

TITMCS_HANDBOOK

(compiled by Mike ">P&A<" Gafurov)

1. Простір елементарних подій. Випадкові події, операції над ними.

Стохастичний експеримент – експеримент, який можна повторювати будь-яку кількість разів і наслідки якого не можна передбачити наперед. (наприклад підкидання монети, грального кубика)

Ω (омега велике) – усі можливі наслідки стохастичного експерименту – простір елементарних подій.

$$\Omega = \{\omega_1, \omega_2 \dots \omega_n\}$$

ω_i – елементарний наслідок або елементарна подія

Операції над подіями:

1. Доповнення: $\bar{A} = \{\omega : \omega \notin A\}$

2. Об'єднання: $A \cup B = \{\omega : \omega \in A \vee \omega \in B\}$ *

* \vee – диз'юнкція – „або”

3. Перетин/переріз: $A \cap B = \{\omega : \omega \in A \wedge \omega \in B\}$ **

** \wedge – кон'юнкція – „і”

4. Різниця: $A \setminus B = \{\omega : \omega \in A \wedge \omega \notin B\}$

Зв'язок між подіями

1. Правило Де-Моргана:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

$$(\bar{\bar{A}} = A)$$

$$A \cup B = \overline{\bar{A} \cap \bar{B}}$$

$$A \cap B = \overline{\bar{A} \cup \bar{B}}$$

2. (різниця через перетин): $A \setminus B = A \setminus (A \cap B) = A \cap \bar{B}$

2. Імовірності в дискретних просторах елементарних подій.

Ω – дискретна множина

ω_i має імовірність P_i

$$0 \leq P_i \leq 1$$

$$\sum_{i=1}^n P_i = 1$$

Нехай $A = \{\omega_{i_1}; \omega_{i_2}; \dots \omega_{i_m}\}$

$$P(A) = \sum_{\omega_i \in A} P_i = P_{i_1} + P_{i_2} + \dots + P_{i_m}$$

3. Дати частотне та класичне означення імовірності та вказати властивості імовірностей.

Нехай є довільний стохастичний експеримент Ω

$\nu_n(A)$ – частота події A ; Скільки разів подія A з'явилася при n повтореннях цього експерименту; $0 \leq \nu_n(A) \leq n$

$$\text{Імовірність } P_n(A) = \frac{\text{частота}}{n} = \frac{\nu_n(A)}{n}$$

Властивості:

1. $0 \leq P_n(A) \leq 1$

2. $P_n(\Omega) = \frac{n}{n} = 1$ – достовірна подія

3. Нехай є дві події A, B , $A \cap B \neq \emptyset$, то

$$\begin{aligned} \nu_n(A \cup B) &= \nu_n(A) + \nu_n(B) \\ P_n(A \cup B) &= P_n(A) + P_n(B) \end{aligned}$$

Частотне означення імовірності (або озн. імов. Мізеса)

Якщо існує границя $\lim_{n \rightarrow \infty} P_n(A) = P(A)$ – імовірність події A ; при цьому подія A – стохастично стійка.

Недоліки такого означення: $n \rightarrow \infty$ (неможливо зробити ∞ експериментів); якщо не існує границі – не існує імовірності.

Класичне означення імовірності (XVI-XVII ст.)

Нехай Ω – скінченний простір. ($\Omega = \{\omega_1; \omega_2 \dots \omega_n\}, n < \infty$) і всі елементарні наслідки рівноможливі ($p_i = \frac{1}{n}$); тоді якщо $A = \{\omega_{i_1}; \omega_{i_2} \dots \omega_{i_m}\}$, то $P(A) = \underbrace{\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}}_{m \text{ разів}} = \frac{m}{n}$

$P(A) = \frac{m}{n}$ – імовірність довільної події A дорівнює числу m події A поділити на загальну кількість точок.

4. Дати геометричне означення імовірності. Задача Бюффона.

$\Omega < R^n$ (Евклідовий простір)

$R^1 = (-\infty; +\infty)$, міра – довжина

R^2 – 2х мірний, міра – площа

R^3 – 3х мірний, міра – об'єм

Міра Лебега $m(A)$

$$m(\Omega) < \infty$$

$$P(A) = \frac{m(A)}{m(\Omega)}$$

Властивості:

1) $0 \leq P(A) \leq 1$

2) $P(\Omega) = \frac{m(\Omega)}{m(\Omega)} = 1$

3) $P(A \cup B) = \frac{m(A) + m(B)}{m(\Omega)} = P(A) + P(B)$

Задача Бюффона

Нехай є простір розчерчений паралельними прямими, відстань між прямими – $2a$. На цю площину довільним чином кидається голка довжиною $2l$. Яка імовірність того, що голка перетне одну з прямих?

$$\Omega = \{(x, \varphi) : 0 \leq x \leq a; 0 \leq \varphi \leq \pi\}$$

$$A = \{..., x < l \sin \varphi\}$$

$$m(\Omega) = a\pi$$

$$m(A) = \int_0^\pi l \sin \varphi d(\varphi) = l(\cos \varphi) \Big|_0^\pi = l(1 + 1) = 2l$$

$$P(A) = \frac{2l}{a\pi} = \frac{m}{n}$$

(що можна використати для наближеного обчислення π)

$$\pi = \frac{2l n}{a m}$$

5. Аксиоми теорії імовірностей.

Аксиома1: $P(A), \forall A \in F$

A2: $P(\Omega) = 1$

A3 (зліченої адитивності):

$$A_i \in F, A_i \cap A_j = \emptyset, i \neq j$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

A4 (скінченної адитивності): $P(A_1 \cup A_2) = P(A_1) + P(A_2); A_1 \cap A_2 = \emptyset$

6. Властивості імовірностей.

1) $P(\bar{A}) = 1 - P(A)$

2) $P(\emptyset) = 0$

3) $A \subset B$

$$P(B \setminus A) = P(B) - P(A)$$

4) $A \subset B$

$$P(A) < P(B)$$

5) (наслідок 4ї): $P(A) \leq 1$

6) Теорема додавання імовірностей: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

7. Умовні імовірності. Приклад.

$P(A/B)$ – A при умові B

Формальне означення: $\frac{P(A \cap B)}{P(B)}, P(B) > 0$

Формула множення імовірностей: $P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$

8. Довести формулу повної імовірності.

H_1, H_2, \dots, H_n – повна група подій, якщо:

1) $H_i \cap H_j = \emptyset, i \neq j, \bigcup_{i=1}^n H_i = \Omega$

2) $P(H_i) > 0$

$$P(A) = \sum_{i=1}^n P(A/H_i)P(H_i)$$

Доведення:

$$P(A) = P(A \cap \Omega) = P\left(A \cap \bigcup_{i=1}^n H_i\right) = P\left(\bigcup_{i=1}^n (A \cap H_i)\right) = \sum_{i=1}^n P(A \cap H_i) = \sum_{i=1}^n P(A/H_i)P(H_i)$$

9. Довести формулу Байеса.

Формула Байеса

$$P(H_i/A) = \frac{P(A \cap H_i)}{P(A)} = \frac{P(A/H_i)P(H_i)}{\sum_{j=1}^n P(A/H_j)P(H_j)}$$

$P(H_i)$ – апріорні імовірності

$P(H_i / A)$ – апостеріорні імовірності

10. Незалежні події. Властивості незалежних подій.

1) A і B – незалежні, якщо $P(A \cap B) = P(A)P(B)$

$$2) P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

Незалежність в сукупності

A_1, A_2, \dots, A_n називаються незалежними в сукупності, якщо для будь-якого цілого

$k = 2, 3, \dots, n$ для будь-яких k подій $A_{i_1}, A_{i_2}, A_{i_3}, \dots, A_{i_k}$ імовірність перетину

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \dots \cap A_{i_n}) = P(A_{i_1})P(A_{i_2})P(A_{i_3}) \dots P(A_{i_n})$$

11. Математичне сподівання дискретних випадкових величин.

Математичне сподівання – середнє ймовірнісне значення випадкової величини.

$$M\xi = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

12. Властивості.

Властивості математичного сподівання:

1. Якщо $\xi = C$, то $M(\xi) = C$
2. $M(\xi \pm C) = M(\xi) \pm C$
3. $M(C \cdot \xi) = C \cdot M(\xi)$
4. $\xi \geq 0 (x_i \geq 0) \Rightarrow M\xi \geq 0$
5. $M(\xi \pm \eta) = M(\xi) \pm M(\eta)$
6. Якщо ξ і η – незалежні, то $M(\xi \cdot \eta) = M(\xi) \cdot M(\eta)$
7. Математичне сподівання функції від випадкової величини:

$$Mf(\xi) = \sum_{i=1}^n (x_i) p_i$$

13. Дисперсія дискретних випадкових величин. Властивості.

Дисперсія – середньоквадратичне відхилення випадкової величини від свого математичного сподівання

$$D\xi = M(\xi - M\xi)^2 = M\xi^2 - (M\xi)^2$$

Властивості дисперсії

1. $D\xi \geq 0$
2. $D\xi = 0 \Leftrightarrow \xi = C$
3. $D(\xi \pm C) = D\xi$

$$4. D(C\xi) = C^2 D\xi$$

$$5. \text{ Якщо } \xi \text{ і } \eta - \text{ незалежні, то } D(\xi \pm \eta) = D\xi + D\eta$$

14. Коефіцієнт кореляції випадкових величин. Властивості.

Коефіцієнт кореляції – міра залежності випадкових величин

$$\rho(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{D(\xi)} \cdot \sqrt{D(\eta)}} \quad , D(\xi), D(\eta) \geq 0$$

$$\text{cov}(\xi, \eta) = M(\xi - M\xi)(\eta - M\eta) = M(\xi\eta) - M\xi \cdot M\eta$$

Властивості:

$$1. \text{ Якщо } \xi, \eta - \text{ незалежні, то } \rho = 0$$

$$2. |\rho| \leq 1$$

$$3. |\rho| = 1 \Leftrightarrow \eta = a\xi + b - \text{ лінійна залежність (найсильніша)}$$

$$\rho = 1, a > 0 - \text{ пряма залежність}$$

$$\rho = -1, a < 0 - \text{ обернена залежність}$$

Зауваження: З того, що $\rho = 0$ не випливає незалежність випадкових величин.

15. Локальна та інтегральна теорема Муавра-Лапласа та їх застосування.

Локальна теорема Муавра-Лапласа:

$$\text{Якщо } n \rightarrow \infty, \left| \frac{i - np}{\sqrt{npq}} \right| \in C, \text{ тоді } \lim_{n \rightarrow \infty} p_n(i) \sqrt{2\pi npq} \cdot e^{-x^2/2} = 1, \text{ де } x = \frac{i - np}{\sqrt{npq}}$$

$$p_n(i) \approx \frac{1}{\sqrt{2\pi npq}} \cdot e^{-x^2/2} = \frac{\varphi(x)}{\sqrt{npq}}, \text{ де } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ (функція Лапласа)}$$

Інтегральна теорема Муавра-Лапласа:

$$p\{a \leq \xi \leq b\} = \sum_{i=a}^b P_n(i)$$

$$\lim_{n \rightarrow \infty} P\left\{x_1 \leq \frac{\xi - np}{\sqrt{npq}} \leq x_2\right\} = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-t^2/2} dt = \int_{x_1}^{x_2} \varphi(t) dt$$

$$p\{a \leq \xi \leq b\} = p\left\{\frac{a - np}{\sqrt{npq}} \leq \frac{\xi - np}{\sqrt{npq}} \leq \frac{b - np}{\sqrt{npq}}\right\} \approx \int_{\frac{a - np}{\sqrt{npq}}}^{\frac{b - np}{\sqrt{npq}}} \varphi(t) dt$$

$\varphi(x)$ – інтегральна функція Лапласа

$$x_1 = \frac{a - np}{\sqrt{npq}}$$

$$x_2 = \frac{b - np}{\sqrt{npq}}$$

Використовується якщо $np < 0,1$ або $np > 10$ (конспект)

16. Теорема Пуассона та її застосування.

Якщо $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \lambda > 0$, то

$$\lim_{n \rightarrow \infty} p_n(i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

$$p_n(i) \approx \frac{\lambda^i}{i!} e^{-\lambda}, \lambda = np$$

Використовується якщо $0,1 \leq np \leq 10$ (конспект)

17. Біноміальний розподіл. Обчислити $M\xi, D\xi$.

n – кількість експериментів

p – імовірність успіху

q – імовірність невдачі

$$p + q = 1$$

ξ – загальне число успіхів в n експериментах

$$\xi \in \{0 \dots n\}$$

$$P_i = P\{\xi = i\} = C_n^i p^i \cdot q^{n-i}$$

m – найбільш імовірне число успіхів

$$np - q \leq m \leq np + p$$

Математичне сподівання:

$$\begin{aligned} M\xi &= \sum_{i=1}^n i \cdot C_n^i p^i q^{n-i} = \sum_{i=1}^n i \cdot \frac{n!}{i!(n-i)!} p^i q^{n-i} = n \sum_{i=1}^n \frac{(n-1)!}{(i-1)!(n-i)!} p^i q^{n-i} = \\ &= np \sum_{i=1}^n C_{n-1}^{i-1} p^{i-1} q^{n-1-(i-1)}, (i-1=k) = np \sum_{k=0}^{n-1} C_{n-1}^k p^k q^{n-1-k} = np(p+q)^{n-1} = np \end{aligned}$$

Дисперсія:

$$M\xi_i = 0 \cdot q + 1 \cdot p = p$$

$$M\xi_i^2 = 0^2 \cdot q + 1^2 \cdot p = p$$

$$D\xi_i = p - p^2 = p(1-p) = pq$$

$$D\xi = D(\xi_1 + \xi_2 + \dots + \xi_n) = D\xi_1 + D\xi_2 + \dots + D\xi_n = \underbrace{pq + pq + \dots + pq}_{n \text{ разів}} = npq$$

18.Геометричний розподіл. Обчислити $M\xi, D\xi$.

Імовірність успіху – p , невдачі – q . Експерименти проводяться до першої появи успіху;
 ξ – кількість послідовних невдач.

$$P_i = P\{\xi = i\} = q^i p$$

$$M\xi = \sum_{i=1}^{\infty} i \cdot q^i p = p \sum_{i=1}^{\infty} i q^i = pq \sum_{i=1}^{\infty} i q^{i-1} = pq \cdot \frac{1}{(1-q)^2} = \frac{q}{p}$$

$$\begin{aligned} M\xi^2 &= \sum_{i=1}^{\infty} i^2 \cdot q^i p = p \sum_{i=1}^{\infty} i^2 \cdot q^i p = p \sum_{i=1}^{\infty} i^2 q^i = p \sum_{i=1}^{\infty} [i(i-1) + i] q^i = p \sum_{i=1}^{\infty} [i(i-1) + i] q^i = \\ &= p \sum_{i=1}^{\infty} i(i-1) q^i + p \sum_{i=1}^{\infty} i \cdot q^i = pq^2 \frac{2}{(1-q)^3} + \frac{q}{p} = \frac{2q^2}{p^2} + \frac{q}{p} \end{aligned}$$

$$D\xi = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q^2}{p^2} + \frac{q}{p} = \frac{q^2 + qp}{p^2} = \frac{q(q+p)}{p^2} = \frac{q}{p^2}$$

19.Пуассонівський розподіл. Обчислити $M\xi, D\xi$.

$$\xi \in \{0, 1, 2, 3, \dots, n\}$$

$$p_i = P\{\xi = i\} = \frac{\lambda^i}{i!} e^{-\lambda}$$

λ - додатне число, параметр розподілу

$$M\xi = \sum_{i=1}^{\infty} i p_i = \sum_{i=1}^{\infty} i \cdot \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda} \lambda (1 + \lambda + \frac{\lambda^2}{2} \dots) = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda$$

$$\begin{aligned} M\xi^2 &= \sum_{i=1}^{\infty} i^2 \cdot \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \cdot \sum_{i=1}^{\infty} \frac{i \cdot \lambda^i}{(i-1)!} = e^{-\lambda} \cdot \sum_{i=1}^{\infty} \frac{(i-1+1)\lambda^i}{(i-1)!} = e^{-\lambda} \sum_{i=1}^{\infty} \frac{(i-1)\lambda^i}{(i-1)!} + e^{-\lambda} \sum_{i=1}^{\infty} \frac{1 \cdot \lambda^i}{(i-1)!} = \\ &= e^{-\lambda} \cdot \lambda^2 \sum_{i=1}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} + \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda} \cdot \lambda^2 (1 + \lambda + \frac{\lambda^2}{2} \dots) + e^{-\lambda} \lambda (1 + \lambda + \frac{\lambda^2}{2} \dots) = \\ &= e^{-\lambda} \cdot \lambda^2 \cdot e^{\lambda} + e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda^2 + \lambda \end{aligned}$$

$$D\xi = M\xi^2 - (M\xi)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

20.Функції розподілу випадкових величин. Властивості.

$F(x) = P\{\xi < x\}$ – функція розподілу

Властивості:

- 1) $F(+\infty) = 1; F(-\infty) = 0$
- 2) $F(x)$ – неперервна зліва
- 3) $F(x)$ – неспадна: $a < b \rightarrow F(a) \leq F(b)$
- 4) $P\{a \leq \xi < b\} = F(b) - F(a)$
 $P\{a < \xi < b\} = F(b) - F(a+0)$
 $P\{a \leq \xi \leq b\} = F(b+0) - F(a)$
 $P\{a < \xi \leq b\} = F(b+0) - F(a+0)$
- 5) Якщо a, b – точки неперервності $F(x)$, то

$$P\{a \leq \xi < b\} = P\{a \leq \xi \leq b\} = P\{a < \xi < b\} = P\{a < \xi \leq b\} = F(b) - F(a)$$

- 6) $P\{\xi = a\} = F(a+0) - F(a)$
якщо a – точка неперервності, то $P\{\xi = a\} = 0$

21. Щільність випадкових величин. Властивості.

$F'(x) = p(x)$, де $p(x)$ – щільність неперервної випадкової величини.

Властивості:

- 1) $F'(x) \geq 0, p(x) \geq 0$
- 2) $P\{a \leq \xi < b\} = F(b) - F(a) = \int_a^b p(x)dx$
- 3) $F(x) = \int_{-\infty}^x p(u)du, F(+\infty) = 1, \int_{-\infty}^{+\infty} p(u)du = 1$
- 4) $P\{x \leq \xi < x + \Delta x\} = F(x + \Delta x) - F(x) = F'(x)\Delta x + o(\Delta x) = p(x)\Delta x + o(\Delta x)$

22. Математичне сподівання та дисперсія неперервних випадкових величин.

Математичне сподівання:

$$M\xi = \int_{-\infty}^{+\infty} x \cdot p(x)dx$$

Властивості:

- 1) $MC = C$
- 2) $M(C\xi) = CM\xi$
- 3) $M(\xi \pm C) = M\xi \pm C$

$$4) \xi \geq 0 \Rightarrow M\xi \geq 0$$

$$5) M(\xi \pm \eta) = M(\xi) \pm M(\eta)$$

$$6) \text{ Якщо } \xi \text{ і } \eta \text{ – незалежні, то } M(\xi \cdot \eta) = M\xi \cdot M\eta$$

$$Mf(\xi) = \int_{-\infty}^{+\infty} f(x)p(x)dx$$

Дисперсія:

$$M^2\xi = \sum_{-\infty}^{+\infty} x^2 p(x)dx$$

$$D\xi = M(\xi - M\xi)^2$$

23.Рівномірний розподіл на відрізку $[a, b]$. Обчислити $M\xi, D\xi$.

$$F(x) = \begin{cases} 0, x \leq a \\ \frac{x-a}{b-a}, a < x \leq b \\ 1, x > b \end{cases}$$

$$p(x) = \begin{cases} 0, x \leq a \\ \frac{1}{b-a}, a < x \leq b \\ 0, x > b \end{cases}$$

$$M\xi = \frac{a+b}{2}$$

$$D\xi = \frac{(b-a)^2}{12}$$

24.Показниковий розподіл. Обчислити $M\xi, D\xi$.

$$F(x) = \begin{cases} 0, x \leq 0 \\ 1 - e^{-\lambda x}, x > 0, \lambda > 0 \end{cases}$$

$$p(x) = \begin{cases} 0, x \leq 0 \\ \lambda e^{-\lambda x}, x > 0 \end{cases}$$

$$M\xi = \frac{1}{\lambda}$$

$$D\xi = \frac{1}{\lambda^2}$$

25.Нормальний розподіл. Обчислити $M\xi$.

$$N(a, \sigma^2)$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(u-a)^2}{2\sigma^2}} du, \sigma > 0$$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} - \text{дзвіноподібна функція}$$

Якщо $a = 0$, $\sigma = 1$, то розподіл стандартний

$$M\xi = a$$

26.Нормальний розподіл. Обчислити $D\xi$.

$$N(a, \sigma^2)$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(u-a)^2}{2\sigma^2}} du, \sigma > 0$$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} - \text{дзвіноподібна функція}$$

Якщо $a = 0$, $\sigma = 1$, то розподіл стандартний

$$D\xi = \sigma^2$$

27.Довести нерівність Чебишева.

Нерівність Чебишева:

Нехай для ξ існує $M\xi, D\xi$

$$P\{|\xi - M\xi| > \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}, D\xi < \varepsilon^2$$

Тоді для $\forall \varepsilon > 0$

$$P\{|\xi - M\xi| \leq \varepsilon\} \geq 1 - \frac{D\xi}{\varepsilon^2}$$

Правило 3σ : Нехай $D\xi = \sigma^2$, тоді $\varepsilon = 3\sigma$

28. Закон великих чисел. (Теорема Хінчина, теорема Чебишева).

Будь-яке твердження про збіжність середньоарифметичних випадкових величин називається законом великих чисел.

$$\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$$

Теорема Хінчина:

Нехай $\xi_1, \xi_2, \dots, \xi_n$ – послідовність незалежних, однаково розподілених випадкових величин (п.н.о.р.в.в.)

$$M\xi_k = a < +\infty$$

$$\text{Тоді } \frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow[n \rightarrow \infty]{P} a$$

Теорема Чебишева:

Нехай $\xi_1, \xi_2, \dots, \xi_n$ – п.н.в.в.

$$M\xi_k = a_k < +\infty$$

$$D\xi_k = \sigma_k^2 \leq C < +\infty$$

$$\text{Тоді: } \frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n a_k \xrightarrow[n \rightarrow \infty]{P} 0$$

Доведення теореми Чебишева:

$$\lim_{n \rightarrow \infty} P\left\{ \underbrace{\frac{1}{n} \sum_{k=1}^n \xi_k}_{=\bar{\xi}} - \underbrace{\frac{1}{n} \sum_{k=1}^n a_k}_{=M\bar{\xi}} > \varepsilon \right\} = 0 \quad \forall \varepsilon > 0$$

$$P\{|\bar{\xi} - M\bar{\xi}| > \varepsilon\} \leq \frac{D\bar{\xi}}{\varepsilon^2} = \frac{D(\frac{1}{n} \sum_{k=1}^n \xi_k)}{\varepsilon^2} = \frac{1}{n^2 \xi^2} D(\sum_{k=1}^n \xi_k) = \frac{\sum_{k=1}^n D\xi_k}{n^2 \xi^2} \leq \frac{C}{n \xi^2} \rightarrow 0, \text{ коли } n \rightarrow \infty$$

29. Центральна гранична теорема.

Нехай $\xi_1, \xi_2 \dots \xi_n$ – п.н.о.р.в.в.

$$M\xi_n = a < +\infty; D\xi = \sigma^2 < \infty; \sigma^2 > 0 (\xi \neq const)$$

Тоді:

$$\lim_{n \rightarrow \infty} P\left\{ \frac{\sum_{k=1}^n \xi_k - na}{\sigma\sqrt{n}} < x \right\} = \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt}_{\text{Функція розподілу } N(0,1)}$$

$$\frac{\sum_{k=1}^n \xi_k - na}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} N(0;1)$$

$$\sum_{k=1}^n \xi_k \approx na + \underbrace{\sqrt{n} \cdot \sqrt{\sigma} \cdot N(0;1)}_{\text{Стандартне збурення}}$$

30. Емпірична функція розподілу.

$F_\xi(x)$ – невідома функція розподілу

$$\zeta = (\xi_1, \xi_2 \dots \xi_n) \quad X = (x_1, x_2 \dots x_n)$$

$\nu_n(x) = \xi_1, \xi_2 \dots \xi_n < x$ (випадкова частота)

$$F_n^*(x) = \frac{\nu_n(x)}{n} \quad \text{— емпірична функція розподілу}$$

$$\left\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$$

$$P\left\{F_n^*(x) = \frac{k}{n}\right\} = C_n^k (F_\xi(x))^k (1 - F_\xi(x))^{n-k}, \quad k = 0, 1 \dots n$$

Асимптотичні властивості емпіричної функції розподілу:

$$\text{Теорема Глівєнка: } F_n^*(x) \xrightarrow[n \rightarrow \infty]{1} F_\xi(x)$$

Теорема Колмогорова:

ξ – неперервна

$$\lim_{n \rightarrow \infty} P\left\{ \sqrt{n} \cdot \max_{-\infty < x < +\infty} |F_n^*(x) - F_\xi(x)| < t \right\} = K(t)$$

$K(t)$ – функція розподілу Колмогорова

$$K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 t^2}$$

Побудова емпіричної функції розподілу:

I. Для дискретної статистичної таблиці

$$F_n^*(x) = \begin{cases} 0, & x \leq y_1 \\ \frac{m_1}{n}, & y_1 < x \leq y_2 \\ \frac{m_1 + m_2}{n}, & y_2 < x \leq y_3 \\ \frac{m_1 + m_2 + m_3}{n}, & y_3 < x \leq y_4 \\ \dots \\ \frac{n}{n} = 1, & x > y_s \end{cases}$$

II. Для інтервальної статистичної таблиці

$$F_n^*(x) = \begin{cases} 0, & x \leq y_1 \\ \frac{m_1}{2n}, & y_1 < x \leq y_1^* \\ \frac{m_1 + \frac{m_2}{2}}{n}, & y_1^* < x \leq y_2^* \\ \frac{m_1 + m_2 + \frac{m_3}{2}}{n}, & y_2^* < x \leq y_3^* \\ \dots \\ 1, & x > y_s^* \end{cases}$$

31. Гістограма та полігон частот.

Полігон частот:

Для дискретної статистичної таблиці:

Ордината точки – $\frac{m_i}{n}$, абсциса – y_i ; Справа/зліва опускати ламану до нуля, останній

інтервал дорівнює попередньому (якщо є обмеження, наприклад не менше нуля, опускати в точку обмеження)

Для інтервальної статистичної таблиці:

Ордината – $\frac{m_i}{n \cdot h_i}$ ($h_i = y_{i+1} - y_i$, довжина інтервалу), абсциса – середина інтервалу.

Справа/зліва опускати ламану до нуля, перший/останній інтервал дорівнює другому/передостанньому (якщо є обмеження, наприклад не менше нуля, опускати в точку обмеження)

Гістограма – фігура, яка складається з прямокутників, основою яких є інтервали довжини

h_i , а висоти – $\frac{m_i}{n \cdot h_i}$.

Площа гістограми:

$$S = h_1 \frac{m_1}{nh_1} + h_2 \frac{m_2}{nh_2} + \dots + h_s \frac{m_s}{nh_s} = \frac{m_1 + m_2 + \dots + m_s}{n} = \frac{n}{n} = 1$$
 (так само як площа під щільністю)

32.Мода.

Мода – число, яке найчастіше зустрічається у вибірці. (ака „найпопулярніше” число)

Для дискретної таблиці $Mo = y_i, m_i = \max m_k$; Якщо два послідовних значення мають макс m_k , то береться сер. Арифметичне цих значень.

Для інтервальної таблиці:

$$m_i = \max m_k [y_i, y_{i+1}]$$

$$Mo = y_i + h_i \cdot \frac{m_i - m_{i-1}}{2m_i - m_{i-1} - m_{i+1}}$$

Якщо два максимуми – порахувати Mo для кожного окремо і взяти сер. арифметичне.

33.Медіана.

$x_{1/2}$ – медіана ξ , якщо $P\{\xi < x_{1/2}\} = P\{\xi \geq x_{1/2}\} = \frac{1}{2}$

$$F_{\xi}(x_{1/2}) = \frac{1}{2}$$

1. Для дискретної таблиці

а) $n = 2k + 1$

$$x_{k+1} = Me$$

б) $n = 2k$

$$Me = \frac{x_k + x_{k+1}}{2}$$

2. Для інтервальної таблиці

$$\left. \begin{array}{l} a) n = 2k + 1 \quad x_{k+1} \\ b) n = 2k \quad x_k, x_{k+1} \end{array} \right\} \Rightarrow (y_i, y_{i+1})$$

$$Me = y_i + h_i \cdot \frac{\frac{n}{2} - (m_1 + m_2 + \dots + m_{i-1})}{m_i}$$

34. Статистичне оцінювання невідомих параметрів розподілів. Властивості оцінок.

Нехай є вибірка ξ з функцією розподілу $F_\xi(x, \theta)$. Оцінка невідомого параметра θ (функція від вибірки) називається статистикою.

$$\hat{\theta}_n = f\{\xi_1, \xi_2, \dots, \xi_n\}$$

Класифікація оцінок:

1. Оцінка $\hat{\theta}_n$ невідомого параметра θ називається *незміщеною*, якщо $M \hat{\theta}_n = \theta$; якщо $M \hat{\theta}_n \neq \theta$ – оцінка *зміщена*.

Зміщена оцінка: $M \hat{\theta}_n = \theta + b(\theta)$, $b(\theta) \neq 0$ – зсув оцінки

2. Незміщена $\hat{\theta}_n$ називається *спроможною*, якщо: $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{P} \theta$

3. $\hat{\theta}_n$ параметра θ називається *асимптотично нормальною*, якщо $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{} N(0, \sigma^2)$

4. Незміщена $\hat{\theta}_n$ параметра θ називається *оптимальною*, якщо дисперсія цієї оцінки найменша серед дисперсій усіх можливих незміщених оцінок.

35. Вибіркове математичне сподівання.

$$\theta = M\xi = a$$

$$\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$$

\bar{x} – вибіркове середнє

1. $M\bar{\xi} = \frac{1}{n} \sum_{k=1}^n M\xi_k = \frac{na}{n} = a \Rightarrow$ незміщенна

2. $\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow[\text{закон великих чисел}]{P} a \Rightarrow$ спроможна

3.

$$\sqrt{n}(\bar{\xi} - a) = \sqrt{n}\left(\frac{1}{n} \sum_{k=1}^n \xi_k - a\right) = \frac{\sum_{k=1}^n \xi_k - na}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} N(0, \sigma^2) \Rightarrow \text{ас. нормальна}$$

центральна гранична теорема

$$4. D\bar{\xi} = D\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right) = \frac{1}{n^2} \sum_{k=1}^n D\xi_k = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \Rightarrow D\bar{\xi} \xrightarrow[n \rightarrow \infty]{} 0$$

Зауваження: для більшості класичних розподілів $\bar{\xi}$ буде оптимальною оцінкою, але є випадки коли вона не є оптимальною (наприклад для рівномірного розподілу)

36. Вибіркова дисперсія.

$$\sigma^2 = D\xi = \theta$$

1. $a = M\xi$ – відоме

$$S^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - a)^2$$

$$MS^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - a)^2 = \frac{n\sigma^2}{n} = \sigma^2 \Rightarrow \text{незміщена}$$

2. a – невідоме

$$\bar{S}^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})^2$$

$$\begin{aligned} M\bar{S}^2 &= M\left(\frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})^2\right) = \frac{1}{n} \sum_{k=1}^n M(\xi_k - a + a - \bar{\xi})^2 = \frac{1}{n} \sum_{k=1}^n \underbrace{M(\xi_k - a)^2}_{=\sigma^2} + \\ &+ \frac{2}{n} \sum_{k=1}^n M(\xi_k - a)(a - \bar{\xi}) + \frac{1}{n} \sum_{k=1}^n \underbrace{M(a - \bar{\xi})^2}_{=\sigma^2} = \frac{n\sigma^2}{n} - \frac{2}{n} M(\bar{\xi} - a) \cdot \sum_{k=1}^n (\xi_k - a) + \\ &+ \frac{1}{n} \cdot n \cdot \frac{\sigma^2}{n} = \sigma^2 - \frac{2}{n} M(\bar{\xi} - a)(n\bar{\xi} - na) + \frac{\sigma^2}{n} = \sigma^2 - \frac{2}{n} n \cdot \underbrace{M(\bar{\xi} - a)^2}_{=D\bar{\xi}} + \frac{\sigma^2}{n} = \\ &= \sigma^2 - 2 \cdot \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \sigma^2 + \frac{\sigma^2}{n} \neq \sigma^2 - \text{оцінка зміщена, } b(\sigma^2) = -\frac{\sigma^2}{n} - \text{зсув оцінки} \end{aligned}$$

$$M\bar{S}^2 = \sigma^2 \left(\frac{n-1}{n}\right)$$

тому:

$$\hat{S}^2 = \frac{n}{n-1} \bar{S}^2 = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})^2 = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2$$

$$M\hat{S}^2 = \frac{n}{n-1} M\bar{S}^2 = \frac{n}{n-1} \sigma^2 \left(\frac{n-1}{n}\right) = \sigma^2 - \text{незміщена}$$

\hat{S}^2 – спроможна і асимптотично нормальна оцінка, але не для всіх розподілів оптимальна.

37.Метод максимальної правдоподібності. Приклад.

$$\text{ММП: } p(x, \theta) = \begin{cases} F'_\xi(x, \theta) - \xi \text{ неперервна} \\ p\{\xi = x\} - \xi \text{ дискретна} \end{cases}$$

$$\text{Функція правдоподібності } L(x, \theta) = \prod_{k=1}^n p(x_k, \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot \dots \cdot p(x_n, \theta)$$

$$\max L(x, \theta) = L(x, \hat{\theta}_n)$$

$\hat{\theta}_n$ – точка максимуму функції правдоподібності або оцінка МПП

Спрощення:

$$\ln L(x, \theta) = \ln L(x, \hat{\theta}_n)$$

$$(\ln L)'_{\theta} = \frac{\partial \ln L(x, \theta)}{\partial \theta} = 0 \text{ – рівняння правдоподібності. Розв'язок рівняння - } \hat{\theta}_n.$$

$$\text{Перевірка } (\ln L)''_{\theta} = \frac{\partial^2 \ln L(x, \theta)}{\partial^2 \theta}, \text{ якщо } \theta = \hat{\theta}_n < 0 \text{ – max}$$

38.Надійні інтервали для математичного сподівання нормального розподілу.

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

$$M\xi = a \quad D\xi = \sigma^2$$

1) σ^2 – відомий

$$\bar{x} - C_{\alpha} \frac{\sigma}{\sqrt{n}} < a < \bar{x} + C_{\alpha} \frac{\sigma}{\sqrt{n}}$$

2) σ^2 – невідомий

$$\bar{\xi} - t_{\alpha} \frac{\hat{S}}{\sqrt{n}} < a < \bar{\xi} + t_{\alpha} \frac{\hat{S}}{\sqrt{n}}$$

39.Надійні інтервали для дисперсії нормального розподілу.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

$$M\xi = a \quad D\xi = \sigma^2$$

1) a – відомий

$$\frac{nS^2}{\chi_2^2} < \sigma^2 < \frac{nS^2}{\chi_1^2}$$

2) a – невідомий

$$\frac{(n-1)S^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_1^2}$$

40. Надійні інтервали для параметрів пуассонівського та біноміального розподілів.

Пуассонівський розподіл:

$$p\{\xi = k\} = \frac{\theta^k}{k!} e^{-\theta} \quad M = \theta \quad \hat{\theta}_n = \bar{\xi}$$

$$\bar{x} - C_\alpha \sqrt{\frac{\bar{x}}{n}} < \theta < \bar{x} + C_\alpha \sqrt{\frac{\bar{x}}{n}}$$

Біноміальний розподіл:

$$P\{\xi = k\} = C_n^k p^k (1-p)^{n-k}$$

$\theta = p - ?$, x – число успіхів n експериментів

$$\hat{p}_n = \frac{x}{n}$$

$$\underbrace{\frac{x}{n} - \frac{C_\alpha}{n} \sqrt{\frac{x(n-x)}{n}}}_{\text{Якщо } < 0, \text{ ставимо } 0} < p < \underbrace{\frac{x}{n} + \frac{C_\alpha}{n} \sqrt{\frac{x(n-x)}{n}}}_{\text{Якщо } > 1, \text{ ставимо } 1}$$

41. Критерії згоди. Критерій χ^2 .

Критерій згоди Пірсона (хі-квадрат) використовується для перевірки гіпотез у поліноміальній схемі. А саме: Нехай проводиться n незалежних випробувань, кожне з яких може мати r різних результатів A_1, A_2, \dots, A_r . Необхідно перевірити гіпотезу про те, що імовірності цих результатів дорівнюють p_1, p_2, \dots, p_r , якщо в послідовності випробувань ці результати зустрілися m_1, m_2, \dots, m_r разів.

Критерій χ^2 для перевірки гіпотези про розподіл:

Нехай є випадкова величина ξ з функцією розподілу $F_\xi(x)$

$H_0 : F_\xi(x) = F(x)$ (проста гіпотеза – без параметрів)

$H_1 : F_\xi(x) \neq F(x)$

$$n = m_1 + m_2 + \dots + m_s$$

$$p_i = P\{\xi \in \Delta_i / H_0^+\}$$

$$\chi_n^2 = \sum_{i=1}^s \frac{(m_i - np_i)^2}{np_i} \xrightarrow{n \rightarrow \infty} \chi^2(S-1)$$

Розрахунок: $1 - \alpha$ (стовпчик); $k = S - 1$ (рядок);

Якщо:

$$\chi_n^2 < \chi_\alpha^2 \Rightarrow H_0^+$$

$$\chi_n^2 \geq \chi_\alpha^2 \Rightarrow H_0^-$$

Якщо перевіряється складна гіпотеза:

$$H_0 : F_\xi(x) = F(x, \theta)$$

$$H_1 : F_\xi(x) \neq F(x, \theta)$$

$$\theta \rightarrow \hat{\theta}_n \text{ Оцінка МПП}$$

l – кількість невідомих параметрів. Тоді число степенів свободи $k = S - 1 - l$

42. Критерій χ^2 для перевірки гіпотези однорідності вибірок.

Є k випадкових величин n_1, n_2, \dots, n_k

$$n = n_1 + n_2 + \dots + n_k$$

$$H_0 : F_1(x) = F_2(x) = \dots = F_k(x) \text{ (всі функції розподілу однакові)}$$

$$H_1 : F_i(x) \neq F_j(x)$$

$$V_{ij}$$

$$j = 1, 2, \dots, k$$

$$i = 1, 2, \dots, l$$

$$V_{i0} = \sum_{j=1}^k V_{ij}$$

$$m_{ij} = \frac{V_{i0} \cdot n_j}{n} - \text{теоретична частота}$$

$$\chi_n^2 = \sum_{k=1}^l \sum_{j=1}^k \frac{(V_{ij} - m_{ij})^2}{m_{ij}} \xrightarrow{n \rightarrow \infty} \chi^2((l-1)(k-1))$$

$$\alpha \quad (1-\alpha) \quad (l-1)(k-1) \rightarrow \chi_\alpha^2$$

$$\chi_n^2 < \chi_\alpha^2 \Rightarrow H_0^+$$

$$\chi_n^2 \geq \chi_\alpha^2 \Rightarrow H_0^-$$

43. Критерій χ^2 для перевірки гіпотези незалежності двох вибірок.

$$(\xi, \eta) \quad (x_i, y_j) \quad i = 1, 2, 3 \dots l; j = 1, 2, 3 \dots k$$

$$V_{ij} = (x_i, y_j)$$

$$H_0 : \xi \text{ і } \eta - \text{незалежні}$$

$$H_1 : \xi \text{ і } \eta - \text{залежні}$$

$$V_{i0} = \sum_{j=1}^k V_{ij}$$

$$V_{0j} = \sum_{i=1}^l V_{ij}$$

$$n = \sum_{i=1}^l \sum_{j=1}^k V_{ij}$$

$$m_{ij} = \frac{V_{i0} \cdot V_{0j}}{n}$$

$$\chi_n^2 = \sum_{k=1}^l \sum_{j=1}^k \frac{(V_{ij} - m_{ij})^2}{m_{ij}} \xrightarrow{n \rightarrow \infty} \chi^2((l-1)(k-1))$$

Прийняття рішення про справедливість гіпотези аналогічне критерію однорідності.

44. Критерії значимості для параметрів нормального розподілу.

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

α (задана); будуємо $1-\alpha$ надійний інтервал $(\underline{\theta}_n; \bar{\theta}_n)$, якщо

$$\theta_0 \in (\underline{\theta}_n; \bar{\theta}_n) \Rightarrow H_0^+$$

$$\theta_0 \notin (\underline{\theta}_n; \bar{\theta}_n) \Rightarrow H_0^-$$

$$N(a, \sigma^2)$$

1) σ^2 – відоме; $H_0 : a = a_0; H_1 : a \neq a_0$

$$a_0 \in (\bar{x} - C_\alpha \frac{\sigma}{\sqrt{n}}, \bar{x} + C_\alpha \frac{\sigma}{\sqrt{n}}) \Rightarrow H_0^+$$

$$a_0 \notin (\bar{x} - C_\alpha \frac{\sigma}{\sqrt{n}}, \bar{x} + C_\alpha \frac{\sigma}{\sqrt{n}}) \Rightarrow H_0^-$$

2) σ^2 – невідоме; $H_0 : a = a_0; H_1 : a \neq a_0$

$$a_0 \in (\bar{x} - t_\alpha \frac{\hat{S}}{\sqrt{n}}, \bar{x} + t_\alpha \frac{\hat{S}}{\sqrt{n}}) \Rightarrow H_0^+$$

$$a_0 \notin (\bar{x} - t_\alpha \frac{\hat{S}}{\sqrt{n}}, \bar{x} + t_\alpha \frac{\hat{S}}{\sqrt{n}}) \Rightarrow H_0^-$$

3) a – відоме; $H_0 : \sigma^2 = \sigma_0^2; H_1 : \sigma^2 \neq \sigma_0^2$

$$\sigma_0^2 \in (\frac{nS^2}{\chi_2^2}; \frac{nS^2}{\chi_1^2}) \Rightarrow H_0^+$$

$$\sigma_0^2 \notin (\frac{nS^2}{\chi_2^2}; \frac{nS^2}{\chi_1^2}) \Rightarrow H_0^-$$

4) a – невідоме; $H_0 : \sigma^2 = \sigma_0^2; H_1 : \sigma^2 \neq \sigma_0^2$

$$\sigma_0^2 \in (\frac{(n-1)\hat{S}^2}{\chi_2^2}; \frac{(n-1)\hat{S}^2}{\chi_1^2}) \Rightarrow H_0^+$$

$$\sigma_0^2 \notin (\frac{(n-1)\hat{S}^2}{\chi_2^2}; \frac{(n-1)\hat{S}^2}{\chi_1^2}) \Rightarrow H_0^-$$

45.Перевірка гіпотези про рівність математичних сподівань двох нормальних сукупностей.

$$(\xi, \eta)$$

$$X = (x_1, x_2 \dots x_{n_1})$$

$$Y = (y_1, y_2 \dots y_{n_2})$$

$$\xi \sim N(a_1, \sigma_1^2)$$

$$\eta \sim N(a_2, \sigma_2^2)$$

а) a_1, a_2 – невідомі, σ_1^2, σ_2^2 – відомі

$$H_0 : a_1 = a_2$$

$$H_1 : a_1 \neq a_2$$

$$R^n = R_{n_0} \cup R_{n_1}$$

$$R_{n_1} \text{ (критична область)} = \{(x, y) : \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq C_\alpha\}$$

б) $a_1, a_2, \sigma_1^2, \sigma_2^2$ – невідомі

$$H_0 : a_1 = a_2$$

$$H_1 : a_1 \neq a_2$$

Припущення: $\sigma_1^2 = \sigma_2^2$

$$\hat{S}_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$$

$$\hat{S}_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2$$

$$R_{n_1} = \{(x, y) : \frac{|\bar{x} - \bar{y}|}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})(\frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2})}} \geq t_\alpha\}$$

46.Перевірка гіпотези про рівність дисперсій двох нормальних сукупностей.

в) a_1, a_2 – відомі, σ_1^2, σ_2^2 – невідомі

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$S_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - a_1)^2$$

$$S_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - a_2)^2$$

1) Якщо $S_1^2 \geq S_2^2$

$$R_{n_1} = \{(x, y) : \frac{S_1^2}{S_2^2} \geq F_\alpha\}$$

2) Якщо $S_1^2 < S_2^2$

$$R_{n_1} = \{(x, y) : \frac{S_2^2}{S_1^2} \geq F_\alpha\}$$

г) $a_1, a_2, \sigma_1^2, \sigma_2^2$ – невідомі

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Знайти \hat{S}_1^2, \hat{S}_2^2

1) Якщо $\hat{S}_1^2 \geq \hat{S}_2^2$

$$R_{n_1} = \{(x, y) : \frac{\hat{S}_1^2}{\hat{S}_2^2} \geq F_\alpha\}$$

2) Якщо $\hat{S}_1^2 < \hat{S}_2^2$

$$R_{n_1} = \{(x, y) : \frac{\hat{S}_2^2}{\hat{S}_1^2} \geq F_\alpha\}$$

47. Лінійна регресія.

$$\varphi(x) = a_0 + a_1 x$$

$$\eta = a_0 + a_1 \xi$$

$$y_i = a_0 + a_1 x_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2 \rightarrow \min a_0, a_1$$

$$\frac{\partial Q}{\partial a_0} = \sum_{i=1}^n 2(y_i - a_0 - a_1 x_i)(-1) = 0$$

$$\frac{\partial Q}{\partial a_1} = \sum_{i=1}^n 2(y_i - a_0 - a_1 x_i)(-x_i) = 0$$

$$\begin{cases} na_0 + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

$$\Delta = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = n \sum x_i^2 - (\sum x_i)^2$$

Якщо $\Delta = 0$ будувати регресію не можна

$$\hat{a}_0 = \frac{\sum y_i \sum x_i^2 - \sum x_i y_i \sum x_i}{\Delta}$$

$$\hat{a}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\Delta}$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$$

$$\bar{y} = \hat{a}_0 + \hat{a}_1 \bar{x}$$

$$Q_{\min} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\hat{\sigma}^2 = \frac{Q_{\min}}{n-2} \quad (2 - \text{кількість невідомих параметрів})$$

$$\hat{\sigma}_{a_0}^2 = \hat{\sigma}^2 \cdot \frac{\sum x_i^2}{\Delta} \quad \hat{\sigma}_{a_1}^2 = \hat{\sigma}^2 \cdot \frac{n}{\Delta}$$

$$\hat{a}_0 - t_{\alpha} \hat{\sigma}_{a_0} < a_0 < \hat{a}_0 + t_{\alpha} \hat{\sigma}_{a_0}$$

$$\hat{a}_1 - t_{\alpha} \hat{\sigma}_{a_1} < a_1 < \hat{a}_1 + t_{\alpha} \hat{\sigma}_{a_1}$$

Гіпотеза

$H_0 : a_1 = 0$ (тоді $y = a_0$, тобто ξ і η - некорельовані)

$H_1 : a_1 \neq 0$

$$t_1 (t\text{-статистика}) = \frac{\hat{a}_1}{\hat{\sigma}_{a_1}} \geq t_{\alpha} \Rightarrow H_1^+, a_1 \neq 0$$

Якщо $< t_{\alpha} \Rightarrow H_1^-, a_1 = 0$, величини некорельовані

$$\rho \quad \eta = a_0 + a_1 \xi \Leftrightarrow \rho = \pm 1$$

$$\rho = +1, a_1 > 0$$

$$\rho = -1, a_1 < 0$$

Якщо $|\hat{\rho}| \approx 1$ – тоді регресія побудована правильно

$$\hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \hat{a}_1 \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}}$$

$$\hat{\rho}^2 \geq \frac{1}{1 + \frac{n-2}{t_\alpha^2}} \Rightarrow \text{лінійна регресія вірна}$$

48. Коефіцієнт кореляції рангів Спірмена.

Припущення: Нехай всі ранги різні

$$d_i = X_i - Y_i$$

$$\hat{\rho} = 1 - \frac{\sigma \sum d_i^2}{n^3 - n} \quad (\text{коеф. Кореляції рангів Спірмена})$$

49. Коефіцієнт кореляції рангів Кандела.

Деякі ранги не можна розрізнити

$$T_x = \sum_i \frac{(t_i^3 - t_i)}{12}$$

$$T_y = \sum_i \frac{((t'_i)^3 - t'_i)}{12}$$

$$\hat{\rho} = \frac{\frac{n^3 - n}{6} - (T_x + T_y) - \sum_{i=1}^n d_i^2}{\sqrt{(\frac{n^3 - n}{6} - 2T_x)(\frac{n^3 - n}{6} - 2T_y)}}$$

Якщо додатне – позитивний зв'язок і навпаки.

Ближче до 1 – сильний, до 0 – слабкий.

50. Однофакторний дисперсійний аналіз.

Дисперсійний аналіз – статистичний метод аналізу результатів спостережень які залежать від різних одночасно діючих факторів, вибір найбільш важливих із них та оцінка їх впливу.

В залежності від кількості факторів – n-факторний дисперсійний аналіз

Однофакторний дисперсійний аналіз

$$x_{ij} = a + \alpha_i + \varepsilon_{ij}$$

α_i – вплив фактору a , $i = 1, 2, \dots, m$

i – рівень фактору a

m – число рівнів фактору a

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

j – номер спостереження

$$n_1 + n_2 + \dots + n_m = n \text{ (загальна кількість спостережень)}$$

$$H_0^A : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

$$H_1^A : \alpha_i \neq 0$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}$$

$$Q = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = Q_1 + Q_2 \text{ (тотожність дисперсійного аналізу)}$$

$$Q_1 = \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2 \text{ (відхилення по фактору } a \text{ або міжгрупове відхилення)}$$

$$Q_2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

$$S_1^2 = \frac{Q_1}{m-1} \text{ (дисперсія фактору } a)$$

$$S_2^2 = \frac{Q_2}{n-m} \text{ (залишкова дисперсія)}$$

$$S^2 = \frac{Q}{n-1} \text{ – загальна незміщенна оцінка } \sigma^2 \text{ (} \varepsilon_{ij} \sim N(0, \sigma^2) \text{)}$$

$$\text{Якщо } \frac{S_1^2}{S_2^2} \geq F_\alpha \Rightarrow H_0^{A-} \text{ – фактор } a \text{ впливає; чим більше } \frac{S_1^2}{S_2^2} \text{ – тим більший вплив}$$

$$\text{Якщо } \frac{S_1^2}{S_2^2} < F_\alpha \Rightarrow H_0^{A+} \text{ – фактор } a \text{ не впливає}$$

51. Двофакторний дисперсійний аналіз.

$$x_{ij} = a + \alpha_i + \beta_j + \varepsilon_{ij}$$

α_i – вплив фактору A , $i = 1, 2 \dots m$

i – рівень фактору A

m – число рівнів фактору A

β_j – вплив фактору B , $j = 1, 2 \dots k$

j – рівень фактору B

k – число рівнів фактору B

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\bar{x}_i = \frac{1}{k} \sum_{j=1}^k x_{ij}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^k x_{ij} \text{ (загальне середнє)}$$

$$H_0^A : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

$$H_1^A : \alpha_i \neq 0$$

$$H_0^B : \beta_1 = \beta_2 = \dots = \beta_m = 0$$

$$H_1^B : \beta_i \neq 0$$

$$Q = \sum_{i=1}^m \sum_{j=1}^k (x_{ij} - \bar{x})^2 = Q_1 + Q_2 + Q_3 \text{ (тотожність дисперсійного аналізу)}$$

$$Q_1 = \sum_{i=1}^m k (\bar{x}_{i*} - \bar{x})^2 \text{ (відхилення по фактору } A)$$

$$Q_2 = m \sum_{j=1}^k (\bar{x}_{*j} - \bar{x})^2 \text{ (відхилення по фактору } B)$$

$$Q_3 = \sum_{i=1}^m \sum_{j=1}^k (x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2$$

$$S_1^2 = \frac{Q_1}{m-1}$$

$$S_2^2 = \frac{Q_2}{k-1}$$

$$S_3^2 = \frac{Q_3}{(m-1)(k-1)}$$

$$S^2 = \frac{Q}{mk-1} - \text{незміщенна оцінка } \sigma^2$$

$$\text{Якщо } \frac{S_1^2}{S_3^2} \geq F_\alpha \Rightarrow H_0^{A-} - \text{фактор } A \text{ впливає}$$

$$\text{Якщо } \frac{S_1^2}{S_3^2} < F_\alpha \Rightarrow H_0^{A+} - \text{фактор } A \text{ не впливає}$$

$$\text{Якщо } \frac{S_2^2}{S_3^2} \geq F_\alpha \Rightarrow H_0^{B-} - \text{фактор } B \text{ впливає}$$

$$\text{Якщо } \frac{S_2^2}{S_3^2} < F_\alpha \Rightarrow H_0^{B+} - \text{фактор } B \text{ не впливає}$$

Якщо обидва фактора впливають, сильніше впливає той для кого відношення більше, наприклад якщо $\frac{S_1^2}{S_3^2} > \frac{S_2^2}{S_3^2}$, то сильніше впливає фактор A .

Аналіз часових рядів. Виділення тренду.

Часовий ряд – спостереження, які залежать від часу.

$$x_t, t = 1, 2, 3, \dots, T$$

1) Аддитивна модель часового ряду

$$x_t = g_t + S_t + \varepsilon_t$$

g_t – тренд часового ряду

S_t – сезонні коливання

ε_t – похибка спостережень

2) мультиплікативна модель часового ряду

$$x_t = g_t \cdot S_t \cdot \varepsilon_t$$

Якщо всі члени додатні – $\ln x_t = \ln g_t + \ln S_t + \ln \varepsilon_t$

Знаходження тренду:

1. Регресійний метод

$$g_t = a_0 + a_1 t + a_2 t^2 + \dots + a_m t^m$$

$$x_t = g_t + \varepsilon_t$$

$$Q = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (x_t - a_0 - a_1 t - \dots - a_m t^m)^2 \rightarrow \min_{a_0, a_1, \dots, a_m}$$

$$g_t = a_0 + a_1 t$$

$$\hat{a}_1 = \frac{n \sum x_i y_i - \sum x_i \cdot \sum y_i}{\Delta}$$

$$\Delta = n \sum x_i^2 - (\sum x_i)^2$$

$$\sum x_i^2 = \sum_{t=1}^T t^2 = \frac{T(2T+1)(T+1)}{6}$$

$$\sum x_i = \sum_{t=1}^T t = \frac{T(T+1)}{2}$$

$$T = T \frac{T(2T+1)(T+1)}{6} - \frac{T^2(T+1)^2}{4} = \frac{T^2(T+1)[4T+2-3T-3]}{12} = \frac{T^2(T+1)(T-1)}{12} =$$

$$= \frac{T^2(T^2-1)}{12}$$

$$\hat{a}_1 = \frac{12(T \sum_{t=1}^T t \cdot x_t - \frac{T(T+1)}{2} \cdot \sum x_t)}{T^2(T^2-1)} = \frac{12(\sum_{t=1}^T t \cdot x_t - \frac{(T+1)}{2} \cdot \sum x_t)}{T(T^2-1)}$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} = \frac{\sum x_t}{T} - \hat{a}_1 \cdot \frac{T+1}{2}$$

2) Методи згладжування часових рядів

2.1) Метод рухомого середнього

$$x_1, x_2, \dots, x_T$$

$$l = 2k + 1 - \text{(довільне непарне число, мінімум 3)}$$

$$\hat{x}_1 = \frac{x_1 + x_2 + \dots + x_l}{l}; \hat{x}_2 = \frac{x_2 + x_3 + \dots + x_{l+1}}{l};$$

$$\hat{x}_3 = \frac{x_3 + x_4 + \dots + x_{l+2}}{l}; \dots \hat{x}_{T-l+1} = \frac{x_{t-l+1} \dots + x_T}{l}$$

2.2) Метод експоненціального згладжування

$$0 < \alpha < 1$$

$$y_1 = x_1$$

$$y_2 = \alpha \cdot x_2 + (1 - \alpha) y_1 = \alpha \cdot x_2 + (1 - \alpha) x_1$$

$$y_3 = \alpha \cdot x_3 + (1 - \alpha) y_2 = \alpha \cdot x_3 + \alpha \cdot (1 - \alpha) x_2 + (1 - \alpha)^2 x_1$$

Якщо $\alpha \approx 1 \rightarrow$ найбільша вага біля $x_{1,2,\dots}$

Якщо $\alpha \approx 0 \rightarrow$ найбільша вага біля x_1

Рекомендується брати значення близькі до 1 (0.8;0.9) або 0.5

2.3) Метод Холта-Вінтерса

$0 < \omega < 1$ – довільні; принцип – як для α (див вище)

$0 < \nu < 1$

$$E_1 = x_1$$

$$E_2 = x_2$$

$$S_2 = E_2 - E_1$$

$$E_3 = \omega \cdot x_3 + (1 - \omega)(E_2 + S_2)$$

$$S_3 = \nu(E_3 - E_2) + (1 - \nu)S_2$$

$$E_4 = \omega \cdot x_4 + (1 - \omega)(E_3 + S_3)$$

$$S_4 = \nu(E_4 - E_3) + (1 - \nu)S_3$$

....

$$E_t$$

$$S_t$$

прогноз

$$\hat{x}_{T+k} = E_T + k \cdot S_T \text{ (не більше трьох кроків)}$$

Міри точності прогнозів

$$MSE = \frac{1}{k} \sum_{i=1}^k (x_{T+i} - \hat{x}_{T+i})^2$$

$$RMSE = \sqrt{MSE} \text{ (наприклад якщо } MSE = \text{грн}^2)$$

$$RMSPE = 100\% \cdot \sqrt{\frac{1}{k} \sum_{i=1}^k \left(\frac{x_{T+i} - \hat{x}_{T+i}}{x_{T+i}} \right)^2} \text{ (якщо } < 10\% \text{ – прогноз годиться)}$$

$$MAD = \frac{1}{k} \sum_{i=1}^k |x_{T+i} - \hat{x}_{T+i}|$$

$$MAPE = 100\% \cdot \frac{1}{k} \sum_{i=1}^k \frac{|x_{T+i} - \hat{x}_{T+i}|}{|x_{T+i}|} \text{ (<10\% – годиться)}$$

Thanks for reading

©Panda 2008