Лабораторная работа 1.01

Исследование распределения случайной величины

Цель работы

1. Провести многократные измерения определенного интервала времени.
2. Построить гистограмму распределения результатов измерения.
3. Вычислить среднее значение и дисперсию полученной выборки.
4. Сравнить гистограмму с графиком функции Гаусса с такими же как и у экспериментального распределения средним значением и дисперсией.

Введение

Случайной называется величина, изменяющаяся от опыта к опыту нерегулярно и на первый взгляд беспорядочно. Так, при бросании игральной кости (кубик с нумерованными гранями) может выпасть любое число от 1 до 6. Радиоактивное ядро может распасться в любую наперед избранную секунду, время жизни ядра до распада — случайная величина. При массовом изготовлении любой продукции все изделия оказываются не вполне идентичными по параметрам. Таким образом, те или иные параметры для совокупности таких изделий также являются случайными величинами.

Результат каждого отдельного измерения случайной величины непредсказуем. Но при многократном повторении измерений в неизменных условиях совокупность их результатов описывается статистическими закономерностями. Если бросать игральную кость сотни раз, каждое определенное число (например, два) выпадает примерно в $1/6$ части общего числа попыток; для радиоактивного вещества, содержащего очень большое число одинаковых ядер, можно надежно предсказать число распадов за любой (но не слишком малый) наперед заданный промежуток времени. Удается подметить закономерности и в распределении по тому или иному параметру изделий, изготовляемых в массовом производстве по определенной технологии.

Наряду со случайными встречаются величины неслучайные. Таковы прежде всего фундаментальные физические постоянные: скорость света в вакууме, заряд и масса электрона и т.п. К неслучайным величинам относятся и свойства конкретного образца в конкретных условиях, например, его плотность или теплоемкость. Но когда в реальном эксперименте измеряется даже неслучайная величина, из-за совместного действия многочисленных неконтролируемых причин результат отдельного измерения подвергается искажениям и становится величиной случайной. Поэтому изучение статистических закономерностей служит одной из основ теории и практики физического и инженерного эксперимента. Часто принимается, что результаты многократных измерений описываются функцией Гаусса (см. ниже) — так называемым законом нормального распределения.

В этой работе необходимо получить выборку (выборочную совокупность) для дискретной случайной величины и исследовать закон распределения этой случайной величины. В качестве исследуемой случайной величины выбран результат измерения заданного промежутка времени. При помощи обычных часов с секундной стрелкой или стрелочного секундомера задают некоторый промежуток t времени и многократно измеряют его достаточно точным цифровым секундомером.

Закономерности в распределении значений созданной таким образом случайной величины можно надежно выявить при соблюдении двух условий:

1. число измерений достаточно велико;
2. чувствительность хронометра достаточна для регистрации из-

менений t от опыта к опыту.

Обычно при задании человеком промежутка времени по секундомеру разброс t не превышает нескольких десятых секунды, следовательно, на табло цифрового хронометра должны высвечиваться сотые и тысячные секунды. Регистрация следующих разрядов бесполезна и лишь усложняет обработку данных.

Закономерности распределения значений изучаемой случайной величины становятся наглядными, если построить гистограмму - ступенчатую диаграмму, показывающую, как часто при измерениях появляются значения, попадающие в тот или иной из равных интервалов Δt, лежащих между наименьшим t_{min} и наибольшим t_{max} из измеренных значений величины t. Гистограмму строят в следующих координатах (см. рис. 1): ось абсцисс — измеряемая величина t, ось ординат — значения $\frac{\Delta N}{N \Delta t}$.

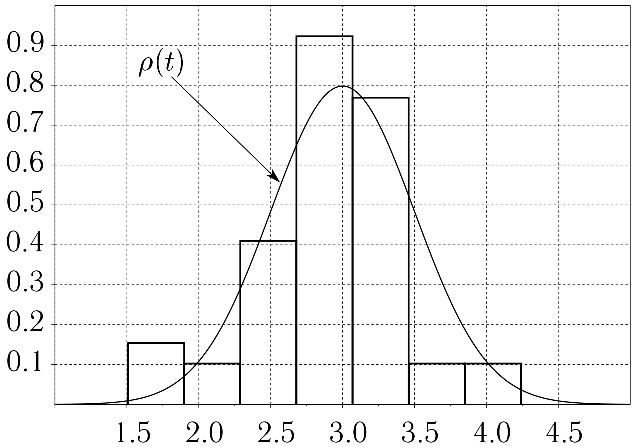

Рис. 1. Гистограмма и функция Гаусса

Здесь N - полное количество измерений, ΔN - количество результатов, попавших в интервал $[t;\ t + \Delta t]$. Частное $\frac{\Delta N}{N}$ есть доля результатов, попавших в указанный интервал, и характеризует вероятность попадания в него результата отдельного измерения. Отношение этой величины к ширине интервала Δt характеризует некоторую «плотность вероятности». При очень большом числе измерений ($N \to \infty$) весь диапазон значений t в принципе можно разбить на «бесконечно малые» интервалы dt и подсчитать число результатов dN в каждом из них. Тогда вместо ступенчатой гистограммы получится плавная кривая соответствующая новой функции t:

$$\rho\left(t\right)=\lim_{\substack{N\to\infty\\ \Delta t\to 0}}\frac{\Delta N}{N\Delta t}=\frac{1}{N}\frac{dN}{dt}. \tag{1}$$

Функцию $\rho(t)$ называют плотностью вероятности или законом распределения исследуемой величины. Нормальное распределение описывается функцией Гаусса:

$$\rho\left(t\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(t-\langle t\rangle\right)^2}{2\sigma^2}\right). \tag{2}$$

Как следует из формулы (2), вид нормального распределения определяется значениями двух параметров: математическим ожиданием $\langle t\rangle$ и среднеквадратичным (стандартным) отклонением σ. Нормальное распределение симметрично относительно вертикали с абсциссой $\langle t\rangle$. Параметр σ характеризует «ширину» распределения: как несложно убедиться, точки перегиба графика нормального распределения (в них вторая производная функции $\rho\left(t\right)$ обращается в ноль) отстоят от линии симметрии графика именно на σ.

Чтобы сравнить исследуемое вами распределение с нормальным, проще всего, найти по данным измерений параметры $\langle t\rangle$ и σ (приближенно, так как число измерений ограничено), вычислить для них функцию $\rho(t)$ по формуле (2) и построить ее в тех же координатах, что и гистограмму (см. рис. 1).

Строго говоря, параметры $\langle t\rangle$ и σ могут быть определены только на основе результатов бесконечно большого числа измерений (генеральной совокупности). Однако, в соответствии с теорией вероятности из выборочной совокупности, содержащей N значений случайной величины $(t_1, t_2, t_3, \ldots t_N)$, можно найти приближенные значений этих параметров: выборочное среднее как среднеарифметическое всех результатов измерений:

$$\langle t \rangle_N = \frac{1}{N}(t_1 + t_2 + \ldots + t_N) = \frac{1}{N}\sum_{i=1}^{N} t_i, \tag{3}$$

и выборочное среднеквадратичное отклонение

$$\sigma_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(t_i - \langle t \rangle_N)^2}. \tag{4}$$

В пределе величины $\langle t \rangle_N$ и σ_N стремятся к $\langle t \rangle$ и σ, соответственно.

Из формулы (2) сразу видно, что плотность нормального распределения имеет максимум при $t = \langle t \rangle$ и симметрична относительно этого значения: следует ожидать, что примерно так же будет выглядеть гистограмма. Можно сравнить максимальную «высоту» гистограммы и $\rho_{max}(t)$:

$$\rho_{\max} = \frac{1}{\sigma\sqrt{2\pi}}. \tag{5}$$

Последнее соотношение получается, если в (2) положить $t = \langle t \rangle$.

Для количественной проверки того, насколько хорошо полученные результаты описываются нормальным распределением, воспользуемся соотношением для вероятности попадания результата измерения в интервал $[t_1, t_2]$:

$$P(t_1 < t < t_2) = \int_{t_1}^{t_2} \rho(t)dt \approx \frac{N_{12}}{N} \tag{6}$$

Смысл соотношения (6) заключается в том, что вероятность $P(t_1 < t < t_2)$ попадания результата каждого измерения в интервал $[t_1; t_2]$, с одной стороны, может быть вычислена как интеграл функции распределения в этих пределах, а с другой — найдена, как относительное число измерений N_{12}, результаты которых попали в этот интервал.

В случае наиболее употребительных на практике интервалов (так называемых стандартных) эта вероятность при условии реализации нормального распределения случайной величины имеет следующие значения:

$$\begin{aligned} & t \in [\langle t\rangle - \sigma, \langle t\rangle + \sigma]\,, && P_{\sigma} \approxeq 0{,}683 \\ & t \in [\langle t\rangle - 2\sigma, \langle t\rangle + 2\sigma]\,, && P_{2\sigma} \approxeq 0{,}954 \\ & t \in [\langle t\rangle - 3\sigma, \langle t\rangle + 3\sigma]\,, && P_{3\sigma} \approxeq 0{,}997 \end{aligned} \tag{7}$$

Из экспериментальной выборки объема N можно найти приближенные значения вероятностей (7), как отношения $\frac{\Delta N_{\sigma}}{N}$, $\frac{\Delta N_{2\sigma}}{N}$, $\frac{\Delta N_{3\sigma}}{N}$ где ΔN_{σ}, $\Delta N_{2\sigma}$, $\Delta N_{3\sigma}$ — количества результатов измерений, попавших в интервалы

$$\begin{aligned} & [\langle t\rangle_N - \sigma_N, \langle t\rangle_N + \sigma_N]\,, \\ & [\langle t\rangle_N - 2\sigma_N, \langle t\rangle_N + 2\sigma_N]\,, \\ & [\langle t\rangle_N - 3\sigma_N, \langle t\rangle_N + 3\sigma_N]\,, \end{aligned} \tag{8}$$

соответственно.

Лабораторная установка

В работе используются устройство или прибор, в котором происходит периодический процесс с частотой порядка нескольких десятых долей герца (часы с секундной стрелкой, стрелочный секундомер, математический или физический маятник) и цифровой секундомер, с ценой деления не более 0,01 *с*. Первый прибор задает интервал времени, который многократно измеряется цифровым секундомером.

Измерения и обработка результатов

1. Выберите устанавливаемый по часам или секундомеру промежуток времени: рекомендуется целое число секунд от 5 до 10. Многократно устанавливая этот промежуток времени, проведите не менее 50 измерений. Результат каждого измерения (показания цифрового хронометра) заносите во второй столбец Табл. 1.

Таблица 1: Результаты прямых измерений

№	t_i, с	$t_i - \langle t \rangle_N$, с	$(t_i - \langle t \rangle_N)^2$, с^2
1			
2			
...			
...			
...			
50			
	$\langle t \rangle_N = \ldots$ с	$\sum_{i=1}^{N} (t_i - \langle t \rangle_N) = \ldots$ с	$\sigma_N = \ldots$ с $\rho_{max} = \ldots$ с$^{-1}$

2. Постройте гистограмму, выполнив для этого следующие операции:

– отыщите в Табл. 1 наименьший t_{min} и наибольший t_{max} из результатов измерений;
– промежуток $[t_{min};\ t_{max}]$ разбейте на m равных интервалов Δt, соблюдая следующие условия; m должно быть целым, близким к $\sqrt{N}$ (напомним, N - полное число измерений). Измеренные значения t_{min} и t_{max} должны попадать внутрь «крайних» интервалов; граничные значения, разделяющие соседние интервалы, должны быть по возможности «круглыми» числами — это облегчит построение гистограммы. Границы выбранных интервалов заносите в первый столбец Табл. 2 (см. Приложение).
– подсчитайте число результатов измерений ΔN_i, из Табл. 1, попавших в каждый из интервалов Δt, заполнив таким образом

второй столбец Табл. 2;
— вычислите опытное значение плотности вероятности (третий столбец Табл. 2);
— постройте на миллиметровой бумаге гистограмму.

3. По данным Табл. 1 с помощью формул (3) и (4) вычислите выборочное значение среднего $\langle t \rangle_N$ и выборочное среднеквадратичное отклонение σ_N;

4. Запишите результаты в «подвал» Табл. 1.
Вычисление $\sum\limits_{i=1}^{N}(t_i - \langle t \rangle_N)$ хороший способ контроля правильности нахождения $\langle t \rangle_N$.

5. По формуле (5) вычислите максимальное значение плотности распределения ρ_{max}, соответствующее $t = \langle t \rangle$, занесите его в «подвал» Табл. 1.

6. Найдите значения t, соответствующие серединам выбранных ранее интервалов, занесите их в четвертый столбец Табл. 2. Для этих значений, используя параметры $\langle t \rangle_N$ и σ_N в качестве $\langle t \rangle$ и σ, вычислите по формуле (2) значения плотности распределения $\rho(t)$, занесите их в пятый столбец Табл. 2. Нанесите все расчетные точки на график, на котором изображена гистограмма, и проводите через них плавную кривую.

7. Проверьте, насколько точно выполняется в ваших опытах соотношение между вероятностями (7) и долями $\frac{\Delta N_\sigma}{N}$, $\frac{\Delta N_{2\sigma}}{N}$, $\frac{\Delta N_{3\sigma}}{N}$. Для этого вычислите границы интервалов (8) для найденных вами значений $\langle t \rangle_N$ и σ_N, занесите их во второй и третий столбцы Табл. 3 (см. Приложение).

8. По данным Табл. 1 подсчитайте и занесите в Табл. 3 количество ΔN измерений, попадающих в каждый из этих интервалов, и отношение $\frac{\Delta N}{N}$ этого количества к общему числу измерений. Сравните их с соответствующими нормальному распределению значениями P вероятности (7).

9. Рассчитайте среднеквадратичное отклонение среднего значения

по формуле:

$$\sigma_{\langle t \rangle} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} \left(t_i - \langle t \rangle_N\right)^2} \tag{9}$$

10.Найдите табличное значение коэффициента Стьюдента $t_{\alpha,N}$ для доверительной вероятности $\alpha = 0{,}95$. Запишите доверительный интервал для измеряемого в работе промежутка времени.

$$\Delta t = t_{\alpha,N} \cdot \sigma_{\langle t \rangle}, \tag{10}$$

где $t_{\alpha,N}$ — коэффициент Стьюдента, зависящий от числа измерений N и доверительной вероятности α:

$$\alpha = P\left(t \in [\langle t \rangle - \Delta t, \langle t \rangle + \Delta t]\right). \tag{11}$$

Контрольные вопрсы

1. Являются ли, по вашему мнению, случайными следующие физические величины:

– плотность алмаза при $20^\circ C$
– напряжение сети
– сопротивление резистора, взятого наугад из партии с одним и тем же номинальным сопротивлением
– число молекул в 1см^3 при нормальных условиях?

Приведите другие примеры случайных и неслучайных физических величин.

2. Изучая распределение ЭДС партии электрических батареек, студент использовал цифровой вольтметр. После нескольких измерений получились такие результаты (в вольтах): 1,50; 1,49; 1,50; 1,50; 1,49. Имеет ли смысл продолжать измерения? Что бы вы изменили в методике этого эксперимента?

3. При обработке результатов измерений емкости партии конденсаторов получено: $\langle C \rangle = 1{,}1$ мкФ, $\sigma = 0{,}1$ мкФ. Если взять коробку со 100 конденсаторами из этой партии, то сколько среди

них можно ожидать конденсаторов с емкостью меньше 1 мкФ? больше 1,3 мкФ?

Литература

1. Методическое пособие «Обработка экспериментальных данных» / Курепин В.В., Баранов И.В., под ред. В.А. Самолетова, 2012.

2. Элементарные оценки ошибок измерений / Зайдель А.Н., Изд. 3-е испр. и доп. Л., "Наука Ленинградское отделение, 1968.

3. Практическая физика / Сквайрс Дж., М.: Мир, 1971.

Приложение

Таблица 2: Данные для построения гистограммы

Границы интервалов, *с*	ΔN	$\frac{\Delta N}{N\Delta t}$, с$^{-1}$	t, с	ρ, с$^{-1}$

Таблица 3: Стандартные доверительные интервалы

	Интервал, c		ΔN	$\dfrac{\Delta N}{N}$	P
	от	до			
$\langle t \rangle_N \pm \sigma_N$					
$\langle t \rangle_N \pm 2\sigma_N$					
$\langle t \rangle_N \pm 3\sigma_N$					