Random numbers and histograms

モデル化とシミュレーション特論
2023 年度前期
佐賀大学理工学研究科　只木進一

Random numbers (乱数)

- Sequences of randomly generated numbers
- Examples: Sequences of numbers generated by a dice
  - Random sequences of numbers: The next number is not predictable.
  - Frequencies of appearance of a number are uniform.

Stochastic processes (確率過程)

- An event E will happen with probability p
  - Events does not happen deterministically.
- Reasons for unpredictable events
  - Unpredictable external effects
  - Unpredictable internal factors
  - Essential stochastic factors
    - thermal effects
    - quantum effects

What probatility p means

- Run sufficiently large number N of trials
- The number N_E of an event E

$$\frac{N_E}{N} \sim p \tag{1.1}$$

The law of large number.

- Example: the probability $p = 1/6$ for each roll of a dice.
    - Throwing a dice sufficiently large number N times.
    - The number of roll 1 happening tends to $N/6$
    - What "tend to" means?

Histograms for simple cases

- Histograms: Frequencies or relative frequencies of an event
- Example: counting the appearances of each roll of a dice.

$$N_i \ (i = \{1, 2, 3, 4, 5, 6\})$$
$$N = \sum_{i=1}^{6} N_i$$

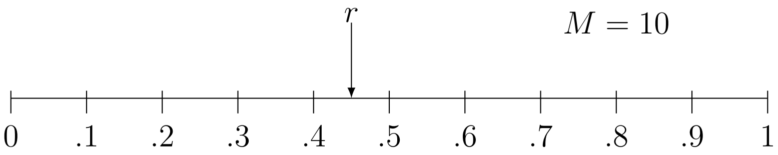
Histograms for cases of huge variety of events

- Example: integer random numbers in $[0, N-1]$
  - If $N = 2^{31} \simeq 2 \times 10^9$
  - How many times each integer happens?
  - What we can know from frequencies?
- Frequencies themselves are meaningless for continuous random variables.

Using bins: Example for $r \in [0, 1)$

- Divide the section into M equal-length subsections, which are called bins.
- Count the number of occurrences in each bin
- The number of occurrences in each bin must be large.

Deciding the subsection of a occurrence



- How to know the bin containing r.
- w: the length of a bin
- k-th bin contains r!

$$k = \left\lfloor \frac{r}{w} \right\rfloor$$

$$w = \frac{1}{M}$$

Example 2.1: Generalization

- Dividing a section $[a, b)$ into M equal-length subsections.
- How to know the bin containing r
- Example
    - Divide a section $[-0.1, .9)$ into 5 subsections.
    - Which subsection contains a element of $\{-0.02, 0.04, 0.22, 0.57\}$?

Ideal uniform numers in $[0, 1)$

- N: the number of random numbers
- M: the number of subsections (bins)
- The probability $p = 1/M$ which a bin contains a generated random number.
- A bin contains k occurrences.

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \tag{3.1}$$

Binomial coefficient

- Binomial expansion

$$(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k} \tag{3.2}$$

- Binomial coefficient

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \tag{3.3}$$

The mean and standard deviation

Deriving the mean using probability.

$$
\begin{aligned}
\langle k \rangle &= \sum_{k=0}^{N} k P(k) = \sum_{k=0}^{N} k \frac{N!}{k!(N-k)!} p^k \left(1-p\right)^{N-k} \\
&= \sum_{k=1}^{N} \frac{N(N-1)!}{(k-1)!(N-k)!} p p^{k-1} \left(1-p\right)^{N-k} \\
&= Np \sum_{\ell=0}^{N-1} \frac{(N-1)!}{\ell!(N-1-\ell)!} p^{\ell} \left(1-p\right)^{N-1-\ell} \\
&= Np \left[p + (1-p)\right]^{N-1} = Np
\end{aligned}
$$

The mean and standard deviation

$$\left\langle k^2 \right\rangle = \langle k(k-1) \rangle = N(N-1)p^2 + Np = Np\left[(N-1)p + 1\right]$$

$$\langle k(k-1) \rangle = \sum_{k=0}^{N} k(k-1) \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}$$

$$= \sum_{k=2}^{N} \frac{N(N-1)(N-2)!}{(k-2)!(N-k)} p^2 p^{N-2} (1-p)^{N-k}$$

$$= N(N-1)p^2$$

$$\sigma^2 = \left\langle k^2 \right\rangle - \langle k \rangle^2 = Np(1-p)$$

Tedious! Need more efficient methods!!

Probability generating function (確率母関数)

$$G(z) = \sum_{k=0}^{N} P(k) z^k$$

$$G(1) = \sum_{k=0}^{N} P(k) = 1$$

$$G'(z) = \sum_{k=1}^{N} kP(k) z^{k-1}$$

$$G'(1) = \sum_{k=1}^{N} kP(k) = \sum_{k=0}^{N} kP(k) = \langle k \rangle$$

$$G''(z) = \sum_{k=2}^{N} k(k-1)P(k)z^{k-2}$$

$$G''(1) = \sum_{k=2}^{N} k(k-1)P(k) = \sum_{k=0}^{N} k(k-1)P(k) = \langle k(k-1) \rangle$$

- We can obtain an average and a squared average, if we know the concrete form of $G(z)$.

$G(z)$ for binomial distribution

$$G(z) = \sum_{k=0}^{N} P(k) z^k = \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} z^k$$
$$= \sum_{k=0}^{N} \binom{N}{k} (zp)^k (1-p)^{N-k} = (zp + 1 - p)^N$$

$$G'(z) = Np\,(zp + 1 - p)^{N-1}$$
$$G''(z) = N(N-1)p^2\,(zp + 1 - p)^{N-2}$$

$$
\begin{aligned}
G'(1) &= Np = \langle k \rangle \\
G''(1) &= N(N-1)p^2 = \left\langle k^2 \right\rangle - \langle k \rangle \\
\left\langle k^2 \right\rangle &= N(N-1)p^2 + Np \\
\sigma^2 &= \left\langle k^2 \right\rangle - \langle k \rangle^2 = N(N-1)p^2 + Np - N^2p^2 \\
&= Np(1-p) \\
\frac{\sigma}{\langle k \rangle} &= \left(\frac{1-p}{p} \frac{1}{N} \right)^{1/2}
\end{aligned}
$$

- Important points
    - $\langle k \rangle \sim N$
    - $\sigma / \langle k \rangle \sim N^{-1/2}$

Continuous random variables

- If random variables are real numbers
    - We can not define the probability that a variable takes some specific real number value.
- Consider a real-valued random variable $X \in [a, b)$
- We can define the probability for cases such as $a \leq X < x < b$

$$F(x) = P(a \leq X < x) \tag{4.1}$$

- $F(x)$ is called *probability distribution.*

Probability density

- Probability X in a short subsection: $X \in [x, x + \Delta x)$

$$f(x)\Delta x = \frac{\mathrm{d}F}{\mathrm{d}x}\Delta x \tag{4.2}$$

$$F(x) = \int_a^x f(y)\mathrm{d}y \tag{4.3}$$

- $f(x)$: *probability density*

Pseudo random number generators

- Pseudo random numbers (疑似乱数)
  - Sequences of numbers generated by algorithms
  - You can generate the same sequences any times.
- `AbstractRandom` class
  - `java.util.Random` class inside
  - Generate next random `getNext()`
- Methods for generating non-uniform random numbers
  - Transform method (変換法): `Transform` class
  - Rejection method (棄却法): `Rejection` class

Transform Method (変換法)

- Probability density $f(x)$ ($x \in [a, b)$) and probability distribution F
$$F(x) = \int_a^x f(z)\mathrm{d}z \tag{5.1}$$
- Transform method is available if the inverse of $F(x)$ is obtained.
- Process
  - Generate a random number $r \in [0, 1)$.
  - $x = F^{-1}(r)$
  - $\{x\}$ distribute with $f(x)$

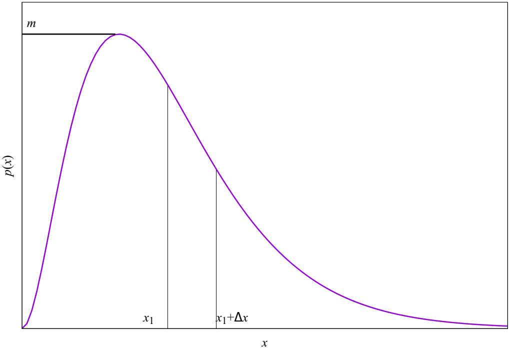
Example: Exponential distribution

$$f(x) = Ae^{-x} \tag{5.2}$$
$$0 \le x < 1$$
$$A = \frac{e}{e-1} \tag{5.3}$$

$$F(x) = \int_0^x f(z)\mathrm{d}z = A\left(1 - e^{-x}\right) \tag{5.4}$$
$$F^{-1}(r) = -\ln\left(1 - \frac{r}{A}\right) \tag{5.5}$$

Rejection Method (棄却法)

- Probability density $f(x)$ defined in $[a, b)$.
- Generate a random number pair $(x, y) \in [a, b) \times [0, m)$.
    - $m \geq f(x), \forall x \in [a, b)$
- Probability entering $[x_i, x_i + \Delta x]$ is proportional to $f(x)\Delta x$.

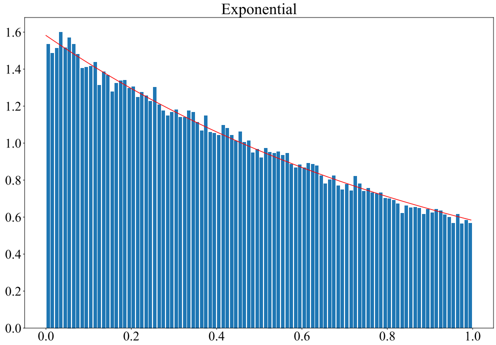


Classes for generating random numbers

- `randomNumbers` package
  - `AbstractRandom.java`
  - `Transform.java`
  - `Rejection.java`
- Using `java.util.Function.DoubleFunction`
  - Define the inverse of $F(x)$ for the Transform method.
  - Define $f(x)$ for the Rejection method.
  - Using lambda expressions

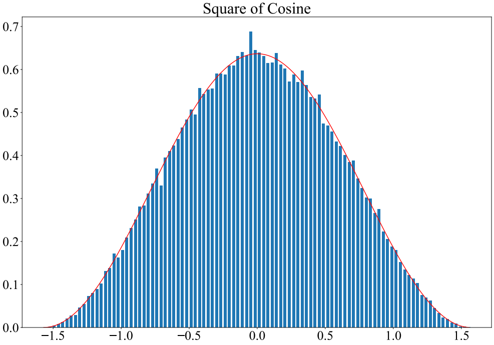
Example of Transform method: exponential distribution

```java
//指数分布に対応した分布関数の逆関数を定義
// A * exp (-x)
double A = Math.E / (Math.E - 1);
DoubleFunction<Double> invProDist = (x) -> {
    return -Math.log(1 - x / A);
};
//変換法による乱数生成のインスタンス
AbstractRandom aRandom = new Transform(invProDist);
```


Exponential
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Example of Rejection Method: Square of Cosine

$$f(x) = \frac{2}{\pi} \cos^2(x), \quad x \in [-\frac{\pi}{2}, \frac{\pi}{2}] \tag{6.1}$$



Law of Large Numbers (大数の法則)

- What does it mean that the probability of getting 1 on a dice is $1/6$?
- Consider the relative frequency of getting 1 on a dice.
  - It approaches $1/6$ with a large number of trials.
- Law of Large Numbers
- Let us take a closer look of this phenomenon.

Sample mean

- Consider a probabilistic variable X with the mean μ and deviation σ^2.
- Sample mean with size n.

$$\bar{X} = \frac{1}{n} \sum_{k=0}^{n-1} X_k \tag{7.1}$$

- Evaluate the population mean and deviation of $\bar{X}$.
    - Evaluate the mean and deviation of $\bar{X}$ with the probability of the population (母集団).
    - Equivalent to the mean of a large number of samples.

Population mean of sample means

The mean equals to the population mean μ.

$$\begin{aligned} E\left(\bar{X}\right) &= \frac{1}{n} E\left(\sum X_k\right) \\ &= \frac{1}{n} \sum E\left(X_k\right) = \frac{1}{n} n\mu \\ &= \mu \end{aligned} \tag{7.2}$$

Population deviation of sample means

The deviation reduces with n^{-1}.

$$\begin{aligned} V\left(\bar{X}\right) &= E\left(\left(\bar{X}-\mu\right)^2\right) = E\left(\frac{1}{n^2}\left(\sum_k \left(X_k-\mu\right)\right)^2\right) \\ &= \frac{1}{n^2}E\left(\sum_k \left(X_k-\mu\right)^2 + \sum_{i\neq j}\left(X_i-\mu\right)\left(X_j-\mu\right)\right) \\ &= \frac{1}{n^2}E\left(\sum_k \left(X_k-\mu\right)^2\right) + \frac{1}{n^2}E\left(\sum_{i\neq j}\left(X_i-\mu\right)\left(X_j-\mu\right)\right) \\ &= \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n} \end{aligned} \tag{7.3}$$

Confirm the law of large numbers by simulations

- Generate samples of size n.
- Instead evaluating the mean using the population distribution
  - Generate a large number m of samples with the same size.
  - Evaluate the mean and deviation for samples
- Changing n and observe n dependence.

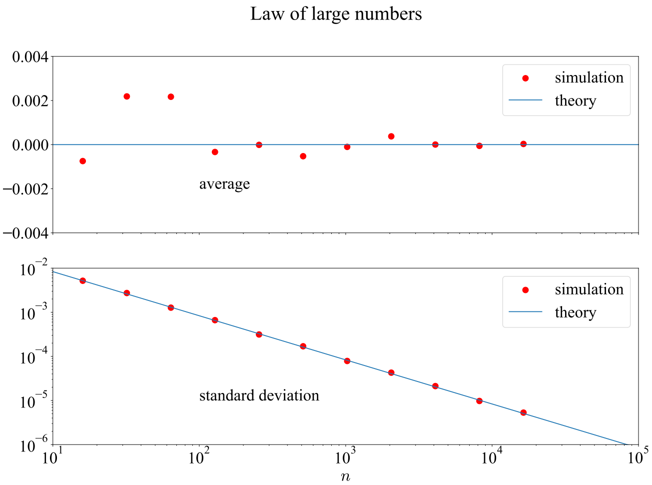
Example : uniform

$$f(x) = \begin{cases} 1 & -\frac{1}{2} \leq x < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{7.4}$$

$$\langle x \rangle = \int_{-1/2}^{1/2} x f(x) \mathrm{d}x = \int_{-1/2}^{1/2} x \mathrm{d}x = \left[\frac{1}{2} x^2 \right]_{-1/2}^{1/2} = 0 \tag{7.5}$$

$$\left\langle x^2 \right\rangle = \int_{-1/2}^{1/2} x^2 f(x) \mathrm{d}x = \int_{-1/2}^{1/2} x^2 \mathrm{d}x = \left[\frac{1}{3} x^3 \right]_{-1/2}^{1/2} = \frac{1}{12} \tag{7.6}$$

$$\sigma^2 = \left\langle x^2 \right\rangle - \langle x \rangle = \frac{1}{12} \tag{7.7}$$



$m = 1000$ for each sample size.