbb{E} h(X) = f \quad (n \to \infty)$

so Monte Carlo converges to the true f.

Special Cases: when I want to estimate

$\mathbb{P}(A) = \mathbb{E}\, I_A$

Monte Carlo (100 simulations),

we are sampling from $B(600, \frac{1}{6})$ for 100 time, so we have ① Sampling

$$Y_1, Y_2, ---, Y_{100} \text{ i.i.d. } \sim B(600, \tfrac{1}{6})$$

but now want to estimate $\underbrace{\mathbb{P}(X<90)}_{\mathbb{E} I_{\{X<90\}}} = p$

so consider

$$\left\{\begin{array}{l} Z_1 = I_{\{Y_1<90\}} \\ Z_2 = I_{\{Y_2<90\}} \\ \vdots \\ Z_{100} = I_{\{Y_{100}<90\}} \end{array}\right\} \text{ i.i.d.}$$

Applying transformation

② : Average

estimator is $\underbrace{\dfrac{Z_1+Z_2+\cdots+Z_{100}}{100}}_{\text{i.i.d. sum, has CLT}} = \hat{p} \Rightarrow$ blue line in notebook

variance of $\hat{p}$:

$$\frac{\hat{p}-\mathbb{E}\hat{p}}{\sqrt{\operatorname{Var}\hat{p}}} \overset{\text{approx}}{\sim} N(0,1)$$

$$\mathbb{E}\hat{p} = \mathbb{E}Z_1 = \mathbb{E}I_{\{Y_1<90\}} = \mathbb{P}(Y_1<90) = p$$

$\hat{p}$ is unbiased

$$\mathrm{Var}\,\hat{p} = \frac{100 \cdot \mathrm{Var} Z_1}{100^2} = \frac{\mathrm{Var} Z_1}{100} = \boxed{\frac{p(1-p)}{100}}$$

appears in CI.

$$Z_1 = I_{\{Y_1 < 90\}} \sim B(1, p)$$

$$\underline{\mathrm{Var} Z_1 = p(1-p)}$$

(General Monte Carlo)

e.g.: We want to figure out expectation of $B(600, \frac{1}{6})$, called f.

sample 100 obs $Y_1, \ldots, Y_{100} \sim B(600, \frac{1}{6})$ i.i.d.

then estimator is $\dfrac{Y_1 + \cdots + Y_{100}}{100} = \hat{f}$.

Build 95% CI for p:

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{100}}} \overset{\text{approx}}{\sim} N(0,1) \qquad \text{(result of CLT)}$$

find a, b such that

$$\mathbb{P}\left(a \le \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{100}}} \le b \right) = 0.95$$

$$\Downarrow$$

$$\mathbb{P}\left(a\sqrt{\frac{p(1-p)}{100}} \le \hat{p}-p \le b\cdot\sqrt{\frac{p(1-p)}{100}} \right) = 0.95$$

$$\Downarrow$$

$$\mathbb{P}\left(\hat{p}-b\cdot\sqrt{\frac{p(1-p)}{100}} \le p \le \hat{p}-a\sqrt{\frac{p(1-p)}{100}} \right) = 0.95$$

So: CI is $\left[\hat{p}-b\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{100}},\ \hat{p}-a\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{100}}\right]$