TDAB01 Probability and Statistics

Maryna Prus
IDA, Linköping University

Lecture 5: Central Limit Theorem, Simulations, Monte Carlo Methods

Overview

- **Law of large numbers**
- **Central Limit Theorem**
- **Simulation of random variables**
- **Monte Carlo methods**

Law of large numbers

- Mean: $\bar{X}_n = \frac{X_1+X_2+\cdots+X_n}{n}$
- Mean of n independent random variables with the same expectation μ and the same variance $\sigma^2 < \infty$ is very close to μ for large n

Law of large numbers

$$\lim_{n\to\infty} \boldsymbol{P}\left(\left|\bar{X}_n - \mu\right| > \varepsilon\right) = 0$$

for all $\varepsilon > 0$

- **Proof** with Chebyshev's inequality: $\mathbb{E}(\bar{X}) = \mu$ and then

$$\boldsymbol{P}(\left|\bar{X}_n - \mu\right| > \varepsilon) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2}$$

but $Var(\bar{X}_n) = Var(X_i)/n \to 0$ for $n \to \infty$

Central Limit Theorem

- Distribution of $\bar{X}_n$ -?

Central Limit Theorem. Let $X_1, X_2, \dots, X_n$ be independent random variables with the same expectation μ and the same variance σ^2 (standard deviation σ), and let

$$S_n = X_1 + X_2 + \cdots + X_n.$$

As $n \to \infty$ the standardized sum

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\mathrm{Std}(S_n)}$$

converges in distribution to a standard normally distributed random variable, i. e.

$$F_{Z_n}(z) = \boldsymbol{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z \right) \to \Phi(z).$$

- S_n and $\bar{X}_n$ converge in distribution to $N(n\mu, n\sigma^2)$ and $N(\mu, \sigma^2/n)$
- For large n ($n > 30$) normal distribution can be used
- Example: See Example 4.13 in textbook

Central Limit Theorem

- $\mathit{Binomial}(n, p) \rightarrow N(np, np(1-p))$ for large n
  $\mathit{NegativBinomial}(k, p)$ and $\mathit{Gamma}(\alpha, \lambda)$ can also be approximated to normal distribution
- For normal distribution $\boldsymbol{P}(X = x) = 0$
  $\Rightarrow$ correction for approximation of **discrete distributions**:
  $\boldsymbol{P}(X = x) \rightarrow \boldsymbol{P}(x - 0.5 < X < x + 0.5)$
- For normal distribution $\boldsymbol{P}(X < x) = \boldsymbol{P}(X \leq x)$
  $\Rightarrow$ correction for approximation of **discrete distributions**:
  $\boldsymbol{P}(X < x) \rightarrow \boldsymbol{P}(X < x - 0.5)$

Simulation of random variables

- **Pseudo random number generator**: Computers can generate sequences of numbers that look like independent $U(0,1)$ random numbers
  $\rightarrow$ Good enough
- R: `runif(n)` simulates n random variables with $U(0,1)$ distribution
- Using $U \sim U(0,1)$ other distributions can be generated
- Example: **Bernoulli distribution** with success probability p:

$$X = \begin{cases} 1 & \text{if } U < p \\ 0 & \text{if } U \geq p \end{cases}$$

  - R code for Bernoulli distribution: `U=runif(1); X=1*(U<p)`
- Example: **Binomially distributed** random variables - Sum of Bernoulli distributed random variables
  - R code for Binomial(n,p): `U=runif(n); X=sum(U<p)`
- Example: **Geometric distribution** - number of trials for first success
  - R code for Geo(p): `X<-1; U=runif(1); while (U>p){X<-X+1;U=runif(1)}; X`

Simulation of discrete distributions

- General approach for simulation of discrete distributions, i. e.

$$p_i = \boldsymbol{P}(X = x_i) \text{ and } \sum_{\text{all } i} p_i = 1$$

  - Divide the interval $[0, 1]$ into sub-intervals:
    - $A_0 = [0, p_0)$
    - $A_1 = [p_0, p_0 + p_1)$
    - $A_2 = [p_0 + p_1, p_0 + p_1 + p_2)$
    - $\vdots$
  - $U \sim U(0, 1)$.
  - If $U \in A_i$ then $X = x_i$
- Example: Poisson distribution (see Example 5.9 in textbook)
  - $x_i = i$, $A_i = [F(i-1), F(i))$
  - R code for Po(λ): `λ<-5; U=runif(1); i<-0; F<-exp(-λ);`
    `while (U>=F){F<-F+exp(-λ)λ^(i+1)/factorial(i+1); i<-i+1}`
    `X<-i; X`

Inverse cdf method

Theorem. Let X be a continuous variable with cdf $F_X(x)$ and let $U = F_X(X)$ (random variable). Then $U \sim U(0,1)$.

Cdf of $U \sim U(0,1)$:

$$F_U(u) = P(U \leq u) = \begin{cases} 0, & u < 0 \\ u, & 0 \leq u \leq 1 \\ 1, & u > 1 \end{cases}$$

Then for $U = F_X(X)$ and $Y = F_X^{-1}(U)$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F_X^{-1}(U) \leq y) \\ &= P(F_X(F_X^{-1}(U)) \leq F_X(y)) \\ &= P(U \leq F_X(y)) \\ &= F_U(F_X(y)) = F_X(y) \end{aligned}$$

as $0 \leq F_X(y) \leq 1$ and $F_U(u) = u$ for $0 \leq u \leq 1$

Then Y has same probability distribution as X

Inverse cdf method

- **Inverse cdf method** (or inverse transform method):

  X with cdf $F(X)$ can be simulated using $U \sim U(0,1)$:
  - Generate values for $U \sim U(0,1)$
  - Compute values for X from $X = F^{-1}(U)$

- Example: $X \sim Exp(\lambda) \Rightarrow$ CDF of X:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

  To determine F_X^{-1} solve $y = 1 - e^{-\lambda x}$:

$$\Rightarrow \quad e^{-\lambda x} = 1 - y \quad \Rightarrow \quad x = -\frac{1}{\lambda}\ln(1-y) \quad \Rightarrow \quad F_X^{-1}(y) = -\frac{1}{\lambda}\ln(1-y)$$

  Then for $U \sim \mathrm{U}(0,1)$

$$X = -\frac{1}{\lambda}\ln(1-U) \sim \exp(\lambda)$$

Monte Carlo methods

- From Lecture 1:
  For large number of trials, relative frequency $\rightarrow$ probability,
  i. e. $\boldsymbol{P}(X = x) = 0.25$ means that $X = x$ occurs in 25 % of cases
- Simulation from distributions can be used to approximate probabilities
- Let $X_1, X_2, \ldots, X_N$ be generated values from distribution of X
  Then probability $p = \boldsymbol{P}(X < 0.5)$ can be approximated by

$$\hat{p} = \hat{\boldsymbol{P}}(X < 0.5) = \frac{\text{number of } X_1, X_2, \ldots, X_N \text{ which are less than } 0.5}{N}$$

- Notation: $\hat{p}$ - **estimator** (estimate) of probability p
- R code:

```
x = runif(10000, mean = 1, sd = 0.5)
pHat = sum(x<0.5)/10000
```

Monte Carlo methods

- But $\hat{p}$ is just **estimate** of p
  Can be different for different samples
- Estimate p several times, each time with new sample of size N
  Is average estimation value p?
  $\rightarrow$ Is $\mathbb{E}(\hat{p}) = p$? (Is $\hat{p}$ **unbiased**?)
  How much will $\hat{p}$ vary from sample to sample? $\rightarrow$ $Var(\hat{p})$-?
- Y = number of $X_1, \ldots, X_N$ which are less than 2
  $\Rightarrow$ $Y \sim Binomial(N, p)$ and then

$$\mathbb{E}(\hat{p}) = \mathbb{E}\left(\frac{Y}{N}\right) = \frac{1}{N} N \cdot p = p$$

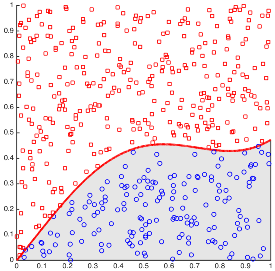
$\Rightarrow$ $\hat{p}$ is unbiased estimator of p, and

$$Var(\hat{p}) = Var\left(\frac{Y}{N}\right) = \frac{1}{N^2} Np(1-p) = \frac{p(1-p)}{N}$$

Monte Carlo integration

- To estimate:

$$\mathcal{I} = \int_0^1 g(x)dx,\ 0 \le x \le 1,\ 0 \le g(x) \le 1$$



- Simulate uniformly distributed

$$U_1, \ldots, U_N \quad \& \quad V_1, \ldots, V_N$$

- Consider pairs (U_i, V_i)

$$\hat{\mathcal{I}} = \frac{\text{Number of pairs for which } V_i < g(U_i)}{N}$$

- R code:
  - Define function g, for example `g=function(x)return(sin(x))`
    `u = runif(10000); v = runif(10000); IHat = mean(v < g(u))`

Simulation in R

- Generate *n* **values** from $N(\mu = 2, \sigma^2 = 3^2)$:

  `rnorm(n, mean = 2, sd = 3)`

- Generate *n* **values** from $Gamma(\alpha = 2, \lambda = 3)$: med

  `rgamma(n, shape = 2 , rate = 3)`

- Compute **pdf** at point $x = 1.5$ for $N(\mu = 2, \sigma^2 = 3^2)$

  `dnorm(x=1.5, mean = 2, sd = 3)`

- Compute **cdf** at point $x = 1.5$ for $N(\mu = 2, \sigma^2 = 3^2)$

  `pnorm(x=1.5, mean = 2, sd = 3)`

- For **other distributions** see Appendix in textbook

Thank you for your attention!