УДК 004.02+519.213

Т. О. САВЧУК, В. В. КОЛОДНИЙ, А. В. КОЗАЧУК

Вінницький національний технічний університет, Вінниця

СИСТЕМА МОДЕЛЮВАННЯ НЕПЕРЕРВНИХ ВИПАДКОВИХ ВЕЛИЧИН З ГРАФІЧНО ЗАДАНИМИ ОСОБЛИВОСТЯМИ «GENGRAPH»

Анотація: У статті поставлено задачу розробки інструментальних засобів моделювання неперервних випадкових величин з графічно заданими особливостями. Розроблена відповідна система, що отримала назву «GenGraph». Проведене дослідження розробленої системи показало її придатність для моделювання неперервних випадкових величин з графічно заданими законами розподілу в умовах недостовірності, неповноти та відсутності статистичних даних.
Ключові слова: неперервна випадкова величина, закон розподілу, девіантні розподіли, графічно задані особливості випадкових величин, імітаційне моделювання.
Аннотация: В статье поставлена задача разработки инструментальных средств моделирования непрерывных случайных величин с графически заданными особенностями. Разработана соответствующая система, получившая название «GenGraph». Проведенное исследование разработанной системы показало ее пригодность для моделирования непрерывных случайных величин с графически заданными законами распределения в условиях недостоверности, неполноты и отсутствия статистических данных.
Abstract: The article deals with development of tools that allows to simulate continuous random variables with graphically given characteristics. System, called «GenGraph» was developed as a result of research. Investigation of developed system has shown its validity to be used for simulating continuous random variables with graphically given distribution laws under conditions of unreliability, incompleteness and lack of statistical data.

Вступ

В імітаційному моделюванні складних систем загальноприйнятим є представлення випадкових факторів з допомогою випадкових чисел, розподілених за певним законом розподілу ймовірностей з відомими параметрами. Якщо закони розподілу або значення параметрів випадкових величин на час розробки імітаційної моделі невідомі, то все одно, явно чи неявно, робиться припущення щодо виду розподілу та значень параметрів. Фахівці відзначають [1,2], що ці припущення далеко не завжди є досить обгрунтованими і тому згенеровані випадкові величини можуть неадекватно відображати реальні випадкові фактори. В цьому випадку іноді спрацьовує «магія комп'ютерного моделювання», при якій результати моделювання невиправдано вважаються достовірними, тому що дійсно генеруються певні випадкові величини згідно з розробленою моделлю, а той факт, що апріорні знання про закон розподілу випадкової величини є неповними, а прийняті функції розподілу – не завжди вірогідними, не береться до уваги.

В цій статті розглядаються інструментальні засоби моделювання випадкових величин, щодо яких точно невідомі закони розподілу ймовірностей та їхні параметри, але є певна апріорна інформація, яка може висловлюватися в нестрогій, вербальній, числовій або графічній формі. Мається на меті максимально можливе врахування всієї апріорної інформації в таких випадках та зниження впливу суб'єктивізму на адекватність імітаційних моделей, що розробляються.

Мета дослідження

Метою дослідження є розробка інструментальних засобів моделювання неперервних випадкових величин з графічно заданими особливостями для максимальної автоматизації процесів побудови і редагування ескізів графіків щільності розподілу та подальшої генерації нестандартних випадкових величин.

Постановка задач

1. Розробити математичні моделі та алгоритми для роботи з графічно заданими законами розподілу.
2. Програмно реалізувати отримані математичні моделі та алгоритми у вигляді системи моделювання неперервних випадкових величин з графічно заданими особливостями.
3. Провести дослідження отриманої системи моделювання неперервних випадкових величин з графічно заданими особливостями розподілу.

Передумови для розробки системи GenGraph

Проведені дослідження говорять, що моделювання випадкових величин з графічно заданими особливостями доцільне, якщо закони розподілу цих величин є нестандартними, наприклад, девіантними [3,4] або багатомодальними.

Девіантний розподіл – це близький до відомого типового (базового) закону розподіл ймовірностей, який має певні визначені суттєві особливості (відмінності від базового закону), що називаються девіантностями.

Девіантні розподіли, як правило, не виражаються аналітичними формулами для щільності ймовірностей, а генеруються за допомогою комп'ютера методами імітаційного моделювання, тобто будьякий девіантний розподіл – це емпіричний розподіл, який отримано певним алгоритмічним перетворенням базових випадкових чисел $R \in (0;1)$.

Відомі такі види опису девіантностей [3]:

1) числовий опис;
2) графічний опис;
3) вербальний опис.

Також можливі комбінації цих видів: числового, графічного та вербального, числового та вербального, числового та графічного, графічного та вербального. Таким чином, можна виділити 7 різних видів описів девіантностей.

Виявилося, що цю класифікацію доцільно застосовувати не тільки для девіантних законів, а взагалі для будь-яких нестандартних законів розподілу. В цьому більш загальному випадку замість терміна «девіантність» будемо використовувати термін «особливість». Наведемо приклади числового опису особливостей розподілу:

1) площа під графіком щільності розподілу, що обмежена інтервалом (a;b);
2) площа під графіком від точки x_1 до точки X_{ma};
3) коефіцієнт асиметрії;
4) точки зрізу x_a, x_b;
5) інші числові параметри розподілу.

Прикладами вербального опису особливостей можуть бути:

1) розтягнутість хвостів (лівого, правого або обох);
2) невелика асиметричність (ліва, права);
3) зрізаність (зліва, справа або з обох боків);
4) підвищена ймовірність виникнення окремих рідких подій (дуже великих або дуже малих значень випадкової величини);
5) кусочна дискретність, та ін.

Можливі також комбінації різних особливостей.

В якості графічного опису особливостей розподілу найбільш наочним і зручним є опис графіком щільності розподілу або гістограмою [1,2]. На жаль, наявної інформації не завжди достатньо для побудови точних графіків, але виявилось можливим в багатьох випадках поєднувати числові, вербальні та графічні описи особливостей і будувати ескізи графіку щільності розподілу.

Зрозуміло, що це досить трудомісткий процес, тому була поставлена задача максимально автоматизувати процес побудови і редагування ескізів графіків щільності розподілу та подальшої генерації випадкових величин згідно з вербально, чисельно та графічно заданими особливостями розподілів. Для розв'язання цієї задачі і була розроблена комп'ютерна система **GenGraph**.

Розробка алгоритму генерації випадкової величини з графічно заданим законом розподілу

Розіб'ємо процес розробки алгоритму генерації випадкової величини з графічно заданим законом розподілу на два послідовні етапи: виділення графіку щільності розподілу заданої випадкової величини, побудова функції генерації випадкових величин, розподілених за заданим законом розподілу. Розглянемо послідовно ці етапи.

Виділення графіку функції щільності розподілу випадкової величини X з множини графічних примітивів, введених користувачем.

Необхідно, маючи множину графічних примітивів Q, введених користувачем, розмір та координати границь робочої області, на яку вводяться примітиви, отримати вектор *H*, що представляє дискретизовану функцію щільності девіантного розподілу випадкової величини *X*.

При роботі з немасштабованими зображеннями неможливо подолати обмеження дискретизації, що накладається роздільною здатністю монітора комп'ютера. Тому за одиницю дискретизації прийнято розмір пікселя.

Кожен елемент q множини Q являє собою ламану лінію, задану у системі координат, пов'язаних із робочою областю. Початок координат знаходиться у правому верхньому куті робочої області, вісь абсцис направлена вліво, вісь ординат – зверху вниз. Одиниця системи координат співпадає з розміром пікселя.

Для досягнення максимальної простоти системи моделювання неперервних випадкових величин з графічно заданими законами розподілу введено додаткову систему координат, яку використовує користувач. Ця система координат має такі характеристики: вість абсцис направлена зліва направо і співпадає з нижньою гранню робочої області. Вісь ординат направлена знизу вгору. Задаються координати лівої та правої граней робочої області по вісі абсцис. Одиниця по осі ординат співпадає з розміром пікселя.

Під час роботи система моделювання неперервних випадкових величин з графічно заданими законами розподілу повинна виконувати перетворення координат із системи, з якою працює користувач в робочу і навпаки.

Розглянемо алгоритм, що використовується для отримання *H* із інформації, представленої у Q, позначивши ширину робочої області у пікселях за L.

1. Створити вектори *HasValue*, *H* розміром L.
2. $\forall i, 0 \le i < L : HasValue[i] \leftarrow false; H[i] \leftarrow 0;$
3. $\forall q \in Q$:
   Для усіх сусідніх точок ламаної *p1*, *p2*:
   якщо *p1.x > p2.x*
   то
   поміняти *p1*та *p2* місцями
   якщо *p1.x = p2.x*
   то
   точки лежать на вертикальній прямій
   $H[p1.x] \leftarrow \max(p1.y, p2.y, H[p1.x])$
   $y \leftarrow p1.y$
   для *cx* від *p1.x* до *p2.x*, крок 1:
   $HasValue \leftarrow true$
   $H[cx] \leftarrow \max(H[cx], y)$
   $y \leftarrow y + (p2.y - p1.y)/(p2.x - p1.x)$
4. для $\forall i, 0 \le i < L$:
   конвертувати значення *H [i]* з урахуванням *HasValues[i]* та висоти робочої області
5. Вивести *H.*

Таким чином, отримано вектор *H*, необхідний для подальшої роботи з графічно заданим законом розподілу.

Побудова функції генерації випадкових величин на основі графіку функції щільності розподілу.

Розглянемо алгоритм, що дозволяє згенерувати вибірку випадкових величин на основі таблично заданої функції щільності розподілу.

Необхідно маючи вектор дискретизованої функції щільності розподілу *H*, координати лівої та правої граней робочої області *minValue* та *maxValue*, розмір результуючого вектора *n*, отримати функцію *CustomRand(H, minValue, maxValue, n)*, що дозволяє здійснити генерацію випадкової величини *X*.

функція *CustomRand(H, minValue, maxValue, n)*

1. $V \leftarrow GetValues(H, \min Value, \max Value, n)$
2. $index \leftarrow round(R * (n-1))$
3. вивести $\mathbf{V}[index]$.

де *round(x)* – функція, що округлює число *x* до найближчого цілого значення.

функція *GetValues(H, minValue, maxValue, n)*

1. $sum \leftarrow 0$
   $\forall h, h \in H$:
   $sum \leftarrow sum + h$
2. $curIndex \leftarrow 0$
   $growth \leftarrow (\max Value - \min Value)/ H.size$
   $curValue \leftarrow \min Value$
   ініцаціалізувати результуючий вектор **V** розміром *n*
3. $\forall h, h \in H$:
   $numberOf \leftarrow h$
   якщо *n≠sum*
   то
   $numberOf \leftarrow round(h * n / sum)$
   для і від *curIndex* до *curIndex + numberOf:*

$V[i] \leftarrow curValue$

$curIndex \leftarrow curIndex + numberOf$

$curValue \leftarrow curValue + growth$

4. вивести V.

Отже, генерація випадкових величин, розподілених за законами розподілу з графічно заданими особливостями, може бути виконана за допомогою вищенаведених функцій.

Інструментальні засоби системи GenGraph

На основі вищенаведених алгоритмів розроблено систему моделювання випадкових величин з графічно заданими законами розподілу. Система отримала назву **GenGraph**. Створений програмний продукт включає в себе такі інструментальні засоби:

- введення графіка щільності розподілу за допомогою миші;
- завдання ділянки лінійності по координатам крайніх точок;
- проведення відрізків мишою;
- завдання координат крайніх точок робочої області;
- побудова графіку щільності одного із стандартних законів розподілу;
- апроксимація виділеної ділянки графіка методом найменших квадратів, або методом гнучкої нитки;
- визначення ймовірності потрапляння випадкової величини у певний інтервал;
- побудова гістограми отриманого розподілу;
- визначення статистичних параметрів отриманого розподілу;
- збереження отриманого розподілу із можливістю подальшого застосування у .Net додатках.

Система є простою у використанні і має зручний покроковий інтерфейс користувача з можливістю повернення до будь-якого попереднього кроку. Інтерфейс системи (рис.1) доступний українською та англійською мовами.

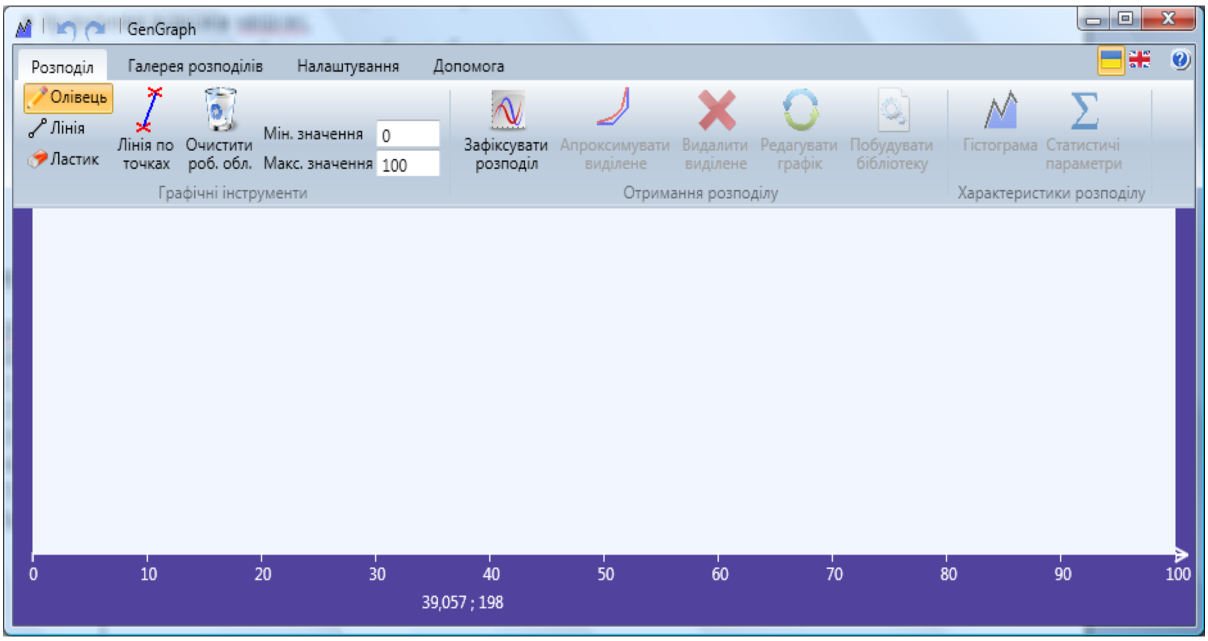

Рисунок 1 – Інтерфейс системи **GenGraph**

Приклад практичного застосування системи GenGraph

Розглянемо приклад практичного застосування системи. Припустимо, що статистичних даних немає, але експертним шляхом були визначені такі особливості неперервної випадкової величини X:

- X>0;
- найбільш імовірними є значення в діапазоні $20 \le x \le 70$;
- значення $x < 20$ є більш імовірними, ніж значення $x > 70$;
- значення $x > 100$ практично не зустрічаються;
- значення $20 \le x \le 30$ є трохи менш імовірними, ніж значення $60 \le x \le 70$.

- Значення x<10 є приблизно рівноймовірними.

Можливий вигляд ескізу графіка функції щільності розподілу неперервної випадкової величини X наведено на рисунку 2.

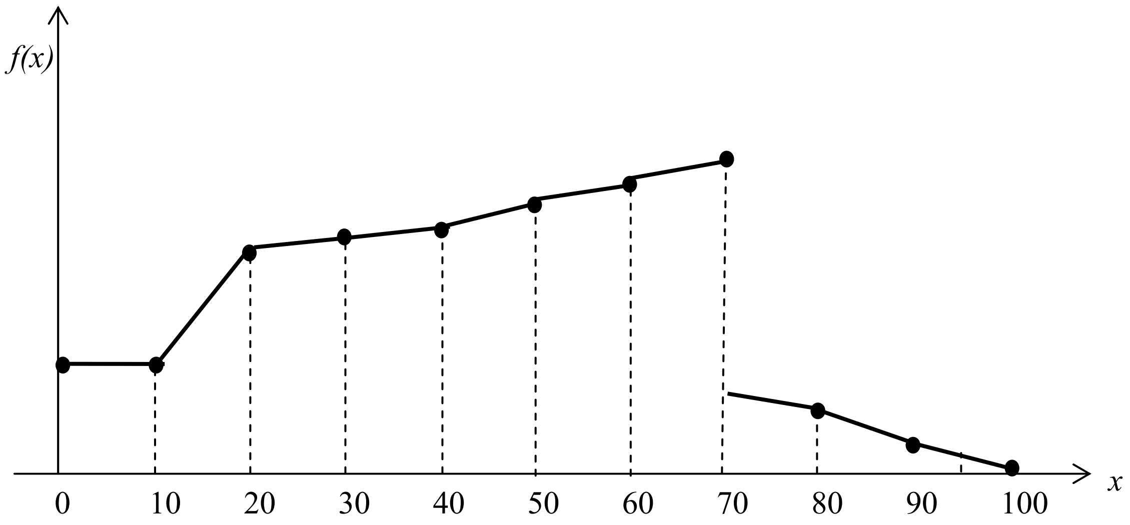

Рисунок 2 – Ескіз графіка функції щільності описаного закону розподілу

Згенеруємо випадкову величину, ескіз графіка щільності якої представлений на рисунку 2. Результат моделювання зображений на рисунку 3.

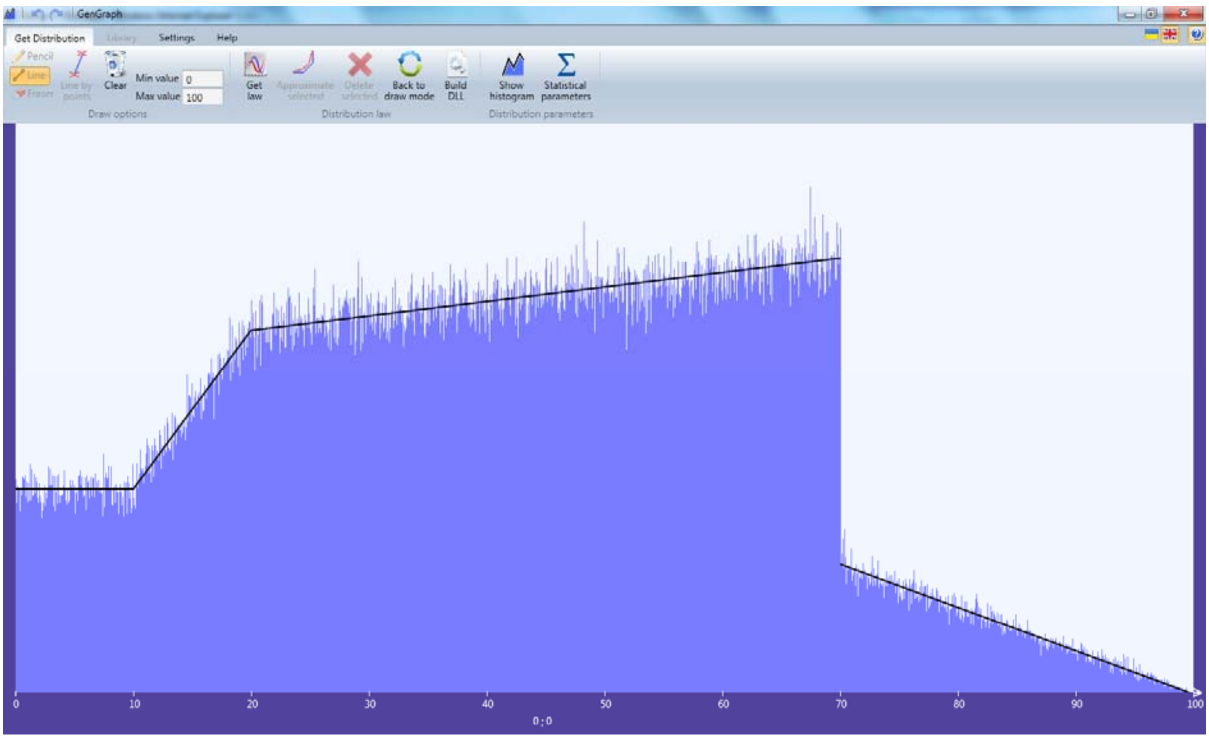

Рисунок 3 – Результат моделювання

Порівняння статистичних параметрів заданого розподілу та вибірки згенерованих величин (гістограма вибірки представлена на рис. 3) наведено на рисунку 4.

statistics

Name of parameter	Drawn value	Generated value
Expected value, μ Мат. сподівання	41.906179675245	41.855097349234977
Variance, σ2 Дисперсія	452.231146882600630	451.754357177037262
Standard deviation, σ С. К. В.	21,2657270480602	21,2545138071196
Skewness, Y1 Коефіцієнт асиметрії	0.012221884376756	0.017147033039889
Kurtosis, Y2 Коефіцієнт ексцесу	2.21819437455848	2.21937556977279
Min value Мінімальне значення	0.035460	0.035460
Max value Максимальне значення	99.680851	99.397163

Рисунок 4 – Порівняння статистичних параметрів згенерованої вибірки та заданої випадкової величини

Припустимо, що експерт додатково відзначає, що графік щільності розподілу даної випадкової величини має бути гладкою кривою із збереженням усіх вищезазначених особливостей. Для, того, щоб привести згенеровану величину до такого вигляду, проведемо апроксимацію заданого графіку та повторно визначимо статистичні параметри. При цьому, імовірність потрапляння випадкової величини у окіл точки 70 не зміниться, завдяки особливості алгоритму апроксимації, що використовується в GenGraph.

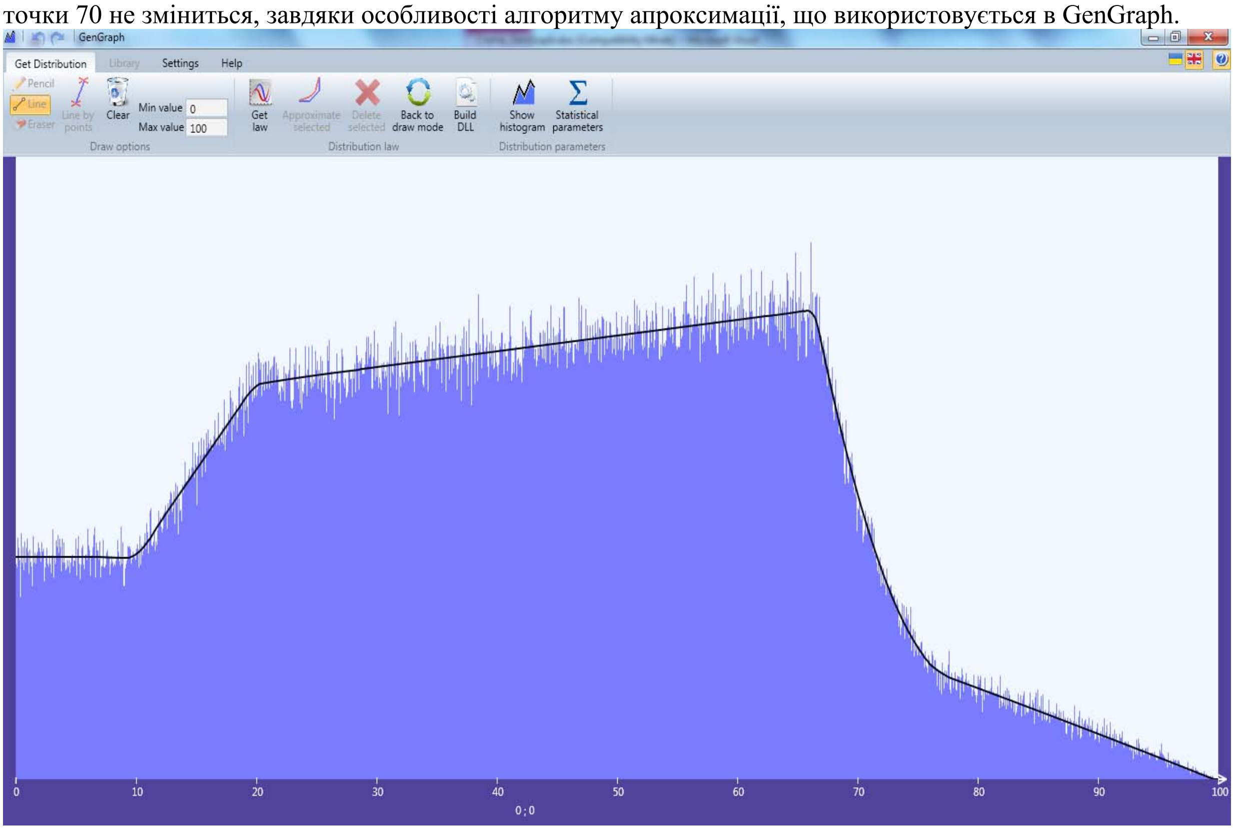

Рисунок 5 – Результат апроксимації випадкової величини

На рисунку 6 показані статистичні параметри випадкової величини з апроксимованим графіком щільності розподілу.

statistics

Name of parameter	Drawn value	Generated value
Expected value, μ Мат. сподівання	41.921037453064	41.930483160380058
Variance, σ2 Дисперсія	453.649692469600623	454.379378161017888
Standard deviation, σ С. К. В.	21,2990537928238	21,3161764432794
Skewness, Y1 Коефіцієнт асиметрії	0.015819432686929	0.017885142321292
Kurtosis, Y2 Коефіцієнт ексцесу	2.21554378691759	2.21738864005578
Min value Мінімальне значення	0.035460	0.035460
Max value Максимальне значення	99.680851	99.539007

Рисунок 6 – Порівняння статистичних параметрів згенерованої вибірки та апроксимованої випадкової величини

Порівняння даних з рисунків 6 та 4 показує, що згенерована випадкова величина не чутлива до розривів графіку щільності в околі точки 70.

Висновки

У статті розглянуто особливості девіантних законів розподілу та способи їх опису. На основі отриманого алгоритму генерації випадкових величин було розроблено систему моделювання неперервних випадкових величин з графічно заданими особливостями **GenGraph**. Розроблена система дає можливість проводити моделювання випадкових величин з нестандартними законами розподілу, а також випадкових величин, розподіли яких близькі до стандартних, але мають певні особливості (девіантності). Також з’являється можливість оцінити чутливість виходу моделі до незначних змін графіка функції щільності розподілу. За допомогою системи **GenGraph** можна моделювати випадкові величини, задані малими вибірками або взагалі без наявної статистичної інформації. Проведене дослідження розробленої системи показало, що її використання є досить ефективним для імітаційного моделювання складних систем.

Список використаної літератури

1. Кельтон В. Имитационное моделирование. Классика CS / В. Кельтон, А. Лоу – СПб: Питер; Киев: Издательская группа BHV, 2004. – 847 с.
2. Мостеллер Ф. Анализ данных и регрессия / Ф. Мостеллер, Дж. Тьюки – Вып. 1. - М: Финансы и статистика, 1982. – 317 с.
3. Колодний В. В. Застосування класу девіантно-нормальних розподілів в імітаційному моделюванні// Вісник ВПІ. – 2001. – №2. – С. 82-85.
4. Колодний В. В. Методи генерації девіантно-нормальних розподілів ймовірностей/ В. В. Колодний, В. В. Седлецький // Вісник ВПІ, – 2004. – № 5. – С. 72-77.

Відомості про авторів

Савчук Тамара Олександрівна – к.т.н., професор кафедри Комп'ютерних наук, Вінницький національний технічний університет, Вінниця, Україна.

Колодний Володимир Володимирович – к.т.н., доцент кафедри Комп'ютерних наук, Вінницький національний технічний університет, Вінниця, Україна.

Козачук Андрій Валерійович – магістрант кафедри Комп'ютерних наук, Вінницький національний технічний університет, Вінниця, Україна.