ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Сборник трудов
1965 г. Института математики СО АН СССР Выпуск I9

ОБ ОДНОЙ ЗАДАЧЕ ОПТИМАЛЬНОГО РАСПОЗНАВАНИЯ ОБРАЗОВ

А.Е. Заславский, Н.М. Сычева

В разного рода прикладных вопросах часто возникает задача проверки гипотез в условиях, когда вероятностная природа альтернатив представлена не априори известными распределениями, а лишь выборками весьма ограниченного объема, причем результат эксперимента является реализацией случайного процесса. Одной такой задаче распознавания образов для процессов и посвящается настоящая работа.

Изучается возможность построения алгоритмического критерия на примере диагностики сердечных заболеваний по баллистокардиограмме (БКГ). Метод калибровки БКГ, введенный К.П.Бутейко [I], позволяет учитывать при диагностике не только форму БКГ, но и величину отклонений по ординате. Удается выделить конечную совокупность параметров, достаточно полно характеризующих БКГ. Использованный экспериментальный материал позволяет считать, что каждый из этих параметров имеет распределение, близкое к нормальному. Это дает возможность, оперируя стандартными методами теории проверки статистических гипотез, построить близкий к оптимальному алгоритмический критерий для постановки диагноза "стенокардия" при альтернативной гипотезе "здоров". Вероятность того, что поставленный диагноз будет ошибочным, не превышает, в среднем, I0%.

Аналогично может быть построен универсальный алгоритмический критерий для постановки диагноза в случае, когда заранее выбирается не один, а несколько видов сердечных заболеваний ([2], стр. 165-171). Этим же способом может быть решен ряд других задач распознавания образов, представленных конечным числом реализаций случайного процесса.

В § I рассмотрены некоторые общие вопросы построения наиболее мощных критериев для проверки статистических гипотез, относящихся к случайным процессам. В §§ 2-3 излагаются способы описания рассматриваемых процессов с помощью конечного числа параметров и обсуждаются возможности сокращения количества таких параметров. В §§ 4-5 строится критерий, близкий к наиболее мощному в статистической задаче, возникающей при замене реализаций процессов наборами выбранных параметров. Приводятся таблицы, позволяющие использовать построенные критерии, и числовой пример. В § 6 изложены соображения вероятностного характера, заставляющие стремиться к сокращению числа параметров, если объем экспериментального материала ограничен.

Экспериментальным материалом служили БКГ, предоставленные Институтом цитологии и генетики СО АН СССР: 100 БКГ больных людей с диагнозом "стенокардия" и 93 БКГ здоровых людей. Диагноз "стенокардия" установлен различными способами работниками института, предоставившего БКГ. Здоровыми считались люди, у которых не обнаружены стандартные симптомы сердечных заболеваний. У нас не было иного выхода, как принять допущение, что у этих людей действительно нет сердечных заболеваний, хотя, может быть, это и не всегда было так.

Мы считаем своим долгом выразить глубокую благодарность заведующему отделом теории вероятностей и математической статистики Института математики СО АН СССР доктору физико-математических наук А.А.Боровкову за математическую постановку задачи и руководство ее решением. Рядом ценных бесед по вопросам, связанным с медицинской стороной задачи, мы обязаны заведующему лабораторией функциональной диагностики Института цитологии и генетики СО АН СССР кандидату медицинских наук К.П.Бутейко. Его горячая убежденность в необходимости разработки точных методов диагностики, собственно, и побудила нас рассмотреть эту задачу. Сотруднице лаборатории функциональной диагностики А.Г.Крисковец мы благодарны за помощь в обработке экспериментального материала.

§ I. Общие вопросы построения критерия

БКГ представляет собой (рис. I) запись смещения тела человека, скорости этого смещения и его ускорения, вызываемых сокращением сердечной мышцы и непостоянством тока крови в сосудах. Одновременно с этими тремя кривыми на ленте записывается электрокардиограмма, позволяющая с достаточной точностью определять начало и конец цикла сердечной деятельности. Так как три величины - смещение, скорость и ускорение - связаны функционально, то исследование достаточно проводить по одной из них. Всюду в дальнейшем под БКГ подразумевается запись скорости смещения тела человека.

БКГ является реализацией некоторого случайного процесса. Критерий для различения двух гипотез (больной, здоровый) можно построить в том случае, когда мера (распределение вероятностей) μ_0 в пространстве реализаций этого процесса для больных не совпадает с мерой μ_1 в пространстве реализаций для здоровых. Как показывает врачебная практика, такое несовпадение мер действительно имеет место.

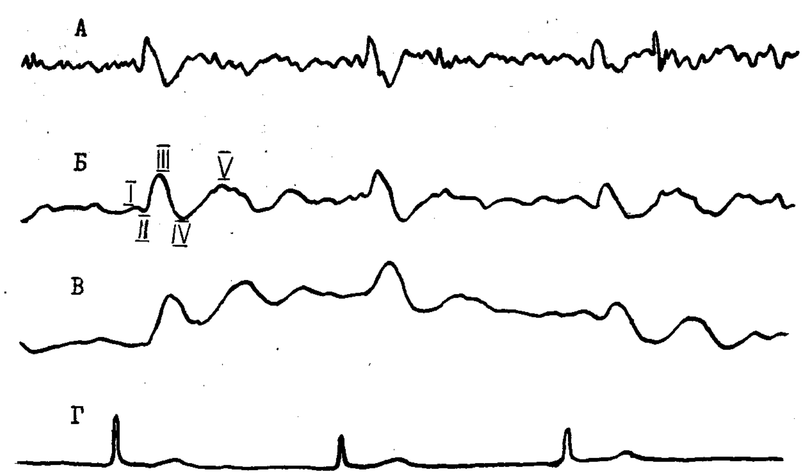

Рис. I. Баллистокардиограмма:

А - ускорение; Б - скорость; В - смещение;
Г - электрокардиограмма.

Задача построения наиболее мощного - оптимального - критерия (н.м.к.) для различения двух альтернативных гипотез заключается в следующем ([3], стр. 36-37).[x] При заданном α $(0 \leq \alpha \leq 1)$ рассматривается класс $\mathcal{M}$ таких множеств $\mathfrak{Z}$ реализаций процесса, для которых $\int\limits_{\mathfrak{Z}} d\mu_0$ существует и равен α. В этом классе выбирается критическое множество $\mathfrak{Z}_{кр}$, для которого максимизируется величина $1-\beta = \int\limits_{\mathfrak{Z}_{кр}} d\mu_1$ (для простоты можно считать, что этот интеграл существует для всех $\mathfrak{Z} \in \mathcal{M}$); иными словами, при всех $\mathfrak{Z} \in \mathcal{M}$ выполняется неравенство

$$1-\beta = \int\limits_{\mathfrak{Z}_{кр}} d\mu_1 \geq \int\limits_{\mathfrak{Z}} d\mu_1 .$$

Альтернативные гипотезы в нашей конкретной задаче означают диагнозы "стенокардия" (H^0) и "здоров" (H^1). Критическое множество $\mathfrak{Z}_{кр}$ полностью определяет оптимальный критерий для различения этих двух гипотез. Именно, если наблюдённая реализация (БКГ) оказывается принадлежащей множеству $\mathfrak{Z}_{кр}$, то принимается гипотеза H^1 - "здоров", если реализация не принадлежит $\mathfrak{Z}_{кр}$, то принимается гипотеза H^0 - диагноз "стенокардия". При этом вероятность того, что в действительности больной человек будет признан здоровым (ошибка первого рода) равна заданному значению α; вероятность того, что здоровый человек будет признан больным (ошибка второго рода) равна β. Если задаваться различными значениями α вероятности ошибки первого рода, будут получаться различные критические множества $\mathfrak{Z}_{кр}$ и различные значения β вероятности ошибки второго рода. Однако критическое множество $\mathfrak{Z}_{кр}$ определяет критерий, обладающий минимальным среди всех возможных (при данном α) значением β вероятности ошибки второго рода - в этом и заключается оптимальность наиболее мощного критерия. Если построенный таким способом н.м.к. дает неудовлетворительную (слишком большую) вероятность β ошибки второго рода при заданном α, то это означает, что различие между мерами μ_0 и μ_1 невелико, и никакой другой критерий для различения двух гипотез не сможет дать лучших результатов.

При построении н.м.к. основную роль играет функция $f(\xi) = \dfrac{d\mu_1(\xi)}{d\mu_0(\xi)}$ - производная меры μ_1 по мере μ_0 (ξ - реализация процесса). Предварительное рассмотрение предоставленных нам БКГ и сам способ их

x) Математическая часть этого параграфа при первом чтении может быть опущена.

получения дают основания отнести нашу конкретную задачу к категории случаев, когда меры μ_0 и μ_1 абсолютно непрерывны друг относительно друга. В формальных выкладках это позволяет считать, что функция $f(\xi)$ измерима относительно мер μ_0 и μ_1. Тогда, как известно ([3], стр. 36-37), множество $\mathfrak{Z}_{кр}$, определяющее н.м.к., описывается при помощи функции $f(\xi)$:

$$\mathfrak{Z}_{кр} = \{ \xi : f(\xi) \geq c \} .$$

Участвующая в определении $\mathfrak{Z}_{кр}$ константа c находится из условия $\int\limits_{\mathfrak{Z}_{кр}} d\mu_0 = \alpha$. Для существования этого интеграла, а также интеграла $\int\limits_{\mathfrak{Z}_{кр}} d\mu_1$ и нужна предполагаемая нами измеримость функции $f(\xi)$ относительно мер μ_0 и μ_1.

Таким образом, для построения н.м.к. мы должны определить меры μ_0 и μ_1 в бесконечномерном пространстве реализаций случайного процесса. Определить эти меры точно не представляется возможным. Однако используя тот факт, что реализации ξ непрерывны, можно найти приближение к н.м.к. любой степени точности. Иными словами, при заданной вероятности ошибки первого рода можно построить критерий другой природы, не использующий всю информацию, содержащуюся в реализации ξ процесса, но такой, что для него вероятность ошибки второго рода будет сколь угодно близка к вероятности β ошибки второго рода, которую дает н.м.к., основанный на функции $f(\xi)$. Для этого следует выделить некоторые n параметров, достаточно полно характеризующих БКГ, определить меры (распределения вероятностей) m_0 и m_1 в n -мерном пространстве значений этой совокупности параметров соответственно для больных и здоровых и построить н.м.к. для различения двух гипотез, основанный на этих n -мерных случайных величинах. Требуемые оценки мер m_0 и m_1 мы надеялись получить с помощью предоставленного нам экспериментального материала.

Заметим, что практическое использование оптимального н.м.к. может потребовать большой вычислительной работы. В то же время незначительное отклонение от оптимальности может позволить существенно сократить объем необходимых вычислений. В связи с этим мы руководствовались следующим: с одной стороны, критерий должен быть близок к оптимальному н.м.к. и, с другой стороны, объем вычислений, необходимых при его использовании, должен быть не слишком большим, таким, чтобы критерий нашел применение в медицинской практике. Важно также, чтобы используемые в критерии параметры имели четкие и достаточно простые правила определения их по БКГ.

§ 2. Выделение параметров

В выделении характеризующих БКГ параметров возможны различные подходы. БКГ представляет собой случайный процесс с явно выраженной периодичностью. Можно выделить какой-либо определенный участок реализации, например цикл сердечной деятельности, выбрать систему координат (о способе построения системы координат будет сказано ниже) и взять в качестве упомянутых параметров ординаты БКГ в точках разбиения выделенного участка на n равных частей. Таким образом пространство реализаций ξ будет заменено пространством ступенчатых функций. Очевидно, что, взяв n достаточно большим, можно добиться возможности сколь угодно точного восстановления БКГ по этим n параметрам. Отсюда, как естественно ожидать ([3], стр. 37-40), следует возможность построения критерия, сколь угодно близкого к н.м.к. Однако такой способ выбора параметров неудовлетворителен, потому что (в силу непрерывности реализаций ξ процесса) коэффициент корреляции между любыми двумя соседними значениями ординат с увеличением n стремится к единице. Стало быть, информация, которую несет каждая такая ордината, становится чрезвычайно малой, и, следовательно, большой объем вычислений, необходимый для использования критерия с большим n, является неоправданным. Целесообразным представляется такой выбор параметров БКГ, когда каждый из них будет нести возможно бóльшую информацию относительно случайной величины $f(\xi)$ при гипотезах H^0 и H^1. Отсюда с необходимостью вытекает, что чем менее коррелированы эти параметры между собой и чем существеннее различия в их распределениях при гипотезах H^0 и H^1, тем лучше. Кроме того, есть соображения принципиального характера, заставляющие строить критерий для различения двух гипотез на возможно меньшем числе параметров. Об этих соображениях будет сказано в § 6.

После опробования большого числа различных вариантов совокупности параметров БКГ мы находим удовлетворительными следующие два способа определения этой совокупности. Отметим, что некоторые из указанных ниже параметров давно используются специалистами по БКГ при визуальной диагностике.

Первый более простой способ предусматривает выбор некоторого цикла БКГ. Различные циклы БКГ, вообще говоря, несколько отличаются друг от друга. Среди них выбирается наиболее "типичный", наиболее характерный для данной БКГ цикл; процедура выбора такого цикла не всегда очевидна, однако ради простоты мы принимаем, что такой выбор всегда возможен. Затем на выбранном цикле строится система координат. Ось абсцисс проводится визуально, так, чтобы затухающая часть цикла была расположена по обе стороны от оси абсцисс и отклонения в положительном и отрицательном направлениях по ординате были в среднем одинаковы. Некоторый элемент произвола, здесь содержащийся, на результатах окончательных вычислений не сказывается. За начало координат принимается начало цикла сердечной деятельности, определяемое моментом скачка электрокардиограммы. Конец цикла определяется следующим моментом скачка.

Из рис. I видно, что на БКГ отчетливо выделяются пять положительных и отрицательных пиков I, II, III, IУ, У. Остальные пики (затухающая часть цикла) выделяются менее отчетливо и на некоторых БКГ отсутствуют вообще. В качестве параметров БКГ (см. рис. 2,3) берутся следующие I3 величин (единица измерения – миллиметр):

I) Разности x_1, x_2, x_3, x_4, x_5 между абсциссами следующих друг за другом пиков.

2) Ординаты x_6, x_7, x_8, x_9, x_{10} пиков.

3) Расстояние x_{11} от точки Q до конца цикла. (Точка Q определяется следующим образом: расстояние от точки P пересечения БКГ с осью абсцисс на участке между пиками IУ и У до проекции на ось абсцисс вершины пика У откладывается вправо от этой проекции; полученная точка и есть точка Q).

4) Вариации x_{12} и x_{13} на участках от начала цикла до вершины пика У и от точки Q до конца цикла соответственно.

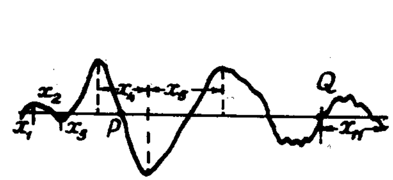

Рис.2. БКГ больного с диагнозом "стенокардия".

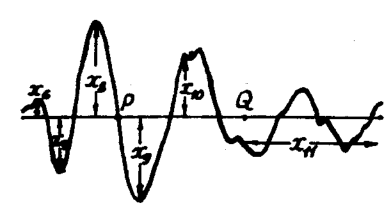

Рис. 3. БКГ здорового человека.

В силу достаточной гладкости БКГ начальный участок цикла восстанавливается по перечисленным параметрам достаточно точно. Как мы убедились, наиболее существенная информация, которую несет конечный участок цикла, содержится в вариации БКГ на этом участке. Например, попытки включить в совокупность параметров абсциссы и ординаты следующих двух пиков и спектральные характеристики БКГ не привели к значительному улучшению результатов. Точка Q введена для того, чтобы учитывать вариацию БКГ именно на затухающем участке цикла.

Если учесть сказанное, то станет ясным, что построенный на этих параметрах критерий должен быть близок к н.м.к.

Второй способ определения параметров БКГ заключается в следующем. В центральной части БКГ выбирается пять соседних циклов, на каждом из них указанным выше способом строится система координат и замеряются те же I3 величин. В качестве параметров берутся средние арифметические этих величин по пяти их замерам. Обозначения для параметров остаются теми же, что и в первом способе, лишь сверху появляется черта.

§ 3. Дополнительные соображения об отборе параметров

Критерий для различения двух гипотез может быть основан как на полной совокупности выделенных параметров, так и на любой части этой совокупности. Критерии IA и IБ основаны на полных совокупностях второго и первого способов соответственно. Однако практическое использование этих критериев требует большой вычислительной работы. Поэтому мы построили также критерии, основанные на меньшем количестве параметров. Ниже приводятся некоторые соображения, которыми мы руководствовались при их отборе.

Исследование эмпирических распределений параметров - как первого, так и второго способов - показывает, что распределение каждого из них в первом приближении можно считать нормальным. Стало быть, то же можно сказать и о совместном распределении всех I3 параметров. Но многомерное нормальное распределение полностью определяется вектором средних значений и матрицей вторых моментов.

В таблице I приведены эмпирические средние и эмпирические дисперсии (диагональные элементы матрицы вторых моментов) по совокупностям больных (I00 БКГ) и здоровых (93 БКГ) для I3 параметров, определенных по второму способу. В таблицах 2,3 приведены оценки коэффициентов корреляции между параметрами, определенными вторым способом с помощью тех же совокупностей БКГ больных и здоровых. Отметим, что для более достоверного суждения о коэффициентах корреляции необходимо исследовать больший объем экспериментального материала (500 - I000 наблюдений).

Из рассмотрения указанных таблиц можно сделать следующие выводы. Эмпирические средние параметров $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5$ для больных и здоровых различаются мало, существенно различаются только эмпирические дисперсии этих параметров (у больных эмпирические дисперсии больше, чем у здоровых). Оценки для коэффициентов корреляции показывают, что эти пять параметров мало зависят друг от друга; в этом случае оценка для дисперсии суммы параметров (такой суммой для разностей абсцисс является абсцисса $\bar{x}_{14} = \bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4 + \bar{x}_5$ пика У) приближенно равна сумме оценок для дисперсий отдельных слагаемых.

Таблица I

№ параметра	Эмпирич. средние		Эмпирич. дисперсии	
	для больных	для здоровых	для больных	для здоровых
I	4,73	4,86	2,04	I,46
2	3,29	3,I6	I,08	0,79
3	3,36	3,7I	0,89	0,39
4	5,37	5,04	I,44	0,8I
5	6,43	6,03	3,40	I,I2
6	2,36	2,23	0,9I	I,I0
7	-2,57	-4,85	I,95	3,00
8	4,93	9,97	3,26	6,78
9	-4,07	-7,87	2,74	4,9I
I0	3,54	6,09	I,5I	3,66
II	I2,54	I5,20	50,I7	46,68
I2	33,7I	60,05	97,43	I68,96
I3	7,39	I2,59	2I,40	55,23

Распределения параметров $\bar{x}_6$ и $\bar{x}_{11}$ для больных и здоровых различаются мало. Параметр $\bar{x}_{12}$ сильно коррелирован с ординатами пиков и вследствие этого не несет существенной информации о БКГ, которая не содержалась бы уже в совокупности значений ординат.

Т а б л и ц а 2

Коэффициенты корреляции между параметрами, определенными вторым способом (I00 БКГ больных)

	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I	I,00	- 0,07	-0,04	-0,04	0,05	0,06	0,29	-0,30	0,3I	-0,04	-0,I2	-0,24	-0,I4
2	-0,07	I,00	0,43	-0,I9	-0,I7	0,29	-0,24	-0,I6	0,I0	-0,I3	-0,I2	0,0I	-0,I7
3	-0,04	0,43	I,00	-0,I5	-0,I5	0,30	-0,38	0,II	-0,I8	0,08	-0,09	0,29	-0,07
4	-0,04	-0,I9	-0,I5	I,00	0,2I	-0,02	0,00	-0,05	-0,0I	0,30	-0,24	0,04	-0,I9
5	0,05	-0,I7	-0,I5	0,2I	I,00	-0,02	-0,06	0,I5	-0,04	0,27	-0,59	0,2I	-0,57
6	0,06	0,29	0,30	-0,02	-0,02	I,00	-0,27	0,I6	-0,I4	0,36	-0,27	0,47	0,04
7	0,29	-0,24	-0,38	0,00	-0,06	-0,27	I,00	-0,43	0,4I	-0,29	0,20	-0,68	-0,03
8	-0,30	-0,I6	0,II	-0,05	0,I5	0,I6	-0,43	I,00	-0,58	0,4I	0,02	0,79	0,27
9	0,3I	0,I0	-0,I8	-0,0I	-0,04	-0,I4	0,4I	-0,58	I,00	-0,60	0,I0	-0,75	-0,I8
I0	-0,04	-0,I3	0,08	0,30	0,27	0,36	-0,29	0,4I	-0,60	I,00	-0,43	0,68	-0,02
II	-0,I2	-0,I2	-0,09	-0,24	-0,59	-0,27	0,20	0,02	0,I0	-0,43	I,00	-0,30	0,69
I2	-0,24	0,0I	0,29	0,04	0,2I	0,47	-0,68	0,79	-0,75	0,68	-0,30	I,00	0,I3
I3	-0,I4	-0,I7	-0,07	-0,I9	-0,57	0,04	-0,03	0,27	-0,I8	-0,02	0,69	0,I3	I,00

Т а б л и ц а 3

Коэффициенты корреляции между параметрами, определенными вторым способом (93 БКГ здоровых)

	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I	I,00	-0,64	-0,II	0,04	-0,II	0,I5	0,I2	-0,29	0,23	-0,27	0,09	-0,04	0,I9
2	-0,64	I,00	0,20	0,I3	0,I5	-0,II	0,I0	0,02	0,0I	-0,03	-0,06	-0,I2	-0,04
3	-0,II	0,20	I,00	-0,08	0,28	-0,07	0,I4	-0,I3	-0,25	0,05	0,08	-0,02	-0,02
4	0,04	0,I3	-0,08	I,00	-0,0I	0,I0	-0,05	-0,I3	0,I7	0,07	-0,03	-0,02	-0,05
5	-0,II	0,I5	0,28	-0,0I	I,00	0,07	-0,02	0,I2	-0,20	-0,0I	-0,35	0,I7	-0,33
6	0,I5	-0,II	-0,07	0,I0	0,07	I,00	-0,34	0,34	-0,I6	0,32	0,02	0,46	0,07
7	0,I2	0,I0	0,I4	-0,05	-0,02	-0,34	I,00	-0,6I	0,42	-0,32	-0,03	-0,70	-0,05
8	-0,29	0,02	-0,I3	-0,I3	0,I2	0,34	-0,6I	I,00	-0,64	0,65	-0,I9	0,82	-0,02
9	0,23	0,0I	-0,25	0,I7	-0,20	-0,I6	0,42	-0,6I	I,00	-0,59	0,20	-0,72	0,03
I0	-0,27	-0,03	0,05	0,07	-0,0I	0,32	-0,32	0,65	-0,59	I,00	-0,I7	0,67	0,05
II	0,09	-0,06	0,08	-0,03	-0,35	0,02	-0,03	-0,I9	0,20	-0,I7	I,00	-0,25	0,45
I2	-0,04	-0,I2	-0,02	-0,02	0,I7	0,46	-0,70	0,82	-0,72	0,67	-0,25	I,00	0,I9
I3	0,I9	-0,04	-0,02	-0,05	-0,33	0,07	-0,05	-0,02	0,03	0,05	0,45	0,I9	I,00

Изложенные соображения относятся ко второму способу определения параметров, который, как естественно ожидать, дает лучшие критерии. Аналогичные соображения справедливы и для первого способа определения параметров.

Руководствуясь этими соображениями, наряду с критериями IА и IБ, основанными на полных совокупностях из I3 параметров второго и первого способов, мы построили критерии 3А и 3Б, основанные на совокупностях параметров $\bar{x}_7, \bar{x}_8, \bar{x}_9, \bar{x}_{10}, \bar{x}_{13}, \bar{x}_{14}$ и $x_7, x_8, x_9, x_{10}, x_{13}, x_{14}$ соответственно. Кроме того, приводятся критерии, основанные на некоторых других неполных совокупностях параметров. Отметим, что с увеличением числа параметров критерий улучшается; критерии, основанные на совокупностях параметров, определенных вторым способом, значительно лучше критериев, основанных на аналогичных совокупностях параметров, определенных первым способом. Поэтому второй способ определения параметров мы считаем основным.

§ 4. Метод построения критерия

Метод построения критерия один и тот же для любой совокупности параметров.

Предполагается, что значение совокупности n параметров БКГ является реализацией n - мерной нормальной случайной величины. Распределение этой случайной величины задается характеристической функцией

$$\varphi(t)=\varphi(t_1,t_2,\dots,t_n)=$$
$$=exp\left\{ i\sum_{j=1}^{n}a_j t_j-\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\tau_{jk}t_j t_k\right\},$$

где a_j - средние значения, $\tau_{jk}(=\tau_{kj})$ - центральные вторые моменты ([4], стр. 342). Если распределение n -мерной нормальной случайной величины невырожденно (определитель матрицы (τ_{jk}) отличен от нуля), то плотность распределения выражается формулой:

$$p(x/H)=p(x_1,x_2,\dots,x_n/a_j,\lambda_{jk})=$$
$$=\frac{1}{(2\pi)^{n/2}\sqrt{\Lambda}}exp\left\{-\frac{1}{2\Lambda}\sum_{j=1}^{n}\sum_{k=1}^{n}\Lambda_{jk}(x_j-a_j)(x_k-a_k)\right\},$$

где $\Lambda_{jk}/\Lambda=\lambda_{jk}$ - элементы матрицы, обратной к матрице (τ_{jk}), Λ - определитель матрицы (τ_{jk}), H означает одну из гипотез, очевидно, определяемую в этом случае набором чисел a_j, λ_{jk}.

Аналогом производной $f(\xi)$ меры μ_1 по мере μ_0 в конечномерном случае является отношение правдоподобия

$$\frac{p(x/H^1)}{p(x/H^0)}$$

(x - реализация конечномерной случайной величины). Если известны наборы чисел a_j^0, λ_{jk}^0 и a_j^1, λ_{jk}^1, определяющие гипотезы H^0 и H^1, то н.м.к. строится без труда ([4], стр. 575-577).

Нам приходится оценивать эти числа с помощью экспериментального материала. Эффективные оценки (см., например, [4], стр. 547) дают

$$\hat{a}_j^i=\frac{1}{n_i}\sum_{m=1}^{n_i}x_{jm}^i, \qquad \hat{\tau}_{jk}^i=\sum_{m=1}^{n_i}(x_{jm}^i-\hat{a}_j^i)(x_{km}^i-\hat{a}_k^i).$$

Здесь x_{jm}^i - значение j -го параметра для m -й БКГ из i -й группы (i = 0 - больной, i = I - здоровый). Оценки $\hat{\lambda}_{jk}^i$ параметров λ_{jk}^i получаются путем вычисления матриц, обратных к матрицам $\hat{\tau}_{jk}^i$.

В качестве статистики для построения н.м.к. может быть использована любая монотонная функция от отношения правдоподобия, например:

$$\ell(x)=\ell(x_1,x_2,\dots,x_n)=2\ln\frac{p(x/H^1)}{p(x/H^0)}-\frac{1}{2}\ln\frac{\Lambda^0}{\Lambda^1}=$$
$$=\sum_{j=1}^{n}\sum_{k=1}^{n}\hat{\lambda}_{jk}^0(x_j-\hat{a}_j^0)(x_k-\hat{a}_k^0)-\sum_{j=1}^{n}\sum_{k=1}^{n}\hat{\lambda}_{jk}^1(x_j-\hat{a}_j^1)(x_k-\hat{a}_k^1)=$$
$$=\Sigma^0-\Sigma^1. \tag{I}$$

Построение критерия завершается теперь выбором константы c , зависящей от α , такой, что в случае $\ell(x)\geqslant c$ принимается гипотеза H^1 , в противном случае - гипотеза H^0. Соотношение $\ell(x)=c$ определяет в n - мерном пространстве параметров некоторую гиперповерхность второго порядка, разделяющую все пространство на две области - s_{kp} и дополнение к s_{kp}. Интегрирование функции $p(x/H^0)$ по области s_{kp} должно дать вероятность α ошибки первого рода - этим условием определяется величина c (и вместе с ней область s_{kp}). Такой путь точного определения c является сложным. Для упрощения можно оценивать распределения самой статистики $\ell(x)$ при гипотезах H^0 и H^1 . Так как $\ell(x)$ есть сумма большого числа слагаемых, то ее распределение должно хорошо описываться нормальным распределением. Исследования эмпирических распределе-

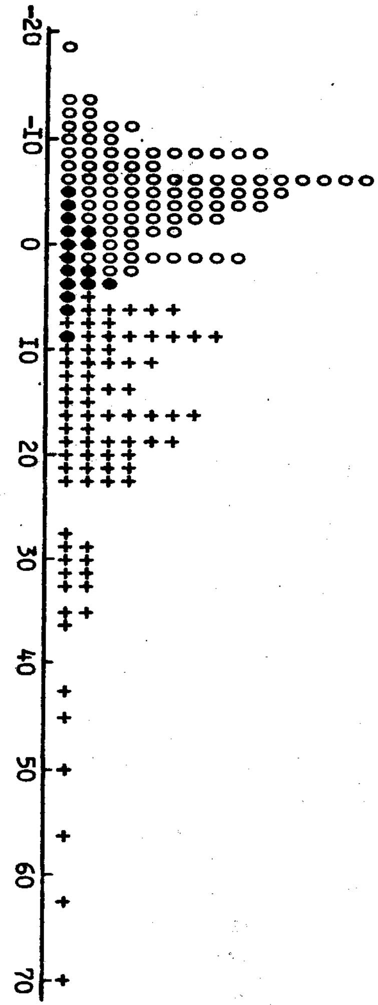

Рис. 4. Гистограмма статистики $\ell(x)$ для критерия 3А. Число знаков о (соответственно +) в столбце равно числу ЕКГ больных (соответственно здоровых), для которых статистика $\ell(x)$ попадает в данный интервал значений. Знак ● означает попадание в один и тот же интервал значений статистики $\ell(x)$ для больного и статистики $\ell(x)$ для здорового.

ний статистики $\ell(x)$ показали, что это действительно так, если не учитывать отдельные выбросы. Поэтому принимается, что с вероятностью p_i $\ell = \xi^i$, с вероятностью $q_i = 1-p_i$ $\ell = \eta^i$, где ξ^i – нормальная случайная величина со средним a_i и дисперсией σ_i^2, а η^i – случайная величина, относительно которой предполагается лишь, что ее значения располагаются значительно левее a_i при $i = 0$ и значительно правее a_i при i = I. Гистограмма статистики $\ell(x)$ для критерия 3А приведена на рис. 4.

В сделанных предположениях свойства распределения случайных величин η^i не оказывают влияния на критическое значение c и вероятность β ошибки второго рода. Критическое значение c определяется по вероятности α ошибки первого рода соотношением

$$\frac{p_0}{\sqrt{2\pi}\,\sigma_0}\int_c^{\infty} e^{-\frac{(t-a_0)^2}{2\sigma_0^2}}\,dt = \alpha, \qquad (2)$$

а вероятность ошибки второго рода

$$\beta = \frac{p_1}{\sqrt{2\pi}\,\sigma_1}\int_{-\infty}^{c} e^{-\frac{(t-a_1)^2}{2\sigma_1^2}}\,dt. \qquad (3)$$

Параметры p_i, a_i, σ_i оценивались следующим образом. С помощью критерия для непринятия резко выделяющихся наблюдений такие наблюдения исключались. Если m_i – число исключенных наблюдений, а n_i – число всех наблюдений, то $p_i = (n_i - m_i)/n_i$, a_i и σ_i – среднее и стандартное отклонение из неисключенных наблюдений.

§ 5. Числовые данные для применения критерия

Для применения критерия, как видно из формулы (I), необходимы следующие величины: оценки $\hat{a}_j^i$ средних значений параметров для больных и здоровых; оценки $\hat{\lambda}_{jk}^i$ элементов матрицы, обратной к матрице вторых моментов параметров; критическое значение c, зависящее от заданной вероятности α ошибки первого рода. Эти величины и приводятся ниже. Значения c и β приведены для двух значений α : α = 5% и α = 10%. Для вычисления значений c (формула (2)) и β (формула (3)) при других значениях α приводятся оценки параметров p_i, a_i, σ_i. Отметим еще раз, что нижний индекс j означает номер параметра, а верхний индекс i = 0 (соответственно i = I), если параметр характеризует ЕКГ больных (соответственно здоровых).

КРИТЕРИЙ IA

Критерий основан на параметрах $\bar{x}_1 - \bar{x}_{13}$.

I. Оценки $\hat{a}_j^i$.

$i \backslash j$	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
0	4,73	3,29	3,36	5,37	6,43	2,36	-2,57	4,93	-4,07	3,54	I2,54	33,7I	7,39
I	4,86	3,I6	3,7I	5,04	6,03	2,23	-4,85	9,97	-7,87	6,09	I5,20	60,05	I2,59

2. Оценки $\hat{\lambda}^o_{jk}$ (для больных)

$j \backslash k$	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I	0,67	0,I5	-0,07	0,I3	0,05	0,09	-0,30	0,27	-0,3I	0,07	0,00	-0,09	0,03
2	0,I5	I,5I	-0,45	0,I8	0,I7	-0,39	0,20	0,I3	-0,07	0,I6	0,03	0,0I	0,06
3	-0,07	-0,45	I,72	0,I0	0,2I	-0,03	0,08	0,I6	-0,02	0,09	-0,04	-0,I0	0,I0
4	0,I3	0,I8	0,I0	0,9I	0,03	0,I2	0,0I	0,I2	-0,I9	-0,42	0,00	-0,02	0,04
5	0,05	0,I7	0,2I	0,03	0,72	0,34	-0,27	0,04	-0,3I	-0,0I	0,02	-0,I2	0,I6
6	0,09	-0,39	-0,03	0,I2	0,34	2,94	-I,I2	I,20	-I,22	0,02	-0,06	-0,60	0,08
7	-0,30	0,20	0,08	0,0I	-0,27	-I,I2	2,03	-I,05	0,86	-0,53	0,I0	0,58	-0,I8
8	0,27	0,I3	0,I6	0,I2	0,04	I,20	-I,05	2,00	-0,8I	0,47	-0,I6	-0,63	0,I2
9	-0,3I	-0,07	-0,02	-0,I9	-0,3I	-I,22	0,86	-0,8I	I,67	0,24	0,I0	0,48	-0,I3
I0	-0,07	0,I6	0,09	-0,42	-0,0I	0,02	-0,53	0,47	0,24	2,04	0,06	-0,25	-0,03
II	0,00	0,03	-0,04	0,00	0,02	-0,06	0,I0	-0,I6	0,I0	0,06	0,08	0,07	-0,07
I2	-0,09	0,0I	-0,I0	-0,02	-0,I2	-0,60	0,58	-0,63	0,48	-0,25	0,07	0,29	-0,09
I3	0,03	0,06	0,I0	0,04	0,I6	0,08	-0,I8	0,I2	-0,I3	-0,03	-0,07	-0,09	0,I7

3. Оценки $\hat{\lambda}^1_{jk}$ (для здоровых)

$j \backslash k$	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I	2,I8	I,63	-0,03	-0,25	0,23	-0,22	-0,69	0,45	-0,3I	0,59	-0,06	-0,22	0,04
2	I,63	2,86	-0,67	-0,5I	-0,02	-0,25	-0,32	-0,08	-0,II	0,45	0,03	-0,03	-0,04
3	-0,03	-0,67	4,0I	0,26	-0,64	0,I6	-0,29	0,59	0,38	-0,20	-0,I2	-0,08	0,05
4	-0,25	-0,5I	0,26	I,56	-0,II	-0,03	0,25	0,26	-0,I6	-0,35	0,0I	0,00	0,00
5	0,23	-0,02	-0,64	-0,II	I,4I	-0,05	-0,3I	0,06	-0,08	0,36	0,0I	-0,II	0,08
6	-0,22	-0,25	0,I6	-0,03	-0,05	I,45	-0,I3	0,I6	-0,29	-0,I4	-0,09	-0,I4	0,07
7	-0,69	-0,32	-0,29	0,25	-0,3I	-0,I3	I,35	-0,24	0,27	-0,48	0,I5	0,28	-0,I3
8	0,45	-0,08	0,59	0,26	0,06	0,I6	-0,24	0,88	-0,I7	-0,07	-0,09	-0,2I	0,I0
9	-0,3I	-0,II	0,38	-0,I6	-0,08	-0,29	0,27	-0,I7	0,7I	0,03	0,06	0,I6	-0,07
I0	0,59	0,45	-0,20	-0,35	0,36	-0,I4	-0,48	-0,07	0,03	0,88	-0,03	-0,I2	0,04
II	-0,06	0,03	-0,I2	0,0I	0,0I	-0,09	0,I5	-0,09	0,06	-0,03	0,06	0,05	-0,04
I2	-0,22	-0,03	-0,08	0,00	-0,II	-0,I4	0,28	-0,2I	0,I6	-0,I2	0,05	0,I2	-0,06
I3	0,04	-0,04	0,05	0,00	0,08	0,07	-0,I3	0,I0	-0,07	0,04	-0,04	-0,06	0,05

4. Параметры для вычисления c и β по α

i	ρ_i	a_i	σ_i
0	0,94	-I3,84	II,87
I	0,94	24,94	I5,02

5. Значения c и β для двух значений α

α	c	β
5%	5,34	8,9%
I0%	0,97	5,I%

КРИТЕРИЙ IБ

Критерий основан на параметрах $x_1 - x_{13}$.

I. Оценки $\hat{a}^i_j$.

i \ j	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
0	4,59	3,28	3,36	5,55	6,59	2,73	-2,72	4,83	-4,06	3,55	I2,I9	34,26	6,87
I	4,88	3,23	3,63	5,04	5,94	2,I9	-4,77	9,87	-7,82	6,26	I5,I2	59,39	I2,05

2. Оценки $\hat{\lambda}^0_{jk}$ (для больных)

j \ k	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I	0,5I	0,I0	0,07	0,II	-0,0I	0,00	-0,I7	0,I7	-0,08	0,0I	0,00	-0,04	0,00
2	0,I0	0,76	-0,I0	0,I9	0,I7	-0,I0	0,II	0,I3	-0,II	-0,05	0,02	-0,02	0,02
3	0,07	-0,I0	0,0I	0,I8	0,I8	-0,07	0,04	0,I2	-0,03	0,07	0,00	-0,06	0,05
4	0,II	0,I9	0,I8	0,46	0,I7	0,00	-0,06	0,I3	-0,02	-0,03	0,02	-0,04	0,03
5	-0,0I	0,I7	0,I8	0,I7	0,45	0,II	-0,05	0,I0	-0,I4	-0,05	0,05	-0,06	0,07
6	0,00	-0,I0	-0,07	0,00	0,II	0,0I	-0,34	0,34	-0,34	0,02	-0,0I	-0,I7	0,04
7	-0,I7	0,II	0,04	-0,06	-0,05	-0,34	0,76	-0,28	0,I5	-0,I9	0,02	0,I6	-0,03
8	0,I7	0,I3	0,I2	0,I3	0,I0	0,34	-0,28	0,87	-0,22	0,I0	-0,05	-0,2I	0,05
9	-0,08	-0,II	-0,03	-0,02	-0,I4	-0,34	0,I5	-0,22	0,6I	0,03	0,00	0,I4	-0,0I
I0	0,0I	-0,05	0,07	-0,03	-0,05	0,02	-0,I9	0,I0	0,03	0,7I	0,03	-0,08	-0,05
II	0,00	0,02	0,00	0,02	0,05	-0,0I	0,02	-0,05	0,00	0,03	0,04	0,0I	-0,03
I2	-0,04	-0,02	-0,06	-0,04	-0,06	-0,I7	0,I6	-0,2I	0,I4	-0,08	0,0I	0,09	-0,02
I3	0,00	0,02	0,05	0,03	0,07	0,04	-0,03	0,05	-0,0I	-0,05	-0,03	-0,02	0,08

3. Оценки $\hat{\lambda}^1_{jk}$ (для здоровых)

j \ k	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I	I,II	0,75	0,3I	-0,I8	0,0I	-0,I2	-0,37	0,27	-0,I0	0,24	-0,04	-0,II	0,02
2	0,75	I,I4	0,23	-0,3I	-0,I4	-0,22	-0,I2	-0,04	0,09	0,20	0,03	0,0I	-0,02
3	0,3I	0,23	2,I8	0,35	-0,II	-0,2I	-0,33	0,27	0,I5	-0,0I	-0,05	-0,04	0,0I
4	-0,I8	-0,3I	0,35	I,09	0,26	0,II	0,0I	0,26	-0,I5	-0,I4	-0,02	-0,06	0,02
5	0,0I	-0,I4	-0,II	0,26	0,69	-0,0I	-0,I3	0,I4	-0,I0	-0,0I	0,0I	-0,06	0,03
6	-0,I2	-0,22	-0,2I	0,II	-0,0I	0,98	-0,09	0,I3	-0,3I	-0,02	-0,07	-0,I2	0,06
7	-0,37	-0,I2	-0,33	0,0I	-0,I3	-0,09	0,83	-0,32	0,I5	-0,2I	0,II	0,I9	-0,07
8	0,27	-0,04	0,27	0,26	0,I4	0,I3	-0,32	0,76	-0,I8	-0,03	-0,09	-0,20	0,07
9	-0,I0	0,09	0,I5	-0,I5	-0,I0	-0,3I	0,I5	-0,I8	0,52	0,05	0,06	0,I3	-0,05
I0	0,24	0,20	-0,0I	-0,I4	-0,0I	-0,02	-0,2I	-0,03	0,05	0,45	0,00	-0,05	0,00
II	-0,04	0,03	-0,05	-0,02	0,0I	-0,07	0,II	-0,09	0,06	0,00	0,05	0,04	-0,03
I2	-0,II	0,0I	-0,04	-0,06	-0,06	-0,I2	0,I9	-0,20	0,I3	-0,05	0,04	0,09	-0,03
I3	0,02	-0,02	0,0I	0,02	0,03	0,06	-0,07	0,07	-0,05	0,00	-0,03	-0,03	0,03

4. Параметры для вычисления C и β по α

i	ρ_i	a_i	θ_i
0	0,94	-II,82	I0,57
I	0,95	I2,56	9,I0

5. Значения C и β для двух значений α

α	C	β
5%	5,26	20,I%
I0%	I,37	I0,I%

КРИТЕРИЙ 2А

Критерий основан на параметрах $\bar{x}_7, \bar{x}_8, \bar{x}_9, \bar{x}_{10}, \bar{x}_{11}, \bar{x}_{13}, \bar{x}_{14}$.

I. Оценки $\hat{a}_j^i$

$i \backslash j$	7	8	9	I0	II	I3	I4
0	-2,57	4,93	-4,07	3,54	I2,54	7,39	23,I8
I	-4,85	9,97	-7,87	6,09	I5,20	I2,59	22,80

2. Оценки $\hat{\lambda}_{jk}^0$ (для больных)

$j \backslash k$	7	8	9	I0	II	I3	I4
7	0,69	0,I7	-0,I6	-0,08	-0,02	0,00	0,03
8	0,I7	0,53	0,22	-0,I0	-0,0I	-0,04	-0,0I
9	-0,I6	0,22	0,8I	0,54	0,00	-0,0I	-0,I0
I0	-0,08	-0,I0	0,54	I,52	0,I3	-0,I3	-0,I4
II	-0,02	-0,0I	0,00	0,I3	0,06	-0,05	0,02
I3	0,00	-0,04	-0,0I	-0,I3	-0,05	0,I2	0,05
I4	0,03	-0,0I	-0,I0	-0,I4	0,02	0,05	0,22

3. Оценки $\hat{\lambda}_{jk}^1$ (для здоровых)

$j \backslash k$	7	8	9	I0	II	I3	I4
7	0,57	0,25	-0,06	-0,09	0,02	0,00	0,0I
8	0,25	0,43	0,I0	-0,22	0,02	0,00	0,06
9	-0,06	0,I0	0,38	0,I5	-0,0I	0,00	0,03
I0	-0,09	-0,22	0,I5	0,56	0,0I	-0,0I	0,02
II	0,02	0,02	-0,0I	0,0I	0,03	-0,0I	0,02
I3	0,00	0,00	0,00	-0,0I	-0,0I	0,02	0,00
I4	0,0I	0,06	0,03	0,02	0,02	0,00	0,29

4. Параметры для вычисления C и β по α

i	ρ_i	a_i	σ_i
0	0,99	-4,29	5,26
I	0,97	I7,82	II,84

5. Значения C и β для двух значений α

α	C	β
5%	4,36	I2,3%
I0%	2,45	9,4%

КРИТЕРИЙ 2Ь

Критерий основан на параметрах $x_7, x_8, x_9, x_{10}, x_{11}, x_{13}, x_{14}$

I. Оценки $\hat{a}_j^i$

$i \backslash j$	7	8	9	I0	II	I3	I4
0	-2,72	4,83	-4,06	3,55	I2,I9	6,87	23,37
I	-4,77	9,87	-7,82	6,26	I5,I2	I2,05	22,72

2. Оценки $\hat{\lambda}_{jk}^0$ (для больных)

$j \backslash k$	7	8	9	I0	II	I3	I4
7	0,38	0,II	-0,I0	-0,05	-0,0I	0,00	0,0I
8	0,II	0,35	0,I0	-0,II	-0,0I	0,0I	0,02
9	-0,I0	0,I0	0,36	0,I3	-0,0I	0,02	-0,03
I0	-0,05	-0,II	0,I3	0,6I	0,05	-0,06	-0,06
II	-0,0I	-0,0I	-0,0I	0,05	0,04	-0,03	0,03
I3	0,00	0,0I	0,02	-0,06	-0,03	0,07	0,02
I4	0,0I	0,02	-0,03	-0,06	0,03	0,02	0,I7

3. Оценки $\hat{\lambda}_{jk}^1$ (для здоровых)

$j \backslash k$	7	8	9	I0	II	I3	I4
7	0,34	0,I2	-0,07	-0,06	0,0I	0,00	-0,02
8	0,I2	0,28	0,08	-0,I3	0,0I	0,00	0,03
9	-0,07	0,08	0,29	0,I0	0,00	-0,0I	0,00
I0	-0,06	-0,I3	0,I0	0,36	0,02	-0,02	0,00
II	0,0I	0,0I	0,00	0,02	0,03	-0,0I	0,02
I3	0,00	0,00	-0,0I	-0,02	-0,0I	0,02	0,00
I4	-0,02	0,03	0,00	0,00	0,02	0,00	0,22

4. Параметры для вычисления C и β по α

i	ρ_i	a_i	σ_i
0	0,99	-3,23	4,72
I	0,99	II,20	8,77

5. Значения C и β для двух значений α

α	C	β
5%	4,35	2I,6%
I0%	2,79	I6,7%

КРИТЕРИЙ 3А

Критерий основан на параметрах $\bar{x}_7, \bar{x}_8, \bar{x}_9, \bar{x}_{10}, \bar{x}_{13}, \bar{x}_{14}$.

I. Оценки $\hat{a}_j^i$

$i \backslash j$	7	8	9	I0	I3	I4
0	-2,57	4,93	-4,07	3,54	7,39	23,I8
I	-4,85	9,97	-7,87	6,09	I2,59	22,80

2. Оценки $\hat{\lambda}^0_{jk}$ (для больных)

$j \backslash k$	7	8	9	I0	I3	I4
7	0,68	0,I7	-0,I5	-0,02	-0,02	0,04
8	0,I7	0,53	0,22	-0,09	-0,05	-0,0I
9	-0,I5	0,22	0,8I	0,54	-0,0I	-0,II
I0	-0,02	-0,09	0,54	I,25	-0,02	-0,I9
I3	-0,02	-0,05	-0,0I	-0,02	0,08	0,07
I4	0,04	-0,0I	-0,II	-0,I9	0,07	0,22

3. Оценки $\hat{\lambda}^1_{jk}$ (для здоровых)

$j \backslash k$	7	8	9	I0	I3	I4
7	0,55	0,24	-0,06	-0,I0	0,0I	-0,0I
8	0,24	0,42	0,II	-0,22	0,0I	0,05
9	-0,06	0,II	0,38	0,I6	0,00	0,04
I0	-0,I0	-0,22	0,I6	0,55	-0,0I	0,02
I3	0,0I	0,0I	0,00	-0,0I	0,0I	0,00
I4	-0,0I	0,05	0,04	0,02	0,00	0,2I

4. Параметры для вычисления c и β по α

i	ρ_i	a_i	σ_i
0	I,00	-4,24	4,97
I	0,99	I5,I6	I0,97

5. Значения c и β для двух значений α

α	c	β
5%	3,94	I5,0%
I0%	2,I3	II,4%

КРИТЕРИЙ 3Б

Критерий основан на параметрах $x_7, x_8, x_9, x_{10}, x_{13}, x_{14}$.

I. Оценки $\hat{a}_j^i$

$i \backslash j$	7	8	9	I0	I3	I4
0	-2,72	4,83	-4,06	3,55	6,87	23,37
I	-4,77	9,87	-7,82	6,26	I2,05	22,72

2. Оценки $\hat{\lambda}^0_{jk}$ (для больных)

$j \backslash k$	7	8	9	I0	I3	I4
7	0,37	0,II	-0,I0	-0,04	-0,0I	0,02
8	0,II	0,34	0,I0	-0,09	0,00	0,03
9	-0,I0	0,I0	0,35	0,I4	0,0I	-0,02
I0	-0,04	-0,09	0,I4	0,55	-0,03	-0,I0
I3	-0,0I	0,00	0,0I	-0,03	0,05	0,04
I4	0,02	0,03	-0,02	-0,I0	0,04	0,I5

3. Оценки $\hat{\lambda}^1_{jk}$ (для здоровых)

$j \backslash k$	7	8	9	I0	I3	I4
7	0,34	0,II	-0,07	-0,06	0,0I	-0,03
8	0,II	0,28	0,09	-0,I4	0,00	0,02
9	-0,07	0,09	0,29	0,II	-0,0I	0,00
I0	-0,06	-0,I4	0,II	0,34	-0,0I	-0,0I
I3	0,0I	0,00	-0,0I	-0,0I	0,02	0,0I
I4	-0,03	0,02	0,00	-0,0I	0,0I	0,2I

4. Параметры для вычисления c и β по α

i	ρ_i	a_i	σ_i
0	0,99	-3,0I	4,I4
I	0,99	I0,33	8,60

5. Значения c и β для двух значений α

α	c	β
5%	3,80	22,2%
I0%	2,30	I7,4%

КРИТЕРИЙ 4А

Критерий основан на параметрах $\overline{x}_7, \overline{x}_8, \overline{x}_9, \overline{x}_{10}, \overline{x}_{13}$.

I. Оценки $\hat{a}_j^i$

$i \backslash j$	7	8	9	I0	I3
0	-2,57	4,93	-4,07	3,54	7,39
I	-4,85	9,97	-7,87	6,09	I2,59

2. Оценки $\hat{\lambda}_{jk}^0$ (для больных)

$j \backslash k$	7	8	9	I0	I3
7	0,68	0,I7	-0,I4	0,0I	-0,03
8	0,I7	0,53	0,22	-0,09	-0,04
9	-0,I4	0,22	0,76	0,44	0,03
I0	0,0I	-0,09	0,44	I,08	0,04
I3	-0,03	-0,04	0,03	0,04	0,05

3. Оценки $\hat{\lambda}_{jk}^1$ (для здоровых)

$j \backslash k$	7	8	9	I0	I3
7	0,55	0,24	-0,06	-0,09	0,0I
8	0,24	0,4I	0,I0	-0,22	0,0I
9	-0,06	0,I0	0,37	0,I5	-0,0I
I0	-0,09	-0,22	0,I5	0,55	-0,0I
I3	0,0I	0,0I	-0,0I	-0,0I	0,02

4. Параметры для вычисления C и β по α

i	ρ_i	a_i	σ_i
0	I,00	-2,85	4,44
I	0,98	I4,22	I0,I6

5. Значения C и β для двух значений α

α	C	β
5%	4,45	I6,6%
I0%	2,84	I2,8%

КРИТЕРИЙ 4Б

Критерий основан на параметрах $x_7, x_8, x_9, x_{10}, x_{13}$.

I. Оценки $\hat{a}_j^i$

$i \backslash j$	7	8	9	I0	I3
0	-2,72	4,83	-4,06	3,55	6,87
I	-4,77	9,87	-7,82	6,26	I2,05

2. Оценки $\hat{\lambda}_{jk}^0$ (для больных)

$j \backslash k$	7	8	9	I0	I3
7	0,37	0,II	-0,I0	-0,03	-0,0I
8	0,II	0,34	0,I0	-0,07	-0,0I
9	-0,I0	0,I0	0,35	0,I3	0,02
I0	-0,03	-0,07	0,I3	0,49	0,00
I3	-0,0I	-0,0I	0,02	0,00	0,03

3. Оценки $\hat{\lambda}_{jk}^1$ (для здоровых)

$j \backslash k$	7	8	9	I0	I3
7	0,34	0,II	-0,07	-0,06	0,0I
8	0,II	0,27	0,09	-0,I4	0,00
9	-0,07	0,09	0,29	0,II	-0,0I
I0	-0,06	-0,I4	0,II	0,34	-0,0I
I3	0,0I	0,00	-0,0I	-0,0I	0,02

4. Параметры для вычисления C и β по α

i	ρ_i	a_i	σ_i
0	I,00	-2,05	3,80
I	0,99	9,76	7,97

5. Значения C и β для двух значений α

α	C	β
5%	4,20	24,0%
I0%	2,82	I9,0%

КРИТЕРИЙ 5А

Критерий основан на параметрах $\overline{x}_7, \overline{x}_8, \overline{x}_9, \overline{x}_{10}$.

I. Оценки $\hat{a}^i_j$

$i \backslash j$	7	8	9	I0
0	-2,57	4,93	-4,07	3,54
I	-4,85	9,97	-7,87	6,09

2. Оценки $\hat{\lambda}^o_{jk}$ (для больных)

$j \backslash k$	7	8	9	I0
7	0,66	0,I4	-0,I2	0,04
8	0,I4	0,50	0,24	-0,06
9	-0,I2	0,24	0,74	0,42
I0	0,04	-0,06	0,42	I,04

3. Оценки $\hat{\lambda}^1_{jk}$ (для здоровых)

$j \backslash k$	7	8	9	I0
7	0,55	0,24	-0,05	-0,09
8	0,24	0,40	0,I0	-0,22
9	-0,05	0,I0	0,39	0,I5
I0	-0,09	-0,22	0,I5	0,55

4. Параметры для вычисления c и β по α

i	ρ_i	a_i	σ_i
0	I,00	-2,22	3,00
I	0,97	II,20	8,94

5. Значения c и β для двух значений α.

α	c	β
5%	2,98	I7,9%
I0%	I,70	I4,I%

КРИТЕРИЙ 5Б

Критерий основан на параметрах x_7, x_8, x_9, x_{10}

I. Оценки $\hat{a}^i_j$

$i \backslash j$	7	8	9	I0
0	-2,72	4,83	-4,06	3,55
I	-4,77	9,87	-7,82	6,26

2. Оценки $\hat{\lambda}^o_{jk}$ (для больных)

$j \backslash k$	7	8	9	I0
7	0,37	0,I0	-0,09	-0,03
8	0,I0	0,34	0,II	-0,07
9	-0,09	0,II	0,34	0,I3
I0	-0,03	-0,07	0,I3	0,49

3. Оценки $\hat{\lambda}^1_{jk}$ (для здоровых)

$j \backslash k$	7	8	9	I0
7	0,33	0,I2	-0,07	-0,06
8	0,I2	0,27	0,09	-0,I4
9	-0,07	0,09	0,29	0,I0
I0	-0,06	-0,I4	0,I0	0,34

4. Параметры для вычисления c и β по α

i	ρ_i	a_i	σ_i
0	I,00	-2,22	3,00
I	0,99	8,II	7,94

5. Значения c и β для двух значений α

α	c	β
5%	2,72	24,6%
I0%	I,63	20,4%

С увеличением числа параметров, на которых основан критерий, возрастает объем необходимой вычислительной работы и уменьшается вероятность ошибочного решения. Этим можно руководствоваться при выборе того или иного критерия для практического использования.

Объем вычислительной работы может быть проиллюстрирован следующим примером. Пусть решено использовать критерий 5А с вероятностью ошибки первого рода α = I0%, которой соответствует критическое значение C = I,70. По БКГ определены значения параметров: $\bar{x}_7$ = 0,00; $\bar{x}_8$ = 3,40; $\bar{x}_9$ = −4,80; $\bar{x}_{10}$ = 3,40. Вычисления удобно располагать следующим образом.

i = 0 (гипотеза H^0) i = I (гипотеза H^1)

Разности $\bar{x}_j - \hat{a}_j^i$

j	7	8	9	I0
	2,57	−I,53	−0,73	−0,I4

j	7	8	9	I0
	4,85	−6,57	3,07	−2,69

Произведения $(\bar{x}_j - \hat{a}_j^i)(\bar{x}_k - \hat{a}_k^i)$

$j \backslash k$	7	8	9	I0
7	6,60	−3,93	−I,88	−0,36
8	−3,93	2,34	I,I2	0,2I
9	−I,88	I,I2	0,53	0,I0
I0	−0,36	0,2I	0,I0	0,02

$j \backslash k$	7	8	9	I0
7	23,52	−3I,86	I4,89	−I3,05
8	−3I,86	43,I6	20,I7	I7,67
9	I4,89	20,I7	9,42	−8,26
I0	−I3,05	I7,67	−8,26	7,23

Произведения $\hat{\lambda}_{jk}^i(\bar{x}_j - \hat{a}_j^i)(\bar{x}_k - \hat{a}_k^i)$

$j \backslash k$	7	8	9	I0
7	4,36	−0,55	0,23	−0,0I
8	−0,55	I,I7	0,27	−0,0I
9	0,23	0,27	0,39	0,04
I0	−0,0I	−0,0I	0,04	0,02

$j \backslash k$	7	8	9	I0
7	I2,94	−7,65	−0,74	I,I7
8	−7,65	I7,26	−2,02	−3,89
9	−0,74	−2,02	3,48	−I,24
I0	I,I7	−3,89	−I,24	3,98

Суммы по столбцам

4,03	0,88	0,93	0,04	5,72	3,70	−0,52	0,02

Σ^0 = 5,88 Σ^1 = 8,92

$\ell(x)$ = 5,88 − 8,92 = −3,04

Сравнивая полученное значение $\ell(x)$ с критическим значением C = I,70, принимаем решение, что обследуемый пациент болен стенокардией. Отметим, что решение, предоставленное критерием, совпадает с диагнозом, поставленным работниками Института цитологии и генетики.

В заключительный период работы Институт цитологии и генетики предоставил нам дополнительно 32 БКГ пациентов с диагнозом "стенокардия". Эти БКГ были проверены по критерию 3А с вероятностями ошибок α = I0%, β = II,4%. 28 из них критерий отнес к категории больных стенокардией, 4 – к категории здоровых людей. Этот результат неплохо согласуется с принятым значением α = I0% вероятности ошибки первого рода.

§ 6. Дополнительные замечания

Рассмотрим соображения, заставляющие при ограниченном объеме экспериментального материала стремиться к сокращению числа параметров, на которых основывается критерий для различения двух альтернативных гипотез.

Соотношение $\ell(x_1, x_2, \dots, x_n) = C$ определяет в n-мерном пространстве параметров некоторую гиперповерхность порядка не выше второго. Эта гиперповерхность разделяет все пространство на две части, одна из которых принимается за критическое множество $S_{кр}$. Гипотеза H^0 отвергается, если вновь наблюденная точка попадает в $S_{кр}$, и принимается – в противном случае. Сама гиперповерхность строится по экспериментальному материалу так, чтобы наилучшим образом разделить уже наблюденные точки, даваемые экспериментальным материалом, – реализации для больных и здоровых. Именно, рассматриваются все гиперповерхности порядка не выше второго, для которых в $S_{кр}$ попадает доля α наблюденных точек для больных, и затем из них выбирается такая гиперповерхность, для которой в $S_{кр}$ попадает максимальная доля I − β наблюденных точек для здоровых. Выборочные (точечные) распределения "разделяются" лучше, чем непрерывные теоретические распределения(и в том и в другом случае построение н.м.к. и есть построение гиперповерхности, разделяющей распределения наилучшим образом). Так, для двух одинаковых распределений в n -мерном пространстве при α = 50% с необходимостью β = 50%. Если же рассмотреть две независимые выборки объема m реализаций одной и той же n -мерной случайной величины и по ним (так же, как мы это делали выше) построить н.м.к., то при $\alpha = 50\%$ с вероятностью I будет $\beta < 50\%$. При этом если зафиксировать n (число параметров) и неограниченно увеличивать m (число наблюдений в группе), то, очевидно, $\beta \to 50\%$. Если же фиксировать m и увеличивать n, то β убывает и уже при $n = m$ с вероятностью I будет $\beta = 0$ (не только при $\alpha = 50\%$, но даже и при $\alpha = 0$). Действительно, n точек в n-мерном пространстве лежат на $(n-I)$-мерной гиперплоскости. Вероятность того, что хотя бы одна из $2n$ реализаций будет находиться на пересечении двух гиперплоскостей, содержащих первую и вторую группы по n реализаций, равна нулю. Поэтому с вероятностью I найдется гиперповерхность порядка не выше второго, полностью разделяющая эти группы (например, две достаточно близкие друг к другу параллельные гиперплоскости, такие что одна из указанных выше гиперплоскостей лежит между ними).

В связи с изложенным применительно к нашей конкретной задаче возникает следующий вопрос. Действительно ли полученные нами критерии основаны на существенных различиях между БКГ больных стенокардией и БКГ здоровых людей, или причиной хорошего разделения больных и здоровых по БКГ является слишком большое для использованного экспериментального материала число параметров. Это прежде всего относится к критериям IА и IБ, основанным на I3 параметрах каждый.

Чтобы проиллюстрировать положение, которое здесь может возникнуть, приведем результаты одного эксперимента. Исследовались наборы из 4,6 и I3 параметров. В каждом случае были получены I93 реализации нормальной n-мерной случайной величины из одной и той же генеральной совокупности с матрицей вторых моментов и вектором средних, равными эмпирической матрице вторых моментов и эмпирическому вектору средних для здоровых. Эти реализации были разделены на две группы - I00 "больных" и 93 "здоровых" - и по ним строились н.м.к. для различения двух гипотез, точно так же, как это было сделано по параметрам БКГ. Полученные в математическом эксперименте критерии характеризуются следующими вероятностями ошибок:

Число параметров	α	β
4	39,4 %	39,4 %
6	37,8 %	37,8 %
I3	23,9 %	23,9 %

Эти данные показывают, что эффект, вызываемый ограниченностью экспериментального материала, может играть весьма существенную роль. Но если выборки получены из совокупностей с "разделенными" распределениями, то этот эффект будет слабее и при значениях вероятностей ошибок 5-I0% будет мало искажать результат.

Однако для уточнения критериев необходимо провести расчеты по гораздо большему (скажем, по 500 - I000 в каждой из групп) числу БКГ. Это связано с тем, что достаточно надежные оценки вторых центральных моментов многомерной случайной величины, которые используются при построении критериев, получаются именно при таком числе наблюдений ([5], гл. 4).

Таким образом, алгоритмическая диагностика сердечных заболеваний по БКГ безусловно возможна. При альтернативе "болен стенокардией" или "здоров" она обеспечивает в среднем 9 правильных решений из I0 при диагнозе "стенокардия" и столько же при диагнозе "здоров". Критерий, обеспечивающий такие вероятности ошибок, определяется формулой (I). Он близок к оптимальному статистическому критерию, и заметно лучший результат при использовании БКГ невозможен. Аналогично может быть построен универсальный алгоритмический критерий для нескольких (а не одного) видов сердечных заболеваний.

Изложенным методом могут быть решены также некоторые другие практические задачи распознавания образов, представленных реализациями случайных процессов.

ЛИТЕРАТУРА

I. Бутейко К.П. Баллистокардиограф смещения, скорости и ускорения с калибровочным устройством. Первая Всесоюзная конференция по баллистокардиографии. Тез. докл. М., I959, стр. I7.
2. Блекуэлл Д., Гиршик М.А. Теория игр и статистических решений. М., I958.
3. Гренандер У. Случайные процессы и статистические выводы. М., I96I.
4. Крамер Г. Математические методы статистики. М., I948.
5. Андерсон Т. Введение в многомерный статистический анализ. М., I963.

Поступила в редакцию
2.IX.I965