Recap [Chapter 5]

- Key to simulation: $U \sim \text{Uniform}(0,1)$
- Simulating from a discrete distribution
- Simulating from a continuous distribution.
    - suppose X has cdf $F(x)$
    - $X = F^{-1}(U)$
- Simulate a large number (N) of independent draws to solve problems using Monte Carlo method.

Solving problems by Monte Carlo methods

Estimating $\mu = E(X)$ and $\sigma^2 = var(X) = E(X-\mu)^2$:

Simulate a large number (N) of independent draws from the distribution of X, say, $X_1, X_2, \ldots, X_N$

MC estimator of μ: $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ → long-run average

- $E[\bar{X}] = \mu$
- $var[\bar{X}] = \frac{\sigma^2}{N}$
- If N is large, $\bar{X} \approx N[\mu, \frac{\sigma^2}{N}]$.

- LLN: $\bar{X} \approx \mu$ because N is large.
- often $N = 1000$ is good enough.

MC estimator of $E[g(X)]$ *where* g *is a given function:*

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i) \approx E[g(X)]$$

→ long-run average.

MC estimator of σ^2: $\sigma^2 = E[(X-\mu)^2]$

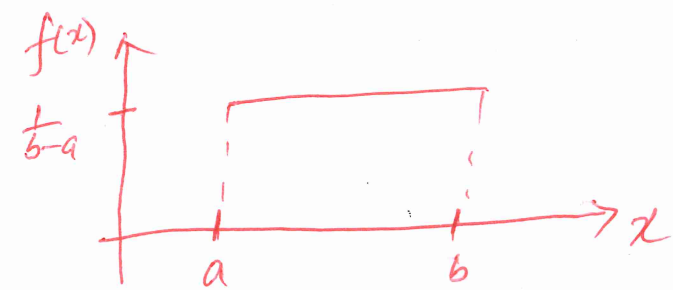
is $\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$ or $\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2$ → sample variance

Note: Divisor N or $N-1$ doesn't matter here since N is large.

Estimating an integral $I = \int_a^b g(x)dx$:

Recall: $X \sim \text{Uniform}(a,b)$, then

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$



Note that:

$$I = \int_a^b g(x)\,dx = \frac{1}{b-a}\int_a^b (b-a)\,g(x)\,dx$$

$$= \int_a^b (b-a)\,g(x)\cdot f(x)\,dx$$

$$= E\left[(b-a)\,g(X)\right]$$

$$= (b-a)\,E\left[g(X)\right]$$

already know how to estimate this using draws from Uniform(a, b) dist.

MC Estimator of I: $\hat{I} = (b-a)\,\frac{1}{N}\sum_{i=1}^{N} g(X_i)$

To simulate $X \sim \text{Uniform}(a,b)$: $a + (b-a)\,U$ ← Verify this.

Solving problems by Monte Carlo methods (continued)

Estimating $p = P(X \in A)$ for a given region A:

Simulate a large number (N) of independent draws from the distribution of X, say, $X_1, X_2, \ldots, X_N$

<u>NOTE</u>: $p = \int_A f(x)\,dx$

Define $Y_1, \ldots, Y_N$ as:

$$Y_i = \begin{cases} 1, & X_i \in A \\ 0, & \text{otherwise} \end{cases}$$

$Y_i \sim \text{Bernoulli}(p)$

NOTE: $E[Y_i] = p$. $\text{var}[Y_i] = p(1-p)$.

MC estimator of p: $\hat{p} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ — proportion of draws that fall in A.

↑ average

Properties of $\hat{p}$:

- LLN: $\hat{p} \approx p$. — for large N
- $E[\hat{p}] = p$, $\text{var}[\hat{p}] = \frac{p(1-p)}{N}$.
- If N is large: $\hat{p} \overset{\cdot}{\sim} N\left[p, \frac{p(1-p)}{N}\right]$.

$$\Rightarrow \frac{\hat{p} - p}{\sqrt{p(1-p)/N}} \overset{\cdot}{\sim} N(0,1).$$

Accuracy of a Monte Carlo study:

Error in estimation: $|\hat{p}-p|$ — a random quantity.

Specify a small *margin of error* ϵ and a small probability α.
Want N such that

small.

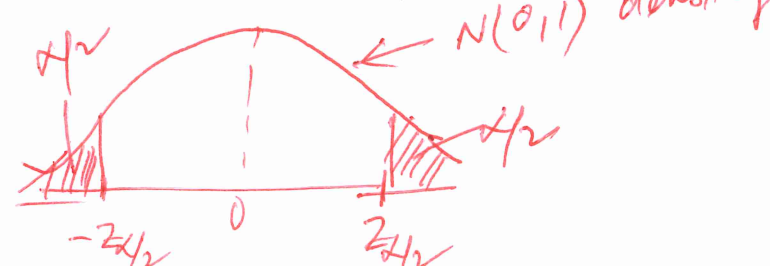
$$P(|\hat{p}-p| > \epsilon) \leq \alpha \qquad (1)$$

or equivalently $P[|\hat{p}-p| \leq \epsilon] \geq 1-\alpha$.

large.

- Error exceeds ϵ with probability α or less
- Error is ϵ or less with probability more than $1-\alpha$.

To derive a formula for N, suppose $Z \sim N(0,1)$ and $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$.



computing $z_{\alpha/2}$: use a normal table or R.

NOTE: $P[|Z| > z_{\alpha/2}] = \alpha$. — (2)

From the symmetry,

$$P(|Z| > z_{\alpha/2}) = \tag{2}$$

Now, let's derive an expression for $P(|\hat{p} - p| > \epsilon)$:

$$P[|\hat{p}-p| > \epsilon] = P\left[\frac{|\hat{p}-p|}{\sqrt{p(1-p)/N}} > \frac{\epsilon}{\sqrt{p(1-p)/N}}\right]$$

$$\approx P\left[|Z| > \frac{\epsilon}{\sqrt{p(1-p)/N}}\right]. \quad — (3).$$

Comparing (2) and (3), and noticing that $P(|Z| > x)$ is decreasing in x, we can conclude that (1) approximately holds if

Find N such that

$$P[|\hat{p}-p| > \epsilon] \le \alpha.$$

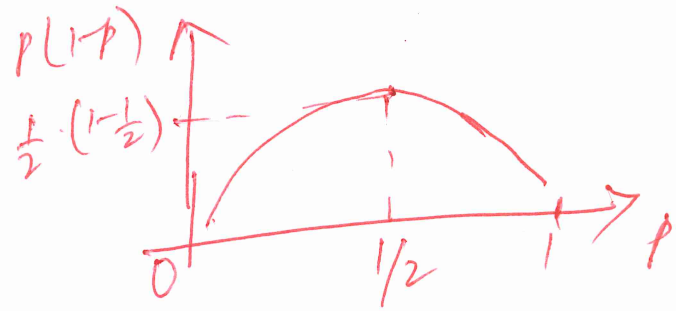
Need to have $\frac{\epsilon}{\sqrt{p(1-p)/N}} \geq z_{\alpha/2} \iff N \geq \left(\frac{z_{\alpha/2}}{\epsilon}\right)^2 p(1-p).$

A practical problem: p is unknown.

Alternative 1: Plug in an intelligent guess for p, say, p^*.

Then: $N \geq \left(\frac{z_{\alpha/2}}{\epsilon}\right)^2 p^*(1-p^*)$.

Alternative 2: Replace $p(1-p)$ by its maximum possible value



max. of $p(1-p) = \frac{1}{4}$.

$N \geq \frac{1}{4}\left(\frac{z_{\alpha/2}}{\epsilon}\right)^2$.

Note: This formula is valid only if N is large.

Ex: Suppose the desired accuracy is $(\epsilon, \alpha) = (0.03, 0.05)$. $N = ?$

$$P\left[\,|\hat{p} - p| \cancel{\geq} 0.03\,\right] \leq 0.05. \qquad z_{\alpha/2} = 1.96.$$

$$N \geq \frac{1}{4}\left[\frac{1.96}{0.03}\right]^2 = 1067.11$$

$$\Rightarrow \quad N \geq 1068.$$