

ТЕОРИЯ ИНФОРМАЦИИ

УДК 621.396.96.01

Прокофьев В. П.**РАЗРАБОТКА МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ.
ИСПОЛЬЗОВАНИЕ МЕТОДОВ ТЕОРИИ ИНФОРМАЦИИ ПРИ
АППРОКСИМАЦИИ МНОГОМЕРНЫХ ПЛОТНОСТЕЙ
РАСПРЕДЕЛЕНИЯ НАБЛЮДЕНИЙ**

*Международная ассоциация ученых, специалистов и деятелей науки — «Наука»
04050, Киев-119, ул. Мельникова, 81
e-mail: vprokofev@scomplex.com.ua, necit@ukr.net*

При решении огромного количества задач в различных областях науки и техники используются сигналы, описываемые с помощью моделей, к которым предъявляются требования адекватности реальным сигналам. Широко известны модели на основе гауссовских случайных процессов. Однако в реальных информационных системах сигналы носят негауссовский характер и, потому, должны описываться негауссовскими законами распределения мгновенных значений. В настоящей работе дан анализ наиболее распространенных мер количества информации, введенных Фишером, Шенноном, Кульбаком и Лейбнером. При проверке математических гипотез применено отношение правдоподобия как более информативное. Результаты, полученные Кульбаком, имеют ограничения и недостатки, заключающиеся в том, что в исследованиях использовались одномерные величины, а искомая плотность распределения представлена в неявном виде, что затрудняет анализ и использование этих результатов. В работе проведены исследования, на основе метода множителей Лагранжа, в определенной степени устраняющие указанные ограничения и недостатки. Путем обобщения известных результатов для одномерных распределений получены новые результаты, относящиеся к многомерному случаю.

Ключевые слова: информационные системы, гауссовские случайные процессы, многомерные распределения, меры количества информации (Фишера, Шеннона, Кульбака-Лейбнера и др.), одномерные и многомерные модели сигналов.

Сигналы, как и другие процессы, характеризующие реальные явления, представляют собой довольно сложные функции времени. При их описании обращаются к некоторым моделям. Важность адекватности моделей и реальных сигналов очевидна. Неудачный выбор моделей приводит к малоэффективным алгоритмам обработки сигналов.

Общие требования к моделям сигналов хорошо известны. Модели должны хорошо отражать закономерности процессов, быть содержательными с точки зрения физики, допускать исследования структуры сигналов с приемлемой полнотой, быть гибкими при описании преобразований сигналов. Вместе с этим, модели сигналов должны быть достаточно простыми, чтобы исключить ненужное усложнение алгоритмов их обработки. Следовательно, выбор моделей сигналов связан с преодолением противоречивых требований к ним. Это, в принципе, возможно на основе стройных математических методов.

Процесс разработки моделей включает в себя следующие основные этапы: выявление ситуаций, встречающихся в практике, когда традиционно используемые модели оказываются не вполне пригодными; установление причин, порождающих такого рода ситуации; разработку моделей, исходя из общетеоретических положений и накопленного опыта; создание математического аппарата по использованию предлагаемых моделей.

При решении огромного количества задач широкое распространение получили модели на основе гауссовских случайных процессов. Гауссовские процессы довольно просто описываются и полностью определяются двумя первыми моментами. Однако практика показывает, что большинство реальных сигналов в радиотехнических, гидроакустических и других информационных системах представляет собой совокупность сигналов с негауссовскими законами распределения мгновенных значений.

Таким образом, реальные сигналы характеризуются довольно сложной внутренней структурой.

Сигналы, наблюдаемые на определенных временных интервалах, обычно обрабатываются либо как непрерывные процессы, либо как дискретные выборки.

Как правило, с целью уменьшения потерь информации, выборки осуществляют настолько часто, что их нельзя считать независимыми. Это означает, что при описании сигналов недостаточно опираться на использование одномерных плотностей распределения наблюдений, поэтому принципиально необходимо знание многомерных плотностей.

В математической статистике такого рода задачи не являются новыми и им всегда уделялось должное внимание. Применение методов теории информации для разработки многомерных статистических моделей ранее не использовались.

Информационное количество Кульбака-Лейблера. Аппроксимация одномерных плотностей распределения наблюдений

Наиболее распространенной мерой количества информации является мера, введенная Р. Фишером. Если $W_\theta(x)$ — одноименная плотность распределения случайной величины ξ со значениями на интервале a, b , зависящая от одномерного параметра θ , то количеством информации по Фишеру относительно этого параметра, содержащимся в одном измерении, называется следующая величина:

$$J_k \theta = \int_a^b W_\theta(x) \left[\frac{d}{d\theta} \ln W_\theta(x) \right]^2 dx = E \left(\frac{d}{d\theta} \ln W_\theta(x) \right)^2. \quad (1)$$

Не менее распространенной информационной мерой является мера, введенная Шенноном. Если $W_\theta(x)$ — одномерная плотность на интервале a, b , то энтропия соответствующей случайной величины численно определяется следующим интегралом:

$$H \theta = - \int_a^b W_\theta(x) \ln W_\theta(x) dx = - E \ln W_\theta(x). \quad (2)$$

Величина, равная разности значений энтропии, характеризующих используемые наблюдения в предположении, что параметр плотности распределения, стоящий под знаком логарифма, может принимать значения θ и $\theta + \Delta\theta$, называется информационным количеством Шеннона:

$$J_s \theta, \Delta\theta = \int_a^b W_\theta(x) \ln \frac{W_{\theta+\Delta\theta}(x)}{W_\theta(x)} dx = E \left(\ln \frac{W_\theta(x)}{W_{\theta+\Delta\theta}(x)} \right). \quad (3)$$

Количество информации по Фишеру характеризует поведение функции $\frac{d}{d\theta} \ln W_\theta(x)$ в окрестности наблюдений, в которой дифференцируемая функция принимает наибольшие значения. Зависимость $W_\theta(x)$, рассматриваемая как функция параметра, как известно, представляет собой функцию правдоподобия. Оценка неизвестного параметра θ , полученная методом максимума правдоподобия, находится как корень уравнения:

$$\frac{d}{d\theta} \ln W_\theta(x) = 0,$$

если верхняя граница функции правдоподобия достигается во внутренней точке области возможных значений параметра. Если незначительные изменения параметра θ приводят к значительным изменениям функции правдоподобия, то оценка этого параметра оказывается достаточно точной, а указанное математическое ожидание, численно характеризующее информацию о параметре, — довольно большим. Поэтому естественно, что количество информации по Фишеру характеризует качество оценивания параметра. Действительно, если θ^* — некоторая оценка параметра θ , то в силу неравенства Крамера-Рао дисперсия данной оценки удовлетворяет условию

$$D(\theta^*) \geq \left[1 + \frac{d}{d\theta} b(\theta) \right]^2 \frac{1}{J_F(\theta)}, \quad (4)$$

где $b(\theta)$ — смещение оценки.

Таким образом, при оценивании параметров методом максимального правдоподобия основной статистикой является производная логарифма функции правдоподобия, а при характеристике получаемых оценок — информация Фишера, равная математическому ожиданию квадрата данной статистики.

При проверке гипотез относительно параметров распределения, как известно, широко используется отношение правдоподобия. Если H — основная гипотеза о параметре $\theta: H = \{\theta = \theta_H\} = x \in W_H(x)$, где $W_H(x) = W_\theta(x)$, K — конкурирующая с ней гипотеза

$K = \{\theta = \theta_K\} = x \in W_K(x)$, то решающее правило, полученное по критерию отношения правдоподобия при простых гипотезах, имеет вид:

$$\lambda(x) = \frac{W_H(x)}{W_K(x)} \geq d_0, \quad (5)$$

где d_0 — пороговая величина

При построении решающего правила может успешно использоваться не только отношение правдоподобия, но и любая функция, обеспечивающая его монотонное преобразование. Наиболее часто используют логарифмическую и степенную функции:

$$\ln \frac{W_H(x)}{W_K(x)} \underset{K}{\overset{H}{\neq}} C_0, \quad \left(\frac{W_H(x)}{W_K(x)} \right)^q \underset{K}{\overset{H}{\neq}} C_0, \quad 0 < q < 1. \quad (6)$$

Качество проверки основной гипотезы в значительной мере определяется математическими ожиданиями статистик (6), сравниваемых с пороговыми величинами, которые вычисляются в предположении, что выполняются гипотезы:

$$J(H:K) = \int_{W_H(x)} \ln \frac{W_H(x)}{W_K(x)} dx, \quad (7)$$

$$\mu(H:K) = -\ln \int_{W_H(x)} \left(\frac{W_K(x)}{W_H(x)} \right)^q dx,$$

где полагается, что интегрирование ведется по области значений x , в которой $W_H(x) \neq 0$.

Первая из приведенных величин $J(H:K)$ представляет собой количество информации Шеннона (3). В связи с тем, что она детально изучена С. Кульбаком и Р. Лейбрером, ее называют количеством информации Кульбака–Лейбрера.

Второе математическое ожидание (7) нами названо информационным количеством Бхаттачария по той причине, что именно им в качестве расстояния между распределениями $W_H(x)$ и $W_K(x)$ предложено использовать величину $-\ln \int W_H^{1-q}(x) W_K^q(x) dx$.

Количество информации по Кульбаку–Лейбреру и методы, основанные на его использовании, довольно детально изучены [1–4].

Количество информации по Бхаттачарию и методы, основанные на его использовании, также хорошо изучены. В статье, результаты, полученные в монографиях [1–4] обобщаются и трансформируются на случай решения задач разработки математических моделей.

Известно, что величины, рассматриваемые как информационные меры, характеризуют расстояние между точками $W_H(x)$ и $W_K(x)$ в области возможных плотностей распределения вероятностей наблюдений. Примечательным свойством таких расстояний является их направленный характер, выражающийся в асимметрии функционалов: $J_{H:K} \neq J_{K:H}$, $\mu_{H:K} \neq \mu_{K:H}$. Полезность этого свойства логично следует из таких рассуждений: нельзя (в общем случае) требовать, чтобы количество информации в наблюдениях в пользу гипотезы H против гипотезы K в точности равнялось количеству информации в пользу K против H .

Таким образом, при проверке статистических гипотез информативным является отношение правдоподобия. Поэтому в качестве информационных мер, численно характеризующих

качество процесса проверки гипотез, целесообразно использовать математические ожидания монотонных функций указанного отношения при условии, что справедлива основная гипотеза. Именно на их основе выводится информационное неравенство, являющееся неравенством Крамера-Рао. Если неравенство Крамера-Рао характеризует качество оценок параметров распределений, то информационное неравенство характеризует качество проверки гипотез об этих параметрах.

Получаемое в результате вывода информационного неравенства выражение для плотности распределения наблюдений предлагается использовать в качестве аппроксимирующей плотности. Ниже приводится данный вывод.

Как уже отмечалось, введенные информационные меры:

$$J(H : K) = \int W_H(x) \ln \frac{W_H(x)}{W_K(x)} dx \text{ и } \mu(H : K) = -\ln \int W_H(x) \left(\frac{W_K(x)}{W_H(x)} \right)^q dx, \quad 0 < q < 1 \quad (8)$$

имеют направленный характер, они, соответственно, означают количество информации в пользу гипотезы основной H против конкурирующей K и наоборот. Большее количество информации свидетельствует о большей отличимости плотности $W_H(x)$ от плотности $W_K(x)$.

Во многих случаях при аппроксимации одной плотности, например, $W_H(x)$, важно, чтобы она была по возможности сильнее всего «похожей» на заданную плотность $W_K(x)$. Пусть степень «похожести» плотности $W_H(x)$ на плотность $W_K(x)$ характеризуется информационным количеством $J H : K$. При этом замена $W_H(x)$ на $W_K(x)$ является наилучшей с точки зрения «похожести» распределений, т. к. их информационная различимость минимальна: $J H : K_{W_H=W_K} = 0$. Однако аппроксимация тривиальна и не имеет практической ценности по той причине, что специфические особенности аппроксимируемой плотности не учитываются, она просто механически заменяется на другую.

Оказывается, что во многих случаях достаточно, чтобы при аппроксимации $W_H(x)$ с помощью $W_K(x)$ выполнялись условия, формулируемые в виде равенства

$$\int T(x) W_H(x) dx = \theta_H, \quad (9)$$

где $T(x)$ — некоторая измеримая статистика; θ_H — заданный параметр; интегрирование ведется по области определения плотности $W_H(x)$, т. е. по области значений, в которой данная плотность отлична от нуля.

Таким образом, задача аппроксимации $W_H(x)$ с помощью $W_K(x)$ сводится к нахождению $\min J H : K$ при условии, что выполняются следующие соотношения:

$$\int T(x) W_H(x) dx = \theta_H, \quad \int W_K(x) dx = 1, \quad W_K(x) \geq 0. \quad (10)$$

Если при этом $T(x)$ и θ_H — одномерные величины, то результат решения этой задачи известен [1]: аппроксимирующая плотность имеет вид:

$$W_H^*(x) = W_K(x) G^{-1}(u^*) e^{u^* T(x)}, \quad (11)$$

где $G(u) = \int W_K(x) e^{u T(x)} dx$, естественно, при условии что существует интеграл; u^* - корень уравнения

$$\frac{d}{du} \ln G(u) = \theta_H, \quad (12)$$

При этом минимальное расстояние между исходной $W_K(x)$ и аппроксимирующей $W_H(x)$ плотностями, обозначаемое через $J(* : K)$, удовлетворяет неравенству

$$J(* : K) \leq J(H : K), \quad J(* : K) = \int W_H^*(x) \ln \frac{W_H^*(x)}{W_K(x)} dx. \quad (13)$$

При этом за показатель качества аппроксимации плотности распределения можно взять следующую величину:

$$V_j = J(H:K)/J(*:K), \quad 1 \leq V_j < \infty. \quad (14)$$

Если данная величина мала, то качество аппроксимации низкое, в противном случае — оно высокое.

Таким образом, аппроксимирующая плотность (11) с помощью исходной плотности $W_K(x)$ порождает экспоненциальное семейство распределений с параметром θ_H и статистикой $T(x)$.

Представленному результату аппроксимации (он получен в [1]) присущи следующие ограничения и недостатки.

Во-первых, использовались одномерные величины $T(x)$ и θ_H , хотя в приложениях довольно часто информативными являются многомерные статистики и параметры.

Во-вторых, искомая плотность представлена в неявном виде, что затрудняет ее анализ и использование.

Ниже приводятся исследования, в определенной степени устраняющие указанные выше ограничения и недостатки.

Сформулированную выше задачу логично решить с помощью метода множителей Лагранжа. Неравенство в ограничивающих условиях (10) можно исключить, если отыскивать не плотность $W_H(x)$, а ее логарифм $\ln W_H(x) = g_H(x)$.

Таким образом, задача сводится к минимизации

$$J(H:K) \text{ по } g_H(x), \quad (15)$$

при ограничениях, задаваемых в виде равенств

$$\int T(x) \exp(g_H(x)) dx = \theta_H, \quad \int \exp(g_H(x)) dx = 1. \quad (16)$$

Обычная функция Лагранжа, подлежащая минимизации по $g_H(x)$, равна

$$L_{g_H, \lambda, \eta} = \int e^{g_H(x)} g_H(x) - g_K(x) + \lambda T(x) + \eta dx,$$

где $g_H(x) = \ln W_K(x)$; λ, η — множители Лагранжа.

Данная функция является регулярной по отношению к $g_H(x), \lambda, \eta$, ее производная Фреше по $g_H(x)$ совпадает с соответствующей производной Гато. Поэтому задача сводится к решению системы уравнений:

$$\begin{aligned} e^{g_H(x)} + e^{g_K(x)} g_H(x) - g_K(x) + \lambda T(x) e^{g_H(x)} + e^{g_H(x)} &= 0, \\ \int T(x) e^{g_H(x)} dx = \theta_H, \quad \int e^{g_H(x)} dx = 1, \end{aligned}$$

которая переписывается в следующем виде:

$$\begin{aligned} W_H(x) + W_K(x) \ln \frac{W_H(x)}{W_K(x)} + \lambda T(x) W_H(x) + \eta W_H(x) &= 0, \\ \int T(x) W_H(x) dx = \theta_H, \quad \int W_H(x) dx = 1. \end{aligned} \quad (17)$$

Интегрируя обе части первого уравнения этой системы по x и учитывая два оставшихся уравнения, получаем:

$$1 + \eta = -J(H:K) - \lambda \theta_H.$$

После подстановки этого решения в первое уравнение (17) имеем:

$$W_H(x) \ln \frac{W_H(x)}{W_K(x)} + \lambda T(x) - \theta_H W_H(x) - J(H:K) W_H(x) = 0. \quad (18)$$

Разделим обе части этого уравнения на $W_H(x)$, естественно, полагая, что $W_H(x) \neq 0$, и умножим их на $W_K(x)$:

$$-W_K(x) \ln \frac{W_K(x)}{W_H(x)} + \lambda T(x) - \theta_H W_K(x) - J(H:K) W_K(x) = 0.$$

Интегрируя обе части данного уравнения по x , получим:

$$-J(K:H) - J(H:K) + \lambda \theta_K - \theta_H = 0, \quad (19)$$

где $\theta_K = \int T(x)W_K(x) dx$.

Отсюда следует, что $\lambda = \theta_K - \theta_H^{-1} J_{H,K}$, где $J_{H,K} = J_{H:K} + J_{K:H}$.

Подставляя полученное значение λ в (18), находим:

$$W_H^{**}(x) = CW_K(x)e^{-\gamma T(x)}, \quad (20)$$

$$\text{где } C = \exp \left\{ -\frac{\theta_K}{\theta_H - \theta_K} J_{H:K} - \frac{\theta_H}{\theta_H - \theta_K} J_{K:H} \right\}, \quad \gamma = \frac{J_{K:H} + J_{H:K}}{\theta_H - \theta_K}.$$

То обстоятельство, что полученная плотность составляет минимум (а не максимум) рассматриваемого расстояния между W_H и W_K , следует из выпуклости вверх функции $u \ln u$, если $u = W_H(x)$.

Покажем, что плотности распределения (11) и (20) тождественно равны. Для этого полагаем, что искомое распределение имеет вид $W_H^{**}: K$. Проинтегрируем обе части (20) по x , в результате получим: $G = \left[\int W_H(x) e^{-\gamma T(x)} dx \right]^{-1}$,

т. е. $c = G^{-1}(u^*)/u^* = -\gamma$.

Теперь найдем $\frac{d}{d\gamma} \ln G(-\gamma)$:

$$\frac{d}{d(-\gamma)} \ln G(-\gamma) = G^{-1}(-\gamma) \int T(x) W_K(x) e^{-\gamma T(x)} dx = \int T(x) W_H^*(x) dx = \theta_H.$$

Отсюда следует, что γ — корень уравнения $\frac{d}{du} \ln G(u) = \theta_H$.

Следовательно, $W_H^{**} \equiv W_H(x)$ и $C = G^{-1}(u^*)$, $u^* = -\gamma$, что и требовалось доказать.

В дальнейшем, учитывая полученное, вместо (20) будем писать:

$$W_H^*(x) = CW_K(x)e^{-\gamma T(x)},$$

$$C = \frac{\theta_K J_{H:K}}{\theta_H - \theta_K} + \frac{\theta_H J_{K:H}}{\theta_H - \theta_K}, \quad \gamma = \theta_H - \theta_K^{-1} J_{H,K}, \quad (21)$$

что полностью эквивалентно представлению плотности в виде (11).

Таким образом, в отличие от известного результата, плотность распределения, наиболее близкая к исходной плотности $W_K(x)$, при учете ограничения (9), представлена в явном виде. При этом указанное ограничение переписывается следующим образом:

$$\int T(x) W_H^*(x) dx = \int T(x) W_H(x) dx, \quad (22)$$

т. е. аппроксимирующая и аппроксимируемая плотности W_H^* , W_H дают одинаковый результат усреднения выбранной статистики $T(x)$.

Важность этого результата, помимо его полезности, состоит в том, что он демонстрирует эффективность использованного нами метода минимизации направленного информационного количества: этот метод столь же эффективен, как и метод, примененный Кульбаком [1].

Обобщение полученных результатов на многомерный случай

Решим рассмотренную задачу применительно к случаю, когда величины $T(x)$ и θ_H являются векторными. В этом случае до получения (19) решение задачи не меняется за исключением того, что коэффициент λ становится векторным:

$$\lambda^T \theta - \theta_K^{-1} = -J_{H,K}. \quad (23)$$

Отыскание из этого одномерного уравнения решения относительно вектора λ обычным путем невозможно, т. к. число неизвестных превосходит количество имеющихся уравнений для их определения. Это является следствием введения большого числа ограничений на одну функцию $W_H(x)$.

В связи с отмеченным, при решении уравнения (23) относительно вектора λ воспользуемся псевдообращением прямоугольных матриц. Из большого числа псевдообращенных матриц выберем матрицу, обратную в смысле Мура (2). Такие матрицы всегда существуют, они единственны и являются наилучшими в том смысле, что имеют наименьшую «длину». То-есть вектор, получаемый в результате такого решения уравнения (23), обозначим его через λ^* , характеризуется наименьшим значением $\lambda^{*T}\lambda^*$.

Можно убедиться в том, что матрица, обратная в смысле Мура по отношению к матрице $\theta_H - \theta_K$, равна:

$$\theta_H - \theta_K^{-T} = \frac{\theta_H - \theta_K^{-T}}{\theta_H - \theta_K^{-T} \theta_H - \theta_K^{-T}}. \quad (24)$$

Поэтому решение уравнения (23) записывается в виде:

$$\lambda^* = - \theta_H - \theta_K^{-T} J_{H,K} = - \frac{\theta_H - \theta_K^{-T}}{\theta_H - \theta_K^{-T} \theta_H - \theta_K^{-T}} J_{H,K}. \quad (25)$$

Далее это решение должно быть подставлено вместо λ в следующее уравнение, являющееся аналогом (18):

$$\ln \frac{W_H(x)}{W_K(x)} + \lambda^T [T(x) - \theta_H] - J_{H,K} = 0.$$

В результате получаем:

$$W_H^*(x) = C_0 W_H(x) e^{-\gamma_0 T(x)}, \quad (26)$$

$$\text{где } C_0 = \exp \left\{ - \frac{(\theta_H - \theta_K)^T \theta_K}{(\theta_H - \theta_K)^T (\theta_H - \theta_K)} J_{H,K} - \frac{(\theta_H - \theta_K) \theta_H}{(\theta_H - \theta_K)^T (\theta_H - \theta_K)} J_{K,H} \right\};$$

$$\gamma_0 = \frac{\theta_H - \theta_K}{\theta_H - \theta_K^{-T} \theta_H - \theta_K^{-T}}.$$

Действуя аналогично, получим выражение для искомой плотности в случае, когда величины $T(x)$ и θ_H представляют собой матрицы:

$$W_H^*(x) = C_{00} W_K(x) e^{-S_p \Gamma_{00}^T T(x)}, \quad (27)$$

$$\text{где } \Gamma_{00} = \frac{\theta_H - \theta_K}{S_p (\theta_H - \theta_K)^T (\theta_H - \theta_K)};$$

$$C_{00} = \exp \left\{ - \frac{S_p (\theta_H - \theta_K)^T \theta_K}{S_p (\theta_H - \theta_K)^T (\theta_H - \theta_K)} J_{H,K} - \frac{S_p (\theta_H - \theta_K) \theta_H}{S_p (\theta_H - \theta_K)^T (\theta_H - \theta_K)} J_{K,H} \right\},$$

где $\theta_H, \theta_K, \theta(x)$ — матрицы, S_p - след матрицы.

Таким образом, получены выражения для плотностей распределения вероятности (26) и (27) для случаев, когда ограничения на аппроксимируемую плотность заданы, соответственно, в виде векторного и матричного равенств. Данные плотности, как и в случае одномерного ограничивающего равенства (20), относятся к экспоненциальному семейству распределений, порожденному $W_K(x)$ с параметром θ_H и статистикой $T(x)$, являющимися векторными или матричными величинами.

В заключение приведем пример.

Пусть известно, что наблюдения распределены по закону Стьюдента:

$$W_S(\bar{x}) = \frac{\Gamma \left[\frac{1}{2} (v + m) \right]}{\pi^{\frac{1}{2}m} \Gamma \left(\frac{1}{2} v \right) |V|^{\frac{1}{2}}} \left[1 + \bar{x} - \mu^T V^{-1} \bar{x} - \mu \right]^{-\frac{1}{2} (v+m)}, \quad (28)$$

где $V = v \sigma_0^2 \Sigma_0$.

Как известно, данная плотность при малом v имеет «толстые» хвосты, что проявляется

в увеличении моментов распределения. Зачастую это усложняет анализ, но этого можно избежать, если аппроксимировать данную плотность с помощью другого распределения. Выберем в качестве аппроксимирующего распределения полигауссовское распределение с плотностью

$$W_{PH} \bar{x} = \frac{1}{2\pi^{\frac{1}{2}m}} \left\{ q \sigma_1^2 \Sigma_0^{-\frac{1}{2}} \exp -A_1 + 1-q \sigma_2^2 \Sigma_0^{-\frac{1}{2}} \exp -A_2 \right\}, \quad (29)$$

$$A_n = \frac{1}{2\sigma_n^2} \bar{x} - \mu^T \Sigma_0^{-1} \bar{x} - \mu, \quad n=1,2$$

Задача сводится к такому выбору величин q , σ_1^2 и σ_2^2 , при котором плотность распределения (28) была бы в некотором смысле близка к плотности (29). Понятно, что за счет большого значения σ_1^2 либо σ_2^2 можно увеличить «хвосты» распределения $W_{PH} \bar{x}$.

Пусть известно также, что получаемые наблюдения имеют распределение (28), параметры которого известны полностью. Исключение составляет вектор μ . Он может принимать только два значения $\bar{0}, \mu_0$. В связи с этим выдвинем две гипотезы: основную $G_0 = \mu = \mu_0$, $\mu_0 \neq 0$, и конкурирующую $G_1 = \mu = \bar{0}$..

Для проверки этих гипотез воспользуемся критерием отношения правдоподобия. В результате получим следующее решающее правило:

$$\lambda_{\bar{x}} = \frac{1 + \bar{x} - \mu_0^T V^{-1} \bar{x} - \mu_0}{1 + \bar{x}^T V^{-1} \bar{x}} G_0 < \lambda_0, \quad (30)$$

где λ_0 — пороговая величина.

Следовательно, статистика $\lambda_{\bar{x}}$ является информативной, ее средние значения могут использоваться при решении сформулированной выше задачи аппроксимации распределения.

Таким образом, аппроксимируемая и аппроксимирующая плотности вероятности имеют вид:

$$W_H \bar{x} = \frac{\Gamma\left[\frac{1}{2}v+m\right]}{\pi^{\frac{1}{2}m} \Gamma\left(\frac{1}{2}v|V|^{\frac{1}{2}}\right)} 1 + \bar{x}^T V^{-1} \bar{x}^{-\frac{1}{2}} v+m, \quad V = v\Gamma_0^2 \Sigma_0 \quad (31)$$

$$W_K \bar{x} = \frac{1}{2\pi^{\frac{1}{2}m} |\Sigma_0|^{\frac{1}{2}}} \left[\frac{q}{\Gamma_1^m} e^{-B_1} + \frac{1-q}{\Gamma_2^m} e^{-B_2} \right],$$

$$B_n = \frac{-1}{2\Gamma_n^2} \bar{x}^T \Sigma_0^{-1} \bar{x}, \quad n=1,2$$

а ограничивающие равенства записываются следующим образом:

$$\theta_H = \int_{-\infty}^{\infty} \lambda_{\bar{x}} W_H \bar{x} d\bar{x}, \quad \theta_K = \int_{-\infty}^{\infty} \lambda_{\bar{x}} W_K \bar{x} d\bar{x}. \quad (32)$$

Если далее воспользоваться формулой (13), то получим, что искомая плотность имеет вид:

$$W_H^* \bar{x} = C \left[\frac{q}{\Gamma_1^m} e^{-B_1} + \frac{1-q}{\Gamma_2^m} e^{-B_2} \right] e^{-\gamma \frac{1 + \bar{x} - \mu_0^T V^{-1} \bar{x} - \mu_0}{1 + \bar{x}^T V^{-1} \bar{x}}}. \quad (33)$$

Итак, установлена структура искомой плотности распределения. Что касается коэффициентов C и γ , то их отыскание в данном случае и в большинстве других случаев, имеющих место в приложениях, связано с определенными вычислительными трудностями. При использовании плотности (33) проще всего полагать данные коэффициенты неизвестными величинами.

Заключение

Использование информационных методов позволяет сделать заметный шаг в направлении разработки предложений по рациональной аппроксимации плотностей распределения реально получаемых наблюдений с помощью различных не известных ранее многомерных распределений. Предложения существенным образом опираются на применение хорошо себя зарекомендовавшей в математической статистике информационной меры Кульбака-Лейбрера. Вводимые при этом ограничения на отыскиваемые плотности распределения вероятностей формируются в виде линейных равенств и практически устраняют усложнение выражений для данных плотностей. Это позволяет достаточно строго количественно охарактеризовать качество аппроксимации распределений и тем самым исключить риск, связанный с эвристическим предложением выбора аппроксимирующих плотностей распределения.

Л и т е р а т у р а :

1. *Кульбак С.* Теория информации и статистика. — М: Наука, 1967 — 408с.
2. *Рао С. Р.* Линейные статистические методы и их применения. — М.: Наука, 1968 — 578с.
3. *Тихонов В. И., Харисов В. Н.* Статистический анализ и синтез радиотехнических устройств и систем. — М: «Радио и связь», 1981.
4. *Прокофьев В. П.* Модели сигналов и поиск, используемые в традиционных задачах загоризонтного обнаружения. // Радиоэлектроника, информатика, управление. — 2004. — №1.

Статья поступила в редакцию 23.01.2007 г.

Prokofiev V. P.

The mathematical models development.

The usage of the information theory methods

under the approximation of the multi-measures densities distributions of observations.

Under the decisions of the huge quantity of problems in the different scientific and technical branches the used signals are described by models, which must be adequate to real signals. The well-known models are based on the Gauss random processes. However in the real information systems the signals have non-gaussian character and they have to be described by the non-gaussian distributions of momentary values. In this work it is given the analysis of the wide-spread information quantity measures, introduced by Fisher, Shannon, Kulback and Liebrer. Under the mathematical hypothesis verification the similarity criterion was used as the most informative. The Kulback results have the limitations and shortcomings, consist in the single-measure quantities usage and the non-evident presentation of the searched density distribution, what produces the difficulties under these results analysis and usage. The work includes the investigations, carried out on the base of the Lagrange factors methods and eliminate by some degree the above mentioned limitations and shortcomings. The known results for the single-measure distributions were generalized and thus the new results were obtained in relation to the multi-measure case.

Key words: information systems, gauss random processes, multi-measure distributions, information quantity measures (Fisher, Shannon, Kulback-Leibrer et sed), single-measure and multi-measures models of signals.