Примеры решения домашней самостоятельной работы №3 по математической статистике.

1. (1.00) . Найти точечную оценку дисперсии генеральной совокупности для данной выборки

X_i	-2	-1	0	1	2
n_i	1	1	2	3	3

$\hat{M}[X] = \ldots\ldots;\quad \hat{D}[X] = \ldots\ldots$

Решение.

Точечные оценки математического ожидания и дисперсии генеральной совокупности получаются на основе выборочных:

$$\hat{E}[X] = \overline{X} = \langle X\rangle = \frac{\sum n_i X_i}{\sum n_i} = \frac{1\times(-2)+1\times(-1)+2\times 0+3\times 1+3\times 2}{1+1+2+3+3} = \frac{6}{10} = 0,60;$$

$$\widetilde{S}^2[X] = \frac{\sum n_i X_i^2}{\sum n_i} - \overline{X}^2 = \frac{1\times(-2)^2+1\times(-1)^2+2\times 0^2+3\times 1^2+3\times 2^2}{1+1+2+3+3} - \left(\frac{6}{10}\right)^2 =$$

$$= \frac{20}{10} - \frac{36}{100} = \frac{164}{100} = \frac{41}{25} = 1,64;$$

$$\hat{D}[X] = S^2[X] = \widetilde{S}^2[X]\frac{N}{N-1} = \frac{41}{25}\frac{10}{9} = \frac{82}{45} \simeq 1,82.$$

Ответ: $\hat{E}[X] = 0,60;\quad \hat{D}[X] = 1,82.$

2.-3.

Метод моментов

Для большинства распределений (но не для всех) существуют конечные моменты распределения, т.е. величины $\nu_m = \sum X_i^m p_i$ для дискретных и $\nu_m = \int x^m f(x) dx$ для непрерывных распределений. Эти величины могут быть найдены (аналитически или численно) для любых значений параметров. Например, для нормального закона аналитически найдены все центральные моменты

$$\mu_m = \int (x - E[X])^m f(x) dx = \begin{cases} 0 & \text{при} \quad m = 2k+1 \\ (m-1)!!\sigma^m & \text{при} \quad m = 2k \end{cases}$$

Остается приравнять теоретические моменты к выборочным, и мы получим систему уравнений для определения оптимального набора параметров распределения. При таком подходе надо учитывать некоторые тонкости

- Очевидно, что число уравнений должно быть равно числу параметров, иначе мы рискуем получить несовместную систему.
- Надо учитывать возможную смещенность выборочных моментов и в уравнениях использовать уточненные (поправленные) оценки вместо непосредственного использования выборочных моментов.

Недостатки метода моментов состоят в том, что

- существует неопределенность (произвол) при выборе порядков моментов
- требуется, чтобы эти моменты существовали

Пример 1.

По данной выборке

X_i	1	2	3	4	5
n_i	20	20	30	20	10

найти оценки параметров *равномерного* распределения, используя метод моментов.

Решение.

Равномерное распределение характеризуется двумя параметрами a и b, которые соответствуют левой и правой границе.

Приравняем теоретические моменты к оценкам, в качестве которых будем использовать выборочные характеристики, поправленные на величину смещения:

$$\nu_1 = \int_a^b \frac{x dx}{b-a} = \frac{a+b}{2} = \hat{\nu}_1 = \bar{x}$$

$$\nu_2 = \int_a^b \frac{x^2 dx}{b-a} = \frac{a^2+ab+b^2}{3} = \overline{x^2}$$

$$\mu_2 = \nu_2 - \nu_1^2 = \frac{(b-a)^2}{12} = \hat{\mu}_2 = S^2 \Leftrightarrow \frac{b-a}{2} = \sqrt{3S^2}$$

$$\hat{a} = \bar{x} - \sqrt{3S^2}; \qquad \hat{b} = \bar{x} + \sqrt{3S^2}$$

$$\bar{x} = \frac{20\cdot1+20\cdot2+30\cdot3+20\cdot4+10\cdot5}{20+20+30+20+10} = \frac{14}{5} = 2.8$$

$$\overline{x^2} = \frac{20\cdot1^2+20\cdot2^2+30\cdot3^2+20\cdot4^2+10\cdot5^2}{20+20+30+20+10} = \frac{47}{5} = 9.4$$

$$\tilde{S}^2 = 9.4 - 2.8^2 = 1.56$$

$$S^2 = \frac{n}{n-1}\tilde{S}^2 = \frac{100}{99}\cdot 1.56 = 1.575757\ldots$$

$$\sqrt{3S^2} = \sqrt{3\cdot\frac{100}{99}\cdot 1.56} = 2.17422923$$

$$\hat{a} = \bar{x} - \sqrt{3S^2} = 2.8 - 2.17422923 = 0.625770\ldots \simeq 0.626$$

$$\hat{b} = \bar{x} + \sqrt{3S^2} = 2.8 + 2.17422923 = 4.974229\ldots \simeq 4.974$$

Пример 2.

По данной выборке

X_i	1	2	3	4	5
n_i	20	20	30	20	10

найти оценки параметров *экспоненциального* распределения $F(x) = (1 - \exp(-\lambda(x-a)))\eta(x-a)$, используя метод моментов.

Решение.

Для нахождения теоретических выражений моментов как функций от параметров надо найти плотность распределения вероятностей $f(x) = \frac{d}{dx}F(x) = \lambda\exp(-\lambda(x-a))\eta(x-a)$.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} x\lambda\exp(-\lambda(x-a))\eta(x-a)dx =$$

$$= \int_{a}^{\infty} x\lambda\exp(-\lambda(x-a))dx = \exp(\lambda a)\int_{a}^{\infty} x\lambda\exp(-\lambda x)dx =$$

$$= -\exp(\lambda a)\int_{a}^{\infty} xd\exp(-\lambda x) = -\exp(\lambda a)\left[x\exp(-\lambda x)|_a^\infty - \int_a^\infty \exp(-\lambda x)dx \right] =$$

$$= -\exp(\lambda a)\left[-a\exp(-\lambda a) + \frac{1}{\lambda}\int_a^\infty \exp(-\lambda x)d(-\lambda x) \right] =$$

$$= -\exp(\lambda a)\left[-a\exp(-\lambda a) + \frac{1}{\lambda}\exp(-\lambda x)|_a^\infty \right] =$$

$$= -\exp(\lambda a)\left[-a\exp(-\lambda a) - \frac{1}{\lambda}\exp(-\lambda a) \right] = a + \frac{1}{\lambda}.$$

Этот же результат можно получить проще, используя свойства несмещенного экспоненциального распределения:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} (x-a+a)\lambda\exp(-\lambda(x-a))\eta(x-a)dx =$$

$$= \int_{-\infty}^{\infty} (x-a)\lambda\exp(-\lambda(x-a))\eta(x-a)dx + a\int_{-\infty}^{\infty} \lambda\exp(-\lambda(x-a))\eta(x-a)dx =$$

$$= \int_{-\infty}^{\infty} t\lambda\exp(-\lambda t)\eta(t)dt + a\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\lambda} + a.$$

Или еще проще, если заметить, что требуемое распределение отличается от стандартного экспоненциального распределения $F(y) = (1 - \exp(-\lambda y))\eta(y)$ только сдвигом начала координат, что соответствует простейшему преобразованию случайной величины $Y = X - a$, и воспользоваться свойством математического ожидания $E[Y] = E[X-a] = E[X] - a \Leftrightarrow E[X] = E[Y] + a = \frac{1}{\lambda} + a$.

Для второго центрального момента $D[Y] = E\left[(Y - E[Y])^2\right] = \frac{1}{\lambda^2}$, и можно воспользоваться свойством дисперсии $D[Y] = D[X-a] = D[X] = \frac{1}{\lambda^2}$.

Приравняем теоретические моменты к оценкам, в качестве которых будем использовать выборочные характеристики, поправленные на величину смещения:

$$\nu_1 = E[X] = \hat{E}[X] = \frac{1}{\hat{\lambda}} + \hat{a} = \bar{x} = 2.80$$

$$\mu_2 = D[X] = \frac{1}{\hat{\lambda}^2} = S^2 = \frac{n}{n-1}\tilde{S}^2 = 1.575757\ldots \simeq 1.5758$$

$$\hat{\lambda} = \frac{1}{s} = \frac{1}{\sqrt{\frac{n}{n-1}\tilde{S}^2}} = \frac{1}{\sqrt{\frac{100}{99}\cdot 1.56}} = 0.7966275\ldots \simeq 0.797$$

$$\hat{a} = \bar{x} - \frac{1}{\hat{\lambda}} = \bar{x} - s = 2.80 - \sqrt{\frac{100}{99}\cdot 1.56} = 1.5447\ldots \simeq 1.545$$

Пример 3.

По данной выборке

X_i	1	2	3	4	5
n_i	20	20	30	20	10

найти оценки параметров *нормального* распределения, используя метод моментов.

Решение.

Нормальное распределение характеризуется двумя параметрами μ и σ, которые соответствуют математическому ожиданию и среднеквадратичному отклонению.

Приравняем теоретические моменты к оценкам, в качестве которых будем использовать выборочные характеристики, поправленные на величину смещения:

$$\nu_1 = \mu = \hat{\mu} = \bar{x} = 2.80$$

$$\mu_2 = \sigma^2 = \hat{\mu}_2 = S^2 = \frac{n}{n-1}\tilde{S}^2 = 1.575757\ldots \simeq 1.5758$$

$$\hat{\sigma} = s = \sqrt{\frac{n}{n-1}\tilde{S}^2} = \sqrt{\frac{100}{99} \cdot 1.56} = 1.255\,291\ldots \simeq 1.26$$

Примечание. Надо внимательно отнестись к обозначениям.

Одной и той же греческой буквой μ обозначаются и центральные моменты (всегда с нижним индексом, указывающим порядок момента) и параметр нормального распределения μ (без индекса), совпадающий с математическим ожиданием $\mathrm{E}(X)$ и первым начальным моментом ν_1. К сожалению многие студенты, выполняя эту работу, в качестве оценки параметра μ нимало не задумавшись выдавали оценку $\hat{\mu}_2$.

Пример 4. Предположим, что случайная величина X распределена по *биномиальному* закону с неизвестным параметром p. Статистическое распределение выборки представлено в таблице.

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

Методом моментов найти оценку $\hat{p}$ параметра p, если другой известен другой параметр $n = 10$.

Решение.

Первый начальный момент биномиального распределения $E[X] = np$.

Оценка математического ожидания $\hat{E}[X] = \bar{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} =$

$$= \frac{2 \cdot 0 + 3 \cdot 1 + 10 \cdot 2 + 22 \cdot 3 + 26 \cdot 4 + 20 \cdot 5 + 12 \cdot 6 + 5 \cdot 7}{2 + 3 + 10 + 22 + 26 + 20 + 12 + 5} = 4$$

$np = 4 \Rightarrow \hat{p} = \dfrac{4}{10}$.

Примечание. Параметр найден, однако не факт, что верным является исходное предположение, т.е., что само распределение подчиняется биномиальному закону. И хотя проверка такого рода предположения составляет предмет самостоятельного исследования (статистическая проверка гипотез), которое будет рассмотрено в следующих главах, всё же не мешает выяснить, имеется ли для данного статистического ряда та связь между дисперсией и математическим ожиданием, которая характерна для биномиального распределения. А именно:

$$\begin{cases} E[X] = np \\ D[X] = npq = np(1-p) \end{cases} \Rightarrow D[X] = E[X]\left(1 - \frac{E[X]}{n}\right)$$

Найдём второй начальный выборочный момент $\overline{x^2} = \dfrac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} =$

$$= \frac{2 \cdot 0^2 + 3 \cdot 1^2 + 10 \cdot 2^2 + 22 \cdot 3^2 + 26 \cdot 4^2 + 20 \cdot 5^2 + 12 \cdot 6^2 + 5 \cdot 7^2}{2 + 3 + 10 + 22 + 26 + 20 + 12 + 5} = \frac{1834}{100} = 18.34$$

Выборочная дисперсия $\tilde{S}^2 = \overline{x^2} - (\bar{x})^2 = 18.34 - 4^2 = 2.34$

Оценка дисперсии $\hat{D}[X] = S^2 = \tilde{S}^2 \dfrac{100}{99} = \dfrac{234}{99} = \dfrac{26}{11} = 2.36363636$

Это надо сравнить с $D[X] = E[X]\left(1 - \dfrac{E[X]}{n}\right) = 4 \cdot \dfrac{6}{10} = 2.4.$

Согласие, похоже, есть, но настораживает тот факт, что в статистическом ряде отсутствовали варианты 8, 9 и 10, которые должны были бы присутствовать в выборке. Это свидетельствует о каких-то неполадках в организации исследования, в которых надо обстоятельно разобраться.

Пример 5. Предположим, что случайная величина X распределена по *закону Пуассона* с неизвестным параметром λ. Статистическое распределение выборки представлено в таблице.

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

Методом моментов найти оценку $\hat{\lambda}$ параметра λ.

Решение.

Первый начальный момент распределения Пуассона $E[X] = \lambda$.

Оценка математического ожидания $\hat{E}[X] = \bar{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} =$

$$= \frac{2 \cdot 0 + 3 \cdot 1 + 10 \cdot 2 + 22 \cdot 3 + 26 \cdot 4 + 20 \cdot 5 + 12 \cdot 6 + 5 \cdot 7}{2 + 3 + 10 + 22 + 26 + 20 + 12 + 5} = 4$$

$\hat{\lambda} = \bar{x} = 4.$

Примечание. Параметр найден, однако не факт, что верным является исходное предположение, т.е., что само распределение подчиняется закону Пуассона. Дело в том, что для закона Пуассона характерно совпадение $E[X] = \lambda = D[X]$.

Между тем, как явствует из результатов анализа предыдущего примера, в данном случае $\hat{E}[X] = 4$, а $\hat{D}[X] = \dfrac{26}{11} = 2.36$, что заметно меньше. Вопрос о значимости такого различия остаётся открытым и может быть проверен с помощью одного из т.н. критериев согласия, которые будут рассматриваться в одной из последующих лекций.

Пример 6. Предположим, что случайная величина X — количество неудач до первого успеха — распределена по варианту $\mathrm{Geom}(0,p)$ *геометрического* закона распределения вида $P_X = p(1-p)^x$ с неизвестным параметром p. Статистическое распределение выборки представлено в таблице.

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

Методом моментов найти оценку параметра p.

Решение.

Первый начальный момент такого варианта геометрического распределения $E[X] = \dfrac{1-p}{p}$.

$$E[X] = \bar{x} \Rightarrow \frac{1-p}{p} = \bar{x} \Leftrightarrow \frac{1}{p} - 1 = \bar{x} \Leftrightarrow p = \frac{1}{1+\bar{x}} = \hat{p}$$

Оценка математического ожидания $\hat{E}[X] = \bar{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} =$

$$= \frac{2\cdot 0 + 3\cdot 1 + 10\cdot 2 + 22\cdot 3 + 26\cdot 4 + 20\cdot 5 + 12\cdot 6 + 5\cdot 7}{2+3+10+22+26+20+12+5} = 4$$

$$\hat{p} = \frac{1}{4+1} = \frac{1}{5} = 0.2.$$

Пример 6а. Предположим, что случайная величина X — количество испытаний по схеме Бернулли до первого "успеха" включительно — распределена по *геометрическому* закону распределения Фарри $\mathrm{Geom}(1,p)$: $P_X = p(1-p)^{x-1}$ с неизвестным параметром p. Статистическое распределение выборки представлено в таблице.

x_i	1	2	3	4	5	6	7	8
n_i	2	3	10	22	26	20	12	5

Методом моментов найти оценку параметра p.

Решение.

Первый начальный момент геометрического распределения $E[X] = \dfrac{1}{p}$.

$$E[X] = \bar{x} \Rightarrow \frac{1}{p} = \bar{x} \Leftrightarrow p = \frac{1}{\bar{x}} = \hat{p}$$

Оценка математического ожидания $\hat{E}[X] = \bar{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} =$

$$= \frac{2\cdot 1 + 3\cdot 2 + 10\cdot 3 + 22\cdot 4 + 26\cdot 5 + 20\cdot 6 + 12\cdot 7 + 5\cdot 8}{2+3+10+22+26+20+12+5} = 5$$

$$\hat{p} = \frac{1}{5} = 0.2.$$

Примечание. Семейство геометрических распределений содержит два общих и два частных случая:

Название	Случайная величина X=	Вероятность	min	$E(X)$	$D(X)$
Отрицательное биномиальное	k — количество неудач до появления m-го успеха	$P_{m,p}^{\overline{Bin}}(k) = C_{k+m-1}^{k} p^m q^k$	$k = 0$	$\frac{mq}{p}$	$\frac{mq}{p^2}$
Паскаля	n — количество испытаний до появления m-го успеха	$P_{m,p}^{Pa}(n) = C_{n-1}^{m-1} p^m q^{n-m}$	$n = m$	$\frac{m}{p}$	$\frac{mq}{p^2}$
Геометрическое	k — количество неудач до появления 1-го успеха	$P_p^G(k) = pq^k$	$k = 0$	$\frac{q}{p}$	$\frac{q}{p^2}$
Фарри	n — количество испытаний до появления 1-го успеха	$P_p^F(n) = pq^{n-1}$	$n = 1$	$\frac{1}{p}$	$\frac{q}{p^2}$

Примечание. Ответ получен, но для геометрического распределения часто́ты статистического ряда должны убывать с ростом вариант. За счёт флуктуаций (случайных отклонений) строгая монотонность может нарушаться, но эти отклонения практически не могут быть настолько явно выражены, как это имеет место в данном примере, где присутствует сначала устойчивый рост, затем столь же устойчивый спад. Похоже, когда исследователь выбирал подходящий вид распределения, ему пришла в голову не самая удачная идея остановить свой выбор на геометрическом распределении.

Пример 7. Предположим, что случайная величина X распределена по *гипергеометрическому* закону с неизвестными параметрами M и N. В статистическом исследовании фиксировалась случайная величина X, представляющая количество появления объектов заданного типа в серии из 7 наблюдений.

Статистический ряд выборки (100 серий) представлен в таблице.

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

Методом моментов найти оценку параметров M и N.

Теоретическое введение.

Гипергеометрическое распределение

Лотерея — это не охота за удачей,
это охота за неудачниками.

Рассмотрим совокупность N объектов, которую по каким-то признакам можно разделить на две части, одна из которых целиком состоит из M объектов типа 1, а другая — только из $N-M$ объектов типа 2. Если из такой совокупности осуществить случайную *бесповторную* выборку объемом n объектов, то вероятность того, что эта выборка содержит ровно m элементов типа 1 (и ровно $n-m$ типа 2), определяется формулой $P_n(m) = \dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n}$. Этот закон распределения называется гипергеометрическим. Формула $\dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n}$ для вероятности $P_n(m)$ следует из классического определения. Она же может быть получена из теоремы умножения.

Действительно, количество способов, которыми можно разделить множество из N различимых объектов на две части, одна из которых составляет n объектов, попавших в выборку, а другая состоит из $N-n$ объектов, в выборку не попавших, составляет величину C_N^n. Другими словами, это полное количество всех возможных исходов опыта по осуществлению случайной выборки объема n из числа N различимых объектов.

Полное количество выборок, для которых исходы "благоприятны" для интересующего нас события (ровно m объектов, которые мы классифицируем, как относящиеся к объектам типа 1) также может быть подсчитано. Прежде всего, найдем количество способов поделить группу из всех M объектов первого типа на две части, одна из которых составляет m объектов, попавших в выборку, а другая состоит из $M-m$ объектов, в выборку не попавших — это составляет величину C_M^m. Каждая такая выборка объектов первого типа дополняется до полной выборки любой выборкой объема $n-m$ объектов второго типа из $N-M$ возможных. Количество таких дополнений, очевидно, равно C_{N-M}^{n-m}. Таким образом, всего получаем $C_M^m C_{N-M}^{n-m}$ благоприятных исходов. В соответствие с классическим определением $P_n(m) = \dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n}$.

Примечание. $\dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \dfrac{M!}{m!(M-m)!} \cdot \dfrac{(N-M)!}{(n-m)!(N-M-n+m)!} \cdot \dfrac{n!(N-n)!}{N!} =$

$= \dfrac{M!(N-M)!}{N!} \cdot \dfrac{(N-n)!}{(M-m)!(N-M-n+m)!} \cdot \dfrac{n!}{m!(n-m)!} = \dfrac{C_n^m C_{N-n}^{M-m}}{C_N^M}$

Можно дать теоретико-множественное толкование полученному выражению.
Общее количество способов разделить множество из N различимых объектов на две части, одна из которых составляет M объектов первого типа, а другая состоит из $N - M$ объектов второго типа, составляет величину C_N^M. Из этих способов благоприятными являются любые пары способов — когда в первой группе из n (извлечённых объектов) оказывается ровно m объектов первого типа, а в оставшейся части (из $N - n$ объектов) оказываются все неизвлечённые объекты первого типа в количестве $M - m$ штук.

Примечание. Формулу $P_n(m) = \dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n}$ легко обобщить на случай нескольких типов объектов:
Рассмотрим совокупность N объектов, которую по каким-то признакам можно разделить на k частей, каждая из которых целиком состоит из M_i объектов типа i. Если из такой совокупности осуществить случайную бесповторную выборку объемом n объектов, то вероятность того, что эта выборка содержит ровно m_i элементов типа i ($i = 1, \dots, k$), определяется формулой
$P_n(m_1; m_2; \dots; m_k) = \dfrac{C_{M_1}^{m_1} \dots C_{M_k}^{m_k}}{C_N^n}$. Очевидно, что при этом должны быть выполнены условия $\sum_{i=1}^{k} M_i = N$ и $\sum_{i=1}^{k} m_i = n$.

Очевидно, что $0 \le n \le N$

Кроме того, должны выполняться ограничения и для m :

$$\begin{cases} 0 \le m \le M \\ 0 \le n - m \le N - M \end{cases} \Leftrightarrow \begin{cases} 0 \le m \le M \\ n - (N - M) \le m \le n \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \max(0; n - (N - M)) \le m \le \min(n; M)$$

Примечание. $\sum_{m=0}^{n} P_n(m) = \sum_{m=0}^{n} \dfrac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \dfrac{\sum_{m=0}^{n} C_M^m C_{N-M}^{n-m}}{C_N^n} = 1 \Rightarrow$
$\Rightarrow \sum_{m=0}^{n} C_M^m C_{N-M}^{n-m} = C_N^n$

Производящая функция моментов $M_X(s) = \sum_{m=0}^{n} P_n(m) e^{ms} = \dfrac{\sum_{m=0}^{n} C_M^m C_{N-M}^{n-m} e^{ms}}{C_N^n}$

Обозначим $p = \dfrac{M}{N}$ и $q = 1 - p$, тогда

$E[k] = np = n \cdot \dfrac{M}{N}$ — математическое ожидание;

$D[k] = npq \dfrac{N - n}{N - 1} = n \cdot \dfrac{M}{N}\left(1 - \dfrac{M}{N}\right)\dfrac{N - n}{N - 1}$ — дисперсия.

Дисперсия гипергеометрического распределения ограничена сверху величиной $\dfrac{n}{4} \ge npq \ge npq \dfrac{N - n}{N - 1} = D[k]$.

Решение примера 7.

Первый начальный момент гипергеометрического распределения — это математическое ожидание

$$E[X] = n\frac{M}{N} \Leftrightarrow \frac{M}{N} = \frac{E[X]}{n}.$$

Второй центральный момент — это дисперсия

$$D[X] = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1} = E[X]\left(1-\frac{E[X]}{n}\right)\frac{N-n}{N-1} \Leftrightarrow$$

$$\Leftrightarrow (N-1)D[X] = E[X]\left(1-\frac{E[X]}{n}\right)(N-n) \Leftrightarrow$$

$$\Leftrightarrow N\left(D[X] - E[X]\left(1-\frac{E[X]}{n}\right)\right) = D[X] - nE[X]\left(1-\frac{E[X]}{n}\right) \Leftrightarrow$$

$$\Leftrightarrow N = \frac{D[X] - nE[X]\left(1-\frac{E[X]}{n}\right)}{\left(D[X] - E[X]\left(1-\frac{E[X]}{n}\right)\right)} = \frac{D[X] - nE[X] + E^2[X]}{\left(D[X] - E[X] + \frac{E^2[X]}{n}\right)} \Rightarrow M = \frac{E[X]}{n}N$$

Приравниваем теоретические моменты эмпирическим оценкам $E[X] = \hat{E}[X] = \bar{x}$ и $D[X] = \hat{D}[X] = S^2$, подставляем в полученные выше формулы и получаем искомые оценки.

Итак, $\hat{N} = \dfrac{S^2 - n\bar{x} + (\bar{x})^2}{S^2 - \bar{x} + \dfrac{(\bar{x})^2}{n}} = n\dfrac{\bar{x}(n-\bar{x}) - S^2}{\bar{x}(n-\bar{x}) - nS^2}$; $\hat{M} = \dfrac{\bar{x}}{n}\hat{N}$

Оценка математического ожидания $\hat{E}[X] = \bar{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} =$

$$= \frac{2\cdot 0 + 3\cdot 1 + 10\cdot 2 + 22\cdot 3 + 26\cdot 4 + 20\cdot 5 + 12\cdot 6 + 5\cdot 7}{2+3+10+22+26+20+12+5} = 4 \Rightarrow \frac{M}{N} = \frac{\bar{x}}{n} = \frac{4}{7}$$

Дисперсия $D[X] = n\dfrac{M}{N}\dfrac{N-M}{N}\dfrac{N-n}{N-1}$

Второй начальный выборочный момент $\overline{x^2} = \dfrac{\sum_{i=1}^{k} n_i x_i^2}{\sum_{i=1}^{k} n_i} =$

$$= \frac{2\cdot 0^2 + 3\cdot 1^2 + 10\cdot 2^2 + 22\cdot 3^2 + 26\cdot 4^2 + 20\cdot 5^2 + 12\cdot 6^2 + 5\cdot 7^2}{2+3+10+22+26+20+12+5} = \frac{1834}{100} = 18.34$$

Выборочная дисперсия $\tilde{S}^2 = \overline{x^2} - (\bar{x})^2 = 18.34 - 4^2 = 2.34$

Оценка дисперсии $\hat{D}[X] = S^2 = \tilde{S}^2\dfrac{100}{99} = \dfrac{234}{99} = \dfrac{26}{11} = 2.363\,636\,36$

$$D[X] = \hat{D}[X] \Rightarrow n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1} = \frac{26}{11} \Rightarrow 4\cdot\frac{3}{7}\cdot\frac{N-7}{N-1} = \frac{26}{11} \Leftrightarrow$$

$$\Leftrightarrow 66(N-7) = 7\cdot 13(N-1) \Leftrightarrow (91-66)N = 7\cdot 13 - 66\cdot 7 \Leftrightarrow 25N = -371.$$

Результат получился бессмысленным, т.к. исходное предположение оказалось неверным. Для гипергеометрического распределения дисперсия не может превосходить $\dfrac{n}{4} = 1.75$.

По-видимому, в реальных условиях исследователю пришлось бы подбирать заведомо другой закон распределения, с которым наблюдаемый статистический ряд был бы лучше согласован.