УДК 519.2

П.О. Єндовицький
Національний технічний університет України “КПІ”, Київ, Україна

ДВІ МОДИФІКАЦІЇ ЗАДАЧІ ПРО ДНІ НАРОДЖЕННЯ

Background. Scheme of particle allocation in cells is studied in probability theory as well as in mathematical statistics. In probability theory it goes about limit theorems, in mathematical statistics – of construction statistical criteria's. Birthday problem is one of main questions in this theory.
Objective. In the paper two modifications of the birthday problem are considered. One was formulated in Fermi statistic scheme, another – in uniform and independent random allocation scheme. In both cases the objective was to solve birthday problem.
Methods. Standard asymptotical methods were used. At first we needed to prove one limit theorem and to estimate rapidity of convergence in it. Using these results numerical calculation of probabilities from birthday problem was made. Also formulas for the group size from birthday problem were obtained.
Results. As a result numerical estimates for birthday problem probability and group size were obtained.
Conclusions. For both modifications asymptotic main value coincides both in the formula for probability calculation and the formula for the group size. But second terms from their asymptotic series are already different.
Keywords: birthday problem; birthday paradox; random allocations; Fermi statistic; Uval attack.

Вступ

Задача про дні народження є класичною в теорії ймовірностей [1]. Відповідь у цій задачі є (на перший погляд) несподіваною, що дає привід для вживання терміну “парадокс” днів народжень. “Парадоксальним” є явище, яке полягає у високій імовірності збігу днів народжень у деякої пари осіб з певної, невеликої за розміром групи осіб.

Цікаво, що це явище малості розміру групи, який необхідний для досягнення певної ймовірності існування пари народжених в один день осіб, має застосування у криптографії: в алгоритмі [2] підробки цифрового підпису. Трудомісткість цього алгоритму визначатиметься саме розміром групи із задачі про дні народження і буде значно меншою за трудомісткість повного перебору. Платою за зменшення трудомісткості є імовірнісний характер алгоритму, і ймовірність успішної підробки цифрового підпису буде визначатися саме ймовірністю існування пари народжених в один день із задачі про дні народження.

Тобто математична модель криптографічної атаки [2] на цифровий підпис зводиться до імовірнісної задачі про дні народження. Але ця математична задача, що виникає в [2], дещо відрізняється від класичної, яка описана в [1]. У статті мова піде про дві математичні моделі атаки [2], які є певними модифікаціями класичної задачі про дні народження.

Актуальним є питання, чи будуть ці різні математичні моделі однієї криптографічної атаки давати однакову відповідь щодо ймовірності успіху атаки та її трудомісткості. Це питання є цікавим і з суто математичної точки зору.

Історичний огляд питання, сучасний стан та перспективи досліджень у задачі про дні народження можна знайти, наприклад, у [3, 4].

Постановка задачі

Метою роботи є розв’язання задачі про дні народження для двох модифікацій класичного випадку цієї задачі.

Пряма та обернена задачі про дні народження

У статі розглядатимуться дві ймовірнісні схеми, які ми будемо називати статистикою Фермі та статистикою Ювала. Обидві схеми формулюються у термінах розміщення частинок по комірках.

Статистика Фермі. Нехай у m комірках незалежно розміщуються два комплекти частинок, по n частинок у кожному комплекті. Частинки кожного комплекту розміщуються у комірках по одній, при цьому всі можливі C_m^n розміщень вважаються рівноймовірними. Позначимо $F_m(n)$ ймовірність того, що існує комірка, яка містить частинки з обох комплектів.

Статистика Ювала. Нехай у m комірках розміщуються незалежно і рівноймовірно частинки двох типів, по n частинок кожного типу. Позначимо $U_m(n)$ ймовірність того, що існує комірка, яка містить частинки обох типів.

На відміну від статистики Фермі, у статистиці Ювала частинки одного типу можуть потрапити в одну комірку.

Для обох статистик будемо розглядати два питання, які можна називати відповідно прямою та оберненою задачами про дні народження.

Задача А. Нехай відомі параметри m та n – кількість комірок та кількість частинок в одному комплекті, і потрібно підрахувати ймовірності $F_m(n)$ та $U_m(n)$.

Задача Б. Нехай заданий параметр m – кількість комірок та число $p \in (0,1)$, і потрібно знайти кількість частинок $n_F(m)$ та $n_U(m)$, які задовольняють рівності:

$$n_F(m) = \min\{n \geq 1: \quad F_m(n) \geq p\},$$
$$n_U(m) = \min\{n \geq 1: \quad U_m(n) \geq p\}.$$

Певну відповідь на ці питання можна отримати з такої теореми.

Теорема. Якщо у статистиках Фермі та Ювала виконується співвідношення $\frac{n^2}{m} \to \lambda > 0$, $n, m \to \infty$, то

$$\lim_{n,m\to\infty} F_m(n) = \lim_{n,m\to\infty} U_m(n) = 1 - e^{-\lambda}.$$

Доведення. Позначимо $G_m(n) := 1 - F_m(n)$, $V_m(n) := 1 - U_m(n)$ – ймовірності того, що комірки, які містять частинки обох типів, відсутні.

Розглянемо спочатку граничну поведінку ймовірності $G(n)$ (іноді, для спрощення запису, будемо писати $G(n)$ замість $G_m(n)$). Маємо

$$G(n) = \frac{C_m^n C_{m-n}^n}{C_m^n C_m^n} =$$
$$= \frac{(m-n)(m-n-1)\cdot\ldots\cdot(m-2n+1)}{m(m-1)\cdot\ldots\cdot(m-n+1)} =$$
$$= \prod_{k=0}^{n-1}\left(1 - \frac{n}{m-k}\right)$$

і звідси

$$\left(1 - \frac{n}{m-n+1}\right)^n \leq \prod_{k=0}^{n-1}\left(1 - \frac{n}{m-k}\right) \leq \left(1 - \frac{n}{m}\right)^n.$$

Але, якщо $\frac{n^2}{m} \to \lambda$, $n, m \to \infty$, то

$$\lim_{n,m\to\infty}\left(1 - \frac{n}{m-n+1}\right)^n = \lim_{n,m\to\infty}\left(1 - \frac{n}{m}\right)^n = e^{-\lambda}.$$

Отже,

$$\lim_{n,m\to\infty} G(n) = \lim_{n,m\to\infty}\prod_{k=0}^{n-1}\left(1 - \frac{n}{m-k}\right) = e^{-\lambda}.$$

Рівність $\lim\limits_{n,m\to\infty} F(n) = 1 - e^{-\lambda}$ доведено.

Далі знайдемо граничну поведінку ймовірності $V(n)$. Для цього спочатку виведемо формулу для $V(n)$.

Позначимо μ_0 випадкову величину, яка дорівнює кількості порожніх комірок, що залишилися після розміщення n частинок першого типу. Тоді з формули повної ймовірності випливає, що

$$V(n) = \sum_{k=0}^{n-1} P(\mu_0 = m-n+k)\left(\frac{m-n+k}{m}\right)^n =$$
$$= \sum_{k=0}^{n-1} P(\mu_0 = m-n+k)\left(1 - \frac{n-k}{m}\right)^n,$$

де $\left(\frac{m-n+k}{m}\right)^n$ – це ймовірність того, що всі частинки другого типу розмістяться у порожні комірки, які залишилися після розміщення частинок першого типу.

Якщо позначити $\xi_n = \mu_0 - m + n$, то ймовірність $V(n)$ виразиться через математичне сподівання:

$$V(n) = M\left(1 - \frac{n-\xi_n}{m}\right)^n.$$

Але відомо [3], що при $\frac{n^2}{m} \to \lambda$, $n, m \to \infty$, випадкова величина ξ_n слабко збігається до пуассонівського розподілу з параметром λ. Звідси матимемо збіжність за ймовірністю:

$$\frac{n-\xi_n}{m}\cdot n = \frac{n^2}{m} - \frac{n}{m}\cdot\xi_n \xrightarrow{P} \lambda, \quad n, m \to \infty,$$

і, отже,

$$\left(1 - \frac{n-\xi_n}{m}\right)^n \xrightarrow{P} e^{-\lambda}, \quad n, m \to \infty.$$

Також справджується нерівність $\left(1-\frac{n-\xi_n}{m}\right)^n \le 1$, тобто з теореми Лебега про обмежену збіжність буде випливати, що

$$V(n) = M\left(1-\frac{n-\xi_n}{m}\right)^n \to e^{-\lambda},\ n,m\to\infty.$$

Рівність $\lim\limits_{n,m\to\infty} U(n) = 1-e^{-\lambda}$ доведено. Теорему доведено.

З теореми випливають наближені рівності для ймовірностей $F(n)$ та $U(n)$ (які, звичайно, не можна назвати розв'язком прямої задачі А через відсутність оцінки похибки):

$$F(n) \approx U(n) \approx 1-\exp\left(-\frac{n^2}{m}\right),$$

а також асимптотичні рівності для величин $n_F(m)$ та $n_U(m)$ із оберненої задачі Б:

$$n_F(m) = \sqrt{a\,m} + o(\sqrt{m}),\ m\to\infty, \qquad (1)$$

$$n_U(m) = \sqrt{a\,m} + o(\sqrt{m}),\ m\to\infty, \qquad (2)$$

де $a = -\ln(1-p)$.

Далі, оцінивши швидкість збіжності у теоремі, надамо уточнення співвідношень (1) і (2). При цьому надалі вважаємо, що виконується співвідношення $\frac{n^2}{m} = O(1),\ n,m\to\infty$.

Це співвідношення, яке береться з теореми, є важливим, бо воно виражає зв'язок між кількістю комірок m та кількістю частинок n у задачі про дні народження. Воно дає розуміння "парадоксу" днів народжень: чому кількість частинок є незначною порівняно з кількістю комірок (наприклад, $m = 365$ та $n = 23$ (див. [1])).

Статистика Фермі

Почнемо з розв'язання задач А і Б для статистики Фермі. Маємо

$$G(n) = \prod_{k=0}^{n-1}\left(1-\frac{n}{m-k}\right),$$

$$\ln G(n) = \sum_{k=0}^{n-1}\ln\left(1-\frac{n}{m-k}\right).$$

Для розв'язання задачі А, тобто знаходження числового значення ймовірності $G(n)$, спочатку отримаємо оцінку для доданка $\ln\left(1-\frac{n}{m-k}\right)$, $k=\overline{0,n-1}$, з допомогою нерівності

$$-x-\frac{x^2}{2(1-x)} \le \ln(1-x) \le -x-\frac{x^2}{2},\ x\in(0,1). \qquad (3)$$

Одержимо таку подвійну нерівність:

$$-\frac{n}{m-k}-\frac{n^2}{2(m-k)^2} \ge \ln\left(1-\frac{n}{m-k}\right) \ge$$

$$\ge -\frac{n}{m-k}-\frac{n^2}{2(m-k)^2\left(1-\frac{n}{m-k}\right)} \ge$$

$$\ge -\frac{n}{m-k}-\frac{n^2}{2(m-k)^2\left(1-\frac{n}{m-n+1}\right)} =$$

$$= -\frac{n}{m-k}-\frac{n^2}{2(m-k)^2}\cdot\frac{m-n+1}{m-2n+1}.$$

Звідси маємо і оцінку для $\ln G(n)$:

$$-\sum_{k=0}^{n-1}\frac{n}{m-k}-\frac{m-n+1}{m-2n+1}\cdot\sum_{k=0}^{n-1}\frac{n^2}{2(m-k)^2} =$$

$$= -\sum_{k=0}^{n-1}\left(\frac{n}{m-k}+\frac{n^2}{2(m-k)^2}\right)-\frac{n}{m-2n+1}\times$$

$$\times\sum_{k=0}^{n-1}\frac{n^2}{2(m-k)^2} \le \sum_{k=0}^{n-1}\ln\left(1-\frac{n}{m-k}\right) \le$$

$$\le -\sum_{k=0}^{n-1}\left(\frac{n}{m-k}+\frac{n^2}{2(m-k)^2}\right).$$

Різниця між правою та лівою частинами в цій нерівності дорівнює

$$\frac{n}{m-2n+1}\cdot\sum_{k=0}^{n-1}\frac{n^2}{2(m-k)^2} \le$$

$$\le \frac{n^3}{2(m-2n+1)}\sum_{k=0}^{n-1}\int_{m-k-1}^{m-k}\frac{dx}{x^2} =$$

$$= \frac{n^4}{2m(m-n)(m-2n+1)},$$

отже,

$$\ln G(n) =$$

$$= -\sum_{k=0}^{n-1}\left(\frac{n}{m-k}+\frac{n^2}{2(m-k)^2}\right)-\theta\,\Delta_1,\ \theta\in(0,1), \qquad (4)$$

де

$$\Delta_1 = \frac{n^4}{2m(m-n)(m-2n+1)} = O\left(\frac{n^2}{m^2}\right),\ m \to \infty.$$

Далі оцінимо суму в (4):

$$\sum_{k=0}^{n-1}\left(\frac{n}{m-k} + \frac{n^2}{2(m-k)^2}\right) =$$

$$= \sum_{k=0}^{n-1}\left\{\left(\frac{m}{n} - \frac{k}{n}\right)^{-1} + \frac{1}{2}\left(\frac{m}{n} - \frac{k}{n}\right)^{-2}\right\} =$$

$$= \sum_{k=0}^{n-1} f\left(\frac{k}{n}\right),$$

де

$$f(x) = \frac{1}{c-x} + \frac{1}{2(c-x)^2},\ c = \frac{m}{n},\ x \in (0,1).$$

Очевидно, що $f(x)$ зростає при $x \in (0,1)$. Отже, для $k = \overline{0, n-1}$ буде виконуватися нерівність (далі A_k позначено інтервал $\left(\frac{k}{n}, \frac{k+1}{n}\right)$, $k = \overline{0, n-1}$):

$$0 \le n \cdot \int_{A_k} f(x)\,dx - f\left(\frac{k}{n}\right) \le$$

$$\le n \cdot \int_{A_k} f\left(\frac{k+1}{n}\right)dx - f\left(\frac{k}{n}\right) =$$

$$= f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right).$$

Звідси

$$0 \le n \cdot \int_0^1 f(x)\,dx - \sum_{k=0}^{n-1} f\left(\frac{k}{n}\right) \le f(1) - f(0)$$

або

$$\sum_{k=0}^{n-1} f\left(\frac{k}{n}\right) = n \cdot \int_0^1 f(x)\,dx - \theta(f(1) - f(0)).$$

Тут і надалі θ буде позначати число, що належить інтервалу $[0,1]$. При цьому писати індекс при θ, щоб розрізняти різні числа, іноді не будемо (для спрощення запису).

Щоб підрахувати суму в лівій частині останньої рівності, знайдемо

$$f(1) - f(0) = \frac{n^2}{m(m-n)} + \frac{n^2}{2} \cdot \frac{2mn - n^2}{m^2(m-n)^2} <$$

$$< \frac{n^2}{m(m-n)} + \frac{n^3}{m(m-n)^2} := \Delta_2$$

та

$$n \cdot \int_0^1 f(x)\,dx = n\,\ln\frac{m}{m-n} + \frac{n^3}{2m(m-n)}.$$

Тепер можемо записати, що

$$\ln G(n) = n\,\ln\left(1 - \frac{n}{m}\right) - \frac{n^3}{2m(m-n)} + \theta_2\Delta_2 - \theta_1\Delta_1,$$

і далі оцінити логарифм з допомогою нерівності (3):

$$-\frac{n}{m} - \frac{n^2}{2m^2} \ge \ln\left(1 - \frac{n}{m}\right) \ge -\frac{n}{m} - \frac{n^2}{2m(m-n)},$$

тобто

$$\ln\left(1 - \frac{n}{m}\right) = -\frac{n}{m} - \frac{n^2}{2m(m-n)} +$$

$$+ \theta\left(\frac{n^2}{2m(m-n)} - \frac{n^2}{2m^2}\right) =$$

$$= -\frac{n}{m} - \frac{n^2}{2m(m-n)} + \theta\frac{n^3}{2m^2(m-n)}.$$

Звідси

$$\ln G(n) = -\frac{n^2}{m} - \frac{n^3}{m(m-n)} + \theta_3\Delta_3 + \theta_2\Delta_2 - \theta_1\Delta_1, \quad (5)$$

де

$$\Delta_1 = \frac{n^4}{2m(m-n)(m-2n+1)} = O\left(\frac{n^2}{m^2}\right),\ m \to \infty,$$

$$\Delta_2 = \frac{n^2}{m(m-n)} + \frac{n^3}{m(m-n)^2} = O\left(\frac{n^2}{m^2}\right),\ m \to \infty,$$

$$\Delta_3 = \frac{n^4}{2m^2(m-n)} = O\left(\frac{n^2}{m^2}\right),\ m \to \infty,$$

$$\theta_1, \theta_2, \theta_3 \in (0,1).$$

Таким чином величину $\ln G(n)$, а отже, і ймовірність $G(n)$ можна підраховувати з точністю $O\left(\frac{n^2}{m^2}\right),\ m \to \infty$. Це добра точність бо, як побачимо далі, різниця сусідніх ймовірностей $G(n-1)$ і $G(n)$ має порядок лише $O\left(\frac{n}{m}\right)$:

$$G(n-1)-G(n)=O\left(\frac{n}{m}\right),\ m\to\infty.$$

Формула (5) дає розв'язок прямої задачі А для статистики Фермі. Далі знайдемо розв'язок для оберненої задачі Б.

Отже, нехай задано число $p\in(0,1)$, тоді кількість частинок $n(m)$ із задачі Б задовольняє співвідношення

$$F(n-1)<p\le F(n),$$

звідки

$$G(n)\le 1-p<G(n-1),$$

$$-\ln G(n-1)<-\ln(1-p)\le-\ln G(n).$$

Позначимо $a=-\ln(1-p)>0$, тоді

$$a=-\ln G(n)-\theta_m(\ln G(n-1)-\ln G(n)). \qquad (6)$$

З формули (6) далі отримаємо неявне співвідношення між кількістю комірок m та шуканою кількістю частинок $n(m)$.

З рівності (5) випливає, що

$$\ln G(n)=-\frac{n^2}{m}-\frac{n^3}{m^2}+O\left(\frac{n^2}{m^2}\right),\ m\to\infty.$$

Отже, вираз у дужках в (6) допускає оцінку:

$$-\ln G(n)+\ln G(n-1)=$$

$$=\frac{n^2}{m}+\frac{n^3}{m^2}-\frac{(n-1)^2}{m}-\frac{(n-1)^3}{m^2}+O\left(\frac{n^2}{m^2}\right)=$$

$$=2\frac{n}{m}+O\left(\frac{n^2}{m^2}\right),\ m\to\infty,$$

тобто тепер, підставивши це значення в (6), отримаємо

$$a=\frac{n^2}{m}+\frac{n^3}{m^2}-2\theta_m\frac{n}{m}+O\left(\frac{n^2}{m^2}\right),\ m\to\infty.$$

З цієї рівності, що пов'язує неявним чином змінні m та $n(m)$, можна знайти і явний вираз для $n(m)$:

$$n(m)=\sqrt{a\,m}-\frac{a}{2}+\theta_m+\alpha_m, \qquad (7)$$

де $\theta_m\in(0,1)$, $\alpha_m=o(1)$, $m\to\infty$.

Зробивши заміну $\lambda_m=\theta_m+\alpha_m$, зведемо формулу (7) до вигляду

$$n(m)=\sqrt{a\,m}-\frac{a}{2}+\lambda_m, \qquad (8)$$

де $0\le\underline{\lim}\,\lambda_m,\ \overline{\lim}\,\lambda_m\le 1$.

Насправді для послідовності $\{\lambda_m : m\ge 1\}$ виконуються рівності

$$\underline{\lim}\,\lambda_m=0,\ \overline{\lim}\,\lambda_m=1,$$

бо дробові частини $\{\sqrt{a\,m}\}$, $m\ge 1$, всюди щільно заповнюють відрізок $(0,1)$, а у лівій частині формули (8) стоїть ціле число.

Формула (8) дає розв'язок задачі Б для статистики Фермі і є уточненням формули (1) із теореми 1.

Отже, маємо такі результати для статистики Фермі. Для розв'язання прямої задачі А можна користуватися формулою (5):

$$F_m(n)=1-\exp\left(-\frac{n^2}{m}-\frac{n^3}{m^2}\right)+O\left(\frac{n^2}{m^2}\right),$$

а для розв'язання оберненої задачі Б можна користуватися формулою (8):

$$n(m)=\sqrt{a\,m}-\frac{a}{2}+\lambda_m,$$

де $\underline{\lim}\,\lambda_m=0,\ \overline{\lim}\,\lambda_m=1$, припускаючи, що $\lambda_m\in(0,1)$, і потім перевіряючи виконання нерівностей $F(n-1)<p\le F(n)$ для знайденого, у припущенні $\lambda_m\in(0,1)$, з формули (8) значення $n(m)$.

***Приклад* 1**. Нехай у задачі Б $m=10^{12}$, $p=0,5$, тоді з формули (8), у припущенні $\lambda_m\in(0,1)$, отримаємо значення для $n_F(m)=832\,555$. Далі з допомогою формули (5) знаходимо значення для $F(n)$ та $F(n-1)$:

$$F(n)=0,500000612274...,$$

$$F(n-1)=0,49999977971...$$

Отже, формула (8) дала правильну відповідь (припущення $\lambda_m\in(0,1)$ виявилося у цьому випадку правильним).

Статистика Ювала

Далі знайдемо розв'язок задач А і Б для статистики Ювала і порівняємо отримані формули з аналогічними формулами для статистики Фермі.

Отже, розглядаємо рівномірне та незалежне розміщення частинок двох типів: по n частинок кожного типу в m комірках. Тоді ймовірність $V(n)$ того, що відсутня комірка, яка містить частинки обох типів, дорівнює (див. вище)

$$V(n) = \sum_{k=0}^{n-1} P(\mu_0 = m - n + k)\left(1 - \frac{n-k}{m}\right)^n = M\left(1 - \frac{n - \xi_n}{m}\right)^n,$$

де μ_0 – кількість порожніх комірок, що залишилися після розміщення n частинок першого типу, $\xi_n = \mu_0 - m + n$.

Для числового підрахунку ймовірності $V(n)$ нам знадобиться наведена нижче лема, що оцінює різницю виразів $(1-x)^n$ та e^{-nx}. З допомогою цієї леми можна буде зробити перехід

$$V(n) = M\left(1 - \frac{n-\xi_n}{m}\right)^n = M\, e^{-\frac{n-\xi_n}{m}\cdot n} + R$$

з певним залишковим членом R.

Лема. Нехай $x \in (0,1)$, $n \geq 1$, тоді

$$(1-x)^n = e^{-nx} - \frac{n}{2}x^2 e^{-nx} - \theta_1 \frac{n}{2}x^3 e^{-nx} + \theta_2 \frac{n^2}{2}x^4 e^{-nx},$$

де $\theta_1, \theta_2 \in (0,1)$.

Доведення. Маємо

$$0 \leq e^{-nx} - (1-x)^n = e^{-nx}(1 - e^{nx}(1-x)^n) = e^{-nx}(1 - (1-y)^n), \qquad (9)$$

де

$$y = y(x) = 1 - e^x(1-x) = \sum_{k=2}^{\infty} \frac{k-1}{k!}x^k.$$

Зазначимо, що при $x \in (0,1)$ для функції $y(x)$ виконуються такі нерівності:

$$y(x) = x^2 \sum_{k=2}^{\infty} \frac{k-1}{k!}x^{k-2} \leq x^2 \sum_{k=2}^{\infty} \frac{k-1}{k!} = x^2$$

і

$$y(x) = \frac{x^2}{2} + \sum_{k=3}^{\infty} \frac{k-1}{k!}x^k \leq \frac{x^2}{2} + x^3 \sum_{k=3}^{\infty} \frac{k-1}{k!} = \frac{x^2}{2} + \frac{x^3}{2}.$$

Або: $y = \theta x^2$ та $y = \frac{x^2}{2} + \theta\frac{x^3}{2}$.

Далі, з формули Тейлора для функції $f(y) = (1-y)^n$, маємо

$$(1-y)^n = 1 - ny + \theta\frac{n(n-1)}{2}y^2.$$

Підставимо це значення у формулу (9) і потім застосуємо доведені вище нерівності для функції $y(x)$:

$$e^{-nx} - (1-x)^n = e^{-nx}(1 - (1-y)^n) = e^{-nx}\left(ny - \theta\frac{n(n-1)}{2}y^2\right) =$$

$$= n\,e^{-nx}\left(\frac{x^2}{2} + \theta_1\frac{x^3}{2}\right) - \theta_2\, e^{-nx}\frac{n^2}{2}x^4 =$$

$$= \frac{n}{2}x^2 e^{-nx} + \theta_1\frac{n}{2}x^3 e^{-nx} - \theta_2\frac{n^2}{2}x^4 e^{-nx},$$

з останньої рівності випливає твердження леми. Лему доведено.

З леми отримуємо такий вираз для ймовірності $V(n)$:

$$V(n) = M\left(1 - \frac{n-\xi_n}{m}\right)^n =$$

$$= M\, e^{-\frac{n-\xi_n}{m}\cdot n} - \frac{n}{2}M\left(\frac{n-\xi_n}{m}\right)^2 e^{-\frac{n-\xi_n}{m}\cdot n} -$$

$$- \theta\frac{n}{2}M\left(\frac{n-\xi_n}{m}\right)^3 e^{-\frac{n-\xi_n}{m}\cdot n} +$$

$$+ \theta\frac{n^2}{2}M\left(\frac{n-\xi_n}{m}\right)^4 e^{-\frac{n-\xi_n}{m}\cdot n}. \qquad (10)$$

Для спрощення запису тут і, іноді, в подальшому не пишемо індекси за різних значень $\theta \in (0,1)$.

Метою подальших перетворень є, як і у випадку статистики Фермі, підрахунок ймовірності $V(n)$ з точністю $O\left(\frac{n^2}{m^2}\right)$, $m \to \infty$. Для цього спочатку оцінимо третє і четверте математичні сподівання у формулі (10). Нагадаємо,

що ці оцінки отримуються у припущенні $\frac{n^2}{m} = O(1)$, $n, m \to \infty$. Маємо

$$\frac{n}{2} M\left(\frac{n-\xi_n}{m}\right)^3 e^{-\frac{n-\xi_n}{m}\cdot n} \le \frac{n}{2} M\left(\frac{n-\xi_n}{m}\right)^3 \le$$

$$\le \frac{n^4}{2m^3} = O\left(\frac{n^2}{m^2}\right), \quad m \to \infty, \tag{11}$$

і аналогічно

$$\frac{n^2}{2} M\left(\frac{n-\xi_n}{m}\right)^4 e^{-\frac{n-\xi_n}{m}\cdot n} \le \frac{n^6}{2m^4} = O\left(\frac{n^2}{m^2}\right), \tag{12}$$

$$m \to \infty.$$

Друге математичне сподівання з (10) підрахуємо з точністю $O\left(\frac{n^2}{m^2}\right)$ таким чином:

$$\frac{n}{2} M\left(\frac{n-\xi_n}{m}\right)^2 e^{-\frac{n-\xi_n}{m}\cdot n} = \frac{n}{2} M\left(\frac{n}{m}\right)^2 e^{-\frac{n-\xi_n}{m}\cdot n} +$$

$$+ \frac{n}{2} M\left\{\left(\frac{n-\xi_n}{m}\right)^2 - \left(\frac{n}{m}\right)^2\right\} e^{-\frac{n-\xi_n}{m}\cdot n},$$

але

$$\frac{n}{2} M\left\{\left(\frac{n-\xi_n}{m}\right)^2 - \left(\frac{n}{m}\right)^2\right\} e^{-\frac{n-\xi_n}{m}\cdot n} \le \frac{n}{2} M \frac{2n\xi_n}{m^2} =$$

$$= \frac{n^2}{m^2} M\xi_n = O\left(\frac{n^2}{m^2}\right), \quad m \to \infty,$$

бо відомо [3], що $M\mu_0 = m\left(1-\frac{1}{m}\right)^n$ і звідси

$$M\xi_n = M(\mu_0 - m + n) = m\left(1-\frac{1}{m}\right)^n - m + n =$$

$$= m\left(1 - \frac{n}{m} + \frac{n(n-1)}{2m^2} - \theta\frac{n^3}{6m^3}\right) - m + n =$$

$$= \frac{n(n-1)}{2m} - \theta\frac{n^3}{6m^2} = O(1), \quad m \to \infty.$$

З останньої рівності також маємо й оцінку для $M\xi_n$:

$$\frac{n(n-1)}{2m} - \frac{n^3}{6m^2} \le M\xi_n \le \frac{n(n-1)}{2m}, \tag{13}$$

яка знадобиться в подальшому.

Продовжимо підрахунок другого математичного сподівання з (10). Зараз маємо:

$$\frac{n}{2} M\left(\frac{n-\xi_n}{m}\right)^2 e^{-\frac{n-\xi_n}{m}\cdot n} =$$

$$= \frac{n}{2} M\left(\frac{n}{m}\right)^2 e^{-\frac{n-\xi_n}{m}\cdot n} - \theta\frac{n^2}{m^2} M\xi_n.$$

Далі з формули $e^x = 1 + \theta x e^x$, $x \ge 0$, отримаємо

$$\frac{n}{2} M\left(\frac{n}{m}\right)^2 e^{-\frac{n-\xi_n}{m}\cdot n} = \frac{n^3}{2m^2} e^{-\frac{n^2}{m}} M e^{\frac{n}{m}\xi_n} =$$

$$= \frac{n^3}{2m^2} e^{-\frac{n^2}{m}} M\left(1 + \theta\frac{n}{m}\xi_n e^{\frac{n}{m}\xi_n}\right) =$$

$$= \frac{n^3}{2m^2} e^{-\frac{n^2}{m}} + \theta\frac{n^4}{2m^3} M\xi_n e^{-\frac{n-\xi_n}{m}\cdot n} =$$

$$= \frac{n^3}{2m^2} e^{-\frac{n^2}{m}} + \theta\frac{n^4}{2m^3} M\xi_n,$$

звідки маємо оцінку з точністю $O\left(\frac{n^2}{m^2}\right)$ для другого доданка з формули (10):

$$\frac{n}{2} M\left(\frac{n-\xi_n}{m}\right)^2 e^{-\frac{n-\xi_n}{m}\cdot n} =$$

$$= \frac{n^3}{2m^2} e^{-\frac{n^2}{m}} + \theta\frac{n^4}{2m^3} M\xi_n - \theta\frac{n^2}{m^2} M\xi_n =$$

$$= \frac{n^3}{2m^2} e^{-\frac{n^2}{m}} + O\left(\frac{n^2}{m^2}\right), \quad m \to \infty. \tag{14}$$

Перший доданок із формули (10) підрахуємо з допомогою співвідношення $e^x = 1 + x + \theta\frac{x^2}{2} e^x$, $x \ge 0$:

$$M e^{-\frac{n-\xi_n}{m}\cdot n} = e^{-\frac{n^2}{m}} M e^{\frac{n}{m}\xi_n} =$$

$$= e^{-\frac{n^2}{m}} M\left(1 + \frac{n}{m}\xi_n + \theta\frac{n^2}{2m^2}\xi_n^2 e^{\frac{n}{m}\xi_n}\right) =$$

$$= e^{-\frac{n^2}{m}} + \frac{n}{m} e^{-\frac{n^2}{m}} M\xi_n + \theta\frac{n^2}{2m^2} M\xi_n^2 =$$

$$= e^{-\frac{n^2}{m}} + \frac{n}{m} e^{-\frac{n^2}{m}} M\xi_n + O\left(\frac{n^2}{m^2}\right), \quad m \to \infty,$$

бо для $M\xi_n^2$ можна отримати, так само, як і для $M\xi_n$, оцінку типу $O(1)$:

$$M\xi_n^2 \le \frac{n^4}{4m^2} + \frac{n^2}{2m} + \frac{1}{10}\frac{n^5}{4m^3} = O(1),\ m \to \infty. \quad (15)$$

Зберемо тепер всі оцінки з формул (11), (12) і (14) в одну формулу:

$$V(n) = e^{-\frac{n^2}{m}} + \frac{n}{m}e^{-\frac{n^2}{m}}M\xi_n - \frac{n^3}{2m^2}e^{-\frac{n^2}{m}} + \\ + \theta\frac{n^2}{2m^2}M\xi_n^2 - \theta\frac{n^4}{2m^3}M\xi_n + \\ + \theta\frac{n^2}{m^2}M\xi_n - \theta\frac{n^4}{2m^3} + \theta\frac{n^6}{2m^4}. \quad (16)$$

Цю формулу, з нерівностями (13) та (15), можна використовувати для числового підрахунку ймовірності $V(n)$, тобто для розв'язання прямої задачі А для статистики Ювала.

Аналогічно випадку статистики Фермі можна тепер розв'язати і обернену задачу Б для статистики Ювала. Маємо з (13):

$$M\xi_n = \frac{n^2}{2m} + O\left(\frac{n}{m}\right),\ m \to \infty.$$

Отже,

$$V(n) = e^{-\frac{n^2}{m}}\left(1 + \frac{n}{m}M\xi_n - \frac{n^3}{2m^2}\right) + O\left(\frac{n^2}{m^2}\right) = \\ = e^{-\frac{n^2}{m}} + O\left(\frac{n^2}{m^2}\right),\ m \to \infty,$$

звідки випливає, що кількість частинок $n(m)$ із задачі Б дорівнює

$$n_U(m) = \sqrt{a\,m} + \lambda_m, \quad (17)$$

де $\underline{\lim}\,\lambda_m = 0$, $\overline{\lim}\,\lambda_m = 1$, $a = -\ln(1-p)$.

Відзначимо, що для статистики Фермі відповідна кількість частинок, з формули (8), буде меншою (не більшою):

$$n_F(m) = \sqrt{a\,m} - \frac{a}{2} + \lambda_m.$$

Це можна було передбачити заздалегідь, бо у статистиці Фермі частинки одного типу розміщуються по одній і займають "більше" комірок, ніж частинки одного типу в статистиці Ювала, які можуть потрапляти в одну комірку. Тому і ймовірність існування комірки з частинками обох типів, при однакових n та m, для статистики Фермі вища, і тому $n_F(m) \le n_U(m)$.

***Приклад* 2**. Нехай у задачі Б $m = 10^{12}$, $p = 0,5$, тоді з формули (17), у припущенні $\lambda_m \in (0,1)$, отримаємо значення для $n_U(m) =$ $= 832\,555$. Далі з допомогою формули (16) знаходимо значення для $U(n)$ та $U(n-1)$:

$$U(n) = 0,500000323732...,$$

$$U(n-1) = 0,49999949117...$$

Отже, формула (17) дала правильну відповідь (припущення $\lambda_m \in (0,1)$ виявилося у цьому випадку правильним).

Цікаво, що в цьому випадку (див. приклад 1) величини $n_F(m)$ та $n_U(m)$ є рівними: $n_F(m) = n_U(m) = 832\,555$, хоча, звичайно, ймовірність $F(n)$ є більшою за ймовірність $U(n)$:

$$F(n) = 0,500000612274...,$$

$$U(n) = 0,500000323732...,$$

$$F(n-1) = 0,49999977971...,$$

$$U(n-1) = 0,49999949117...$$

Якщо ж у задачі Б $m = 10^{12}$, $p = 0,99$, то з формул (8) та (17), у припущенні $\lambda_m \in (0,1)$, отримаємо значення

$$n_F(m) = 2\,145\,964,$$

$$n_U(m) = 2\,145\,967.$$

Тут уже $n_F(m) < n_U(m)$. Далі з формул (5) і (16) отримаємо числові значення для ймовірностей $F(n)$, $F(n-1)$, $U(n)$ і $U(n-1)$:

$$F(n) = 0,990000011858...,$$

$$U(n) = 0,990000041...,$$

$$F(n-1) = 0,98999996893...,$$

$$U(n-1) = 0,98999998...$$

Бачимо, що зі зростанням параметра $\frac{n^2}{m} = O(1)$ точність оцінок погіршується. Також у випадку $m = 10^{12}$, $p = 0,99$ не виконується (!) нерівність

$$F(n_F(m)) > U(n_U(m)).$$

Це відбувається тому, що тут $n_F(m) \neq n_U(m)$.

Висновки

У статті отримано формули (5) та (16), які дають розв'язок прямої задачі про дні народження для статистик Фермі та Ювала відповідно. Формули (8) і (17) дають розв'язок вже оберненої задачі про дні народження для тих же статистик.

Як і слід було очікувати, кількість частинок n_F у формулі (8) для статистики Фермі виявилася меншою за аналогічну кількість частинок n_U у формулі (17) для статистики Ювала. Асимптотична різниця цих виразів дорівнює певній додатній константі. Це показує, що статистики, які розглядаються, є "різними" (незважаючи на результат теореми), і дає оцінку їх "віддаленості".

Наведену у статті методику можна в подальшому для розв'язання задачі про дні народження й для інших статистик (не лише для статистик Фермі та Ювала).

Список літератури

1. *Секей Г.* Парадоксы в теории вероятностей и мат. статистике. – М.: Мир, 1990. – 240 с.
2. *Основы* криптографии / А.П. Алферов, А.Ю. Зубов, А.С. Кузьмин, А.В. Черемушкин. – М.: Гелиос АРВ, 2001. – 480 с.
3. *Колчин В.Ф., Севастьянов Б.А., Чистяков В.П.* Случайные размещения. – М.: Наука, 1976. – 224 с.
4. DasGupta A. The matching, birthday and strong birthday problem: a contemporary review // J. Stat. Plan. Inference. – 2005. – **130**. – P. 377–389.

References

1. G. Szekely, *Paradoxes in Probability* Theory and Mathematical Statistics. Moscow, USSR: Mir, 1990, 240 p. (in Russian).
2. A.P. Alferov *et al.*, *Handbook of Cryptography*. Moscow, Russia: Gelios, 2001, 480 p. (in Russian).
3. V.F. Colchin *et al.*, *Random Allocations*. Moscow, USSR: Nauka, 1976, 224 p. (in Russian).
4. A. DasGupta, "The matching, birthday and strong birthday problem: a contemporary review", *J. Stat. Plan. Inference*, vol. 130, pp. 377–389, 2005.

П.О. Єндовицький

ДВІ МОДИФІКАЦІЇ ЗАДАЧІ ПРО ДНІ НАРОДЖЕННЯ

Проблематика. Схема розміщення частинок по комірках досліджується як у теорії ймовірностей, так і в математичній статистиці. В теорії ймовірностей мова йде про доведення граничних теорем для цієї схеми, в математичній статистиці – про побудову статистичних критеріїв. Одним із важливих питань у цій теорії є задача про дні народження.

Мета дослідження. У статті розглядаються дві модифікації класичної задачі про дні народження. Одна модифікація формулюється у схемі статистики Фермі, інша – в схемі рівномірного та незалежного розміщення частинок по комірках. В обох випадках метою дослідження є розв'язок задачі про дні народження.

Методика реалізації. Використовувалися стандартні асимптотичні методи. При цьому спочатку було доведено певну граничну теорему та знайдено швидкість збіжності в ній. З допомогою цих результатів було проведено числовий підрахунок ймовірностей у задачі про дні народження та отримано формули для розміру групи в цій задачі.

Результати дослідження. У результаті були отримані числові оцінки для ймовірності та розміру групи із задачі про дні народження.

Висновки. Для обох модифікацій збігається головний член асимптотики як у формулі для підрахунку ймовірності, так і у формулі для розміру групи, але вже другі доданки в отриманих асимптотичних формулах відрізняються.

Ключові слова: задача про дні народження; парадокс днів народжень; випадкові розміщення; статистика Фермі; атака Ювала.

П.А. Ендовицкий

ДВЕ МОДИФИКАЦИИ ЗАДАЧИ ПРО ДНИ РОЖДЕНИЯ

Проблематика. Схема размещения частиц по ячейкам исследуется как в теории вероятностей, так и в математической статистике. В теории вероятностей речь идет о предельных теоремах для этой схемы, в математической статистике – о построении статистических критериев. Одним из важных вопросов в этой теории является задача про дни рождения.

Цель исследования. В статье рассмотрены две модификации классической задачи про дни рождения. Одна модификация формулировалась в схеме статистики Ферми, вторая – в терминах равновероятного и независимого размещения частиц по ячейкам. В обоих случаях целью исследования было решение задачи про дни рождения.

Методика реализации. Использовались стандартные асимптотические методы. При этом сначала была доказана определенная предельная теорема и найдена скорость сходимости в ней. С помощью этих результатов был проведен численный подсчет вероятностей в задаче про дни рождения и получены формулы для размера группы в этой задаче.

Результаты исследования. В результате были получены числовые оценки для вероятностей и размера группы из задачи про дни рождения.

Выводы. Для обеих модификаций совпадает главный член асимптотики как в формуле для подсчета вероятностей, так и в формуле для размера группы, но уже вторые слагаемые в полученных асимптотических формулах отличаются.

Ключевые слова: задача про дни рождения; парадокс дней рождений; случайные размещения; статистика Ферми; атака Ювала.

Рекомендована Радою
фізико-математичного факультету
НТУУ “КПІ”

Надійшла до редакції
28 травня 2015 року