

ТЕХНИЧЕСКИЕ НАУКИ

EDN FVKWFF
DOI 10.26163/RAEN.2022.32.23.003
УДК 621.519.23:658.562.3

П.М. Винник, Т.В. Винник, Е.А. Еськова

МАТЕМАТИЧЕСКИЕ ЗАДАЧИ, ВОЗНИКАЮЩИЕ ПРИ СТАТИСТИЧЕСКОМ КОНТРОЛЕ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ МАШИНОСТРОИТЕЛЬНЫХ ПРОИЗВОДСТВ

Петр Михайлович Винник – зав. кафедрой высшей математики, БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова, доктор технических наук, доцент, г. Санкт-Петербург; **e-mail:** vinnik_pm@voenmeh.ru.

Татьяна Викторовна Винник – доцент кафедры математики, Санкт-Петербургский государственный технологический институт (Технический университет), кандидат технических наук, г. Санкт-Петербург; **e-mail:** vinnik.tv92@voenmeh.ru.

Екатерина Александровна Еськова – ассистент кафедры высшей математики, БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова, г. Санкт-Петербург; **e-mail:** eskova_ea@voenmeh.ru.

При статистическом контроле массового технологического процесса возникает ряд математических задач, методы решения которых не во всем удовлетворительны. В статье сформулированы некоторые такие задачи и обсуждаются возможные подходы к их решению.

Ключевые слова: теоретический закон распределения; группированные выборки; объединение выборок; зависимость формируемых свойств от исходных свойств.

P.M. Vinnik, T.V. Vinnik, E.A. Eskova

MATHEMATICAL PROBLEMS ARISING IN STATISTICAL CONTROL OF TECHNOLOGICAL PROCESSES IN MACHINE-BUILDING INDUSTRIES

Petr Vinnik – Head of the Department of Higher Mathematics, D. Ustinov Baltic State Technical University (VOENMEH), Doctor of Engineering, associate professor, St. Petersburg; **e-mail:** vinnik_pm@voenmeh.ru.

Tatyana Vinnik – senior lecturer, the Department of Mathematics, St. Petersburg State Technological Institute (Technical University), PhD in Engineering, St. Petersburg; **e-mail:** vinnik.tv92@voenmeh.ru.

Ekaterina Eskova – assistant, the Department of Higher Mathematics, D. Ustinov Baltic State Technical University (VOENMEH), St. Petersburg; **e-mail:** eskova_ea@voenmeh.ru.

A number of mathematical problems arise during the statistical control of a mass technological process. The methods of solving such problems are not entirely acceptable. We present some of the problems in question and discuss possible approaches to their solution.

Keywords: theoretical distribution law; grouped samples; pooling of samples; dependence of generated properties on initial properties.

Ресурс и надежность изделий машиностроительных производств характеризуется соответствием геометрических, механических и прочих свойств продукции технической документации, а также ста-

бильностью этого соответствия [3; 6].

При проверке указанного соответствия в технологическом процессе на контрольных операциях по выборке, взятой из партии изделий, оценивают годность

партии и текущее состояние оборудования

Типовыми математическими задачами, возникающими при таком контроле, являются:

- подбор теоретического закона распределения, с заданной точностью соответствующего эмпирическому закону;
- оценка качества отдельной технологической операции по имеющимся выборкам заготовок до и после технологической операции.

Установление теоретического закона распределения позволяет обоснованно определить условия признания партии годной и уточнить объем выборки, необходимый для принятия партии с требуемым уровнем доверия.

Заметим, что традиционно контрольная операция после технологической операции должна только подтвердить корректность работы технологического оборудования и годность партии изготовленных деталей, что в случае возникновения брака в недопустимых объемах приводит к немедленному вмешательству в технологический процесс.

Однако в последнее время развитие измерительного оборудования, компьютеризация позволяют ставить задачу комплексного анализа технологического процесса с отслеживанием траектории детали через все технологические операции ее изготовления и даже вплоть до завершения ее функционирования.

С такой точки зрения представляет значительный интерес оценка зависимости статистических свойств характеристики изделия после технологической операции от таких же свойств характеристики до этой операции.

Причем результатом такой оценки должна являться возможность расчета того, как уменьшение разброса характеристики до операции скажется на разбросе характеристики после операции.

Подбор теоретического закона распределения для продукции машиностроительного производства

В идеальной математической постановке задачу можно сформулировать так: есть выборка $X = (X_1, \dots, X_k)$, характери-

зующая ту или иную характеристику изделия, выборка считается однородной, ее элементы – независимыми друг от друга. Необходимо подобрать теоретический закон распределения.

Различные критерии согласия (хи-квадрат, отношения правдоподобия, Колмогорова, Смирнова, ω^2 и т.д.) дают возможность проверить гипотезу о теоретическом законе распределения против различных альтернатив.

Выбор теоретического закона в качестве нулевой гипотезы может опираться как на рекомендации общего типа [9] о классе закона для тех или иных конкретных задач, так и на подбор подходящей аппроксимации по кривым Джонсона S_L , S_B и S_U или Пирсона [4].

К сожалению, при реальном статистическом анализе машиностроительной продукции возникает ряд трудностей.

Для простоты будем говорить об операции вытяжки с утонением стенки, широко распространенной в специальном машиностроении и осуществляемой в массовом объеме:

1) в силу индивидуальных особенностей оборудования принятие рекомендованного типа закона, как минимум, требует проверки (см. например, [7, начало гл. 5);

2) выборка существенно зависит от износа инструмента и степени наладки оборудования. Одно и то же оборудование с течением времени будет выдавать изделия с разными статистическими характеристиками. Причем степени износа и наладки трудно измеряемы и в лучшем случае отражаются в документации выборки словами «сразу после наладки», «новый пуансон» или «изношенный пуансон» (см., например, [2, рис. на с. 224).

Эта ситуация показана на рис. 1. По горизонтальной оси отложено общее количество изготовленных на оборудовании деталей (обработанных заготовок) между его переналадками. Интервал A_1 , A_2 обозначает место извлечения первой выборки, интервал B_1 , B_2 – место извлечения второй выборки. Выборки условно показаны как сплошные, хотя могут не предполагать извлечения всех обработанных

заготовок подряд. По вертикальной оси отложена точность изготовления, которую можно примерно оценить величиной, обратно пропорциональной дисперсии элементов выборки. В общем случае неизвестны не только точные значения A_1, A_2, B_1, B_2 , но даже и разности $(A_1 - 0), (B_1 - A_2)$. Напротив, разности $(A_2 - A_1)$ и $(B_2 - B_1)$ для сплошных выборок известны и равны количествам элементов в первой и второй выборках.

Заметим, что изменение точности означает изменение закона распределения рассматриваемой характеристики обработанных заготовок. Предположение, что изменяется тип закона распределения, представляется маловероятным. Такое предположение означало бы, что до заготовки с таким-то номером тип закона распределения один, а начиная с этого номера, – другой, что придавало бы заготовке с таким-то номером особое пограничное значение.

Более разумным кажется предположение, что по мере износа изменяются параметры (параметр) закона распределения;

1) особенности реального производства (применение калибров, округление, оценка годности по альтернативному признаку и т.п.) приводят к тому, что имеющиеся выборки, особенно большого объема, являются группированными (в лучшем случае – частично группированными [5]);

2) как правило, для анализа предлагаются давно извлеченные выборки, границы интервалов группировки которых фиксированы и были установлены из неиз-

вестных производственных соображений (для конкретной выборки обычно интервалы имеют одну и ту же длину).

Оценка качества отдельной технологической операции по имеющимся выборкам заготовок до и после технологической операции

Традиционный подход рассмотрен в [8, с. 231, девятая задача]. Он состоит в построении линии регрессии в предположении наличия двумерной выборки, состоящей из пар измерений свойств одной и той же заготовки до и после технологической операции.

Однако при массовом производстве выборки $X = (X_1, \dots, X_k)$ и $Y = (Y_1, \dots, Y_m)$ могут иметь существенно разное количество элементов, причем отследить изменение свойств конкретного изделия крайне трудно. Потому следует считать, что двумерную выборку вышеуказанного типа получить невозможно.

Математические задачи

Таким образом, возникают следующие математические задачи.

Задача 1. Имеются две группированные выборки разного объема, принадлежащие одному закону распределения с одними и теми же параметрами, но с разными границами группировки. Требуется их объединить в одну выборку.

Задача 2. Определить тип закона распределения по нескольким группированным выборкам, имеющим разный объем (и возможно, разные границы интервалов группировки), и извлеченным в произвольные неизвестные моменты времени между последовательными наладками

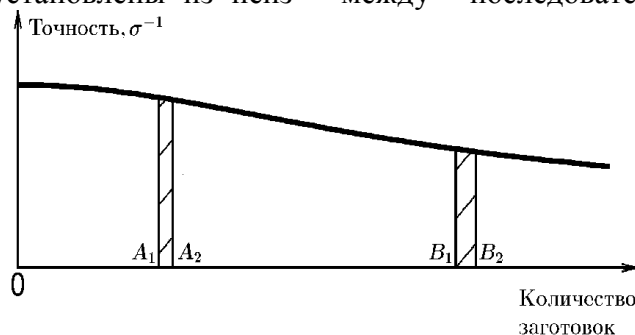


Рис. 1. Извлечение двух выборок из потока обрабатываемых заготовок. Штриховкой показаны места извлечения выборок. Интервал A_1, A_2 обозначает место извлечения первой выборки, интервал B_1, B_2 – место извлечения второй выборки. Выборки условно показаны как сплошные, хотя могут не предполагать извлечения всех обработанных заготовок подряд. Значения A_1, A_2, B_1, B_2 обычно неизвестны

оборудования.

Задача 3. По несогласованным между собой выборкам $X = (X_1, \dots, X_k)$ – исходных свойств и $Y = (Y_1, \dots, Y_m)$ – формируемых в ходе технологической операции свойств требуется установить приближенную аналитическую зависимость рассматриваемой характеристики заготовки после технологической операции от этой же характеристики до операции, позволяющую произвести расчет того, как уменьшение разброса характеристики до операции скажется на разброс характеристики после операции.

Соображения по решению Задачи 1

Заметим, что напрашивающийся путь, состоящий в применении критериев согласия к каждой выборке по отдельности, очевидно, может приводить к ситуации, когда по одной выборке нулевая гипотеза с вычисленным по выборке значением параметра θ_1 принимается, а нулевая гипотеза с вычисленным по другой выборке значением параметра θ_2 отвергается. Считая безусловным совпадение типа распределения выборок, такой ситуации хотелось бы избежать.

Разный объем выборок, разное количество интервалов группировки делают принятие решения в этой ситуации особенно затруднительным.

Как известно, при применении критерия хи-квадрат [1] рекомендуется следить за тем, чтобы в каждом интервале группировки количество наблюдений было бы не слишком малым (например, не менее 10 [7, с. 174]) и сравнимым между разными интервалами, объединяя в случае необходимости, те интервалы, где наблюдений недостаточно, с другими соседними интервалами. В [1, с. 6] указывается, что «на практике обычно строят интервалы равной длины или, в лучшем случае, интервалы равной вероятности».

Поэтому для решения Задачи 1 можно предложить такой план действий.

Для простоты предположим, что среди границ интервалов $x_1 < x_2 < \dots < x_n$ и $y_1 < y_2 < \dots < y_k$ первой и второй выборок нет ни одной совпадающей.

- по разным границам интервалов исходных выборок составить объединенный

упорядоченный перечень границ $z_1 < z_2 < \dots < z_{n+k}$, где каждое z_i является или одним из чисел x_i , или одним из чисел y_i ;

- раздробить каждую из исходных выборок на новые интервалы границами $z_1 < z_2 < \dots < z_{n+k}$. При этом некоторые (или все) исходные интервалы разобьются на две части, а число наблюдений, входивших в исходный интервал, разделится между образуемыми частями пропорционально длинам этих частей. В результате этого получатся две выборки с одинаковыми интервалами разной длины с дробным числом наблюдений в интервалах;

- объединить обе получившиеся выборки в одну;

- выбрать желательное число интервалов группировки;

- определить количество наблюдений (возможно, дробное), которое должно попасть в каждый интервал равной вероятности при выбранном количестве интервалов

- объединить полученные ранее интервалы объединенной выборки, деля в случае необходимости тем же вышеуказанным образом некоторые из этих интервалов, чтобы обеспечить попадание нужного количества наблюдений в каждый интервал.

Поясним этот план действий на примере.

На рис. 2 показаны гистограммы исходных группированных выборок (количество интервалов взято малым для простоты изложения). Границы интервалов первой выборки 0, 10, 20, 30.

Границы интервалов второй выборки – 1, 11, 23, 35.

Объединенный перечень границ интервалов: 1, 0, 10, 11, 20, 23, 30, 35.

Дробление исходных выборок по новым границам интервалов показано на рис. 3.

Заметим, что объединение всех образованных частей каждого интервала приводит к исходной выборке.

Объединение получившихся выборок показано на рис. 4.

Всего в двух исходных выборках имелось 61 наблюдение. Возьмем количество интервалов в новой выборке равным

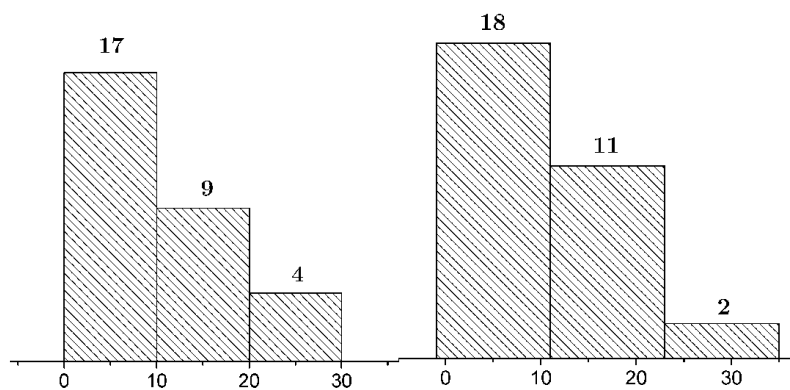


Рис. 2. Исходные выборки

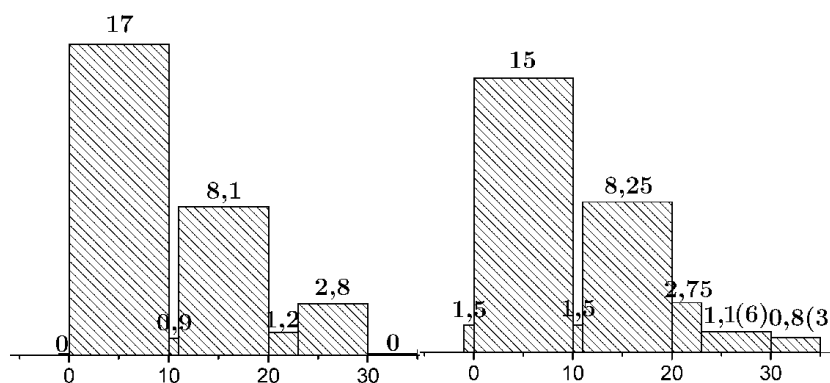


Рис. 3. Дробление исходных выборок по объединенному множеству границ интервалов

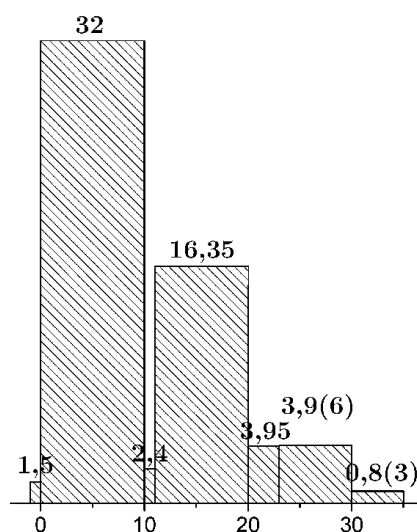


Рис. 4. Объединение полученных раздробленных группированных выборок

четырем и построим равновероятные интервалы. В этом случае в каждом из них должно быть по 15,25 наблюдений.

Делим столбец № 2 так, чтобы в его левой части было $(15,25 - 1,5 = 13,75)$ наблюдений.

Тогда левая часть должна занимать $\frac{13,75}{32} \cdot 10 = 4,296875$. Таким образом, первый новый интервал будет $\langle -1; 4,296875 \rangle$.

От разделенного второго интервала осталось еще 19,75. Значит, второй новый интервал будет $\langle 4,296875; 9,0625 \rangle$. Продолжая аналогично, получаем третий и четвертый интервалы: $\langle 9,0625; 16,422018349 \rangle$ и $\langle 16,422018349; 35 \rangle$. Разумеется, длины интервалов 5,296875; 4,765625; 7,35951835; 18,57798165 обратно пропорциональны величине эмпирической плот-

ности на интервалах.

Получившаяся финальная группированная выборка с равновероятными интервалами показана на рис. 5.

Соображения по решению Задачи 2

Видятся два пути решения Задачи 2.

Во-первых. Безусловное предположение о совпадении типа закона распределения обеих выборок означает, что элементы первой и второй выборок различаются только параметрами закона распределения. Следовательно, вообще говоря, числовые характеристики у них будут различными. Выбрав одну из них, и масштабируя одну или обе выборки, можно добиться совпадения этой эмпирической числовой характеристики.

Например, это может быть какой-нибудь параметр положения или разброса. Однако масштабирование приведет к изменению границ группировки выборки. Полученные масштабированные выборки остается объединить так, как предложено при решении Задачи 1. Затем к полученной объединенной выборке применяется критерий согласия (например, хи-квадрат).

Очевидными недостатками этого пути решения Задачи 2 являются, во-первых, невозможность масштабированием добиться попарного совпадения всех параметров двух выборок (если параметров распределения больше одного), во-вторых, произвол в выборе числовой характеристики, при совпадении которой в результате масштабирования выборки считаются имеющими одно распределение.

Во-вторых. Можно построить «синтетическую» статистику из статистик критериев, построенных по каждой выборке.

Например, вычисляем к каждой выборке по отдельности соответственно статистики $\chi^2_{(1)}$ и $\chi^2_{(2)}$ критерия хи-квадрат. Пусть первая выборка содержит n наблюдений, а вторая – k наблюдений. В качестве «синтетической» статистики возьмем величину

$$\frac{n}{n+k} \chi^2_{(1)} + \frac{k}{n+k} \chi^2_{(2)}. \quad (1)$$

Недостатком такого пути в докомпьютерную эпоху являлось бы то, что предельное распределение статистики (1) будет зависеть от n и k , что требовало бы составления специальных статистических таблиц. Однако в настоящее время компьютерные математические пакеты позволяют производить расчет критического значения статистики критерия при фиксированных n , k и объемах выборок и заданном уровне значимости.

Вторым спорным моментом является то, что статистика (1) не учитывает напрямую количество интервалов группировки в выборках. Например, если обе выборки состоят из 200 наблюдений, заключенных между 0 и 50, но первая имеет 10 интервалов равной длины, а вторая – только 5, то очевидно, что первая выборка содержит больше информации об исследуемой характеристике, чем вторая. Однако в статистике (1) вес выборок будет одинаковым.

Соображения по решению Задачи 3

Заметим, что в идеальном случае, когда вместо выборок свойств $X = (X_1, \dots, X_k)$ и $Y = (Y_1, \dots, Y_n)$ известны их законы распределения $F_{X_i}(x)$ и $F_{Y_j}(x)$, установление функциональной зависимости при допущении ее монотонного характера осуще-

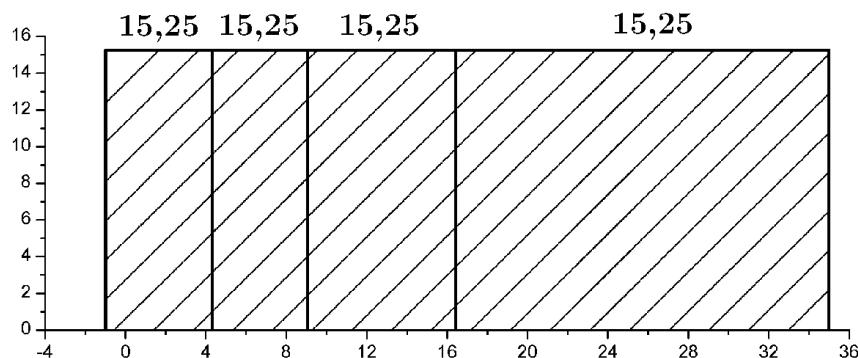


Рис. 5. Финальная группированная выборка с равновероятными интервалами

ствляется по определению функции распределения – из соотношения

$$F_{Y_1}(Y(x)) = F_{X_1}(x).$$

при строго монотонно возрастающей функции $Y = Y(X)$, и из соотношения

$$F_{Y_1}(Y(x)) = 1 - F_{X_1}(x).$$

при строго монотонно убывающей функции $Y = Y(X)$.

Поэтому восстановление (подбор) по выборкам $X = (X_1, \dots, X_k)$ и $Y = (Y_1, \dots, Y_n)$ законов распределения позволит получить искомую зависимость.

Выводы

1. Авторами рассмотрены некоторые типичные актуальные задачи статистического контроля технологического процесса машиностроительных производств.

2. Сформулированы математические задачи, решение которых способствовало бы решению вышеуказанных задач статистического контроля.

3. Предложены некоторые возможные пути решения этих математических задач.

ЛИТЕРАТУРА

1. Р 50.1.033-2001 Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть 1. Критерии типа хи-квадрат // Нормативные базы ГОСТ/СП/СНиП. URL: <http://files.stroyinf.ru> (дата обращения: 11.10.2022).

2. Данилин Г.А., Затеруха Е.В., Филин

Д.С., Терещенко В.Г. Проектно-технологическое обеспечение надежности функционирования патронов стрелкового оружия / Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова. СПб., 2017. 490 с.

3. Ефремов А.Н., Юдин С.В. Анализ практического применения инструментария контроля качества на машиностроительном предприятии // Известия ТулГУ. Технические науки. 2021. Вып. 12.

4. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.

5. Кулльдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам. М.: Наука, 1966. 176 с.

6. Миттаг Х.-Й., Ринне Х.М. Статистические методы обеспечения качества. М.: Машиностроение, 1995. 616 с.

7. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. Изд. 2-е, перераб. и доп. Л.: Энергоатомиздат; Ленинградское отделение, 1991. 304 с.

8. Точность производства в машиностроении и приборостроении / под ред. А.Н. Гаврилова. М.: Машиностроение, 1973. 567 с.

9. Хан Г., Шапиро С. Статистические модели в инженерных задачах. М.: Мир, 1969. 370 с.