Статистический анализ данных. Спецкурс.

Лекция 2. Оценивание параметров.

Ботанический сад-институт ДВО РАН

Кислов Д.Е.

13 ноября 2016 г.

Положение №1

Набор однородных измерений (значений) x_1, x_2,. . . ,x_n некоторого параметра отождествляется с набором случайных величин, имеющих одинаковое распределение. Если $F_\xi(x)$ – функция этого распределения, то полагается, что $x_i \Subset F_\xi(x), \forall i$.

Положение №1

Набор однородных измерений (значений) x_1, x_2,...,x_n некоторого параметра отождествляется с набором случайных величин, имеющих одинаковое распределение. Если $F_\xi(x)$ – функция этого распределения, то полагается, что $x_i \Subset F_\xi(x)$, $\forall i$.

Следствие 1

Любая характеристика выборки, функция G от значений x_1, x_2,...,x_n, является случайной величиной, распределение которой определяется видом функций $F_\xi(x)$ и G.

Принципы статистического вывода

Положение №1

Набор однородных измерений (значений) x_1, x_2,. . . ,x_n некоторого параметра отождествляется с набором случайных величин, имеющих одинаковое распределение. Если $F_\xi(x)$ – функция этого распределения, то полагается, что $x_i \Subset F_\xi(x)$, $\forall i$.

Следствие 2

Если функция распределения для случайной величины $G = G(x_1, x_2, \ldots, x_n)$ будет найдена, то можно будет строить суждения относительно широты гипотетического рассеяния совокупности значений $\{G\}$, как если бы G вычислялось вновь и вновь по новым выборочным значениям x_1, x_2,. . . ,x_n.

Принципы статистического вывода

Следствие 2

Если функция распределения для случайной величины $G = G(x_1, x_2, \ldots, x_n)$ будет найдена, то можно будет строить суждения относительно широты гипотетического рассеяния совокупности значений $\{G\}$, как если бы G вычислялось вновь и вновь по новым выборочным значениям x_1, x_2,...,x_n.

???

В общем случае (при произвольных F и G) это крайне сложная задача...

Оценка параметров

Постановка задачи

Имеется выборка $(x_1, x_2, \ldots, x_n)$ из распределения $F_\xi(x, \theta)$. Необходимо построить оценку неизвестного параметра θ.

Оценка параметров

Постановка задачи

Имеется выборка $(x_1, x_2, \ldots, x_n)$ из распределения $F_\xi(x, \theta)$. Необходимо построить оценку неизвестного параметра θ.

Набросок решения

- построение функции $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$;

Оценка параметров

Постановка задачи

Имеется выборка $(x_1, x_2, \ldots, x_n)$ из распределения $F_\xi(x, \theta)$. Необходимо построить оценку неизвестного параметра θ.

Набросок решения

- построение функции $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$;
- $\hat{\theta}_n$ — случайная величина;

Оценка параметров

Постановка задачи

Имеется выборка $(x_1, x_2, \ldots, x_n)$ из распределения $F_\xi(x, \theta)$. Необходимо построить оценку неизвестного параметра θ.

Набросок решения

- построение функции $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$;
- $\hat{\theta}_n$ — случайная величина;
- из множества оценок $\hat{\theta}_n$ необходимо выбрать ту, которая обладает желаемыми статистическими свойствами;

Оценка параметров

Постановка задачи

Имеется выборка $(x_1, x_2, \ldots, x_n)$ из распределения $F_\xi(x, \theta)$. Необходимо построить оценку неизвестного параметра θ.

Набросок решения

- построение функции $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$;
- $\hat{\theta}_n$ — случайная величина;
- из множества оценок $\hat{\theta}_n$ необходимо выбрать ту, которая обладает желаемыми статистическими свойствами;

Желаемые статистические свойства

- $P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0,\ n \to \infty$ (состоятельность);

Оценка параметров

Постановка задачи

Имеется выборка $(x_1, x_2, \ldots, x_n)$ из распределения $F_\xi(x, \theta)$. Необходимо построить оценку неизвестного параметра θ.

Набросок решения

- построение функции $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$;
- $\hat{\theta}_n$ — случайная величина;
- из множества оценок $\hat{\theta}_n$ необходимо выбрать ту, которая обладает желаемыми статистическими свойствами;

Желаемые статистические свойства

- $P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0$, $n \to \infty$ (состоятельность);
- $\mathrm{E}(\hat{\theta}_n) = \theta$ (несмещенность);

Оценка параметров

Постановка задачи

Имеется выборка $(x_1, x_2, \ldots, x_n)$ из распределения $F_\xi(x, \theta)$. Необходимо построить оценку неизвестного параметра θ.

Набросок решения

- построение функции $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$;
- $\hat{\theta}_n$ — случайная величина;
- из множества оценок $\hat{\theta}_n$ необходимо выбрать ту, которая обладает желаемыми статистическими свойствами;

Желаемые статистические свойства

- $P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0$, $n \to \infty$ (состоятельность);
- $\mathrm{E}(\hat{\theta}_n) = \theta$ (несмещенность);
- $\mathrm{D}(\hat{\theta}_n)$ — минимальна (эффективность);

Оценка математического ожидания и дисперсии

Задача

По выборке $(x_1, x_2, \ldots, x_n)$ оценить математическое ожидание (a) и дисперсию (σ^2) распределения, из которого сформирована выборка.

Решение

1. $\hat{a}_1 = \dfrac{1}{n}\sum\limits_i x_i$, то $\hat{a}$ несмещенная и состоятельная;
2. $\hat{a}_2 = \sum\limits_i \alpha_i x_i$, где $\sum\limits_i \alpha_i = 1$, также является состоятельной и несмещенной оценкой a;
3. $\hat{\sigma}_1^2 = \dfrac{1}{n}\sum\limits_i (x_i - \hat{a}_1)^2$ — состоятельна, но смещена;
4. $\hat{\sigma}_2^2 = \dfrac{1}{n-1}\sum\limits_i (x_i - \hat{a}_1)^2$ — состоятельна и несмещена.

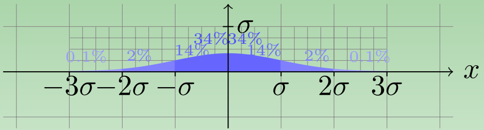
Задача

Дана выборка $x_1, x_2, \dots, x_n$ из нормального распределения. Необходимо построить оценки математического ожидания и дисперсии распределения.

Решение

- $\hat{a}_1 = \dfrac{1}{n} \sum_i x_i$, то $\hat{a}$ несмещенная и состоятельная;
- $\hat{\sigma}_2^2 = \dfrac{1}{n-1} \sum_i (x_i - \hat{a}_1)$ — состоятельна и несмещена.

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$



Формулы для доверительных интервалов

Пусть $x_1, x_2, \ldots, x_n$ выборка из нормального распределения. a, σ^2 – неизвестные математическое ожидание и дисперсия этого распределения.

Интервальная оценка среднего:

$$\overline{x} - t_\gamma \frac{\hat{\sigma}}{\sqrt{n}} \leq a \leq \overline{x} + t_\gamma \frac{\hat{\sigma}}{\sqrt{n}}$$

где

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\overline{x} - x_i)^2$$

t_γ – γ-квантиль распределения Стьюдента с $n-1$ степенями свободы. $\gamma = \dfrac{1+p}{2}$, $p = 1 - \alpha$, α – уровень значимости.

Формулы для доверительных интервалов (дисперсия)

Пусть $x_1, x_2, \dots, x_n$ выборка из нормального распределения. a, σ^2 – неизвестные математическое ожидание и дисперсия этого распределения.
Интервальная оценка дисперсии:

$$\frac{1}{\chi^2_{\gamma_1}} \sum_{i=1}^{n} (\overline{x} - x_i)^2 \le \sigma^2 \le \frac{1}{\chi^2_{\gamma_2}} \sum_{i=1}^{n} (\overline{x} - x_i)^2$$

где

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\overline{x} - x_i)^2$$

χ^2_γ – γ-квантиль распределения χ^2 с $n-1$ степенями свободы.
$\gamma_1 = \dfrac{1+p}{2}$, $\gamma_1 = \dfrac{1-p}{2}$, $p = 1 - \alpha$, α – уровень значимости.

Вычисление доверительных интервалов на R

Пример вычислений

Пусть x = (1.62081382, -1.52406491, -0.90221069, 0.34349028, 0.61838820, -0.80183081, -0.55995198, -0.71210125, -0.30742587, 0.23734660, 0.63370975, -1.04691828, 0.96165012, -0.35549459, -0.61876437, 1.36677559, 0.09070588, -0.94914323, 0.97321593, -1.73956125) выборка из нормального распределения.
Необходимо найти интервальные оценки дисперсии и математического ожидания.

код на R. Среднее

```
n<-length(x)
alpha<-0.05
error <- qt((1-alpha / 2.0), df=(n-1)) * sd(x) / sqrt(n)
l <- mean(x) - error
r <- mean(x) + error
```

Доверительный интервал для среднего

```
> l
[1] -0.5789119
> r
[1] 0.3117748
> mean(x) # point estimation of the m-expectation
[1] -0.1335686
```

Вычисление доверительных интервалов на R

Код на R. Дисперсия

```
df <- length(x)-1
alpha <- 0.05
l <-  df * var(x)/qchisq(1-alpha/2.0)
r <-  df * var(x)/qchisq(alpha/2.0)
```

```
> r
[1] 1.9316
> l
[1] 0.5236715
> var(x) # point estimation of the variance
[1] 0.9054645
```

Оценка всхожести семян

Формулировка задачи

Имеется образцы семян в количестве n-штук. Необходимо оценить вероятность всхожести семян.

Оценка всхожести семян

Формулировка задачи

Имеется образцы семян в количестве n-штук. Необходимо оценить вероятность всхожести семян.

Формулировка в терминах ТВ

Пусть p –вероятность того, что семя взойдет; $1 - p$ – вероятность того, что оно не взойдет. В результате n-испытаний (проращивания семян), имеется серия экспериментов по схеме Бернулли с неизвестной вероятностью «успеха» p. Эту вероятность и требуется найти.

Оценка всхожести семян

Решение

- Обозначим k – количество проросших семян;

Оценка всхожести семян

Решение

- Обозначим k – количество проросших семян;
- Можно получить оценку возможных значений k, решив: $\sum\limits_{i=k-s}^{i=k+s} C_n^i p^i (1-p)^{n-k} = 1-\alpha$, где α (уровень значимости) задается, а s – находится из уравнения, но

Оценка всхожести семян

Решение

- Обозначим k – количество проросших семян;
- Можно получить оценку возможных значений k, решив:
  $\sum_{i=k-s}^{i=k+s} C_n^i p^i (1-p)^{n-k} = 1-\alpha$, где α (уровень значимости) задается, а s – находится из уравнения, но
  - такое уравнение не всегда просто решить

Оценка всхожести семян

Решение

- Обозначим k – количество проросших семян;
- Можно получить оценку возможных значений k, решив:
$$\sum_{i=k-s}^{i=k+s} C_n^i p^i (1-p)^{n-k} = 1-\alpha,$$
где α (уровень значимости) задается, а s – находится из уравнения, но
    - такое уравнение не всегда просто решить
- Выход?

Оценка всхожести семян

Решение

- Обозначим k – количество проросших семян;
- Можно получить оценку возможных значений k, решив: $\sum_{i=k-s}^{i=k+s} C_n^i p^i (1-p)^{n-k} = 1-\alpha$, где α (уровень значимости) задается, а s – находится из уравнения, но
  - такое уравнение не всегда просто решить
- Выход?

Оценка всхожести семян

Решение

- Обозначим k – количество проросших семян;
- Можно получить оценку возможных значений k, решив:
  $\sum\limits_{i=k-s}^{i=k+s} C_n^i p^i (1-p)^{n-k} = 1-\alpha$, где α (уровень значимости) задается, а s – находится из уравнения, но
  - такое уравнение не всегда просто решить
- Выход?

- Использование предельных теорем

Оценка всхожести семян

Решение

- Обозначим k – количество проросших семян;
- Можно получить оценку возможных значений k, решив: $\sum\limits_{i=k-s}^{i=k+s} C_n^i p^i (1-p)^{n-k} = 1-\alpha$, где α (уровень значимости) задается, а s – находится из уравнения, но
  - такое уравнение не всегда просто решить
- Выход?

- Использование предельных теорем
- Теорема Муавра-Лапласа (интегральная) Если p вероятность успеха в серии n испытаний по схеме Бернулли (k – число успехов в n испытаниях), то

$$P\left(\frac{k-np}{\sqrt{np(1-p)}} < x\right) \xrightarrow[\text{при } n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2}dt.$$

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(-\varepsilon < \frac{k - np}{n} < \varepsilon\right) =$$

Оценка всхожести семян

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(-\varepsilon < \frac{k-np}{n} < \varepsilon\right) =$$

$$= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} < \frac{k-np}{\sqrt{npq}} < \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) =$$

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(-\varepsilon < \frac{k - np}{n} < \varepsilon\right) =$$

$$= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} < \frac{k - np}{\sqrt{npq}} < \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) =$$

$$\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - \Phi_0\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 2 \cdot \Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - 1$$

Оценка всхожести семян

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(-\varepsilon < \frac{k - np}{n} < \varepsilon\right) =$$

$$= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} < \frac{k - np}{\sqrt{npq}} < \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) =$$

$$\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - \Phi_0\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 2 \cdot \Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - 1$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 2\Phi_0(2\varepsilon\sqrt{n}) - 1$$

Оценка всхожести семян

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(-\varepsilon < \frac{k - np}{n} < \varepsilon\right) =$$

$$= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} < \frac{k - np}{\sqrt{npq}} < \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) =$$

$$\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - \Phi_0\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 2 \cdot \Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - 1$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 2\Phi_0(2\varepsilon\sqrt{n}) - 1$$

Что дальше?

Оценка всхожести семян

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(-\varepsilon < \frac{k-np}{n} < \varepsilon\right) =$$

$$= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} < \frac{k-np}{\sqrt{npq}} < \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) =$$

$$\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - \Phi_0\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 2 \cdot \Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - 1$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 2\Phi_0(2\varepsilon\sqrt{n}) - 1 = 1 - \alpha, (\alpha = 0.05)$$

Что дальше?

Оценка всхожести семян

Решение

$$\Phi_0 = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(-\varepsilon < \frac{k - np}{n} < \varepsilon\right) =$$

$$= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} < \frac{k - np}{\sqrt{npq}} < \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) =$$

$$\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - \Phi_0\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 2 \cdot \Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - 1$$

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 2\Phi_0(2\varepsilon\sqrt{n}) - 1 = 1 - \alpha, (\alpha = 0.05)$$

Что дальше? Значения для Φ_0^{-1} могут вычислять любые современные программные пакеты статистического анализа, либо его можно определить по таблицам.

Оценка всхожести семян

Вычисления на R

```
> n<-100
> k<-40
> alpha <- 0.05
> error <- 0.5*qnorm(1-alpha/2)/sqrt(n)
> left  <- k/n - error
> right <- k/n + error
> left
[1] 0.3020018
> right
[1] 0.4979982
```

Вычисления на Python

Python:

```
>>> from scipy.stats import norm
>>> import numpy as np
>>> n = 100.0
>>> k = 40.0
>>> alpha = 0.05
>>> error = 0.5*norm().ppf(1-alpha/2.0)/np.sqrt(n)
>>> left = k / n  - error
>>> right = k / n  + error
>>> left
0.30200180077299732
>>> right
0.49799819922700272
```

Оценка необходимого числа экспериментов/размера выборки

Задача

Сколько нужно взять семян, чтобы оценить их всхожесть с точностью $\varepsilon = 0.01$ и доверительной вероятностью $p_v = 0.99$.

Оценка необходимого числа экспериментов/размера выборки

Задача

Сколько нужно взять семян, чтобы оценить их всхожесть с точностью $\varepsilon = 0.01$ и доверительной вероятностью $p_v = 0.99$.

Решение

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 2\Phi_0(2\varepsilon\sqrt{n}) - 1$$

Оценка необходимого числа экспериментов/размера выборки

Задача

Сколько нужно взять семян, чтобы оценить их всхожесть с точностью $\varepsilon = 0.01$ и доверительной вероятностью $p_v = 0.99$.

Решение

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 2\Phi_0(2\varepsilon\sqrt{n}) - 1$$
$$2\Phi_0(2 \cdot \varepsilon\sqrt{n}) - 1 = p_v$$

Оценка необходимого числа экспериментов/размера выборки

Задача

Сколько нужно взять семян, чтобы оценить их всхожесть с точностью $\varepsilon = 0.01$ и доверительной вероятностью $p_v = 0.99$.

Решение

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 2\Phi_0(2\varepsilon\sqrt{n}) - 1$$

$$2\Phi_0(2 \cdot \varepsilon\sqrt{n}) - 1 = p_v$$

$$n = \frac{1}{4\varepsilon^2}\left(\Phi_0^{-1}\left(\frac{1+p_v}{2}\right)\right)^2$$

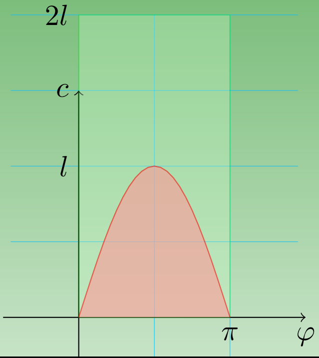
Оценка необходимого числа экспериментов/размера выборки

$$n = \frac{1}{4\varepsilon^2}\left(\Phi_0^{-1}\left(\frac{1+p_v}{2}\right)\right)^2$$

Реализация на R:

```
epsilon <- 0.01
pv <- 0.99
n = 1/4/epsilon^2*qnorm((1+pv)/2)^2
n
[1] 16587.24
```

На плоскости расчерчивают параллельные линии. Показать, что если расстояние между линиями вдвое больше иглы, подбрасываемой над плоскостью, то отношение общего числа бросаний иглы к числу пресечений ею одной из линий стремится при увеличении количества экспериментов к $\pi = 3.14159\ldots$.

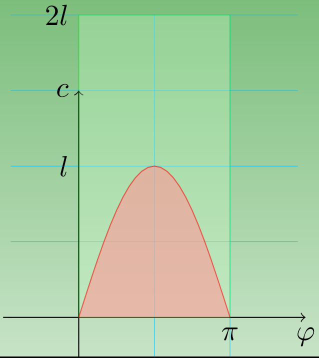


Решение:
$\{c, \varphi\} = [0, 2l] \times [0, \pi]$, откуда

$$P(\text{«пер-ние»}) = \frac{S_{\sin}(\bullet)}{S_{2l\times\pi}};$$

$$P(\text{«пер-ние»}) = \frac{1}{\pi}.$$

На плоскости расчерчивают параллельные линии. Показать, что если расстояние между линиями вдвое больше иглы, подбрасываемой над плоскостью, то отношение общего числа бросаний иглы к числу пресечений ею одной из линий стремится при увеличении количества экспериментов к $\pi = 3.14159\ldots$.



Решение:
$\{c, \varphi\} = [0, 2l] \times [0, \pi]$, откуда

$$P(\text{«пер-ние»}) = \frac{S_{\sin}(\bullet)}{S_{2l\times\pi}};$$

$$P(\text{«пер-ние»}) = \frac{1}{\pi}.$$

Вопрос

Сколько раз необходимо осуществлять эксперимент, чтобы достичь заданной точности?

Сколько необходимо испытаний? (грубая оценка по неравенству Чебышева)

$$pq = \frac{\pi - 1}{\pi^2}$$

$$P\left(\left|\frac{\nu_n}{n} - p\right| < \alpha\right) > 1 - \varepsilon,$$
$$\alpha = 0.001, \qquad \varepsilon = 0.001,$$
$$\Downarrow$$
$$\mathbf{n > 232013}$$

$$pq \leq 1/4$$

$$P\left(\left|\frac{\nu_n}{n} - p\right| < \alpha\right) > 1 - \varepsilon,$$
$$\alpha = 0.001, \qquad \varepsilon = 0.001,$$
$$\Downarrow$$
$$\mathbf{n > 267311}$$

Вывод

Априорное знание оцениваемых вероятностей позволяет сэкономить число необходимых экспериментов для достижения установленной точности

Продолжение решения задачи Бюффона

Сколько необходимо испытаний? (грубая оценка по неравенству Чебышева)

$pq = \dfrac{\pi - 1}{\pi^2}$

$$P\left(\left|\frac{\nu_n}{n} - p\right| < \alpha\right) > 1 - \varepsilon,$$
$$\alpha = 0.001, \qquad \varepsilon = 0.001,$$
$$\Downarrow$$
$$\mathbf{n > 232013}$$

$pq \leq 1/4$

$$P\left(\left|\frac{\nu_n}{n} - p\right| < \alpha\right) > 1 - \varepsilon,$$
$$\alpha = 0.001, \qquad \varepsilon = 0.001,$$
$$\Downarrow$$
$$\mathbf{n > 267311}$$

Упражнение

Получить оценку необходимого числа экспериментов используя предельные теоремы