УДК 519.2

Об одном подходе к проблеме экологического прогнозирования на основе модификаций метода максимума статистической энтропии

И. А. Соловьев, И. М. Петрушко*, М. И.Петрушко

Проблема планирования показателей различных характеристик в экологии, технике, технологиях, экономике и т.д. заключается в их информационной достоверности. На практическую реализацию плановых показателей влияет множество случайных факторов. В настоящей работе предложены новые подходы к использованию принципа максимума статистической энтропии к предсказанию исполнения планируемых показателей и информационной достоверности в применении к прогнозированию.

Ключевые слова: вероятность, статистическая энтропия, максимум энтропии, функция Лагранжа, решение систем уравнений, прогнозирование.

Опыт проведения прогнозных исследований в различных областях общественной жизни, науки и техники позволил выявить ряд методов, которые могут эффективно применяться для прогнозирования развития экологической ситуации. Любая типовая методика прогнозирования включает такие необходимые элементы, как выполнение предпрогнозной ориентации (определение предмета, целей, задач и периода упреждения); создание предпрогнозного фона (сбор и анализ данных в интервале ретроспекции); формирование исходной базовой и конструирование поисковой моделей, их верификация, а при необходимости уточнение (корректировка), подготовка, обоснование и принятие необходимых решений.

Поскольку узловым этапом является построение модели прогноза, то известные методы прогнозирования удобно классифицировать, разделив их на три группы:

1.Эвристические.

2. Прогнозных моделей.

3. Статистические.

К статистическим относятся методы, основу которых составляет формирование стохастических моделей прогнозирования. Предпосылкой их применения является наличие необходимых статистических данных, характеризующих период ретроспекции, и сведений, необходимых для определения модели прогноза. Широкое применение в прогнозировании статистических методов объясняется тем, что предметом статистики служит изучение методов выявления закономерностей массовых процессов.

Развитый математический аппарат и накопленный опыт применения делают привлекательным обращение в решаемой проблеме к статистическим прогнозным методам и моделям.

В настоящей работе предложены модификации одного из статистических прогнозных методов, основанного на принципе максимума энтропии. Этот подход дает возможность определять вероятность прогнозов, а цифры и их количество лучше говорят, чем интуитивные или прагматические предсказания. Разбор предлагается провести на основе решения двух проблем: прогнозирования показателей урожая и его стоимости и прогнозирования показателей добычи нефти.

Рассмотрим вначале проблему планирования урожая.

* petrushkoim@mpei.ru

В масштабах всей страны, в различных регионах, в каждом отдельном хозяйстве ежегодно планируется урожай различных сельскохозяйственных культур, их стоимость и общая предопределяемая выручка за произведенную продукцию. Также планируется возможная стоимость земельных участков, как частного владения, так и юридических организаций. Каждый руководитель планирует ожидаемые результаты своих предположений, исходя из разного рода качества и объема статистики прежних лет и интуитивного представления о возможности влияния случайных погодных явлений: наводнений, пожаров, нашествия вредителей и других неопределенных внешних воздействий. При наблюдаемых в настоящее время природных катаклизмах, когда наводнения, засуха, землетрясения, извержения вулканов, а также политические ситуации и другие явления не позволяют точно прогнозировать объем урожая и себестоимость произведенной сельскохозяйственной продукции, а также стоимость земельных участков, есть необходимость вероятностного подхода к оценкам такого рода влияния случайных явлений. Именно в этом и состоит необходимость применения стохастических математических моделей для прогнозирования вероятностей ожидаемых величин. Один из таких методов определения вероятности ожидаемых результатов основан на принципе максимума статистической энтропии. Он позволяет с учетом смысла энтропии как среднего значения информации оценить вероятности прогноза. Существует много инструментов, позволяющих оценивать как объемы, так и стоимость произведенной продукции. В большинстве случаев такие прогнозы строятся на основе многолетнего опыта прагматических оценок. Эти оценки носят слишком приблизительный характер и не позволяют делать математически обоснованные выводы. В настоящей работе предлагается один из вариантов произведения таких оценок, основанный на принципе максимума энтропии. Этот подход дает возможность определять вероятность прогнозов, а цифры и их количества лучше говорят, чем интуитивные или прагматические предсказания.

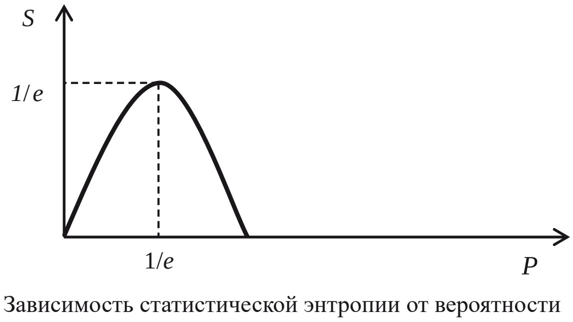
Настоящий анализ основан на классических работах Шеннона [1], который ввел понятие статистической энтропии, и А. К. Прица [4], который, изучая возрастную структуру популяции рыб, видел большое значение подобных исследований для рыболовного промысла (предсказание будущих уловов и предотвращение переуловов). В качестве статистической характеристике состояния популяции, по которой можно судить об изменениях возрастной структуры, А. К. Приц использовал энтропию S в следующем виде: $S = \sum_{k=1}^{n} (x_k / N)\ln(x_k / N)$, где $\sum_{k=1}^{n} x_k = N$. Здесь N — общая численности популяции; x_k — численность популяций рыб возраста k лет, а n — предельный возраст особей популяции. Вклад каждой возрастной группы в общую численность представляет собой относительную частоту x_k/N встретить особь k-го возраста среди N особей популяции. Фиксированная масса популяции определяется выражением $M = -\sum_{k=1}^{n} m_k x_k$, где m_k — масса одной особи k-го возраста.

Цель применения метода максимума статистической энтропии состоит в нахождении оптимальных вероятностных характеристик планов и определении стратегии долгосрочного планирования.

Основные результаты

Статистическая энтропия в предлагаемом варианте обозначается так: $S = -\sum_{k=1}^{n} p_k \ln p_k$, где p_k — вероятность наступления случайной дискретной случайной величины x_k, или так: $S = -\sum_{k=1}^{n} (x_k / N)\ln(x_k / N)$, где x_k — планируемый показатель, а N — общий объем плана. Рассматриваемая задача о максимуме статистической энтропии оценок объема урожая и стоимости произведенной сельскохозяйственной продукции, оценки земельных участков или экологических показателей региона подобна задаче о вероятностных оценках абсолютной погрешности функции нескольких переменных. Поскольку все погрешности аргументов x_k, имеющие смысл объема или стоимости произведенной сельскохозяйственной продукции или экологических показателей, носят случайный характер, то в стохастической постановке задается ряд распределения (x_k, p_k), $k = 1, ..., n$. Требуется при известном априори среднем значении $\overline{X} = p_1x_1 + p_2x_2 + ... + p_nx_n = \text{const}$ стоимости произведенного урожая региона найти значения вероятностей прогноза этих величин (p_k, $k = 1, ..., n$) из условия максимума статистической энтропии $S = \sum_{k=1}^{n} p_k \ln p_k \to \max$, когда выполняется условие нормировки вероятности на единицу $\sum_{k=1}^{n} p_k = 1$.

Статистическая энтропия обладает важным экстремальным свойством. В случае одной переменной $S(p) = -p\ln p$ максимум $S(p) = S_{\max} = 1/e \approx 1/2{,}71$ достигается при $p = 1/e \approx 1/2{,}71$. Графически зависимость $S(p)$ представлена на рисунке.



Зависимость статистической энтропии от вероятности

Смысл ее заключается в том, что наибольшая информационная составляющая случайного явления достигается при $p = 1/e \approx 1/2{,}71$ и стремится к нулю либо при $p \to 0$, либо при $p \to 1$. Вывод: для планирования каких-либо показателей только вдали от $p \approx 0$ или $p \approx 0$ будут получаться наиболее объективные оценки вероятностных прогнозов. Повторить аналогичный результат в случае многих переменных, когда $S(p_1, p_2 ..., p_n) = \sum_{k=1}^{n} p_k \ln p_k \to \max$ не получается. Действительно, необходимое условие экстремума $\partial S(p_1, p_2 ..., p_n) \partial p_k = -\ln p_k - 1 = 0$ приводят к решению $p_k = 1/e$, $S_{\max} = n/e$, но условие нормировки $\sum_{k=1}^{n} p_k = 1$ не соблюдается. Отсюда можно сделать вывод: для некоторой ситуации «выброс» экстремальной вероятности $p_k = 1/e$ возможен в отдельных хозяйствах в конкуренции со всеми остальными. С другой стороны, представляя энтропию в виде $S(p_1, p_2 ..., p_n) = \ln \prod_{k=1}^{n} p_k^{p_k}$, придем к известной задаче о максимуме произведения при заданной сумме $\sum_{k=1}^{n} p_k = 1$. Ее решение имеет вид $S_{\max}(p_1, p_2 ..., p_n) = \ln n,\ p_k = 1/n,\ k = 1, 2, ..., n.$

Другое важное свойство дифференциальной энтропии, которое мы назовем принципом симметрии, состоит в том, что для двух рядов распределения с одинаковыми плановыми показателями $x_1 < x_2 < ... < x_n$, но в первом случае с возрастающими (x_1, p_1), (x_2, p_2), (x_n, p_n), $p_1 < p_2 < ... < p_n$, а во втором случае с убывающими вероятностями (x_1, p_n), (x_2, p_{n-1}), (x_n, p_1), $p_1 < p_2 < ... < p_n$ будут получаться одинаковые значения статистической энтропии

$$S_{1\max} = -p_1 \ln p_1 - p_2 \ln p_2 - p_n \ln p_n =$$
$$= S_{2\min} - p_n \ln p_n - p_{n-1} \ln p_{n-1} - p_1 \ln p_{1n}.$$

При этом в первом случае среднее значение плановых показателей максимально $\bar{X}_{\max} = x_1p_1 + x_2p_2 + ... + x_n p_n$, а во втором минимально $\bar{X}_{\min} = x_1p_n + x_2 p_{n-2} + ... + x_n p_1$. Все остальные случаи с такими же наборами плановых показателей и их вероятностей находятся в промежутках между максимальным и минимальным значениями статистической энтропии: $S_{2\min} < S < S_{\max}$.

Возникает вопрос, а при каком априорном задании среднего значения плановых показателей будет обеспечена наибольшая энтропия, иными словами, какому плану можно доверять более всего в его информационном прогнозировании? Чтобы ответить на него, приведем следующие рассуждения.

Рассмотрим некую задачу плана региона с двумя хозяйствующими субъектами. Требуется найти максимальную статистическую энтропию

$$S = -p_1 \ln p_1 - p_2 \ln p_2 \qquad (1)$$

и вероятности p_1 и p_2 плановых показателей x_1 и x_2 при априори заданном среднем значении плана $\bar{X}$

$$p_1x_1 + p_2x_2 = \bar{X} \qquad (2)$$

и соблюдении условия нормировки

$$p_1 + p_2 = 1. \qquad (3)$$

Решение системы (1) — (3) можно получить методом неопределенных множителей Лагранжа [2].

Введем функцию Лагранжа:

$$L(p_1, p_2, \lambda_1, \lambda_2) = \sum_{k=1}^{2} p_k \ln p_k +$$
$$+\lambda_1\left(\sum_{k=1}^{2} p_k x_2 - \bar{X}\right) + \lambda_2\left(\sum_{k=1}^{2} p_k - 1\right) \to \max.$$

Решение задачи (1) — (3) имеет следующий вид:

$$p_1 = (x_2 - \bar{X})/(x_2 - x_1);$$
$$p_2 = (\bar{X} - x_1)/(x_2 - x_1); \qquad (4)$$

$$\lambda_1 = \ln(p_1/p_2)/(x_1 - x_2);$$
$$\lambda_2 = x_1 \ln(p_1/p_2)/(x_1 - x_2) + \ln p_1 + 1. \qquad (5)$$

Энтропия в этом случае получается такой:

$$S = -(x_2 - \bar{X})/(x_2 - x_1) \ln[(x_2 - \bar{X})/(x_2 - x_1)] -$$
$$-(\bar{X} - x_1)/(x_2 - x_1) \ln[(\bar{X} - x_1)/(x_2 - x_1)]. \qquad (6)$$

Справедливо следующее утверждение.

Утверждение 1. Максимум максимальной статистической энтропии (6) достигается при среднем значении равном среднему арифметическому плановых показателей $\bar{X} = (x_1 + x_2)/2$ и вероятностях $p_1 = p_2 = 0{,}5$.

Доказательство. Воспользуемся необходимым условием экстремума

$$S_{\bar{X}}^{/} = (1/(x_2 - x_1)) \ln[(x_2 - \bar{X})/(\bar{X} - x_1)] = 0.$$

Отсюда $(x_2 - \bar{X})/(\bar{X} - x_1) = 1$ и $\bar{X} = (x_1 + x_2)/2$ есть среднее арифметическое плановых показателей. Воспользуемся уравнением для среднего значения: $p_1x_1 + p_2x_2 = 0{,}5\,(x_1 + x_2)$. Из него получим $(p_1 - 0{,}5) + (p_2 - 0{,}5)x_2 = 0$, отсюда следует, что $p_1 = p_2 = 0{,}5$.

Вывод. При задании среднего значения плана в виде среднего арифметического значения плановых показателей будет достигаться максимум статистической энтропии, то есть наибольшие информационная достоверность плана и вероятность его реального исполнения.

Рассмотрим три конкретных случая планирования в двух хозяйственных субъектах. В первом случае стоимость первого земельного участка в условных единицах (это могут быть сотни тысяч или миллионы рублей) $x_1 = 1$, а второго — в пять раз больше: $x_2 = 5$. Пусть априори заданное среднее значение планируемых выручек $\bar{X} = 1{,}4$. Вероятности при этом, получающиеся из формул (4), таковы $p_1 = 0{,}9$, $p_2 = 0{,}1$. Энтропия при этом $S_1 = -0{,}9 \cdot \ln 0{,}9 - 0{,}1 \cdot \ln 0{,}1 \approx 0{,}329$. Во втором случае $x_1 = 1$

и $x_1 = 5$, а априори заданное среднее значение $\bar{X} = 4{,}6$. Вероятности получаются следующие $p_1 = 0{,}1$, $p_2 = 0{,}9$. Удивительным представляется результат сравнения энтропии первого случая S_1 с энтропией второго случая $S_2 = -0{,}1\cdot\ln 0{,}1 - 0{,}9\cdot\ln 0{,}9 \approx 0{,}329$. Они одинаковы, что соответствует описанному выше принципу симметрии статистической энтропии и эта ситуация предсказуема с учетом изменения статистической энтропии, которая представлена на рисунке. В третьем случае зададим априорно задаваемое среднее значение плановых показателей $\bar{X} = (1 + 5)/2 = 3$ как среднее арифметическое. Найдем в этом случае вероятности и энтропию из формул (4) — (5). Получаются предсказанные выше результаты: вероятности близки к 0,5: $p_1 = 0{,}5$, $p_2 = 0{,}5$, а статистическая энтропия $S_3 = -0{,}5\cdot\ln 0{,}5 - 0{,}5\cdot\ln 0{,}5 \approx 0{,}69$ имеет наибольшее значение по сравнению с предыдущими вариантами, то есть, наибольшая достоверность плана достигается, во-первых, при задании среднего значения вблизи среднего арифметического значения планируемых показателей, а во вторых, при вероятностях, равных 0,5.

Рассмотрим теперь задачу о прогнозировании плановых показателей урожайности зерновых культур x_1 и x_2, находимых из принципа максимума статистической энтропии, которая записывается с помощью относительных частот x_1/N и x_2/N, где $N = x_1 + x_2$.

Обозначим через x_1 и x_2 — предполагаемые (искомые) объемы урожая соответственно пшеницы 3-го и 4-го классов в одной области. За $N = x_1 + x_2$ примем заданный руководителем области будущий общий объем урожая пшеницы 3-го и 4-го классов в тысячах тонн, за $\bar{X} = m_1x_1 + m_2x_2$ — средний объем выручки урожая в этой области (в тысячах рублей) по результатам прошлых лет, где себестоимости одной тонны пшеницы 3-го класса — m_1 и 4-го класса — m_2 рублей за тонну. Тогда постановка задачи примет следующий вид.

Найдем максимум статистической энтропии:

$$S = -(x_1 / N)\ln(x_1 / N) - (x_2 / N)\ln(x_2 / N) \to \max \quad (7)$$

и величины x_1 и x_2 при условии выполнения заданного объема урожая:

$$x_1 + x_2 = N, \quad (8)$$

когда априори определена общая стоимость произведенной пшеницы 3-го и 4-го классов:

$$m_1x_1 + m_2x_2 = \bar{X}. \quad (9)$$

Решение задачи, полученное методом неопределенных множителей Лагранжа, имеет вид:

$$\begin{aligned} x_1 &= (Nm_2 - \bar{X}) / (m_2 - m_1); \\ x_2 &= (\bar{X} - Nm_1) / (m_2 - m_1); \end{aligned} \quad (10)$$

$$\begin{aligned} S_1 = &-((Nm_2 - \bar{X}) / (m_2 - m_1))\ln[(Nm_2 - \bar{X}) / (m_2 - m_1)] - \\ &-((\bar{X} - Nm_1) / (m_2 - m_1))\ln[(\bar{X} - Nm_1) / (m_2 - m_1)]. \end{aligned} \quad (11)$$

Довольно интересен результат анализа полученных формул. В случае отрицательности одной из величин $Nm_2 - \bar{X}$ или $\bar{X} - Nm_1$ можно получить отрицательные значения прогнозируемых показателей плана, что будет свидетельствовать о неправильных оценках задаваемых величин N или $\bar{X}$. В особенности это касается стоимости предполагаемого объема урожая $\bar{X}$. Проведем статистический анализ следующих данных. В качестве N гипотетически выберем заданную руководителем хозяйства в планируемом году величину $N \approx 4$ тыс. тонн. Представляющейся реальной суммарной величиной средней плановой выручки от производства пшеницы 3-го и 4-го классов определим как $\bar{X} \approx 4{,}5$ тыс. рублей. По некоторым статистическим данным Роскомстата прошлых лет были приведены следующие себестоимости одной тонны пшеницы: 3-го класса — $m_1 = 3$ тыс. рублей за тонну и 4-го класса — $m_1 = 2{,}5$ тыс. рублей. Формулы (10) — (11) приводят к следующим результатам:

$$x_1 = (Nm_2 - \bar{X}) / (m_2 - m_1) \approx 1{,}5, \text{ тыс. тонн;}$$
$$x_2 = (\bar{X} - Nm_1) / (m_2 - m_1) \approx 2{,}5 \text{ тыс. тонн.}$$

Заметим, что в случае всего двух плановых показателей задача может быть решена проще, без привлечения метода неопределенных множителей Лагранжа. В самом деле, из системы только двух уравнений (2), (3) можно получить вероятности, приведенные в формулах (4) и из них найти статистическую энтропию (6). Такое простое решение получается в случае только двух переменных. Если их число больше двух, то приходится применять итерационный метод решения системы уравнений, получающихся из необходимого условия экстремума функции Лагранжа:

$$L(p_1, p_2, ..., p_n, \lambda_1, \lambda_2) = \sum_{k=1}^{n} p_k \ln p_k +$$
$$+\lambda_1(\sum_{k=1}^{n} p_k x_k - M_z^{(1)}) + \lambda_2(\sum_{k=1}^{n} p_k - 1).$$

Вероятности и множители Лагранжа λ_1, λ_2 находятся из системы уравнений, выражающих необходимое условие экстремума:

$$\begin{cases} \partial L / \partial p_k = \ln p_k + 1 + \lambda_1 x_k + \lambda_2 = 0\ , k = 1....n; & (12) \\ \partial L / \partial \lambda_1 = \sum_{k=1}^{n} p_k x_k - \bar{X} = 0; & (13) \\ \partial L / \partial \lambda_2 = \sum_{k=1}^{n} p_k - 1 = 0. & (14) \end{cases}$$

Сформулируем и докажем следующее утверждение.

Утверждение 2. Максимальная энтропия связана с неопределенными множителями Лагранжа и средним значение, следующим образом $S_{\max} = 1 - \lambda_1\bar{X} - \lambda_2$.

Доказательство. Умножим каждое из уравнений (12) на соответствующее значение p_k и сложим все полученные уравнения, тогда:

$$S_{\max} = 1 - \lambda_1\bar{X} - \lambda_2. \tag{15}$$

Казалось бы, максимальная энтропия линейно зависит от среднего значения $\bar{X}$. Однако это не так, поскольку λ_1 и λ_2 нелинейно связаны с $\bar{X}$. Продемонстрируем это, применив предлагаемые в настоящей статье методы суммирования, основанные на интегрировании и дифференцировании функций, соответствующих конечным суммам. Выразим p_k из уравнения (7) через λ_1 и λ_2. Имеем:

$$p_k = \exp(\lambda_2 - 1)\exp(\lambda_1 x_k),\ k = 1, 2, ..., n. \tag{16}$$

Подставим значения p_k в уравнение для среднего значения (13). Получим:

$$\exp(\lambda_2 - 1)\sum_{k=1}^{n}\exp(\lambda_1 x_k)x_k = \bar{X}. \tag{17}$$

Введем функцию

$$f(t) = \exp(\lambda_2 - 1)\sum_{k=1}^{n}\exp(\lambda_1 t)x_k - \bar{X} = 0. \tag{18}$$

Проинтегрируем $f(t)$ в пределах от 0 до t:

$$\int_0^{\lambda_1} f(t)dt = \exp(\lambda_2 - 1)\sum_{k=1}^{n}\int_0^{\lambda_1}\exp(\lambda_1 t)x_k dt - \int_0^{\lambda_1}\bar{X}dt = 0.$$

Получим:

$$\exp(\lambda_2 - 1)\sum_{k=1}^{n}\exp(\lambda_1 x_k) - \exp(\lambda_2 - 1)n = \bar{X}\lambda_1. \tag{19}$$

С учетом условия нормировки (14)

$$1 - n\exp(\lambda_2 - 1) = \bar{X}\lambda_1.$$

Отсюда можно выразить либо λ_2 через λ_1

$$\lambda_2 = 1 + \ln((1 - \bar{X}\lambda_1)/n), \tag{20}$$

либо λ_1 через λ_2:

$$\lambda_1 = (1 - n\exp(\lambda_2 - 1))\,\bar{X}. \tag{21}$$

Из условия нормировки (14) получим нелинейное уравнение для λ_1. Для этого введем функцию

$$F(z) = \exp(\lambda_2(z) - 1)\sum_{k=1}^{n}\exp(zx_k) - 1 = 0.$$

Продифференцируем $F(z)$:

$$dF(z)/dz = \exp(\lambda_2(z) - 1)\times$$
$$\times\sum_{k=1}^{n}\exp(zx_k)d\lambda_2/dz + \exp(\lambda_2(z) - 1)\sum_{k=1}^{n}\exp(zx_k)x_k = 0.$$

Отсюда

$$d\lambda_2/dz = -(\exp(\lambda_2(z) - 1)\times$$
$$\times\sum_{k=1}^{n}\times\exp(zx_k)x_k)/(\exp(\lambda_2(z) - 1)\sum_{k=1}^{n}\exp(zx_k)). \tag{22}$$

При $z = \lambda_1$ выражение (17) с учетом уравнений (13), (14) приобрете вид:

$$d\lambda_2/d\lambda_1 = -\bar{X}. \tag{23}$$

Начальное условие для λ_1 получим из уравнения (15):

$$\lambda_2(0) = 1 - S_{\max}. \tag{24}$$

После интегрирования задачи Коши для уравнений (23) — (24), получим следующее решение:

$$\lambda_2(\lambda_1) = -\bar{X}\lambda_1 + 1 - S_{\max}. \tag{25}$$

Подставив (25) в (21) получим нелинейное уравнение относительно λ_1:

$$\lambda_1 = (1 - n\exp(-\bar{X}\lambda_1 - S_{\max}))/\bar{X}. \tag{26}$$

Таким образом, λ_1, λ_2 и $S_{\max}$ можно найти из нелинейной системы уравнений (15, (25), (26), которая не содержит вероятности p_k. Эта система может быть решена, например, методом Ньютона [3].

Проанализируем полученное решение. Из уравнения (15), так как выполняется условие неотрицательности энтропии, следует неравенство $1 - \lambda_1\bar{X} - \lambda_2 \geq 0$. Отсюда $\lambda_1 \leq (1 - \lambda_2)/\bar{X}$. Если $(1 - \lambda_2) < 0$, то $\lambda_1 \leq 0$ и из уравнений (16) следует, что чем больше величина планового показателя x_k, тем меньше вероятность ее реального исполнения, т. е, тем больше риска в предсказуемости результата реализации плана.

В терминах относительных частот аналог задачи (12) — (14) об определении производства зерновой культуры разного класса или множества разных зерновых культур в многомерном случае формулируется следующим образом. Найти объемы производства зерна отдельных категорий x_k, $k = 1, 2, ..., n$ из условия максимума статистической энтропии

$$S(x_1, x_2, ..., x_n) = -\sum_{k=1}^{n}(x_k/N)\ln(x_k/N) \to \max \tag{27}$$

при известном значении общего объема урожая N

$$\sum_{k=1}^{n}x_k = N, \tag{28}$$

известных себестоимостях отдельных объемов продукции m_i и общей стоимости всего произведенного урожая M

$$\sum_{k=1}^{n}m_k x_k = M. \tag{29}$$

После введения функции Лагранжа

$$L(x_1, x_2, ..., x_n, \lambda_1, \lambda_2) = -\sum_{k=1}^{n}(x_k/N)\ln(x_k/N) +$$
$$+\lambda_1\left(\sum_{k=1}^{n}m_k x_k - M\right) + \lambda_2\left(\sum_{k=1}^{n}x_k - N\right)$$

и применения необходимого условия экстремума решение находится из системы уравнений

$$-(1/N)\ln(x_k/N) - (1/N) + \lambda_1 m_k + \lambda_2 = 0;$$
$$\sum_{k=1}^{n}x_k = N;$$
$$\sum_{k=1}^{n}m_k x_k = M.$$

Заметим, что можно доказать утверждение, аналогичное утверждению 2.

Другой подход к методу максимума статистической энтропии основан на замене формулы (2), учитывающей математическое ожидание случайных величин, на формулу (31), учитывающую их дисперсию.

Предложим новый вариант реализации классической постановки задачи о максимуме статистической энтропии.

Найдем максимум статистической энтропии

$$S = -\sum_{k=1}^{n} p_k \ln p_k, \qquad (30)$$

когда априори задается дисперсия σ^2 случайной величины X

$$\sum_{k=1}^{n} p_k x_k^2 - (\sum_{k=1}^{n} p_k x_k)^2 = \sigma^2 \qquad (31)$$

и соблюдается условие нормировки

$$\sum_{k=1}^{n} p_k = 1. \qquad (32)$$

Возникает два вопроса:

1) из каких соображений задаются среднее значение и дисперсия;

2) при каких сочетаниях в этих двух постановках получатся одинаковые вероятности.

Рассмотрим случай подобной случайной величины: (x_1, p_1), (x_2, p_2). Здесь задача нахождения условного экстремума для статистической энтропии $S = -p_1\ln p_1 - p_2\ln p_2$ удивительно сводится к решению двух уравнений с двумя неизвестными без привлечения метода неопределенных множителей Лагранжа. В первом классическом случае априори задается среднее значение

$$p_1x_1 + p_2x_2 = \bar{X}. \qquad (33)$$

Во втором случае считается известной дисперсия

$$p_1x_1^2 + p_2x_2^2 - (p_1x_1 + p_2x_2)^2 = \sigma^2. \qquad (34)$$

Естественно и в первом и во втором случаях должно выполняться условие нормировки $p_1 + p_2 = 1$

Решение первой задачи, когда считается известным среднее значение, имеет следующий вид.

$$p_1 = (x_2 - \bar{X})/(x_2 - x_1);\ \ p_1 = (x_2 - \bar{X})/(x_2 - x_1); \qquad (35)$$

Если бы задача решалась методом неопределенных множителей Лагранжа, то они имели бы такой вид:

$$\lambda_1 = \ln(p_1/p_2)/(x_1 - x_2);\ \lambda_2 = x_1\ln(p_1/p_2)/(x_1 - x_2) + \ln p_1 + 1. \qquad (36)$$

Энтропия в этом случае записывается так:

$$\begin{aligned} S = -(x_2 - \bar{X})/(x_2 - x_1)\ln[(x_2 - \bar{X})/(x_2 - x_1)] - \\ -(\bar{X} - x_1)/(x_2 - x_1)\ln[(\bar{X} - x_1)/(x_2 - x_1)]. \end{aligned} \qquad (37)$$

Решение задачи с априори заданной дисперсией имеет поиска вероятностей прогнозируемых случайных плановых величин:

$$p_{1,2} = (1 \pm \sqrt{(1 + 4(x_2^2 - \sigma^2)/(x_1 - x_2)^2}\,)/2,\ p_2 = 1 - p_1. \qquad (38)$$

Энтропия при заданном значении дисперсии имеет вид:

$$S = -p_1 \ln p_1 - (1 - p_1)\ln(1 - p_1). \qquad (39)$$

Для рассматриваемого варианта, когда дискретная случайная величина имеет всего два значения, ответ yf этот вопрос может быть таким. Задание среднего значения $\bar{X}$ должно удовлетворять условию $\bar{X} < x_1 + x_2$. Лучше всего выбирать средним значением $\bar{X} = 0{,}5(x_1 + x_2)$ при этом вероятности $p_1 = p_2 = 0{,}5$, однако не всегда такие вероятности реально сочетаются с относительной частотой и соответствуют этой желанной действительности. Другой вопрос. Почему лучше ставить задачу для максимума статистической энтропии на основе заданности среднего квадратического отклонения? Дело в том, что априори заданное среднее квадратическое отклонение σ, выбираемое, например, так $2\sigma/(x_1 + x_2) \approx 0{,}05 - 0{,}10$ часто наблюдается в физических экспериментах. Такое значение для σ более реально, чем выбираемое произвольно среднее значение $\bar{X}$.

Ответ на второй вопрос представляется таким. Из уравнений $p_1x_1 + p_2x_2 = \bar{X}$, $p_1x_1^2 + p_2x_2^2 = \sigma^2 - (\bar{X})^2$ получаются следующие условия равенства вероятностей двух приведенных задач, одной на основе заданности среднего значения, другой — при априорной определенности среднего квадратического отклонения:

$$p_1 = (M[X]x_2^2 - x_2(\sigma^2 - (M[X])^2))/(x_1x_2^2 - x_2x_1^2);$$
$$p_2 = (x_1(\sigma^2 - (M[X])^2) - (M[X]x_1^2))/(x_1x_2^2 - x_2x_1^2).$$

При этом должно соблюдаться условие нормировки $p_1 + p_2 = 1$.

Рассмотрим теперь задачу о прогнозировании плановых показателей x_1 и x_2, находимых из принципа максимума статистической энтропии, которая записывается с помощью относительных частот x_1/N и x_2/N, где $N = x_1 + x_2$, когда задается средняя стоимость продукции $m_1x_1 + m_2x_2 = \bar{X}$ — средний объем выручки по результатам предыдущего временного периода, где себестоимости плановых показателей x_1 и x_2 соответственно равны m_1 и m_2 за единицу каждой величины. Постановка задачи имеет следующий вид. Найти максимум статистической энтропии

$$S = -(x_1/N)\ln(x_1/N) - (x_2/N)\ln(x_2/N) \to \max \qquad (40)$$

и величины x_1 и x_2 при условии выполнения заданного объема

$$x_1 + x_2 = N, \qquad (41)$$

когда априори определена общая стоимость $\bar{X}$ произведенной продукции первой и второй планируемой характеристики

$$m_1x_1 + m_2x_2 = \bar{X}. \qquad (42)$$

Решение задачи (40) — (42) выглядит как

$$x_1 = (Nm_2 - \bar{X})/(m_2 - m_1);\ x_2 = (\bar{X} - Nm_1)/(m_2 - m_1). \quad (43)$$

$$S_1 = -((Nm_2 - \bar{X})/(m_2 - m_1))\ln[(Nm_2 - \bar{X})/(m_2 - m_1)] - \\ - ((\bar{X} - Nm_1)/(m_2 - m_1))\ln[(\bar{X} - Nm_1)/(m_2 - m_1)]. \quad (44)$$

Постановка второй задачи при априорно задаваемом значении σ среднего квадратического отклонения имеет такой вид. Найдем x_1 и x_2 из условия максимума статистической энтропии

$$S = -(x_1 / N)\ln(x_1 / N) - (x_2 / N)\ln(x_2 / N) \quad (45)$$

при соблюдении условия известности дисперсии σ^2

$$(x_1 / N)m_1^2 + (x_2 / N)m_2^2 - \\ -((x_1 / N)m_1 + (x_2 / N)m_2)^2 = \sigma^2 \quad (46)$$

и выполнения равенства, аналогичного равенству нормировки,

$$x_1 + x_2 = N, \quad (47)$$

Решение этой задачи таково.

$$x_1 = (Nm_2 - \bar{X})/(m_2 - m_1);\ x_2 = (\bar{X} - Nm_1)/(m_2 - m_1);$$

$$p_1 = ((x_1^2 - 2p_2x_1x_2) + \\ +\sqrt{(x_1^2 - 2p_2x_1x_2)^2 - 4x_1^2(p_2^2x_2^2 - p_2x_2^2 - \sigma^2)})/(2x_1^2). \quad (48)$$

Подставим в данное равенство вместо p_1 величину $1 - p_2$ и решим полученное уравнение относительно p_2, получим оптимальные вероятностные характеристики p_1 и p_2 при которых достигается максимум статистической энтропии:

$$p = \frac{1}{2} \pm \sqrt{\frac{1}{4} + \frac{\sigma^2}{(x_2 - x_1)^2}}$$

Заметим, что в случае всего двух плановых показателей задача легко решается, если же число переменных больше двух, приходится применять итерационные методы решения системы уравнений, получающихся из необходимого условия условного экстремума функции статистической энтропии

$$S = -\sum_{k=1}^{n} p_k \ln p_k \quad (49)$$

при условиях

$$\sum_{k=1}^{n} p_k x_k^2 - \left(\sum_{k=1}^{n} p_k x_k\right)^2 = \sigma^2, \quad (50)$$

$$\sum_{k=1}^{n} p_k = 1 \quad (51)$$

методом неопределенных множителей Лагранжа:

$$L(p_1, p_2, ..., p_n, \lambda_1, \lambda_2) = -\sum_{k=1}^{n} p_k \ln p_k + \\ +\lambda_1\left(\sum_{k=1}^{n} p_k x_k^2 - \left(\sum_{k=1}^{n} p_k x_k\right)^2 - \sigma^2\right) + \lambda_2\left(\sum_{k=1}^{n} p_k - 1\right).$$

Вероятности и множители Лагранжа λ_1, λ_2 находятся из системы уравнений, выражающих необходимое условие экстремума:

$$\begin{cases} \partial L / \partial p_k = -\ln p_k - 1 + \lambda_1(2p_kx_k - 2x_k) \times \\ \times \sum_{k=1}^{n} p_k x_k + \lambda_2 = 0\ ,\ k = 1....n; \\ \partial L / \partial \lambda_1 = \sum_{k=1}^{n} p_k x_k^2 - \left(\sum_{k=1}^{n} p_k x_k\right)^2 - \sigma^2 = 0; \\ \partial L / \partial \lambda_2 = \sum_{k=1}^{n} p_k - 1 = 0. \end{cases}$$

Практическая реализация принципа максимума статистической энтропии

Рассмотрим задачу прогнозирования показателей добычи нефти x_1 — x_4, находимых из принципа максимума статистической энтропии. Приведем таблицу добычи и стоимости различных сортов нефти в мире за период 2009 — 2013 гг. (добыча измеряется в миллионах баррелей в сутки, а цена — в долларах за баррель).

Производство и цена нефти в мире

Год	2009	2010	2011	2012	2013
Производство, всего	85,6	87,5	88,4	91,3	92,8
Цена нефти					
Brent	61,67	79,63	110,94	111,71	108,7
WTI	61,92	79,40	95,05	95,13	94,65
Urals	61,22	78,21	109,35	110,53	107,88
нефтяная корзина ОПЕК	61,76	77,38	107,44	106,31	105,87

В качестве исходных данных взяты стоимости барреля различных сортов нефти в 2011 — 2013 гг. при этом x_1 — x_3 — цены нефти марок Brent, WTI, Urals; x_4 — усредненная цена нефтяной корзины ОПЕК.

Используя метод неопределенных множителей Лагранжа для нахождения максимума статистической энтропии (24) получим
если примем добычу нефти всех сортов в 2011 г. за 100%, то оптимальный относительный объем добычи составит:

Год	2011	2012	2013
$\boldsymbol{P_1}$	0,304	0,327	0,276
$\boldsymbol{P_2}$	0,151	0,126	0,196
$\boldsymbol{P_3}$	0,284	0,306	0,271
$\boldsymbol{P_4}$	0,261	0,240	0,258
$\boldsymbol{S}$	1,356	1,332	1,378

Марка нефти	Год		
	2011	**2012**	**2013**
Brent	100%	111,43%	103,84%
WTI	100%	86,30%	135,58%
Urals	100%	111,80%	100,10%
Нефтяная корзина ОПЕК	100%	95,36%	103,56%

Таким образом, в условиях стабильной экономической и политической ситуации информационное среднее значение (статистическая энтропия) растет, что свидетельствует о здоровом росте экономики и производства нефти. С другой стороны изменения максимума статистической энтропии позволяет нам предсказывать и анализировать изменения в экономической и политической ситуациях.

Заключение

Предложенные модификации метода максимума статистической энтропии позволяют прогнозировать как медленно текущие случайные процессы, так и быстрые. Среднее значение информации (статистической энтропии) в будущем зависит от многих факторов, которые не могут быть использованы без учета внешних условий. Тем не менее, прогнозы нужно делать, иногда из субъективных оценок политологов и экономистов.

Один из основных выводов заключается в следующем: чем сильнее отличается априори задаваемое из прагматических оценок хозяйствующих субъектов среднее значение показателей $\bar{X} = p_1x_1 + p_2x_3 + ... + p_nx_n$ = const от среднего арифметического значения $\tilde{X} = (x_1 + x_2 + ... + x_n)/n$, тем меньше вероятность его истинной реализации.

Авторы считают также необходимым указать трудности использования предложенной методики. Самое сложное — обеспечить достаточный и достоверный объем исходных статистических данных При планировании вероятностных характеристик, обозначенных как x_k, продуктивным представляется строить статистические данные на базе результатов работ о глобальной и локальной цикличностях показателей роста или спада экономики.

Литература

1. **Shannon C.E.** Mathematical Theory of communication. // Bell System Tech. J. 1948. V. 27. P. 379 — 423.

2. **Соловьев И.А., Шевелев В.В., Червяков А.В., Репин А.Ю.** Практическое руководство к решению задач по высшей математике. Интегрирование функций одной переменной, функции многих переменных, ряды. СПб.: Изд-во «Лань», 2009.

3. **Соловьев И.А., Червяков А.В., Репин А.Ю.** Вычислительная математика на смартфонах, коммуникаторах и ноутбуках с использованием программных сред Python. СПб.: Изд-во «Лань», 2011.

4. **Приц А.К.** Принцип стационарных состояний открытых систем и динамика популяций. Калининград, 1974.

5. **Кондратьев Н.Д.** Большие циклы конъюнктуры и теория предвидения. Избранные труды. Изд-во Экономика, 2002.

Статья поступила в редакцию 30.03.2016