

Лекція 21.  
**Статистичне моделювання**

Статистичне моделювання — базовий метод моделювання, що полягає в тім, що модель випробовується множиною випадкових сигналів із заданою щільністю ймовірності. Метою є статистичне визначення вихідних результатів. В основі статистичного моделювання лежить *метод Монте-Карло*. Нагадаємо, що імітацію використають тоді, коли інші методи застосувати неможливо.

---

# Метод Монте-Карло

Розглянемо метод Монте-Карло на прикладі обчислення інтеграла, значення якого аналітичним способом знайти не вдається.

Задача 1. Знайти значення інтеграла:

$$y = \int_{x_1}^{x_2} f(x) dx$$

На мал. 21.1 представлений графік функції  $f(x)$ . Обчислити значення інтеграла цієї функції - виходить, знайти площу під цим графіком.

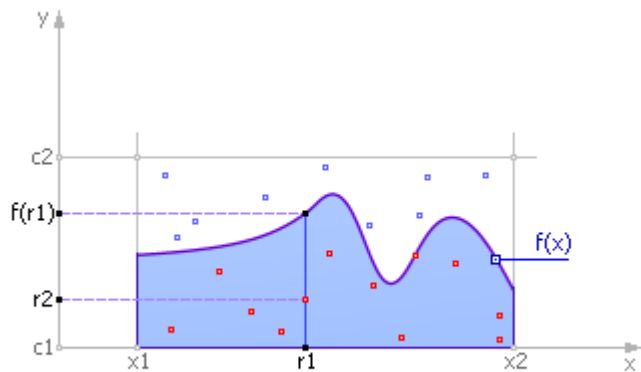


Рис. 21.1. Визначення значення інтеграла методом Монте-Карло

Обмежуємо криву зверху, праворуч і ліворуч. Випадковим образом розподіляємо крапки в прямокутнику пошуку. Позначимо через  $N_1$  кількість крапок, прийнятих для випробувань (тобто попавших у прямокутник, ці крапки зображені на мал. 21.1 червоним і синім кольорами), і через  $N_2$  — кількість крапок під кривою, тобто, що потрапили в зафарбовану площу під функцією (ці крапки зображені на мал. 21.1 червоними кольорами). Тоді природно припустити, що кількість крапок, що потрапили під криву стосовно загального числа крапок пропорційно площі під кривій (величині інтеграла) стосовно площі випробуваного прямокутника. Математично це можна виразити так:

$$\frac{N_2}{N_1} = \frac{y}{(x_2 - x_1)(c_2 - c_1)}$$

Міркування ці, звичайно, статистичні й тим більше вірні, чим більше число випробуваних крапок ми візьмемо.

Фрагмент алгоритму методу Монте-Карло у вигляді блок-схеми виглядає так, як показано на мал. 21.2.

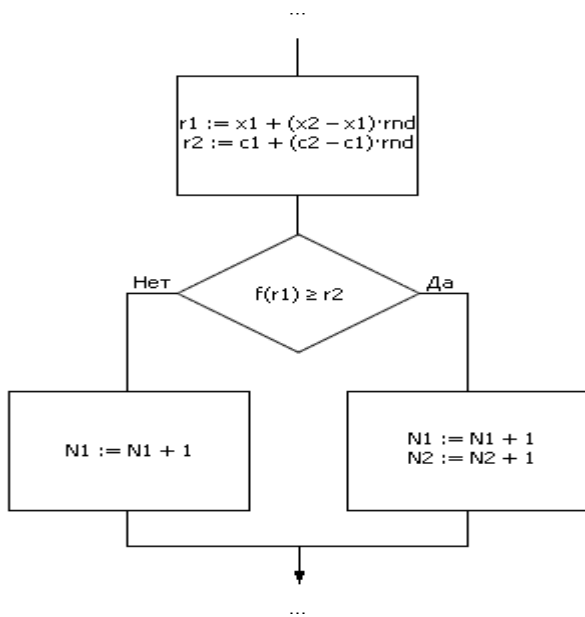


Рис. 21.2. Фрагмент алгоритму реалізації методу Монте-Карло

Значення  $r_1$  й  $r_2$  на мал. 21.2 є рівномірно розподіленими випадковими числами з інтервалів  $(x_1; x_2)$  і  $(c_1; c_2)$  відповідно.

Метод Монте-Карло надзвичайно ефективний, простий, але необхідний «гарний» генератор випадкових чисел. Друга проблема застосування методу полягає у визначенні об'єму вибірки, тобто кількості крапок, необхідних для забезпечення рішення із заданою точністю. Експерименти показують: щоб збільшити точність в 10 разів, об'єм вибірки потрібно збільшити в 100 разів; тобто точність приблизно пропорційна кореню квадратному з об'єму вибірки:

# Схема використання методу Монте-Карло при дослідженні систем з випадковими параметрами

Побудувавши модель системи з випадковими параметрами, на її вхід подають вхідні сигнали від генератора випадкових чисел (ГВЧ), як показано на мал. 21.3. ГВЧ улаштований так, що він видає *рівномірно розподілені* випадкові числа  $r_{pp}$  із інтервалу  $[0; 1]$ . Тому що одні події можуть бути більше ймовірними, інші — менш ймовірними, те рівномірно розподілені випадкові числа від генератора подають на перетворювач закону випадкових чисел (ПЗСЧ), що перетворить їх у *заданий* користувачем закон розподілу ймовірності, наприклад, у нормальний або експонентний закон. Ці перетворені випадкові числа  $x$  подають на вхід моделі. Модель відпрацьовує вхідний сигнал  $x$  за деяким законом  $y = \varphi(x)$  і одержує вихідний сигнал  $y$ , що також є випадковим.



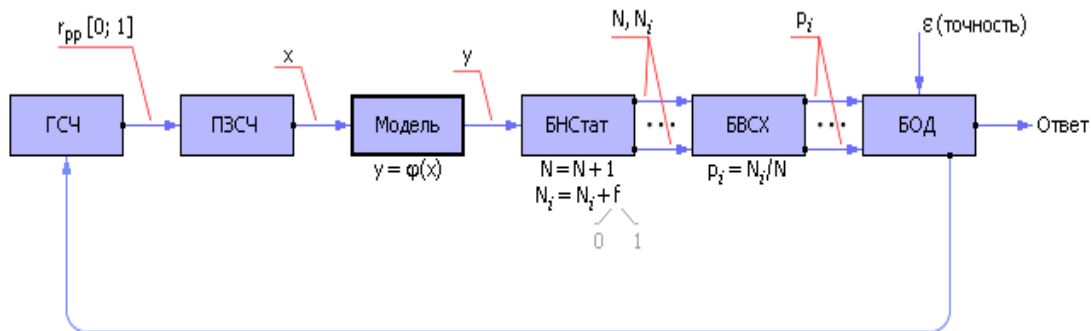


Рис. 21.3. Загальна схема методу статистичного моделювання

У блоці нагромадження статистики (БНСтат) установлені фільтри й лічильники. Фільтр (деяка логічна умова) визначає за значенням  $u$ , чи реалізувалося в конкретному досвіді деяка подія (виконалася умова,  $f = 1$ ) чи ні (умова не виконалася,  $f = 0$ ). Якщо подія реалізувалася, то лічильник події збільшується на одиницю. Якщо подія не реалізувалася, то значення лічильника не міняється. Якщо потрібно стежити за декількома різними типами подій, то для статистичного моделювання знадобиться кілька фільтрів і лічильників  $N_i$ . Завжди ведеться лічильник кількості експериментів —  $N$ .

Далі відношення  $N_i$  до  $N$ , що розраховує в блоці обчислення статистичних характеристик (БОСХ) по методу Монте-Карло, дає оцінку ймовірності  $p_i$  появи події  $i$ , тобто вказує на частоту його випадання в серії з  $N$  досвідів. Це дозволяє зробити висновки про статистичні властивості модельованого об'єкта.

Наприклад, подія  $A$  відбулося в результаті проведених 200 експериментів 50 разів. Це означає, відповідно до методу Монте-Карло, що ймовірність здійснення події дорівнює:  $p = 50/200 = 0.25$ . Імовірність того, що подія не відбудеться, дорівнює, відповідно,  $1 - 0.25 = 0.75$ .

**Зверніть увагу:** коли говорять про ймовірності, отриманої експериментально, то її називають частотою; слово ймовірність уживають, коли хочуть підкреслити, що мова йде про теоретичне поняття.

При великій кількості досвідів  $N$  частота появи події, отримана експериментальним шляхом, прагне до значення теоретичної ймовірності появи події.

У блоці оцінки вірогідності (БОВ) аналізують ступінь вірогідності статистичних експериментальних даних, знятих з моделі (беручи до уваги точність результату  $\varepsilon$ , задану користувачем) і визначають необхідне для цього кількість статистичних випробувань. Якщо коливання значень частоти появи подій щодо теоретичної ймовірності менше заданої точності, то експериментальну частоту приймають як відповідь, інакше генерацію випадкових вхідних впливів продовжують, і процес моделювання повторюється. При малому числі випробувань результат може виявитися недостовірним. Але чим більше випробувань, тим точніше відповідь, відповідно до центральної граничної теореми.

Помітимо, що оцінювання ведуть по гіршій із частот. Це забезпечує достовірний результат відразу по всім найма характеристикам, що, моделі.

Приклад 1. Вирішимо просту задачу. Яка ймовірність випадання монети орлом догори при падінні її з висоти випадковим образом?

Почнемо підкидати монетку й фіксувати результати кожного кидка (див. табл. 21.1).

Таблиця 21.1.

Результати випробувань кидання монети

Кількість досвідів $N$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Значення лічильника випадання орла $N_o$	0	0	1	1	2	3	4							
Значення лічильника випадання решки $N_p$	1	2	2	3	3	3	3							
Частота випадання орла $P_o = N_o/N$	0	0	0.33	0.25	0.4	0.5	0.57							
Частота випадання решки $P_p = N_p/N$	1	1	0.66	0.75	0.6	0.5	0.43							

Будемо підраховувати частість випадання орла як відношення кількості випадків випадання орла до загального числа спостережень. Подивитися в табл. 21.1. випадки для  $N = 1$ ,  $N = 2$ ,  $N = 3$  — спочатку значення частоті не можна назвати достовірними. Спробуємо побудувати графік залежності  $P_0$  від  $N$  — і подивимося, як міняється частість випадання орла залежно від кількості проведених досвідів. Зрозуміло, при різних експериментах будуть виходити різні таблиці й, отже, різні графіки. На мал. 21.4 показаний один з варіантів.

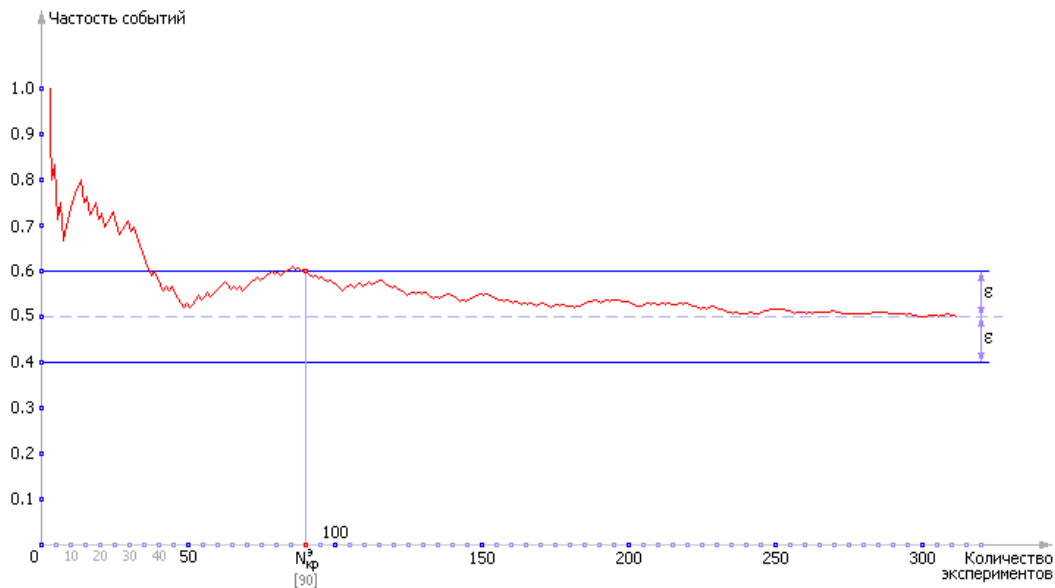


Рис. 21.4. Експериментальна залежність **частоти** появи випадкової події від кількості спостережень й її прагнення до теоретичної ймовірності

Зробимо деякі висновки.

1. Видно, що при малих значеннях  $N$ , наприклад,  $N = 1$ ,  $N = 2$ ,  $N = 3$  відповіді взагалі довіряти не можна. Наприклад,  $P_0 = 0$  при  $N = 1$ , тобто ймовірність випадання орла при одному кидку дорівнює нулю! Хоча всім добре відомо, що це не так. Тобто поки ми одержали дуже грубу відповідь. Однак, подивитися на графік: у процесі **нагромадження** інформації відповідь повільно, але вірно наближається до правильного (він виділений пунктирною лінією). На щастя, у даному конкретному випадку правильна відповідь нам відомий: в ідеалі, ймовірність випадання орла дорівнює 0.5 (в інші, більш складних задачах, відповідь нам, звичайно, буде невідомий). Допустимо, що відповідь нам треба знати з точністю  $\varepsilon = 0.1$ . Проведемо дві паралельні лінії, що відстоять від правильної відповіді 0.5 на відстань 0.1 (див. мал. 21.4). Ширина коридору, що утворився, буде дорівнює 0.2. Як тільки крива  $P_0(N)$  увійде в цей коридор так, що вже ніколи його не покине, можна зупинитися й подивитися, для якого значення  $N$  це відбулося.



Це і є **експериментально обчислене критичне значення** необхідної кількості досвідів  $N_{кр}^{\varepsilon}$  для визначення відповіді з точністю  $\varepsilon = 0.1$ ;  $\varepsilon$ -околиця в наших міркуваннях відіграє роль своєрідної трубки точності. Помітьте, що відповіді  $P_o(91)$ ,  $P_o(92)$  і так далі вже не міняють сильно своїх значень (див. мал. 21.4); принаймні, у них не змінюється перша цифра після коми, який ми зобов'язані довіряти за умовами задачі.

2. Причиною такого поведіння кривої є дія **центральної граничної теореми**. Поки тут ми сформулюємо її в найпростішому варіанті «Сума випадкових величин є величина не випадкова». Ми використали середню величину  $P_o$ , що несе в собі інформацію про суму досвідів, і тому поступово ця величина стає усе більше достовірною.
3. Якщо проробити ще раз цей досвід спочатку, те, звичайно, його результатом буде інший вид випадкової кривої. І відповідь буде іншим, хоча приблизно таким же. Проведемо целую серію таких експериментів (див. мал. 21.5). Така серія називається ансамблем реалізацій. Якій же відповіді в підсумку варто вірити? Адже вони, хоч й є близькими, все-таки відрізняються. На практиці надходять по-різному. Перший варіант — обчислити середнє значення відповідей за кілька реалізацій (див. табл. 21.2).

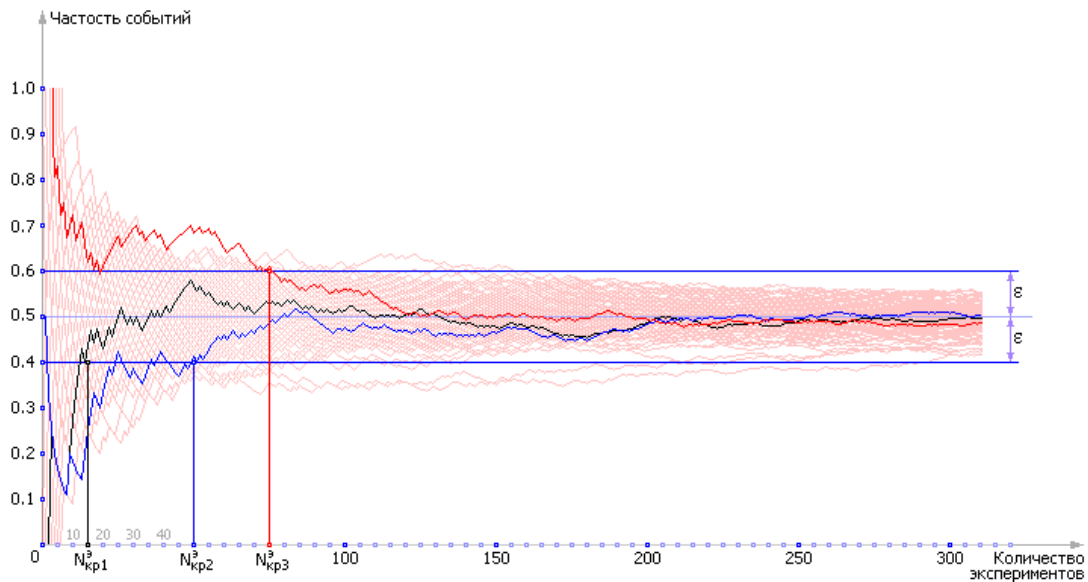


Рис. 21.5. Експериментально знятий ансамбль випадкових залежностей **частоти** появи випадкової події від кількості спостережень

Ми поставили кілька експериментів і визначали щораз, скільки необхідно було зробити досвідів, тобто  $N_{\text{кр}}^{\text{э}}$ . Було пророблено 10 експериментів, результати яких були зведені в табл. 21.2. За результатами 10-ти експериментів було обчислено середнє значення  $N_{\text{кр}}^{\text{э}}$ .

Таблиця 21.2.

Експериментальні дані необхідної кількості кидків монети для досягнення точності  $\varepsilon = 0.1$  при обчисленні ймовірності випадання орла

Досвід	$N_{кр}^{\varepsilon}$
1	288
2	95
3	50
4	29
5	113
6	210
7	30
8	42
9	39
10	48
Середнє $N_{кр.}^{\varepsilon}$	94

Таким чином, провівши 10 реалізацій різної довжини, ми визначили, що досить у **середньому** було зробити 1 реалізацію довжиною в 94 кидка монети.

Ще один важливий факт. Уважно розглянете графік на мал. 21.5. На ньому намальовано 100 реалізацій — 100 червоних ліній. Відзначте на ньому абсцису  $N = 94$  вертикальною рисою. Є якийсь відсоток червоних ліній, які не встигли перетнути  $\varepsilon$ -околиця, тобто  $(P^{\text{експ}} - \varepsilon \leq P^{\text{теор}} \leq P^{\text{експ}} + \varepsilon)$ , і ввійти в коридор точності до моменту  $N = 94$ . Зверніть увагу, таких ліній 5. Це значить, що 95 з 100, тобто 95%, ліній вірогідно ввійшли в позначений інтервал.

Таким чином, провівши 100 реалізацій, ми домоглися приблизно 95%-ного довіри до отриманого експериментально величині ймовірності випадання орла, визначивши її з точністю 0.1. Для порівняння отриманого результату обчислимо теоретичне значення  $N_{кр}^T$  теоретично. Однак для цього прийде ввести поняття довірчої ймовірності  $Q_F$ , що показує, наскільки ми готові вірити відповіді. Наприклад, при  $Q_F = 0.95$  ми готові вірити відповіді в 95% випадків з 100. Формула теоретичного розрахунку числа експериментів, що буде докладно вивчатися в лекції 34, має вигляд:  $N_{кр}^T = k(Q_F) \cdot p \cdot (1 - p) / \varepsilon^2$ , де  $k(Q_F)$  — коефіцієнт Лапласа,  $p$  — імовірність випадання орла,  $\varepsilon$  — точність (довірчий інтервал). У табл. 21.3 показані значення теоретичної величини кількості необхідних досвідів при різних  $Q_F$  (для точності  $\varepsilon = 0.1$  й імовірності  $p = 0.5$ ).

Таблиця 21.3.

Теоретичний розрахунок необхідної кількості кидків монети для досягнення точності  $\varepsilon = 0.1$  при обчисленні ймовірності випадання орла

Довірча ймовірність $Q_F$	Коефіцієнт Лапласа $k(Q_F)$	Необхідне число досвідів $N_{кр}^T = k(Q_F) \cdot p \cdot (1 - p) / \varepsilon^2$
0.90	2.72	68
0.95	3.84	96
0.99	6.66	167

Як бачите, отримана нами оцінка довжини реалізації, рівним 94 досвідам дуже близька до теоретичної, рівної 96. Деяка розбіжність порозумівається тим, що, видимо, 10 реалізацій недостатньо для точного обчислення  $N_{кр}^3$ . Якщо ви вирішите, що вам потрібний результат, якому варто довіряти більше, те змініте значення довірчої ймовірності. Наприклад, теорія говорить нам, що якщо досвідів буде 167, то всього 1-2 лінії з ансамблю не ввійдуть у запропоновану трубку точності. Але майте на увазі, кількість експериментів з ростом точності й вірогідності росте дуже швидко.

Другий варіант, використовуваний на практиці — провести **одну** реалізацію й **збільшити отримане для неї  $N_{кр}^3$  в 2 рази**. Це вважають гарною гарантією точності відповіді (див. мал. 21.6).



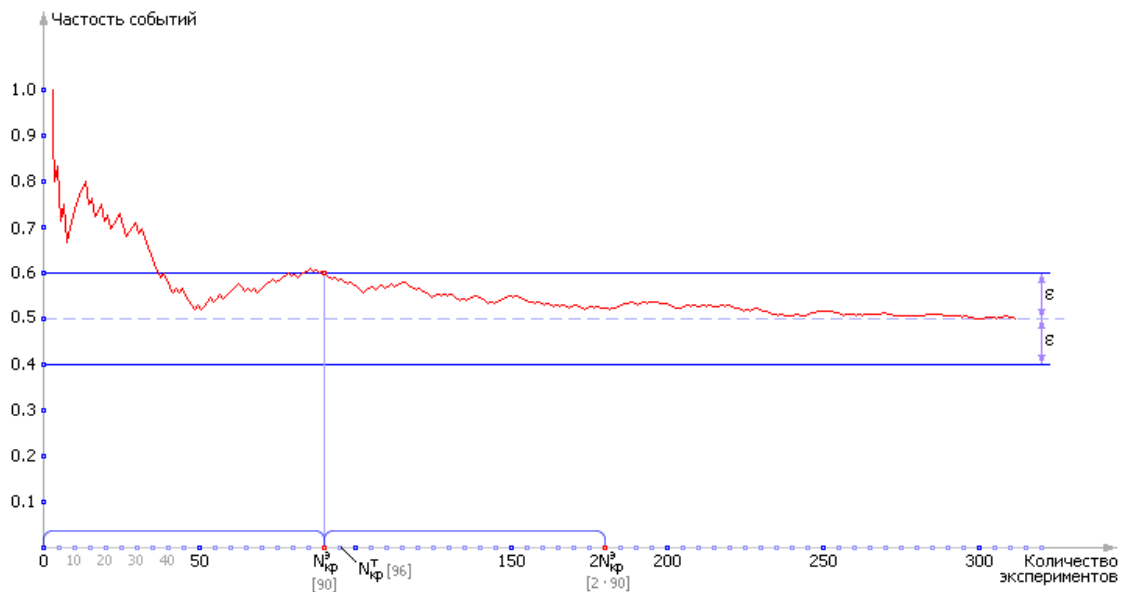


Рис. 21.6. Ілюстрація експериментального визначення  $N_{кр}^9$  за правилом  
 «ПОМНОЖ НА ДВА»

Якщо придивитися до **ансамблю випадкових реалізацій**, то можна виявити, що збіжність частоти до значення теоретичної ймовірності відбувається по кривій, що відповідає зворотної квадратичної залежності від числа експериментів (див. мал. 21.7).

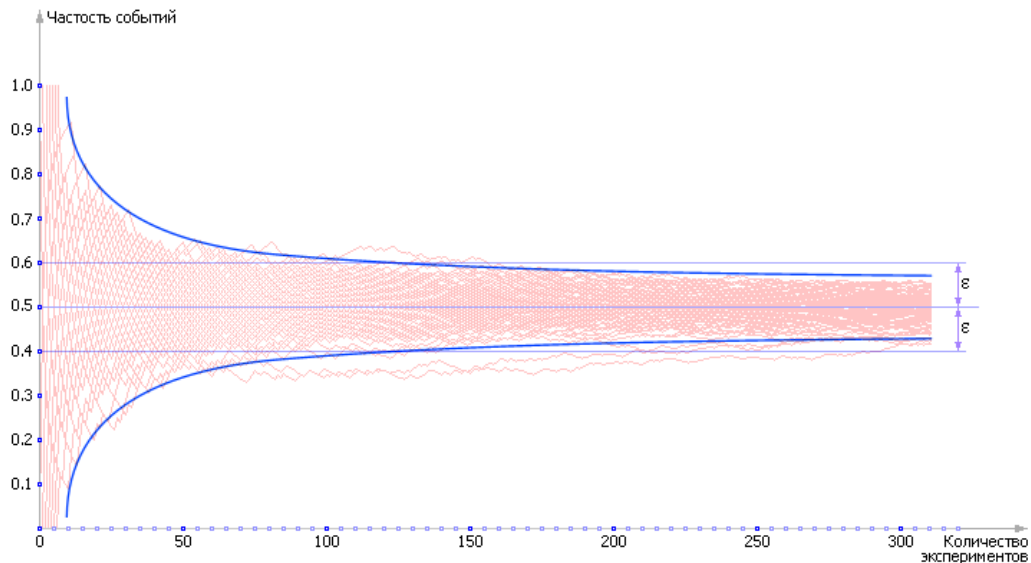


Рис. 21.7. Ілюстрація швидкості сходження експериментально одержуваної частоти до теоретичної ймовірності

Це дійсно так виходить і теоретично. Якщо змінювати задану точність  $\varepsilon$  і досліджувати кількість експериментів, необхідних для забезпечення кожної з них, то вийде табл. 21.4.

Таблиця 21.4.  
Теоретична залежність  
кількості експериментів, необхідних  
для забезпечення заданої точності при  $Q_F = 0.95$

Точність $\varepsilon$	Критичне число експериментів $N_{кр}^T$
0.1	96
0.01	9600
0.001	960000

Побудуємо по табл. 21.4 графік залежності  $N_{кр}^T(\varepsilon)$  (див. мал. 21.8).

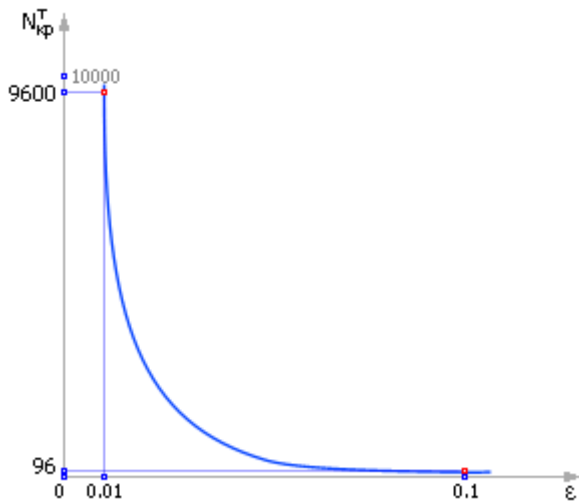


Рис. 21.8. Залежність числа експериментів, **необхідних** для досягнення заданої точності  $\epsilon$  при **фіксованому**  $Q_F = 0.95$

Отже, розглянуті графіки підтверджують наведену вище оцінку:

$$\text{точность} \cong \sqrt{\text{объем выборки}}$$

Помітимо, що оцінок точності може бути кілька. Деякі з них будуть ще обговорюватися в лекції 34.

Приклад 2. Знаходження площі фігури методом Монте-Карло. Визначите методом Монте-Карло площа п'ятикутника з координатами кутів  $(0, 0)$ ,  $(0, 10)$ ,  $(5, 20)$ ,  $(10, 10)$ ,  $(7, 0)$ .

Намалюємо у двомірних координатах заданий п'ятикутник, вписавши його в прямокутник, чия площа, як неважко догадатися, становить  $(10 - 0) \cdot (20 - 0) = 200$  (див. мал. 21.9).

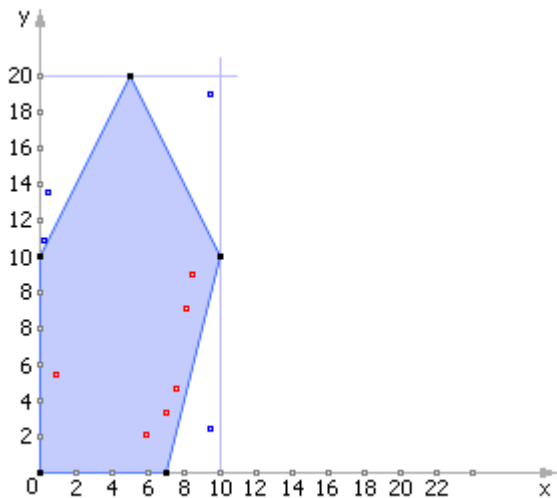


Рис. 21.9. Ілюстрація до рішення задачі про площі фігури методом Монте-Карло

Використаємо таблицю випадкових чисел для генерації пар чисел  $R$ ,  $G$ , рівномірно розподілених в інтервалі від 0 до 1. Число  $R$  буде імітувати координату  $X$  ( $0 \leq X \leq 10$ ), отже,  $X = 10 \cdot R$ . Число  $G$  буде імітувати координату  $Y$  ( $0 \leq Y \leq 20$ ), отже,  $Y = 20 \cdot G$ . Згенеруємо по 10 чисел  $R$  й  $G$  і відобразимо 10 крапок  $(X; Y)$  на мал. 21.9 й у табл. 21.5.



Таблиця 21.5.  
Рішення задачі методом Монте-Карло

<b>Номер крапки</b>	<b>R</b>	<b>G</b>	<b>X</b>	<b>Y</b>	<b>Крапка (X; Y) потрапила в прямокутник?</b>	<b>Крапка (X; Y) потрапила в п'ятикутник?</b>
1	0. 8109	0. 3557	8. 109	7. 114	Так	Так
2	0. 0333	0. 5370	0. 333	10. 740	Так	Немає
3	0. 1958	0. 2748	1. 958	5. 496	Так	Так
4	0. 6982	0. 1652	6. 982	3. 304	Так	Так
5	0. 9499	0. 1090	9. 499	2. 180	Так	Немає
6	0.	0.	7.	4. 388	Так	Так

	7644	2194	644			
7	0. 8395	0. 4510	8. 395	9. 020	Так	Так
8	0. 0415	0. 6855	0. 415	13. 710	Так	Немає
9	0. 5997	0. 1140	5. 997	2. 280	Так	Так
10	0. 9595	0. 9595	9. 595	19. 190	Так	Немає
<b>Усього:</b>					10	6

Статистична гіпотеза полягає в тім, що кількість крапок, що потрапили в контур фігури, пропорційно площі фігури:  $6:10 = S:200$ . Тобто, по формулі методу Монте-Карло, одержуємо, що площа  $S$  п'ятикутника дорівнює:  $200 \cdot 6/10 = 120$ .

Простежимо, як мінялася величина  $S$  від досвіду до досвіду (див. табл. 21.6).

Таблиця 21.6.  
Оцінка точності відповіді

<b>Кількість випробувань N</b>	<b>Оцінка ймовірності влучення випадкової крапки у випробувану область</b>	<b>Оцінка площі S методом Монте-Карло</b>
1	$1/1 = 1.00$	200
2	$1/2 = 0.50$	100
3	$2/3 = 0.67$	133
4	$3/4 = 0.75$	150
5	$3/5 = 0.60$	120
6	$4/6 = 0.67$	133
7	$5/7 = 0.71$	143
8	$5/8 = 0.63$	125
9	$6/9 = 0.67$	133
10	$6/10 = 0.60$	120

Оскільки у відповіді усе ще міняється значення другого розряду, те можлива неточність становить поки більше 10%. Точність розрахунку може бути збільшена з ростом числа випробувань (див. мал. 21.10).

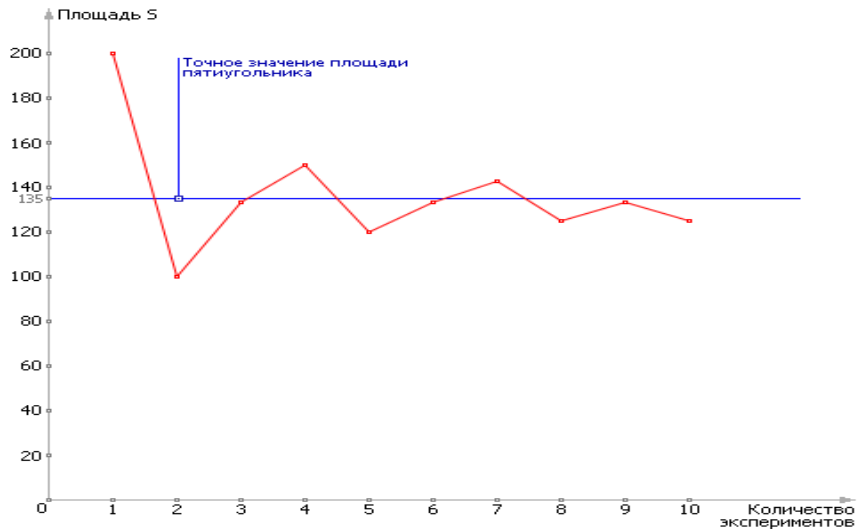


Рис. 21.10. Ілюстрація процесу збіжності обумовленого експериментально відповіді до теоретичного результату