Lecture 2 - Review of Random Variables and Introduction to Computational Statistics

Computational Statistics and Applications

Vu Quoc Hoang (vqhoang@fit.hcmus.edu.vn)
Tran Thi Thao Nhi (thaonhitt2005@gmail.com)

FIT - HCMUS

Ngày 6 tháng 2 năm 2023

Agenda

1. The coupon collector's problem
2. Review of random variables
3. Probabilistic approximation by simulation
4. Zipf's law and Truyện Kiều - Nguyễn Du
5. Review of continuous random variable
6. Limit theorems
7. Probabilistic approximation by simulation

Agenda

1. The coupon collector's problem
2. Review of random variables
3. Probabilistic approximation by simulation
4. Zipf's law and Truyện Kiều - Nguyễn Du
5. Review of continuous random variable
6. Limit theorems
7. Probabilistic approximation by simulation

The coupon collector's problem

The coupon collector's problem A local retailer has a promotion in which they issue a set of n different coupons and place randomly one of the coupons in boxes of their product. To get a special gift from the retailer, the customer has to collect all n of coupons.

The question is: how many boxes need to be bought to collect all n coupons in order to receive the special gift?

Agenda

1. The coupon collector's problem
2. **Review of random variables**
3. Probabilistic approximation by simulation
4. Zipf's law and Truyện Kiều - Nguyễn Du
5. Review of continuous random variable
6. Limit theorems
7. Probabilistic approximation by simulation

Random variables

A numerical aspect X, whose value is determined by the outcome ω of an underlying random experiment T, is called the random variable (associated with T)

- We only know that X takes its value in set A before getting the final outcome,
- After getting ω, we determine the specific value of X, $x \in A$, which is denoted by $X(\omega) = x$.

Random variable is a function on the sample space Ω

- $X : \Omega \to A$, assigns to each possible outcome $\omega \in \Omega$ a numerical value $X(\omega) \in A$,
- A is called the **range** of X and is generally the subset of $\mathbb{R}$ (or $\mathbb{R}^d$).

Random variables are the main tools used for modeling the events. Consider a (numerical) random variable X associated with T on the sample space Ω. Let $C \subset \mathbb{R}$, the event "X takes its value in C" is defined as

$$(X \in C) = \{\omega \in \Omega : X(\omega) \in C\}.$$

The ditribution of a random variable

Consider a random variable X associated with the experiment T on the sample space Ω. The collection of all probabilities $\{P(X \in C) : C \subset \mathbb{R}\}$ specifies a probability measure on (the new sample space) $\mathbb{R}$ which is called **distribution** of X.

- The distribution of X shows the possibility that X might take on different values.
- Knowing the distribution of X makes it possible to analyze X without worrying about T or Ω.
- In general, set $\{P(X \in C) : C \subset \mathbb{R}\}$ in the definition above is "unpredictable". It will be useful to find alternative ways to specify the distribution of X, in order to make it "calculable".

Discrete random variable and its probability function

- We say that a random variable X has a discrete distribution or that X is a **discrete random variable** if its range is a **discrete** set (**finite** or **countably infinite**).
- If random variable X has a discrete distribution, the **probability function** (probability mass function - pmf) of X is defined as $f : \mathbb{R} \to \mathbb{R}$, and

$$f(x) = f_X(x) = P(X = x), x \in \mathbb{R}.$$

  - The probability function f is a probability measure that gives us probabilities of the possible values for the random variable X .
  - The closure of the real set $\{x \in \mathbb{R} : f(x) > 0\}$ is called the support of X, denoted by $\mathrm{Sup}(X)$.
  - The probability function satisfies these properties: $f(x) \geq 0, \forall x \in \mathbb{R}$ and $\sum_{x \in \mathrm{Sup}(X)} f(x) = 1$.
- Probability function that determines the distribution of a random variable

$$P(X \in C) = \sum_{x \in C} f(x), C \subset \mathbb{R}.$$

Independent random variables

Consider two discrete random variables X, Y. We say that X and Y are independent if for all sets $A, B \subset \mathbb{R}$,

$$P((X \in A) \cap (Y \in B)) = P(X \in A)P(Y \in B).$$

Intuitively, two random variables are independent if knowing the value of one of them does not change the probabilities for the other one.

Proposition. Two discrete random variables X, Y are independent only if

$$P((X = x) \cap (Y = y)) = P(X = x)P(Y = y)$$

for all $x, y \in \mathbb{R}$.

Mean of random variable

Let X be a discrete random variable with the probability function f, mean (or expectation) of X, denoted by $E(X)$, is defined as (if "calculable")

$$E(X) = \mu_X = \mu = \sum_x xP(X = x) = \sum_x xf(x).$$

The mean of X is the weighted average of the values with weights given by their respective probabilities.

Consider a random variable $X : \Omega \to \mathbb{R}$ and function $r : \mathbb{R} \to \mathbb{R}$, we say that $Y : \Omega \to \mathbb{R}$ is transformed from X by the function r. Y is called **transformation**, denoted by $Y = r(X)$, if Y is determined as

$$Y(\omega) = r(X(\omega)), \omega \in \Omega.$$

Then,

$$E(Y) = E(r(X)) = \sum_x r(x)f(x).$$

Variance and Standard deviation

Let X be a discrete random variable with the probability function f and mean $\mu = E(X)$, **variance** of X, denoted by $Var(X)$, is calculated as (if "calculable")

$$Var(X) = \sigma_X^2 = \sigma^2 = E\left((X-\mu)^2\right) = \sum_x (x-\mu)^2 P(X=x) = \sum_x (x-\mu)^2 f(x).$$

Then, $\sigma = \sigma_X = \sqrt{\sigma_X^2} = \sqrt{Var(X)}$ is called **standard deviation** of X. *Note:* standars deviation has the same unit as X but the variance has not.

Variance (and standard deviation) measures how **spread out** the distribution of a random variable is.

Proposition. Let X be a random variable (with variance), then

$$Var(X) = E(X^2) - (E(X))^2\,.$$

Essential properties of variance and stardard deviation

Let $X_1, X_2, ..., X_n$ be random variables (with variance), then

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) \quad \textbf{(linearity of expectation)}$$

Let X be a random variable and a, b are constant real number, then

1. $E(aX + b) = aE(X) + b$,
2. $Var(aX + b) = a^2 Var(X)$.

Let X, Y be independent random variables, then

1. $E(XY) = E(X)E(Y)$,
2. $Var(X + Y) = Var(X) + Var(Y)$.

Characteristic funtion of event

Given event A associated with an experiment T and the sample space Ω, we say that **characteristic function** (or indicator function) of A is $\mathbb{I}_A : \Omega \to \mathbb{R}$ and defined as

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

The characteristic function provides an alternative route to analyze an event as a random variable.

Proposition. For all events A,

$$E(\mathbb{I}_A) = P(A).$$

Bernoulli distribution

A discrete random variable X has the **Bernoulli distribution** with parameter p $(0 \leq p \leq 1)$, denoted by $X \sim \text{Bernoulli}(p)$, if its range only includes $\{0, 1\}$ and

$$f(x) = P(X = x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases}$$

Then, X has the variance $E(X) = p$ and standard deviation $Var(X) = p(1 - p)$.

Consider a tossing coin trial in which the probability of heads outcome is p, let X be "the number of times the coin lands on heads", then $X \sim \text{Bernoulli}(p)$. In case of fair coin, $X \sim \text{Bernoulli}(0.5)$.

Consider an trial T with event A has $P(A) = p$, then $\mathbb{I}_A \sim \text{Bernoulli}(p)$.

Binomial distribution

A discrete random variable X is said to be a **binomial distribution** with parameter n $(n \in \mathbb{N})$, p $(0 \le p \le 1)$, denoted by $X \sim \mathcal{B}(n, p)$, if its range includes $\{0, 1, ..., n\}$ and

$$f(x) = P(X = x) = C_n^x p^x (1 - p)^{n-x}, x \in \{0, 1, ..., n\}.$$

Then, X has the variance $E(X) = np$ and standard deviation $Var(X) = np(1 - p)$.

Given a trial T with event A and $P(A) = p$. Consider another trial R that "repeating T n times independently", let X be "the number of time event A occurs" then $X \sim \mathcal{B}(n, p)$.

Proposition. If $X_1, X_2, ..., X_n$ are **independent** random variable and **identically distributed** (iid) Bernoulli with parameter p, denoted by $X_1, X_2, ..., X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, and $X = \sum_{i=1}^{n} X_i$ then $X \sim \mathcal{B}(n, p)$.

Geometric distribution

A discrete random variable X is said to be a **geometric distribution** with parameter p $(0 < p \leq 1)$, denoted by $X \sim \text{Geometric}(p)$, if its range includes $\{1, 2, ...\}$ and

$$f(x) = P(X = x) = (1 - p)^{x-1}p, x \in \{1, 2, ...\}.$$

Then, X has the variance $E(X) = \frac{1}{p}$ and standard deviation $Var(X) = \frac{1-p}{p^2}$.

Given a trial T with event A and $P(A) = p$. Consider another trial R "repeating T independently until A occurs", let X be "the number of trials until observing A" then $X \sim \text{Geometric}(p)$.

Proposition (memoryless). Given $X \sim \text{Geometric}(p)$, for all $n = 1, 2, ...$ and $k = 0, 1, ...$,

$$P(X = k + n | X > k) = P(X = n).$$

Poisson distribution

A discrete random variable X is said to be a **Poisson distribution** with parameter λ ($\lambda > 0$), denoted by $X \sim \text{Poisson}(\lambda)$, if its range includes $\{0, 1, 2, ...\}$ and

$$f(x) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, x \in \{0, 1, 2, ...\}.$$

Then, X has variance $E(X) = \lambda$ and standard deviation $Var(X) = \lambda$.

Proposition. Given $X_1 \sim \text{Poisson}(\lambda_1), X_2 \sim \text{Poisson}(\lambda_2)$ and X_1, X_2 are independent, then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Proposition. Given $X \sim \mathcal{B}(n, p = \frac{\lambda}{n})$ with constant $\lambda > 0$, then

$$\lim_{n\to\infty} f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \text{ với mọi } x \in \{0, 1, 2, ...\}.$$

When n is very large and p is very small, distribution $\mathcal{B}(n, p)$ can be approximated by distribution $\text{Poisson}(\lambda)$.

Agenda

1. The coupon collector's problem
2. Review of random variables
3. **Probabilistic approximation by simulation**
4. Zipf's law and Truyện Kiều - Nguyễn Du
5. Review of continuous random variable
6. Limit theorems
7. Probabilistic approximation by simulation

Probabilistic approximation by simulation

To approximate the variance $E(X)$ of random variable X associated with an experiment T, we can execute an analytical calculation below

- Perform the experiment T N times repetitively (and independently), record all values X takes $x_1, x_2, \ldots, x_N$ (which is called **sample**), and calculate the **average**

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N}.$$

- When we execute this experiment a large number of times, $\bar{x} \approx E(X)$.
- Performing this experiment N times repetitively can be implemented by a computer simulation program.

The coupon collector's problem - Theoretically

Let X be the number of boxes that need to be bought to receive the special gift, which means the number of boxes barely enough to collect n coupons.
Let X_i be the time to collect the first i^{th} coupon right after $i-1$ coupons have been collected $(i = 1, 2, ..., n)$.
Then, $X = \sum_{i=1}^{n} X_i$ and X_i has geometric distribution with parameter

$$p_i = \frac{n-(i-1)}{n} = \frac{n-i+1}{n}\ (i = 1, 2, ..., n).$$

And we have

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{n}{n-i+1} = n\sum_{i=1}^{n} \frac{1}{i} = nH_n,$$

where $H_n = \sum_{i=1}^{n} \frac{1}{i}$ is called as the n^{th} **harmonic number**.

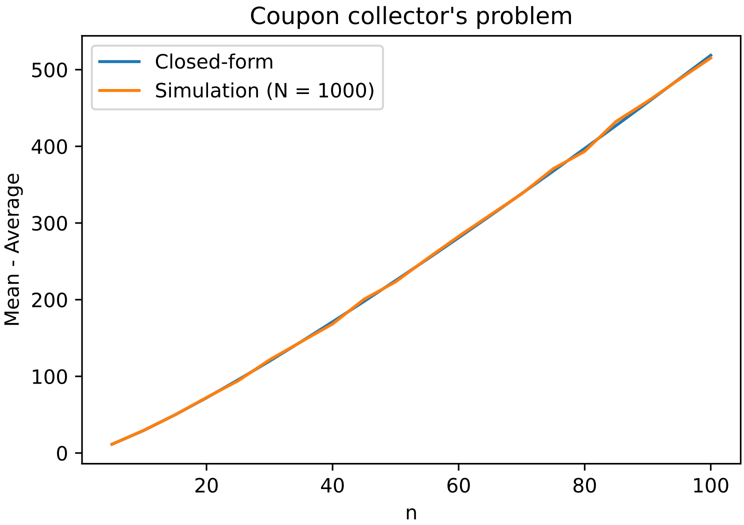
The coupon collector's problem - Simulation

```
def num_buy_to_win(n):
    coupons = []
    while len(set(coupons)) < n:
        coupons.append(random.randint(1, n))
    return len(coupons)

def average(n, N, X):
    m = sum(X(n) for _ in range (N))
    return m/N

average(10, 1000, num_buy_to_win)
#29.175
```

The coupon collector's problem - Result



Agenda

1. The coupon collector's problem
2. Review of random variables
3. Probabilistic approximation by simulation
4. **Zipf's law and Truyện Kiều - Nguyễn Du**
5. Review of continuous random variable
6. Limit theorems
7. Probabilistic approximation by simulation

Zipf's law

Zipf's law in term of quantitative linguistic: the **frequency** of words is inversely proportional to its **rank** (in many natural language corpus)

$$f(r) = c \times \frac{1}{r^s} \text{ hay } \log f(r) = \log c - s \log r$$

where constant c is the ratio factor, constant $s \approx 1$ is the value of exponent, $f(r)$ is the frequency of word of rank r ($r = 1, 2, ...$).

Zipf-Mandelbrot's law the generalization of Zipf's law

$$f(r) = c \times \frac{1}{(r+q)^s} \text{ hay } \log f(r) = \log c - s \log(r+q).$$

(https://en.wikipedia.org/wiki/Zipf%27s_law.)

Truyện Kiều - Nguyễn Du

“ ... Dưới cầu nước chảy trong veo
Bên cầu tơ liễu bóng chiều thướt tha ...”

Truyện Kiều is an epic poem written by Nguyễn Du, which is considered as the most famous poem in Vietnamese literature

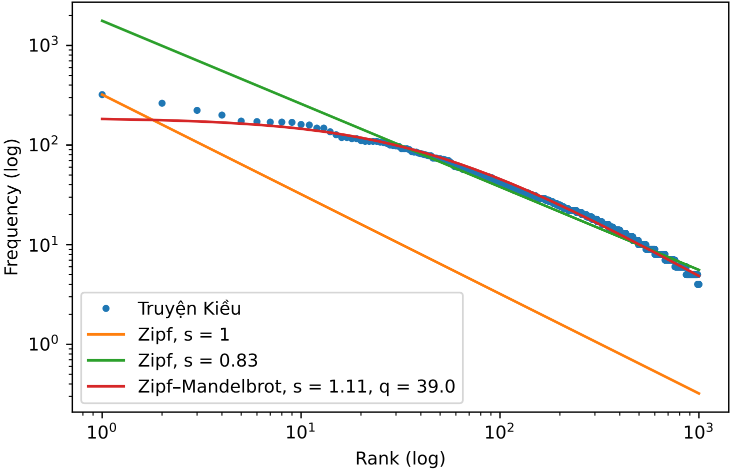
- lục bát (six-eight) meter,
- 3,254 verses,
- 22,778 words,
- 2,383 unique words.

(`https://vi.wikipedia.org/wiki/Truy%E1%BB%87n_Ki%E1%BB%81u`.)

Truyện Kiều - Nguyễn Du (cont.)

Rank	Word	Frequency	Rank	Word	Frequency
1	một	321	11	rằng	159
2	đã	263	12	lại	148
3	người	223	13	ra	148
4	nàng	200	14	hoa	136
5	lòng	174	15	tình	127
6	lời	172	16	còn	119
7	là	170	17	mới	119
8	cho	170	18	ai	116
9	cũng	169	19	đâu	116
10	có	161	20	chẳng	111

Truyện Kiều - Nguyễn Du (cont.)



Zipf's law

Question: why are **frequency** and **rank** of words in an inverse relation based on the **power law**?

Agenda

Continuous random variable and Probability denstity function

X is a **continuous random variable** if there exists a nonnegative function $f : \mathbb{R} \to \mathbb{R}$ such that for every interval of real numbers $[a, b]$ in $\mathbb{R}$, we have

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

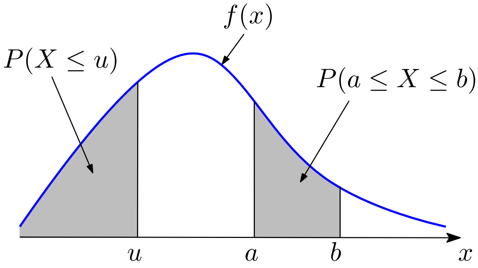
- f is called the **probability density function** of X which shows the probability that X can take value in the interval of $\mathbb{R}$
$$P(a \leq X \leq a+\epsilon) = \int_a^{a+\epsilon} f(x)dx \approx \epsilon f(a) \text{ when } \epsilon \text{is tiny.}$$
- The closure of the set $\{x \in \mathbb{R} : f(x) > 0\}$ is called the support of X, denoted by $\mathrm{Sup}(X)$.
- The probability density function satisfies two properties: $f(x) \geq 0, \forall x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f(x)dx = 1$.

Probability density function (cont.)

The probability density function determines the distribution of continuous random variable

$$P(X \in C) = \int_C f(x)dx, C \subset \mathbb{R}.$$

- $P(X = u) = \int_u^u f(x)dx = 0$,
- $P(X < u) = P(X \leq u) = \int_{-\infty}^u f(x)dx$,
- $P(X > u) = P(X \geq u) = \int_u^\infty f(x)dx$,
- $P(a \leq X \leq b) = \int_a^b f(x)dx$.



As can be seen from the note above $P(X = u) = 0$, it is possible that one event occurs though its probability is 0 (with E and $P(E) = 0$ but $E \neq \varnothing$).

Distribution function

(Cumulative) distribution function of a random variable X where $F : \mathbb{R} \to \mathbb{R}$ is defined by

$$F(x) = P(X \leq x) = \begin{cases} \sum_{t \leq x} f(t) & \text{if } X \text{ is discrete with probability function } f, \\ \int_{-\infty}^{x} f(t)dt & \text{if } X \text{ is continuous with probability density function } f. \end{cases}$$

F determines probability of X.

Distribution function F has the following properties:

1. Increasing: if $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$,
2. Standardizing: $lim_{x \to -\infty} F(x) = 0$ and $lim_{x \to \infty} F(x) = 1$,
3. Right-continuous: $F(x) = F(x^+) = lim_{t \to x, t > x} F(t)$.
4. If X is a continuous random variable, then F is continuous function and if F has continuous derivative at x then $F'(x) = f(x)$.

Joint distribution function

Joint distribution function of two random variables X, Y where $F_{XY} : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \ (x, y \in \mathbb{R}).$$

Proposition. Two random variables X, Y are independent if and only if $F_{XY}(x, y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$.

Two random variables X, Y are called **jointly continuous** if there exists a nonnegative function $f_{XY} : \mathbb{R}^2 \to \mathbb{R}$, such that, for any set $C \in \mathbb{R}^2$ we have

$$P((X, Y) \in C) = \iint_C f_{XY}(x, y) dxdy.$$

Proposition. Two joint continuous random variable X, Y are independent only if $f_{XY}(x, y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$.

Mean and variance

Given continuous random variable X with the probability density function f

- **Mean** of X is defined by

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

- **Variance** of X is defined by

$$\sigma^2 = Var(X) = E\left((X-\mu)^2\right) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx,$$

- Function $r : \mathbb{R} \to \mathbb{R}$ and $Y = r(X)$

$$E(Y) = E(r(X)) = \int_{-\infty}^{\infty} r(x) f(x) dx.$$

Uniform distribution

Continuous random variable X is said to be a **uniform distribution** over $[a, b]$ with $a < b$, denoted by $X \sim \mathcal{U}(a, b)$, if X ranges $[a, b]$ and

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{others.} \end{cases}$$

Then, X has mean $E(X) = \frac{a+b}{2}$ and variance $Var(X) = \frac{(b-a)^2}{12}$.

Let X be result of the trial "choosing randomly one point from $[a, b]$" then $X \sim \mathcal{U}(a, b)$.

Proposition. Given $X \sim \mathcal{U}(a, b)$ and $d \in (a, b)$, distribution of X knowing that $X \leq d$ is the uniform distribution over $[a, d]$, denoted by $(X|X \leq d) \sim \mathcal{U}(a, d)$.

Exponential distribution

Continuous random variable X is said to be a **exponential distribution** with parameter λ ($\lambda > 0$), denoted by $X \sim \text{Exp}(\lambda)$, if X ranges $[0, \infty)$ và

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{others.} \end{cases}$$

Then, X has mean $E(X) = \frac{1}{\lambda}$, variance $Var(X) = \frac{1}{\lambda^2}$ and distribution function

$$F(x) = 1 - e^{-\lambda x}, x \geq 0.$$

The exponential distribution may be viewed as a "continuous counterpart" of the geometric distribution.

Proposition. (Memoryless property). Given $X \sim \text{Exp}(\lambda)$, for all $t, s \geq 0$ then

$$P(X > t + s | X > s) = P(X > t).$$

Normal distribution

Continuous random variable X is said to be a **normal distribution** with mean μ and variance σ^2 ($\sigma > 0$), denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$, if X has probability density function

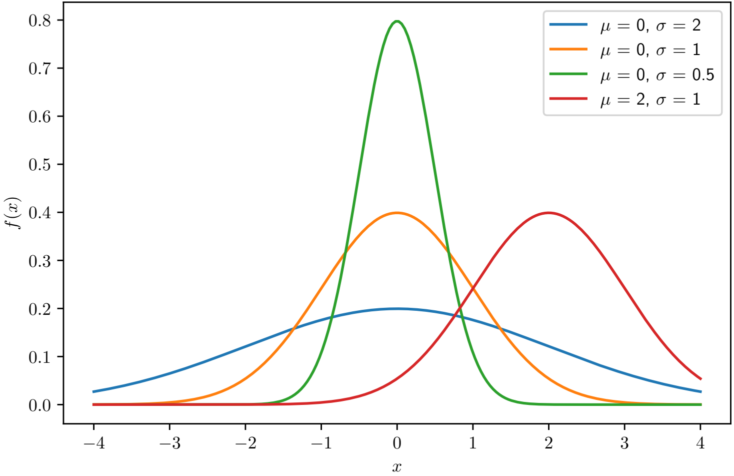
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

Then, X has mean $E(X) = \mu$ and variance $Var(X) = \sigma^2$.

In case $Z \sim \mathcal{N}(0, 1)$, then Z is said to be **standard normal distribution**. The probability density function and probability function of Z is usually denoted by ϕ, Φ, respectively, which means

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

Normal distribution (cont.)



Normal distribution (cont.)

Normal distribution has these essential properties

1. If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$ $(a \neq 0)$ then $X \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$,
2. If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and X_1, X_2 are independent then $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,
3. If $Z \sim \mathcal{N}(0, 1)$ và $X = \sigma Z + \mu$ then $X \sim \mathcal{N}(\mu, \sigma^2)$,
4. If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$ then $Z \sim \mathcal{N}(0, 1)$, and

$$F_X(x) = P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

Zipf's law and "monkey random texts"

Question: why are **frequency** and **rank** of words in an inverse relation based on the **power law**?

Answer: maybe just because of random!

- Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, 38(6), 1842–1845.
- Strategy: a monkey type some words by his special keyboard including M symbols and "blank space" button. Any "non-blank" symbol string between two blank spaces is called a "word" whereas a string of blank spaces is not. For example, string `a_mdf__pwell_` creates 3 words including `a`, `mdf`, `pwell`. Suppose that the monkey is uneducated (so he randomly types words) and has a lot of free time (so he types a very long document). Zipf's law also exists in the document created by that monkey!

Zipf's law and "monkey strategy" - Simulation

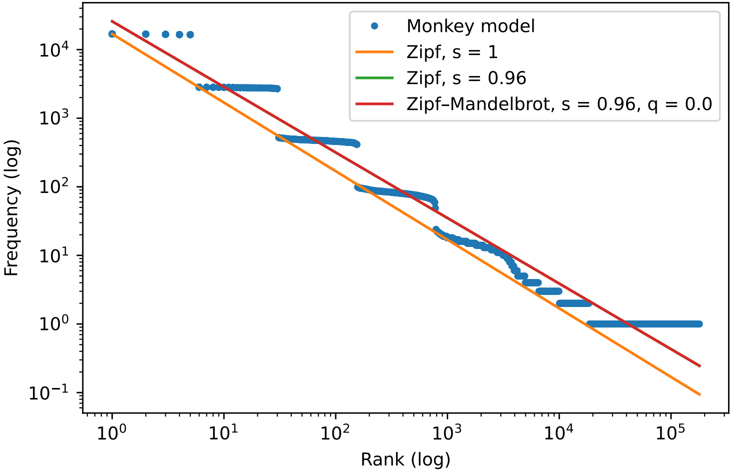
```
def monkey(N, k, alphabet, space=" "):
    alphabet += space; words = []; curWord = ""
    while len(words) < N:
        letter = random.choice(alphabet)
        if letter == space:
            if curWord == "":
                continue
            words.append(curWord)
            curWord = ""
        else:
            curWord += letter
    #...
```

Zipf's law and "monkey strategy" - Simulation (cont.)

```
def monkey(N, k, alphabet, space=" "):
    #...
    word_freqs = collections.Counter(words).most_common()
    freq = np.array([f for _, f in word_freqs])
    rank = np.array([int(r) for r in np.logspace(0,
                        np.log10(len(freq)), num=k)])
    return rank, freq[rank - 1]

M = 5 # alphabet size
N = 500_000 # number of word for simulation
rank, freq = monkey(N, 1000,
                       alphabet=string.ascii_lowercase[:M])
```

Zipf's law and "monkey strategy" - Result



Agenda

The law of large numbers (LLN)

Strong law of large numbers. Given iid random variables $X_1, X_2, \dots$ with expected value μ, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu \text{ (with probability1).}$$

Then, with N "large enough", we have $\mu \approx \frac{1}{N} \sum_{i=1}^{N} X_i$. Generally, let f be the real-valued function and iid random variables $X_1, X_2, \dots$ (same as X), then

$$E(f(X)) \approx \frac{1}{N} \sum_{i=1}^{N} f(X_i).$$

Especially, given event A, we have

$$P(A) = E(\mathbb{I}_A) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_A(X_i).$$

Central limit theorem

Central limit theorem. Given independent random variables $X_1, X_2, \ldots$ with finite expected value μ and variance $\sigma^2 > 0$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1),$$

where, $\xrightarrow{d}$ denotes **convergence in distribution**, which means

$$\lim_{n \to \infty} P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \leq x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, \forall x \in \mathbb{R}.$$

Then, with N "large enough", we have $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$ "approximate" standard normal distribution.

Agenda

Probabilistic approximation by simulation
Discrete random variable

To approximate the probability function f_X of discrete random variable X associated with an experiment T, we can execute an analytical calculation below

- Perform the experiment T N times repetitively (and independently) and calculate the frequency p_x of event "X takes x value".
- When we execute this experiment a large number of times, $p_x \approx P(X = x) = f_X(x)$.
- Performing this experiment N times repetitively can be implemented by a computer simulation program.

Probabilistic approximation by simulation
Continuous random variable

To approximate the probability density function f_X of continuous random variable X associated with an experiment T, we can execute an analytical calculation below

- Perform the experiment T N times repetitively (and independently), record all values X takes $x_1, x_2, ..., x_N$ (which is called **sample**).
- When we execute this experiment the large number of times, we can use **histogram** or **kernel density estimation (KDE)** over sample to approximate f_X.
- Performing this experiment N times repetitively can be implemented by a computer simulation program.

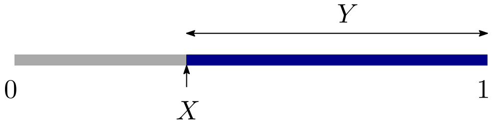
Probabilistic approximation by simulation
Example 1

Problem. Randomly pick a point on a segment of length 1. What is the expectation and distribution of the length of the longer part?

Solution. Let X be the point randomly picked on the segment, then $X \sim \mathcal{U}(0,1)$. Then X is the continuous random variable with probability density function

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{others.} \end{cases}$$

Let Y be the length of longer part, then $Y = \max\{X, 1 - X\}$.



Probabilistic approximation by simulation
Example 1 (cont.)

The expectation of the length of the longer part:

$$
\begin{aligned}
E(Y) &= E\left(\max\{X, 1-X\}\right) = \int_{-\infty}^{\infty} \max\{x, 1-x\} f_X(x)dx = \int_0^1 \max\{x, 1-x\}dx \\
&= \int_0^{1/2} \max\{x, 1-x\}dx + \int_{1/2}^1 \max\{x, 1-x\}dx = \int_0^{1/2} (1-x)dx + \int_{1/2}^1 xdx \\
&= \frac{3}{4}.
\end{aligned}
$$

Probabilistic approximation by simulation
Example 1 (cont.)

Let's find the probability density function of random variable $Y = \max\{X, 1-X\}$

$$F_Y(y) = P(Y \leq y) = P(\max\{X, 1-X\} \leq y), y \in \mathbb{R}.$$

Consider the following cases of y

1. $y < 1/2$: $(\max\{X, 1-X\} \leq y) = \varnothing$ since $0 \leq X \leq 1$ so $1/2 \leq \max\{X, 1-X\}$,
$$P\left(\max\{X, 1-X\} \leq y\right) = P(\varnothing) = 0.$$
2. $1/2 \leq y \leq 1$: $(\max\{X, 1-X\} \leq y) = (1-y \leq X \leq y)$,
$$P\left(\max\{X, 1-X\} \leq y\right) = P\left(1-y \leq X \leq y\right) = \int_{1-y}^{y} f_X(x)dx = \int_{1-y}^{y} dy = 2y.$$
3. $y > 1$: $(\max\{X, 1-X\} \leq y) = \Omega$ since $0 \leq X \leq 1$ so $\max\{X, 1-X\} \leq 1$,
$$P\left(\max\{X, 1-X\} \leq y\right) = P(\Omega) = 1.$$

Probabilistic approximation by simulation Example 1 (cont.)

Then,

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 1/2, \\ 2y & \text{if } 1/2 \leq y \leq 1, \\ 1 & \text{if } 1 < y. \end{cases}$$

Taking the derivative of distribution function, the probability density function of Y is defined by

$$f_Y(y) = F_Y'(y) = \begin{cases} 2 & \text{if } 1/2 \leq x \leq 1, \\ 0 & \text{others.} \end{cases}$$

In conclusion, Y have the uniform distribution on the interval $[1/2, 1]$, which means $Y \sim \mathcal{U}(1/2, 1)$.
Note that, from the distribution of Y, $Y \sim \mathcal{U}(1/2, 1)$, we also have $E(Y) = \frac{1/2+1}{2} = \frac{3}{4}$.

Probabilistic approximation by simulation
Example 1 (cont.)

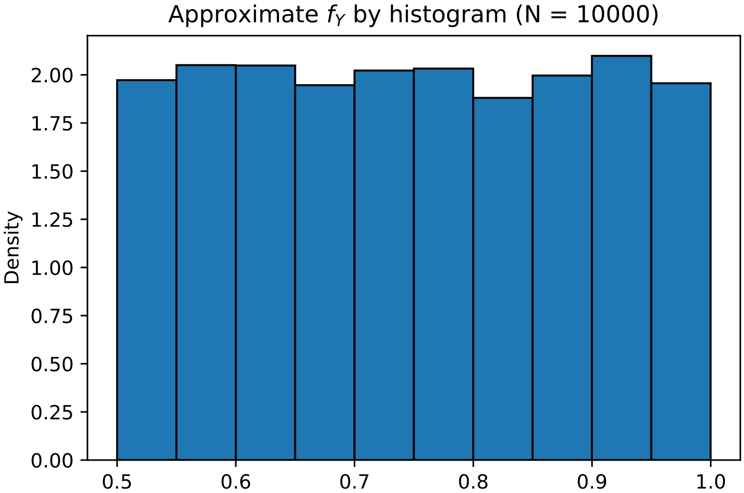
```
def greater_len(N):
    X = np.random.uniform(size=N)
    Y = np.maximum(X, 1 - X)
    return Y

N = 10000

np.mean(greater_len(N))
#0.7499721269808018

plt.hist(greater_len(N), density=True, edgecolor="black")
```

Probabilistic approximation by simulation
Example 1 (cont.)



Probabilistic approximation by simulation
Example 2

Problem. Let $X_1, X_2, ..., X_n$ be n random variables drawn from the uniform distribution $\mathcal{N}(\mu, \sigma^2)$. Suppose that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

($X_1, ..., X_n$ is usually seen as a sample of size n, with expected mean value $\bar{X}$ and variance S^2.)
Find the distribution of random variables $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\bar{X}-\mu}{S/\sqrt{n}}$.

Solution. $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ has standard normal distribution $\mathcal{N}(0,1)$ and $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ has **Student's t-distribution** with $n-1$ degrees of freedom.
(`https://en.wikipedia.org/wiki/Student%27s_t-distribution`.)

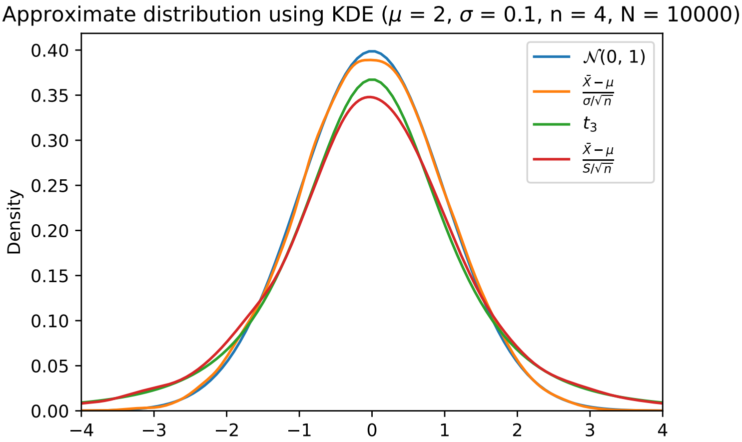
Probabilistic approximation by simulation
Example 2 (cont.)

```python
def sample(mu, sigma, n, N):
    X = np.random.normal(mu, sigma, size=(N, n))
    X_bar = np.mean(X, axis=1)
    S2 = np.var(X, axis=1, ddof=1)
    return X_bar, S2

X_bar, S2 = sample(mu, sigma, n, N)

plt.plot(x, scipy.stats.norm.pdf(x))
sns.kdeplot((X_bar - mu)/(sigma/np.sqrt(n)))
plt.plot(x, scipy.stats.t.pdf(x, n - 1))
sns.kdeplot((X_bar - mu)/(np.sqrt(S2)/np.sqrt(n)))
```

Probabilistic approximation by simulation
Example 2 (cont.)



References

Chapter 3-5. Morris H. DeGroot, Mark J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012.

Chapter 3-5. H. Pishro-Nik. *"Introduction to probability, statistics, and random processes"*, available at `https://www.probabilitycourse.com`. Kappa Research LLC, 2014.