сообщения
объединенного
института
ядерных
исследований
дубна

5-80-579

В.М.Горожанкин, В.М.Колобашкин, М.В.Лепешкин,
В.Н.Покровский, Т.М.Телевинова, В.П.Чистяков

К ВОПРОСУ ОБЪЕДИНЕНИЯ ОЦЕНОК

1980

Перед исследователем часто возникает задача объединения оценок одной и той же величины (как правило, многокомпонентной) с тем, чтобы найти новую, более точную оценку этой величины. Объединяемые оценки обычно получаются в результате довольно сложной процедуры обработки собственно данных эксперимента.

Настоящая статья посвящена рассмотрению ряда статистических задач, возникающих в связи с объединением оценок. В частности, исследуется влияние учета корреляции оценок на качество объединенной оценки, влияние на оценку замены теоретических весов их приближенными значениями, приводятся приближенные формулы для вычисления ковариационной матрицы вектора сложных оценок, предлагается метод оценки неизвестных параметров при случайных регрессионных переменных.

I. Случай одной компоненты. Объединение с постоянными весами

Рассмотрим сначала случай, когда оценивается единственная скалярная величина θ. Пусть по результатам эксперимента было получено k оценок $y_1, \ldots, y_k$ параметра θ. Будем предполагать, что эти оценки являются несмещенными (т.е. $My_i = \theta$), их ковариационная матрица $C = \{cov(y_i y_j)\}$ известна и ее определитель отличен от 0. Наши предположения можно записать в векторной форме:

$$\vec{y} = \vec{e}\,\theta + \vec{\delta} \quad ,$$

где $\vec{y}$ - вектор-столбец оценок $\{y_i\}$; $\vec{e}$ - вектор-столбец из I; $\vec{\delta}$ - вектор ошибок. При этом очевидно, что $M\vec{\delta} = 0$, $\mathcal{D}\vec{\delta} = \mathcal{D}\vec{y} = C$.

Здесь и далее символом $\mathcal{D}$, стоящим перед случайным вектором, обозначается ковариационная матрица вектора. Ниже мы часто будем использовать известные формулы (см., например, /I/):

$$\mathcal{D}(A\vec\xi) = A\,\mathcal{D}\vec\xi\, A^T; \qquad \mathcal{D}(\vec\xi+\vec\eta) = \mathcal{D}\vec\xi + \mathcal{D}\vec\eta, \qquad (I.I)$$

где $\vec\xi$, $\vec\eta$ – независимые случайные векторы; A – постоянная матрица; T – знак транспонирования.

Оценка $\hat\theta$ величины θ, полученная методом наименьших квадратов, имеет следующий вид (см., например, /2/):

$$\hat\theta = (\vec e^{\,T} C^{-1}\vec e)^{-1}(\vec e^{\,T} C^{-1})\vec y. \qquad (I.2)$$

Отсюда, воспользовавшись первой из формул (I.I), находим

$$\mathcal{D}\hat\theta = (\vec e^{\,T} C^{-1}\vec e)^{-1} = \left(\sum_{i,j=1}^{k} C^{-1}_{ij}\right)^{-1},$$

где через C^{-1}_{ij} обозначены элементы обратной матрицы C^{-1}. Если положить

$$W_i = \sum_{j=1}^{k} C^{-1}_{ij}, \qquad W = \sum_{i=1}^{k} W_i = \sum_{i,j=1}^{k} C^{-1}_{ij},$$

то можно записать

$$\hat\theta = W^{-1}\sum_{i=1}^{k} W_i y_i, \qquad \mathcal{D}\hat\theta = W^{-1}. \qquad (I.3)$$

Элементы обратной матрицы, входящие в W^{-1} и W_i, вычисляются довольно сложно, если оценки y_i коррелированы.

Если корреляции отсутствуют, то ковариационная матрица имеет вид $C_0 = diag(\sigma_1^2, \dots, \sigma_k^2)$, где $\sigma_j^2 = \mathcal{D}y_j$, и $C_0^{-1} = diag(1/\sigma_1^2, \dots, 1/\sigma_k^2)$. Получим оценку

$$\hat\theta_0 = W_0^{-1}\sum_{i=1}^{k} W_i^{(0)} y_i, \qquad \mathcal{D}\hat\theta_0 = W_0^{-1},$$

где $W_i^{(0)} = 1/\sigma_i^2$, $W_0 = \sum_{i=1}^{k} W_i^{(0)} = \sum_{i=1}^{k} 1/\sigma_i^2$.

Поскольку веса $W_i^{(0)} W_0^{-1}$ вычисляются значительно проще, чем $W_i W^{-1}$, рассмотрим, к чему приведет пренебрежение корреляциями при вычислении весов в том случае, когда корреляции между данными хотя и существуют, но малы и по порядку величины равны $\varepsilon \ll 1$.

Итак, возьмем оценку

$$\tilde\theta_0 = (\vec e^{\,T} C_0^{-1}\vec e)^{-1}(\vec e^{\,T} C_0^{-1})\vec y = W_0^{-1}(\vec e^{\,T} C_0^{-1})\vec y,$$

дисперсия которой будет, согласно первой из формул (I.I),

$$\mathcal{D}\tilde\theta_0 = W_0^{-2}(\vec e^{\,T} C_0^{-1} C\, C_0^{-1}\vec e) = W_0^{-2}\sum_{i,j=1}^{k} W_i^{(0)} C_{ij} W_j^{(0)} = W_0^{-2}\sum_{i,j=1}^{k} \frac{C_{ij}}{\sigma_i^2\sigma_j^2}. \qquad (I.4)$$

Ухудшение оценки можно охарактеризовать величиной $(\mathcal{D}\tilde\theta_0 - \mathcal{D}\hat\theta)/\mathcal{D}\hat\theta$, которую в общем случае оценить довольно сложно. При наших предположениях имеем

$$C = C_0 + \varepsilon R,$$

где $R = \{r_{ij}\}$, $r_{ij} = \rho_{ij}\sigma_i\sigma_j$ $(i \ne j)$, $r_{ii} = 0$, причем εr_{ij} и $\varepsilon\rho_{ij}$ – соответствующие ковариации и коэффициенты корреляции.

Воспользовавшись приближенным выражением

$$C^{-1} = \left[(I + \varepsilon R C_0^{-1})C_0\right]^{-1} = C_0^{-1}(I + \varepsilon R C_0)^{-1} \simeq$$

$$\simeq C_0^{-1}\left[I - \varepsilon R C_0^{-1} + \varepsilon^2 (R C_0^{-1})^2\right],$$

(I – единичная матрица), которое справедливо при одной из норм матрицы $\|\varepsilon R C_0^{-1}\| < 1$, например, $\sup_i \sum_{j=1}^{k} \left|\frac{\varepsilon r_{ij}}{\sigma_j^2}\right| < 1$, после некоторых преобразований получаем

$$\frac{\mathcal{D}\tilde\theta_0 - \mathcal{D}\hat\theta}{\mathcal{D}\hat\theta} \simeq \frac{\varepsilon^2}{2} W_0^{-2} \sum_{i,j=1}^{k} \frac{(\Delta_i - \Delta_j)^2}{\sigma_i^2\sigma_j^2},$$

где $\Delta_i = \sum_{\ell \ne i} \rho_{i\ell}\,\sigma_i/\sigma_\ell$. При малом отличии $\mathcal{D}\tilde\theta_0$ от $\mathcal{D}\hat\theta$ оценка $\tilde\theta_0$ может оказаться предпочтительнее.

Отметим, что в обеих оценках веса $W_i W^{-1}$ и $W_i^{(0)} W_0^{-1}$ зависят от теоретических ковариаций и дисперсий оценок y_i. При замене теоретических значений $cov(y_i y_j)$ их оценками мы получим случайные веса. Исследование влияния подобной замены на оценку θ проводится в следующем разделе.

2. Случай одной компоненты. Объединение со случайными весами

Здесь мы ограничимся следующим случаем. Будем считать, что оценки y_i ($i = 1, \dots, k$) параметра θ связаны следующим образом с собственно экспериментальными данными y_{ij}:

$$y_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}. \qquad (2.I)$$

Величины y_{ij} независимы, нормально распределены и $My_{ij}=\theta$, $\mathcal{D}y_{ij}=\beta_i^2$. Таким образом,

$$My_i=\theta\,,\quad \mathcal{D}y_i=\frac{\beta_i^2}{n_i}=\sigma_i^2\,.$$

Так как теперь $cov(y_i y_j)=0$ при $i\neq j$, то оценки $\hat\theta$ и $\hat\theta_0$ совпадают

$$\hat\theta=\hat\theta_0=\Big(\sum_{i=1}^{k}\frac{n_i}{\beta_i^2}y_i\Big)\Big/\Big(\sum_{i=1}^{k}\frac{n_i}{\beta_i^2}\Big).$$

Заменив в этой формуле величины β_i^2 их оценками

$$b_i^2=\frac{1}{n_i-1}\sum_{j=1}^{n_i}(y_{ij}-y_i)^2\,,$$

мы получим новую объединенную оценку

$$\bar\theta=\sum_{i=1}^{k}h_i y_i\,,$$

где нормированные случайные веса h_i определяются формулами

$$h_i=\frac{1/S_i^2}{\bar W_0}\,,\quad \bar W_0=\sum_{i=1}^{k}\frac{1}{S_i^2}\,;\quad S_i^2=\frac{b_i^2}{n_i}\,.$$

Практически используется именно оценка $\bar\theta$, так как она вычисляется по известным величинам y_i и по оценкам их дисперсий S_i^2. Исследуем ее свойства.

Используя взаимную независимость $y_1,\ldots,y_k,\ b_1^2,\ldots,b_k^2$ /3/, нетрудно показать, что в $\bar\theta$ не была внесена систематическая ошибка, т.е. $M\bar\theta=\theta$. Кроме того, проверяется также, что

$$\mathcal{D}\bar\theta=\sum_{i=1}^{k}\frac{\beta_i^2}{n_i}Mh_i^2=\sum_{i=1}^{k}\sigma_i^2 Mh_i^2\,.$$

Для Mh^2 нет простой точной формулы, однако, если $n_i=\alpha_i n$ ($\alpha_i=const$), то при $n\to\infty$ можно найти разложение $\mathcal{D}\bar\theta$ по степеням $1/n$. Такое разложение с двумя членами с учетом порядка малости отброшенных членов (ограниченность k не предполагается) имеет вид

$$\mathcal{D}\bar\theta=\frac{1}{W_0}+\frac{2}{W_0^2}\sum_{i=1}^{k}\frac{1}{n_i\sigma_i^2}\Big(1-\frac{1}{W_0}\cdot\frac{1}{\sigma_i^2}\Big)+O\Big(\frac{1}{n^{5/2}}+\frac{k^2}{n^t}\Big)\,,$$

где $t>0$, или

$$\mathcal{D}\bar\theta=\frac{1}{W_0}\Big[1+2\sum_{i=1}^{k}\frac{\eta_i(1-\eta_i)}{n_i}\Big]\,,$$

где $\eta_i=W_0^{-1}(1/\sigma_i^2)$ – теоретические нормированные веса. Строгое обоснование приведенного выше разложения выходит за рамки настоящей работы. Отметим, что разложение $\mathcal{D}\bar\theta$ с двумя членами было получено в работе /4/. В качестве оценки $\mathcal{D}\bar\theta$ естественно использовать

$$\hat{\mathcal{D}}\bar\theta\approx\frac{1}{\bar W_0}\Big[1+2\sum_{i=1}^{k}\frac{h_i(1-h_i)}{n_i}\Big]\,.$$

Таким образом, при малых объемах выборок n_i главный член $\bar B^2=\bar W_0^{-1}$ в оценке $\hat{\mathcal{D}}\bar\theta$ может заметно отличаться от $\mathcal{D}\bar\theta$ и в этом случае поправочный член нужно учитывать.

Исследуем теперь асимптотическую интервальную оценку параметра θ и выясним границы ее применимости. Можно показать (см. теорему п.28.4 /3/), что при $n_i=\alpha_i n$ оценка $\bar\theta$ асимптотически нормальна с математическим ожиданием θ и дисперсией $B^2=1/W_0$. Отсюда, так как b_i^2 сходятся по вероятности к β_i^2, следует, что

$$\lim_{n\to\infty}P\Big(\Big|\frac{\bar\theta-\theta}{\bar B}\Big|\leqslant u_\alpha\Big)=\frac{1}{\sqrt{2\pi}}\int_{-u_\alpha}^{u_\alpha}e^{-\frac{u^2}{2}}dU=1-2\alpha\,.$$

Таким образом, можно написать асимптотический доверительный интервал для θ. Неравенство

$$\bar\theta-u_\alpha\bar B\leqslant\theta\leqslant\bar\theta+u_\alpha\bar B \tag{2.2}$$

верно с вероятностью $1-2\alpha$. При небольших значениях n_i вероятность выполнения неравенства может отличаться от $1-2\alpha$. Для оценки границ применимости интервальной оценки при помощи случайных чисел для заданных θ, n_i, β_i 100 раз моделировались значения y_{ij} и определялась частота нарушений $(2\alpha)^*$ неравенства. Результаты приведены в таблице (u_α выбирались для 5% уровня значимости). Оказалось, что в качестве параметра, характеризующего близость распределения $\bar\theta$ к нормальному, можно предложить $\Delta=(\hat{\mathcal{D}}\bar\theta-\bar B^2)/\bar B^2$. При $\Delta\leqslant 0{,}04$ частоты $(2\alpha)^*_{\bar B}$; $(2\alpha)^*_{\hat{\mathcal{D}}}$ нарушений неравенства (2.2) мало отличаются от вероятности 2α (см. таблицу). В приведенных примерах это неравенство выполняется обычно при всех $n_i>25$.

Часто используемая оценка дисперсии среднего взвешенного

$$S_R^2=\frac{1}{k-1}\sum_{i=1}^{k}h_i(y_i-\bar\theta)^2$$

(так называемая погрешность разброса) дает в большинстве случаев неправильный доверительный интервал (см. таблицу). Таким об-

Т А Б Л И Ц А

$(2\alpha)^{\circ}_R$	$(2\alpha)^{\theta}_B$	$(2\alpha)^{\theta}_D$	S_R	$\bar{B}$	$\sqrt{D\bar{\theta}}$	$\sqrt{\hat{D}\bar{\theta}}$	Δ	1		2		3		4		5		6		7		8		9		10	
			$\times 10^{-5}$					β_1	n_1	β_2	n_2	β_3	n_3	β_4	n_4	β_5	n_5	β_6	n_6	β_7	n_7	β_8	n_8	β_9	n_9	β_{10}	n_{10}
20	23	20	1148	898	1165	1028	0,308		3		4		3		5		5										
12	7	7	709	696	756	737	0,121		10		10		8		9		9										
14	5	5	478	507	529	522	0,060	0,0822	20	0,0300	20	0,0982	18	0,0950	19	0,0590	19										
16	8	8	265	275	276	277	0,015		69		72		69		75		75										
15	5	5	149	159	159	160	0,013		207		216		207		225		225										
21	32	26	517	397	589	462	0,354		5		5		3		4		4		5		5		4		4		3
13	15	15	328	312	340	331	0,126		15		15		9		12		12		15		15		12		12		9
6	3	2	237	221	230	228	0,064	0,0822	23	0,0300	24	0,0982	23	0,0950	25	0,0590	25	0,0871	12	0,0650	15	0,0212	20	0,0423	14	0,0200	35
17	8	6	185	190	196	194	0,043		45		45		27		36		36		45		45		36		36		27
10	2	2	73	76	77	77	0,027		207		216		207		225		225		108		135		180		126		315
24	29	26	3314	2532	3495	2877	0,291		3		4		3		5		5										
13	6	5	1894	1897	2018	1988	0,098		9		12		9		15		15										
14	7	6	1324	1396	1434	1430	0,055	0,2466	23	0,0900	24	0,2346	23	0,2850	25	0,1770	25										
12	5	5	841	894	902	901	0,016		30		72		39		24		60										
17	4	4	437	477	478	479	0,008		207		216		207		225		225										

разом, часто встречающаяся практика использования в качестве оценки $\hat{\mathcal{D}}\bar{\theta}$ наибольшей из двух оценок $\bar{B}^2=\bar{W}_0^{-1}$ и S_R^2 статистически не оправдана и проистекает, по-видимому, из стремления перестраховаться. Для исследуемого случая показано, что интервал (2.2) содержит оцениваемый параметр θ с вероятностью, близкой к теоретической.

Мы рассмотрели случай, когда объединяемые оценки y_i связаны с данными эксперимента y_{ij} простыми формулами (2.I). В практически интересных случаях y_i вычисляются по данным эксперимента посредством довольно сложной процедуры. Сами экспериментальные данные могут и не быть гауссовскими. Однако даже в этом случае, если оценки y_i и $\hat{\sigma}_i^2$ окажутся состоятельными и асимптотически нормальными оценками θ и $\mathcal{D}y_i$, то предлагаемая оценка $\bar{\theta}$ тоже будет асимптотически нормальной. Параметры близости распределения $\bar{\theta}$ к нормальному распределению могут быть исследованы в практически важных случаях на типичных примерах по методике, примененной для случая (2.I).

3. Общий случай объединения оценок

Пусть полный вектор оценок $\vec{y}$ состоит из векторов $\vec{y}_i$ ($i=1,\ldots,\kappa$), а каждый вектор $\vec{y}_i$ является несмещенной оценкой вектора неизвестных величин $\vec{\theta}$ ($\theta_1,\ldots,\theta_m$) (например, κ раз измерялся набор из m γ-переходов при разных условиях, в разных лабораториях и т.д.). Тогда линейная модель имеет вид

$$y_{ij}=\theta_j+\delta_{ij} \qquad (j=1,\ldots,m\ ;\ i=1,\ldots,\kappa),$$

или, в матричной форме,

$$\vec{y}=E\vec{\theta}+\vec{\delta},$$

где $\vec{y}$ – вектор размерности $n=\kappa\times m$; E – матрица размерности $n\times m$, составленная из κ единичных матриц I_m ; $\vec{\delta}$ – n-мерный вектор ошибок, для которого $M\vec{\delta}=0$.

Будем предполагать, что ковариационная матрица $C=\mathcal{D}\vec{y}=\mathcal{D}\vec{\delta}$ невырождена и известна. В нашем случае для объединенной оценки $\hat{\vec{\theta}}$ сохранится формула типа (I.2), в которой $\vec{e}$ заменяется на матрицу E .

Ковариационную матрицу C и ее обратную C^{-1} можно представить в виде

$$C = \{ C_{ij} \} \quad , \quad C^{-1} = \{ C_{ij}^{-1} \} \quad ,$$

где C_{ij} , C_{ij}^{-1} – подматрицы размерности $m \times m$. Отметим, что в общем случае подматрица C_{ij}^{-1} не является обратной к C_{ij}. В этих обозначениях для $\hat{\vec{\theta}}$ и $\mathcal{D}\hat{\vec{\theta}}$ легко выписываются явные выражения:

$$\hat{\vec{\theta}} = W^{-1} \sum_{i=1}^{k} W_i \vec{y}_i \quad , \quad \mathcal{D}\hat{\vec{\theta}} = W^{-1} , \qquad (3.I)$$

где матрицы $W_i = \sum_{j=1}^{k} C_{ij}^{-1}$ и $W^{-1} = \left(\sum_{i=1}^{k} W_i \right)^{-1}$ имеют размерность $m \times m$. Если предположить, что матрица C близка к диагональной, то можно для "весов" $W_i W^{-1}$ написать приближенные формулы и провести исследования влияния корреляций на объединенную оценку.

Заметим, что здесь и далее под m понимается максимальная размерность вектора $\vec{\theta}$. Если некоторые из векторов $\vec{y}_i$ содержат оценки не всех параметров $\theta_1, \dots, \theta_m$, они дополняются нулями, причем в матрице C также появляются строки и столбцы из нулей. Тем не менее, формула (3.I) остается справедливой, но вместо обратных матриц в соответствующие выражения входят псевдообратные матрицы.

4. Вычисление ковариационной матрицы объединяемых оценок

При объединении оценок нужно знать их ковариационную матрицу. Корреляция между оценками обычно появляется в результате использования при их вычислении общих случайных параметров. Пусть полный вектор оценок $\vec{y} = (\vec{y}_1, \dots, \vec{y}_k)^T$ определяется формулой

$$\vec{y} = \vec{f}(\vec{x}, \vec{\varphi}) = \left[f_1(\vec{x}_1, \vec{\varphi}), \dots, f_k(\vec{x}_k, \vec{\varphi}) \right]^T ,$$

где $\vec{x} = (\vec{x}_1, \dots, \vec{x}_k)^T$ – вектор регрессионных переменных, а $\vec{\varphi} = (\varphi_1, \dots, \varphi_N)^T$ – вектор общих случайных параметров, причем векторы $\vec{x}_1, \dots, \vec{x}_k$, $\vec{\varphi}$ взаимно независимы. Размерность вектора $\vec{x}_i$ обозначим через m_i , т.е. $\vec{x}_i = (x_{i1}, \dots, x_{im_i})^T$. Заметим также, что вектор $\vec{\varphi}$ размерности N составляется, вообще говоря, из векторов $\vec{a}_i$ общих параметров, специфических для i-ого блока данных, и вектора $\vec{b}$ параметров, общих для всех блоков.

Найдем приближенную формулу для $\mathcal{D}\vec{y}$. Разложим функцию $\vec{f}$

в ряд Тейлора в окрестности точки $M\vec{x}, M\vec{\varphi}$. Ограничившись линейным приближением, получим

$$\vec{y} \simeq \vec{f}_0 + F_0^{(x)}(\vec{x} - M\vec{x}) + F_0^{(\varphi)}(\vec{\varphi} - M\vec{\varphi}) , \qquad (4.I)$$

где $\vec{f}_0 = \left[f_1(M\vec{x}_1, M\vec{\varphi}) , \dots , f_k(M\vec{x}_k, M\vec{\varphi}) \right]^T$, а $F_0^{(x)}$, $F_0^{(\varphi)}$ – матрицы неслучайных величин вида

$$F_0^{(x)} = \begin{pmatrix} (grad_{\vec{x}_1} f_1)^T & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & (grad_{\vec{x}_k} f_k)^T \end{pmatrix}$$

$$F_0^{(\varphi)} = \left((grad_{\vec{\varphi}} f_1)^T, \dots , (grad_{\vec{\varphi}} f_k)^T \right) ,$$

причем

$$(grad_{\vec{x}_i} f_i)^T = \left(\frac{\partial f_i}{\partial x_{i,1}} , \dots , \frac{\partial f_i}{\partial x_{i,m_i}} \right)$$

$$(grad_{\vec{\varphi}} f_i)^T = \left(\frac{\partial f_i}{\partial \varphi_1} , \dots , \frac{\partial f_i}{\partial \varphi_N} \right) .$$

Следует отметить, что элементы матриц $F_0^{(x)}$, $F_0^{(\varphi)}$, соответствующие неиспользованным компонентам x_{ij} и/или общим случайным параметрам φ_e , будут равны нулю.

Так как $\vec{x}$ и $\vec{\varphi}$ независимы, то из формулы (I.I) и из формулы (4.I) получим

$$\mathcal{D}\vec{y} \approx F_0^{(x)} \mathcal{D}\vec{x} \, F_0^{(x)T} + F_0^{(\varphi)} \mathcal{D}\vec{\varphi} \, F_0^{(\varphi)T} . \qquad (4.2)$$

Отсюда, учитывая независимость $\vec{x}_1, \dots, \vec{x}_k$, получим

$$Cov(y_i y_j) \approx \begin{cases} \sum_{r,s=1}^{N} \frac{\partial f_i}{\partial \varphi_r} cov(\varphi_r \varphi_s) \frac{\partial f_i}{\partial \varphi_s} , & \text{если } j \neq i \\ \sum_{r,s=1}^{N} \frac{\partial f_i}{\partial \varphi_r} cov(\varphi_r \varphi_s) \frac{\partial f_i}{\partial \varphi_s} + \sum_{r,s=1}^{m_i} \frac{\partial f_i}{\partial x_{i,r}} cov(x_{i,r} x_{i,s}) \frac{\partial f_i}{\partial x_{i,s}} , & \text{если } j = i . \end{cases}$$

Таким образом, $cov(y_i y_j)$ при $i \neq j$ прямо пропорциональна сумме ковариаций общих используемых параметров. Общими параметрами, определенными с ошибкой, могут быть, например, коэффициенты градуировочной кривой спектрометра. Некоторые статистические задачи, связанные с градуировкой, рассматриваются в следующем разделе.

5. Некоторые задачи градуировки

Определение энергий переходов на некотором спектрометре включает в себя две статистические задачи:

I. градуировка спектрометра (оценка коэффициентов градуировочной кривой);

2. определение энергий и вычисление их ковариационной матрицы с использованием оценок коэффициентов градуировочной кривой. На втором этапе в качестве общих параметров (типа вектора $\vec{a}_i$ из раздела 4) выступают коэффициенты градуировочной кривой, что обуславливает корреляцию между компонентами вектора энергий, найденными при данной градуировке. На первом же этапе роль общих параметров (типа вектора $\vec{b}$ раздела 4) играют энергии реперных переходов, использованных (в разных опытах и т.п.) для градуировки. Положение дополнительно осложняется тем, что энергии реперных переходов в свою очередь могут быть закоррелированы между собой благодаря использованию при их определении общих случайных параметров – энергий нормалей.

Рассмотрим первый этап. Для идеальной градуировочной кривой справедливо соотношение

$$\vec{\gamma} = \Pi \vec{\varphi}, \tag{5.I}$$

где

$\vec{\gamma} = (\gamma_1, \dots, \gamma_m)^T$ – вектор энергий;

$\vec{\varphi} = (\varphi_1, \dots, \varphi_n)^T$ – вектор коэффициентов градуировочной кривой;

$\Pi = \{\pi_{ij}\}$, $i = 1,\dots,m;\ j = 1,\dots,n$ – матрица, элементы которой есть функции номеров каналов идеального спектрометра, где зарегистрированы энергии γ_i.

Если, например, идеальная градуировочная кривая является полиномом степени $n-1$, то $\pi_{ij} = \pi_i^{j-1}$, где π_i – номер канала идеального спектрометра, в котором зарегистрирована энергия γ_i.

В реальном случае нам известны только оценки $\vec{g}$ и P энергии $\vec{\gamma}$ и матрицы Π. Используя эти оценки и уравнение связи (5.I), можно найти оценку $\hat{\vec{\varphi}}$ вектора $\vec{\varphi}$. Будем предполагать, что оценки $\vec{g}$ и P – независимые и несмещенные ($M\vec{g} \equiv \vec{\gamma}$, $MP \equiv \Pi$). Когда распределение $\vec{g}$ и P является гауссовским, мы найдем оценки максимального правдоподобия для $\vec{\varphi}$, которые в этом случае совпадают с оценками по методу наименьших квадратов.

Для вычисления ковариаций и записи формулы для функции правдоподобия нам удобно будет преобразовывать по определенному правилу матрицу в вектор и вектор в матрицу. Пусть имеется матрица $\Pi = \{\pi_{ij}\}$ размерностью $m \times n$ и вектор $\vec{\varphi} = (\varphi_1, \dots, \varphi_n)^T$, из элементов которых образуем вектор $\vec{\pi}$ и матрицу Φ:

$$\vec{\pi} = (\pi_{11}, \dots, \pi_{1n}, \pi_{21}, \dots, \pi_{2n}, \dots, \pi_{m1}, \dots, \pi_{mn})^T,$$

$$\Phi = \begin{pmatrix} \vec{\varphi}^T & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \vec{\varphi}^T \end{pmatrix}$$

размерности соответственно $n \times m$ и $m \times (n \times m)$. Нетрудно убедиться, что

$$\Pi \vec{\varphi} = \Phi \vec{\pi}. \tag{5.2}$$

Используя уравнение связи (5.I) и соотношение (5.2), получим логарифм функции правдоподобия в следующем виде:

$$L(\vec{g}, P) = (\vec{g} - \Pi\vec{\varphi})^T W_g (\vec{g} - \Pi\vec{\varphi}) + (\vec{p} - \vec{\pi})^T W_p (\vec{p} - \vec{\pi}), \tag{5.3}$$

где $W_g = (\mathcal{D}\vec{g})^{-1}$ и $W_p = (\mathcal{D}\vec{p})^{-1}$. Оценки коэффициентов $\vec{\varphi}$ и величин $\vec{\pi}$ находятся из минимизации L (5.3) по $\vec{\varphi}$ и $\vec{\pi}$, т.е. являются решениями уравнений:

$$\frac{\partial L}{\partial \vec{\varphi}} = 0, \quad \frac{\partial L}{\partial \vec{\pi}} = 0.$$

Получаем следующие решения:

$$\vec{\varphi} = [\Pi^T W_g \Pi]^{-1} \Pi^T W_g \vec{g}, \tag{5.4}$$

$$\vec{\pi} = [\Phi^T W_g \Phi + W_p]^{-1} [\Phi^T W_g \vec{g} + W_p \vec{p}] \tag{5.5}$$

или, учитывая (5.2),

$$\Pi = \Phi [\Phi^T W_g \Phi + W_p]^{-1} [\Phi^T W_g \vec{g} + W_p \vec{p}] \vec{\varphi}^T (\vec{\varphi}\vec{\varphi}^T)^{-1}. \tag{5.6}$$

Подставляя (5.6) в (5.4), находим искомое уравнение для $\vec{\varphi}$, которое в явном виде в общем случае не решается. Поэтому для нахождения вектора коэффициентов градуировочной кривой можно использовать итерационный процесс:

$$\Pi^{(0)} = P; \quad \vec{\varphi}^{(0)} = [P^T W_g P]^{-1} P^T W_g \vec{g} \tag{5.7}$$

$$\Pi^{(k+1)} = \Phi^{(k)}\left[\Phi^{(k)T} W_q \Phi^{(k)} + W_p\right]^{-1}\left[\Phi^{(k)T} W_q \vec{q} + W_p \vec{p}\right]\vec{\varphi}^{(k)T}\left(\vec{\varphi}^{(k)}\vec{\varphi}^{(k)T}\right)^{-1}$$

$$\varphi^{(k+1)} = \left[\Pi^{(k+1)T} W_q \, \Pi^{(k+1)}\right]^{-1} \Pi^{(k+1)T} W_q \vec{q} \quad .$$

Нетрудно указать условия, при которых этот процесс сходится.

Определенные трудности возникают при нахождении ковариационной матрицы $\mathcal{D}\vec{\varphi}^{(n)}$ вектора оценок коэффициентов градуировочной кривой на k-м ($k = 0, 1, \dots$) итерационном шаге из-за случайности элементов матрицы P . Воспользуемся разложением в ряд в окрестности точки $[M\vec{p}, M\vec{q}]$ с точностью до членов I-го порядка:

$$\vec{\varphi}^{(k)} \approx \vec{\varphi}_0^{(k)} + G_0(\vec{q} - M\vec{q}) + H_0(\vec{p} - M\vec{p}) .$$

Тогда

$$\mathcal{D}\vec{\varphi}^{(k)} = G_0 W_q G_0^T + H_0 W_p H_0^T \quad , \tag{5.8}$$

где G_0 – матрица размерности $n \times (n \times m)$ со строками вида $(grad_{\vec{p}}\ \varphi_i^{(k)})^T$, $i = 1, \dots, n$, а H_0 – матрица размерности $n \times m$ со строками вида $(grad_{\vec{q}}\ \varphi_i^{(k)})^T$, $i = 1, \dots, n$, причем все производные берутся в точке $[M\vec{p}, M\vec{q}]$.

Для решения второй статистической задачи можно привлечь метод максимального правдоподобия, аналогично тому, как это было сделано выше. Вообще говоря, для решения обеих задач можно воспользоваться также менее точными, но более простыми методами. Выбор наиболее приемлемого с практической точки зрения метода заслуживает дальнейшего исследования

Авторы выражают благодарность Ц.Д.Вылову за плодотворные обсуждения.

Литература

1. Г.Шеффе. Дисперсионный анализ. М., "Наука", 1980.
2. Д.Худсон. Статистика для физиков. М., "Мир", 1970.
3. Г.Крамер. Математические методы статистики. М., "Мир", 1975.
4. P.Meier. Variance of a weighted mean. Biometrics, 1953, v.9, p.59.

Рукопись поступила в издательский отдел
21 августа 1980 года.