

**Об оптимальном размытии точечных экспертных
оценок и их агрегировании в рамках теории
свидетельств**

А.Е. Лепский

МЦАВР НИУ ВШЭ, Москва

Общемосковский семинар «Математические методы анализа решений
в экономике, бизнесе и политике»,
20 декабря 2023, Москва

Агрегирование экспертной информации

- экспертное оценивание и прогнозирование широко применяется при разработке программ социально-экономического развития, принятии управленческих решений и т. д.;
- агрегирование прогнозной информации из разных источников — одна из важных методик экспертного оценивания:
 - алгоритм Hedge [Freund & Schapire 1997], использующий метод обновления мультипликативных весов экспертов;
 - консенсус-прогноз в макроэкономическом прогнозировании, как метод усреднения с весами или использование медианы.
- показано (см., н/р, [Kim et al 2001]), что к агрегированным прогнозам надо относиться с осторожностью;
- актуальна задача разработки новых методов агрегирования с учетом исторических данных источников, противоречивости отдельных прогнозов, их неопределенности и т. д.
- эти факторы удобно моделировать в рамках **теория свидетельств** [Dempster 1967, Shafer 1976].

План презентации

- инструментарий теории свидетельств;
- построение тел свидетельств-прогнозов:
 - статистика смещений экспертных оценок;
 - формирование фокальных элементов свидетельств-прогнозов;
 - оптимизация тела свидетельств;
- выбор и агрегирование экспертных оценок;
- численный пример:
 - описание данных;
 - построение тел свидетельств-прогнозов;
 - агрегирование прогнозов;
 - оценка качества агрегирования;
- заключение.

Инструментарий теории свидетельств

Пусть

- X — некоторый промежуток вещественной оси \mathbb{R} ;
- \mathcal{A}_X — множество всех отрезков на X ;
- \mathcal{A} — конечное подмножество (фокальных элементов) множества \mathcal{A}_X ;
- $m : \mathcal{A}_X \rightarrow [0, 1]$, $\sum_{A \in \mathcal{A}} m(A) = 1$ — **функция масс**, $m(A) > 0 \Leftrightarrow A \in \mathcal{A}$;
- пара $F = (\mathcal{A}, m)$ называется **телом свидетельств**;
- свидетельство $F_A = (A, 1)$ называется **категоричным**, F_X — **бессодержательным**.

Любое свидетельство можно представить в виде

$$F = \sum_{A \in \mathcal{A}} m(A) F_A.$$

Тело свидетельств вида $F_A^m = m F_A + (1 - m) F_X$, $m \in [0, 1]$ называется **простым**.

Противоречивость свидетельств

Например, свидетельство {рост ВВП в след. году составит 2.8–3.2% с достоверностью 0.6, 3.2–3.4% с достоверностью 0.3 или «неизвестно какой»} запишется так

$$F = 0.6F_{[2.8;3.2]} + 0.3F_{[3.2;3.4]} + 0.1F_X,$$

где (например) $X = [-5; 5]$.

Пусть два независимых источника информации описываются двумя свидетельствами $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$. Для оценивания их противоречивости (конфликтности) будем использовать меру

$$Con(F_1, F_2) = \sum_{A \in \mathcal{A}_1, B \in \mathcal{A}_2} \gamma(A, B) m_1(A) m_2(B), \quad (1)$$

где $\gamma(A, B) = 1 - s(A, B)$ и $s(A, B)$ — коэффициент сходства, удовлетворяющий условиям: 1) $0 \leq s(A, B) \leq 1$; 2) $s(A, B) = 0$, если $A \cap B = \emptyset$; 3) $s(A, A) = 1$.

Агрегирование свидетельств

Например, Jaccard index $s(A, B) = \frac{L(A \cap B)}{L(A \cup B)}$, который и будем использовать ниже, где $L(C)$ — длина отрезка $C \subseteq \mathbb{R}$ (сумма длин, если C — объединение нескольких отрезков).

С мерой конфликта (1) связано следующее обобщение правила Демпстера — **центральное правило комбинирования Чжана [Zhang 1994]**: $F_1 \otimes F_2 = (\mathcal{A}, m)$, где $\mathcal{A} = \{C = A \cap B \neq \emptyset : A \in \mathcal{A}_1, B \in \mathcal{A}_2\}$ и

$$m(C) = \frac{1}{K} \sum_{A \cap B = C} s(A, B) m_1(A) m_2(B), \quad (2)$$

если $K = 1 - \text{Con}(F_1, F_2) \neq 0$ (в этом случае $\mathcal{A} \neq \emptyset$).

Построение тел свидетельств-прогнозов

Пусть:

- $x(t)$ — некоторый показатель, зависящий от времени (рост ВВП, стоимость нефти, инфляция и пр.);
- $x_i = x(t_i)$ — этот показатель в момент времени t_i , $i = 1, \dots, N$;
- f_i — прогноз эксперта о значениях этого показателя в момент времени t_i , сделанный в момент времени t_{i-1} .

В результате мы имеем две выборки:

- прогнозных значений $\{f_i\}_{i=1}^N$;
- реальных значений $\{x_i\}_{i=1}^N$ показателя за N предшествующие моменты времени.

Предположим, что эксперт в момент времени t_N сделал прогноз f_{N+1} о значениях показателя в момент времени t_{N+1} . Поставим в соответствие этому прогнозу и с учетом исторических данных тело свидетельств вида

$$F_{[l,r]}^m = mF_{[l,r]} + (1 - m) F_X, \quad (3)$$

где $[l, r]$ — промежуток, содержащий истинное значение показателя со степенью доверия $m \in [0, 1]$, X — множество всех возможных значений показателя.

Промежуток $[l, r]$ получим из прогнозного значения f_{N+1} путем его «размытия» с учетом информации о предшествующих прогнозах. Степень доверия m получим из оценки надежности прогноза.

Статистика смещений экспертных оценок

С выборками $\{f_i\}_{i=1}^N$ и $\{x_i\}_{i=1}^N$ можно связать выборку $\Delta = \{\delta_i\}_{i=1}^N$ абсолютных смещений прогнозных значений

$$\delta_i = x_i - f_i, \quad i = 1, \dots, N$$

для безразмерных показателей (процент прироста ВВП, процент инфляции и пр.) или выборку относительных смещений

$$\delta_i = \frac{x_i - f_i}{f_i}, \quad i = 1, \dots, N.$$

Ниже рассмотрим только случай безразмерных показателей и абсолютных смещений.

Кроме того, будем рассматривать выборочное среднее (относительное) смещение

$$\bar{\delta} = \frac{1}{N} \sum_{i=1}^N \delta_i$$

и выборочное среднее квадратичное отклонение смещений

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\delta_i - \bar{\delta})^2}.$$

Размытие точечных оценок

Необходимо «размыть» точечные прогнозные значения $\{f_i\}_{i=1}^N$, т. е. построить такую систему отрезков $\{[l_i, r_i]\}_{i=1}^N$, чтобы $x_i \in [l_i, r_i]$, $i = 1, \dots, N$ и выполнялось некоторое оптимальное свойство.

1. Статистический подход

Найдем доверительный интервал, содержащий все выборочные значения $\tilde{x}_i = x_i - f_i - \bar{\delta} = \delta_i - \bar{\delta}$, $i = 1, \dots, N$ с некоторой вероятностью. Если считать, что смещения распределены по нормальному закону, то

$$P \left\{ \tilde{x}_i \in \left[-\frac{s}{\sqrt{N}} t_{\alpha, N-1}, \frac{s}{\sqrt{N}} t_{\alpha, N-1} \right] \right\} = 2\alpha - 1,$$

где $t_{\alpha, N-1}$ — α -квантиль распределения Стьюдента. Тогда промежуток «размытия» значения f_i равен

$$\left[f_i + \bar{\delta} - \frac{s}{\sqrt{N}} t_{\alpha, N-1}, f_i + \bar{\delta} + \frac{s}{\sqrt{N}} t_{\alpha, N-1} \right], \quad i = 1, \dots, N, \quad \alpha \in [0.5, 1].$$

В качестве массы этого фокального элемента можно взять величину $m = 2\alpha - 1$, равную вероятности попадания случайного прогноза в данный промежуток.

2. Оптимизационный подход

Будем искать отрезки $[l_i, r_i]$, $i = 1, \dots, N$ в виде

$$\begin{cases} l_i = f_i + \bar{\delta} - as, \\ r_i = f_i + \bar{\delta} + bs, \end{cases} \quad (4)$$

где $a, b \geq 0$ находятся из условия

$$x_i \in [l_i, r_i] \Leftrightarrow \tilde{x}_i = x_i - f_i - \bar{\delta} \in [-as, bs], \quad i = 1, \dots, N \quad (5)$$

и условия минимизации некоторого критерия. Например, в качестве критерия может быть выбрана длина отрезка $[l_i, r_i]$ (все отрезки равны по длине)

$$L(a, b) = r_i - l_i = (a + b)s \rightarrow \min. \quad (6)$$

Нетрудно видеть, что решением задачи (5), (6) будут границы

$$a_0 = \max \left\{ 0, \max_{1 \leq i \leq N} \left\{ \frac{\bar{\delta} - \delta_i}{s} \right\} \right\}, \quad b_0 = \max \left\{ 0, \max_{1 \leq i \leq N} \left\{ \frac{\delta_i - \bar{\delta}}{s} \right\} \right\}. \quad (7)$$

Заметим, что в оптимизационном подходе условие (5) является жестким. Поэтому такой способ размытия будет сильно неустойчивым к выбросам данных. Робастность оптимизационного алгоритма может быть достигнута, если вместо условия (5) использовать «мягкое» условие

$$l_i - \xi_i^- \leq x_i \leq r_i + \xi_i^+, \quad i = 1, \dots, N, \quad (8)$$

где вспомогательные переменные $\xi_i^- \geq 0$, $\xi_i^+ \geq 0$ имеют смысл величины ошибки «выхода» x_i за пределы промежутка $[l_i, r_i]$, $i = 1, \dots, N$.

Задача «мягкой» оптимизации нахождения границ промежутков размытия $[l_i, r_i]$, $i = 1, \dots, N$: необходимо найти a , b , $\xi^- = (\xi_i^-)$, $\xi^+ = (\xi_i^+)$, которые удовлетворяли бы условиям (4), (8) и минимизировали бы функционал

$$G(a, b, \xi^-, \xi^+) = L(a, b) + c \sum_{i=1}^N (\xi_i^- + \xi_i^+), \quad (9)$$

где слагаемое $\sum_{i=1}^N (\xi_i^- + \xi_i^+)$ характеризует штраф за суммарную ошибку выхода прогнозных значений за пределы промежутков $[l_i, r_i]$, $i = 1, \dots, N$, $c \geq 0$ — параметр, регулирующий отношение между минимизацией длины отрезка $L(a, b)$ и штрафом за непопадание в него.

Этот метод «мягкой» оптимизации аналогичен известному методу SVM регрессии [Drucker et al 1997].

Оптимизация тела свидетельств

В качестве оценки функции масс, отражающей степень принадлежности истинного значения прогнозируемому интервалу, можно взять относительную частоту попадания реальных значений показателя x_i в промежутки $[l_i, r_i]$, $i = 1, \dots, N$:

$$m = \frac{|i : x_i \in [l_i, r_i], i = 1, \dots, N|}{N}. \quad (10)$$

В результате «размытия» прогнозного значения f_{N+1} и оценивания массы получим тело свидетельств, которое будет зависеть только от одного параметра $c \geq 0$: $F = F(c)$. Этот параметр можно оценить, вводя дополнительные ограничения на F . В качестве такого ограничения рассмотрим минимизацию меры неточности информации, задаваемой F — **обобщенную кардинальность** (cardinality), которая для $F = (\mathcal{A}, m)$ определяется как

$$H(F) = \sum_{A \in \mathcal{A}} m(A)L(A).$$

Почему должно существовать оптимальное значение c , минимизирующее $H(F)$?

1. "Правдоподобные" рассуждения

Пусть $L(c) = (a + b)s$ — длина оптимального промежутка. Из общего вида G нетрудно видеть, что $L(c) \rightarrow L_{max} = (a_0 + b_0)s$ при $c \rightarrow \infty$ («жесткая» оптимизация) и $L(c) \rightarrow 0$ при $c \rightarrow 0$.

2. Вероятностная модель: равномерное распределение

Если считать, что $\{\tilde{x}_i\}$ являются выборочными значениями с.в. θ , то для большого N имеем $m \approx P\{\theta \in [-as, bs]\}$.

В частности, для равномерного распределения на отрезке $[-a_0s, b_0s]$ получим, что $m = m(c) \approx \frac{L(c)}{L_{max}}$. Тогда

$$H(F) \approx \frac{L^2(c)}{L_{max}} + \left(1 - \frac{L(c)}{L_{max}}\right) L(X), \quad 0 \leq L(c) \leq L_{max}$$

и c_{opt} должно удовлетворять условию $L(c_{opt}) = \frac{L(X)}{2}$, если $L(X) \leq 2L_{max}$ и $L(c_{opt}) = L_{max}$ в противном случае.

3. Вероятностная модель: нормальное распределение

Если выборочные точки $\{\tilde{x}_i\}$ распределены по нормальному закону $N(\bar{\delta}, s^2)$, то

$$m \approx P\{\theta \in [-as, bs]\} = \Phi(a) + \Phi(b),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ — функция Лапласа. Тогда мера неточности $H(F)$ будет приближенно равна

$$H(F) \approx (\Phi(a) + \Phi(b)) (a + b) s + (1 - \Phi(a) - \Phi(b)) L(X).$$

Можно показать, что эта функция достигает глобального минимума в точке (a_{opt}, b_{opt}) , где $a = a_{opt} = b_{opt}$ является единственным корнем уравнения

$$2\Phi(a)s = \Phi'(a) (L(X) - 2as), \quad \text{если } 0 \leq 2as \leq L(X).$$

На практике оптимальное значение c_{opt} можно найти, построив график зависимости $H(F(c))$.

4. Статистический подход: оптимальное α

Если же для формирования тела свидетельств использовался статистический подход, основанный на вычислении доверительного интервала, то длина этого интервала и оптимальная масса будут равны соответственно

$$L_{opt} = \frac{2st_{\alpha, N-1}}{\sqrt{N-1}}, \quad m_{opt} = 2\alpha - 1.$$

Поэтому оптимальная мера неточности

$$H(F) = \frac{2st_{\alpha, N-1}}{\sqrt{N-1}} (2\alpha - 1) + 2(1 - \alpha)L(X)$$

будет зависеть от параметра $\alpha \in [0.5, 1]$ и достигать наименьшего значения в некоторой внутренней точке этого промежутка.

Выбор и агрегирование экспертных оценок

Известны M прогнозов $f_{N+1}^{(k)}$, $k = 1, \dots, M$ о значениях показателя в момент t_{N+1} и информация о прошлых прогнозах, полученных из разных независимых источников.

Общая схема агрегирования экспертных прогнозов

- 1 Для каждого k -го источника информации найдем среднее смещение $\bar{\delta}^{(k)}$, выборочное среднее квадратичное отклонение $s^{(k)}$, параметры границ оптимальных промежутков $a_{opt}^{(k)}$ и $b_{opt}^{(k)}$, функции масс m_k , $k = 1, \dots, M$.
- 2 Для каждого k -го прогнозного значения $f_{N+1}^{(k)}$, $k = 1, \dots, M$, сформируем простые тела свидетельств нового прогноза вида $F_k = F_{[l_k, r_k]}^{m_k} = m_k F_{[l_k, r_k]} + (1 - m_k) F_X$, где

$$l_k = f_{N+1}^{(k)} + \bar{\delta}^{(k)} - a_{opt}^{(k)} s^{(k)}, \quad r_k = f_{N+1}^{(k)} + \bar{\delta}^{(k)} + b_{opt}^{(k)} s^{(k)}.$$

- 1 Найдем значения попарных конфликтов $Con(F_k, F_p)$, $k = 1, \dots, M - 1, p = k + 1, \dots, M$.
- 2 Пару тел свидетельств с минимальным конфликтом агрегируем в новое тело свидетельств по правилу комбинирования Чжана (2).

Численный пример. Описание данных

Применим эту методику для агрегирования прогнозов о темпе роста ВВП России с учетом исторических прогнозов за период 2010–2019 годов [Bulletin CDI 2023].

Источниками прогнозов:

- 1) Международный валютный фонд (IMF);
- 2) Институт центра развития НИУ ВШЭ (HSE Centre of Development Institute, HSE);
- 3) Минэкономразвития РФ (MED).

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
GDN	4,5	4,3	4,0	1,8	0,7	-2,0	0,2	1,8	2,8	2,2
IMF	1,5	4,8	3,5	3,8	2,0	0,5	-1,0	1,0	1,6	1,6
HSE	2,6	4,0	3,5	3,2	2,1	0,2	-0,7	1,0	1,7	1,4
MED	1,6	4,2	3,4	3,7	3,0	1,2	-0,6	1,0	2,1	1,3

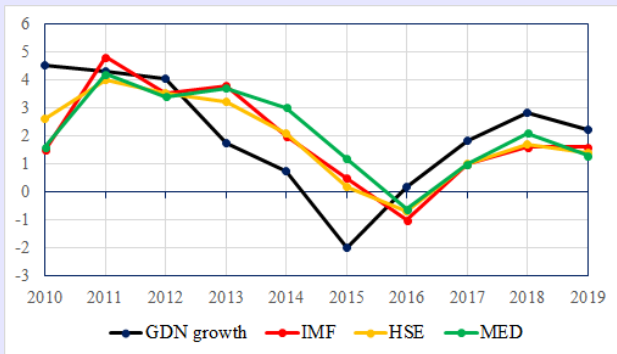


Рис.: Реальный ВВП и прогнозы, % прироста

Численный пример. Построение тел свидетельств-прогнозов

Для прогнозных значений 2019 года построим соответствующие тела свидетельств, используя статистику 2010–2018 годов.

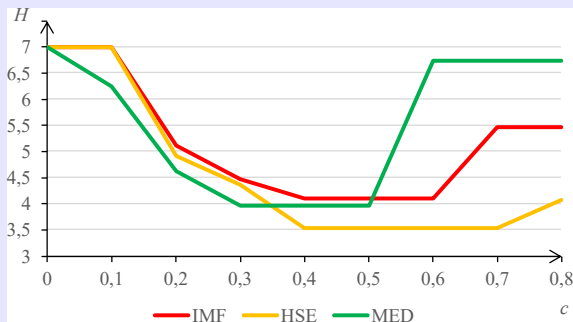


Рис.: Зависимости меры неточности $H(F)$ от значений параметра c

Результаты «жесткого» и «мягкого» оптимизационных «размытий» точечных прогнозов 2019 года и непараметрического оценивания масс приведены в Табл.

	$\bar{\delta}$	s	$-a_0s$	b_0s	$-a_{opt}s$	$b_{opt}s$	m	H	c_{opt}
IMF	0,05	1,16	-2,53	2,95	-2,1	1,16	0,78	4,08	$0,4 \div 0,6$
HSE	0,06	1,3	-2,24	1,84	-1,51	1,04	0,78	3,54	$0,4 \div 0,7$
MED	-0,16	1,81	-3,01	3,06	-2,11	1,0	0,78	3,96	$0,3 \div 0,5$

Таким образом, мы имеем три прогноза-свидетельства:

$$F_1 = 0,78F_{[-0,45;2,8]} + 0,12F_X, \quad F_2 = 0,78F_{[-0,05;2,5]} + 0,12F_X,$$

$$F_3 = 0,78F_{[-0,96;2,12]} + 0,12F_X,$$

которые будем рассматривать теперь на базовом множестве $X = [-1; 3]$.

Численный пример. Агрегирование прогнозов

Применим общую схему агрегирования прогнозов. Матрица попарных конфликтов $Con = \{Con(F_k, F_p)\}$ будет равна

$$Con = \begin{pmatrix} 0,064 & 0,226 & 0,265 \\ 0,226 & 0,125 & 0,329 \\ 0,265 & 0,329 & 0,079 \end{pmatrix}.$$

Наименьший конфликт у прогнозов IMF и HSE. Но найдем агрегирование всех трех пар — получим (1 – IMF, 2 – HSE, 3 – DEM):

$$F_{12} = 0,76F_{[-0,05;2,5]} + 0,18F_{[-0,45;2,8]} + 0,06F_X,$$

$$F_{13} = 0,56F_{[-0,45;2,12]} + 0,19F_{[-0,45;2,8]} + 0,18F_{[-0,96;2,12]} + 0,07F_X,$$

$$F_{23} = 0,56F_{[-0,05;2,12]} + 0,16F_{[-0,05;2,5]} + 0,2F_{[-0,96;2,12]} + 0,08F_X.$$

Численный пример. Оценка качества агрегирования

Качество агрегирования будем оценивать по двум параметрам:

- изменение меры близости между прогнозом и точным значением после агрегирования;
- изменения степени неточности после агрегирования.

В качестве меры близости между точечным значением x и прогнозным в виде тела свидетельств $F = (\mathcal{B}, m)$ будем использовать величину

$$P_F(x) = \sum_{B \in \mathcal{B}: x \in B} \frac{m(B)}{L(B)}.$$

	F_1	F_2	F_3	$F_1 \otimes F_2$	$F_1 \otimes F_3$	$F_2 \otimes F_3$
$P_F(x_{N+1}) \uparrow$	0,294	0,361	0,055	0,368	0,075	0,083
$H(F) \downarrow$	3,42	2,87	3,28	2,72	2,89	2,54

Нетрудно видеть, что после агрегирования близость прогноза к точному значению немного возросла относительно меры P_F и достигается на комбинировании наименее конфликтной пары тел свидетельств F_1 и F_2 . Кроме того, точность прогнозов после комбинирования всех пар также увеличилась.

Заклучение

- 1 на основе метода SVR регрессии предложена процедура оптимального «размытия» точечных значений и построения тел свидетельств, учитывающая историческую прогностическую информацию источника;
- 2 предложена процедура нахождения оптимальных значений параметров процедуры «размытия» на основе минимизации меры неточности полученного тела свидетельств;
- 3 предложена процедура выбора прогнозов (тел свидетельств) для комбинирования на основе минимизации меры противоречивости между телами свидетельств;
- 4 применимость указанных процедур продемонстрирована на примере выбора для агрегирования прогнозов о росте ВВП.

СПИСОК ИСТОЧНИКОВ



Bulletin of macroeconomic forecasts of the HSE Centre of Development Inst.
https://dcenter.hse.ru/consensus_forecast_archive. Last accessed 18.12.2023



Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Ann. Math. Statist. 38, 325–339 (1967)



Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V.: Support Vector Regression Machines. Advances in Neural Inf. Proc. Systems 9, 155–161 (1997)



Freund, Y., Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. of computer and system sciences 55, 119–139 (1997)



Kim, O., Lim, S.C., Shaw, K.W.: The inefficiency of the mean forecast as a summary forecast of earnings. Journal of Accounting Research 39, 329–336 (2001)



Lepskiy, A.: On Optimal Blurring of Point Expert Estimates and their Aggregation in the Framework of Evidence Theory. Procedia Computer Science 214, 573–580 (2022)



Shafer, G.: A mathematical theory of evidence. Princeton Univ. Press (1976)



Zhang, L.: Representation, independence and combination of evidence in the Dempster-Shafer theory. In: Yager, R.R. et al. (eds.) Advances in the Dempster-Shafer Theory of Evidence, pp. 51–69. John Wiley & Sons, New York (1994)

Спасибо за внимание

alepskiy@hse.ru

<https://www.hse.ru/en/org/persons/10586209>