CHAPTER 4

Data Analysis and Probability Theory

Data analysis is the process by which we examine test results and draw conclusions from them. Using data analysis, we can evaluate DUT design weaknesses, identify DIB and tester repeatability and correlation problems, improve test efficiency, and expose test program bugs. As mentioned in Chapter 1, debugging is one of the main activities associated with mixed-signal test and product engineering. Debugging activities account for about 20% of the average workweek. Consequently, data analysis plays a very large part in the overall test and product-engineering task.

Many types of data visualization tools have been developed to help us make sense of the reams of test data that are generated by a mixed-signal test program. In this chapter we will review several common data analysis tools, such as the datalog, histogram, lot summary, shmoo plot, and the wafer map. We shall also review some statistical and probability concepts as they apply to analog and mixed-signal test. Of particular interest to the engineer is the ability to model large quantity of data with simple mathematical models. Often, these are based on an assumption of the structure of the underlying random behavior and several quantities obtained from the measured data, such as mean and standard deviation. Once a model of the randomness of the data is found, questions about the likelihood an event will occur can be quantified. This helps the test engineer identify with confidence the meaning of a measurement and, ultimately, the conditions under which to perform a measurement in minimum time.

4.1 DATA VISUALIZATION TOOLS

4.1.1 Datalogs (Data Lists)

A datalog, or data list, is a concise listing of test results generated by a test program. Datalogs are the primary means by which test engineers evaluate the quality of a tested device. The format of a datalog typically includes a test category, test description, minimum and maximum test limits, and a measured result. The exact format of datalogs varies from one tester type to another, but datalogs all convey similar information.

Figure 4.1. Example datalog from a Teradyne Catalyst tester.

```
Sequencer:  S_continuity
 1000 Neg PPMU Cont      Failing Pins:  0
Sequencer:  S_VDAC_SNR
 5000 DAC Gain Error      T_VDAC_SNR     -1.00  dB    <    -0.13    dB        < 1.00    dB
 5001 DAC S/2nd           T_VDAC_SNR     60.0  dB     <=   63.4     dB
 5002 DAC S/3rd           T_VDAC_SNR     60.0  dB     <=   63.6     dB
 5003 DAC S/THD           T_VDAC_SNR     60.00  dB    <=   60.48    dB
 5004 DAC S/N             T_VDAC_SNR     55.0  dB     <=   70.8     dB
 5005 DAC S/N+THD         T_VDAC_SNR     55.0  dB     <=   60.1     dB
Sequencer:  S_UDAC_SNR
 6000 DAC Gain Error      T_UDAC_SNR     -1.00  dB    <    -0.10    dB        < 1.00    dB
 6001 DAC S/2nd           T_UDAC_SNR     60.0  dB     <=   86.2     dB
 6002 DAC S/3rd           T_UDAC_SNR     60.0  dB     <=   63.5     dB
 6003 DAC S/THD           T_UDAC_SNR     60.00  dB    <=   63.43    dB
 6004 DAC S/N             T_UDAC_SNR     55.0  dB     <=   61.3     dB
 6005 DAC S/N+THD         T_UDAC_SNR     55.0  dB     <=   59.2     dB
Sequencer:  S_UDAC_Linearity
 7000 DAC POS ERR         T_UDAC_Lin    -100.0 mV     <    7.2      mV        < 100.0   mV
 7001 DAC NEG ERR         T_UDAC_Lin    -100.0 mV     <    3.4      mV        < 100.0   mV
 7002 DAC POS INL         T_UDAC_Lin     -0.90  lsb   <    0.84     lsb       < 0.90    lsb
 7003 DAC NEG INL         T_UDAC_Lin     -0.90  lsb   <    -0.84    lsb       < 0.90    lsb
 7004 DAC POS DNL         T_UDAC_Lin     -0.90  lsb   <    1.23     lsb (F)   < 0.90    lsb
 7005 DAC NEG DNL         T_UDAC_Lin     -0.90  lsb   <    -0.83    lsb       < 0.90    lsb
 7006 DAC LSB SIZE        T_UDAC_Lin      0.00 mV     <    1.95     mV        < 100.00  mV
 7007 DAC Offset V        T_UDAC_Lin    -100.0 mV     <    0.0      mV        < 100.0   mV
 7008 Max Code Width      T_UDAC_Lin      0.00  lsb   <    1.23     lsb       < 1.50    lsb
 7009 Min Code Width      T_UDAC_Lin      0.00  lsb   <    0.17     lsb       < 1.50    lsb
Bin:  10
```

A short datalog from a Teradyne Catalyst tester is shown in Figure 4.1. Each line of the datalog contains a shorthand description of the test. For example, "DAC Gain Error" is the name of test number 5000. The gain error test is part of the S_VDAC_SNR test group and is executed during the T_VDAC_SNR test routine. The minimum and maximum limits for the test are also listed. Using test number 5000 as an example, the lower limit of DAC gain error is –1.00 dB, the upper limit is +1.00 dB, and the measured value for this DUT is –0.13 dB.

The datalog displays an easily recognizable fail flag beside each value that falls outside the test limits. For instance, test 7004 in Figure 4.1 shows a failure in which the measured value is 1.23 LSBs. Since the upper limit is 0.9 LSBs, this test fails. In this particular example, the failure is flagged with an (F) symbol. Hardware and software alarms from the tester also result in a datalog alarm flag, such as (A). Alarms can occur for a variety of reasons, including mathematical divisions by zero and power supply currents that exceed programmed limits. When alarms are generated, the test program halts (unless instructed by the test engineer to ignore alarms). The tester assumes that the DUT is defective and treats the alarm as a failure.

Because the device in Figure 4.1 fails test 7004, it is categorized into bin 10 as displayed at the bottom of the datalog. Bin 1 usually represents a good device, while other bins usually represent various categories of failures and alarms. Sometimes there are multiple grades of shippable devices, which are separated into different passing bins. For example, a certain percentage of 2-GHz microprocessors may fail at 2 GHz, but may operate perfectly well at 1.8 GHz. The 1.8-GHz processors might be sorted into bin 2 and shipped at a lower cost, while the higher-grade 2-GHz processors are sorted into bin 1 to be sold at full price.

4.1.2 Lot Summaries

Lot summaries are generated after all devices in a given production lot have been tested. A lot summary lists a variety of information about the production lot, including the lot number, product number, operator number, and so on. It also lists the yield loss and cumulative yield associated with each of the specified test bins. The overall lot yield is defined as the ratio of the total number of good devices divided by the total number of devices tested:

$$\text{lot yield} = \frac{\text{total good devices}}{\text{total devices tested}} \times 100\% \tag{4.1}$$

Figure 4.2 shows a simplified lot summary, including yields for a variety of test categories. The lot yield is listed in the lot summary, but it does not tell us everything we need to know. If a particular lot exhibits a poor yield, we want to know *why* its yield was low. We want to know what category or categories of tests dominated the failures so we can look into the problem to determine its cause. A lot summary can help us identify which failures are most common to a particular type of DUT. This allows us to focus our attention on the areas of the design, the process, and the test program that might be causing the most failures in production. For this reason, lot summaries also list test categories and what percentage of devices failed each category. Specifically, two metrics are used called test category yield loss and cumulative yield. Yield loss per category is defined as

$$\text{test category yield loss} = \frac{\text{\# failed devices per specific test}}{\text{total devices tested}} \times 100\% \tag{4.2}$$

and cumulative test yield is defined as

$$\text{cumulative yield} = \left(1 - \frac{\text{total failed devices at end of a specific test}}{\text{total devices tested}}\right) \times 100\% \tag{4.3}$$

The lot summary in Figure 4.2 shows that our highest yield loss is due to the RECV channel AC tests. We might think that our XMIT channel has no problems, because it causes only a 0.30% yield loss. However, we have to be careful in making such judgements based on data collected during production. We have to remember that once a DUT fails any test, the tester immediately rejects it and moves on to the next device. After all, there is no point in continuing to test a DUT once it has been disqualified for shipment to the customer.

Since the test program halts after the first DUT failure, the earlier tests will tend to cause more yield loss than later ones, simply because fewer DUTs proceed to the later tests. The earlier failures mask any failures that would have occurred in later tests. For example, any or all the devices that failed the RECV channel tests in Figure 4.2 might also have failed the XMIT channel tests if given the chance. Therefore, during the device characterization phase we may want to instruct the tester to collect data from all tests whether the DUT passes or not. Of course, the extra testing leads to a longer average test time; thus we do not want to perform continue-on-fail, testing in production unless necessary.

We can sometimes improve our overall production throughput by moving the more Commonly failed tests toward the beginning of the test program. Average test time is reduced by the rearrangement because we do not waste time performing tests that seldom fail only to lose yield to subsequent tests that often fail. Once again, we have to remember that the order of tests may affect the lot summary output. Therefore, whenever we wish to reorder our test program based on yield loss, we would prefer to use lot summaries in which the tester does not halt on the first failure. That way, we know which tests are truly the ones that catch the largest number of defective DUTs.

Figure 4.2. Simplified lot summary.

Lot Number: 122336
Device Number: TLC1701FN
Operator Number: 42
Test Program: F779302.load

Devices Tested: 10233
Passing Devices: 9392
Test Yield: 91.78%

Bin#	Test Category	Devices Tested	Failures	Yield Loss	Cum. Yield
7	Continuity	10233	176	1.72%	98.28%
2	Supply Currents	10057	82	0.80%	97.48%
3	Digital Patterns	9975	107	1.05%	96.43%
4	RECV Channel AC	9868	445	4.35%	92.08%
5	XMIT Channel AC	9423	31	0.30%	91.78%

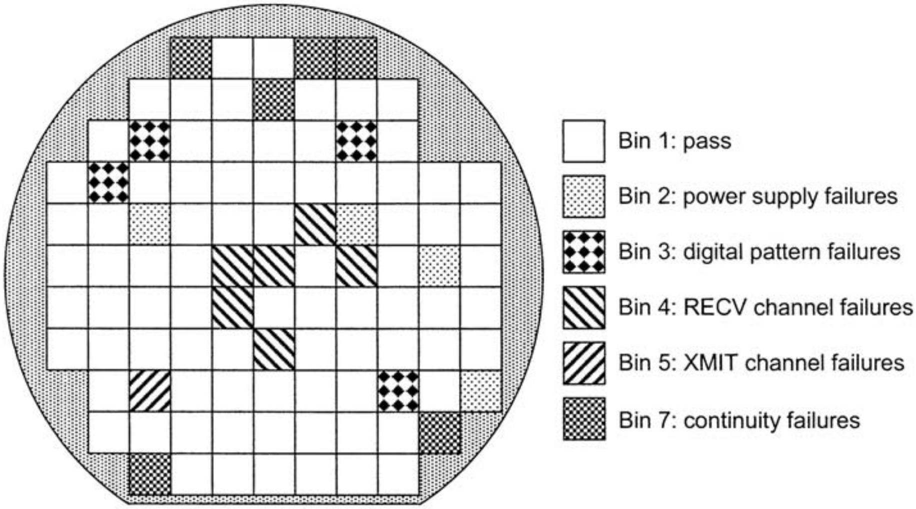
When rearranging test programs based on yield loss, we also have to consider the test time that each test consumes. For example, the RECV channel tests in Figure 4.2 may take 800 ms, while the digital pattern tests only takes 50 ms. The digital pattern test is more efficient at identifying failing DUTs since it takes so little test time. Therefore, it might not make sense to move the longer RECV test to the beginning of the program, even though it catches more defective DUTs than does the digital pattern test. Clearly, test program reordering is not a simple matter of moving the tests having the highest yield loss to the beginning of the test program.

4.1.3 Wafer Maps

A wafer map (Figure 4.3) displays the location of failing die on each probed wafer in a production lot. Unlike lot summaries, which only show the number of devices that fail each test category, wafer maps show the physical distribution of each failure category. This information can be very useful in locating areas of the wafer where a particular problem is most prevalent. For example, the continuity failures are most severe at the upper edge of the wafer illustrated in Figure 4.3. Therefore, we might examine the bond pad quality along the upper edge of the wafer to see if we can find out why the continuity test fails most often in this area. Also, the RECV channel failures are most severe near the center of the wafer. This kind of ring-like pattern often indicates a processing problem, such as uneven diffusion of dopants into the semiconductor surface or photomask misalignment. If all steps of the fabrication process are within allowable tolerances, consistent patterns such as this may indicate that the device is simply too sensitive to normal process variations and therefore needs to be redesigned. Wafer maps are a powerful data analysis tool, allowing yield enhancement through a cooperative effort between the design, test, product, and process engineers.

Naturally, it is dangerous to draw too many conclusions from a single wafer map. We need to examine many wafer maps to find patterns of consistent failure distribution. For this reason, some of the more sophisticated wafer mapping tools allows us to overlay multiple wafer maps on top of one another, revealing consistency in failure distributions. From these composite failure maps, we can draw more meaningful conclusions about consistent processing problems, design weaknesses, and test hardware problems.

Figure 4.3. Wafer map.



4.1.4 Shmoo Plots

Shmoo plots were among the earliest computer-generated graphic displays used in semiconductor manufacturing.[1] A shmoo plot is a graph of test results as a function of test conditions. For example, some of the earliest shmoo plots displayed pass/fail test results for PMOS memory ICs as a function of V_{DD} and V_{SS}. The origins of the name "shmoo plot" are not known for certain. According to legend, some of the early plots reminded the engineers of a shmoo, a squash-shaped cartoon character from Al Capp's comic strip "Li'l Abner." Although few shmoo plots are actually shmoo-shaped, the name has remained with us.

The graph in Figure 4.4 is called a *functional shmoo plot*, since it only shows which test conditions produce a passing (functional) or failing (nonfunctional) test result. This type of plot is commonly used to characterize purely digital devices, since digital test programs primarily produce functional pass/fail results from the digital pattern tests. Measured values such as supply current and distortion cannot be displayed using a functional shmoo plot.

Analog and mixed-signal measurements often require a different type of graph, called a *parametric shmoo plot*. Analog and mixed-signal test programs produce many parametric values, such as gain error and signal-to-distortion ratio. Parametric shmoo plots, such as that shown in Figure 4.5, can be used to display analog measurement results at each combination of test conditions rather than merely displaying a simple pass/fail result. Naturally, we always have the option of comparing the analog measurements against test limits, producing pass/fail test results compatible with a simple functional shmoo plot.

Parametric shmoo plots give the test engineer a more complete picture of the performance of mixed-signal DUTs under the specified range of test conditions. This information can tell the engineering team where the device is most susceptible to failure. Assume, for example, that the DUT of Figure 4.5 needs to pass a minimum S/THD specification of 75 dB. The shmoo plot in Figure 4.5 tells us that this DUT is close to failure at about 40°C and it is somewhat marginal at 30°C if our V_{DD} supply voltage is near either end of the allowable range. Once the device weaknesses are understood, then the device design, fabrication process, and test program can be improved to maximize production yield. Also, shmoo plots can help us identify worst-case test conditions. For example, we may choose to perform the S/THD test at both low and high V_{DD} based on the worst-case test conditions indicated by the shmoo plot in Figure 4.5.

Figure 4.4. Functional shmoo plot.

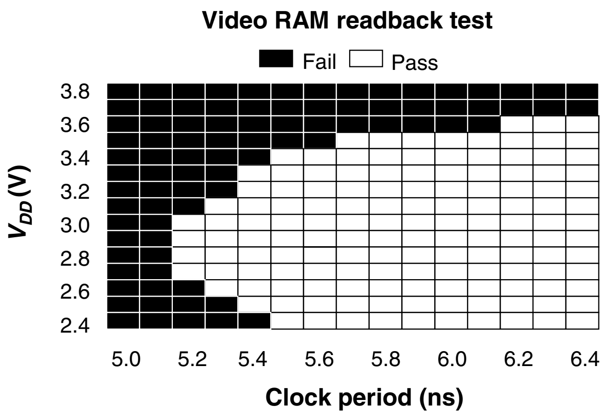
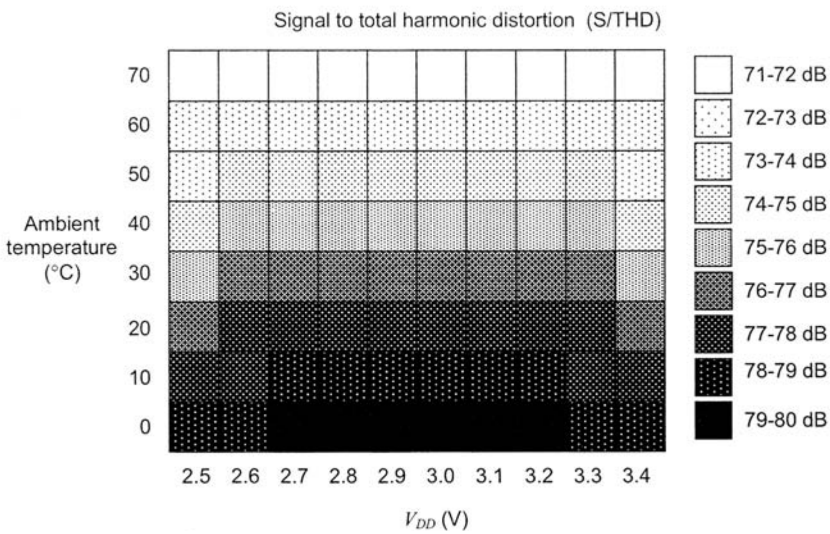


Figure 4.5. Two-dimensional parametric shmoo plot.



Shmoo plots can be generated, at least in principle, using data collected through manual adjustment of test conditions. However, such a process would be extremely tedious. For example, if we wanted to plot pass/fail results for 10 values of V_{DD} combined with 10 values of master clock period, then we would have to run the test 100 times under 100 different test conditions, adjusting the test conditions by hand each time. Clearly, software automation in the tester is required if the shmoo data collection process is to be a practical one. For this reason, modern ATE tester operating systems often include built-in shmoo plotting tools. These tools not only display the shmoo plots themselves, but they also provide automated adjustment of test conditions and automated collection of test results under each permutation of test conditions.

The shmoo plots illustrated in Figures 4.4 and 4.5 only represent a few of the many types of shmoo plots that can potentially be created. For example, we could certainly imagine a 3D shmoo plot showing pass/fail results for combinations of three test conditions instead of two. It is important to note that any of the many factors affecting DUT performance can be used as shmoo plot test conditions.

Common examples of shmoo test conditions include power supply voltage, master clock frequency (or period), setup and hold times, ambient temperature, I_{OL} or I_{OH} load current, and so on. However, we are free to plot any measured values or pass/fail results as a function of any combination of test conditions. This flexibility makes the shmoo plot a very powerful characterization and diagnostic tool whose usefulness is limited only by the ingenuity and skill of the test or product engineer.

4.1.5 Histograms

When a test program is executed multiple times on a single DUT, it is common to get multiple answers. This is largely a result of the additive noise that is present in the system involving the DUT. For example, a DAC gain error test may show slight repeatability errors if we execute the test program repeatedly, as shown in Figure 4.6. (In this example, only the results from test 5000 have been enabled for display.)

We can view the repeatability of a group of measurements using a visualization tool called a *histogram*. A histogram corresponding to the DAC gain error example is shown in Figure 4.7. It shows a plot of the distribution of measured values as well as a listing of several key statistical values. The plot is divided into a number of vertical histogram cells, each indicating the percentage of values falling within the cell's upper and lower thresholds. For example, approximately 5% of the DAC gain error measurements in this example fell between –0.137 and –0.136 dB. The histogram is a very useful graphical tool that helps us visualize the repeatability of measurements. If the measurement repeatability is good, the distribution should be closely packed, as the example in Figure 4.7 shows. But if repeatability is poor, then the histogram spreads out into a larger range of values. Although histograms are extremely useful for analyzing measurement stability, repeatability studies are not the only use for histograms. They are also used to look at distributions of measurements collected from many DUTs to determine the extent of variability from one device to another. Excessive DUT-to-DUT variability indicates a fabrication process that is out of control or a device design that is too susceptible to normal process variations.

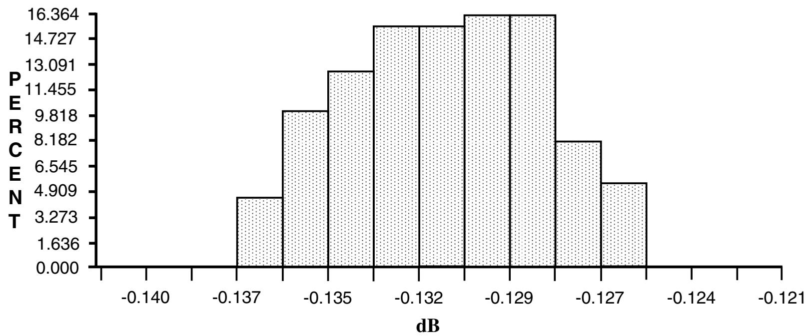
In addition to the numerical results and a plot of the distribution of measured values, the example histogram in Figure 4.7 displays a number of other useful values. For example, the

Figure 4.6. Repeated test executions result in fluctuating measurements.

```
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.127  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.129  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.125  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.131  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.129  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.128  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.132  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.130  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.134  dB     < 1.000  dB
5000 DAC Gain Error     T_VDAC_SNR  -1.000  dB <  -0.131  dB     < 1.000  dB
```

Figure 4.7. Histogram of the DAC gain error test.

Test No	Test Function Name	Test Label	Units
5000	T_VDAC_SNR	DAC Gain Error	db
Lower Test Limit=	-1dB	Upper Test Limit=	1dB
	- DISTRIBUTION	STATISTICS -	
Lower Pop Limit=	-Infinity	Upper Pop Limit=	+Infinity
Total Results=	110	Results Accepted=	110
Underflows=	0	Overflows=	0
Mean=	-0.13003dB	Std Deviation=	0.00292899
Mean – 3 Sigma=	-0.13882dB	Mean + 3 Sigma=	-0.012125dB
Minimum Value=	-0.13594dB	Maximum Value=	-0.12473dB
	- PLOT	STATISTICS –	
Lower Plot Limit=	-0.14dB	Upper Plot Limit=	-0.12dB
Cells=	15	Cell Width=	0.0013333dB
Full Scale Percent=	16.36%	Full Scale Count=	18



population size is listed beside the heading "Total Results=." It indicates how many times the measurement was repeated. In the case of a DUT-to-DUT variability study, the "Total Results=" value would correspond to the number of DUTs tested rather than the number of measurement repetitions on the same DUT. In either case, the larger the population of results, the trustworthier a histogram becomes. A histogram with fewer than 50 results is statistically questionable because of the limited sample size. Ideally a histogram should contain results from at least 100 devices (or 100 repeated test executions in the case of a single-DUT repeatability study).

While the histogram is useful tool for visualizing the data set, we are often looking for a simpler representation of the data set, especially one that lends itself for comparison to other data sets. To this end, the next section will introduce a statistical description of the data set, where specific metrics of the data are used for comparative purposes. Moreover, we shall introduce some basic concepts of probability theory of random variables to help quantify the mathematical structure present in a data set. It is assumed that the reader has been exposed to probability theory elsewhere, and only the concepts relevant to analog and mixed-signal testing will be covered here. For a more in-depth presentation of statistics, including the derivation of fundamental equations and properties of statistics, the reader should refer to a book on the subject of statistics and probability theory.[2–4]

4.2 STATISTICAL ANALYSIS

4.2.1 Mean (Average) and Standard Deviation (Variance)

The frequency or distribution of data described by a histogram characterizes a sample set in great detail. Often, we look for simpler measures that describe the statistical features of the sample set. The two most important measures are the arithmetic mean μ and the standard deviation σ.

The arithmetic mean or simply the mean μ is a measure of the central tendency, or location, of the data in the sample set. The mean value of a sample set denoted by $x(n)$, $n = 0, 1, 2, \ldots, N-1$, is defined as

$$\mu = \frac{1}{N}\sum_{n=0}^{N-1} x(n) \tag{4.4}$$

Exercises

4.1.	A 5-mV signal is measured with a meter 10 times, resulting in the following sequence of readings: 5 mV, 6 mV, 9 mV, 8 mV, 4 mV, 7 mV, 5 mV, 7 mV, 8 mV, 11 mV. What is the mean value? What is the standard deviation?	ANS. 7 mV, 2.0 mV.
4.2.	What are the mean and standard deviation of a set of samples of a coherent sine wave having a DC offset of 5 V and a peak-to-peak amplitude of 1.0 V?	ANS. μ = 5.0 V, σ = 354 mV.
4.3.	A 5-mV signal is measured with a meter 10 times, resulting in the following sequence of readings: 7 mV, 6 mV, 9 mV, 8 mV, 4 mV, 7 mV, 5 mV, 7 mV, 8 mV, 11 mV. What is the mean value? What is the standard deviation?	ANS. 7.2 mV, 1.887 mV.
4.4.	If 15,000 devices are tested with a yield of 63%, how many devices passed the test?	ANS. 9450 devices.

For the DAC gain error example shown in Figure 4.7 the mean value from 110 measurements is –0.1300 dB. The standard deviation σ, on the other hand, is a measure of the dispersion or uncertainty of the measured quantity about the mean value, μ. If the values tend to be concentrated near the mean, the standard deviation is small. If the values tend to be distributed far from the mean, the standard deviation is large. Standard deviation is defined as

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left[x(n) - \mu\right]^2} \tag{4.5}$$

Standard deviation and mean are expressed in identical units. In our DAC gain error example, the standard deviation was found to be 0.0029 dB. Another expression for essentially the same quantity is the variance or mean square deviation. It is simply equal to the square of the standard deviation, that is, variance $= \sigma^2$.

Often, the statistics of a sample set are used to estimate the statistics of a larger group or population from which the samples were derived. Provided that the sample size is greater than 30, approximation errors are insignificant. Throughout this chapter, we will make no distinction between the statistics of a sample set and those of the population, because it will be assumed that sample size is much larger than 30.

There is an interesting relationship between a sampled signal's DC offset and RMS voltage and the statistics of its samples. Assuming that all frequency components of the sample set are coherent, the mean of the signal samples is equal to the signal's DC offset. Less obvious is the fact that the standard deviation of the samples is equal to the signal's RMS value, excluding the DC offset. The RMS of a sample set is calculated as the square root of the mean of the squares of the samples

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left[x(n)\right]^2} \tag{4.6}$$

If the value of μ in Eq. (4.5) is zero (i.e., if the sample set has no DC component), then Eq. (4.5) becomes identical to the RMS calculation above. Thus we can calculate the standard deviation of the samples of a coherent signal by calculating the RMS of the signal after subtracting the average value of the sample set (i.e., the DC offset).

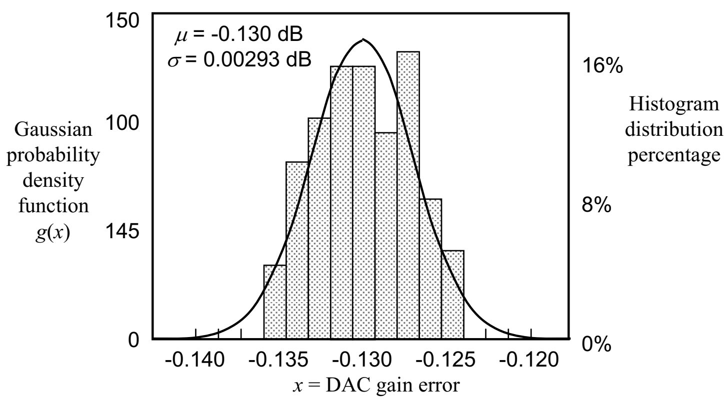
4.2.2 Probabilities and Probability Density Functions

The histogram in Figure 4.7 exhibits a feature common to many analog and mixed-signal measurements. The distribution of values has a shape similar to a bell. The bell curve (also called a *normal distribution* or *Gaussian distribution*) is a common one in the study of statistics. According to the central limit theorem,[5] the distribution of a set of random variables, each of which is equal to a summation of a large number ($N > 30$) of statistically independent random values, trends toward a Gaussian distribution. As N becomes very large, the distribution of the random variables becomes Gaussian, whether or not the individual random values themselves exhibit a Gaussian distribution. The variations in a typical mixed-signal measurement are caused by a summation of many different random sources of noise and crosstalk in both the device and the tester instruments. As a result, many mixed-signal measurements exhibit the common Gaussian distribution.

Figure 4.8 shows the histogram count from Figure 4.7 superimposed on a plot of the corresponding Gaussian probability density function (PDF). The PDF is a function that defines the probability that a randomly chosen sample X from the statistical population will fall near a particular value. In a Gaussian distribution, the most likely value of X is near the mean value, μ. Thus the PDF has a peak at $x = \mu$.

Notice that the height of the histogram cells only approximates the shape of the true Gaussian curve. If we collect thousands of test results instead of the 110 used in this example, the height of the actual histogram cells should more closely approach the shape of the probability density

Figure 4.8. Continuous normal (Gaussian) distribution for DAC gain example.



function. This is the nature of statistical concepts. Actual measurements only approach the theoretical ideal when large sample sets are considered.

The bell-shaped probability density function $g(x)$ for any Gaussian distribution having a mean μ and standard deviation σ is given by the equation

$$g(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{4.7}$$

Since the shape of the histogram in Figure 4.8 approximates the shape of the Gaussian pdf, it is easy to assume that we can calculate the expected percentage of histogram counts at a value by simply plugging the value of into Eq. (4.7). However, a PDF represents the probability *density*, rather than the probability itself. We have to perform integration on the PDF to calculate probabilities and expected histogram counts.

The probability that a randomly selected value in a population will fall between the values a and b is equal to the area under the PDF curve bounded by $x = a$ and $x = b$, as shown in Figure 4.9. Stating this more precisely, for any probability density function $f(x)$, the probability P that a randomly selected value X will fall between the values a and b is given by

$$P(a < X < b)=\int_a^b f(x)dx \tag{4.8}$$

The value of P must fall between 0 (0% probability) and 1 (100% probability). In the case where $a = -\infty$ and $b = \infty$, the value of P must equal 1, since there is a 100% probability that a randomly chosen value will be a number between $-\infty$ and $+\infty$. Consequently, the total area underneath any PDF must always be equal to 1.

As $f(x)$ is assumed continuous, the probability that a random variable X is *exactly* equal to any particular value is zero. In such case we can replace either or both of the signs < in Eq. (4.8) by ≤, allowing us to write

$$P(a \le X \le b)=\int_a^b f(x)dx \tag{4.9}$$

Incorporating the equality in the probability expression is therefore left as a matter of choice.

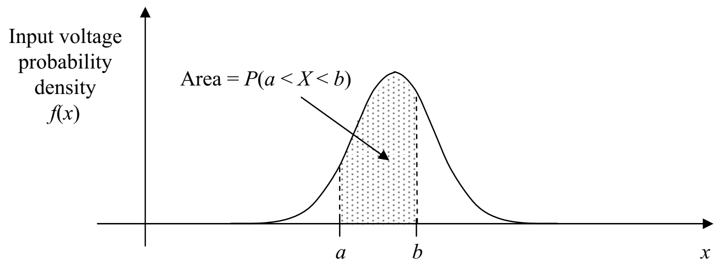
The probability that a Gaussian distributed randomly variable X will fall between the values of a and b can be derived from Eq. (4.8) by substituting Eq. (4.7) to obtain

$$P(a \le X \le b)=\int_a^b g(x)\,dx=\int_a^b \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx \tag{4.10}$$

Unfortunately, Eq. (4.10) cannot be solved in closed form. However, it can easily be solved using numerical integration methods. For instance, in our Gaussian DAC gain example, let us say we want to predict what percentage of measured results should fall into the seventh histogram cell. From the histogram, we see that there are 15 evenly spaced cells between –0.140 and –0.120 dB. The seventh cell represents all values falling between the values

$$a=-0.140\text{ dB}+7\,\frac{\left[-0.120\text{ dB}-(-0.140\text{ dB})\right]}{15}=-0.132\text{ dB}$$

Figure 4.9. The probability over the range ***a*** to ***b*** is the area under the PDF, ***f***(***x***), in that interval.



and

$$b = -0.140\ \text{dB} + 8\,\frac{\left[-0.120\ \text{dB} - \left(-0.140\ \text{dB}\right)\right]}{15} = -0.1307\ \text{dB}$$

We can calculate the probability that a randomly selected DAC gain error measurement X will fall between –0.1320 and –0.1307 using the equation

$$P\left(-0.132 < X < -0.1307\right) = \int_{-0.132}^{-0.1307} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)}{2\sigma^2}} dx$$

where μ = –0.130 dB and σ = 0.00293 dB. Using mathematical analysis software, the value of $P(-0.1320 < X < -0.1307)$ is found to be 0.163, or 16.3%. Theoretically, then, we should see 16.3% of the 110 DAC gain error measurements fall between –0.1307 and –0.1320. Indeed the seventh histogram cell shows that approximately 15.5% of the measurements fall between these values.

This example is fairly typical of applied statistics. Actual distributions never exactly match a true Gaussian distribution. Notice, for example, that the ninth and tenth histogram cells in Figure 4.8 are badly out of line with the ideal Gaussian curve. Also, notice that there are no values outside the fourth and twelfth cells. An ideal Gaussian distribution would extend to infinity in both directions. In other words, if one is willing to wait billions of years, one should eventually see an answer of +200 dB in the DAC gain error example. In reality, of course, the answer in the DAC gain example will never stray more than a few tenths of a decibel away from the average reading of –0.130 dB, since the actual distribution is only near Gaussian.

Nevertheless, statistical analysis predicts actual results well enough to be very useful in analyzing test repeatability and manufacturing process stability. The comparison between ideal results and actual results is close enough to allow some general statements. First, the peak-to-peak variation of the observed data (e.g., maximum observed value – minimum observed value) is roughly equal to six times the standard deviation of a near-Gaussian distribution, that is,

$$\text{peak-to-peak value} \approx 6\sigma \tag{4.11}$$

In the DAC gain distribution example, the standard deviation is 0.00293 dB. Therefore, we would expect to see values ranging from approximately –0.139 dB to –0.121 dB. These values are displayed in the example histogram in Figure 4.7 beside the labels "Mean –3 sigma" and "Mean +3 sigma." The actual minimum and maximum values are also listed. They range from –0.136 to –0.125 dB, which agrees fairly well with the ideal values of $\mu \pm 3\sigma$.

At this point we should note a common misuse of statistical analysis. We have used as our example a gain measurement, expressed in decibels. Since the decibel is based on a logarithmic transformation, we should actually use the equivalent V/V measurements to calculate statistical quantities such as mean and standard deviation. For example, the average of three decibel values, 0, –20, and –40 dB, is 20 dB. However, the true average of these values as calculated using V/V is given by

$$\text{average gain} = \frac{1 + \frac{1}{10} + \frac{1}{100}}{3} = 0.37\ \frac{\text{V}}{\text{V}} = -8.64\ \text{dB}$$

A similar discrepancy arises in the calculation of standard deviation. Therefore, a Gaussian-distributed sample set converted into decibel form is no longer Gaussian. Nevertheless, we often use the nonlinearized statistical calculations from a histogram to evaluate parameters expressed in decibel units as a time-saving shortcut. The discrepancy between linear and logarithmic calculations of mean and standard deviation become negligible as the range of decibel values decreases. For example, the range of values in the histogram of Figure 4.7 is quite small; so the errors in mean and standard deviation are minor. The reader should be careful when performing statistical analysis of decibel values ranging over several decibels.

4.2.3 The Standard Gaussian Cumulative Distribution Function $\Phi(z)$

Computing probabilities involving Gaussian distributions is complicated by the fact that numerical integration methods must be used to solve the definite integrals involved. Fortunately, a simple change of variable substitution can be used to convert the integral equations into one involving a Gaussian distribution with zero mean and unity standard deviation. This then enables a set of tables or approximations that require numerical evaluation to be used to solve the probabilities associated with an arbitrary Gaussian distribution. Hence, the test engineer can completely avoid the need for numerical integration routines. Let us consider how this is done.

The probability that a randomly selected value X will be less than a particular value x can be calculated directly from Eq. (4.9). We set $a = -\infty$ and $b = x$ and write

$$F(x) = P(X < x) = P(-\infty < X < x) = \int_{-\infty}^{x} f(y)\,dy \qquad \textbf{(4.12)}$$

This integral is central to probability theory and is given a special name called the *cumulative distribution function (CDF).* Here we view $F(x)$ as an ordinary function of the variable x. The probability that X lies in the range a to b can then be expressed in terms of the difference of $F(x)$ evaluated at $x = a$ and $x = b$ according to

$$P(a < X < b) = F(b) - F(a) \qquad \textbf{(4.13)}$$

In the case of a Gaussian distribution, $F(x)$ is equal to

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y-\mu)^2}{2\sigma^2}}\,dy \qquad \textbf{(4.14)}$$

If we consider the simple change of variable $z = (y - \mu)/\sigma$, Eq. (4.14) can be rewritten as

$$F(x) = \int_{-\infty}^{\left(\frac{x-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \tag{4.15}$$

Except for the presence of μ and σ in the upper integration limit, the integration kernel no longer depends on these two values. Alternatively, one can view $F(x)$ in Eq. (4.15) as the CDF of a Gaussian distribution having zero mean and unity standard deviation. By tabulating a single function, say

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-u^2}{2}} du \tag{4.16}$$

we can write $F(x)$ as

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{4.17}$$

In other words, to determine the value of a particular CDF involving a Gaussian random variable with mean μ and standard deviation σ at a particular point, say x, we simply normalized x by subtracting the mean value followed by a division by σ, that is, $z = (y - \mu)/\sigma$, and compute $\Phi(z)$.

The function $\Phi(z)$ is known as the *standard* Gaussian CDF. The variable z is known as the *standardized point of reference*. Traditionally, $\Phi(z)$ has been evaluated by looking up tables that list $\Phi(z)$ for different values of z. A short tabulation of $\Phi(z)$ is provided in Appendix A. Rows of $\Phi(z)$ are interleaved between rows of z. A plot of $\Phi(z)$ versus z is also provided in Figure 4.10. For reference, we see from this plot that $\Phi(-\infty) = 0$ and $\Phi(\infty) = 1$. Also evident is the antisymmetry about the point (0, 0.5), giving rise to the relation $\Phi(-z) = 1 - \Phi(z)$.

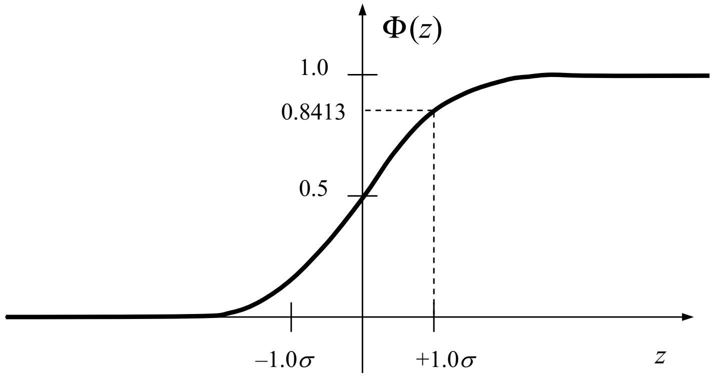
More recently, the following expression[6] has been found to give reasonably good accuracy (less than 0.1%) for $\Phi(z)$ with $\alpha = 1/\pi$ and $\beta = 2\pi$:

$$\Phi(z) \approx \begin{cases} 1 - \left(\dfrac{1}{(1-\alpha)z + \alpha\sqrt{z^2 + \beta}}\right) \dfrac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} & 0 \le z \le \infty \\ \left(\dfrac{1}{(\alpha-1)z + \alpha\sqrt{z^2 + \beta}}\right) \dfrac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} & -\infty < z < 0 \end{cases} \tag{4.18}$$

Equation (4.18) is generally very useful when we require a standardized value that is not contained in the table of Appendix A or any other Gaussian CDF table for that matter. To illustrate the application of the standard cdf, the probability that a Gaussian distributed random variable X with mean μ and standard deviation σ lies in the range a to b is written as

$$P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{4.19}$$

Figure 4.10. Standard Gaussian cumulative distribution function.



Equation (4.19) is a direct result of substituting Eq. (4.17) into Eq. (4.13). Equation (4.19) can be used to generate certain rules of thumb when dealing with Gaussian random variables. The probability that a random variable will fall within:

1σ of its mean is

$$P(\mu-\sigma < X < \mu+\sigma) = \Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right) = \Phi(1) - \Phi(-1) = 0.6826$$

2σ of its mean is

$$P(\mu-2\sigma < X < \mu+2\sigma) = \Phi\left(\frac{\mu+2\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-2\sigma-\mu}{\sigma}\right) = \Phi(2) - \Phi(-2) = 0.9544$$

3σ of its mean is

$$P(\mu-3\sigma < X < \mu+3\sigma) = \Phi\left(\frac{\mu+3\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-3\sigma-\mu}{\sigma}\right) = \Phi(3) - \Phi(-3) = 0.9974$$

It is also instructive to look at several limiting cases associated with Eq. (4.19); specifically, when $a = -\infty$ and b is an arbitrary value, we find

$$P(X < b) = P(-\infty < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{-\infty-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) \qquad \textbf{(4.20)}$$

Conversely, with a arbitrary and $b = \infty$, we get

$$P(a < X) = P(a < X < \infty) = \Phi\left(\frac{\infty-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad \textbf{(4.21)}$$

Of course, as previously mentioned, with $a = -\infty$ and $b = \infty$, we obtain $P(-\infty < X < \infty) = 1$.

Notice that the probability that X will fall outside the range $\mu \pm 3\sigma$ is extremely small, that is, $P(X < \mu - 3\sigma) + P(\mu + 3\sigma < X) = \Phi(-3) + 1 - \Phi(3) = 0.0026$. As a result, one would not expect many measurement results to fall very far beyond $\mu \pm 3.0\sigma$ in a relatively small-size test set.

To summarize, the steps involved in calculating a probability involving a Gaussian random variable are as follows:

1. Estimate the mean μ and standard deviation σ of the random variable from the sample set using

$$\mu = \frac{1}{N}\sum_{n=0}^{N-1} x(n) \quad \text{and} \quad \sigma = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left[x(n) - \mu\right]^2}$$

2. Determine the probability interval limits, a and b, and write a probability expression in terms of the standard Gaussian cumulative distribution function $\Phi(z)$ according to

$$P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

3. Evaluate $\Phi(z)$ through a computer program, or a look-up table (Appendix A) or by using the numerical approximation given in Eq. (4.18).

The following examples will help to illustrate this procedure.

EXAMPLE 4.1

A DC offset measurement is repeated many times, resulting in a series of values having an average of 257 mV. The measurements exhibit a standard deviation of 27 mV. What is the probability that any single measurement will return a value larger than 245 mV?

Solution:

If X is used to denote the Gaussian random variable, then we want to know $P(245 \text{ mV} < X)$. Comparing the probability limits with the expression listed in Eq. (4.21), we can state a = 245 mV and $b = \infty$. Furthermore, since μ = 257 mV and σ = 27 mV, we can write

$$P(245 \text{ mV} < X) = 1 - \Phi\left(\frac{245 \text{ mV} - 257 \text{ mV}}{27 \text{ mV}}\right) = 1 - \Phi(-0.44)$$

Referring to Appendix A we see that the value $\Phi(-0.44)$ is not listed. However, it lies somewhere between 0.3085 and 0.3446; thus either we can make a crude midpoint interpolation of 0.33 or, alternatively, we can use the approximation for $\Phi(z)$ given in Eq. (4.18) and write

$$\Phi(-0.44) \approx \left[\frac{1}{\left(\frac{1}{\pi} - 1\right)(-0.44) + \left(\frac{1}{\pi}\right)\sqrt{(-0.44)^2 + 2\pi}}\right]\frac{1}{\sqrt{2\pi}} e^{\frac{-(-0.44)^2}{2}} = 0.3262$$

We shall select the latter value of 0.3262 and state that the probability that the measurement will be ***greater*** than 245 mV is equal to 1 – 0.3262, or 0.6738. Consequently, there is a 67.38% chance that any individual measurement will exceed 245 mV.

EXAMPLE 4.2

A sample of a DC signal in the presence of noise is to be compared against a reference voltage level. If the sample is greater than the 1-V threshold, the comparator output goes to logic 1. If 100 samples were taken, but only five ones were detected, what is the value of the signal, assuming that the noise is zero mean and has a standard deviation of 10 mV?

Solution:
Modeling the PDF of the signal plus noise at the receiver input, we can write

$$f(v) = \frac{1}{10^{-2}\sqrt{2\pi}} e^{\frac{-(v-\mu)^2}{2(10^{-2})^2}}$$

where the mean value, μ, is equal to the unknown DC signal level. Next, we are told that there is a 5% chance of the sample exceeding the 1-V threshold. This is equivalent to the mathematical statement $\boldsymbol{p}$ $(1.0 \le \boldsymbol{v} < \infty) = 5/100$. Because the noise is Gaussian, we can write the following from Eq. (4.19):

$$P(1.0 \le V < \infty) = 0.05 = 1 - \Phi\left(\frac{1.0-\mu}{10^{-2}}\right)$$

or

$$\Phi\left(\frac{1.0-\mu}{10^{-2}}\right) = 0.95$$

which allows us to write

$$\frac{1.0-\mu}{10^{-2}} = \Phi^{-1}(0.95) \Rightarrow \mu = 1.0 - 10^{-2} \times \Phi^{-1}(0.95)$$

We can estimate the value of $\Phi^{-1}(0.95)$ from Appendix A, to be somewhere between 1.6 and 1.7. or we can make use of a computer program and find $\Phi^{-1}(0.95) = 1.645$. This leads to

$$\mu = 1.0 - 10^{-2} \times 1.645 = 0.98$$

Therefore the signal level is 0.98 V.

Exercises

4.5. A 5-mV signal is measured with a meter ten times resulting in the following sequence of readings: 5 mV, 6 mV, 9 mV, 8 mV, 4 mV, 7 mV, 5 mV, 7 mV, 8 mV, 11 mV. Write an expression for the PDF for this measurement set assuming the distribution is Gaussian.

ANS. $g(v)=\dfrac{1}{\left(2\times10^{-3}\right)\sqrt{2\pi}}e^{\frac{-\left(v-7\times10^{-3}\right)^2}{2\left(2\times10^{-3}\right)^2}}$; $\mu=7.0$ mV; $\sigma=2.0$ mV.

4.6. A set of measured data is modeled as a Gaussian distribution with 1.0-V mean value and a standard deviation of 0.1 V. What is the probability a single measurement will fall between 0.8 and 0.9 V.

ANS. $P(0.8 \le V \le 0.9) = 0.18$.

4.2.4 Verifying Gaussian Behavior: The Kurtosis and Normal Probability Plot

It is fairly common to encounter distributions that are non-Gaussian. Two common deviations from the familiar bell shape are bimodal distributions, such as that shown in Figure 4.11, and distributions containing outliers as shown in Figure 4.12. When evaluating measurement repeatability on a single DUT, these distributions are a warning sign that the test results are not sufficiently repeatable. When evaluating process stability (i.e., consistency from DUT to DUT), these plots may indicate a weak design or a process that needs to be improved. In general, looking at the histogram plot to see if the distribution accurately follows a Gaussian distribution is difficult to do by inspection. Instead, one prefers to work with some form of mathematical measure of normality.

A commonly used metric to see how closely a given data set follows a Gaussian distribution is the kurtosis. The kurtosis is defined as

$$\text{kurtosis}=\frac{1}{N}\sum_{n=1}^{N}\left[\frac{x(n)-\mu}{\sigma}\right]^4 \tag{4.22}$$

where μ is the mean and σ is the standard deviation of a random sequence *x*. A Gaussian distribution regardless of its mean value or standard deviation has a kurtosis of three. Any significant deviation from this value will imply that the distribution is non-Gaussian. However, one generally

Figure 4.11. Bimodal distribution.

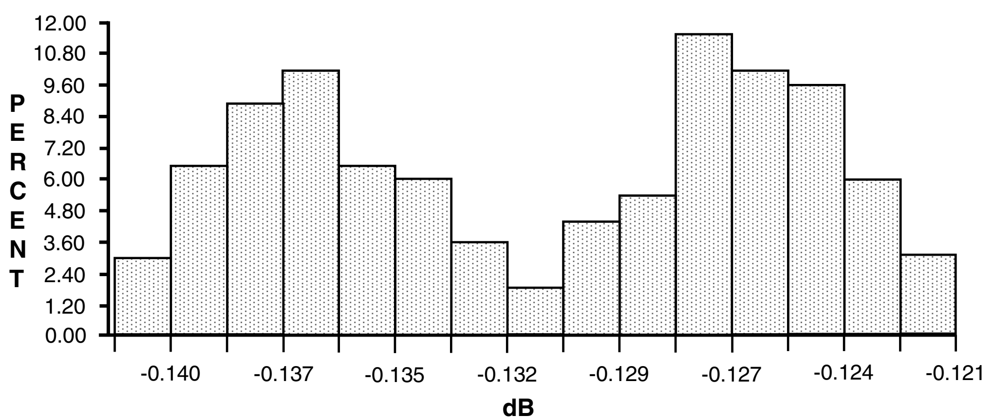
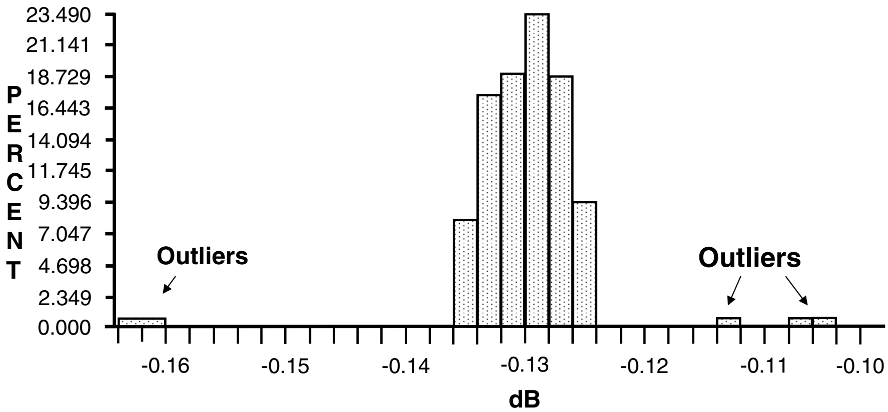


Figure 4.12. Distribution with statistical outliers.



must be cautious of making decisions from a point metric like kurtosis alone. Instead, a normal probability plot[7] is a better approach for observing whether or not a data set is approximately normally distributed. The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality.

The steps for constructing a normal probability plot are best illustrated by way of an example. Let us assume that we have the following 10 data points with a kurtosis of 2.5 (generalizing these steps to larger data sets should be straightforward):

Data	12.04	14.57	12.00	12.02	13.48	12.70	12.86	15.13	10.10	14.14

Step 1: Sort the data from the smallest to the largest.

Data	12.04	14.57	12.00	12.02	13.48	12.70	12.86	15.13	10.10	14.14
Data Sort	10.10	12.00	12.02	12.04	12.70	12.86	13.48	14.14	14.57	15.13

Step 2: Assign each data point a position number, k, from 1 to N (length of data vector, 10).

Data	12.04	14.57	12.00	12.02	13.48	12.70	12.86	15.13	10.10	14.14
Data Sort	10.10	12.00	12.02	12.04	12.70	12.86	13.48	14.14	14.57	15.13
Position k	1	2	3	4	5	6	7	8	9	10

Step 3: Use the position values to calculate the corresponding uniformly distributed probability, where the individual probabilities are found from the expression $p_k = (k - 0.5)/N$, where in this case $N = 10$.

Data	12.04	14.57	12.00	12.02	13.48	12.70	12.86	15.13	10.10	14.14
Data Sort	10.10	12.00	12.02	12.04	12.70	12.86	13.48	14.14	14.57	15.13
Position *k*	1	2	3	4	5	6	7	8	9	10
Prob. p_k	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95

Step 4: Compute the standard z scores corresponding to each probability, that is, $\Phi^{-1}(p_k)$.

Data	12.04	14.57	12.00	12.02	13.48	12.70	12.86	15.13	10.10	14.14
Data Sort	10.10	12.00	12.02	12.04	12.70	12.86	13.48	14.14	14.57	15.13
Position *k*	1	2	3	4	5	6	7	8	9	10
Prob. p_k	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
$\Phi^{-1}(p_k)$	-1.645	-1.036	-0.674	-0.385	-0.126	0.126	0.385	0.674	1.036	1.645

Step 5: Plot the sorted measured data versus the standard normal values, $\Phi^{-1}(p_k)$ as shown in Figure 4.13.

Figure 4.13. A normal probability plot used to visually validate that a data set follows Gaussian behavior (***N*** = 10).

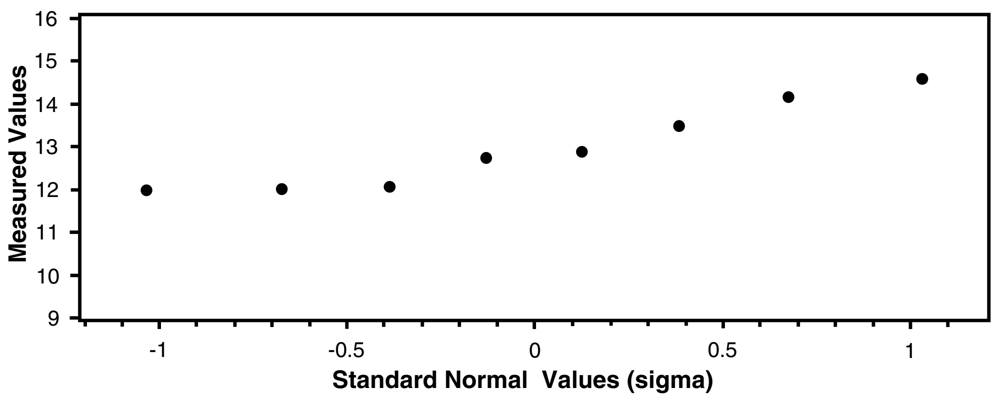
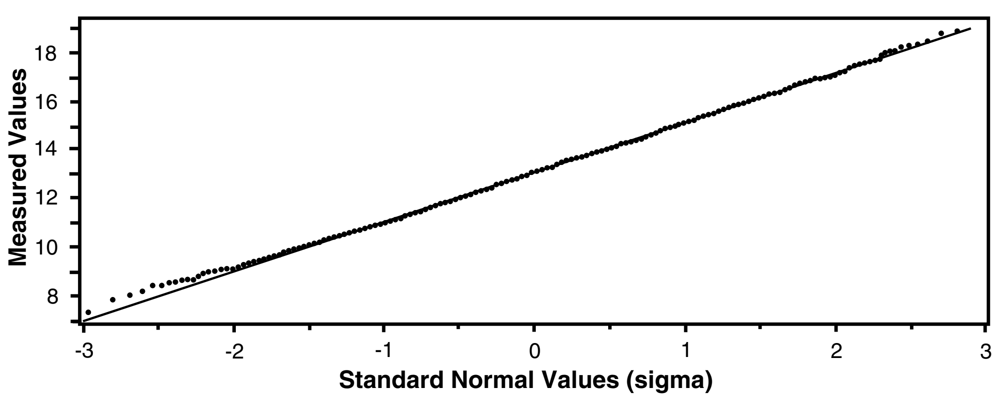


Figure 4.14. A normal probability plot of a near-Gaussian distribution (***N*** = 1000).



As is evident from Figure 4.13, the data appear to roughly follow a straight line. There is no reason to think that the data are severely non-Gaussian even though the kurtosis does not equal to 3. Keep in mind that our confidence in our decision either way is questionable since we are working with only 10 samples. If the sample set is increased to 1000, the straight-line behavior of the probability plot becomes quite obvious as shown in Figure 4.14. We would then say with great confidence that the data behave Gaussian over a range of $\pm 2\sigma$ but less so beyond these values. It is also interesting to note that the kurtosis is very close to 3 at 2.97. This example serves to illustrate the limitation of a single-point metric in judging the normality of the data.

4.3 NON-GAUSSIAN DISTRIBUTIONS FOUND IN MIXED-SIGNAL TEST

A few examples of non-Gaussian probability distributions functions that one encounters in analog and mixed-signal test problems will be described in this section. Both continuous and discrete probability distributions will be described. The reader will encounter these distributions later on in this textbook.

4.3.1 The Uniform Probability Distribution

A random variable X that is equally likely to take on values within a given range [A,B] is said to be uniformly distributed. The random variable may be discrete or continuous. The PDF of a continuous uniform distribution is a rectangular function, as illustrated in Figure 4.15. Note that the height of this PDF must be chosen so that the area enclosed by the rectangular function is unity. Mathematically, the PDF of a uniform distribution is given by

$$f(x) = \begin{cases} \dfrac{1}{B-A}, & A \le x \le B \\ 0, & \text{elsewhere} \end{cases} \tag{4.23}$$

The probability that a uniformly distributed random variable X will fall in the interval a to b, where $-A \le a < b \le B$ is obtained by substituting Eq. (4.23) into Eq. (4.9) to get

$$P(a \le X \le b) = \int_a^b f(x)\,dx = \int_a^b \frac{1}{B-A}\,dx = \frac{b-a}{B-A} \tag{4.24}$$

The mean and standard deviation of an ideal uniform distribution are given by

$$\mu = \frac{A+B}{2} \tag{4.25}$$

and

$$\sigma = \frac{(B-A)}{\sqrt{12}} \tag{4.26}$$

In the discrete case, the PDF (or what is also referred to as the probability mass function) of a uniformly distributed random variable within the interval [A,B] is a series of equally spaced, equally weighted impulse functions described as

$$f(x) = \frac{1}{B - A + 1} \sum_{k=A}^{B} \delta(x - k) \tag{4.27}$$

In much the same way as the continuous case, the probability that a uniformly distributed random integer X will fall in the interval bounded by integers a and b, where $A \le a < b \le B$ is obtained by substituting Eq. (4.27) into Eq. (4.9) to get

$$P(a \le X \le b) = \int_a^b f(x)\, dx = \frac{b - a}{B - A + 1} \tag{4.28}$$

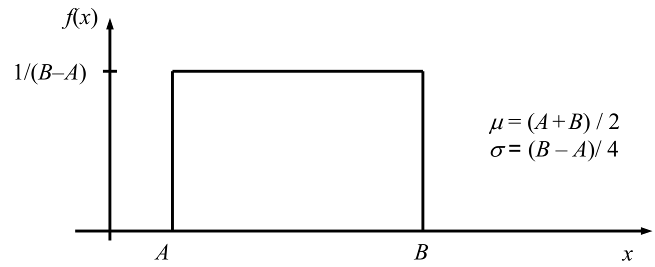
The mean and standard deviation of the discrete uniform distribution are given by

$$\mu = \frac{A + B}{2} \quad \text{and} \quad \sigma = \sqrt{\frac{(B - A + 1)^2 - 1}{12}} \tag{4.29}$$

Uniform distributions occur in at least two instances in mixed-signal test engineering. The first instance is random number generators found in various programming languages, such as the *rand* function in MATLAB and the *random* function in *C*. This type of function returns a randomly chosen number between a minimum and maximum value (typically 0 and 1, respectively). The numbers are supposed to be uniformly distributed between the minimum and maximum values. There should be an equal probability of choosing any particular number, and therefore a histogram of the resulting population should be perfectly flat. One measure of the quality of a random number generator is the degree to which it can produce a perfectly uniform distribution of values.

The second instance in which we commonly encounter a uniform distribution is the errors associated with the quantization process of an analog-to-digital converter (ADC). It is often assumed that statistical nature of these errors is uniformly distributed between –1/2 LSB (least significant bit) and +1/2 LSB. This condition is typically met in practice with an input signal that is sufficiently random. From Eq. (4.25), we see that the average error is equal to zero. Using Eq. (4.26) and, assuming that the standard deviation and RMS value are equivalent, we expect that the ADC will generate $1/\sqrt{12}$ LSB of RMS noise when it quantizes a signal. In the case of a

Figure 4.15. Uniform distribution pdf.



full-scale sinusoidal input having a peak of 2^{N-1} LSBs (or an RMS of $2^{N-1}/\sqrt{2}$), the signal-to-noise ratio (SNR) at the output of the ADC is

$$\text{SNR} = 20\ \log\left(\frac{\text{signal RMS}}{\text{noise RMS}}\right) = 20\ \log\left(\frac{2^{N-1}/\sqrt{2}\ \text{LSB}}{1/\sqrt{12}\ \text{LSB}}\right) \tag{4.30}$$

leading to

$$\text{SNR} = 20\ \log_2\left(\sqrt{6}\times 2^{N-1}\right) \tag{4.31}$$

Simplifying further, we see that the SNR depends linearly on the number of bits, N, according to

$$\text{SNR} = \frac{20\log_2\left(\sqrt{6}\times 2^{N-1}\right)}{\log_2(10)} = 1.761\ \text{dB} + 6.02\ \text{dB}\times N \tag{4.32}$$

In the situation where we know the SNR, we can deduce from Eq. (4.32) that the equivalent number of bits for the ADC, that is, ENOB $= N$, is

$$\text{ENOB} = \frac{\text{SNR} - 1.761\ \text{dB}}{6.02\ \text{dB}} \tag{4.33}$$

Exercises

4.7. A random number generator produces the digits: $N = 0, 1, 2, \ldots, 9$ uniformly. What is the cumulative distribution function of this data set? What is the mean and standard deviation of this pdf?

ANS. $P(0 \le X \le x) = F(x) = 0.1 + 0.1x$, $x \in \{0,1,2,\ldots,9\}$; $\mu = 4.5$; $\sigma = 2.87$.

4.8. A triangular PDF has the form $f(x) = \begin{cases} 4x, & 0 \le x < 0.5 \\ 4-4x, & 0.5 \le x \le 1 \end{cases}$. What is the corresponding cumulative distribution function? What is the mean and standard deviation?

ANS. $F(x) = \begin{cases} 0, & x < 0 \\ 2x^2, & 0 \le x < 0.5 \\ 1-2(1-x)^2, & 0.5 \le x \le 1 \\ 1, & 1 < x \end{cases}$; $\mu = \frac{1}{2}$; $\sigma = \frac{1}{\sqrt{24}}$

4.3.2 The Sinusoidal Probability Distribution

Another non-Gaussian distribution that is commonly found in analog and mixed-signal test engineering problems is the sinusoidal distribution shown in Figure 4.16. The PDF of a sinusoidal signal described by $A\sin(2\pi f \cdot t) + B$ is given by

$$f(x) = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - (x-B)^2}}, & -A + B \le x \le A + B \\ 0, & \text{elsewhere} \end{cases} \tag{4.34}$$

The probability that a sinusoidal distributed random variable X will fall in the interval a to b, where $-A+B \le a < b \le A+B$ is obtained by substituting Eq. (4.34) into Eq. (4.9) to get

$$P\left(a < X < b\right) = \int_a^b \frac{1}{\pi\sqrt{A^2-\left(x-B\right)^2}}\,dx = \frac{1}{\pi}\left[\sin^{-1}\left(\frac{b-B}{A}\right) - \sin^{-1}\left(\frac{a-B}{A}\right)\right] \tag{4.35}$$

The mean and standard deviation of an ideal sinusoidal distribution with offset B are given by

$$\mu = B \tag{4.36}$$

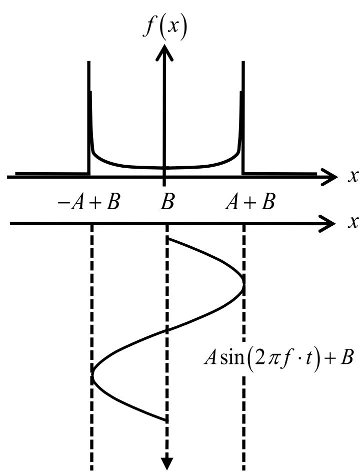
and

$$\sigma = \frac{A}{\sqrt{2}} \tag{4.37}$$

Sinusoidal distributions occur in at least two instances in mixed-signal test engineering. The first instance is in ADC testing where the ADC transfer characteristic is derived from the code distribution that results from sampling a full-scale sine-wave input described by $A\sin\left(2\pi f\cdot t\right)+B$. Because an ideal D-bit ADC has 2^D-1 code edges located at $-A+B+V_{LSB}\times n,\ n=0,1,\ldots 2^D-1$, where V_{LSB} is the distance between adjacent code edges, the probability of a random sample X falling in the region bounded by consecutive code edges can be derived from Eq. (4.35) as

$$\begin{aligned} &P\left[-A+B+n\times V_{LSB} < X < -A+B+(n+1)\times V_{LSB}\right] \\ &= \frac{1}{\pi}\sin^{-1}\left[\frac{(n+1)\times V_{LSB}-A}{A}\right] - \frac{1}{\pi}\sin^{-1}\left[\frac{n\times V_{LSB}-A}{A}\right] \end{aligned} \tag{4.38}$$

Figure 4.16. The PDF of a sinusoidal signal described by $A\sin\left(2\pi f\cdot t\right)+B$.



If N samples of the ADC are collected, we can then predict (based on the above probability argument) that the number of times we should see a code between n and $n + 1$, denoted by $H(n)$, is

$$H(n)=\frac{N}{\pi}\left\{\sin^{-1}\left[\frac{(n+1)\times V_{LSB}}{A}-1\right]-\sin^{-1}\left[\frac{n\times V_{LSB}}{A}-1\right]\right\} \tag{4.39}$$

Another situation where sinusoidal distributions are used is in serial I/O channel testing. A sinusoidal phase modulated signal is created and used to excite the clock recovery circuit of a digital receiver. The amplitude of the sine wave is then varied until a specific level of probability of error (also referred to as bit error rate) is achieved. In this situation, we are looking for the probability that the sinusoidal distribution with zero mean lies in the region between some threshold value, say t_{TH}, and the upper limit of the sine wave, A. Using Eq. (4.35), we can state this as

$$P(t_{TH}<X<A)=\frac{1}{\pi}\left[\frac{\pi}{2}-\sin^{-1}\left(\frac{t_{TH}}{A}\right)\right]=\frac{1}{\pi}\cos^{-1}\left(\frac{t_{TH}}{A}\right) \tag{4.40}$$

EXAMPLE 4.3

A 3-bit ADC is excited by a 1024-point full-scale sine wave described by $1.0\sin\left(2\pi\frac{1}{1024}\cdot n\right)+2.0,\ n=0,1,\ldots,1023$. Arrange the voltage axis between 1.00 and 3.00 V into 2^3 equal regions and count the number of samples from the sine-wave that fall into each region. Compare with the counts given by Eq. (4.39).

Solution:

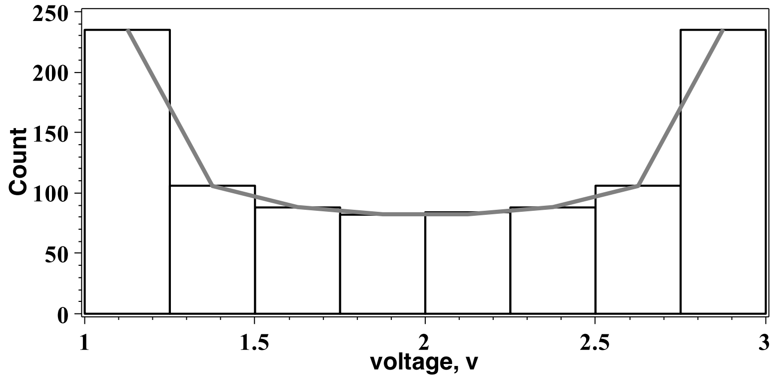
Subdividing the voltage range of 1.00 to 3.00 V into 8 regions suggests that the distance between each code is given by

$$V_{LSB}=\frac{3.00-1.00}{2^3}=0.25\text{ V}$$

Next, organizing all 1024 samples into eight regions we obtain the table of code distribution shown below. In addition, in the rightmost column we list the corresponding code counts as predicted by Eq. (4.39). As is clearly evident, the results compare.

Index, *n*	**Code Region (*V*)**	**Samples Falling in Code Region**	**Predicted Count, *H***
0	1.00–1.25	235	235.57
1	1.25–1.50	106	105.75
2	1.50–1.75	88	88.31
3	1.75–2.00	82	82.36
4	2.00–2.25	84	82.36
5	2.25–2.50	88	88.31
6	2.50–2.75	106	105.75
7	2.75–3.00	235	235.57

A plot of the above two data sets is shown below for comparative purposes. Clearly, the two data sets are in close agreement.



4.3.3 The Binomial Probability Distribution

Consider an experiment having only two possible outcomes, say **A** and **B**, which are mutually exclusive. If the probability of **A** occurring is p, then the probability of **B** will be $q = 1 - p$. If the experiment is repeated N times, then the probability of **A** occurring k times is given by

$$P[X = k] = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \tag{4.41}$$

Exercises

4.9. An 8-bit ADC converts an analog input voltage level into 2^8 code levels ranging from 0 to $2^8 - 1$. If the ADC is excited by a 1024-point full-scale sine wave described by $0.5\sin(2\pi\ 1/1024 \cdot n) + 0.5$, $n = 0,1,\ldots,1023$, what is the expect number of codes appearing at the output corresponding to code 0? What is the code count for code 2^8-1? What is the distance between adjacent code edges?

ANS. $H(0) = H(2^8 - 1) = 40.77$; $LSB = 3.9$ mV.

4.10. What is the probability that a 1-V amplitude sine wave lies between 0.6 and 0.8 V? How about lying between –0.1 and 0.1 V? Again for 0.9 V and 1.0 V?

ANS. $P(0.6 \le V \le 0.8) = 0.09$; $P(-0.1 \le V \le 0.1) = 0.06$; $P(0.9 \le V \le 1.0) = 0.14$

The binomial probability density function can then be written as

$$f(x) = \sum_{k=0}^{N} \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \delta(x-k) \tag{4.42}$$

The probability that a binomial distributed random integer X will fall in the interval bounded by integers a and b, where $0 \le a < b \le N$, is obtained by substituting Eq. (4.42) into Eq. (4.9) to get

$$P\left(a \le X \le b\right) = \int_a^b f\left(x\right)\, dx = \int_a^b \sum_{k=0}^{N} \frac{N!}{k!\left(N-k\right)!} p^k \left(1-p\right)^{N-k} \delta\left(x-k\right) dx \tag{4.43}$$

or, when simplified,

$$P\left(a \le X \le b\right) = \sum_{k=a}^{b} \frac{N!}{k!\left(N-k\right)!} p^k \left(1-p\right)^{N-k} \tag{4.44}$$

The mean and standard deviation of the binomial PDF is

$$\mu = N \cdot p \quad \text{and} \quad \sigma = \sqrt{N \cdot p \cdot \left(1-p\right)} \tag{4.45}$$

The binomial distribution arises in applications where there are two types of objects or outcomes (e.g., heads/tails, correct/erroneous bits, good/defective items) and we are interested in the number of times a specific object type occurs in a randomly selected batch of size *N*. We shall see the application of the binomial distribution to model the bit error rate in data communications.

EXAMPLE 4.4

A lot of 100 devices is tested during production, where it is assumed that a single bad device has a probability of 0.1 of failing. What is the probability that three bad dies will show up during this test?

Solution:
Each die from the lot of 100 has equal chance of failing with a probability $p = 0.1$. Consequently, the probability that three dies out of 100 will fail is determined using Eq. (4.41) as follows:

$$P\left[X = 3\right] = \frac{3!}{3!\left(100-3\right)!)}\left(0.1\right)^3\left(1-0.1\right)^{100-3} = 0.0059$$

Exercises

4.11. A lot of 500 devices is tested during production, where it is assumed that a single device has a probability of 0.99 of passing. What is the probability that two bad dies will show up during this test?

ANS. $P[X = 2] = 0.0836$

4.12. A sample of 2 dies are selected from a large lot and tested. If the probability of passing is 5/6, what is the PDF of two bad dies appearing during this test?

ANS. $f(x) = \frac{25}{36}\delta(x) + \frac{10}{36}\delta(x-1) + \frac{1}{36}\delta(x-2)$

4.13. Data are transmitted over a channel with a single-bit error probability of 10^{-12}. If 10^{12} bits are transmitted, what is the probability of one bit error. What is the average number of errors expected?

ANS. $P[X = 1] = 0.3678$; $\mu = 1$

4.4 MODELING THE STRUCTURE OF RANDOMNESS

At the heart of any measurement problem is the need to quantify the structure of the underlying randomness of the data set. For data that are modeled after a single probability density function, this is generally a straightforward procedure. For instance, the measured data shown in Figure 4.17a was found to have a mean value of 1 and standard deviation of 3. If we assume that the randomness of this data set is Gaussian, then we can describe the PDF of this sequence as

$$g_1(x)=\frac{1}{3\sqrt{2\pi}}e^{\frac{-(x-1)^2}{18}} \tag{4.46}$$

where x represents any value of the data sequence.

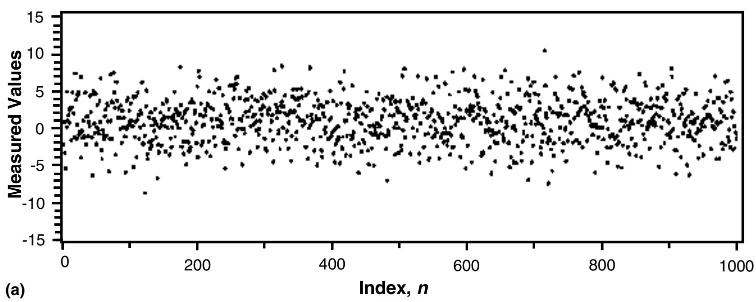
Conversely, if we assume that the randomness of the data set is to be modeled as a uniformly distributed random process, then according to the formulas for mean and standard deviation for the uniform distribution given in Eq. (4.25) and Eq. (4.26), we write

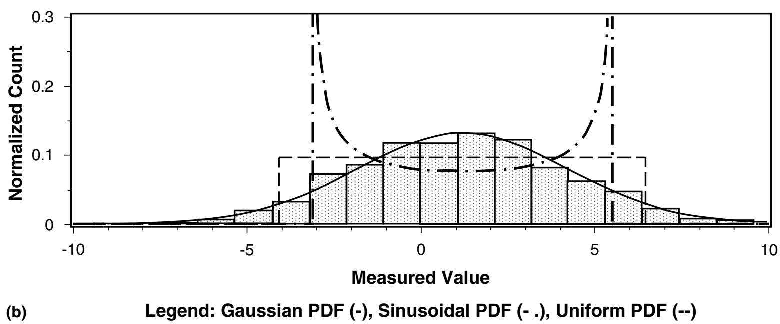
$$1=\frac{A+B}{2} \quad \text{and} \quad 3=\frac{(B-A)}{\sqrt{12}} \tag{4.47}$$

leading to

$$A=1-3\sqrt{3} \quad \text{and} \quad B=1+3\sqrt{3} \tag{4.48}$$

Figure 4.17. (a) A random set of measured data and (b) its corresponding histogram with various PDFs superimposed.





from which we write

$$g_2(x)=\begin{cases}\dfrac{1}{6\sqrt{3}}, & 1-3\sqrt{3}\le x\le 1-3\sqrt{3}\\[2ex] 0, & \text{elsewhere}\end{cases}\tag{4.49}$$

Likewise, if we assume that the randomness is to be modeled as a sinusoidal distribution, then we can immediately see

$$B=1 \quad\text{and}\quad A=3\sqrt{2}\tag{4.50}$$

from Eqs (4.36) and (4.37), allowing us to write

$$g_3(x)=\begin{cases}\dfrac{1}{\pi\sqrt{18-(x-1)^2}}, & 1-3\sqrt{2}\le x\le 1+3\sqrt{2}\\[2ex] 0, & \text{elsewhere}\end{cases}\tag{4.51}$$

Validating one's assumptions of the underlying randomness is, of course, essential to the modeling questions. One method is to look at a plot of the histogram of the measured data and superimpose the PDF over this plot. This is demonstrated in Figure 4.17b for the Gaussian, uniform, and sinusoidal pdfs. As is evident form this plot, the Gaussian PDF models the randomness the best in this case.

4.4.1 Modeling a Gaussian Mixture Using the Expectation-Maximization Algorithm

In some measurement situations, randomness can come from multiple sources, and one that cannot be modeled by a single Gaussian or similar pdf. We saw such a situation in Figure 4.11 and Figure 4.12 where the histogram either was bimodal or contained outliers. To model either situation, we require a different approach than that just described. It is based on the expectation–maximization (EM) algorithm.[8] The algorithm begins by writing the general form of the PDF as

$$f(x)=\alpha_1 f_1(x,\mu_1,\sigma_1)+\cdots+\alpha_i f_i(x,\mu_i,\sigma_i)+\cdots+\alpha_G f_G(x,\mu_G,\sigma_G)\tag{4.52}$$

where $f_i(x,\mu_i,\sigma_i)$ is the ith component PDF and α_i is its corresponding weighting factor. The sum of all weighting factors must total to one, that is, $\sum_{k=1}^{G}\alpha_k=1$. To ensure convergence, $f_i(x,\mu_i,\sigma_i)$ should be continuous in x over the range of interest. In this textbook, we shall limit our discussion to individual PDFs having Gaussian form, that is,

$$f_i(x,\mu_i,\sigma_i)=\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$

It is the objective of the EM algorithm to find the unknown parameters $\alpha_1,\alpha_2,\ldots,\alpha_G$, $\mu_1,\mu_2,\ldots,\mu_G$ and $\sigma_1,\sigma_2,\ldots,\sigma_G$ based on a set of data samples. The EM algorithm begins by first defining the data weighting function Φ_i for each PDF component (i from 1 to G) as follows

$$\Phi_i(x,\mu,\sigma,\alpha)=\frac{\alpha_i f_i(x,\mu_i,\sigma_i)}{\alpha_1 f_1(x,\mu_1,\sigma_1)+\cdots+\alpha_i f_i(x,\mu_i,\sigma_i)+\cdots+\alpha_G f_G(x,\mu_G,\sigma_G)} \tag{4.53}$$

where the model parameters are defined in vector form as

$$\begin{aligned}\mu&=[\mu_1, \quad \ldots, \quad \mu_i, \quad \ldots, \quad \mu_G]\\ \sigma&=[\sigma_1, \quad \ldots, \quad \sigma_i, \quad \ldots, \quad \sigma_G]\\ \alpha&=[\alpha_1, \quad \ldots, \quad \alpha_i, \quad \ldots, \quad \alpha_G]\end{aligned} \tag{4.54}$$

According to the development of the EM algorithm, the following set of compact recursive equations for each model parameter can be written:

$$\begin{aligned}
\alpha_1^{(n+1)}&=\frac{1}{N}\sum_{i=1}^{N}\Phi_1\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right) &\qquad \mu_1^{(n+1)}&=\frac{1}{N\alpha_1^{(n+1)}}\sum_{i=1}^{N}x_i\Phi_1\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)\\
\alpha_2^{(n+1)}&=\frac{1}{N}\sum_{i=1}^{N}\Phi_2\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right) &\qquad \mu_2^{(n+1)}&=\frac{1}{N\alpha_2^{(n+1)}}\sum_{i=1}^{N}x_i\Phi_2\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)\\
\alpha_G^{(n+1)}&=\frac{1}{N}\sum_{i=1}^{N}\Phi_G\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right) &\qquad \mu_G^{(n+1)}&=\frac{1}{N\alpha_G^{(n+1)}}\sum_{i=1}^{N}x_i\Phi_G\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)
\end{aligned}$$

$$\begin{aligned}
\sigma_1^{(n+1)}&=\sqrt{\frac{1}{N\alpha_1^{(n+1)}}\sum_{i=1}^{N}\left(x_i-\mu_1^{(n+1)}\right)^2\Phi_1\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)},\\
\sigma_2^{(n+1)}&=\sqrt{\frac{1}{N\alpha_2^{(n+1)}}\sum_{i=1}^{N}\left(x_i-\mu_2^{(n+1)}\right)^2\Phi_2\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)},\ \ldots,\\
\sigma_G^{(n+1)}&=\sqrt{\frac{1}{N\alpha_G^{(n+1)}}\sum_{i=1}^{N}\left(x_i-\mu_G^{(n+1)}\right)^2\Phi_G\left(x_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)}
\end{aligned} \tag{4.55}$$

The above model parameters are evaluated at each data point, together with an initial guess of the model parameters. The process is repeated until the change in any one step is less than some desired root-mean-square error tolerance, defined as

$$\text{error}^{(n+1)} < \text{TOLERANCE} \tag{4.56}$$

where

$$\text{error}^{(n+1)}=\begin{bmatrix}\left(\mu_1^{(n+1)}-\mu_1^{(n)}\right)^2+\cdots+\left(\mu_G^{(n+1)}-\mu_G^{(n)}\right)^2+\left(\sigma_1^{(n+1)}-\sigma_1^{(n)}\right)^2+\cdots\\ +\left(\sigma_G^{(n+1)}-\sigma_G^{(n)}\right)^2+\left(\alpha_1^{(n+1)}-\alpha_1^{(n)}\right)^2+\cdots+\left(\alpha_G^{(n+1)}-\alpha_G^{(n)}\right)^2\end{bmatrix}^{1/2}$$

To help clarify the situation, let us assume we want to fit two independent Gaussian distributions to a bimodal distribution such as that shown Figure 4.11. In this case, we shall assume that the PDF for the overall distribution is as follows

$$f(x) = \alpha_1 \left\{ \frac{1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} \right\} + \alpha_2 \left\{ \frac{1}{\sigma_2\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}} \right\} \quad \textbf{(4.57)}$$

where the unknown model parameters are $\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha_1$ and α_2. Next, we write the data weighting functions for this particular modeling situation as

$$\begin{aligned}
\Phi_1(x, \mu, \sigma, \alpha) &= \frac{\frac{\alpha_1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}}{\frac{\alpha_1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} + \frac{\alpha_2}{\sigma_2\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}}} \\
\Phi_2(x, \mu, \sigma, \alpha) &= \frac{\frac{\alpha_2}{\sigma_2\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}}}{\frac{\alpha_1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} + \frac{\alpha_2}{\sigma_2\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}}}
\end{aligned} \quad \textbf{(4.58)}$$

From Eq. (4.55), we write the update equations for each model parameter as

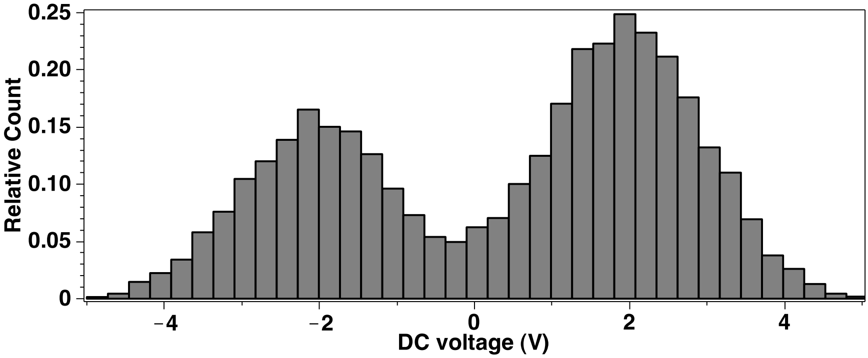
$$\begin{aligned}
\alpha_1^{(n+1)} &= \frac{1}{N}\sum_{i=1}^{N} \Phi_1\left(x_i, \mu^{(n)}, \sigma^{(n)}, \alpha^{(n)}\right), & \mu_1^{(n+1)} &= \frac{1}{N\,\alpha_1^{(n+1)}}\sum_{i=1}^{N} x_i \Phi_1\left(x_i, \mu^{(n)}, \sigma^{(n)}, \alpha^{(n)}\right) \\
\alpha_2^{(n+1)} &= \frac{1}{N}\sum_{i=1}^{N} \Phi_2\left(x_i, \mu^{(n)}, \sigma^{(n)}, \alpha^{(n)}\right), & \mu_2^{(n+1)} &= \frac{1}{N\,\alpha_2^{(n+1)}}\sum_{i=1}^{N} x_i \Phi_2\left(x_i, \mu^{(n)}, \sigma^{(n)}, \alpha^{(n)}\right)
\end{aligned}$$

$$\begin{aligned}
\sigma_1^{(n+1)} &= \sqrt{\frac{1}{N\,\alpha_1^{(n+1)}}\sum_{i=1}^{N}\left(x_i - \mu_1^{(n+1)}\right)^2 \Phi_1\left(x_i, \mu^{(n)}, \sigma^{(n)}, \alpha^{(n)}\right)} \\
\sigma_2^{(n+1)} &= \sqrt{\frac{1}{N\,\alpha_2^{(n+1)}}\sum_{i=1}^{N}\left(x_i - \mu_2^{(n+1)}\right)^2 \Phi_2\left(x_i, \mu^{(n)}, \sigma^{(n)}, \alpha^{(n)}\right)}
\end{aligned} \quad \textbf{(4.59)}$$

The following two examples will help to illustrate the EM approach more clearly.

EXAMPLE 4.5

A set of DC measurements was made on an amplifier where the normalized histogram of the data set revealed a bimodal distribution as shown below:



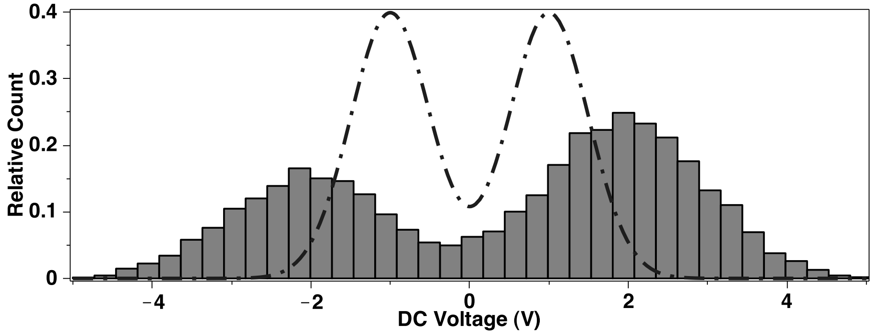
Assuming that the data consist of 10,000 samples and are stored in a vector, model the distribution of these data using a two-Gaussian mixture having the following general form:

$$f(x) = \alpha_1 \left\{ \frac{1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} \right\} + \alpha_2 \left\{ \frac{1}{\sigma_2\sqrt{2\pi}} e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}} \right\}$$

Solution:
To begin, we recognize from the histogram plot that two peaks occur at V = -2 and V = 2.2. We could use these values as our initial guess for the means of the two-Gaussian mixture. However, lets use some values quite a distance from these values, say V = -1.0 and V = 1.0. Next, recognizing that 67% of the area under one Gaussian represents ±1 sigma about the mean value, the standard deviation is about 1 V each. To stress the algorithm, we will start with a standard deviations of 0.5 V each. For lack of any further insight, we shall assume that each Gaussian is equally likely to occur resulting in the two alpha coefficients starting off with a value of 0.5 each. Using the data weighting functions for a two-Gaussian mixture found from Eq. (4.58), together with the update equations of Eq. (4.59), we wrote a computer program to iterate through these equations to find the following model parameters after 20 iterations:

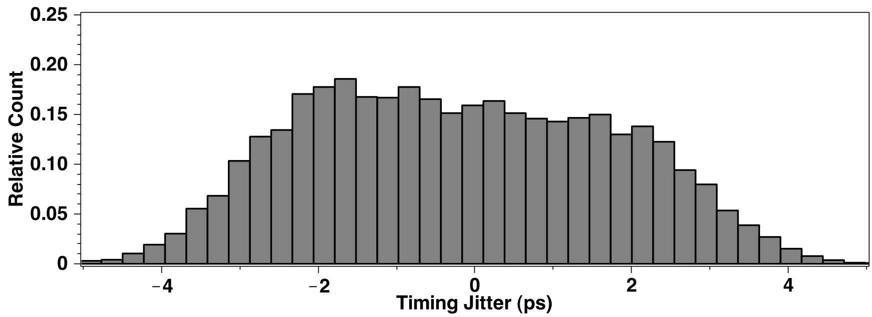
$$\alpha_1^{(21)} = 0.3959, \quad \mu_1^{(21)} = -2.0235, \quad \sigma_1^{(21)} = 0.98805$$
$$\alpha_2^{(21)} = 0.6041, \quad \mu_2^{(21)} = 1.9695, \quad \sigma_2^{(21)} = 1.0259$$

Based on Eq. (4.56), we have an RMS error of 0.97 mV. Superimposing the PDF predicted by the two-Gaussian mixture model onto the histogram data above (solid line) reveals a very good fit as illustrated below. Also, superimposed on the plot is the PDF curve corresponding to the initial guess (dash–dot line).



EXAMPLE 4.6

Ten thousand zero crossings were collected using an oscilloscope. A histogram of these zero-crossings is plotted below:



It is assumed that this data set consists of three independent Gaussian distributions described by

$$f(t)=\alpha_1\left\{\frac{1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(t-\mu_1)^2}{2\sigma_1^2}}\right\}+\alpha_2\left\{\frac{1}{\sigma_2\sqrt{2\pi}}e^{\frac{-(t-\mu_2)^2}{2\sigma_2^2}}\right\}+\alpha_3\left\{\frac{1}{\sigma_3\sqrt{2\pi}}e^{\frac{-(t-\mu_3)^2}{2\sigma_3^2}}\right\}$$

Using the EM algorithm, compute the model parameters of this PDF?

Solution:

We begin by first identifying the three data weighting functions as follows:

$$\Phi_1(t,\mu,\sigma,\alpha)=\frac{\dfrac{\alpha_1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(t-\mu_1)^2}{2\sigma_1^2}}}{\dfrac{\alpha_1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(t-\mu_1)^2}{2\sigma_1^2}}+\dfrac{\alpha_2}{\sigma_2\sqrt{2\pi}}e^{\frac{-(t-\mu_2)^2}{2\sigma_2^2}}+\dfrac{\alpha_3}{\sigma_3\sqrt{2\pi}}e^{\frac{-(t-\mu_3)^2}{2\sigma_3^2}}}$$

$$\Phi_2(t,\mu,\sigma,\alpha)=\frac{\dfrac{\alpha_2}{\sigma_2\sqrt{2\pi}}e^{\frac{-(t-\mu_2)^2}{2\sigma_2^2}}}{\dfrac{\alpha_1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(t-\mu_1)^2}{2\sigma_1^2}}+\dfrac{\alpha_2}{\sigma_2\sqrt{2\pi}}e^{\frac{-(t-\mu_2)^2}{2\sigma_2^2}}+\dfrac{\alpha_3}{\sigma_3\sqrt{2\pi}}e^{\frac{-(t-\mu_3)^2}{2\sigma_3^2}}}$$

$$\Phi_3(t,\mu,\sigma,\alpha)=\frac{\dfrac{\alpha_3}{\sigma_3\sqrt{2\pi}}e^{\frac{-(t-\mu_3)^2}{2\sigma_3^2}}}{\dfrac{\alpha_1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(t-\mu_1)^2}{2\sigma_1^2}}+\dfrac{\alpha_2}{\sigma_2\sqrt{2\pi}}e^{\frac{-(t-\mu_2)^2}{2\sigma_2^2}}+\dfrac{\alpha_3}{\sigma_3\sqrt{2\pi}}e^{\frac{-(t-\mu_3)^2}{2\sigma_3^2}}}$$

Next, we identify the Gaussian mixture model fit recursive equations as

$$\alpha_1^{(n+1)}=\frac{1}{N}\sum_{i=1}^{N}\Phi_1\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right),\quad \mu_1^{(n+1)}=\frac{1}{N\alpha_1^{(n+1)}}\sum_{i=1}^{N}t_i\Phi_1\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)$$

$$\alpha_2^{(n+1)}=\frac{1}{N}\sum_{i=1}^{N}\Phi_2\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right),\quad \mu_2^{(n+1)}=\frac{1}{N\alpha_2^{(n+1)}}\sum_{i=1}^{N}t_i\Phi_2\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)$$

$$\alpha_3^{(n+1)}=\frac{1}{N}\sum_{i=1}^{N}\Phi_3\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right),\quad \mu_3^{(n+1)}=\frac{1}{N\alpha_3^{(n+1)}}\sum_{i=1}^{N}t_i\Phi_3\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)$$

$$\sigma_1^{(n+1)}=\sqrt{\frac{1}{N\alpha_1^{(n+1)}}\sum_{i=1}^{N}\left(t_i-\mu_1^{(n+1)}\right)^2\Phi_1\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)}$$

$$\sigma_2^{(n+1)}=\sqrt{\frac{1}{N\alpha_2^{(n+1)}}\sum_{i=1}^{N}\left(t_i-\mu_2^{(n+1)}\right)^2\Phi_2\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)}$$

$$\sigma_3^{(n+1)}=\sqrt{\frac{1}{N\alpha_3^{(n+1)}}\sum_{i=1}^{N}\left(t_i-\mu_3^{(n+1)}\right)^2\Phi_3\left(t_i,\mu^{(n)},\sigma^{(n)},\alpha^{(n)}\right)}$$

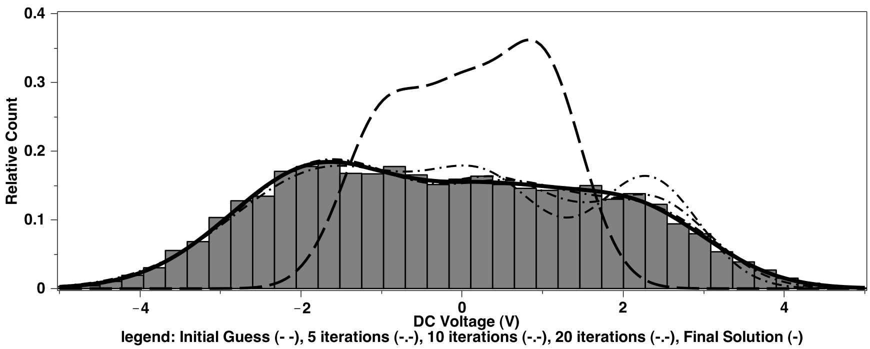
Using the captured data, together with our initial guess of the model parameters,

$$\begin{aligned}
&\alpha_1^{(1)}=0.3,\quad \mu_1^{(1)}=-1.0,\quad \sigma_1^{(1)}=0.5\\
&\alpha_2^{(1)}=0.3,\quad \mu_2^{(1)}=0,\quad \sigma_2^{(1)}=0.5\\
&\alpha_3^{(1)}=0.4,\quad \mu_3^{(1)}=1.0,\quad \sigma_3^{(1)}=0.5
\end{aligned}$$

we iterate through the above equations 100 times and find the following set of parameters:

$$\alpha_1^{[101]} = 0.4782, \quad \mu_1^{[101]} = -1.8150, \quad \sigma_1^{[101]} = 1.0879$$
$$\alpha_2^{[101]} = 0.2644, \quad \mu_2^{[101]} = 0.4486, \quad \sigma_2^{[101]} = 0.9107$$
$$\alpha_3^{[101]} = 0.2572, \quad \mu_3^{[101]} = 2.1848, \quad \sigma_3^{[101]} = 0.9234$$

Double-checking the curve fit, we superimpose the final PDF model on the histogram of the original data below, together with intermediate results after 5, 10, and 20 iterations. We also illustrate our first guess of the PDF. As is clearly evident from below, we have a very good fit, and this occurs after about 20 iterations.



The EM algorithm is a well-known statistical tool for solving maximum likelihood problems. It is interesting to note that a survey of the literature via the Internet reveals that this algorithm is used in a multitude of applications from radar to medical imaging to pattern recognition. The algorithm provides an iterative formula for the estimation of the unknown parameters of the Gaussian mixture and can be proven to monotonically increase the likelihood that the estimate is correct in each step of the iteration. Nonetheless, the EM algorithm is not without some drawbacks. An initial guess is required to start the process; the closer to the final solution, the better, because the algorithm will converge much more quickly. And secondly, the algorithm cannot alter the assumption about the number of Gaussians in the mixture and hence can lead to sub-optimum results. Thirdly, local maxima or saddle points can prematurely stop the search process on account of the nonlinearities associated between the densities and their corresponding parameters μ and σ. Nonetheless, the EM algorithm provides an excellent means to model jitter distribution problems.[8] In fact, due to the underlying importance of the EM algorithm to statistics in general, advances are being made to the EM algorithm regularly, such as that described in Vlassis and Likas.[9]

4.4.2 Probabilities Associated with a Gaussian Mixture Model

The ultimate goal of any pdf-modeling problem is to answer questions related to the probability that a specific type of event can occur. We encounter this question previously in Subsections

4.2.2–4.2.3 for a single Gaussian pdf. Here we shall extend this concept to an arbitrary Gaussian mixture of the general form,

$$f\left(x\right)=\alpha_1\left\{\frac{1}{\sigma_1\sqrt{2\pi}}e^{\frac{-\left(x-\mu_1\right)^2}{2\sigma_1^2}}\right\}+\alpha_2\left\{\frac{1}{\sigma_2\sqrt{2\pi}}e^{\frac{-\left(x-\mu_2\right)^2}{2\sigma_2^2}}\right\}+\cdots+\alpha_G\left\{\frac{1}{\sigma_G\sqrt{2\pi}}e^{\frac{-\left(x-\mu_G\right)^2}{2\sigma_G^2}}\right\} \quad \textbf{(4.60)}$$

The probability that a random event X lies in the interval bounded by a and b is given by

$$P(a\le X\le b)=\int_a^b f\left(x\right)dx \quad \textbf{(4.61)}$$

Substituting Eq. (4-60), we write

$$P(a\le X\le b)=\int_a^b\left[\frac{\alpha_1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}+\frac{\alpha_2}{\sigma_2\sqrt{2\pi}}e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}}+\cdots+\frac{\alpha_G}{\sigma_G\sqrt{2\pi}}e^{\frac{-(x-\mu_G)^2}{2\sigma_G^2}}\right]dx \quad (4.62)$$

$$P(a\le X\le b)=\alpha_1\int_a^b\frac{1}{\sigma_1\sqrt{2\pi}}e^{\frac{-\left(x-\mu_1\right)^2}{2\sigma_1^2}}dx+\alpha_2\int_a^b\frac{1}{\sigma_2\sqrt{2\pi}}e^{\frac{-\left(x-\mu_2\right)^2}{2\sigma_2^2}}dx+\cdots+\alpha_G\int_a^b\frac{1}{\sigma_G\sqrt{2\pi}}e^{\frac{-\left(x-\mu_G\right)^2}{2\sigma_G^2}}dx \quad \textbf{(4.63)}$$

Finally, using the standard Gaussian CDF notation, we write Eq. (4.63) as

$$\begin{aligned}P(a\le X\le b)=&\ \alpha_1\left[\Phi\left(\frac{b-\mu_1}{\sigma_1}\right)-\Phi\left(\frac{a-\mu_1}{\sigma_1}\right)\right]+\alpha_2\left[\Phi\left(\frac{b-\mu_1}{\sigma_2}\right)-\Phi\left(\frac{a-\mu_2}{\sigma_2}\right)\right]+\cdots\\&+\alpha_G\left[\Phi\left(\frac{b-\mu_G}{\sigma_G}\right)-\Phi\left(\frac{a-\mu_G}{\sigma_G}\right)\right]\end{aligned} \quad \textbf{(4.64)}$$

It should be evident from the above result that the probabilities associated with a Gaussian mixture model is a straightforward extension of a single Gaussian PDF and offers very little additional difficulty.

EXAMPLE 4.7

A bimodal distribution is described by the following pdf:

$$f(x)=\frac{\alpha_1}{\sigma_1\sqrt{2\pi}}e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}+\frac{\alpha_2}{\sigma_2\sqrt{2\pi}}e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}}$$

with coefficients,

$$\alpha_1=0.3959,\quad \mu_1=-2.0235,\quad \sigma_1=0.98805$$
$$\alpha_2=0.6041,\quad \mu_2=1.9695,\quad \sigma_2=1.0259$$

A plot of this PDF is provided in Example 4.5 above. What is the probability that some single random event ***X*** is greater than 0? What about greater than 10?

Solution:
According to Eq. (4.64), we can write

$$P(0\le X<\infty)=\alpha_1\left[\Phi\left(\frac{\infty-\mu_1}{\sigma_1}\right)-\Phi\left(\frac{0-\mu_1}{\sigma_1}\right)\right]+\alpha_2\left[\Phi\left(\frac{\infty-\mu_1}{\sigma_2}\right)-\Phi\left(\frac{0-\mu_2}{\sigma_2}\right)\right]$$

which further simplifies to

$$P(0\le X)=2-\alpha_1\Phi\left(\frac{-\mu_1}{\sigma_1}\right)-\alpha_2\Phi\left(\frac{-\mu_2}{\sigma_2}\right)$$

Substituting the appropriate Gaussian mixture model parameters, we write

$$P(0\le X)=2-0.3959\times\Phi\left(\frac{2.0235}{0.98805}\right)-0.6041\times\Phi\left(\frac{-1.9695}{1.0259}\right)$$

Solving each standard CDF term, we write

$$P(0\le X)=2-0.3959\times 0.9797-0.6041\times 0.02744$$

which finally reduces to

$$P(0\le X)=0.5955\,.$$

Likewise, for $P(10\le X)$ we write

$$P(10\le X)=2-0.3959\times\Phi\left(\frac{10+2.0235}{0.98805}\right)-0.6041\times\Phi\left(\frac{10-1.9695}{1.0259}\right)$$

which, when simplified, leads to

$$P(10 \le X) \approx 1 - 0.6041 \times \Phi\left(\frac{10 - 1.9695}{1.0259}\right) = 1.5 \times 10^{-15}$$

Exercises

4.14. A measurement set is found to consist of two Gaussian distributions having the following model parameters: $\alpha_1 = 0.7$, $\mu_1 = -2.$, $\sigma_1 = 1$, $\alpha_2 = 0.3$, $\mu_2 = 2$, and $\sigma_2 = 0.5$. What is the probability that some single random measurement X is greater than 0? What about greater than 2? What about less than –3?

ANS. $P[0 \le X] = 0.3159$; $P[2 \le X] = 0.1500$; $P[X \le -3] = 0.1110$.

4.5 SUMS AND DIFFERENCES OF RANDOM VARIABLES

Up to this point in the chapter we have described various mathematical functions that can be used to represent the behavior of a random variable, such as the bimodal shape of a histogram of data samples. In this section we will go one step further and attempt to describe the general shape of a probability distribution as a sum of several random variables. This perspective provides greater insight into the measurement process and provides an understanding of the root cause of measurement variability. The test engineer will often make use of these principles to identify the root cause of device repeatability problems or product variability issues.

Consider a sum of two independent random variables X_1 and X_2 as follows:

$$Y = X_1 + X_2 \tag{4.65}$$

If the PDF of X_1 and X_2 are described by $f_1(x)$ and $f_2(x)$, then it has been shown[2] that the PDF of the random variable Y, denoted by $f_y(x)$, is given by the convolution integral as

$$f_y(x) = \int_{-\infty}^{\infty} f_1(x - \tau) f_2(\tau) d\tau \tag{4.66}$$

It is common practice to write this convolution expression using a shorthand notation as follows

$$f_y(x) = f_1(x) \otimes f_2(x) \tag{4.67}$$

The above theory extends to the summation of multiple independent random variables. For example,

$$Y = X_1 + X_2 + X_3 + \cdots + X_R \tag{4.68}$$

where the PDF of the total summation is expressed as

$$f_y(x) = f_1(x) \otimes f_2(x) \otimes f_3(x) \otimes \cdots \otimes f_R(x) \quad \textbf{(4.69)}$$

The above theory can also be extended to sums or differences of scaled random variables according to

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_R X_R \quad \textbf{(4.70)}$$

where the PDF of the sum is given by

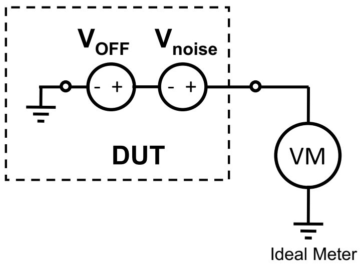
$$f_y(x) = \frac{1}{|\alpha_1|} f_1\left(\frac{x}{\alpha_1}\right) \otimes \frac{1}{|\alpha_2|} f_2\left(\frac{x}{\alpha_2}\right) \otimes \frac{1}{|\alpha_3|} f_3\left(\frac{x}{\alpha_3}\right) \otimes \cdots \otimes \frac{1}{|\alpha_R|} f_R\left(\frac{x}{\alpha_R}\right) \quad \textbf{(4.71)}$$

The following three examples will help to illustrate the application of this theory to various measurement situations.

EXAMPLE 4.8

A DUT with zero input produces an output offset of 10 mV in the presence of a zero mean Gaussian noise source with a standard deviation of 10 mV, as illustrated in the figure accompanying. Write a PDF description of the voltage that the voltmeter reads.

Solution:



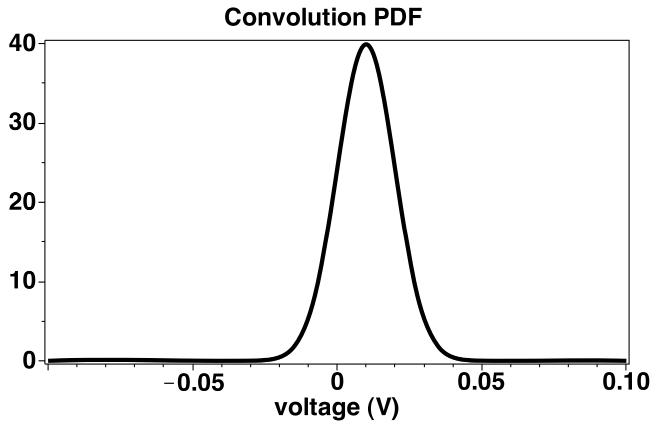
Let us designate the offset voltage V_{OFF} as a random variable and the noise produced by the DUT (V_{noise}) as the other random variable. As the offset voltage is constant for all time, we can model the PDF of this voltage term using a unit impulse centered around a 10-mV level as $f_1(v) = \delta\left(v - 10\times10^{-3}\right)$. The PDF of the noise source is simply $f_2(v) = \frac{1}{\left(10\times10^{-3}\right)\sqrt{2\pi}} e^{-v^2 / 2\left(10\times10^{-3}\right)^2}$. As the noise signal is additive with the offset level, we find the PDF of the voltage seen by the voltmeter as the convolution of the two individual pdfs, leading to

$$f_{VM}(v) = f_1(v) \otimes f_2(v) = \int_{-\infty}^{\infty} \delta\left(v - \tau - 10\times10^{-3}\right) \frac{1}{\left(10\times10^{-3}\right)\sqrt{2\pi}} e^{-\tau^2 / 2\left(10\times10^{-3}\right)^2} d\tau$$

Next, using the sifting property of the impulse function, we write the voltmeter PDF as

$$f_{VM}(x) = \frac{1}{\left(10\times10^{-3}\right)\sqrt{2\pi}} e^{-\left(v - 10\times10^{-3}\right)^2 / 2\left(10\times10^{-3}\right)^2}$$

A plot of this PDF is shown below:



The voltage seen by the voltmeter is a Gaussian distrubed random variable with a mean of 10 mV and a standard deviation of 10 mV.

While the PDF in the above example could have been written directly from the problem description, it serves to illustrate the relationship of the observed data to the physics of the problem. Let us repeat the above example, but this time let us assume that the offset voltage can take on two different levels.

EXAMPLE 4.9

A DUT with zero input produces an output offset that shifts equally between two different levels of 10 mV and –20 mV, depending on the DUT temperature. The noise associated with the DUT is a zero mean Gaussian noise source with a standard deviation of 10 mV. Write a PDF description of the voltage that the voltmeter reads.

Solution:

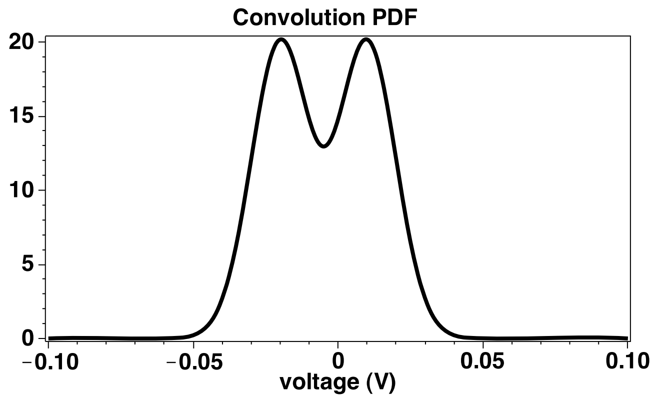
As the offset voltage can take on two different voltage values, the PDF of the offset voltage becomes $f_1(v) = 1/2 \cdot \delta(v + 20 \times 10^{-3}) + 1/2 \cdot \delta(v - 10 \times 10^{-3})$. The PDF of the noise signal is described as $f_2(v) = \dfrac{1}{(10 \times 10^{-3})\sqrt{2\pi}} e^{-v^2 / 2(10 \times 10^{-3})^2}$, leading to the PDF of the voltage seen by the voltmeter as

$$f_{VM}(v) = \int_{-\infty}^{\infty} \left[\frac{1}{2} \cdot \delta(v - \tau + 20 \times 10^{-3}) + \frac{1}{2} \cdot \delta(v - \tau - 10 \times 10^{-3}) \right] \frac{1}{(10 \times 10^{-3})\sqrt{2\pi}} e^{-\tau^2 / 2(10 \times 10^{-3})^2} d\tau$$

which reduces to

$$f_{VM}(x)=\frac{1}{2}\frac{1}{\left(10\times10^{-3}\right)\sqrt{2\pi}}e^{-\left(v+20\times10^{-3}\right)^2 / 2\left(10\times10^{-3}\right)^2}+\frac{1}{2}\frac{1}{\left(10\times10^{-3}\right)\sqrt{2\pi}}e^{-\left(v-10\times10^{-3}\right)^2 / 2\left(10\times10^{-3}\right)^2}$$

A plot of this PDF is shown below:



Here we see that the voltage seen by the voltmeter is a bimodal Gaussian distrubed random variable with centers of –20 mV and 10 mV.

Let us again repeat the above example, but this time let us assume that the offset voltage can take on a range of different levels.

EXAMPLE 4.10

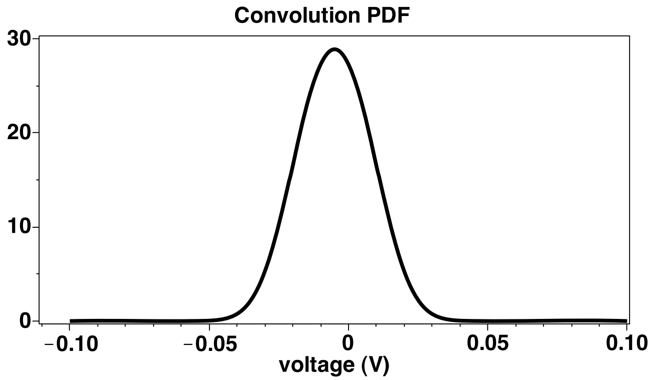
A DUT with zero input produces an output offset that is uniformly distributed between –20 mV and 10 mV. The noise associated with the DUT is a zero mean Gaussian noise source with a standard deviation of 10 mV. Write a PDF description of the voltage that the voltmeter reads.

Solution:

As the offset voltage can take on a range of voltage values between –20 mV and 10 mV, the PDF of the offset voltage becomes $f_1(v)=\frac{1}{30\times10^{-3}}$ for $-20\times10^{-3}\le v\le10\times10^{-3}$. The PDF of the noise signal is described as $f_2(v) = \frac{1}{(10\times10^{-3})\sqrt{2\pi}} e^{-v^2 / 2(10\times10^{-3})^2}$, leading to the PDF of the voltage seen by the voltmeter as

$$f_{VM}(v) = \int_{-20\times10^{-3}}^{10\times10^{-3}} \frac{1}{30\times10^{-3}} \frac{1}{(10\times10^{-3})\sqrt{2\pi}} e^{-(v-\tau)^2 / 2(10\times10^{-3})^2} d\tau$$

Unlike the previous two examples, there is no obvous answer to this intergral equation. Instead, we resort to solving this integral using a numerical integration method. The solution is shown in the plot below.



4.5.1 The Central Limit Theorem

As mentioned previously, the central limit theorem states that if a large number of random variables are independent, then the density of their sum Eq. (4.68) tends to a Gaussian distribution with mean μ and standard deviation σ as $R \to \infty$ described by

$$f_y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2} \tag{4.72}$$

This fact can be viewed as a property of repeat application of the convolution operation on positive functions. For instance, let us consider the sum of three random variables, each with a uniform PDF between 0 and 1, that is, for i from 1 to 3, we obtain

$$f_i(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.73}$$

Applying the convolution on the first two terms of the summation, we write

$$f_{X_1+X_2}(x)=\int_{-\infty}^{\infty}\left[\begin{cases}1, & 0\le x-\tau\le 1\\ 0, & \text{otherwise}\end{cases}\right]\left[\begin{cases}1, & 0\le \tau\le 1\\ 0, & \text{otherwise}\end{cases}\right]d\tau=\begin{cases}0, & x<0\\ x, & x<1\\ 2-x, & x<2\\ 0, & 2\le x\end{cases} \tag{4.74}$$

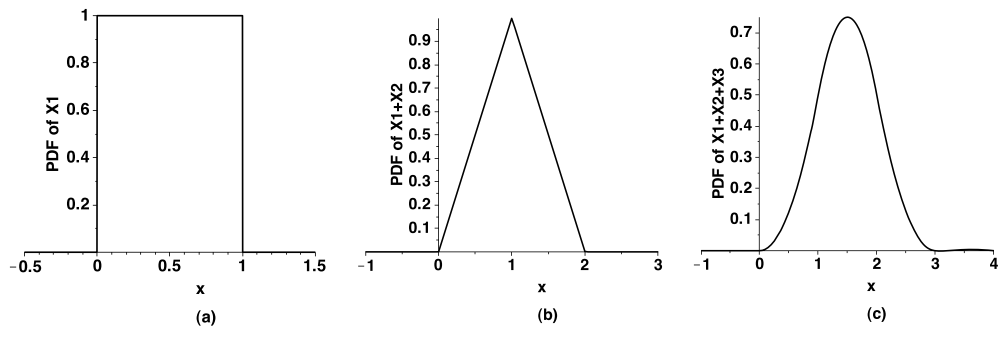
Next we repeat this integration with the above result and write

$$f_{X_1+X_2+X_3}(x)=\int_{-\infty}^{\infty}\left[\begin{cases}1, & 0\le x-\tau\le 1\\ 0, & \text{otherwise}\end{cases}\right]\left[\begin{cases}0, & \tau<0\\ \tau, & \tau<1\\ 2-\tau, & \tau<2\\ 0, & 2\le \tau\end{cases}\right]d\tau=\begin{cases}0, & x<0\\ x^2/2, & x<1\\ -1-(x-1)^2/2+2x-x^2/2, & x<2\\ 4-2x+(x-1)^2/2, & x<3\\ 0, & 3\le x\end{cases} \tag{4.75}$$

Figure 4.18 illustrates how the uniform distribution described above tends toward a Gaussian function by repeated application of the convolution operation.

The importance of the central limit theorem for analog and mixed-signal test is that repeated measurements of some statistic, such as a mean or standard deviation, generally follows a Gaussian distribution, regardless of the underlying random behavior from which the samples were derived. This fact will figure prominently in the concept of measurement repeatability and accuracy of the next chapter.

Figure 4.18. Illustrating repeated application of the convolution operation (a) PDF of X_1, (b) PDF of $X_1 + X_2$, (c) PDF of $X_1 + X_2 + X_3$.



Exercises

4.15. Two random variables are combined as the difference according to $y = X_1 - X_2$, where the PDF for X_1 is given by $f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2 / 2\sigma_1^2}$ and the PDF for X_2 as $f_2(x) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x-\mu_2)^2 / 2\sigma_2^2}$. What is the PDF of Y?

ANS. $f_Y(x) = \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2}\sqrt{2\pi}} e^{-[x-(\mu_1-\mu_2)]^2 / 2(\sigma_1^2+\sigma_2^2)}$

4.6 SUMMARY

There are literally hundreds, if not thousands, of ways to view and process data gathered during the production testing process. In this chapter, we have examined only a few of the more common data displays, such as the datalog, wafer map, and histogram. We reviewed several statistical methods in which to model large quantities of data with simple mathematical models. These models capture the underlying structure of randomness, and not the randomness itself. For the most part, we focused on models based on Gaussian or normal distributions. Probability theory was then introduced to answer questions related to the likelihood that an event will occur. This idea is central to most, if not all, production measurement approaches, because obtaining just the right amount of information with the least effort is the objective of any production engineer. We also looked at several other types of distributions that are commonly found in analog and mixed-signal test applications, such as the uniform, binomial and sinusoidal distribution. Near the end of the chapter the idea of modeling an arbitrary distribution as mixture of multiple Gaussian distributions was introduced. A numerical technique based on the expectation-maximization algorithm was described. This idea will figure prominently in Chapter 14 where we quantify clock and data jitter associated with high-speed digital I/O channels. Finally, we looked at the operation of adding several random variables together from a probability perspective. The theory related to this operation can often help the test engineer identify sources of variability. Finally, we concluded this chapter with a discussion on the central limit theorem, which essentially states that any operation that adds multiple samples together such as a mean or standard deviation metric, will behave as a Gaussian random variable. This fact will figure significantly in the next chapter on measurement repeatability and accuracy.

PROBLEMS

4.1. The thickness of printed circuit boards is an important characteristic. A sample of eight boards had the following thickness (in millimeters): 1.60, 1.55, 1.65, 1.57, 1.55, 1.62, 1.52, and 1.67. Calculate the sample mean, sample variance, and sample standard deviation. What are the units of measurement for each statistic?

4.2. A random variable X has the probability density function

$$f(x) = \begin{cases} ce^{-3x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

(a) Find the value of the constant c.

(b) Find $P(1 < X < 2)$

(c) Find $P(X \le 1)$
(d) Find $P(X < 1)$
(e) Find the cumulative distribution function $F(x)$.

4.3. Compare the tabulated results for $\Phi(z)$ listed in the table of Appendix A with those generated by Eq. (4.18). Provide a plot of the two curves. What is the worst-case error?

4.4. For the following specified values of z, estimate the value of $\Phi(z)$ using Eq. (4.18) and compare the results with those obtained from the table of Appendix A. Use linear interpolation where necessary.
(a) $z = -3.0$, (b) $z = -1.9$, (c) $z = -0.56$, (d) $z = -0.24$, (e) $z = 0.0$,
(f) $z = -0.09$, (g) $z = +0.17$, (h) $z = +3.0$, (i) $z = +5.0$, (j) $z = -5.0$.

4.5. Using first principles, show that the following relationships are true.
$P(X < b) = 1 - P(X > b)$
$P(a < X < b) = 1 - P(X < a) - P(X > b)$
$P(|X| > c) = 1 - P(-c < X < c)$

4.6. Calculate the following probabilities associated with a Gaussian-distributed random variable having the following mean and standard deviation values:
(a) $P(0 < X < 30\text{ mV})$ when $\mu = 0$, $\sigma = 10$ mV
(b) $P(-30\text{ mV} < X < 30\text{ mV})$ when $\mu = 1$ V, $\sigma = 10$ mV
(c) $P(-1.5\text{ V} < X < 1.4\text{ V})$ when $\mu = 0$, $\sigma = 1$ V
(d) $P(-300\text{ mV} < X < -100\text{ mV})$ when $\mu = -250$ mV, $\sigma = 50$ mV
(e) $P(X < 250\text{ mV})$ when $\mu = 100$ mV, $\sigma = 100$ mV
(f) $P(-200\text{ mV} > X)$ when $\mu = -75$ mV, $\sigma = 150$ mV
(g) $P(|X| < 30\text{ mV})$ when $\mu = 0$, $\sigma = 10$ mV
(h) $P(|X| > 30\text{ mV})$ when $\mu = 0$, $\sigma = 10$ mV

4.7. In each of the following equations, find the value of z that makes the probability statement true. Assume a Gaussian distributed random variable with zero mean and unity standard deviation.
(a) $\Phi(z) = 0.9452$
(b) $P(Z < z) = 0.7881$
(c) $P(Z < z) = 0.2119$
(d) $P(Z > z) = 0.2119$
(e) $P(|Z| < z) = 0.5762$

4.8. In each of the following equations, find the value of x that makes the probability statement true. Assume a Gaussian distributed random variable with $\mu = -1$ V and $\sigma = 100$ mV.
(a) $P(X < x) = 0.9$
(b) $P(X < x) = 0.3$
(c) $P(X > x) = 0.3$
(d) $P(|X| < x) = 0.4$

4.9. It has been observed that a certain measurement is a Gaussian-distributed random variable, of which 25% are less than 20 mV and 10% are greater than 70 mV. What are the mean and standard deviation of the measurements?

4.10. If X is a uniformly distributed random variable over the interval (0, 100), what is the probability that the number lies between 23 and 33? What are the mean and standard deviation associated with this random variable?

4.11. It has been observed that a certain measurement is a uniformly distributed random variable, of which 25% are less than 20 mV and 10% are greater than 70 mV. What are the mean and standard deviation of the measurements?

4.12. A DC offset measurement is repeated many times, resulting in a series of values having an average of −110 mV. The measurements exhibit a standard deviation of 51 mV. What is the probability that any single measurement will return a positive value? What is the probability that any single measurement will return a value less than -200 mV? Provide sketches of the pdf, label critical points, and highlight the areas under the PDF that corresponds to the probabilities of interest.

4.13. A series of AC gain measurements are found to have an average value of 10.3 V/V and a variance of 0.1 $(\text{V/V})^2$. What is the probability that any single measurement will return a value less than 9.8 V/V? What is the probability that any single measurement will lie between 10.0 V/V and 10.5 V/V? Provide sketches of the pdf, label critical points, and highlight the areas under the PDF that correspond to the probabilities of interest.

4.14. A noise measurement is repeated many times, resulting in a series of values having an average RMS value of 105 μV. The measurements exhibit a standard deviation of 21 μV RMS. What is the probability that any single noise measurement will return an RMS value larger than 140 μV? What is the probability that any single measurement will return an RMS value less than 70 μV? What is the probability that any single measurement will return an RMS value between 70 and 140 μV? Provide sketches of the pdf, label critical points, and highlight the areas under the PDF that correspond to the probabilities of interest.

4.15. The gain of a DUT is measured with a meter 10 times, resulting in the following sequence of readings: 0.987 V/V, 0.966 V/V, 0.988 V/V, 0.955 V/V, 1.00 V/V, 0.978 V/V, 0.981 V/V, 0.979 V/V, 0.978 V/V, 0.958 V/V. Write an expression for the PDF for this measurement set assuming the distribution is Gaussian.

4.16. The total harmonic distortion (THD) of a DUT is measured 10 times, resulting in the following sequence of readings expressed in percent: 0.112, 0.0993, 0.0961, 0.0153, 0.121, 0.00911, 0.219, 0.101, 0.0945, 0.0767. Write an expression for the PDF for this measurement set assuming the distribution is Gaussian.

4.17. The total harmonic distortion (THD) of a DUT is measured 10 times, resulting in the following sequence of readings expressed in percent: 0.112, 0.0993, 0.0961, 0.0153, 0.121, 0.00911, 0.219, 0.101, 0.0945, 0.0767. Write an expression for the PDF for this measurement set, assuming that the distribution is uniformily distributed.

4.18. A 6-bit ADC is excited by a 1024-point full-scale sine wave described by $1.0\sin\left(2\pi\frac{1}{1024}\cdot n\right)+2.0,\ \ n=0,1,...,1023$. Arrange the voltage axis between 1.00 and 3.00 V into 2^6 equal regions and count the number of samples from the sine-wave that fall into each region. Compare with the actual counts with that predicted by theory.

4.19. An 8-bit ADC operates over a range of 0 – 10 V. Using a 5-V 1024-point input sine wave centered around a 5-V level, determine the expected code distribution. Verify your results using a histogram of captured data.

4.20. A 6-bit ADC operates over a range of 0 – 10 V. Using a 3-V 1024-point input sine wave centered around a 5-V level, how many times does the code $n = 50$ appear at the output?

4.21. A 6-bit ADC operates over a range of 0 – 10 V. Using a 3-V 1024-point input sine wave centered around a 6-V level, how many times does the code $n = 50$ appear at the output?

4.22. A random number generator produces the digits $N = -10, -9, -8, \ldots, 9, 10$ uniformly. What is the cumulative distribution function of this data set? What is the mean and standard deviation of this pdf?

4.23. A random number generator produces the digits: $N = 100, 101, \ldots, 999, 1000$ uniformly. What is the cumulative distribution function of this data set? What is the mean and standard deviation of this pdf?

4.24. A PDF has the form $f(x) = \begin{cases} 4\left(x-x^3\right), & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$. What is the corresponding cumulative distribution function? What is the mean and standard deviation of this pdf?

4.25. A lot of 1000 devices is tested during production, where it is assumed that a single device has a probability of 0.99 of passing. What is the probability that five bad dies will show up during this test?

4.26. A lot of 1000 devices is tested during production. The device is categorized into five bins, where bins 1 to 4 each have a probability of 0.1 of occurring and bin 5 with a probability of 0.6. What is the probability that 75 dies will appear in bin 1 or in bin 2?

4.27. A sample of 4 dies are selected from a large lot and tested. If the probability of passing is 5/6, what is the PDF of four bad dies appearing during this test?

4.28. Data is transmitted over a channel with a single-bit error probability of 10^{-12}. If 10^{13} bits are transmitted, what is the probability of one bit error? What is the average number of errors expected?

4.29. Data are transmitted over a channel with a single-bit error probability of 10^{-10}. If 10^{9} bits are transmitted, what is the probability of one bit error. What is the average number of errors expected?

4.30. A measurement set is found to consist of two Gaussian distributions having the following model parameters: $\alpha_1 = 0.5, \mu_1 = 2.5, \sigma_1 = 0.5,\ \alpha_2 = 0.5, \mu_2 = 3.0,$ and $\sigma_2 = 0.2$. What is the probability that some single random measurement X is greater than 0? What about greater than 2? What about less than 3?

4.31. A measurement set is found to consist of three Gaussian distributions having the following model parameters: $\alpha_1 = 0.3, \mu_1 = 1.0, \sigma_1 = 1,\ \alpha_2 = 0.4, \mu_2 = -1.0, \sigma_2 = 1.5,$ $\alpha_3 = 0.3, \mu_3 = 0,$ and $\sigma_3 = 0.5$. What is the probability that some single random measurement X is greater than 0? What about greater than 2? What about less than -3?

4.32. A DUT with zero input produces an output offset of -5 mV in the presence of a zero mean Gaussian noise source with a standard deviation of 3 mV. Write a PDF description of the voltage that the voltmeter reads.

4.33. A DUT with zero input produces an output offset that shifts equally between two different levels of 2 mV and –2 mV, depending on the DUT temperature. The noise associated with the DUT is a zero mean Gaussian noise source with a standard deviation of 5 mV. Write a PDF description of the voltage that the voltmeter reads.

4.34. A DUT with zero input produces an output offset that is uniformly distributed between -5 mV and 7 mV. The noise associated with the DUT is a zero mean Gaussian noise source with a standard deviation of 5 mV. Write a PDF description of the voltage that the voltmeter reads.

4.35. Two random variables are combined as the difference according to $y = 1/2\ X_1 + X_2$ where the PDF for X_1 is given by $f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2 / 2\sigma_1^2}$ and the PDF for X_2 as $f_2(x) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x-\mu_2)^2 / 2\sigma_2^2}$. What is the PDF of Y?

4.36. Two random variables are combined as the difference according to $y = 1/2\ X_1 + 3X_2$, where the PDF for X_1 is Gaussian with a mean of 1 and a standard deviation of 1. The PDF for X_2 is Gaussian with a mean of -1 and a standard deviation of 2. What is the PDF of Y?

4.37. Two random variables are combined as the difference according to $y = X_1 - 1/2\ X_2$, where the PDF for X_1 is Gaussian with a mean of 2 and a standard deviation of 2. The PDF for X_2 is Gaussian with a mean of –1 and a standard deviation of 1. What is the PDF of Y?

REFERENCES

1. K. Baker and J. van Beers, Shmoo plotting: The black art of IC testing, in *Proceedings International Test Conference*, 1996, pp. 932, 933.
2. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edition, McGraw-Hill, New York, December 1991, ISBN 0070484775.
3. J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, John Wiley & Sons, New York, April 1986, ISBN 0471040002.
4. G. R. Cooper and C. D. McGillem, *Probabilistic Methods of Signal and System Analysis*, 3rd edition, Oxford University Press, New York, 1999, ISBN 0195123549.
5. M. J. Kiemele, S. R. Schmidt and R. J. Berdine, *Basic Statistics, Tools for Continuous Improvement*, 4thedition, Air Academy Press, Colorado Springs, CO, 1997, ISBN 1880156067, pp. 9–71.
6. P.O. Börjesson and C. E. W. Sundberg, Simple approximations of the error function $Q(x)$ for communication applications, *IEEE Transactions On Communications*, pp. 639–43, March 1979.
7. J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey, *Graphical Methods for Data Analysis*, Wadsworth, Belmont, CA, 1983.
8. A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from in-complete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, **39**(1), pp. 1–38, November 1977.
9. N. Vlassis and A. Likas, A kurtosis-based dynamic approach to Gaussian mixture modeling, *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, **29**(4), pp. 393–399, July 1999.