12

Estimation

Designing a good sensor is the first step towards interfacing the physical world with the virtual world. Two additional steps are required before the data obtained from a sensor can be useful. The first step deals with determining whether the data represent the physical reality and the second step deals with understanding the meaning of the sensed data. Erroneous representations or interpretations of sensor data often have detrimental consequences and great care must be taken over the two stages. This chapter deals with the intermediate stage. I will not be dealing with the last stage in this book, as it is application dependent. I shall begin this chapter at the simplest level; you may already know some of the concepts and techniques I will be treating, but my goal is to lay a sound foundation, so that the book is self-contained.

The subjects treated in this chapter are by no means exhaustive. Moreover, I regard *estimation* from a single viewpoint, which is the processing of sensed data. I refer readers wishing to further enrich their knowledge on random variables, stochastic processes, and estimation techniques to the excellent books by Papoulis and Pillai (2002), Ross et al. (1996), Gardiner (1985) and Grewal (2011).

I shall begin this chapter by making a somewhat sensitive statement: we shall never be able to construct a sensor (or any system, for that matter) that captures reality as it is. We can approach reality but never touch it. Even at a quantum level, our approaching of reality is limited by Heisenberg's uncertainty principle. The error with which we perceive reality accumulates as we move away from a quantum reality towards a macro reality. But fortunately for us, this error will never reach a magnitude at which our perception of reality makes existence impossible.

There are two fundamental premises—understood consciously or unconsciously—for relying on sensor data; whether the data come from a biological sensor or a physical sensor constructed by human beings, does not matter. These are:

1) The change in the physical reality (measurand) is a gradual process rather than haphazard and wild; statistically speaking, the measurand is correlated with itself to a certain extent.
2) The state of reality and the output of a sensor are correlated to a certain extent.

The significance of these assertions will become clear when we deal with the mathematics. But for now, consider Figure 12.1, which is a measurement taken from an ordinary temperature sensor having an accuracy (according to the manufacturer) of 1 °C. The sensor was placed outdoors and sampled every second for 30 min. From the reading (and also from intuition) it is clear that even though the measurement fluctuates

Principles and Applications of Ubiquitous Sensing, First Edition. Waltenegus Dargie.

Companion Website: www.wiley.com/go/dargie2017

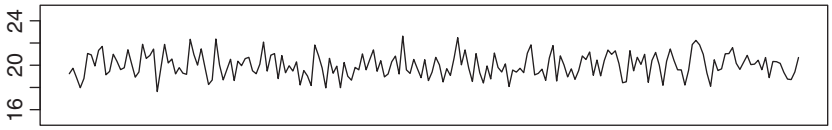


Figure 12.1 The reading of a temperature sensor having an accuracy of 1 °C.

over time, the fluctuation is not haphazard. Secondly, the temperature sensor may not be accurate enough (whatever that means) but it reflects reality. The outside temperature might not have changed at all during the 30 minutes during which we took the reading, or the change might have been quite dissimilar from the one we obtained using the sensor. But there should be certain correlation between the physical reality and the sensor output (unless the sensor is defective). It cannot, for instance, be the case that the temperature fluctuated by around 5 °C much of the time whilst the sensor reading fluctuated by around 20 °C!

12.1 Sensor Error as a Random Variable

The accuracy of a sensor can be quantified but it can also be viewed as a qualitative property. The qualitative and quantitative aspects of a sensor's accuracy can be explained by Figure 12.2. Suppose we expose two sensors having different accuracies to a measurand that does not change; in other words one that is constant over time. For our example, the measurand is a temperature of 20 °C. As can be seen, even if the input is constant, the outputs vary to some extent. Most existing physical sensors share this basic feature. The nearness of the sensor output to the true value of a measurand (as we have already seen in Chapter 11) is what we call accuracy. In Figure 12.2 (top), the output of the sensor appears to be constant, properly reflecting reality, but when we regard it at a fine-grained resolution (in the middle), we see that the samples are different. Indeed, the samples are

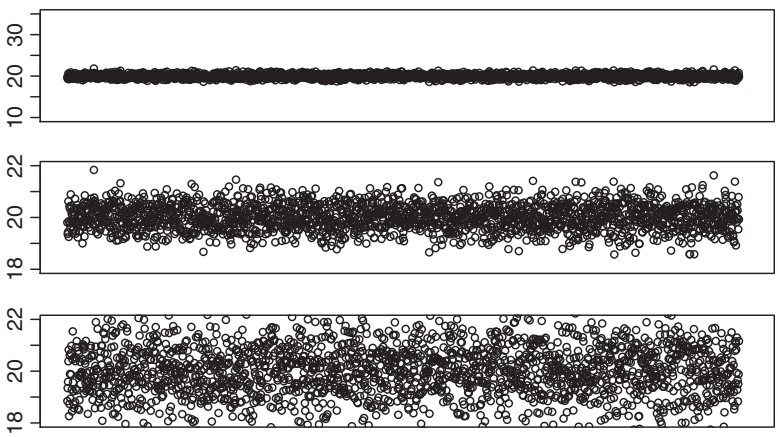


Figure 12.2 A computer generated simulation of the output of two temperature sensors for a fixed (constant) input: (top), (middle) the output of the same sensor at two different granularities; (bottom) the output of a second sensor, having a different accuracy, for the same input.

different not only from the input, but also from each other. The dissimilarity becomes conspicuous in the output of the second sensor (bottom), which has a different accuracy.

Because we cannot be certain of what specific value we will get when we next sample a sensor, we regard its output as a random variable. It is worthwhile emphasising here that the variation in the output of a sensor may have nothing to do with the measurand or the physical process; it can be an inherent property of the sensor itself. The randomness in the output of the sensor nevertheless obeys an underlying probability distribution, because some samples are more likely to occur than others. If we represent the output of a sensor (a random variable) by a small boldface letter (such as $\mathbf{x}$), the function that assigns a probability term to each of the outputs of $\mathbf{x}$ is called a probability density function (pdf), $f(x)$, where x is a real number, representing one of the distinct outcomes of $\mathbf{x}$. Assuming that an infinite number of samples can be obtained if we sample the sensor for long enough, then $f(x)$ provides sufficient information about $\mathbf{x}$.

Figure 12.3 compares the pdfs of the two sensor outputs in Figure 12.2, which happen to be normally distributed. The width of a pdf is an indication of the dissimilarity between the sample outputs of a sensor. The broader it is, the more dissimilar the sample outputs are, and therefore the less reliable the sensor is. If we have the mathematical expression of the pdf, we can ask and answer several questions, such as:

- What is the expected outcome of the random variable (the mean)?
- What is the variance of the random variable (the quantifiable expression of the dissimilarity between the samples, which is also a measure of the error of the sensor)?
- What is the probability that the output of the sensor is between two real numbers, say between 19.5 and 20.5°C, in other words, $P\{19.5 \leq \mathbf{x} \leq 20.5\}$?

The expected outcome or the mean is given by:

$$\eta_x = E[\mathbf{x}] = \int_{-\infty}^{\infty} xf(x)dx \tag{12.1}$$

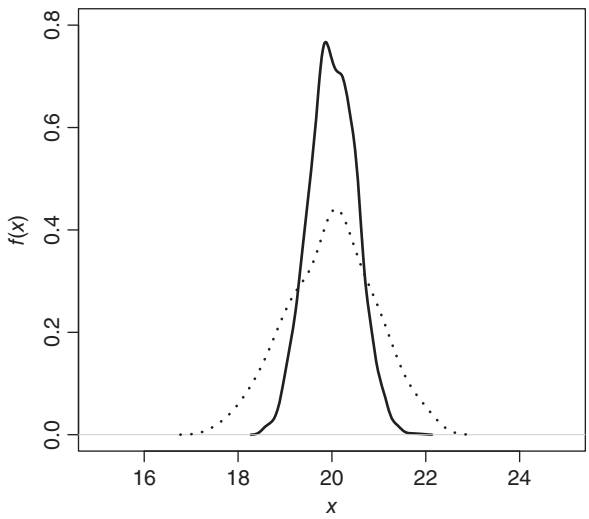


Figure 12.3 The probability density functions of two random variables.

Similarly, the variance of $\mathbf{x}$ is defined as the square of the expected variance of the sample outputs with respect to the mean:

$$\sigma_x^2 = E[(\mathbf{x} - \eta_x)^2] = \int_{-\infty}^{\infty} (\mathbf{x} - \eta_x)^2 f(x) dx \tag{12.2}$$

The reason we consider $E[(\mathbf{x} - \eta_x)^2]$ instead of $E[(\mathbf{x} - \eta_x)]$ is that the latter will always yield a value of zero, since the summation (integration) results in a net negative value for the samples below the mean and a net positive value for the samples above the mean and both values are equal in magnitude (alternatively, $E[(\mathbf{x} - \eta_x)] = E[\mathbf{x}] - \eta_x = \eta_x - \eta_x = 0$). The variance and mean are related with one another and we shall exploit their relationship to solve some important problems later. Since $\sigma_x^2 = E[(\mathbf{x} - \eta_x)^2]$, we can distribute the right-hand term as follows: $E[\mathbf{x}^2 - 2\eta_x \mathbf{x} + \eta_x^2] = E[\mathbf{x}^2] - 2\eta_x E[\mathbf{x}] + \eta_x^2$, from which we have (as a reminder, the expected value of a constant is the constant itself):

$$\sigma_x^2 = E[\mathbf{x}^2] - \eta_x^2 \tag{12.3}$$

The probability that $\{x_1 \leq \mathbf{x} \leq x_2\}$ can be computed using $f(x)$ alone:

$$P\{x_1 \leq \mathbf{x} \leq x_2\} = \int_{x_1}^{x_2} f(x) dx \tag{12.4}$$

The pdf is also useful to visualise the difference between precision (consistency) and accuracy. Suppose the output of two sensors for a known, fixed input (say, 20 °C), is described by the two pdfs shown in Figure 12.4. As can be seen, the first sensor (the solid line) has a mean that overlaps the input (the true value), while the second sensor (the dotted line) has a mean that is different from the input. Hence we can say that the first sensor is more accurate than the second. On the other hand, the pdf of the second sensor is much wider than that of the first, so we can say that it is more precise than the first, because its output is more consistent and hence more predictable.

Another important function by which a random variable can be described is the cumulative distribution function (CDF) or the probability distribution function (PDF), $F(x)$:

$$F(x) = P\{\mathbf{x} \leq x\} \tag{12.5}$$

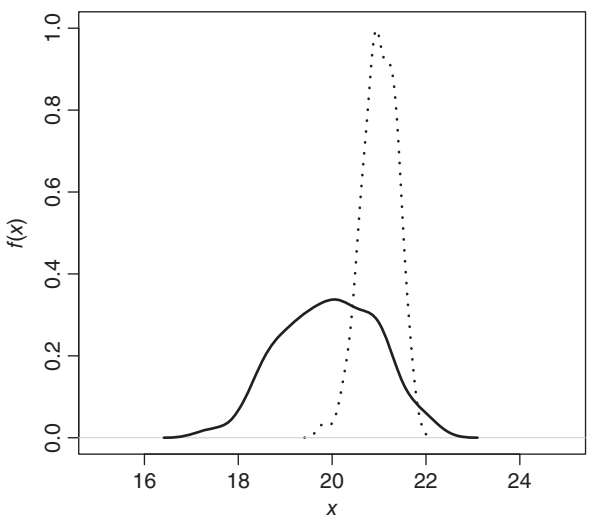


Figure 12.4 The probability density functions of the outputs of two sensors are shown to explain the difference between accuracy and precision. Assuming that the known input is 20 °C, the mean of one of the pdfs (solid line) overlaps with the known input, but its variance is big whereas the mean of the other pdf (dotted line) does not overlap with the input but its variance is small. In the first case, we can say that the sensor is more accurate but in the second case we can say that the sensor is more precise.

where **x** (boldface type) is the random variable and x (normal font) is a real number. $F(x)$ quantifies the probability that the outcome of a random variable is below a certain value x. For example:

$$F(20) = P\{\mathbf{T} \leq 20\}$$

refers to the probability that the temperature is below 20 °C. Since the probability accumulates as x increases, it is called a cumulative function. Therefore, $F(x)$ is a non-decreasing (monotonically increasing), right-continuous function. For example,

$$F(21) = P\{\mathbf{T} \leq 21\} = F(20) + P\{20 < \mathbf{T} \leq 21\}$$

Figure 12.5 compares the distribution functions of the two simulated temperature sensors of Figure 12.2. The slope of the distribution function indicates the variance of a random variable. The steeper the slope, the smaller the variance, and for our case, the more consistent a sensor is, the gentler the slope, the larger the variance, and the more dissimilar are the outcomes of the sensor. The distribution and the density functions are related with one another:

$$f(x) = \frac{dF(x)}{dx} \tag{12.6}$$

$$F(x) = \int_{-\infty}^{x} f(u)du \tag{12.7}$$

We shall use the two functions alternatively to solve different problems. Sometimes solving problems with one is simpler than solving with the other. It is also worthwhile noticing that since the total probability is always one, we have:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{12.8}$$

and

$$1 = P\{\mathbf{x} \leq x\} + P\{\mathbf{x} > x\} \tag{12.9}$$

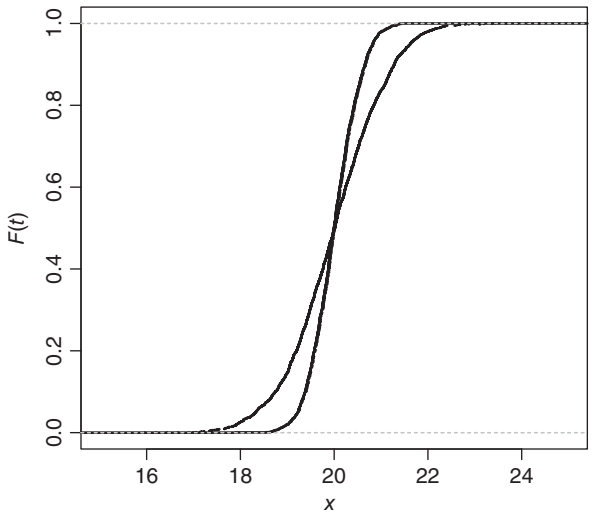


Figure 12.5 A comparison of the cumulative distribution functions of the two simulated temperature sensors, the output of which (for a fixed input) are shown in Figure 12.2.

or

$$1 = F(x) + P\{\mathbf{x} > x\} \tag{12.10}$$

from which we have:

$$P\{\mathbf{x} > x\} = 1 - F(x) \tag{12.11}$$

12.2 Zero-offset Error

One of the inherent errors of physical sensors is the zero-offset error. This error corresponds to the random output of the sensor when there is no input from the measurand. The sources of these outputs can be many, including internal thermal noise, external thermal noise, radiation, and so on. The most significant of these is the internal thermal noise coming from the random vibration of electrons. Most existing sensors have a zero-mean, normal distributed zero-offset error at a given temperature (for example, at room temperature). In other words, the density function of this error can be described as follows:

$$f(e) = \frac{1}{\sqrt{2\pi}\sigma_e} e^{-e^2/2\sigma_e^2} \tag{12.12}$$

where σ_e is the standard deviation or σ_e^2 is the variance of the offset. Figure 12.6 compares the zero-offset errors of two sensors. In general, the inherent error of a sensor due to its internal composition increases as the magnitude and frequency of the measurand increase. You may recall from Chapter 11 how self-heating contributes to the error of a sensor.

Example 12.1 The zero-offset error of a given sensor can be modelled as a uniformly distributed random variable between -2 mV and 2 mV as shown in Figure 12.7. Determine the variance of the random variable. x

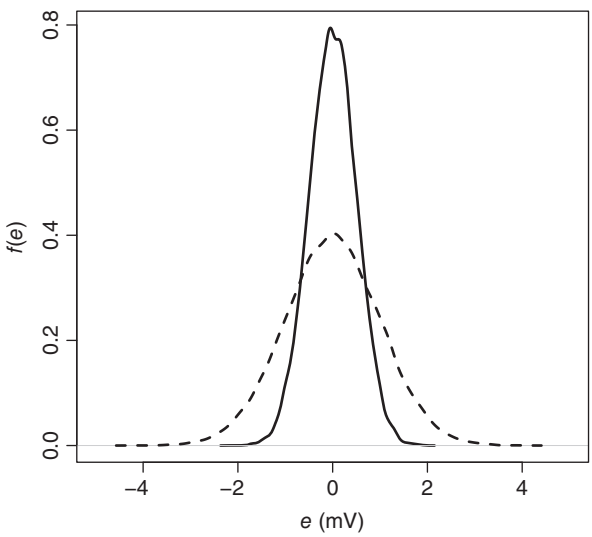


Figure 12.6 The zero-offset (the output being a voltage) of two different sensors described by normally distributed random variables. The zero-offset described by the solid line has a standard deviation of 1 mV whilst the zero-offset described by the dashed line has a standard deviation of 0.5 mV.

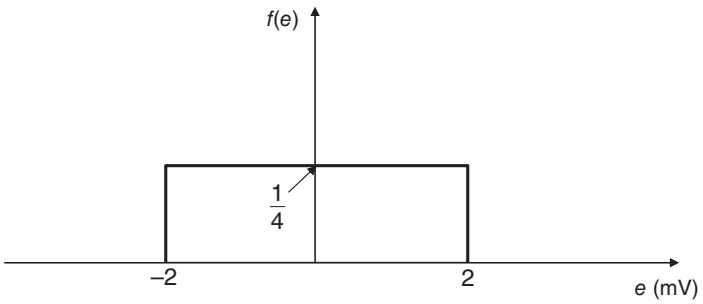


Figure 12.7 The zero-offset of a sensor modelled as a uniformly distributed random variable.

The variance of a random variable is expressed as:

$$\sigma_x^2 = E[(\mathbf{x} - \eta_x)^2] = E[\mathbf{x^2}] - \eta_x^2$$

Since the mean of our random variable is zero, we have:

$$\sigma_x^2 = E[\mathbf{x^2}] = \int_{-2}^{2} x^2 f(x)dx = \frac{1}{12}(2^3 - (-2)^3) = 1.33\,\text{mV}$$

Example 12.2 We wish to build an electronic switch as shown in Figure 12.8. It should respond when the temperature of the measurand crosses a set threshold (specified by V_{REF}). Suppose the RTD has a resistance of 1 kΩ at the reference temperature and a zero-offset voltage distribution as shown in Figure 12.7. Since the sensor contains a zero-offset output, it may trigger the switch erroneously. On the other hand, the sensor may also attenuate an authentic signal due to a negative offset voltage. Assuming the zero-offset is independent of the measurand's temperature and that our priority is reducing false positives, determine a reference voltage that suppresses the contribution of both the biasing voltage and the zero-offset 75% of the time.

At the reference temperature, the voltage across the RTD is a superposition of a portion of the biasing voltage and the zero-offset voltage:

$$V_S = V_B + V_{\text{OFF}}$$

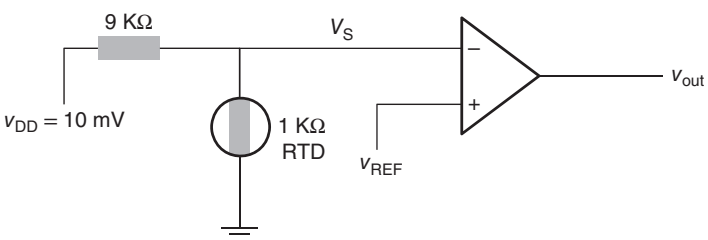


Figure 12.8 The circuit block diagram of a simple comparator. The reference voltage of the comparator is set so that 75 % of the false positive originating from the zero-offset voltage of the sensor can be suppressed.

But,

$$V_B = \left(\frac{1\,\text{k}\Omega}{9\,\text{k}\Omega + 1\,\text{k}\Omega}\right) 10\,\text{mV} = 1\,\text{mV}$$

Since V_S contains a random variable, it too is a random variable. As long as V_S is less than V_{REF}, the output of the comparator is negative. If we simply set the reference voltage to be -3 mV, we can be certain that the switch is turned on only as a result of an increment in the measurand's temperature. By doing so, however, we also increase the possibility of a false negative, because the offset voltage might have suppressed an output voltage due to an increase in temperature. In order to satisfy the specified requirement,

$$0.75 = P\{V_S \leq x\} = P\{(V_B + \mathbf{V}_{\text{OFF}}) \leq x\} = P\{\mathbf{V}_{\text{OFF}} \leq (x - V_B)\}$$

Since we have the distribution of $\mathbf{V}_{\text{OFF}}$, the value of x that yields a probability of 0.75 can be determined as follows:

$$0.75 = \frac{1}{4}\int_{-2}^{x-V_B} dx = \frac{1}{4}\int_{-2}^{x-1} dx$$

From which we have:

$$0.75(4) = (x - 1\,\text{mV}) - (-2\,\text{mV})$$

Rearranging terms will result in,

$$x = 2\,\text{mV}$$

12.3 Conversion Error

The output of a sensor passes through many intermediate stages. Therefore, it is important to understand how these stages influence the accuracy and precision of the sensor output. In this section we shall consider how the randomness of the sensor influences the probability distribution of the output. In the next section, we will consider the accumulation of error in more detail. In order to demonstrate how the statistics of the sensor error influence the output voltage, consider Figure 12.9, where we display the electrical circuit diagram of a temperature-to-voltage converter employing a TC1047A temperature sensor manufactured by Microchip Technology Inc. According to the manufacturer, the module has an accuracy of ±2 °C at 25 °C and can measure a change in temperature between -40 and 125 °C. For the specified temperature range the output voltage varies from 2.7 V to 4.4 V. Assuming that the sensor has a zero-offset voltage that can be characterised by a zero-mean, normally distributed random variable with a variance of 0.5 mV, we can determine the statistics of the output voltage.

Because the reference voltage is a fixed quantity, the output of the operational amplifier is a square wave, positive when $V_{\text{REF}} > V_{\text{S}}$ and negative when $V_{\text{REF}} < V_{\text{S}}$. Moreover, we have,

$$i_s = -i_{\text{f}} = -\frac{\mathbf{V}_{\text{S}}}{jX_C} = \frac{(\mathbf{V}_{\text{S}} - \mathbf{V}_{\text{O}})}{RTD}$$

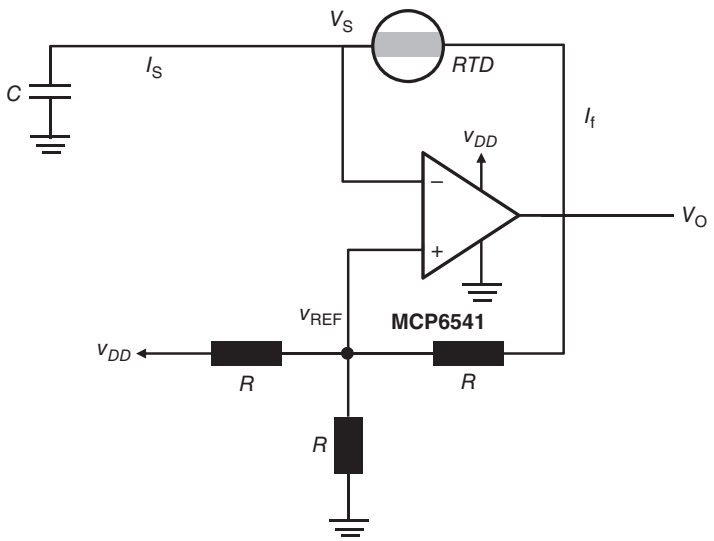


Figure 12.9 A schematic diagram of a temperature-to-voltage converter employing a resistance temperature detector (RTD).

(Recall that because of the high input impedance of the amplifier, we assume that no current flows into the operational amplifier). Hence,

$$\mathbf{V}_O = \mathbf{V}_S \left[1 + \frac{RTD}{jX_C}\right] \tag{12.13}$$

Because the output of the operational amplifier is related to a random variable ($\mathbf{V}_S$), it, too, is a random variable. If we for now disregard the randomness added to the output voltage from the internal noise of the operational amplifier, then the PDF of the output voltage can be expressed in terms of the distribution of the sensor's voltage:

$$F_O(v) = P\{\mathbf{V}_O \le v\} = P\left\{\left[1 + \frac{RTD}{jX_C}\right]\mathbf{V}_S \le v\right\} \tag{12.14}$$

where we use $F_O(v)$ to indicate that the distribution function refers to the output voltage. If we rearrange the terms in Eq. 12.14, we have:

$$F_O(v) = P\left\{\mathbf{v_s} \le \left[1 + \frac{jX_C}{RTD}\right]v\right\} = F_S\left(\left[1 + \frac{jX_C}{RTD}\right]v\right) \tag{12.15}$$

where $F_S(.)$ refers to the PDF of $\mathbf{V}_S$. As can be seen, we managed to express the distribution of $\mathbf{V}_O$ in terms of the distribution of $\mathbf{V}_S$.

The labels $\mathbf{V}_O$, $\mathbf{V}_S$ and v should by now be clear; the boldface letters represent random variables whilst the normal font v represents a real number or an instance of the random variables. With the change in the distribution function of the output voltage, its mean and variance change as well. As far as the error of the sensor is concerned, the change in the mean is not much of an issue (it is zero), but the variance is. So how does the variance of the output voltage change?

$$\sigma^2_{V_o} = E[(\mathbf{V}_o - \eta_{V_O})^2] = E[(\mathbf{V}_O)^2] - \eta^2_{V_o} \tag{12.16}$$

Substituting Eq. (12.13) into Eq. (12.16) yields,

$$\sigma^2_{V_O} = \left[1 + \frac{RTD}{jX_C}\right]^2 E(\mathbf{V}_S^2) = \left[1 + \frac{RTD}{jX_C}\right]^2 \sigma^2_{V_S} \tag{12.17}$$

Notice that $\eta^2_{V_O} = 0$, because $\mathbf{V}_S$ is a zero-mean random variable. From Eq. (12.15), it is possible to determine the probability density of the output voltage, because:

$$f_O(v) = \frac{dF_O(v)}{dv} = \left[1 + \frac{jX_C}{RTD}\right] f_S\left(\left[1 + \frac{jX_C}{RTD}\right] v\right) \tag{12.18}$$

where we applied the chain-rule on Eq. (12.15).

Example 12.3 We wish to determine the pdf of the magnitude of the output voltage of the temperature-to-voltage converter in the absence of a measurand. Suppose we are interested in the response of the converter at 100 Hz, with the circuit elements having the values shown in Figure 12.10.

The probability density function of the output voltage is a complex function because of the capacitive reactance. Its magnitude component is given as:

$$|F_O(v)| = \sqrt{\left[1 + \left(\frac{X_C}{RTD}\right)^2\right]}\,\mathbf{V}_S$$

At the specified frequency, the capacitive reactance is:

$$XC = -j\left(\frac{1}{2\pi \times 100 \times 10^{-6}}\right) = -j1.6\,\text{k}\Omega$$

Thus,

$$|F_O(v)| = \sqrt{\left[1 + \left(\frac{1.6\text{k}\Omega}{1\text{k}\Omega}\right)^2\right]}\,\mathbf{V}_S = 1.9\,\mathbf{V}_S$$

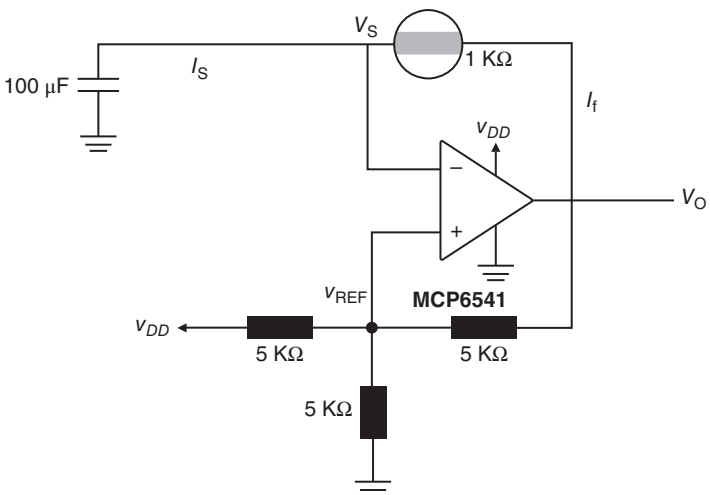


Figure 12.10 The input-output relationship of a temperature-to-voltage converter at 100 Hz.

Moreover, we have:

$$\min(\mathbf{V}_O) = 1.9 \times \min(\mathbf{V}_S) = 1.9 \times -2 = -3.8 \text{ mV}$$

and

$$\max(\mathbf{V}_O) = 1.9 \times \max(\mathbf{V}_O) = 1.9 \times 2 = 3.8 \text{ mV}$$

Figure 12.11 shows the relationship between the two random variables, namely, $\mathbf{V}_S$ and $\mathbf{V}_O$.

In general, if the random variables **y** and **x** are related to one another, then it is possible to determine the statistics of one of the random variables (the unknown) in terms of the statistics of the other (the known), beginning by describing the PDF of the unknown random variable in terms of the known random variable. For example, if the two random variables are related as follows:

$$\mathbf{y} = a\mathbf{x} + b$$

where a and b are known positive constants, the statistics of **y** can be expressed in terms of the statistics of **x** and:

$$F(y) = P\{\mathbf{y} \leq y\} = P\{(a\mathbf{y} + b) \leq y\}$$

$$F(y) = P\left\{\mathbf{x} \leq \frac{(y-b)}{a}\right\}$$

Since

$$\frac{y-b}{a} = C$$

is a constant, then we have:

$$F(y) = P\{\mathbf{x} \leq C\} = F_X(C)$$

where $F_X(C) = P\{\mathbf{x} \leq C\}$ is the distribution of **x**. From here on, it is a matter of making the appropriate derivation to determine the statistics pertaining to **y**.

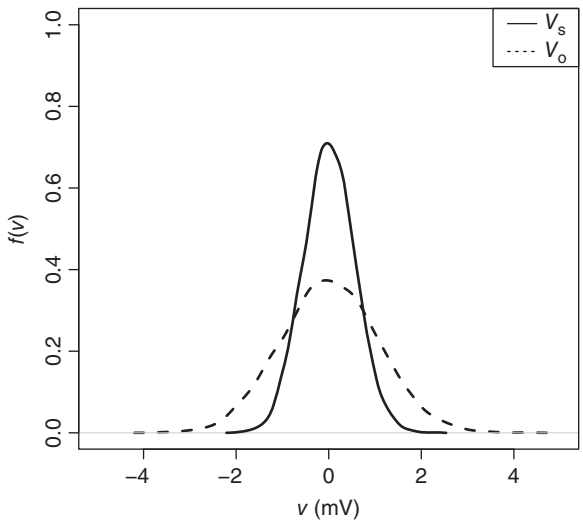


Figure 12.11 The relationship between the PDFs of $\mathbf{V}_S$ and $\mathbf{V}_O$ at 100 Hz frequency.

12.4 Accumulation of Error

A sensing system typically consists of two or more stages. The conditioning circuit, for example, may consist of a Wheatstone bridge, an amplifier, and a filter. Each of these stages will introduce its own error into the signal produced by the sensing stage. These errors should be accounted for in order to meaningfully interpret the sensed signal. Suppose the aggregate error of the conditioning circuit can be modelled by the random variable **c**, having its own distribution and density functions. The combined error of the sensing and the conditioning stages is given as:

$$\mathbf{e} = \mathbf{s} + \mathbf{c} \tag{12.19}$$

In order to explain the addition of random variables, we shall begin with a simple example. Suppose, in the absence of an input (due to an internal thermal noise), the sensing system produces either -0.5 mV or 0.5 mV with equal probabilities. Likewise, the conditioning circuit, independent of the sensing system, produces either -0.5 mV or +0.5 mV with equal probabilities. If we connect the two systems, as shown in Figure 12.12, and measure the output voltage, we may get the following values with the corresponding probabilities:

$$e = -0.5\text{ mV} + -0.5\text{ mV} = -1.0\text{ mV} \quad (P(-0.5) \times P(-0.5))$$
$$e = -0.5\text{ mV} + 0.5\text{ mV} = 0.0\text{ mV} \quad (P(-0.5) \times P(0.5))$$
$$e = 0.5\text{ mV} + -0.5\text{ mV} = 0.0\text{ mV} \quad (P(0.5) \times P(-0.5))$$
$$e = 0.5\text{ mV} + 0.5\text{ mV} = 1.0\text{ mV} \quad (P(0.5) \times P(0.5))$$

So, we see that **e** as a random variable has different outcomes with different probabilities (notice how the new probability distribution should be computed):

$$\mathbf{e} = [-1.0(P = 0.25), 0.0(P = 0.5), 1.0(P = 0.25)]$$

As a result, some values—0.0 mV for example—are more likely to occur—P{0.0 mV} = 0.5—than other values. Figure 12.13 shows the probability mass function of **e**.

Now suppose, **s** and **c** each have five discrete outputs (in mV) as follows:

$$\mathbf{s} = \mathbf{c} = [-0.5, -0.25, 0.0, 0.25, 0.5]$$

Then, $\mathbf{e} = \mathbf{s} + \mathbf{c}$ will have 5×5 elements, but some of the elements have equal values. Carrying out the computation as shown above yields nine unique values with corresponding probabilities:

$$\mathbf{e} = [-1.0, -0.75, -0.5, -0.25, 0.0, 0.25, 0.5, 0.75, 1.0]$$

The corresponding probability of occurrence is:

$$f(e) = [0.04, 0.08, 0.12, 0.16, 0.2, 0.16, 0.12, 0.08, 0.04]$$

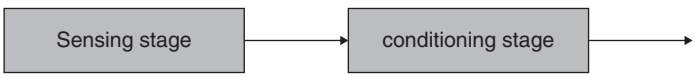


Figure 12.12 The accumulation of error when the sensing and the conditioning systems are connected in series.

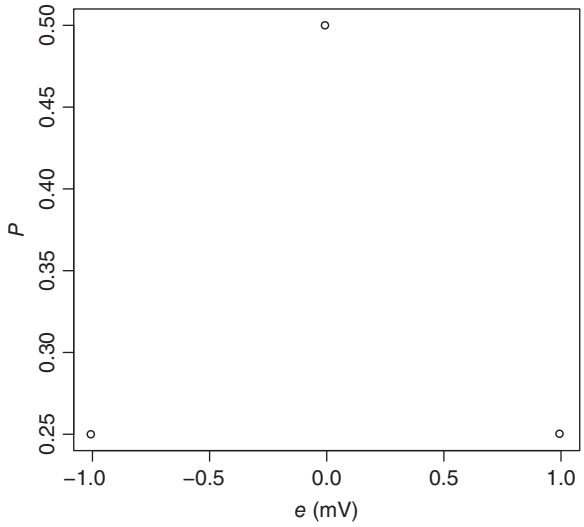


Figure 12.13 The probability mass function of $\mathbf{e} = \mathbf{s} + \mathbf{c}$ when each random variable has only two discrete values with equal probabilities.

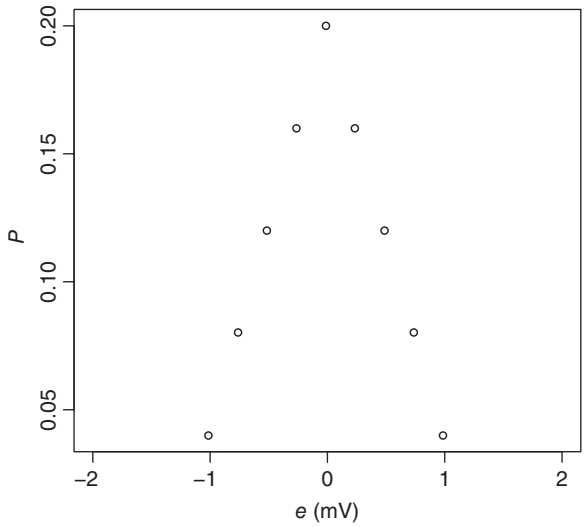


Figure 12.14 The probability mass function(the discrete equivalent of the pdf) of $\mathbf{e} = \mathbf{s} + \mathbf{c}$ when each random variable has only five discrete values with equal probabilities.

Figure 12.14 shows the probability mass function of **e**. In general, if **s** and **c** have I and J discrete values, the elements of **e** can be computed using two `for` loops, as shown in Figure 12.15. Moreover, when some of the elements have equal values, their probabilities should be added. From the above example, it is clear that the computation of the distribution and the density functions for **e**, when **s** and **c** have continuous distributions and density functions, can be carried out as follows:

$$F(e) = P\{\mathbf{s} + \mathbf{c} \leq e\} = \int_{s=-\infty}^{\infty} \int_{c=-\infty}^{e-c} f(s,c) ds\, dc \tag{12.20}$$

where $f(s,c)$ is the joint density function. If **s** and **c** are independent, then $f(s,c) = f(s)f(c)$. You may notice that the two integrations above correspond to the two `for` loops in Figure 12.15 for the case where the two random variables are

```
for(int i; i < sizeof(s);i++){
       for(int j; j < sizeof(c); j++) {
              e[ i * sizeof(c) + j] = s[i] + c[j]
       }
}
```

Figure 12.15 Computing the elements of $\mathbf{e} = \mathbf{s} + \mathbf{c}$ in C. As can be seen, **e** can have sizeof(**s**) × `sizeof`(**c**) distinct elements.

discrete. Similarly, the density function of **e** is given as:

$$f(e) = \int_{-\infty}^{\infty} f(e - c, c)dc \tag{12.21}$$

If the two random variables are independent, then we have:

$$f(e) = \int_{-\infty}^{\infty} f(e - c)f(c)dc \tag{12.22}$$

where $f(e - c)$ is the density of **s** expressed in terms of $s = e - c$. The density function amounts to adding the probabilities of similar values after the elements of **e** are computed in Figure 12.15.

Example 12.4 Suppose both the sensing element and the conditioning circuit of a given sensor generate random outputs **s** and **c**, respectively. In the absence of a measurand, these outputs are described by uniformly distributed voltages, each ranging between 0 mV and 1 mV. Assuming that the two sources of error are independent and their cumulative effect is additive, determine the distribution of the error as $\mathbf{e} = \mathbf{s} + \mathbf{c}$.

As the sum of two random variables, **e** may have any outcomes between 0 (the minimum value) and 2 (the maximum value). Since both random variables are continuous, the PDF of **e** can be determined by integrating the joint density function. Since we are dealing with definite integrals, we should first determine the boundaries of integration. We shall rely on Figure 12.16 to determine the boundaries. The x- and y-intercepts of the equation $\mathbf{e} = \mathbf{s} + \mathbf{c}$ can be determined by setting one of the random variables to zero. Hence, for the x-intercept we have $\mathbf{c} = 0$ and:

$$\mathbf{s} = \mathbf{e}$$

Similarly, for the y-intercept we have $\mathbf{s} = 0$ and,

$$\mathbf{c} = \mathbf{e}$$

But notice that since **s** and **c** are random variables, the x- and y-intercepts are not constant values as they would be for a deterministic function. Instead, the line $\mathbf{e} = \mathbf{s} + \mathbf{c}$ can be located in different places in the first quadrant bounded by the coordinates $(0, 0)$ and $(1, 1)$. For any positive real value $e \leq 1$, the distribution function,

$$F(e) = P\{\mathbf{e} \leq e\} = P\{\mathbf{s} + \mathbf{c} \leq e\}$$

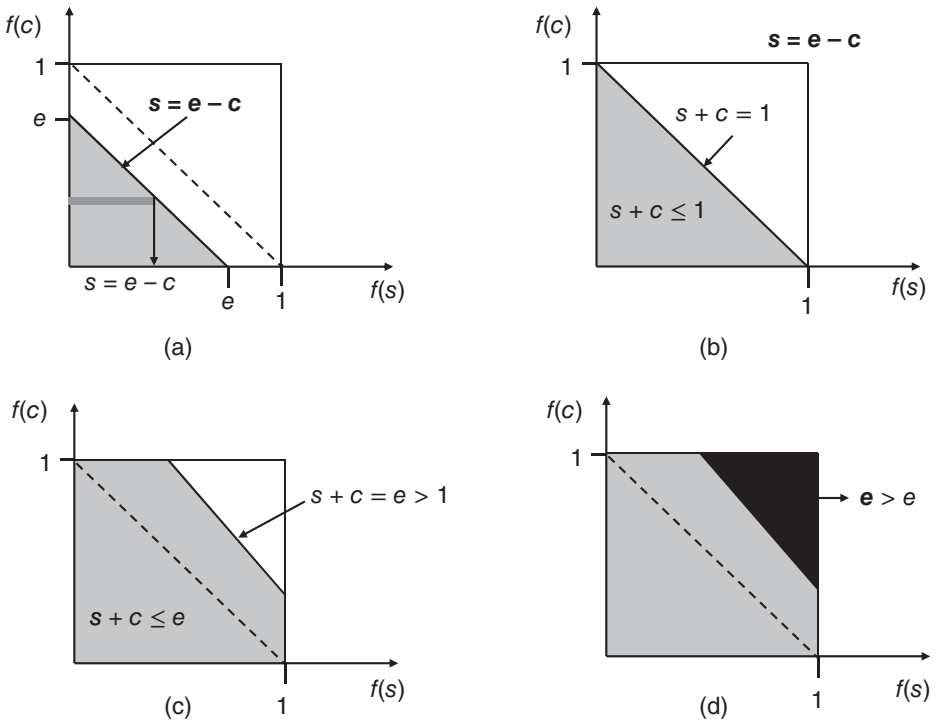


Figure 12.16 Determining the probability distribution function of $\mathbf{e} = \mathbf{s} + \mathbf{c}$: $F(e) = P\{\mathbf{s} + \mathbf{c} \leq e\}$. We are interested in the region where $\mathbf{s} + \mathbf{c} \leq e$. In (a), the shaded region: $\mathbf{s} + \mathbf{c} \leq 1$, and can easily be integrated. Here it is clear that as **c** varies from zero to e, **s** varies from zero to the intersection (e - c). In (b), the shaded region subsumes the boundary $\mathbf{s} + \mathbf{c} = 1$, beyond this boundary, $\mathbf{s} + \mathbf{c} \leq e$ is not simple to integrate on account of the complexity of the geometry of the shaded region, as can be seen in (c). In (d) we can take advantage of the mutual exclusiveness of $\{\mathbf{e} \leq e\}$ and $\{\mathbf{e} > e\}$ and the fact that the probability of the two regions adds to unity.

can be determined by integrating the shaded region of Figure 12.16a, which describes the regions where the two density functions overlap. From the figure, it is apparent that as **c** varies from 0 to e, **s** varies from 0 to $e - c$. Alternatively, we can vary **s** from 0 to e and bind **c** to vary from 0 to $e - s$. Thus, $F(e)$ can be computed as (remember that both random variables are uniform):

$$\begin{aligned} F(e) &= \int_{c=0}^{e} \int_{s=0}^{e-c} f(c) f(e-c)\, dsdc \\ &= \int_{c=0}^{e} \int_{s=0}^{e-c} dsdc \\ &= \int_{c=0}^{e} (e-c)\, dc \\ &= \frac{e^2}{2} \end{aligned}$$

But when $\mathbf{e} > 1$, the graph $\mathbf{e} = \mathbf{s} + \mathbf{c}$, which is shown in Figure 12.16c becomes a little complicated to integrate. However, recall from Eq. (12.11) that:

$$F(e) = P\{\mathbf{e} \leq e\} = 1 - P\{\mathbf{e} > e\}$$

Since the black region in Figure 12.16d is $(\mathbf{s} + \mathbf{c}) > e$, it is relatively easy to integrate. Consequently, for $1 < \mathbf{e} \leq 2$, we have:

$$\begin{aligned} F(e) &= 1 - \int_{c=e-1}^{1} \int_{s=e-c}^{1} f(c)f(e-c)dsdc \\ &= 1 - \int_{c=e-1}^{1} \int_{s=e-c}^{1} dsdc \\ &= 1 - \int_{c=e-1}^{1} (1-e+c)dc \\ &= 1 - \frac{(2-e)^2}{2} \end{aligned}$$

The pdf of the error can be determined by differentiating $F(e)$ with respect to e. Thus:

$$f(e) = \begin{cases} e & 0 \leq e \leq 1 \\ 2-e & 1 < e \leq 2 \end{cases}$$

Figure 12.17 displays the probability density function of $\mathbf{e} = \mathbf{s} + \mathbf{c}$.

12.4.1 The Central Limit Theorem

One interesting aspect of an accumulation of error is that, as the number of random variables that should be added increases, their pdf tends to be normally distributed, regardless of the shape of the density functions of the individual stages. If, for example, we divide the conditioning circuit into a Wheatstone bridge (**w**) and an amplification stage (**a**) and assume that **s**, **w**, and **a**, are zero-mean, uniformly distributed random

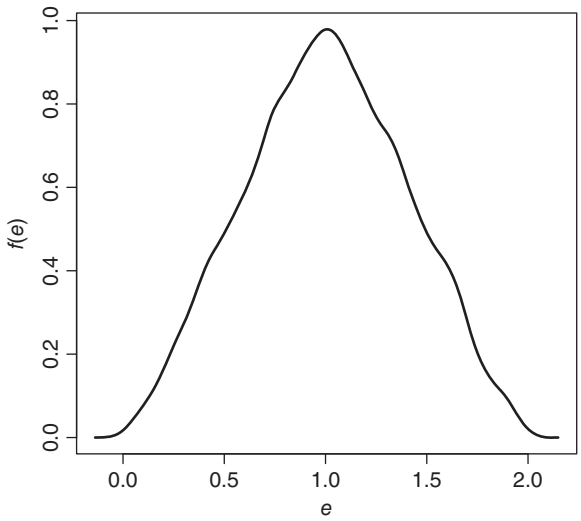


Figure 12.17 The pdf of $\mathbf{e} = \mathbf{s} + \mathbf{c}$ when both random variables are uniformly distributed between 0 and 1. As expected, the plot is similar to the one depicted in Figure 12.14, where we computed the probability mass function for two uniformly distributed discrete random variables.

variables with the following discrete outputs (in mV):

$$\mathbf{s} = \mathbf{w} = \mathbf{a} = [-0.5, -0.25, 0.0, 0.25, 0.5]$$

Then, $\mathbf{e} = \mathbf{s} + \mathbf{w} + \mathbf{a}$ will have $5 \times 5 \times 5 = 125$ entries, but some of the values occur multiple times, as a result of which their frequency of occurrence has to be summed in order to calculate their probability of occurrence. Table 12.1 summarises the statistical parameters of **e**. The approximated probability mass function of **e** is given in Figure 12.18 and reflects a well-studied and statistically well-formulated phenomenon known as the central limit theorem (CLT). The CLT states that the pdf of the sum (or average) of a

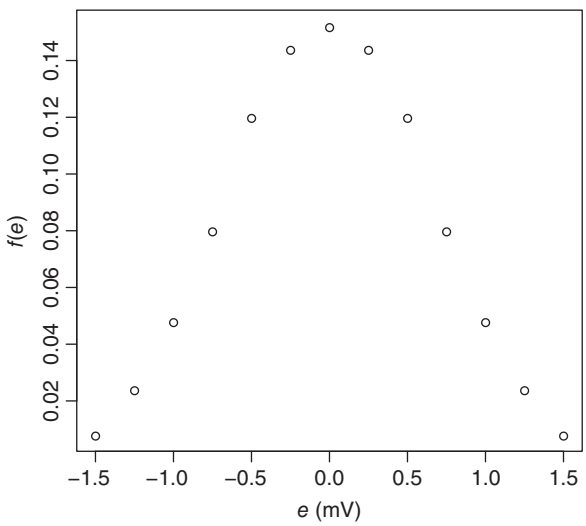


Figure 12.18 The approximate density function of $\mathbf{e} = \mathbf{s} + \mathbf{w} + \mathbf{a}$ is a zero-mean normal distribution function.

Table 12.1 A summary of the statistical parameters of $\mathbf{e} = \mathbf{s} + \mathbf{w} + \mathbf{a}$

Unique entries of e during summing	Frequency of occurrence	Probability of occurrence
−1.5	1	0.08
−1.25	3	0.02
−1.00	6	0.05
−0.75	10	0.08
−0.50	15	0.12
−0.25	18	0.14
0.00	19	0.15
0.25	18	0.14
0.50	15	0.12
0.75	10	0.08
1.00	6	0.05
1.25	3	0.02
1.50	1	0.08

large number of independent random variables with well defined means and variances will be approximately normal, regardless of the underlying distribution of the individual random variables. From Eq. (12.12), a normally distributed density function can be specified by its mean (η) and variance (σ^2). Hence, if we have n independent stages, each of which contributes its own error $\mathbf{e}_i$, having its own mean and variance, then it is possible to sufficiently describe the overall error as $\texttt{norm}(\eta, \sqrt{\sigma^2})$, where:

$$\eta = E\{\mathbf{e}\} = E\{\mathbf{e}_1 + \mathbf{e}_2 + \cdots + \mathbf{e}_n\} = \eta_1 + \eta_2 + \cdots + \eta_n \tag{12.23}$$

To compute the variance (σ^2) of the overall error, $\mathbf{e}$, we shall make use of the relation $\sigma^2 = E[\mathbf{e}^2] - \eta^2$. Moreover,

$$E[\mathbf{e}^2] = E\left[\left(\sum_{i=1}^{n} \mathbf{e}_i\right)^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} E[\mathbf{e}_i\mathbf{e}_j] \tag{12.24}$$

Since the individual random variables are independent:

$$E[\mathbf{e}_i\mathbf{e}_j] = \begin{cases} \sigma_i^2 + \eta_i & i = j \\ \eta_i\eta_j & i \neq j \end{cases} \tag{12.25}$$

The double sum in Eq. (12.24) contains n terms for $i = j$ and $n^2 - n$ terms for $i \neq j$. Notice that we have made use of $E[\mathbf{e}_i^2] = \sigma_i^2 + \eta_i^2$. Meanwhile, what happens if all the errors are zero-mean random variables? In that case, the resulting error will also be a zero-mean normally distributed random variable, the variance of which is the sum total of the variance of the individual random variables, because:

$$\eta = \eta_1 + \eta_2 + \cdots + \eta_n = 0$$

$$\eta_i\eta_j = 0$$

$$E[\mathbf{e}_i\mathbf{e}_j] = \begin{cases} \sigma_i^2 & i = j \\ 0 & i \neq j \end{cases}$$

Hence

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

12.5 Combining Evidence

So far, we have considered that an error

- is an inherent characteristic of a sensing system
- accumulates as the sensed signal advances towards the processing subsystem (due to the error introduced by the conditioning and additional intermediate stages).

Moreover, we have considered that as the number of independent error sources increases, the overall error assumes a normal pdf. One way of reducing the uncertainty stemming from the inherent sources of error is to employ multiple sensors and combine their evidence. Indeed, arrays of sensors are employed in many practical applications to this end. Hence, the next practical question is determining the appropriate techniques for combining the output of multiple sensors. There can be different combining

techniques and we shall examine some of them closely, but one essential aspect to bear in mind when dealing with the combination of evidence is the definition of uncertainty in quantifiable terms. The degree of unreliability of a sensor is directly related to the characteristics of its error. Therefore, some aspects of the error should necessarily be taken into account in the combination equation.

Figure 12.19 shows the outputs (simulated) of three temperature sensors. This figure is similar to Figure 12.2, but there is a slight difference between them. Two of the sensors have the same mean but different variances, whereas two of them have the same variance but different means. A good combination technique is one that takes these aspects into account to minimise the overall error.

12.5.1 Weighted Sum

When all the errors are zero-mean random variables, the simplest way to combine the outputs is as the weighted sum of the individual outputs. The weight given to each sensor output must be inversely proportional to the variance of its error. This is simply because, as we have already seen, the bigger the variance, the bigger our uncertainty. However, uncertainty is a relative term, because it has to be assessed relative to the uncertainty introduced by the other sensors. Suppose we have only two sensors with zero-mean errors and wish to combine their output as follows:

$$\hat{\mathbf{s}} = \alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2 \tag{12.26}$$

where $\mathbf{s}_1 = T + \mathbf{e}_1$ and $\mathbf{s}_2 = T + \mathbf{e}_2$ (and T, we assume, is the true value we are seeking to determine). One way to determine the weight for each sensor output is as follows:

$$\hat{\mathbf{s}} = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right) \mathbf{s}_1 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right) \mathbf{s}_2 \tag{12.27}$$

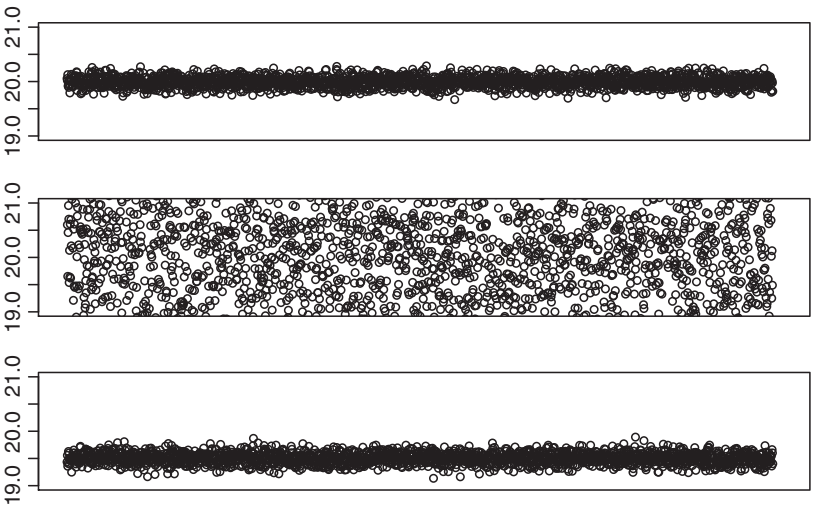


Figure 12.19 The output of three temperature sensors for the same input (20 °C). Two of the sensor outputs have the same mean but different variances (one of them $\sigma_1^2 = 0.5$ and the other $\sigma_2^2 = 1$). Two of the sensor outputs have different means but the same variance. A combination technique should take both aspects into consideration.

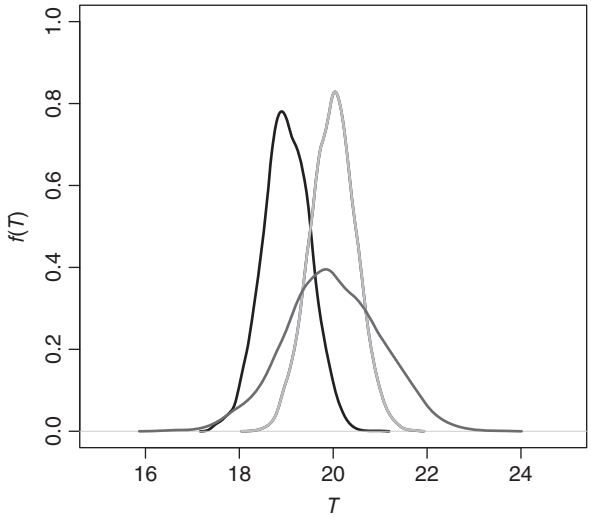


Figure 12.20 The density functions of the three simulated temperature sensors shown in Figure 12.19.

Equation (12.27) fulfils the inverse proportionality requirement as well as the relative significance of the contribution of each sensor (the term $1/(\sigma_1^2 + \sigma_2^2)$, which is a normalisation factor, so that $\alpha_1 + \alpha_2 = 1$, implicitly gives a relative significance to each coefficient). If $\sigma_1^2 > \sigma_2^2$, then we should trust $\mathbf{s}_2$ (which is now multiplied by the larger σ_1^2). If, on the other hand, $\sigma_1^2 < \sigma_2^2$ (which is multiplied by the larger σ_2^2), then we should trust $\mathbf{s}_1$; otherwise, we should trust both equally. We can rewrite Eq. (12.27) as follows:

$$\hat{\mathbf{s}} = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\mathbf{s}_1 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right)\mathbf{s}_2 + (\mathbf{s}_1 - \mathbf{s}_1) \tag{12.28}$$

Taking the $1/(\sigma_1^2 + \sigma_2^2)$ term as the common factor will yield:

$$\hat{\mathbf{s}} = \left(\frac{1}{\sigma_1^2 + \sigma_2^2}\right)[\sigma_2^2\mathbf{s}_1 + \sigma_1^2\mathbf{s}_2 + (\sigma_1^2 + \sigma_2^2)\mathbf{s}_1 - \sigma_1^2\mathbf{s}_1 - \sigma_2^2\mathbf{s}_1] \tag{12.29}$$

Collecting like terms and simplifying and rearranging yields:

$$\hat{\mathbf{s}} = \mathbf{s}_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}[\mathbf{s}_2 - \mathbf{s}_1] \tag{12.30}$$

If we let

$$K = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{12.31}$$

Eq. (12.30) can be expressed in a more compact form:

$$\hat{\mathbf{s}} = \mathbf{s}_1 + K[\mathbf{s}_2 - \mathbf{s}_1] \tag{12.32}$$

The significance of Eq. (12.32) will be clearer later, when we deal with the Kalman filter. For now, it suffices to state the following:

1) Given that we already have evidence from sensor 1, the evidence coming from sensor 2 can be added into the output by properly weighing the difference between sensor 2 and sensor 1.

2) If we trust sensor 1, then the new information (from sensor 2) should not change our belief considerably (K should be small).
3) If, however, we do not trust sensor 1, then, the new evidence should change our belief considerably (K should be large).

If, instead of two, we have n independent sensors from which we can gather evidence, the weighted sum approximating the true value can be described as:

$$\hat{\mathbf{s}} = \alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2 + \cdots + \alpha_n \mathbf{s}_n \tag{12.33}$$

Notice that $\hat{\mathbf{s}}$ is a random variable, because it is the sum of multiple random variables, which means it has its own pdf, mean, and variance. Since our aim is reducing the uncertainty introduced by the error of each sensing element, we must determine the coefficients in such a way that the variance of $\hat{\mathbf{s}}$ is a minimum. We can achieve this goal by describing the variance in terms of the coefficients:

$$\sigma_{\hat{s}}^2 = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \cdots + \alpha_n^2 \sigma_n^2 \tag{12.34}$$

We squared the coefficients merely for reason of convenience. We should select the coefficients such that their sum yields unity:

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1 \tag{12.35}$$

Now we can rewrite Eq. (12.34) as follows:

$$\sigma_{\hat{s}}^2 = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \cdots + \alpha_n^2 \sigma_n^2 - \lambda(\alpha_1 + \alpha_2 + \cdots + \alpha_n - 1) \tag{12.36}$$

Since $\lambda(\alpha_1 + \alpha_2 + \cdots + \alpha_n - 1)$ is zero, Eq. 12.36 is essentially the same as Eq. 12.34. We call λ a Lagrange multiplier and its significance will be clear shortly. The coefficients minimising the variance in Eq. (12.36) can be determined by partial differentiation:

$$\frac{\partial \sigma_{\hat{s}}^2}{\partial \alpha_i} = 0! \tag{12.37}$$

Consequently, the minimum mean square estimation approach aims to set the expected error to a minimum value:

$$\frac{\partial \sigma_{\hat{s}}^2}{\partial \alpha_i} = 2\alpha_i \sigma_i^2 - \lambda = 0 \tag{12.38}$$

and,

$$\alpha_i = \frac{\lambda}{2\sigma_i^2} \tag{12.39}$$

Equation (12.39) fulfils one of our requirements, namely that our level of confidence in a sensor should be inversely proportional to its variance. However, we have also stressed that our uncertainty in a sensor is a relative characteristic, and should be regarded with respect to the uncertainty the others sensors introduce into the equation, which is why we introduced the λ term in Eq. (12.39). Taking the fact that all the coefficients should add to unity,

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = \frac{\lambda}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \cdots + \frac{1}{\sigma_n^2}\right) \tag{12.40}$$

from which we have:

$$\frac{\lambda}{2} = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2 + \cdots + 1/\sigma_n^2} = \sigma_{\hat{s}}^2 \tag{12.41}$$

which now completes our two requirements. We can likewise determine the expected value of $\hat{\mathbf{s}}$:

$$\eta_{\hat{s}} = E[\hat{\mathbf{s}}] = \alpha_1 E[\mathbf{s}_1] + \alpha_2 E[\mathbf{s}_2] + \cdots + \alpha_n [\mathbf{s}_n] = \alpha_1 \eta_1 + \alpha_2 \eta_2 + \cdots + \alpha_n \eta_n \tag{12.42}$$

Substituting Eq. (12.39) for each α_i, we have:

$$\eta_{\hat{s}} = \frac{\eta_1/\sigma_1^2 + \eta_2/\sigma_2^2 + \cdots + \eta_n/\sigma_n^2}{1/\sigma_1^2 + 1/\sigma_2^2 + \cdots + 1/\sigma_n^2} \tag{12.43}$$

Similarly:

$$\hat{\mathbf{s}} = \frac{\mathbf{s}_1/\sigma_1^2 + \mathbf{s}_2/\sigma_2^2 + \cdots + \mathbf{s}_n/\sigma_n^2}{1/\sigma_1^2 + 1/\sigma_2^2 + \cdots + 1/\sigma_n^2} \tag{12.44}$$

Notice that unlike in Eq. (12.32), we have made no assumption about the mean of the individual random variables ($\mathbf{s}_i$) in Eq. (12.44). They can have any value. But what will happen if all of them have the same mean (η) and variance (σ^2)? In this case,

$$\hat{\mathbf{s}} = \frac{1}{n}(\mathbf{s}_1 + \mathbf{s}_2 + \cdots + \mathbf{s}_n) \tag{12.45}$$

and,

$$\sigma_{\hat{s}}^2 = \frac{\sigma^2}{n} \tag{12.46}$$

and,

$$\eta_{\hat{s}} = \eta \tag{12.47}$$

From Eq. (12.46), we can conclude that the more sensors we involve, however imperfect they are when considered individually, the less uncertain we become. Indeed, the variance tends to zero as n tends to infinity.

Example 12.5 An ultrasound scanning system uses an array of 19 microphones arranged as shown in Figure 12.21. Each microphone has a normally distributed, zero-mean, zero-offset output voltage, with a variance of 1 mV. Assuming that the microphones' output voltages are independent of one another and we employ the weighted-sum technique to combine their outputs, determine the pdf of the zero-offset output voltage of the scanning system and compare it with the pdfs of the zero-offset voltage of the individual microphones.

From the CLT we know that the output zero-offset voltage of the entire system will have a normal distribution. Since the output voltages of the sensors are regarded as independent and identically distributed (iid), this suffices for us to determine the variance of $\hat{s}$:

$$\sigma_{\hat{s}}^2 = \frac{\sigma^2}{n} = \frac{1}{19} \approx 0.05 \text{ mV}$$

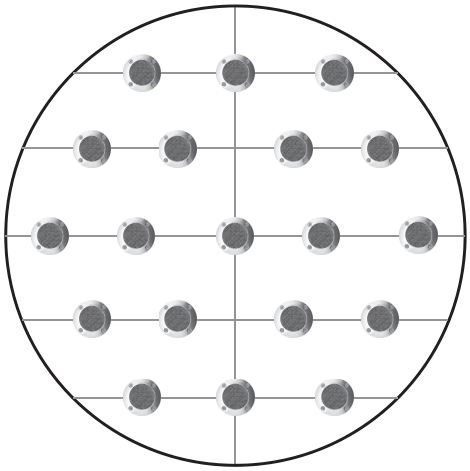


Figure 12.21 An ultrasound scanning system consisting of 19 microphones, which serve as ultrasound receivers. The zero-offset voltages of the microphones are modelled as independent, zero-mean, and normally distributed random variables.

Hence, after inserting the parameters we have into Eq. (12.12), the density function of $\hat{s}$ is given as:

$$f(\hat{s}) = (1.78)e^{-10\hat{s}^2}$$

Figure 12.22 compares the density function of the combined error with the density function of the error of the individual microphones.

Example 12.6 Suppose the manufacturer of the ultrasound scanning system wishes to make its product affordable by mixing two types of microphones. Maintaining that the central seven microphones are more important than the outer 12 microphones, it develops them with an expensive technique, and therefore, each microphone has a normally distributed, zero-mean zero-offset output voltage having a variance of 1 mV. The outer microphones, on the other hand, are cheaper and have a zero-mean, normally distributed zero-offset output voltage with a variance of 5 mV. How can the manufacturer possibly combine the output of the microphones such that the zero-offset error is minimum?

There can be different combination techniques, but one of the plausible approaches is to combine the sensor data in two stages, as shown in Figure 12.23. First the sensors are categorised into two groups, based on their statistical properties, so that we can apply Eq. (12.46). Then we can apply Eq. (12.32) to combine the output of the intermediate stages:

1) *First stage combination (combination based on statistical properties).*

$$\hat{\mathbf{s}}_1 = \alpha_{11}\mathbf{s}_{11} + \alpha_{12}\mathbf{s}_{12} + \cdots + \alpha_{17}\mathbf{s}_{17} \tag{12.48}$$

Since the sensors have identical statistics, the variance of $\hat{\mathbf{s}}_1$ can be computed as:

$$\sigma_1^2 = \frac{1\,\text{mV}}{7} = 0.14\,\text{mV} \tag{12.49}$$

Similarly,

$$\hat{\mathbf{s}}_2 = \alpha_{21}\mathbf{s}_{21} + \alpha_{22}\mathbf{s}_{22} + \cdots + \alpha_{212}\mathbf{s}_{212} \tag{12.50}$$

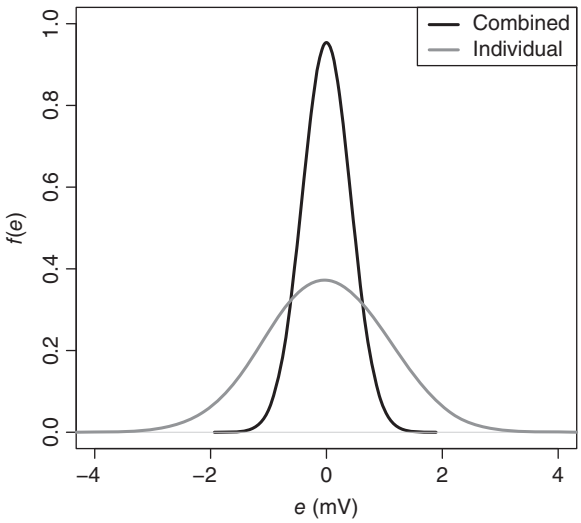


Figure 12.22 A comparison of the zero-offset errors of individual microphones and their combined error.

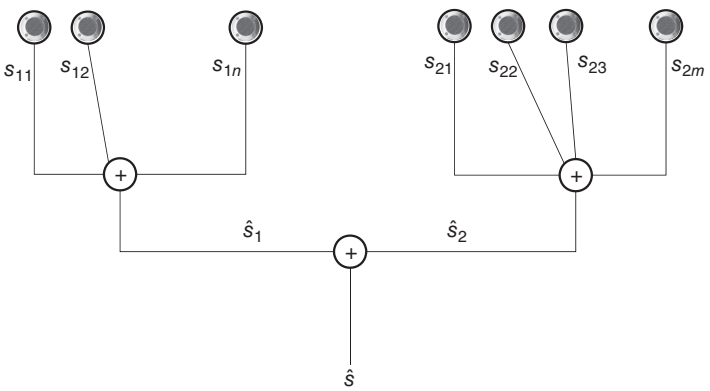


Figure 12.23 Using intermediate stages to systematically combine the outputs of multiple independent sensors having different zero-offset output voltage statistics.

And

$$\sigma_2^2 = \frac{5\text{ mV}}{12} = 0.42\text{ mV} \tag{12.51}$$

2) *Second stage combination.*
We can combine the intermediate stages using Eq. (12.32):

$$\hat{\mathbf{s}} = \hat{\mathbf{s}}_1 + K[\hat{\mathbf{s}}_1 - \hat{\mathbf{s}}_2] \tag{12.52}$$

Using Eq. (12.31) we can determine K:

$$K = \frac{0.14\text{ mV}}{0.14\text{ mV} + 0.42\text{ mV}} = 0.25\text{ mV} \tag{12.53}$$

Hence,

$$\hat{\mathbf{s}} = \hat{\mathbf{s}}_1 + 0.25[\hat{\mathbf{s}}_1 - \hat{\mathbf{s}}_2] \tag{12.54}$$

Likewise, using K, we can compute the variance of $\hat{\mathbf{s}}$ as follows:

$$\sigma_{\hat{s}}^2 = \sigma_1^2 - K\sigma_2^2 = 0.14\,\text{mV} - 0.25(0.42\,\text{mV}) = 0.035 \tag{12.55}$$

Figure 12.24 displays how the error of the ultrasound scanning system is improved stage-by-stage through the systematic combination of the different sensor outputs.

12.5.2 Maximum-likelihood Estimation

In Section 12.5.1, we first decided how to combine the outputs of multiple independent sensors, determined the optimal coefficients for minimising our uncertainty, and produced the joint pdf (Figure 12.24). In this section we shall take the reverse order. Our aim is to determine the optimal and unbiased combination technique that optimises the joint density function.

Recall the two important assertions we made in the beginning of this chapter. One of them was stated as follows:

> The state of reality and the output of a sensor are correlated to a certain extent.

This is the assertion upon which the maximum-likelihood estimation (ML) approach is established. Regardless of the quality of the sensors, ML asserts that the output of each sensor has something to do with the real quantity we wish to determine. Take, as an example, the temperature-to-voltage converter we considered previously. Suppose

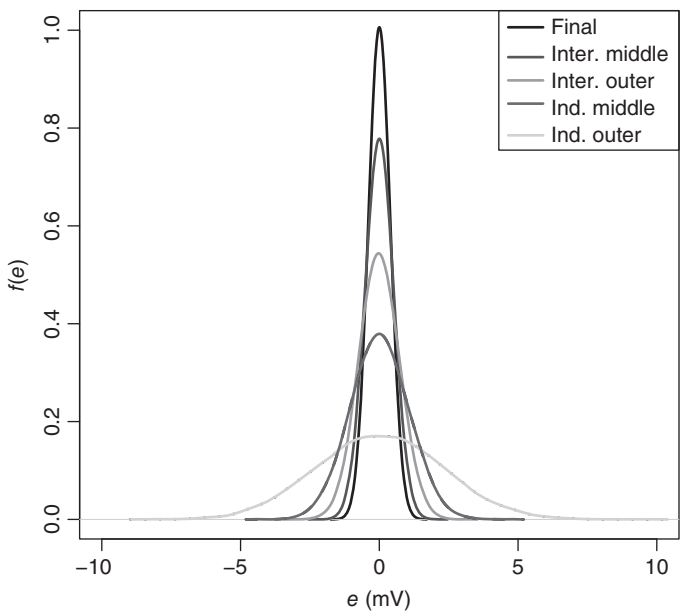


Figure 12.24 Comparison of the pdfs of the different stages during the combination of the zero-offset voltages of different microphones. From bottom to top: The pdfs of the output of the outer microphones, the inner microphones, the intermediate stage combining the outputs of the outer microphones, the intermediate stage combining the outputs of the inner microphones, and the output of the final stage.

between t and $t + dt$ we sample, at a rate of 1 kHz, four independent temperature sensors measuring the temperature of one and the same process and that we obtain the pdfs as shown in Figure 12.25. Assuming that the temperature of the process during this period remains unchanged, we suspect that the most likely temperature the sensors "perceived" should correspond to 2 V. Since the sensors are independent, their joint density function given that they respond to one and the same parameter V – mind you, V is not a random variable, since we assume that the value of the measurand does not change for the time interval (t, dt) – can be computed as:

$$f(\mathbf{s}; V) = f(\mathbf{s}_1; V) f(\mathbf{s}_2; V) \cdots f(\mathbf{s}_n; V) = \prod_{i=1}^{n} f(\mathbf{s}_i; V) \tag{12.56}$$

Since we assume that the output of each sensor has something to do with V, we can express it as follows:

$$\begin{aligned} \mathbf{s}_1 &= V + \mathbf{e}_1 \\ \mathbf{s}_2 &= V + \mathbf{e}_2 \\ &\vdots \\ \mathbf{s}_n &= V + \mathbf{e}_n \end{aligned}$$

If the error introduced by each sensor is a zero-mean, normally distributed random variable, then $\mathbf{s}_i$ is a normally distributed random variable having V as its mean, because:

$$E[\mathbf{s}_i] = V + E[\mathbf{e}_i] = V$$

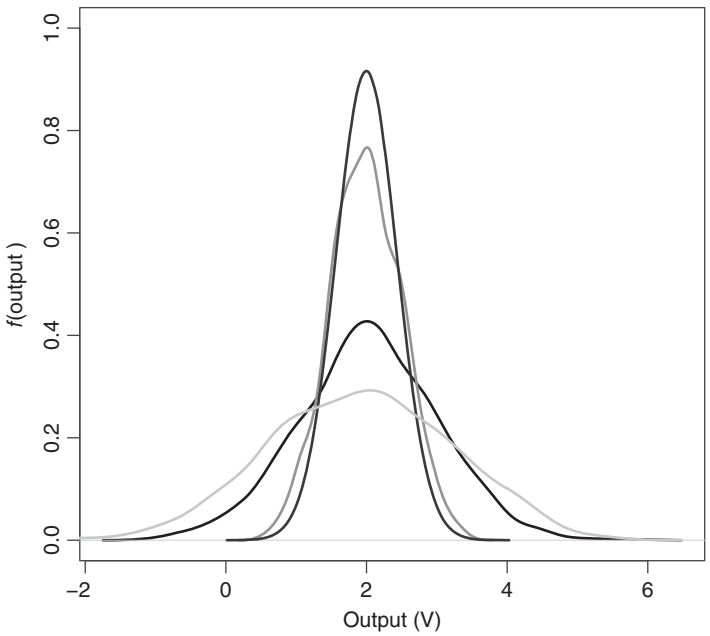


Figure 12.25 Comparison of the probability density functions of four temperature-to-voltage converters measuring the temperature of one and the same process.

and,

$$\sigma_{si}^2 = E[(\mathbf{s}_i - \eta_{si})^2] = E[(V + \mathbf{e}_i - V)^2] = E[(\mathbf{e}_i - \eta_{ei})^2] = \sigma_{ei}^2 = \sigma_i^2$$

Consequently, the density function of the output of each sensor is given as:

$$f(s_i; V) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(s_i - V)^2/2\sigma_i^2} \tag{12.57}$$

The joint density function as a product of the density functions of the individual sensors is given by:

$$f(s_1, s_2, ..., s_n; V) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n}((s_i - V)^2/2\sigma_i^2)} \tag{12.58}$$

Now having the joint density function, we can determine the best combination strategy by differentiating Eq. (12.58) with respect to V, because we are interested in the value of V that results in the highest probability or which is the most likely outcome (hence, the name maximum-likelihood estimation). Alternatively, we can differentiate the logarithmic value of Eq. (12.58) (due to the linearity property of logarithms). Thus, we have:

$$\ln\ (f(s_1, s_2, ..., s_n; V)) = \frac{n}{2} \ln\ (2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(s_i - V)^2}{2\sigma_i^2} \tag{12.59}$$

Differentiating Eq. (12.59) with respect to V and setting the result equal to zero to approximate V in terms of $\mathbf{s}_i$ yields:

$$\sum_{i=1}^{n} \frac{(\mathbf{s}_i - V)}{2\sigma_i^2} = 0 \tag{12.60}$$

Consequently,

$$\hat{V}_{ML} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{s}_i) \tag{12.61}$$

As expected:

$$E[\hat{V}_{ML}(\mathbf{s})] = \frac{1}{n} E\left[\sum_{i=1}^{n} \mathbf{s}_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[\mathbf{s}_i] = V \tag{12.62}$$

The variance of $\hat{V}_{ML}(\mathbf{s})$ can be determined as follows:

$$\sigma_{\hat{V}}^2 = E[(\hat{V}_{ML} - V)^2] = \frac{1}{n^2} E\left[\left(\sum_{i=1}^{n} (\mathbf{s}_i - V)\right)^2\right] \tag{12.63}$$

where V is the true value we wish to approximate and $\hat{V}_{ML}$ is its approximation. Notice that in order to include V into the summation term, we have to divide it by n because it will be added n times as a part of the summation term. The square of the summation term will yield the following:

$$\sigma_{\hat{V}}^2 = \frac{1}{n^2} \left\{ \sum_{i=1}^{n} E[(\mathbf{s}_i - V)^2] + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} E[\mathbf{s}_i - V] E[\mathbf{s}_j - V] \right\} \tag{12.64}$$

The last term of Eq. (12.64) is zero because the sensors are independent and

$$\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} E[\mathbf{s}_i - V]E[\mathbf{s}_j - V] = \sum_{i=1}^{n} E[\mathbf{s}_i - V]\sum_{j=1,j\neq i}^{2} E[\mathbf{s}_j - V] = 0 \tag{12.65}$$

From this we conclude that:

$$\sigma^2_{\hat{V}} = \frac{1}{n^2}\sum_{i=1}^{n} E[(\mathbf{s}_i - V)^2] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \tag{12.66}$$

If all the sensors have the same variance, σ^2, then,

$$\sigma^2_{\hat{V}} = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n} \tag{12.67}$$

As a result,

$$\lim_{n\to\infty}\sigma^2_{\hat{V}} = 0 \tag{12.68}$$

Under the assumption that all sensors produce errors that can be regarded as iid random variables, Eq. (12.67) produces the same result as Eq. (12.46).

12.5.3 Minimum Mean Square Error Estimation

So far we assumed, at least implicitly, that the error has nothing to do with the magnitude of the measurand or any of its properties. In reality, however, this is not always the case; some of the characteristics of the error may change in response to a change in the characteristics of the measurand. As an example, consider Figure 12.26, where the pdf of the error changes with a change in the magnitude of the measurand. Another assumption we have made so far is that the measurand is a fixed or a constant quantity.

Most measurands we wish to sense in real life are themselves continuously changing and should be regarded as random variables. For instance, the temperature, relative humidity, the intensity of light, the quality of air, and the air pressure change in time, however slowly. Through repeated measurements or knowledge of causes and effects we may have the pdfs of these random variables for particular places or processes. So, if we label the measurand as a random variable $\mathbf{m}$ with its own density function, $f(m)$, the minimum mean square error estimation (MMSE) aims to minimise the mean square error between the real $\mathbf{m}$ and its approximation, $\hat{\mathbf{m}}$:

$$\mathbf{e} = \mathbf{m} - \hat{\mathbf{m}} \tag{12.69}$$

If you remember an important assertion made at the beginning of this chapter, at this point you may wish to ask how we can ever measure the error between $\mathbf{m}$ and $\hat{\mathbf{m}}$, as we may never be able to measure $\mathbf{m}$. This is correct; we may not be able to measure $\mathbf{m}$. Nevertheless, let us put this question aside for a while and assume that there is a mechanism to measure the error.

Consequently:

$$E[(\mathbf{m} - \hat{\mathbf{m}})^2] \stackrel{!}{=} minimum \tag{12.70}$$

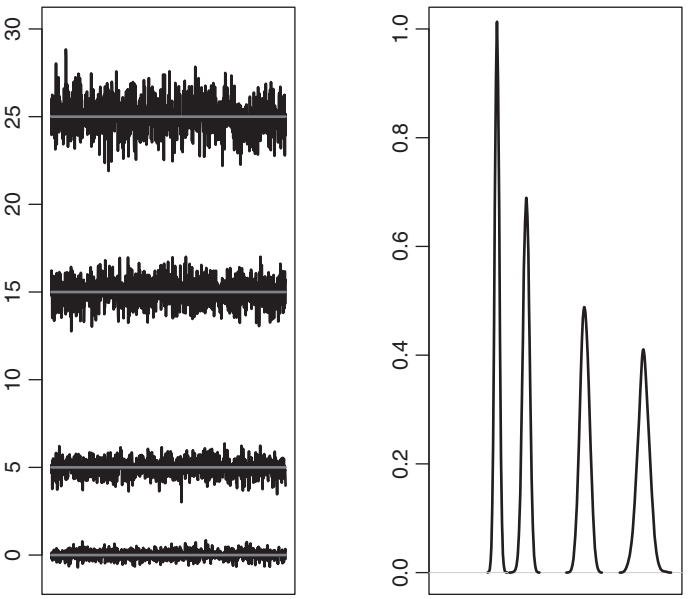


Figure 12.26 An illustration of the change in the pdf of the error of a temperature sensor as the magnitude of the measurand changes.

One way to minimise the error is to employ multiple, independent sensors, just as we did in the previous cases. Thus, if we can approximate the measurand by properly fusing the output of multiple sensors:

$$\hat{\mathbf{m}} = \alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2 + \cdots + \alpha_n \mathbf{s}_n \tag{12.71}$$

We can minimise the error by determining the optimal α_i. In other words, we can differentiate Eq. (12.70) with respect to α_i and set the result to zero:

$$E[(\mathbf{m} - \hat{\mathbf{m}})^2] = E[(\mathbf{m} - (\alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2 + \cdots + \alpha_n \mathbf{s}_n))^2] \tag{12.72}$$

$$\frac{\partial}{\partial \alpha_i} E[(\mathbf{m} - \hat{\mathbf{m}})^2] = E[(\mathbf{m} - (\alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2 + \cdots + \alpha_n \mathbf{s}_n))\mathbf{s}_i] = 0 \tag{12.73}$$

Recalling that,

$$\frac{\partial}{\partial \alpha_i}(\alpha_j \mathbf{s}_j) = 0 \tag{12.74}$$

and, because of our assumption of independence,

$$E[\mathbf{s}_i \mathbf{s}_j] = \eta_i \eta_j \tag{12.75}$$

Listing together all the results, we have,

$$\begin{aligned} E[\mathbf{m}\mathbf{s}_1] &= \alpha_1 E[\mathbf{s}_1^2] + \alpha_2 \eta_1 \eta_2 + \cdots + \alpha_n \eta_1 \eta_n \\ E[\mathbf{m}\mathbf{s}_2] &= \alpha_1 \eta_2 \eta_1 + \alpha_2 E[\mathbf{s}_2^2] \cdots + \alpha_n \eta_2 \eta_n \\ &\vdots \\ E[\mathbf{m}\mathbf{s}_n] &= \alpha_1 \eta_n \eta_1 + \alpha_2 \eta_n \eta_2 + \cdots + \alpha_n E[\mathbf{s}_n^2] \end{aligned} \tag{12.76}$$

If we let $E[\mathbf{ms}_i] = R_{0i}$, $E[\mathbf{s}_i^2] = R_{ii}$ and $E[\mathbf{s}_i\mathbf{s}_j] = \eta_i\eta_j = R_{ij}$, then Eq. (12.76) can be expressed as:

$$\begin{aligned} R_{01} &= \alpha_1 R_{11} + \alpha_2 R_{12} + \cdots + \alpha_n R_{1n} \\ R_{02} &= \alpha_1 R_{21} + \alpha_2 R_{22} + \cdots + \alpha_n R_{2n} \\ &\vdots \\ R_{0n} &= \alpha_1 R_{n1} + \alpha_2 R_{n2} + \cdots + \alpha_n R_{nn} \end{aligned} \tag{12.77}$$

Equation (12.77) can be expressed in matrix form as follows:

$$\begin{bmatrix} R_{01} \\ R_{02} \\ \vdots \\ R_{0n} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & & & \\ R_{n1} & R_{n2} & \cdots & R_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \tag{12.78}$$

At this point, it is imperative to explain some of the variables in Eq. (12.78) and to address the issue of measuring the true value of the measurand. The quantity R_{ii} contains the variance (signifying the error) of the output of the ith sensor ($\mathbf{s}_i$), because $E[\mathbf{s}_i^2] = \sigma_i^2 + \eta_i^2$; σ_i^2 and η_i are usually determined in an environment resembling the typical operating condition of the sensor and by reading the output of the sensor in the absence of a measurand (we have already made reference to this several times). If one can take enough samples, the pdf of the error and, hence, σ_i^2 and η_i, can be determined. The quantity R_{0i} relates the output of the ith sensor to the true value of the measurand. The quantity R_{0i} is determined in a laboratory setting or by the manufacturer of the sensor itself. Here as well, in an environment resembling the typical operation condition and spanning the entire sensing range of the sensor, the sensor is given known inputs ('known' meaning that the controlled input is measured by a highly accurate device, although still an approximation) and for each input, the conditional pdf $f(\mathbf{s}_i|m)$ is carefully determined. Then, the joint pdf can be obtained by multiplying the conditional density function by the density function of the measurand:

$$f(s_i, m) = f(\mathbf{s}_i|m)f(m) \tag{12.79}$$

$f(m)$ is the pdf of $\mathbf{m}$, which we take as a random variable. Our knowledge of $f(m)$ comes from our knowledge of the measurand. Then $E[\mathbf{ms}_i]$ can be determined as:

$$E[\mathbf{ms}_i] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} m\, s_i\, f(s_i, m)\, ds_i\, dm \tag{12.80}$$

Finally, the MMSE coefficients can be determined as:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & & & \\ R_{n1} & R_{n2} & \cdots & R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{01} \\ R_{02} \\ \vdots \\ R_{0n} \end{bmatrix} \tag{12.81}$$

Example 12.7 Suppose we use two sensors to estimate the outside temperature. The two sensors have a zero-mean, normally distributed error with different statistical properties and we wish to determine the best way (in the MMSE sense) of combining evidence from these sensors.

The estimated temperature can be expressed as:

$$\hat{\mathbf{T}} = \alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2 \tag{12.82}$$

$$\mathbf{e} = \mathbf{T} - \hat{\mathbf{T}} = \mathbf{T} - (\alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2) \tag{12.83}$$

$$\frac{\partial}{\partial \alpha_1} E[\mathbf{e}^2] = E[(\mathbf{T} - (\alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2))(-\mathbf{s}_1)] = 0 \tag{12.84}$$

Likewise,

$$\frac{\partial}{\partial \alpha_2} E[\mathbf{e}^2] = E[(\mathbf{T} - (\alpha_1 \mathbf{s}_1 + \alpha_2 \mathbf{s}_2))(-\mathbf{s}_2)] = 0 \tag{12.85}$$

From this we have:

$$\begin{aligned} E[\mathbf{T}\mathbf{s}_1] &= \alpha_1 E[\mathbf{s}_1^2] \\ E[\mathbf{T}\mathbf{s}_2] &= \alpha_2 E[\mathbf{s}_2^2] \end{aligned} \tag{12.86}$$

We obtained simplified expressions in Eq. (12.86) because the errors in both sensors have zero means and, as a result, $E[\mathbf{s}_1 \mathbf{s}_2] = 0$. Moreover, $E[\mathbf{s}_1^2] = \sigma_1^2$ and $E[\mathbf{s}_2^2] = \sigma_2^2$. Therefore,

$$\alpha_1 = \frac{E[\mathbf{T}\mathbf{s}_1]}{\sigma_1^2} \tag{12.87}$$

and,

$$\alpha_2 = \frac{E[\mathbf{T}\mathbf{s}_2]}{\sigma_2^2} \tag{12.88}$$

12.5.4 Kalman Filter

So far, even though our evidence combination strategies gradually became more complex, we have nevertheless been entirely dependent on the measurements we got from the sensors in order to determine the values of a measurand. We can reduce our uncertainty about the measurand if we can add knowledge from a different domain. The second assertion made at the beginning of this chapter can serve us towards this end:

> The change in the physical reality (measurand) is a gradual process rather than being haphazard and wild; statistically speaking, the measurand is correlated with itself to a certain extent.

One way of interpreting this assertion is that the future values of a measurand are, to a certain extent, explainable in terms of the present, in the same way its present value is explainable in terms of its past values. Perhaps the poet T.S. Eliot had this in mind when he composed the opening verses of *Burnt Norton in Four Quarters*:

> Time present and time past
> Are both perhaps present in time future,
> And time future contained in time past.
> If all time is eternally present
> All time is unredeemable.
> What might have been is an abstraction

Remaining a perpetual possibility
Only in a world of speculation.
What might have been and what has been
Point to one end, which is always present.

To give a concrete example, the temperature in the city where I am living begins to decline steadily towards the end of September all the way through to the end of March, even though it goes up and down in between. Suppose, based on the knowledge I have up to time $t-1$ (whatever my source of knowledge may be), I predict the temperature for time t with a certain degree of accuracy. Let us label this measurement as $\mathbf{x}_{\mathrm{p}}(t)$ (notice the indices p and t; p stands for prediction, because I have not yet made a measurement for time t; t indicates that the prediction is made for time t when I am at time $t-1$). When time t arrives, I make a measurement using a temperature sensor. Let's label this measurement as $\mathbf{x}_{\mathrm{m}}(t)$, where the indices m and t represent a measurement taken at time t. Both my prediction and measurement are random variables on account of the uncertainty stemming from prediction and measurement errors. By carefully combining $\mathbf{x}_{\mathrm{p}}(t)$ and $\mathbf{x}_{\mathrm{m}}(t)$ I can get $\hat{\mathbf{x}}_{\mathrm{e}}(t)$, the uncertainty of which is less than if I were to rely on either $\mathbf{x}_{\mathrm{p}}(t)$ or $\mathbf{x}_{\mathrm{m}}(t)$. Indeed, I can now even improve my prediction of the temperature for time $t+1$ due to my improved estimation of the temperature of t. This is illustrated in Figure 12.27. Formally, the Kalman filter is described by two equations:

$$\mathbf{x}_{\mathrm{m}}(t) = \mathbf{x}(t) + \mathbf{v}(t) \tag{12.89}$$

$$\mathbf{x}(t+1) = \mathbf{x}(t) + \mathbf{w}(t) \tag{12.90}$$

where $\mathbf{x}(t)$ is the random variable we are interested in estimating but will never be able to directly measure, $\mathbf{v}(t)$ is the measurement error at time t modelled as a random variable, and $\mathbf{w}(t)$ is the error made in the prediction as a result of the inherent randomness in the measurand (hence, the measurand is also regarded as a random variable).

To illustrate the above relations by example, suppose we wish to estimate the temperature variation of the city of Dresden for the months between the beginning of September and the end of March; a total of 210 days. Each day exactly at noon we measure the temperature of a particular location and predict the temperature of that same location for the next day (at noon). Our aim is to improve our knowledge of the temperature of each day by carefully combining the values of our prediction and measurement. Suppose that scientific evidence shows that the temperature of Dresden falls about 2% daily in the time period we are concerned with:

$$\mathbf{T}(t+1) = 0.98\mathbf{T}(t) \tag{12.91}$$

This is illustrated in Figure 12.28a. The scientific claim is, of course, very optimistic, because the future temperature, even though it is to some extent correlated with the present, entails also some randomness; otherwise we need not employ any sensor at all. Suppose the actual temperature variation looks like the trace shown in Figure 12.28b. The good thing about the correlation in the temperature variation is that we can include a process error in Eq. (12.91) to accommodate the randomness in the temperature (process) variation.

$$\mathbf{T}(t+1) = 0.98\mathbf{T}(t) + \mathbf{w}(t) \tag{12.92}$$

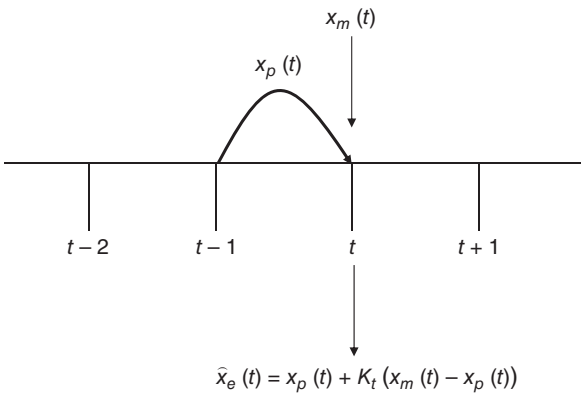
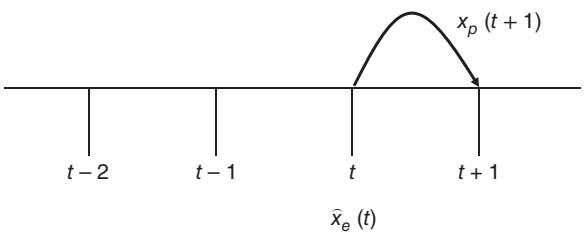


Figure 12.27 The basic principle of a Kalman Filter. The Kalman approach combines two types of evidence: one from knowledge of how the measurand evolves in time ($\mathbf{x}_p(t)$), the other from measurement ($\mathbf{x}_m(t)$). The idea is to reduce the uncertainty in the combined evidence $\hat{\mathbf{x}}_e(t)$ by properly weighing $\mathbf{x}_p(t)$ and $\mathbf{x}_m(t)$. As the bottom part illustrates, as our uncertainty decreases, the propagation of our belief into the future further reduces our uncertainty.



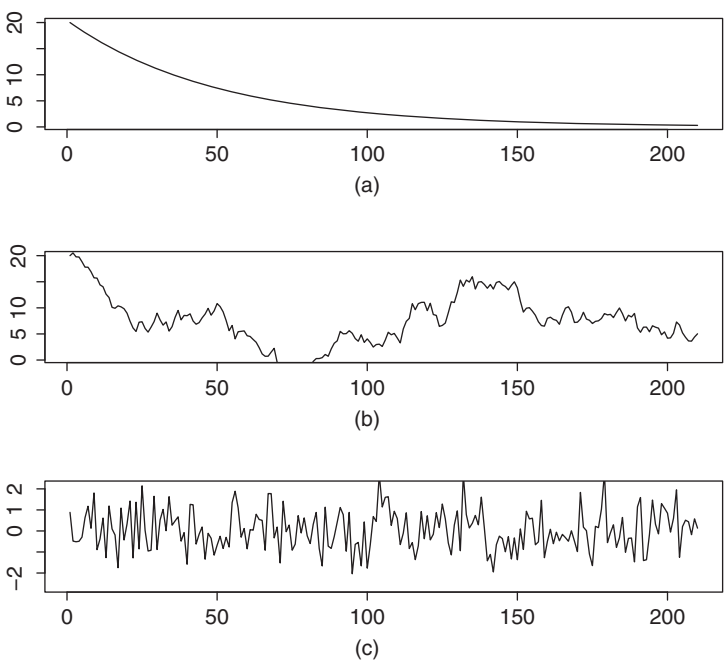


Figure 12.28 An illustration of the temperature variation over time: (a) the temperature at $t + 1$ expressed as a function of the temperature at t, in other words, $T(t + 1) = 0.98T(t)$; (b) the actual temperature; (c) the process error modelled as a zero-mean, normally distributed random variable.

The process error is shown in Figure 12.28c. At this point it should be noted that Eq. (12.92) should not be confused with the prediction for the time $t + 1$, which we label as $\mathbf{T}_{\mathrm{p}}(t + 1)$. Equation (12.92) stems from the second assertion made at the beginning of the chapter and repeated at the beginning of this section. $\mathbf{T}_{\mathrm{p}}(t + 1)$, on the other hand, depends on the knowledge oft the temperature of t, but this should become clear in the subsequent explanation.

The measurement we take each day using a sensor contains the actual temperature, but this is mixed with the inherent error of the sensor. Thus, we can express it as follows:

$$\mathbf{T}_{\mathrm{m}}(t) = \mathbf{T}(t) + \mathbf{v}(t) \tag{12.93}$$

Notice that even though the temperature of the present as well as the future are scalar quantities, due to $\mathbf{w}$ and $\mathbf{v}$, they should be taken as random variables. If we let $\hat{\mathbf{T}}(t)$ be our estimation of the temperature for the time t, then the mean square error at time t is expressed as follows:

$$P(t) = E[\mathbf{e}^2] = E[(\mathbf{T}(t) - \hat{\mathbf{T}}(t))(\mathbf{T}(t) - \hat{\mathbf{T}}(t))] \tag{12.94}$$

The estimated temperature for time t can be expressed in terms of the predicted temperature for time t and the measured temperature for time t using Eq. (12.32):

$$\hat{\mathbf{T}}(t) = \mathbf{T}_{\mathrm{p}}(t) + K(t)[\mathbf{T}_{\mathrm{m}}(t) - \mathbf{T}_{\mathrm{p}}(t)] \tag{12.95}$$

Since $\mathbf{T}_{\mathrm{m}}(t) = \mathbf{T}(t) + \mathbf{v}(t)$, Eq. (12.95) can be rewritten as:

$$\hat{\mathbf{T}}(t) = \mathbf{T}_{\mathrm{p}}(t) + K(t)[\mathbf{T}(t) + \mathbf{v}(t) - \mathbf{T}_{\mathrm{p}}(t)] \tag{12.96}$$

Substituting Eq. (12.96) into Eq. (12.94) yields,

$$P(t) = E\{([1 - K(t)][\mathbf{T}(t) - \mathbf{T}_{\mathrm{p}}(t)] - K(t)\mathbf{v}(t))^2\} \tag{12.97}$$

The term $E[\mathbf{T}(t) - \mathbf{T}_{\mathrm{p}}(t)]$, denoted as $P_{\mathrm{p}}(t)$, quantifies the prediction error and is sometimes known as the error of the priori estimate. This error does not correlate with the measurement error, since the prediction is made before the measurement is taken. Now we can rewrite Eq. (12.97) in terms of the prediction error:

$$P(t) = (1 - K(t))^2 P_{\mathrm{p}}(t) + K^2(t)R \tag{12.98}$$

where $R = E[\mathbf{v}^2]$. If the measurement error is a zero-mean error, then $R = \sigma_v^2$. If we distribute Eq. (12.98), it yields,

$$P(t) = P_{\mathrm{p}}(t) - 2K(t)P_{\mathrm{p}}(t) - K^2(t)(P_{\mathrm{p}}(t) + R) \tag{12.99}$$

We are now in a position to choose the optimal $K(t)$ that can minimise our estimation error at time t. This can be done by differentiating Eq. (12.99) with respect to $K(t)$ and setting the result to zero. The result is:

$$K(t) = \frac{P_{\mathrm{p}}(t)}{P_{\mathrm{p}}(t) + R} \tag{12.100}$$

Finally, substituting Eq. (12.100) into Eq. (12.99) reduces the expression for $P(t)$ into:

$$P(t) = (1 - K(t))P_{\mathrm{p}}(t) \tag{12.101}$$

One way to continuously improve our ability to predict is to propagate or project the estimation error into the future. Hence our prediction of the temperature for time $t + 1$ based on the evidence we have at time t is:

$$\mathbf{T}_{\mathrm{p}}(t+1) = 0.98\hat{\mathbf{T}}(t) \tag{12.102}$$

The prediction error for time $t + 1$ is,

$$\mathbf{e}_{\mathrm{p}}(t+1) = \mathbf{T}(t+1) - \mathbf{T}_{\mathrm{p}}(t+1) = 0.98\mathbf{T}(t) + \mathbf{w}(t) - 0.98\hat{\mathbf{T}}(t) = 0.98\mathbf{e}(t) + \mathbf{w}(t) \tag{12.103}$$

The MMSE of the predicted temperature for time $t + 1$, $P_{\mathrm{p}}(t + 1)$, is:

$$P_{\mathrm{p}}(t+1) = E[\mathbf{e}_{\mathrm{p}}^2(t+1)] = (0.98)^2 E[\mathbf{e}^2(t)] + E[\mathbf{w}^2(t)] = 0.9604P(t) + Q \tag{12.104}$$

where $Q = E[\mathbf{w}^2(t)]$. If the process error is a zero-mean error, then, $Q = \sigma_w^2$. With $P_{\mathrm{p}}(t)$, $P(t)$, and $P_{\mathrm{p}}(t + 1)$ the Kalman Filter connects the past, the present, and the future, in the same way the poet maintains:

> Time present and time past
> Are both perhaps present in time future,
> And time future contained in time past.

Consequently, we can now estimate step by step the temperature of each day. Suppose we define:

$$\mathbf{T}_{\mathrm{p}}(t) = 0.75\hat{\mathbf{T}}(t-1) + 0.25\hat{\mathbf{T}}(t-2)$$

For the first time slot, we shall have no predicted value and therefore the estimated and the measured data are equal:

$$\hat{\mathbf{T}}(1) = \mathbf{T}_{\mathrm{m}}(1)$$

We still do not have sufficient data to make a prediction that satisfies our definition of $\mathbf{T}_{\mathrm{p}}(t)$, but we can make a reasonable estimate using the evidence we have:

$$\mathbf{T}_{\mathrm{p}}(2) = 0.75\hat{\mathbf{T}}(1)$$

With $\mathbf{T}_{\mathrm{p}}(2)$ and $\mathbf{T}_{\mathrm{m}}(2)$ (which we have, because we can always measure them), we can compute $\hat{\mathbf{T}}(2)$, but for that we need $K(2)$, which we do not have. Once again, we can make a reasonable guess; as we don't have any evidence to mistrust either the predicted or the measured values, we can set $K(2) = 0.5$. Thus,

$$\hat{\mathbf{T}}(2) = \mathbf{T}_{\mathrm{p}}(2) + K(2)[\mathbf{T}_{\mathrm{m}}(2) - \mathbf{T}_{\mathrm{p}}(2)]$$

With $K(2)$ determined, we can also determine $P_{\mathrm{p}}(2)$, $P(2)$, and $P_{\mathrm{p}}(2 + 1)$ using Eqs 12.100 –12.102. And with these we can move on to predicting and estimating the temperature for the next time slot, and so on. Figure 12.29 lists the program code in *R* for computing all the remaining values for each parameter we require for determining prediction and estimations. Figure 12.30 displays the three important temperature values. Clearly, the Kalman estimation is more accurate than either the predicted or measured values.

```
# process noise modelled as a zero-mean normal distribution with a variance of 1
w <- rnorm(210, 0, 1)

# measurement noise modelled as a zero-mean normal distribution with a variance of 9
v <- rnorm(210, 0, 3)

# initialising the vector  for the actual temperature
T <- rep(0, 210)

# The temperature of day one being 20 degree Celsius
T[1] <- 20

# The actual temperature initialised
for(t in 2:length(T) ) {
 T[i] <- 0.98 * T[t - 1] + w[t]
 }

# initialising the vector of the predicted temperature
Tp <- rep(0, 210)
# initialising the vector of the measured temperature
Tm <- rep(0, 210)
# initialising the vector of the estimated temperature
Th <- rep(0, 210)
# initialising the vector of the Kalman constants
K <- rep(0, 210)
# initialising the vector of the minimum mean square prediction error
Pp <- rep(0, 210)
# initialising the vector of the minimum mean square estimation error
P <- rep(0, 210)

# Tm as the addition of the actual temperature and the measurement noise
Tm <- T + v
# Th of day one is set to equal the measured temperature of day one
Th[1] <- Tm[1]
# the predicted temperature of day two is set as 0.75 times Th[1]
Tp[2] <-  0.75 * Th[1]
# setting K[2] = 0.5, because I initially trust Tm[2] and Tp[2] equally
K[2] <- 0.5
# with K[2] = 0.5, Pp[2] will equal to 1
Pp[2] <- 1
# The minimum mean square error of t = 2 is computed
P[2] <- (1-k[2]) * Pp[2]

# the temperature of t = 2 is estimated
Th[2] <- Tp[2] + K[2] * (Tm[2] - Tp[2])

# Computing all the parameters of the Kalman filter

for(t in 3:(length(Tm)-1)) {
 Pp[t] <- P[i-1] * var(w)
 K[t] <- Pp[i] / (Pp[i] + var(v))
# I compute Tp[t] by combining the last two estimated values
 Tp[t] <- 0.75 * Th[t-1] + 0.25 * Th[t-2]
 Th[t] <- Tp[i] + K[t] * (Tm[t] - Tp[t])
 P[t] <- (1- K[t]) * Pp[t]
 }
```

Figure 12.29 Code written in *R* to estimate the temperature fluctuation of 210 days using a Kalman filter. Tp: $\mathbf{T}_{\mathrm{p}}(t)$, Tm: $\mathbf{T}_{\mathrm{m}}(t)$, Th: $\hat{\mathbf{T}}(t)$, Pp: $P_{\mathrm{p}}(t)$, and P: $P(t)$.

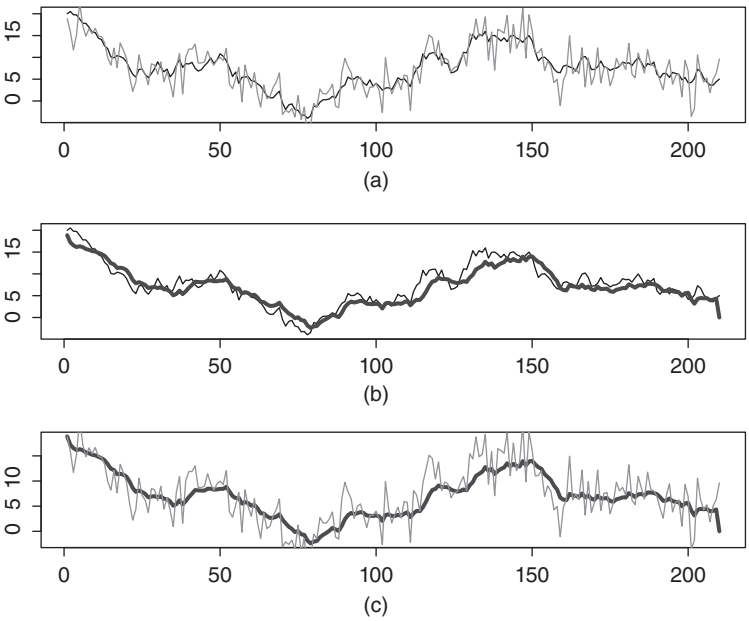


Figure 12.30 Application of the Kalman filter to estimate the temperature variation over time. (a) difference between the actual variation and the measured values; (b) difference between the actual variation and the Kalman estimation (which combines the predicted values and the measured values). (c) difference between the measured and the Kalman estimated values.

12.5.5 The Kalman Filter Formalism

In Eq. (12.89), we assumed (implicitly) that the output of the sensor and the measurand are one and the same type. But this is not generally the case. For the temperature-to-voltage converter sensor, for example, the input is a temperature (**T**) and the output is a voltage (**V**). Therefore, we need a more inclusive approach to formally express the Kalman filter. If **x** is depicted as the measurand which we wish to estimate and **s** is the out put of a sensor (the measurement), the relationship between the present and future values of the measurand are related to one another as follows:

$$\mathbf{x}(t+1) = \Phi\mathbf{x}(t) + \mathbf{w}(t) \tag{12.105}$$

where Φ is the state transition matrix, from t to $t+1$, and is assumed to be statistically stationary and **w**(t) is called the process error at time t. Likewise,

$$\mathbf{s}(t) = H\mathbf{m}(t) + \mathbf{v}(t) \tag{12.106}$$

where H establishes an ideal relationship between **s** (t) and **m**(t). With these adjustments in mind, the estimated value of the measurand for time t can be expressed as:

$$\hat{\mathbf{x}}(t) = \mathbf{x}_{\mathrm{p}}(t) + K(t)[\mathbf{s}(t) - H\mathbf{x}_{\mathrm{p}}(t)] \tag{12.107}$$

where $H\mathbf{x}_{\mathrm{p}}(t)$ predicts the sensor output (no measurement is taken yet for time t). The Kalman constant $K(t)$ should have the appropriate unit so as to enable the addition of the two terms on the right side in Eq. (12.107). In the same way, we can modify the

expressions for the Kalman constant and the MMSE for time t as well as the prediction error for time $t + 1$ as follows:

$$K(t) = \frac{HP_{\mathrm{p}}(t)}{H^2 P_{\mathrm{p}}(t) + R} \tag{12.108}$$

$$P(t) = (1 - HK(t))P_{\mathrm{p}}(t) \tag{12.109}$$

$$P_{\mathrm{p}}(t+1) = \Phi^2 P(t) + Q \tag{12.110}$$

In conclusion, the estimation of a measurand may deal with past, present, or future values, depending on what we wish to do with the sensor data. In his original publication Kalman (1960) uses the term "estimation" to collectively describe the problem of interpolation, filtering, and prediction. He describes the individual problems as follows (I have made a slight adjustment to the parameter depiction to make the text readable):

> We are given signal $\mathbf{x}_1(t)$ and noise $\mathbf{x}_2(t)$. Only the sum $\mathbf{y}(t) = \mathbf{x}_1(t) + \mathbf{x}_2(t)$ can be observed. Suppose we have observed and know exactly the values of $y(0),\ldots, y(n)$. What can we infer from this knowledge in regard to the (unobservable) value of the signal at [time] t, where t may be less than, equal to, or greater than n? If $t < n$, this is a data-smoothing (interpolation) problem. If $t = n$, this is called filtering. If $t > n$, we have a prediction problem. Since our treatment will be general enough to include these and similar problems, we shall use hereafter the collective term estimation.

References

Gardiner CW 1985 *Handbook of Stochastic Methods*, vol. 3. Springer.

Grewal MS 2011 *Kalman Filtering*. Springer.

Kalman RE 1960 A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**(1), 35–45.

Papoulis A and Pillai SU 2002 *Probability, Random Variables, and Stochastic Processes*. Tata McGraw-Hill Education.

Ross SM *et al*. 1996 *Stochastic Processes*, vol. **2**. John Wiley & Sons.