Результаты приведены в табл. 2. Анализ данных этой таблицы показывает, что случайная погрешность измерения площади не превосходит 2,5%, причем при увеличении площади объекта относительная ошибка уменьшается. Погрешность определения размеров, как нетрудно показать, имеет в два раза меньшую величину, и среднеквадратичное отклонение ошибки радиуса частиц составляет 1—1,2%.

## ЛИТЕРАТУРА


1. Автоматический анализ цитологических препаратов. Под ред. А. Я. Хесина. Рига, «Зинатне», 1975.
2. Бурый Л. В., Коронкевич В. П., Нестерихин Ю. Е., Нестеров А. А., Пушной Б. М., Ткач С. Е., Щербаченко А. М. Прецизионный фотограмметрический автомат.— «Автометрия», 1974, № 4, с. 83—89.
3. Киричук В. С., Косых В. П., Перетягин Г. И. Восстановление распределения интенсивности светового потока в считывающем луче сканирующей системы.— «Автометрия», 1977, № 3, с. 57—65.


*Поступила в редакцию 18 января 1977г.*

---

УДК 519.2 : 62-50

**И. В. СМЕРТИНЮК**

*(Новосибирск)*

# ПРОВЕРКА ПОВТОРЯЕМОСТИ В ФИЗИЧЕСКИХ ЭКСПЕРИМЕНТАХ

Начальная стадия постановки научного эксперимента (в частности, физического) характеризуется рядом особенностей, которые на последующих этапах либо отсутствуют, либо выражены гораздо слабее. Поскольку цель научного эксперимента — получение неизвестных заранее характеристик (количественных либо качественных) исследуемых явлений (процессов), достаточно устойчиво повторяющихся при одинаковых условиях, представляется очень важным, прежде чем перейти к определению и уточнению получаемых характеристик, убедиться в том, что они действительно повторяются. Следовательно, необходимость проверки повторяемости — важнейшая особенность начального этапа при проведении экспериментов. Проверка повторяемости необходима также в случаях, когда имеется ряд серий экспериментальных данных, полученных при исследовании одного и того же процесса. Серии могут отличаться друг от друга условиями проведения эксперимента (например, различная степень влияния внешней среды) или просто могут быть получены на разных экспериментальных установках. Совместная статистическая обработка всех имеющихся данных может значительно повысить точность результатов. Однако предварительно нужно убедиться в том, что все серии действительно отличаются друг от друга только случайными вариациями.

Обычно этот этап — проверка повторяемости — производится с помощью качественных соображений и общетеоретических представлений. При обработке больших массивов экспериментальных данных автоматизированными комплексами, включающими в себя ЭВМ, необходимо иметь для той же цели количественные критерии.

В этой заметке ограничимся рассмотрением только таких экспериментов, где результат измерений может быть представлен в виде

$$x_i(\tau)=m_i(\tau)+\xi_i(\tau) \quad (i=\overline{1,\ p}). \tag{1}$$

Здесь $x_i(\tau)$ — измеренная величина; $m_i(\tau)$ — неизвестная нам характеристика исследуемого процесса; $\tau$ — свободный параметр (обычно время); $\xi_i(\tau)$ — случайная величина, распределенная по нормальному закону: $\xi_i(\tau)\in N(0,\ \sigma^2)$; $i$ — номер эксперимента; $p$ — общее число экспериментов. Свободный параметр $\tau$ может быть как непрерывным, так и дискретным. Задача сводится к построению критерия, позволяющего по измеренным значениям $x_i(\tau)$ определить, идентичны ли функции $m_i(\tau)$ друг другу.

Рассмотрим вначале дискретный случай. Выражение (1) запишется следующим образом:

$$x_i(\tau_j)=m_i(\tau_j)+\xi_i(\tau_j) \quad (i=\overline{1,\ p};\ j=\overline{1,\ q}), \tag{2}$$

где $\xi_i(\tau_j)$ попарно не коррелированы. Иными словами, взаимно независимые случайные величины $x_i(\tau_j)$ распределены по нормальному закону с математическим ожиданием $M\{x_i(\tau_j)\}=m_i(\tau_j)$ и дисперсией $\sigma^2$.

Требуется установить справедливость равенств

$$m_1(\tau_j)=m_2(\tau_j)=\ldots=m_p(\tau_j)=m(\tau_j) \quad (j=\overline{1,\ q}), \tag{3}$$

где $m(\tau_j)$ — постоянные, остающиеся неизвестными.

Эту задачу можно решить с помощью следующей статистики [1]:

$$S=\frac{q\sum\limits_{i=1}^{p}\bar{y}_i^2}{\frac{1}{q-1}\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{q}(y_{ij}-\bar{y}_i)^2}. \tag{4}$$

Здесь

$$y_{ij}=x_i(\tau_j)-\frac{1}{p}\sum_{s=1}^{p}x_s(\tau_j); \tag{5}$$

$$\bar{y}_i=\frac{1}{q}\sum_{j=1}^{q}y_{ij}. \tag{6}$$

Гауссовские величины $y_{ij}$ и $\bar{y}_i$ обладают, как нетрудно видеть из выражений (2) и (3), нулевыми математическими ожиданиями и дисперсиями:

$$D\{y_{ij}\}=\frac{p-1}{p}\sigma^2=\sigma_y^2; \quad D\{\bar{y}_i\}=\frac{1}{q}\sigma_y^2. \tag{7}$$

В числителе выражения (4) стоит величина

$$Q_1^2=q\sum_{i=1}^{p}\bar{y}_i^2, \tag{8}$$

которая, если ее предварительно пронормировать по $\sigma^2$, подчиняется распределению $\chi^2$ с $p-1$ степенями свободы. Знаменатель (4), точнее, величина

$$G^2=(q-1)\,Q_2^2=\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij}-\bar{y}_i)^2 \tag{9}$$

после предварительной нормировки по $\sigma^2$ будет распределена по $\chi^2$

с $(p-1)(q-1)$ степенями свободы. Величины $Q_1$ и $Q_2$ статистически независимы.

В таком случае статистика $S$, которую можно представить в виде

$$S = \frac{Q_1^2}{p-1} : \frac{G^2}{(p-1)(q-1)}, \tag{10}$$

будет распределена по закону Фишера $F[(p-1), (p-1)(q-1)]$. Это распределение не зависит ни от $\sigma^2$, ни от $m(\tau)$. Гипотеза (3) принимается при выполнении условия

$$c_1 \leqslant S \leqslant c_2. \tag{11}$$

Числа $c_1$ и $c_2$ выбираются из уравнения

$$\int_{c_1}^{c_2} F(S)\, dS = 1 - \alpha, \tag{12}$$

где $\alpha$ — уровень значимости, $F(S)$ — функция плотности распределения Фишера. Недостающее условие для однозначного решения (12) можно задать следующим образом:

$$\int_0^{c_1} F(S)\, dS = \int_{c_2}^{\infty} F(S)\, dS. \tag{13}$$

Перейдем теперь к рассмотрению непрерывного случая, когда свободный параметр $\tau$ принимает значения в интервале $[0, t]$. Соотношение (3), справедливость которого требуется установить, теперь принимает вид

$$m_1(\tau) = m_2(\tau) = \ldots = m_p(\tau) = m(\tau) \quad (\forall \tau \in [0, t]), \tag{14}$$

где $m(\tau)$ — функция, остающаяся неизвестной. Здесь в качестве исходной модели возьмем гауссовский «белый шум» (1):

$$M\{\xi_i(\tau)\} = 0; \tag{15}$$

$$M\{\xi_i(\tau)\xi_i(\tau')\} = \sigma^2\delta(\tau - \tau'), \tag{16}$$

где $\delta(\tau-\tau')$ — дельта-функция.

Случайный процесс $\int_0^t \xi_i(\tau)\, d\tau$ будет, как известно, винеровским, с нулевым математическим ожиданием и дисперсией

$$D\left\{\int_0^t \xi_i(\tau)\, d\tau\right\} = \sigma^2 t. \tag{17}$$

Приращения винеровского процесса статистически независимы на непересекающихся интервалах:

$$M\left\{\left[\int_{\tau_j}^{\tau_j+\Delta\tau} \xi_i(\tau)\, d\tau\right]\left[\int_{\tau_k}^{\tau_k+\Delta\tau} \xi_i(\tau)\, d\tau\right]\right\} = 0;$$
$$[\tau_j, \tau_j + \Delta\tau] \cap [\tau_k, \tau_k + \Delta\tau] = \varnothing. \tag{18}$$

При определении статистики для непрерывной модели представляется удобным использовать случайные функции

$$\eta_i(\tau) = x_i(\tau) - \frac{1}{p}\sum_{s=1}^{p} x_s(\tau). \tag{19}$$

В соответствии с выражениями (14)÷(16) математическое ожидание $\eta_i(\tau)$ равно нулю, а корреляционная функция задается выражением

$$M\{\eta_i(\tau)\,\eta_i(\tau')\} = \sigma_y^2\delta(\tau-\tau'), \tag{20}$$

где $\sigma_y^2$ определяется аналогично (7):

$$\sigma_y^2 = \frac{p-1}{p}\sigma^2.$$

Далее поступим следующим образом. Разделим интервал $[0, t]$ на $q$ одинаковых частей длительностью $\Delta\tau = \tau_{j+1} - \tau_j$. Здесь $\tau_j$, $\tau_{j+1}$ — границы полученных подынтервалов, $j = \overline{1, q}$, $\tau_1 = 0$, $\tau_{q+1} = t$. Теперь рассмотрим случайные величины вида

$$z_{ij} = \frac{1}{\sqrt{\Delta\tau}}\int\limits_{\tau_j}^{\tau_{j+1}} \eta_i(\tau)\,d\tau; \tag{21}$$

$$\bar{z}_i = \frac{1}{q}\sum_{j=1}^{q} z_{ij}. \tag{22}$$

Из выражений (15)—(20) видно, что статистические свойства случайных величин $z_{ij}$ и $\bar{z}_i$ те же, что и у соответствующих величин $y_{ij}$ (5) и $\bar{y}_i$ (6). Эти свойства заданы соотношениями (7). Следовательно, подставив (21) и (22) в (4), можно получить критерий проверки идентичности $m_i(\tau)$ в цепочке равенств (14) для непрерывной модели. Прежде чем записать этот тест в явном виде, займемся уточнением выражения (22):

$$\bar{z}_i = \frac{1}{q}\sum_{j=1}^{q}\frac{1}{\sqrt{\Delta\tau}}\int\limits_{\tau_j}^{\tau_{j+1}} \eta_i(\tau)\,d\tau. \tag{23}$$

Легко видеть, что сумма интегралов в (23) может быть записана в более компактной форме:

$$\sum_{j=1}^{q}\int\limits_{\tau_j}^{\tau_{j+1}} \eta_i(\tau)\,d\tau = \int\limits_0^t \eta_i(\tau)\,d\tau. \tag{24}$$

Тогда, учитывая очевидное соотношение $q = t/\Delta\tau$, можно записать

$$\bar{z}_i = \frac{\sqrt{\Delta\tau}}{t}\int\limits_0^t \eta_i(\tau)\,d\tau. \tag{25}$$

Окончательно критерий для проверки справедливости равенств (14) в непрерывной модели имеет вид

$$S = \frac{\dfrac{t-\Delta\tau}{t}\sum\limits_{i=1}^{p}\left(\int\limits_0^t \eta_i(\tau)\,d\tau\right)^2}{\sum\limits_{i=1}^{p}\left\{\left[\sum\limits_{j=1}^{q}\left(\int\limits_{\tau_j}^{\tau_{j+1}}\eta_i(\tau)\,d\tau\right)^2\right] - \dfrac{\Delta\tau}{t}\left(\int\limits_0^t \eta_i(\tau)\,d\tau\right)^2\right\}}, \tag{26}$$

где $\eta_i(\tau)$ определяются по формуле (19).

Рассмотрим предельное соотношение для статистики (26), когда $q$ стремится к бесконечности. Легко видеть, что в этом предельном случае выражение (26) может быть записано следующим образом:

$$S = \frac{\sum_{i=1}^{p} \left( \int_{0}^{t} \eta_i (\tau)\, d\tau \right)^2}{(p-1)\, \sigma^2 t}. \tag{27}$$

Действительно, существует лемма [2], которая утверждает, применительно к условиям и обозначениям нашей задачи, что с вероятностью 1

$$\lim_{q\to\infty} \sum_{j=1}^{q} \left( \int_{\tau_j}^{\tau_{j+1}} \eta_i (\tau)\, d\tau \right)^2 = \sigma_y^2 t. \tag{28}$$

Производя предельный переход в выражении (26) с учетом (28), (20) и (7), получаем соотношение (27). Числитель в (27) распределен аналогично (8), откуда следует, что величина $T=(p-1)S$ в (27) распределена по $\chi^2$ с $p-1$ степенями свободы. Непосредственной проверкой можно установить, что плотность распределения величины $S$ в (27) имеет вид

$$F(S) = \frac{\left(\frac{p-1}{2} S\right)^{\frac{p-1}{2}} \exp\left(-\frac{p-1}{2} S\right)}{S\Gamma((p-1)/2)}, \tag{29}$$

где $\Gamma((p-1)/2)$ — гамма-функция. Математическое ожидание $M\{S\}$ равно единице, а дисперсия определяется по формуле $D\{S\}=2/(p-1)$.

Анализ соотношений (26) и (28) показывает, что выражением (29) для предельной плотности распределения $S$ можно пользоваться при больших $q$ (точнее, при $q\geqslant 10^4$), т. е. когда случайные вариации знаменателя в (26) очень малы по сравнению с его математическим ожиданием.

В этом случае для вычисления $S$ можно использовать более простое выражение, чем (26):

$$S = \frac{\sum_{i=1}^{p} \left( \int_{0}^{t} \eta_i (\tau)\, d\tau \right)^2}{\sum_{i=1}^{p} \sum_{j=1}^{q} \left( \int_{\tau_j}^{\tau_{j+1}} \eta_i (\tau)\, d\tau \right)^2}. \tag{30}$$

Если же известна заранее дисперсия $\sigma^2$ либо достаточно точная ее оценка, то можно воспользоваться соотношением (27). Значение $\sigma^2$ или ее оценки подставляется в знаменатель, а затем нужно перейти к вычислению величины $T=(p-1)S$, которая, как было уже отмечено, распределена по $\chi^2$ с $p-1$ степенями свободы.

## ЛИТЕРАТУРА


1. Дюге Д. Теоретическая и прикладная статистика. М., «Наука», 1972, с. 271.
2. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. М., «Наука», 1974, с. 99.


*Поступила в редакцию 17 сентября 1976 г.*