ESTIMATION GIVEN NO OBSERVATIONS

We now initiate our discussions of estimation theory by posing and solving a simple (almost trivial) estimation problem. Thus suppose that all we know about a real-valued random variable x is its mean $\bar{x}$ and its variance σ_x^2, and that we wish to estimate the value that x will assume in a given experiment. We shall denote the *estimate* of x by $\hat{x}$; it is a deterministic quantity (i.e., a number). But how do we come up with a value for $\hat{x}$? And how do we decide whether this value is optimal or not? And if optimal, in what sense? These inquiries are at the heart of every estimation problem.

To answer these questions, we first need to choose a cost function to penalize the estimation error. The resulting estimate $\hat{x}$ will be optimal only in the sense that it leads to the smallest cost value. Different choices for the cost function will in general lead to different choices for $\hat{x}$, each of which will be optimal in its own way.

The design criterion we shall adopt is the so-called *mean-square-error* criterion. It is based on introducing the error signal

$$\tilde{x} \triangleq x - \hat{x}$$

and then determining $\hat{x}$ by minimizing the mean-square-error (m.s.e.), which is defined as the expected value of $\tilde{x}^2$, i.e.,

$$\min_{\hat{x}} \mathsf{E}\, \tilde{x}^2 \tag{1.2.1}$$

The error $\tilde{x}$ is a random variable since x is random. The resulting estimate, $\hat{x}$, will be called the *least-mean-squares estimate* of x. The following result is immediate (and, in fact, intuitively obvious as we explain below).

> **Lemma 1.2.1 (Lack of observations)** The least-mean-squares estimate of x given knowledge of only $(\bar{x}, \sigma_x^2)$ is $\hat{x} = \bar{x}$. The resulting minimum cost is $\mathsf{E}\,\tilde{x}^2 = \sigma_x^2$.

Proof: Expand the mean-square error by subtracting and adding $\bar{x}$ as follows:

$$\mathsf{E}\,\tilde{x}^2 = \mathsf{E}\,(x - \hat{x})^2 = \mathsf{E}\,[(x - \bar{x}) + (\bar{x} - \hat{x})]^2 = \sigma_x^2 + (\bar{x} - \hat{x})^2$$

The choice of $\hat{x}$ that minimizes the m.s.e. is now evident. Only the term $(\bar{x} - \hat{x})^2$ is dependent on $\hat{x}$ and this term can be annihilated by choosing $\hat{x} = \bar{x}$. The resulting minimum mean-square error (m.m.s.e.) is then

$$\text{m.m.s.e.} \triangleq \mathsf{E}\,\tilde{x}^2 = \sigma_x^2$$

An alternative derivation would be to expand the cost function as

$$\mathsf{E}\,(x - \hat{x})^2 = \mathsf{E}\,x^2 - 2\bar{x}\hat{x} + \hat{x}^2$$

and to differentiate it with respect to $\hat{x}$. By setting the derivative equal to zero we arrive at the same conclusion, namely, $\hat{x} = \bar{x}$.

◇

There are several good reasons for choosing the mean-square-error criterion (1.2.1). The simplest one perhaps is that the criterion is amenable to mathematical manipulations, more so than any other criterion. In addition, the criterion is in effect attempting to force the estimation error to assume values close to its mean, which happens to be zero since

$$\mathsf{E}\,\tilde{x} = \mathsf{E}\,(x - \hat{x}) = \mathsf{E}\,(x - \bar{x}) = \bar{x} - \bar{x} = 0$$

Therefore, by minimizing $\mathsf{E}\,\tilde{x}^2$ we are in effect minimizing the variance of the error. And in view of the discussion in Sec. 1.1 regarding the interpretation of the variance of a random variable, we see that the mean-square-error criterion tries to increase the likelihood of small errors.

The effectiveness of this estimation procedure can be measured by examining the value of the resulting minimum cost, which is the variance of the resulting estimation error. The above lemma tells us that the minimum cost is equal to σ_x^2. That is,

$$\sigma_{\tilde{x}}^2 = \sigma_x^2$$

so that the estimate $\hat{x} = \bar{x}$ does not reduce our initial uncertainty about x since the error variable still has the same variance as x itself! We thus find that the performance of the mean-square-error design procedure is rather limited in this case. Of course, we are more interested in estimation procedures that result in error variances that are smaller than the original signal variance. We shall discuss one such procedure in the next section.

The reason for the poor performance of the estimate $\hat{x} = \bar{x}$ lies in the lack of more sophisticated prior information about x. Note that Lemma 1.2.1 simply tells us that the best we can do, in the absence of any other information about a random variable x, other than its mean and variance, is to use the mean value of x as our estimate. This statement is, in a sense, intuitive. After all, the mean value of a random variable is, by definition, an indication of the value that we would expect to occur on average in repeated experiments. Hence, in answer to the question: what is the best guess for x?, the analysis tells us that the best guess is what we would expect for x on average! This is a circular answer, but one that is at least consistent with intuition.

Example 1.2.1 (Binary signal) Assume x represents a BPSK (binary phase-shift keying) signal that is equal to ± 1 with probability $1/2$ each. Then

$$\bar{x} = \frac{1}{2}\cdot(1) + \frac{1}{2}\cdot(-1) = 0$$

and

$$\sigma_x^2 = \mathsf{E}\,x^2 = 1$$

Now given knowledge of $\{\bar{x}, \sigma_x^2\}$ alone, the best estimate for x in the least-mean-squares sense is $\hat{x} = \bar{x} = 0$. This example shows that the least-mean-squares (and, hence, optimal) estimate does not always lead to a meaningful solution! In this case, $\hat{x} = 0$ is not useful in guessing whether x is 1 or -1 in a given realization. If we could incorporate into the design of the estimator the knowledge that x is a BPSK signal, or some other related information, then we could perhaps come up with a better estimate for x.

$\diamond$

Mean-Square-Error Criterion

The criterion we shall use to determine the estimator $\hat{\boldsymbol{x}}$ is still the mean-square-error criterion. We define the error signal

$$\boxed{\tilde{\boldsymbol{x}} \triangleq \boldsymbol{x} - \hat{\boldsymbol{x}}} \tag{1.3.2}$$

and then determine $\hat{\boldsymbol{x}}$ by minimizing the mean-square-error over all possible functions $h(\cdot)$:

$$\boxed{\min_{h(\cdot)} \; \mathsf{E}\, \tilde{\boldsymbol{x}}^2} \tag{1.3.3}$$

The solution is given by the following statement.

Theorem 1.3.1 (Optimal mean-square-error estimator) The least-mean-squares estimator (l.m.s.e.) of $\boldsymbol{x}$ given $\boldsymbol{y}$ is the conditional expectation of $\boldsymbol{x}$ given $\boldsymbol{y}$, i.e., $\hat{\boldsymbol{x}} = \mathsf{E}(\boldsymbol{x}|\boldsymbol{y})$. The resulting estimate is

$$\hat{x} = \mathsf{E}(\boldsymbol{x}|\boldsymbol{y} = y) \;=\; \int_{S_x} x f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)\mathrm{d}x$$

where S_x denotes the support (or domain) of the random variable $\boldsymbol{x}$. Moreover, the estimator is unbiased, i.e., $\mathsf{E}\,\hat{\boldsymbol{x}} = \bar{x}$, and the resulting minimum cost is given by either expression

$$\mathsf{E}\,\tilde{\boldsymbol{x}}^2 = \mathsf{E}\,\boldsymbol{x}^2 \;-\; \mathsf{E}\,\hat{\boldsymbol{x}}^2 \;=\; \sigma_x^2 \;-\; \sigma_{\hat{x}}^2$$

Proof: There are several ways to establish the result. Our argument is based on recalling that for any two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$, it holds that (see Prob. 1.4):

$$\boxed{\mathsf{E}\,\boldsymbol{x} \;=\; \mathsf{E}\,[\mathsf{E}\,(\boldsymbol{x}|\boldsymbol{y})]} \tag{1.3.4}$$

where the outermost expectation on the right-hand side is with respect to $\boldsymbol{y}$, while the innermost expectation is with respect to $\boldsymbol{x}$. We shall indicate these facts explicitly by showing the variables with respect to which the expectations are performed, so that

$$\mathsf{E}\,\boldsymbol{x} \;=\; \mathsf{E}_{y}[\mathsf{E}_{x}(\boldsymbol{x}|\boldsymbol{y})]$$

It now follows that, for any function of $\boldsymbol{y}$, say $g(\boldsymbol{y})$, it holds that

$$\mathsf{E}_{x,y}\, \boldsymbol{x}g(\boldsymbol{y}) = \mathsf{E}_y\left[\mathsf{E}_x(\boldsymbol{x}g(\boldsymbol{y})|\boldsymbol{y})\right] = \mathsf{E}_y\left[\mathsf{E}_x(\boldsymbol{x}|\boldsymbol{y})g(\boldsymbol{y})\right] = \mathsf{E}_{x,y}\left[\mathsf{E}_x(\boldsymbol{x}|\boldsymbol{y})g(\boldsymbol{y})\right]$$

This means that, for any $g(\boldsymbol{y})$,

$$\mathsf{E}_{x,y}\left[\boldsymbol{x} - \mathsf{E}_x(\boldsymbol{x}|\boldsymbol{y})\right] g(\boldsymbol{y}) = 0$$

which we write more compactly as

$$\boxed{\mathsf{E}\,[\boldsymbol{x} - \mathsf{E}\,(\boldsymbol{x}|\boldsymbol{y})]\, g(\boldsymbol{y}) = 0} \tag{1.3.5}$$

Expression (1.3.5) states that the random variable $\boldsymbol{x} - \mathsf{E}\,(\boldsymbol{x}|\boldsymbol{y})$ is uncorrelated with any function $g(\cdot)$ of $\boldsymbol{y}$.[3]

[3] As mentioned before, two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are uncorrelated if, and only if, their cross-correlation is zero, i.e., $\mathsf{E}\,(\boldsymbol{x} - \bar{x})(\boldsymbol{y} - \bar{y}) = 0$. On the other hand, the random variables are said to be *orthogonal* if, and only if, $\mathsf{E}\,\boldsymbol{xy} = 0$. It is easy to verify that the concepts of orthogonality and uncorrelatedness coincide if at least one of the random variables is zero mean. From equation (1.3.5) we conclude that the variables $\boldsymbol{x} - \mathsf{E}\,(\boldsymbol{x}|\boldsymbol{y})$ and $g(\boldsymbol{y})$ are orthogonal. However, since $\boldsymbol{x} - \mathsf{E}\,(\boldsymbol{x}|\boldsymbol{y})$ is zero mean, then we can also say that they are uncorrelated.

Using this intermediate result, we return to the cost function (1.3.3), add and subtract $\mathsf{E}(x|y)$ to its argument, and express it as

$$\mathsf{E}(x-\hat{x})^2 = \mathsf{E}[x - \mathsf{E}(x|y) + \mathsf{E}(x|y) - \hat{x}]^2$$

The term $\mathsf{E}(x|y) - \hat{x}$ is a function of y. Therefore, if we choose $g(y) = \mathsf{E}(x|y) - \hat{x}$, then from the uncorrelatedness property (1.3.5) we conclude that

$$\mathsf{E}(x-\hat{x})^2 = \mathsf{E}[x - \mathsf{E}(x|y)]^2 \; + \; \mathsf{E}[\mathsf{E}(x|y) - \hat{x}]^2$$

Only the second term on the right-hand side is dependent on $\hat{x}$ and the m.s.e. is minimized by choosing $\hat{x} = \mathsf{E}(x|y)$.

To evaluate the resulting m.m.s.e. we first note that the optimal estimator is unbiased since

$$\mathsf{E}\hat{x} \;=\; \mathsf{E}[\mathsf{E}(x|y)] \;=\; \mathsf{E}x \;=\; \bar{x}$$

so that its variance is given by

$$\sigma_{\hat{x}}^2 = \mathsf{E}\hat{x}^2 \; - \; \bar{x}^2$$

Moreover, in view of the uncorrelatedness property (1.3.5), and in view of the fact that the optimal estimator $\hat{x} = \mathsf{E}(x|y)$ is itself a function of y, we have

$$\boxed{\mathsf{E}(x-\hat{x})\hat{x} = 0} \tag{1.3.6}$$

In other words, the estimation error, $\tilde{x}$, is also uncorrelated with the optimal estimator. Using this fact, we can evaluate the m.m.s.e. as follows:

$$\begin{aligned}
\mathsf{E}\tilde{x}^2 &= \mathsf{E}[x-\hat{x}][x-\hat{x}] \\
&= \mathsf{E}[x-\hat{x}]x \qquad \text{(because of (1.3.6))} \\
&= \mathsf{E}x^2 \; - \; \mathsf{E}\hat{x}x \\
&= \mathsf{E}x^2 \; - \; \mathsf{E}\hat{x}[\tilde{x}+\hat{x}] \\
&= \mathsf{E}x^2 \; - \; \mathsf{E}\hat{x}^2 \qquad \text{(because of (1.3.6))} \\
&= \left(\mathsf{E}x^2 \; - \; \bar{x}^2\right) \; + \; \left(\bar{x}^2 \; - \; \mathsf{E}\hat{x}^2\right) \\
&= \sigma_x^2 - \sigma_{\hat{x}}^2
\end{aligned}$$

◇

Theorem 1.3.1 tells us that the least-mean-squares estimator of x is its conditional expectation given y. This result is again intuitive. In answer to the question: what is the best guess for x given that we observed y?, the analysis tells us that the best guess is what we would expect for x given the occurrence of y!