УДК 621.396:621.391

Оценка параметров навигационных сигналов в рамках теории информационной геометрии

В.В. Латышев

Аннотация

Рассмотрены основные понятия дифференциальной геометрии, показано, как выбор координатной системы естественным образом приводит к координатам, являющимся оценками параметров экспоненциального семейства плотностей распределения вероятностей, рассмотрен подход на основе ортогонального разложения наблюдаемых данных, позволяющий сократить размерность пространства наблюдений и одновременно сохранить количество фишеровской информации об оцениваемых параметрах, представлены результаты численного моделирования процедуры сокращения размерности для сигнала в виде 7-битового кода Голда, где в качестве параметров рассмотрены запаздывание и доплеровское смещение частоты.

Ключевые слова

фишеровская информация; оценка параметров; информационная метрика; аффинные связи; дуальность

1. Введение

В последние годы возник повышенный интерес к использованию достижений дифференциальной геометрии в теории вероятностей и математической статистике. С одной стороны дифференциальная геометрия является фундаментальной и хорошо развитой математической дисциплиной. С другой – еще Рао предложил рассматривать фишеровскую информацию как метрику для определения расстояния между различными плотностями распределения вероятностей [1]. Он установил связь между дифференциальной геометрией и статистикой для построения римановой геометрии на многообразии, определяемом диапазоном изменений параметров законов распределений, ввел семейство мер расстояний, базирующееся на энтропийной метрике, обобщающей фишеровскую информационную метрику, и одновременно сопрягающейся с понятием энтропии Шеннона. Кроме того, показал, что квадратичная дифференциальная метрика, характеризующая расстояние между законами распределения вероятностей, обладает привлекательными для практического использования свойствами. Симбиоз дифференциальной геометрии и теории информации привел к созданию новой математической дисциплины – информационной геометрии. Подробный обзор на эту тематику можно найти в [2]. Впоследствии советским математиком Н. Н. Ченцовым разработана аксиоматическая теория информационной геометрии [3]. Он использовал фишеровскую информационную матрицу как метрику, введенную на пространстве параметров, рассматриваемом как дифференцируемое многообразие, и ввел расстояние между параметрами семейства функций распределения вероятностей $G_\theta = \{p(./\theta): \theta \in \Theta\}$, где Θ - пространство параметров, как разницу между логарифмами плотностей распределения.

Информационная геометрия представляет собой новый математический инструмент для анализа данных во многих областях исследований. В частности, она уже показала свою эффективность в таких сферах как нейрокомпьютерная техника, многомерная информационная теория, нелинейные системы и нелинейное предсказание, статистические выводы и теория информации квантовых механических систем [4].

2. Основные понятия информационной геометрии

Информационная геометрия началась с изучения естественных дифференциальных структур, индуцированных семейством распределений вероятностей. Как простой пример можно рассмотреть множество S нормальных плотностей распределения вероятностей

$$p(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Поскольку конкретное распределение задается двумерным параметром (μ,σ), множество S можно рассматривать как двумерное пространство или многообразие (manyfold), в котором параметры (μ,σ) являются координатной системой. Однако это многообразие является не евклидовым пространством, а римановым с метрикой, которая естественным образом следует из свойств функций распределения. В частности, когда S – семейство нормальных распределений, это пространство является пространством постоянной отрицательной кривизны [4]. Фундаментальные свойства распределений приводят не только к римановой структуре указанного многообразия. Анализ свойств множества распределений приводит к новой концепции внутри дифференциальной геометрии: к взаимно дуальным аффинным связям. В свою очередь, структура взаимно-дуальных аффинных связей возникающая в рамках аффинной дифференциальной геометрии, начала привлекать внимание математиков - специалистов дифференциальной геометрии [4]. Информационная геометрия может также рассматриваться в рамках геометрии положительно определенных матриц для комплексных многомерных распределений Лапласа-Гаусса. Некоторые аспекты этого подхода в [5] недавно были представлены на основе функционального анализа.

В соответствии с теорией дифференциальной геометрии фишеровская информационная матрица представляет собой ковариантный симметричный тензор второго порядка, следовательно, ассоциированная с ней метрика инвариантна относительно допустимых преобразований параметров. Дифференциальная геометрия рассматривает множество плотностей распределений как дифференцируемое многообразие, тогда как случайные переменные и их математические ожидания являются векторами, для которых определено скалярное произведение в касательном к этому многообразию пространстве.

Соединение информационного и геометрического подходов при анализе статистических данных приводит к очень интересным параллелям. Так в [6] сформулирована и доказана теорема, утверждающая, что для плотностей распределения вероятностей с двумерным и трехмерным параметрическими пространствами существует оптимальная в смысле границы Крамера-Рао оценка всех параметров, которая эквивалентна решению уравнений Эйнштейна в вакууме.

Одним из центральных в дифференциальной геометрии является понятие дуальных связей, имеющее фундаментальное значение и сравнительно простым путем приводящее к оценкам параметров. Дуальность проявляется не только тогда, когда рассматриваются статистические модели, но и на множестве проблем, связанных с информационной геометрией. Через дуальность можно совместно анализировать эти проблемы, что является одной из основных мотиваций книги [4].

Рассмотрим принцип дуальности на примере экспоненциального семейства распределений. В соответствии с [4] к нему относится гауссовское распределение. В случае n-мерного вектора наблюдений $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ имеем

$$p(\mathbf{x};\mu,\sigma) = \left(2\pi\sigma^2\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right), \qquad (1)$$

где μ и σ математическое ожидание и среднеквадратическое отклонение, соответственно, составляют двумерный вектор (μ, σ) параметров. Суммирование ведется по всем значения индекса i от 1 до n.

В [4] для экспоненциального семейства используется иной вид записи

$$p(\mathbf{x};\theta)=\exp\left[\sum \theta_i F_i(\mathbf{x})-\psi(\theta)\right], \tag{2}$$

где m функций $F_1(\mathbf{x}), F_2(\mathbf{x}),\ldots, F_m(\mathbf{x})$ – зависят от всех составляющих случайного вектора переменных $\mathbf{x}$. Кроме того, новые параметры $[\theta_i]$ рассматриваются как натуральные, которые образуют аффинную координатную систему. Если переписать (1) в виде:

$$p(\mathbf{x};\theta)=\exp\left[-\frac{1}{2\sigma^2}\sum x_i^2+\frac{\mu}{\sigma^2}\sum x_i-\frac{n\mu^2}{2\sigma^2}-n\cdot\ln\left(\sqrt{2\pi}\sigma\right)\right], \tag{3}$$

и ввести обозначения $\theta_1=\frac{\mu}{\sigma^2}$, $\theta_2=-\frac{1}{2\sigma^2}$, $F_1(\mathbf{x})=\sum x_i$, $F_2(\mathbf{x})=\sum x_i^2$ и $\psi(\theta)=\frac{n\mu^2}{2\sigma^2}+n\cdot\ln\left(\sqrt{2\pi}\sigma\right)$, получается представление распределения (3) в виде (2).

Введение переменной

$$\eta_i=E_\theta[F_i]=\int F_i(\mathbf{x})p(\mathbf{x};\theta)d\mathbf{x} \tag{4}$$

дает новую аффинную координатную систему $[\eta_i]$ являющуюся дуальной к $[\theta_i]$ [4]. Фактически $[\eta_i]$ - математические ожидания, которые и являются дуальными параметрами.

Из введенных обозначений видно, что функции $F_i(\mathbf{x})$ являются оценками дуальных параметров η_i: $\hat{\eta}_i(\mathbf{x})=F_i(\mathbf{x})$. Тогда уравнение (3) означает, что $\hat{\boldsymbol{\eta}}=[\hat{\eta}_1,\ldots,\hat{\eta}_m]$ есть несмещенная оценка для координатной системы $\boldsymbol{\eta}=[\eta_1,\ldots,\eta_m]$, с ковариационной матрицей G, элементами которой являются:

$$g_{ij}(\theta)=E_\theta\left[(F_i-\eta_i)(F_j-\eta_j)\right]. \tag{5}$$

В [4] доказано, что G есть фишеровская информационная матрица для координатной системы $\boldsymbol{\theta}=[\theta_i]$, в то же время, ее инверсия - фишеровской информационной матрицей для $\boldsymbol{\eta}=[\eta_i]$. Следовательно, $\hat{\boldsymbol{\eta}}$ обеспечивает знак равенства в неравенстве Крамера-Рао и $\hat{\boldsymbol{\eta}}$ есть эффективная оценка.

3. Оценка параметров радионавигационных сигналов

Сигналы, используемые в глобальных навигационных системах ГЛОНАСС или GPS, хорошо вписываются в рамки экспоненциального семейства. Пусть семейство распределений вероятностей $S = \{p(\mathbf{x};\xi)\}$, зависит от параметра ξ. Множество S может рассматриваться как одномерное многообразие, для которого ξ есть локальная координатная система. Роль параметра ξ может играть задержка сигнала относительно момента его излучения или доплеровское смещение частоты при взаимном перемещении навигационного спутника и навигационного приемника. В простейшем случае пусть $x_1, x_2, ..., x_n$ – n независимых наблюдений случайной переменной x, распределенной в соответствии с $p(\mathbf{x};\xi)$. Предполагается, что вектор наблюдения $\mathbf{x} = (x_1, x_2, ..., x_n)$ представляет собой сумму дискретного варианта квазидетерминированного сигнала $s(t;\xi)$ и шума с нулевым математическим ожиданием и независимыми отсчетными значениями. В рамках этой модели (1) будет иметь вид:

$$p(\mathbf{x};\xi,\sigma) = \left(2\pi\sigma^2\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum\left(x_i - s_i(\xi)\right)^2\right) \qquad (6)$$

Чтобы привести к виду (2) перепишем это выражение:

$$p(\mathbf{x};\xi,\sigma) = \exp\left[-\frac{1}{2\sigma^2}\sum x_i^2 + \frac{1}{\sigma^2}\sum x_i s_i(\xi) - \frac{1}{2\sigma^2}\sum s_i^2(\xi) - n\cdot\ln\left(\sqrt{2\pi}\sigma\right)\right], \qquad (7)$$

Введем обозначения $\theta_1 = \frac{1}{\sigma^2}$, $\theta_2 = -\frac{1}{2\sigma^2}$, $F_1(\mathbf{x}) = \sum x_i s_i(\xi)$, $F_2(\mathbf{x}) = \sum x_i^2$ и $\psi(\xi) = \frac{1}{2\sigma^2}\sum s_i^2(\xi) - n\cdot\ln\left(\sqrt{2\pi}\sigma\right)$. Это дает представление распределения (7) в виде (2). Заметим, что в случае неэнергетического параметра ξ функцию $\psi(\xi)$ при больших n можно рассматривать в качестве константы.

Классическая задача статистического вывода – найти оценку $\hat{\xi}$ для ξ, используя заданную последовательность $x_1, x_2, ..., x_n$. Заметим, что в этой задаче только функция $F_1(\mathbf{x})$ зависит от параметра ξ, но она не дает ее оценки и это связано с нелинейной зависимостью используемых сигналов от параметра. Следовательно, $F_1(\mathbf{x})$ является дуальным параметром, но напрямую не оценивающим ξ.

В теории оценивания для заданного $\mathbf{x}$ распределение $p(\mathbf{x};\xi,\sigma)$ рассматривают как функцию ξ и называют ее функцией правдоподобия. Существуют различные подходы к формированию искомой оценки. Например, наиболее распространенная оценка максимального правдоподобия удовлетворяет уравнению

$$\max_{\xi} p(\mathbf{x};\xi,\sigma) = p(\mathbf{x};\hat{\xi},\sigma). \tag{8}$$

В (7), с учетом замечания о неэнергетическом характере параметра ξ только $F_1(\mathbf{x})$ зависит от него. Поскольку экспонента является монотонной функцией, (8) сводится к нахождению такого значения $\hat{\xi}$, которое обеспечит максимальное значение $F_1(\mathbf{x})$. В этом случае $\hat{\xi}$ будет оценкой максимального правдоподобия.

Заметим, что при нелинейной связи оцениваемого параметра с наблюдениями оценка максимального правдоподобия не является эффективной [7], т.е. не достигает потенциальной границы, которая обратно пропорциональна фишеровской информации об оцениваемом параметре, содержащейся в предъявленных данных. При нелинейной связи параметра и наблюдений количество фишеровской информации о параметре θ, вычисляется следующим образом [8]:

$$G(\xi) = E\left\{\left[\frac{\partial \ln p(\mathbf{x}\,|\,\xi)}{\partial \xi}\right]^2\right\} = -E\left(\frac{\partial^2 \ln p(\mathbf{x}\,|\,\xi)}{\partial \xi^2}\right). \tag{9}$$

Помимо максимального правдоподобия существуют и другие критерии: байесовский, максимальной апостериорной вероятности, минимаксный и другие [7]. Не останавливаясь на особенностях того или иного подхода к получению конкретной оценки параметра ξ, отметим, что при большой размерности вектора наблюдений $\mathbf{x}$ получение оценки связано с большим объемом вычислений. Рассмотрим, как можно сократить этот объем без заметной потери в точности оценивания.

Задача заключается в преобразовании $\mathbf{x}$ в новый вектор $\mathbf{y}$ с меньшей размерностью $m < n$ с помощью линейного преобразования $\mathbf{y} = \mathbf{C}\mathbf{x}$. Матрица преобразования $\mathbf{C}$ должна гарантировать минимальные потери фишеровской информации, чтобы для оценки параметра ξ можно было использовать $\mathbf{y}$. Для этой цели можно использовать теорему, сформулированную и доказанную в [9]. В соответствии с ней при независимых компонентах вектора наблюдения $\mathbf{x}$ строками матрицы $\mathbf{C}$ должны быть комплексно сопряженные собственные векторы матрицы:

$$\mathbf{B} = E_{\theta}\{\mathbf{s}'(\xi)(\mathbf{s}'(\xi))^{H}\}, \qquad (10)$$

где $\mathbf{s}(\xi)$ - вектор столбец ожидаемого навигационного сигнала конкретного спутника, а $\mathbf{s}'(\xi)$ - его производная по оцениваемому параметру, н – символ транспонирования с одновременным комплексным сопряжением. Количество фишеровской информации, передаваемое в вектор $\mathbf{y}$, можно рассчитать по формуле [8]:

$$I_m(\xi) = \sum_{k=1}^{m}\left[\mathbf{c}_k \mathbf{s}'(\xi)\right]^2, \qquad (11)$$

где $\mathbf{c}_k$ - соответствующая строка матрицы $\mathbf{C}$.

Для иллюстрации рассмотрим пример с фазоманипулированным сигналом $\mathbf{s} = (\xi)$, соответствующим с 7-битовому коду Голда. Длительность Т всей 7-битовой последовательности нормирована (Т=1). Размерность исходного вектора наблюдения $\mathbf{x}$ равна 128. Пусть параметром ξ является доплеровское смещение. Собственные векторы матрицы (10) позволяют сформировать матрицу $\mathbf{C}$, а фактически - ортогональное разложение $\mathbf{x}$, коэффициенты которого и образуют вектор $\mathbf{y}$. В этой формуле в качестве вектора $\mathbf{s}'(\xi)$ должна использоваться производная сигнала по доплеровскому смещению. На рис.1 показаны результаты расчета количества фишеровской информации в виде зависимости от размерности m вектора $\mathbf{y}$. Нормированный доплеровский сдвиг менялся в диапазоне [-3.5; 3.5]. По вертикали отложено количество информации, нормированное к количеству, содержащемуся в исходном векторе $\mathbf{x}$. Видно, что всего 8 первых членов ортогонального разложения обеспечивают сохранение практически всей информации о значении доплеровского смещения.

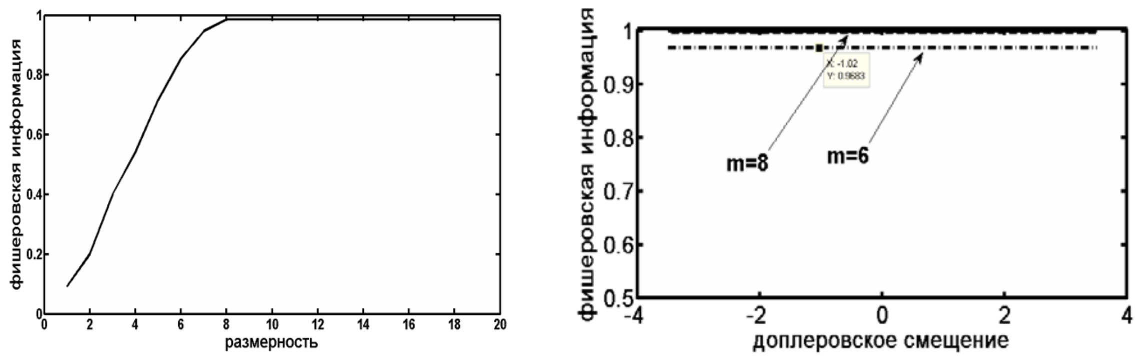

Рис.1. Фишеровская информация как функция вектора $\mathbf{y}$

Рис.2. Зависимость фишеровской информации размерности от значения доплеровского смещения

На рис.2 показана нормированная зависимость фишеровской информации от величины доплеровского смещения. Параметром выступает размерность вектора сокращенной размерности. Видно, что дальнейшее уменьшение размерности от 8 до 6 приводит уже к значительным информационным потерям, следствием чего будет ухудшение точности оценки доплеровского смещения с использованием вектора $\mathbf{y}$.

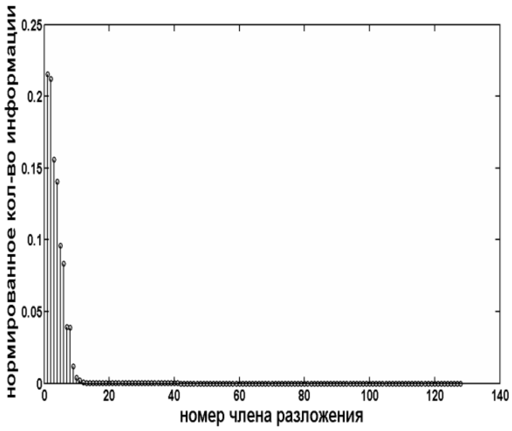

Рис.3. Распределение фишеровской информации по членам разложения (10)

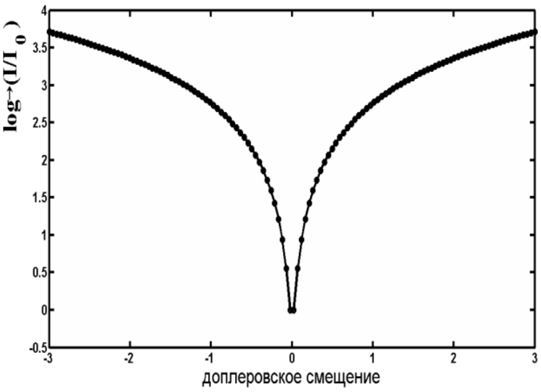

Рис.4. Зависимость количества информации о задержке от значения доплеровского смещения

Пусть теперь параметром ξ является запаздывание сигнала относительно момента излучения. Ортогональное разложение наблюдения, позволяющее концентрировать фишеровскую информацию о задержке, так же получается с использованием собственных векторов матрицы (10), но в этом случае в качестве вектора $\mathbf{s}'(\xi)$ должна использоваться производная сигнала по задержке. Результаты соответствующих расчетов показаны на рис.3 и рис.4. На рис.3 видно, как фишеровская информация о задержке концентрируется в первых членах ортогонального разложения. Здесь показано, какая часть информации о задержке связана с каждым членом полученного разложения. Видно, что практически вся информация сосредоточена в первых 10 членах разложения. На рис.4 показано количество фишеровской информации о задержке в зависимости от доплеровского смещения. По вертикальной оси отложен десятичный логарифм отношения количества фишеровской информации при произвольном доплеровском смещении I к аналогичному количеству при нулевом смещении I_0. Сплошная линия отображает информационное содержание исходного 256-мерного вектора наблюдения. Точками показано количество информации в 10-мерном векторе $\mathbf{y}$. Видно, что существенное уменьшение размерности представления данных со 256 до 10 возможно без потери информации о задержке. Кроме того, рис.4 иллюстрирует интересный факт, заключающийся в том, что фишеровская информация о задержке сигнала, а значит и потенциальная точность ее оценивания существенно зависит от доплеровского смещения. Чем больше доплеровское смещение, тем точнее может быть оценка положения сигнала на оси времени.

4. Заключение

Информационная геометрия является новым и быстро развивающимся направлением статистической теории оценивания и представления данных. Оно базируется на хорошо проработанной классической математической дисциплине - дифференциальной геометрии. Это раздел геометрии, в котором геометрические образы изучаются методами математического анализа. В классической дифференциальной геометрии исследуются локальные свойства геометрических образов, которые присущи сколь угодно малой их части и подчинены некоторым требованиям гладкости. Как правило, эти требования выражаются в том, что функции, задающие указанные объекты, не менее двух раз непрерывно дифференцируемы. В отличие от классической дифференциальной геометрии в информационной геометрии использует фишеровскую информацию как метрику для построения римановой геометрии на многообразии, определяемом диапазоном изменений параметров законов распределений. Геометрический подход позволяет естественным образом определять расстояния между различными статистическими распределениями, как точками риманового пространства. На этой основе получаются легко интерпретируемые оценки различных параметров наблюдаемых сигналов, обеспечивающие точности, близкие к теоретическим границам. Важно отметить, что вычислительная сложность оценок напрямую зависит от размерности обрабатываемых данных (размерности риманового пространства). Как показано в статье, для навигационных сигналов размерность представления данных может быть уменьшена на один – два порядка и более практически без потерь фишеровской информации об оцениваемом параметре, что позволяет уменьшить объем вычислений и, в тоже время, гарантирует сохранение точности оценивания в рамках более простой информационной геометрии.

Библиографический список

1. C. R. Rao. “Information and the accuracy attainable in the estimation of statistical parameters”. Bulletin of the Calcutta Mathematical Society, 1945, vol.37, pp.81-91.
2. F. Barbaresco. Robust Statistical Radar Processing in Fréchet Metric Space: OS-HDR-CFAR and OS-STAP Processing in Siegel Homogeneous Bounded Domains. Proceedings of International Radar Symposium IRS 2011, pp.639-644
3. Н. Н. Ченцов. Статистические решающие правила и оптимальные выводы. Наука, Москва, 1972, .
4. S. Amari, H. Nagaoko. Methods of information geometry. Translations of mathematical monographs, v. 191, AMS, University Press, Oxford, 2000.
5. R. Bhatia. Positive Definite Matrices (Princeton Series in Applied Mathematics). Princeton University Press, 2006.
6. M. Frasca. Optimal Cramer-Rao estimators for dimensions greater than two. Proceedings of International Radar Symposium IRS 2011, pp.657-650.
7. Г. Ван Трис. Теория обнаружения, оценок и модуляции. Сов. Радио, Москва, 1977, т.1.
8. Ибрагимов И. А., Хасьминский Р. З. Асимптотическая теория оценивания. Наука, Москва, 1979.
9. Латышев В.В. Сокращение размерности в задачах оценивания параметров. Радиотехника и электроника, 1988, т. 33, №3, с. 635-637.

Сведения об авторе:

Латышев Вячеслав Васильевич, профессор Московского авиационного института (национального исследовательского университета), д.т.н., тел.: 8 916-610-78-42, e-mail: lvv@mai.ru