

Appendix G

Statistical Errors

Geoffrey N. Smit, *The Aerospace Corporation*

Reproduced with permission from *Orbit and Constellation Design and Management* [Wertz, 2009], Appendix C.

This appendix is an analytic introduction to error analysis. The flow of topics is summarized in Fig. Gweb-1. Section 6.3 provides an introduction to error analysis and a general recipe for adding errors in an error budget. The process of creating an error budget is introduced and discussed in detail with tables of error sources in Sec. 6.3. Chapter 8 provides an introduction to spacecraft position and attitude measurements. Specifically, Sec. 7.2 in OCDM by Wertz [2009] provides a detailed assessment of the evaluation of measurement uncertainty on the celestial sphere and Sec. 7.6 in OCDM describes good and bad measurement sets (in terms of uncertainty) and a set of practical tests to determine them.

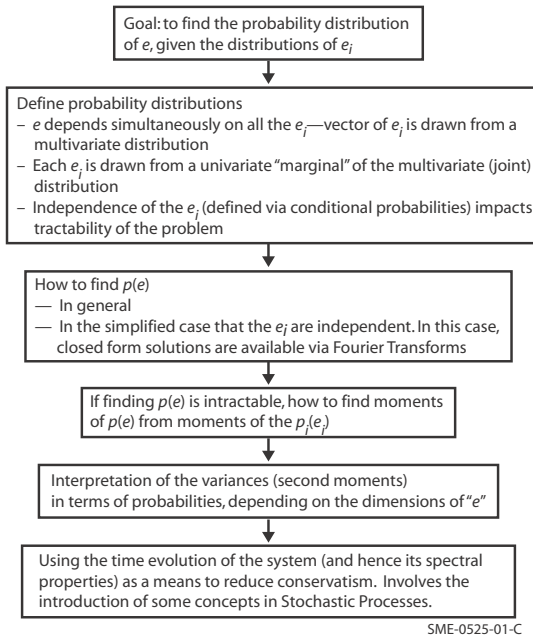


Fig. Gweb-1. Sequence of Topics in this Appendix.

G.1 Probability Considerations

The basic problem to be addressed is:

Given: a set of components with errors e_1, e_2, \dots with known probability distributions. The components are combined in a system such that the output error is

$$e = e_1 + e_2 + \dots + e_n \quad (\text{Gweb-1})$$

Find: the probability distribution of e .

As noted in Sec. 6.3, the more general problem is

$$e = F(e_1, e_2, \dots, e_n) \quad (\text{Gweb-2})$$

F may, in general, depend on the state of the system—i.e., if the component contributions are

$$x_i = d_i + e_i, \quad (i = 1, \dots, n) \quad (\text{Gweb-3})$$

then the system output is some function, $G(x)$, which can be decomposed into the nominal output, $G(d)$, and the error,

$$G(x) = G(d) + F(x) \quad (\text{Gweb-4})$$

G can usually be expanded in a Taylor series. When this is the case, we have

$$G(x) = G(d) + eG'(x) + \text{higher order terms} \quad (\text{Gweb-5})$$

This gives the system error,

$$F = G(x) - G(d) \approx eG'(x) \quad (\text{Gweb-6})$$

which is a weighted sum of the e_i , with coefficients which are functions of the state, x .

A weighted sum can always be scaled to be a pure sum. Therefore we will concentrate on the case of a pure sum, since we have seen that it covers the majority of cases of interest. Also, it allows us to derive a number of concrete results which would be less understandable in a more general setting.

If, for some reason, it is desired to retain a functional form other than a sum, this does not impact the conceptual basis for much that follows. For example, a worst case estimate of the system error can readily be produced (and it typically will be too conservative), and computational schemes can be devised to estimate the probability distribution of e or quantities related to this probability distribution.

Note that e is a function of several random variables. To proceed, we first summarize some of the concepts and definitions of probability distributions involving several variables in order to fix notation and terminology.

Consider first probability on the real line (i.e., we consider “events” labelled by real numbers—such as a measurement, x). The probability that $x \leq y$, $\text{Prob}(\{x \leq y\})$, is denoted $P(y)$. $P(y)$ is defined to be the *probability distribution*. Assuming that the derivative exists

$$dP / dy \equiv p(y) \quad (\text{Gweb-7})$$

is called the *probability density function*, or *pdf*. $p(y) dy$ is the probability that x lies between y and $y + dy$ and

$$\int_{-\infty}^{\infty} p(x) dx = 1 \quad (\text{Gweb-8})$$

In the definition of the probability distribution, there are two variables (I've called them x and y). The variable y is the actual argument of the probability distribution. x is a "dummy variable", introduced for explanatory reasons. Thus, $P(y) = \text{Prob}\{x \leq y\}$, means that any "experiment" performed on the system described by P will yield an outcome less than or equal to y on P fraction of the trials. Thus, as shown in Fig. Gweb-2, P is monotonically increasing. In the case of the probability density, we are focussing attention on the neighborhood of a single point, " y ". It is not uncommon to see terminology like "the probability that y lies between y and $y + dy$ is given by $p(y)dy$ ".

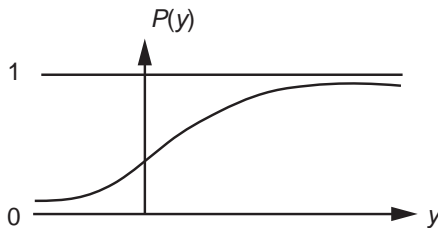


Fig. Gweb-2. Probability Distribution $P(y)$ vs. y .

The variable, x , used to label points or "events" in our probability sample space is called a *random variable* or *variate*. I use the terms "variable," "random variable," and "variate" interchangeably.* The sample space need not be the real line. It could be n -dimensional space (i.e., n copies of the real line), or, in general, any suitable set.

Measurements of angles, distances, brightnesses, or frequencies are all random variables. The result of any particular measurement would be a single real number, the value of which includes some element of chance. The associated probability densities would usually be peaked around the true value of the item being measured, and would be negligible far away from the true value as shown in Fig. Gweb-3.

Examples of multidimensional random variables include measurements of "physical" vectors such as the 3-dimensional velocity vector of an object and the arguments of functions of several 1-dimensional random variables. For example, to determine the area of a rectangle, we have to measure two lengths. This input "vector" would be a 2-dimensional random variable. The pointing errors of a spacecraft typically depend on the value of numerous parameters (such as sensor readings, tempera-

* Intuitively, the idea of a random variable is fairly clear—there is an element of chance in any given "reading" of x , i.e., x is drawn from some probability space ("black bag") at random, and the likelihood that it will have any particular value is given by its probability density. The mathematical theory was originally developed to describe various games of chance. The subject only became mathematically respectable after its foundations had been worked out independent of references to these intuitive notions. A rapid increase in the rate of progress occurred at this point. (See Kac [1989].)

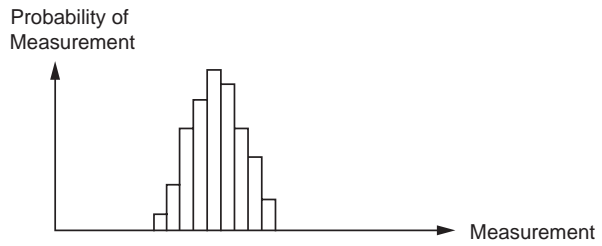


Fig. Gweb-3. A Discrete Probability Density Function.

tures, structural deformations) and this input vector would be a multidimensional random variable. Random variables can also be *discrete*, i.e., limited to a finite number of values. Examples are the outputs of coin or die tossings. The output from a digital sensor is discrete.

For a multi-dimensional random variable, the probability density has the form

$$p(x_1, x_2, \dots, x_n) \tag{Gweb-9}$$

This gives the probability that x will be found in the hypercube

$$\{(x_1, x_1 + dx_1), (x_2, x_2 + dx_2), \dots, (x_n + dx_n)\} \tag{Gweb-10}$$

The multi-dimensional probability distribution is defined as

$$P(y) = \text{Prob}\{x_1 \leq y_1, \dots, x_n \leq y_n\} \tag{Gweb-11}$$

and we have $p(x)$

$$p(x) = \frac{\partial^n P}{\partial x_1 \partial x_2 \dots \partial x_n} \tag{Gweb-12}$$

The system errors that we are trying to estimate depend on the input errors contributed by numerous components, i.e., the system error depends on a multi-variable probability distribution. In general, we don't have any data on this multivariable distribution, per se. What we have is data on the distribution of each of the inputs. We may also have some information on whether some of these inputs interact with each other so that the value measured by one input may be coordinated with that measured by another. We are thus faced with the task of building up the multivariable distribution from its "projection" on each of the input axes. This brings us to the topic of such projections (marginal distributions), and the topic of interdependence between the inputs (conditional probabilities).

Marginal Distributions

One can "integrate out" some of the variables of a multivariate distribution function, leaving a probability distribution in the remaining variates.†

For example,

$$\int_{-\infty}^{\infty} p(x_1, \dots, x_n) dx_2 dx_3 \dots dx_n \tag{Gweb-13}$$

leaves a density

$$p_1(x_1) = \frac{dP_1(x_1)}{dx_1} \tag{Gweb-14}$$

P_1 and p_1 are called the *marginal distribution function* and *marginal density* of x_1 associated with $p(x_1, \dots, x_n)$ and $P(x_1, \dots, x_n)$. Marginal distributions of x_2, x_3, \dots can be similarly defined. The figure below illustrates the marginals of a bivariate probability density. (The geometry of the figure also illustrates why they are called marginals.)

The multivariate density, p , is called the *joint density* associated with these variates. Similarly we can define the *joint distribution*.

While a given multivariate distribution function (or density) will have unique marginals, the inverse problem of finding a joint distribution given the marginals does not have a unique solution. In the error analysis problem being addressed in this section, the information that we have about the joint distribution of the errors, e_i , comes

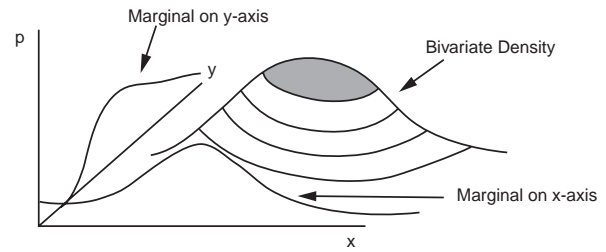


Fig. Gweb-4. Joint (Bivariate) Probability Density Showing Marginals on Each of the Axes.

from its marginals, the p_i , and the correlations between the e_i .^{*} In addition to the uniqueness issue mentioned above, for some marginal distributions not all correlation coefficients (between -1 and $+1$) are possible, so that we must also observe bounds on the correlations we input to the problem. Assuming that we have posed the problem properly, we can in principle generate a joint distribution (e.g., via a Monte Carlo simulation) consistent with the marginals. Table Gweb-1 lists a number of common probability distributions used in statistics.

^{*} The word “correlation” has a technical meaning which will be defined later. It refers to the degree to which the values of random variables are interrelated.

[†] The significance of the integration is as follows. For simplicity consider the two variable case. The multivariate probability distribution function, $p(x_1, x_2)dx_1dx_2$ is the probability that x_1 and x_2 will lie in $\{(x_1, x_1 + dx_1), (x_2, x_2 + dx_2)\}$. $\int p(x_1, x_2)dx_2$ is the probability that x_1 and x_2 will lie in the strip $\{(x_1, x_1 + dx_1), \text{any } x_2\}$, i.e., we’ll “accept” any measurement which has x_1 in $(x_1, x_1 + dx_1)$ regardless of the x_2 value.

Table Gweb-1. Some Common Probability Distributions.

Name, Comments	Probability Density Function	Graph	Mean, M Variance, V	Characteristic Function
Uniform Distribution Arises during quantization	$p(x) = 1/(b - a), a < x < b$ $= 0$, otherwise		$M = (a+b)/2$ $V = (b - a)^2/12$	$(\exp(sb) - \exp(sa))/(b - a)s$
Normal Ubiquitous: arises via sums, also as the limit of some distributions	$(1/\sqrt{2\pi s^2})e^{-(x-a)^2/2s^2}$		$M = a$ $V = s^2$	$\exp(at + s^2/2)$
Rayleigh Distribution of radial error with normal components	$(x/s^2)e^{-x^2/2s^2}$ $x > 0$ (zero for $x < 0$)		$M = s\sqrt{\pi/2}$ $V = ((4 - \pi)/2)s^2$	Complicated
Cauchy Ratio of two normal variates. Also, geometric meaning (tan of a uniformly distributed variate)	$a/\left[\pi\left\{(x-b)^2 + a^2\right\}\right]$		$M = b$ Variance diverges	$\exp(-bs - a s)$
Poisson Prob. of getting k of independent events (discrete)—model of shot noise, etc.	$\lambda^k e^{-\lambda}/k!$ (prob of a negative number of occurrences = 0)		$M = \lambda$ $V = \lambda$	$\exp[\lambda\{\exp(s) - 1\}]$
Binomial* Prob. of getting k wins in m trials, each with prob of success, $p = 1 - q$	$C_m^k p^k q^{m-k}$		$M = mp$ $V = mpq$	$[1 + p\{\exp(s) - 1\}]^m$

*The Bernoulli distribution is a binomial distribution with $m = 1$

Conditional Probabilities and Independence

Random variables can be coupled, so that the values taken on by one may give us information about the others. For example, a spacecraft structural distortion may be related to a temperature change. The probability that an event A occurs, given that event B has occurred, is written $P(A|B)$ and said “probability of A , given B .” From the definition,

$$P(A,B) = P(A|B)P(B) \tag{Gweb-15}$$

i.e., the probability of A and B is the probability of B times the probability of A given B .

The events A and B are called *independent* if

$$P(A|B) = P(A) \tag{Gweb-16}$$

which is the same as

$$P(A,B) = P(A)P(B) \tag{Gweb-17}$$

In this case, the densities can also be written as products, e.g.,*

$$p(x_1, x_2) = p(x_1)p(x_2) \tag{Gweb-18}$$

Later, we will introduce the moments of probability densities. These provide simple measures of various attributes of the density function. The “correlation” will be a moment quantifying the extent to which the variates in a multivariate density are independent.

Functions of Random Variables

Given a random variable, x , with density $p(x)$ or a set of random variables, x_1, x_2, \dots, x_n , with density $p(x_1, x_2, \dots, x_n)$, we frequently need to determine the density corresponding to some function of x or of x_1, x_2, \dots, x_n . This is the central issue of the error combination problem.

First look at the one dimensional case. Let

$$y = y(x) \tag{Gweb-19}$$

If y is a monotonic function, we have

$$P_y(y) = P_x[x(y)] \tag{Gweb-20}$$

and

$$p_y(y) |dy| = p_x[x(y)] |dx| \tag{Gweb-21}$$

i.e.,

$$p_y(y) = p[x(y)] |dx/dy| \tag{Gweb-22}$$

For example, if x has density function $p(x)$, and

$$y = 3x + 1 \tag{Gweb-23}$$

then y has density function

$$q(y) = p((y - 1)/3)/3 \tag{Gweb-24}$$

As a second example, if x has density function $p(x)$,

* Independent variates have many mathematical similarities to orthogonal coordinate systems. In particular, they tend to make the algebra much simpler than coupled variates and non-orthogonal coordinates, respectively.

and $y = x^2$, then y has the density function

$$q(y) = \left[p(\sqrt{y}) + p(-\sqrt{y}) \right] / \left[2\sqrt{y} \right] \tag{Gweb-25}$$

The pairing of terms here derives from the fact that y is not monotonic.

Similarly, in the multivariate case, if

$$x = (x_1, \dots, x_n) \tag{Gweb-26}$$

and

$$y = (y_1, \dots, y_n) \tag{Gweb-27}$$

are related by a one-to-one transformation, then

$$P_y(y) = P_x[x(y)] \tag{Gweb-28}$$

and

$$p_y(y) = p_x[x(y)] |\partial x / \partial y| \tag{Gweb-29}$$

where $|\partial x / \partial y|$ is the Jacobian. The basic idea is that we solve for x in terms of y and substitute this into the argument for the density (in the obvious way) to yield a function of y , and we also rescale the function so that the integral of the density is one.

The process becomes more complicated if we map onto a space of fewer dimensions, e.g.,

$$y = (y_1, \dots, y_m) \tag{Gweb-30}$$

and

$$x = (x_1, \dots, x_n), m < n. \tag{Gweb-31}$$

A special case of this is the sum, $y = x_1 + \dots + x_n$. We are taking a generalized sort of marginal, and we need to integrate out some of the variables. The problem is to determine what needs to be integrated out. This is usually more easily done in the case of the probability distribution, as opposed to the density. For example, assume we have:

$$y_i = y_i(x_1, \dots, x_n), i = 1, \dots, m \tag{Gweb-32}$$

The marginal distribution, $P_{y_1}(y_1)$ is, by definition, the probability that “ $y_1 \leq Y_1$ ” as shown in Fig. Gweb-5, for 2-dimensions. In “ y - space”, the set of all appropriate y 's is the shaded area.

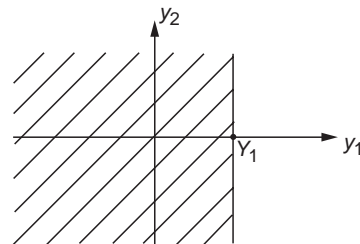


Fig. Gweb-5. Shaded Region Represents the Set of Points for Which the y_1 Component is Less than or Equal to Y_1 .

To find $P_{y_1}(y_1)$ we need to associate a probability density with each part of the shaded region, and then integrate (see Fig. Gweb-6). Since we only know the density function in the x -space, we must find the corresponding region there, and integrate there. Thus, for the distribution, $P_{y_1}(y_1)$ = probability that $y_1 \leq Y_1$, we need to integrate over all combinations of the x 's which yield a $y_1 \leq Y_1$:

$$P_{y_1}(y_1) = \int \dots \int_{y_1(x_1, \dots, x_n) \leq Y_1} p_x(x_1, \dots, x_n) dx_1 \dots dx_n \tag{Gweb-33}$$

(similarly for y_2, y_3, \dots). The corresponding joint probabilities are

$$P_{y_1 \dots y_k}(Y_1 \dots Y_k) = \int \dots \int_{\substack{y_i(x_1, \dots, x_n) \leq Y_i \\ y_k(x_1, \dots, x_n) \leq Y_k}} p_x(x_1, \dots, x_n) dx_1 \dots dx_n \tag{Gweb-34}$$

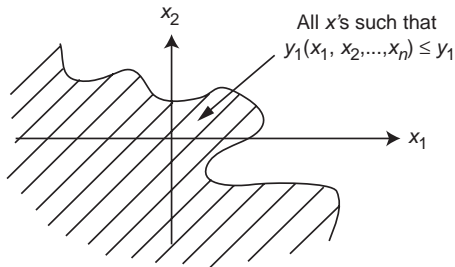


Fig. Gweb-6. Shaded Region is the Set of Points for which $y_1(x_1 \dots x_n) \leq Y_1$.

Depending on the nature of the function y , this can take some effort. Similarly, the densities can be found by either (1) differentiating the P 's, or (2) integrating the p_x 's over suitable subsets of x space. The next section illustrates this for the special case where $y = x_1 + x_2 + \dots + x_n$.

G.2 Addition of Random Variables

Here we have a single function, $y = y(x_1, x_2, \dots, x_n)$ given by

$$y = x_1 + x_2 + \dots + x_n. \tag{Gweb-35}$$

We shall first reproduce the above analysis for this special case. It turns out that when the variables are independent, the analysis can be significantly simplified—in fact, all the standard “textbook” results for the addition of random variables are for independent random variables. We will point out below how these simplifications occur.

For simplicity, first consider the case of two variables. Let the joint distribution of x_1 and x_2 be $P_x(x_1, x_2)$ and the density be $p_x(x_1, x_2)$. As shown graphically in Fig. Gweb-7, the probability that $x_1 + x_2 \leq Y$ say is given by the integral of p_x over the shaded region, i.e., over all points x_1, x_2 such that $x_1 + x_2 \leq Y$.

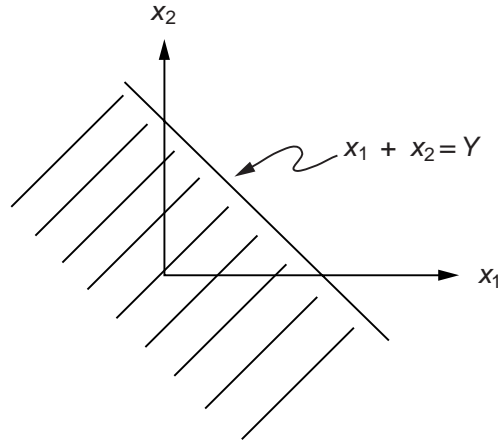


Fig. Gweb-7. Shaded Region is the Set for Which $x_1 + x_2 \leq Y$.

Hence,

$$P_y(Y) = \iint_{x_1 + x_2 \leq y} p_x(x_1, x_2) dx_1 dx_2 \tag{Gweb-36}$$

The density is given by:

$$p_y(Y) = \iint dx_1 dx_2 p_x(x_1, x_2) \delta(Y - x_1 - x_2) = \int p(x_1, Y - x_1) dx_1 \tag{Gweb-37}$$

This type of analysis is easily extended to more than two variables. In general, the formulas get very complicated unless the variables are independent. The main exception is if P_x is multivariate Gaussian. The underlying reason for the normal distribution retaining its tractability, even in the presence of correlations, is the fact that we can always transform to a new set of variates which are independent.

In the case where the x_i are independent, the analysis can be simplified as follows. Starting again with the two variate case, we have

$$p(x_1, x_2) = p_1(x_1) p_2(x_2) \tag{Gweb-38}$$

and

$$y = x_1 + x_2 \tag{Gweb-39}$$

$$P_y(y) = \int p_x(x_1, y - x_1) dx_1 = \int p_1(x_1) p_2(y - x_1) dx_1 \tag{Gweb-40}$$

This is suggestive of the convolution of p_1 and p_2 , and invites the use of Fourier transforms. In probability theory, the Fourier transform of a density is called its *characteristic function*.*

The Fourier Transform of a convolution is the product of the Fourier Transforms.

$$\begin{aligned} FT[p_y(y)] &= FT\left[\int p_1(x_1)p_2(y-x_1)dx_1\right] \\ &= \int e^{i(q+x_1)s} p_1(x_1)p_2(q)dx_1dq \\ &= FT(p_1(x))FT(p_2(x)) \\ &= G_{x_1}(s)G_{x_2}(s) = G_y(s) \end{aligned} \tag{Gweb-41}$$

Hence the distribution of y can be found by inverting the Fourier Transform. This is easily extended to n independent variables—we simply get more terms in the product

$$G = G_1G_2G_3\dots G_n \tag{Gweb-42}$$

Examples

- (1) The sum of n normal independent random variables is a normal random variable:

$$x = x_1 + x_2 + \dots + x_n \tag{Gweb-43}$$

$$\begin{aligned} p(x_i) &= \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\left[\frac{x_i - \xi_i}{\sigma_i}\right]^2\right) \\ \Rightarrow p(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{x - \xi}{\sigma}\right]^2\right) \end{aligned} \tag{Gweb-44}$$

where

$$\begin{aligned} \xi &= \xi_1 + \dots + \xi_n \\ \sigma^2 &= \sigma_1^2 + \dots + \sigma_n^2 \end{aligned} \tag{Gweb-45}$$

The sum of n statistically independent random variables is normal if and only if each of the variables is normal.

- (2) The binomial, Poisson and Cauchy distributions also “reproduce themselves” under addition of independent variates:

If

$$p_i(x_i) = \binom{n_i}{x_i} q^{x_i} (1-q)^{n_i-x_i} \tag{Gweb-46}$$

* Characteristic functions or Fourier Transforms are ubiquitous in probability theory and statistics. In fact it has been said that knowledge of statistics, like Electrical Engineering, reduces to a knowledge of the Fourier Transform.

Then

$$p(x) = \binom{N}{x} q^x (1-q)^{N-x}, \quad N = n_1 + \dots + n_n \tag{Gweb-47}$$

If

$$p_i(x_i) = e^{-\xi_i} \frac{\xi_i^{x_i}}{x_i!} \tag{Gweb-48}$$

Then

$$p(x) = e^{-\xi} \frac{\xi^x}{x!}, \quad \xi = \xi_1 + \dots + \xi_n \tag{Gweb-49}$$

If

$$p_i(x_i) = \frac{1}{\pi\alpha} \frac{1}{1 + \left(\frac{x_i - \xi_i}{\alpha}\right)^2} \tag{Gweb-50}$$

Then

$$p(x) = \frac{1}{\pi\alpha} \frac{1}{1 + \left(\frac{x - \xi}{\alpha}\right)^2} \tag{Gweb-51}$$

- (3) If we add an arbitrary random variable, x , to a *uniformly* distributed random variable, y , the sum, $z = x + y$ yields the moving average of x , assuming x and y are independent. Let

$$\begin{aligned} p(y) &= \frac{1}{2a} \quad -a \leq y < a \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{Gweb-52}$$

Then

$$\begin{aligned} P_z(z) &= \int p_x(x)p_y(z-x)dx = \frac{1}{2a} \int_{z-a}^{z+a} p_x(x)dx \\ &= \frac{1}{2a} [P_x(z+a) - P_x(z-a)] \end{aligned} \tag{Gweb-53}$$

In general, the probability distributions of sums $x_1 + x_2 + \dots$, where the x_i come from different distribution types, can yield messy algebra.

Other Functions of Random Variables

We have already presented the general approach to finding the probability distributions of functions of random variables other than sums. We have emphasized sums because they are important, particularly since we have indicated that most error combination problems can be reduced to sums, and also because they are tractable. In a few cases, analytical results can be obtained for other functions, as the following examples show.

To find the distribution of a product of two independent random variables, let the variables be x and y , with joint density $f(x,y) = p(x)q(y)$ since they are independent. We want to find the density, $w(z)$, of $z = xy$. We have

$$w(z) = \int_{-\infty}^{\infty} p(x)q(z/x) \left| \frac{1}{x} \right| dx \quad (\text{Gweb-54})$$

To find the distribution of a quotient of two independent random variables, with $x, y, p(x), q(y)$ as above, and with $z = x/y$ and the density of z being $w(z)$, we have:

$$W(z) = \int_{-\infty}^{\infty} p(yz)q(y)|y|dy \quad (\text{Gweb-55})$$

For p and q normal, W reduces to a Cauchy distribution.

Once we have the probability distribution of the output (i.e., the system performance), we are in a position to estimate the probability that the system error will lie within the required bounds. The generation of this probability cannot be accomplished analytically, except for some special cases, such as independent, identically distributed e_i , with the p_i given by some of the standard types. It will in general require the use of a Monte Carlo simulation plus some insight to ensure that the basic problem is well defined. This can be time consuming. Also, often the input data (e.g., the p_i) are not particularly well known, and it might not make sense to expend tremendous amounts of computational effort analyzing their consequences. For these reasons, it is often of interest to find simplified ways of assessing the system performance. It turns out that the moments of e can be calculated quite simply in many situations as discussed in the next section.

G.3 Expectations and Moments

It is often convenient to find a few numbers which summarize the key features of a probability distribution. For example, consider the following density functions. In Fig. Gweb-8(A), the density is centered about the origin and is fairly spread out. This indicates that “on average” readings of this variable will give zero, but that there will be a lot of scatter. In Fig. Gweb-8(B), we also have a variable centered about the origin, but in this case there will be much less scatter. In Fig. Gweb-8(C), the density is centered about a non-zero value—i.e., on average, we expect these measurements to give a reading displaced from zero. In Fig. Gweb-8(D), the probability density is “skewed”—there is more scatter for positive than for negative values. The basic point is that the behavior of the random variable can be summarized via a few general features of the density—i.e., measures of location, scatter and symmetry. Invoking the analogy of the mass moments of a rigid body, these features are usually quantified by taking the “moments” of the density function.

We start with some definitions. The *expectation* of a function f of a random variable x with probability density $p(x)$ is defined to be:

$$E[f(x)] \equiv \int f(x)p(x)dx \quad (\text{Gweb-56})$$

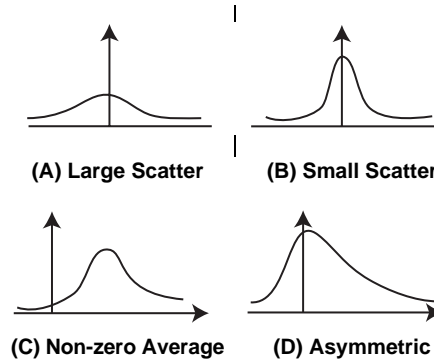


Fig. Gweb-8. Qualitative Features of Distributions.

The n th moment of the probability density $p(x)$ is then defined as

$$m_n(x) \equiv E[x^n] \quad (\text{Gweb-57})$$

The following rules for expectations are easily established:

$$E(c) = c \quad (\text{Gweb-58})$$

$$E(cf) = cE(f) \quad (\text{Gweb-59})$$

$$E(f_1 + f_2) = E(f_1) + E(f_2) \quad (\text{Gweb-60})$$

If

$$f_1(x) \leq f_2(x) \quad (\text{Gweb-61})$$

for all x then,

$$E(f_1) \leq E(f_2) \quad (\text{Gweb-62})$$

$$|E(f(x))| \leq E(|f(x)|) \quad (\text{Gweb-63})$$

These results are completely independent of the probability distributions of the random variables.

The first moment, μ , given by

$$\mu \equiv \int xp(x)dx \quad (\text{Gweb-64})$$

corresponds to the “center of mass” of the probability density. It is called the *mean* or *expected value* of the random variable, x . The second moment is called the variance

$$\text{Var}(x) \equiv \int (x - \mu)^2 p(x)dx \quad (\text{Gweb-65})$$

The variance is the square of the standard deviation, σ . The variance corresponds to the moment of inertia of p , and the standard deviation to the radius of gyration. They give a measure of how spread out p is, and hence how scattered the observations of the random variable, x , will be. Specifications of error bounds are often given in terms of standard deviations, i.e., “ n times the standard deviation should not exceed a certain value,” where n is typically 1, 2 or 3. Implicit in such a specification is a probability figure. For a given density, p , the probability

that a measurement lies within n standard deviations of the mean is well defined. For example, for a normal distribution, these values are 68.3%, 95.5%, and 99.7% respectively as given in Table 7-3 in Sec. 7.2.2.2 in OCDM by Wertz [2009]. We discuss later the fact that these probabilities will change if a different distribution is used.

We now return to the problem of characterizing the probability distribution of the system error, given those of the components. We can use the above rules to calculate some moments of sums:

$$\begin{aligned} E\left[\sum x_i\right] &= \sum E(x_i) \\ E\left[\sum x_i^2\right] &= \sum \left(Ex_i^2\right) + 2\sum_{i>j} E(x_i x_j) \end{aligned} \quad (\text{Gweb-66})$$

Hence,

$$\text{Var}\left[\sum x_i\right] = \sum \left(\text{Var}(x_i)\right) + 2\sum_{i>j} \text{Cov}(x_i x_j) \quad (\text{Gweb-67})$$

If the x_i are independent, the covariances are zero, and we have

$$\text{Var}(\sum (x_i)) = \sum (\text{Var}(x_i)) \quad (\text{Gweb-68})$$

Taking square roots, we have

$$\sigma_{x_1+x_2+\dots+x_n} = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2 + \dots + \sigma_n^2} \quad (\text{Gweb-69})$$

If we have perfect correlation, the standard deviations sum:

$$\begin{aligned} \text{Var}(x_1 + x_2) &= \text{Var}(x_1) + \text{Var}(x_2) + 2\sigma_1\sigma_2 \\ &= \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 = (\sigma_1 + \sigma_2)^2 \end{aligned} \quad (\text{Gweb-70})$$

Hence

$$\begin{aligned} \sigma_{x_1+x_2} &= \sigma_{x_1} + \sigma_{x_2} \\ (\text{for } r = -1, \text{ this becomes } |\sigma_1 - \sigma_2|) \end{aligned} \quad (\text{Gweb-71})$$

Similar analyses can be carried out for more variables. Hence, if we know the means and standard deviations (or variances) of our inputs, it is easy to calculate the mean and standard deviation of the output, particularly if the inputs are either independent or perfectly correlated.

When the variables are uncorrelated, the system standard deviation is the square root of the sum of the squares, called the *root sum square* or *RSS*, of the individual standard deviations. When the variables are correlated, the system error starts to take on more of the appearance of a sum of the errors of the input variables. The interpretation is that the statistical “smoothing” no longer takes place, and the variables add algebraically. Between these two extremes lie systems with partial correlations. The results also lie in between, and the algebra becomes messier.

If the output is a weighted sum of the individual inputs, we can perform a similar analysis. For uncorrelated inputs, we get a weighted sum of squares for the variance, and for perfect correlation, we get a weighted sum.

Note that in this case we can always rescale the inputs so that the weights are all equal to one—i.e., the weights really introduce no new concepts or complications.

These results form the basis for some powerful ideas in probability theory and its applications:

(1) Law of Large Numbers

Take the average of a large number of samples,

$$\mu_n = (x_1 + x_2 + \dots + x_n)/n \quad (\text{Gweb-72})$$

then

$$\text{Var}(\mu_n) = \text{Var}(x)/n \quad (\text{Gweb-73})$$

Thus, the variance of the sample means is $(1/n)$ times the variance of the population. By taking large enough samples we can cluster as close to the mean as desired. However, there are some bounds on “how close”. For example, *Chebyshev's inequality* says that the probability that we lie within $c\sigma^2/\sqrt{n}$ of the mean is greater than $1 - 1/c^2$, for any c . This idea underlies a number of filtering approaches.

(2) Central Limit Theorem

If the x_i are independent and identically distributed, each with mean μ , and variance σ^2 , then

$$z_n = (x_1 + x_2 + \dots + x_n)/\sqrt{n\sigma^2} \quad (\text{Gweb-74})$$

is normally distributed as $n \rightarrow \infty$ **regardless of the distributions of the x_i** . This is used, for example, to justify the use of normal distributions in the theory of errors. The basic idea is that errors of measurement in most physical systems are due to a large number of small influences, so that one can argue that their combination is normally distributed. Experimentally, this is borne out in a wide variety of situations. It is somewhat of a miracle, because the normal distribution also happens to be extremely convenient mathematically. A very large portion of all that has been written on statistics assumes that samples are drawn from normally distributed populations.

(3) Stable Distributions

We have already noted that the binomial, Cauchy, Poisson and normal distributions are self-reproducing under summation (and hence averaging)—i.e., sums of such variates are also so distributed. This concept can be generalized to that of a *stable* distribution.

Let x, x_1, \dots, x_n be independent identically distributed random variables with density $p(x_i)$. Let

$$s_n \equiv x_1 + x_2 + \dots + x_n \quad (\text{Gweb-75})$$

The density $p(x)$ is *stable* if there exist constants $c_n > 0$ and γ_n (for all $n \geq 2$) such that s_n has the same density as $c_n x + \gamma_n$. If $\gamma_n = 0$, $p(x)$ is called *strictly stable*.

For the normal distribution, $c_n = n^{1/2}$. For the Cauchy, $c_n = n$. One can show [Feller, 1957, 1966]: that c_n can only be of the form

$$n^{1/\alpha}, \quad 0 < \alpha \leq 2 \quad (\text{Gweb-76})$$

where $\alpha = 1$ for Cauchy, 2 for Normal. The distribution

$$p(x) = \begin{cases} \frac{1}{\sqrt{2\pi x^3}} e^{-1/2x} & (x > 0) \\ 0 & (x < 0) \end{cases} \quad (\text{Gweb-77})$$

is stable with $\alpha = 1/2$. The stable densities with $\alpha < 2$ do not have variances, i.e., their second moments are unbounded.

(4) Computing Higher Moments

The calculation of the higher moments of sums of independent random variables is usually accomplished using the characteristic function. The log of a characteristic function is easy to differentiate. We define the *n*th cumulant of a random variable, x , as

$$K_n[x] \equiv \frac{1}{i^n} \frac{d^n}{du^n} \log G_x(u) \quad u = 0 \quad (\text{Gweb-78})$$

The cumulants of independent random variables are additive, i.e.,

$$K_r[x_1 + x_2 + \dots + x_n] = K_r[x_1] + K_r[x_2] + \dots + K_r[x_n] \quad (\text{Gweb-79})$$

Even if the Fourier Transform of the sum cannot be inverted, we can evaluate the cumulants and express the moments in terms of them. In fact, the first three moments about the mean equal the first three cumulants. This also

shows that the first three moments are additive for independent summands.

If we take the Laplace transform of a density function (analogously to taking the Fourier transform to form the *characteristic function*) the result is called the *moment generating function (MGF)*. The reason for this is that the moments turn out to be the coefficients in the Taylor series for the MGF. Let $G(s)$ be the MGF of the probability density $p(x)$. Then

$$G(s) = \int \left[e^{sx} p(x) dx \right] = E \left[e^{sx} \right] \quad (\text{Gweb-80})$$

Now note that

$$\begin{aligned} G'(s) &= E \left[x e^{sx} \right], \\ G''(s) &= E \left[x^2 e^{sx} \right], \dots, G^{(n)}(s) = \left[x^n e^{sx} \right] \end{aligned} \quad (\text{Gweb-81})$$

Hence

$$\begin{aligned} G'(0) &= E[x], \\ G''(0) &= E[x^2], \dots, G^{(n)}(0) = \left[x^n \right] \end{aligned} \quad (\text{Gweb-82})$$

Finally note that

$$\begin{aligned} G(s) &= 1 + G'(0)s + G''(0)s^2 / 2! + \dots \\ &= 1 + E[x]s + E[x^2]s^2 / 2! + \dots \end{aligned} \quad (\text{Gweb-83})$$

Two and Three Dimensional Target Spaces

We have seen that if the target location must be specified as a 2 or 3 dimensional vector, we can readily calculate the standard deviation of the error in terms of the standard deviations of the 1-dimensional components. We now discuss the significance of the fact that this system error will have a different probability density from those of the components. To illustrate the ideas involved, we shall present the simplest possible case—a bivariate normal distribution—and then briefly indicate what happens in general.

Suppose the target is nominally at (0,0) in the (x, y) plane, and that the x and y error densities are given by normal densities:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/2\sigma_x^2} \quad (\text{Gweb-84})$$

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-y^2/2\sigma_y^2} \quad (\text{Gweb-85})$$

Suppose, further, that these are independent, and that $\sigma_x = \sigma_y = \sigma$. Then the probability of making a measurement in an element of area, $dxdy$ is given by

$$\begin{aligned} p(x, y) dxdy &= p(x)p(y) dxdy \\ &= \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} dxdy \end{aligned} \quad (\text{Gweb-86})$$

We can convert to polar coordinates (r, θ) and integrate out the θ -dependence, leaving the distribution of radial errors:

$$p(r) dr = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} dr \quad (\text{Gweb-87})$$

This is called a *circular normal* or *Rayleigh* probability density. It is usually tabulated in terms of the *circular error probability*, or *CEP*.

Assume that the cumulative distribution function, $P\{r \leq R\}$ is given by

$$\int_0^R p(r) dr = 1 - e^{-R^2/2\sigma^2} \quad (\text{Gweb-88})$$

The *CEP* is defined as the R such that $P(R) = 0.5$. Hence

$$CEP = \sigma \sqrt{2 \ln 2} = 1.177 \sigma \quad (\text{Gweb-89})$$

We have also seen that the standard deviation of r is

$$\sigma_r = \sqrt{2}\sigma \tag{Gweb-90}$$

Let us now calculate the probability that

$$r \leq \sigma_r = \sqrt{2}\sigma \tag{Gweb-91}$$

We have

$$P(\sqrt{2}\sigma) = 1 - e^{-1} = 0.6321 \tag{Gweb-92}$$

Similarly the “ 2σ ” probability is $1 - e^{-4} = .9817$, and the “ 3σ ” probability is 0.9998. Numerical values for 1, 2, 3, and 4σ variations in 1, 2, and 3-dimensional problems are given in Table 7-3 in Sec. 7.2.2 in OCDM by Wertz [2009]. The large print is that the probabilities associated with n -sigma errors vary with dimension and with the probability distributions. Note that in two dimensions there is an extra “ r ” factor in the area element. For small r we enclose fewer points, while at large r we enclose more. In three dimensions we get a similar effect, except that now the extra factor involves r^2 . **The main message of this discussion is that the probabilistic significance of n -sigma error bounds changes with the dimensionality of the problem and with the assumed probability distributions.**

Adding Errors of Different Frequencies

Any measurement on a system can be regarded as drawing a sample from an *ensemble* of systems, a population with the statistical properties we have derived. Up to now we have not said anything about how these samples are drawn. We now consider to what extent the sampling procedure can affect the errors. In an actual system, this sampling procedure is realized as a time series of measurements. Apart from providing us with a string of readings (and hence the opportunity to do statistics) a time series has extra structure. Readings taken close together in time can be more closely related than those taken far apart. As a result, the probability distributions of the output will depend on the time scales used in taking the readings. Since time domain and frequency domain behavior can be related through the Fourier transform, the spectral structure of the errors can influence the way in which we combine them.

From a systems engineering point of view, one of our goals in error budgeting is to eliminate excessive conservatism. We don’t want to impose unnecessarily strict requirements on the subsystems, since this drives up cost. The “extra degree of freedom” obtained by considering frequency can sometimes be exploited to reduce conservatism. The basic idea is that at any given time only a limited fraction of the total error variance may actually be accessible. This is illustrated schematically in Fig. Gweb-9. We may think of the shaded area (the short term available error variation) as wandering around the total area. At small time scales, our error variance is defined by the shaded area. At long time scales, we can observe errors anywhere in the larger, unshaded area. For this to make sense, there must be two time scales in the problem—a short time scale associated with activity within the shaded area, and a longer one associated with the motion of the shaded area. We also require that the motion of the shaded area be classified as “noise”—i.e., not as a deterministic motion which we should have accounted for in the measurement process.

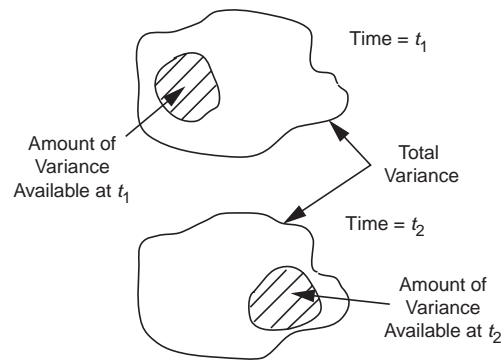


Fig. Gweb-9. The Amount of Variation Available to a System in the Short Term (Shaded Region) May (1) be Smaller than the Total Amount Available in the Long Term (Unshaded Region), and (2) May Move Around with Time.

For a simple illustration, consider adding two pure sinusoids. If the frequencies are far apart, the sum will look like the curve in Fig. Gweb-10.

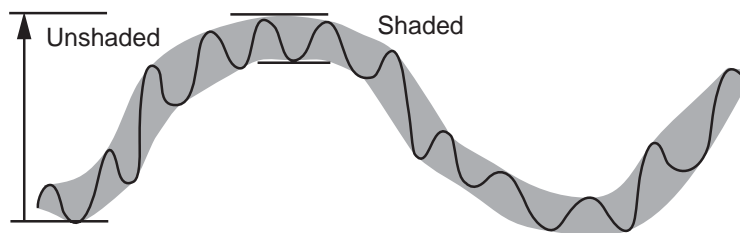


Fig. Gweb-10. Sinusoidal Example of Short Term Variation (Shaded) vs. Long Term Variation (Unshaded).

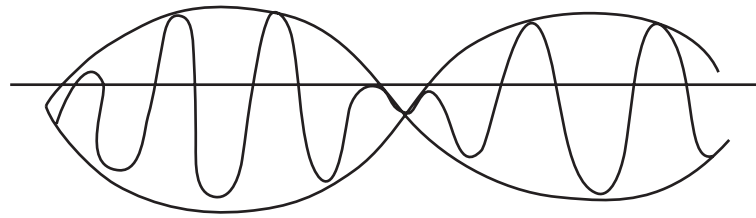


Fig. Gweb-11. Beat Pattern in Mixture of Nearby Frequencies.

Measuring over a long time scale, we'll pick up the total variation. Over a short time scale we'll get a smaller "local" variation, plus a slowly varying bias. However, if the frequencies are close together, we'll get a beat pattern as shown in Fig. Gweb-11. In principle, if the frequencies were very steady (in which case they wouldn't be "noise") we could take advantage of the small amplitudes at the nodes of the beats. In practice, there is a band of frequencies, and the beats don't occur. In both these examples, the total sample space available is large (essentially the total variation of the combined signal). In the first case, however, because of the spread in frequencies, the perceived variance can be less than the total variance if an appropriate time scale is used.

It is possible to have "random analogs" to the sinusoids used in the discussion above—an example would be integrated white noise as it appears on a gyro angle output. Here the bias can shift slowly with time (and can be tracked and eliminated from the error budget), but it is not deterministic.*

Stochastic Processes

If we consider our system to evolve in time but that it is influenced by some random effects, then the state of the system will be given by a time series of random variables. Also, the series of random disturbances, and the measurements with random errors that we use to describe our system will also yield time series of random variables. Such series are called *random processes*. We will show how the frequency content of such processes can be measured, and that processes with various frequency contents can be constructed. We will then give an analog of our sinusoidal example.

A *stochastic process* (stochastic means random) is a family $\{X_t; t \in T\}$ of random variables. For our purposes, the index set, T , will either be a continuous or a discrete time variable. The X_t will typically represent the state of our system. A random process is *strongly stationary* if its probability distributions are invariant under time shifts. It is *weakly stationary* if the first two moments (i.e., the means and covariances) are invariant under time shifts.

For a discrete time variable, a stationary random process is easy to visualize. For each time step, t_i , we draw an X from its probability distribution, $p(X)$, and plot it as shown in Fig. Gweb-12.

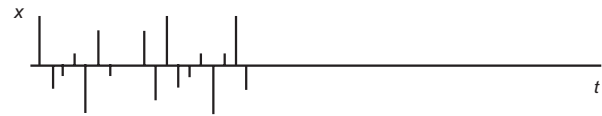


Fig. Gweb-12. A Discrete Time Stationary Random Process, $x = x(t)$. See text for definitions.

If we observe the process for a long time, and count the number of times the ordinate lies between X and $X + dX$ for various X we will find that the probabilities of the various X values match $p(X)$.

Next, we want to see how a frequency spectrum can be associated with such a process. Start in the time domain, and introduce a measure of how much the X 's are correlated with their neighbors. A high positive correlation would imply that X does not change rapidly with time. The *autocorrelation function* of a univariate process is defined as

$$R(t) \equiv E\{X(s-t)X^*(s)\} \tag{Gweb-93}$$

It measures the correlation between readings taken a time interval t apart. Given two processes, $x(t)$ and $y(t)$, their *cross-correlation* is

$$R_{xy}(s) \equiv E\{x(t+s)y^*(t)\} \tag{Gweb-94}$$

We now look at how this temporal behavior translates into the frequency domain. The *power spectral density* of a process $x(t)$ is the Fourier Transform of its autocorrelation:

$$S(\omega) \equiv \int_{-\infty}^{\infty} e^{-i\omega t} R(t) dt \tag{Gweb-95}$$

$$R(t) = \frac{1}{2\pi} \int S(\omega) e^{i\omega t} d\omega \tag{Gweb-96}$$

The *cross-power spectrum*, or *coherence*, of two processes, $x(t)$ and $y(t)$ is

$$S_{xy}(\omega) \equiv \int R_{xy}(t) e^{-i\omega t} dt \tag{Gweb-97}$$

$$R_{xy}(t) = \frac{1}{2\pi} \int S_{xy}(\omega) e^{i\omega t} d\omega \tag{Gweb-98}$$

As the name implies, this measures the power in our signal as a function of frequency, and hence indicates its frequency content.

* There are also a number of deterministic situations in which a slowly varying bias might occur. Typical examples include disturbances that are periodic with the spacecraft orbit, such as thermal or lighting effects.

For example, a theoretical model of *white noise*, $W(t)$, is given by

$$S_{ww}(\omega) = 1, R_{ww}(t) = \delta(t) \quad (\text{Gweb-99})$$

There is no correlation over time. For Gaussian white noise, the probability density of each $W(t)$ is Gaussian.* The total power in white noise is

$$\text{Power} = \frac{1}{2\pi} \int S_{ww}(\omega) d\omega = \infty \quad (\text{Gweb-100})$$

Any real physical system would attenuate the high frequencies. Modifications of white noise by filtering are termed *colored noise*.

As a second example, a *random walk*, also called *Brownian motion* is integrated white noise. Many physical processes can be modeled as linear systems driven by white noise. *Autoregressive models* are an example of this. They have the form

$$\sum a_n x_{t-n} = \omega_t \quad (\text{Gweb-101})$$

Moving average models are another example. They have the form

$$x_t = \sum b_n \omega_{t-n} \quad (\text{Gweb-102})$$

These can be combined to yield *moving average autoregressive* models:

$$\sum a_n x_{t-n} = \sum b_n \omega_{t-n} \quad (\text{Gweb-103})$$

Similar definitions involving stochastic differential equations apply in the continuous case. It should be noted that the “stochastic calculus” involved in the continuous case involves some subtleties. See, for example, Karatzas and Shreve [1997], and Oksendal [1998].

Since many stochastic processes can be built up by passing white noise through various kinds of filters, the inverse problem suggests itself—finding a filter which will reduce a given stochastic process to white noise. Such filters are commonly used in estimation theory—once we get down to white noise, we have essentially squeezed all the information out of the process.

We now return to the “stochastic analog” of the sinusoidal example given previously. We first need to generate a pair of stochastic processes which have very different frequencies, and then discuss how they can be combined. In a real situation, the stochastic processes would be provided by nature. We shall use a pair of mathematically very simple autoregressive processes. (See Box and Jenkins [1970].) Let

$$x_i = ax_{i-1} + u_i \quad (\text{Gweb-104})$$

and

* In the noise literature, “Normal” distributions are usually called “Gaussian”.

$$y_i = -ay_{i-1} + v_i \quad (\text{Gweb-105})$$

be two stationary processes. It can be shown that this requires $|a| < 1$. u and v are supposed to both be white noise with variance σ^2 . As shown in Fig. Gweb-13, heuristically, x tries to match itself, and hence is relatively smooth, with slow meanderings. y reverses sign at each time step, and hence has a much higher frequency appearance. This is confirmed by calculating the power spectra (the Fourier transforms of the autocovariance functions). See Fig. Gweb-14.

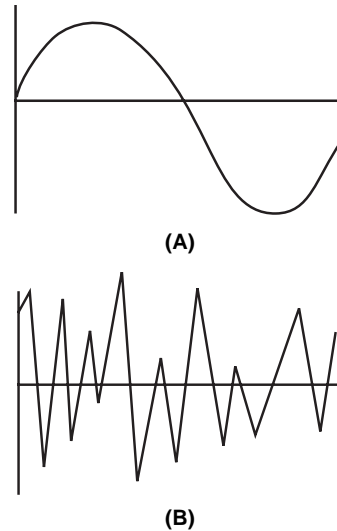


Fig. Gweb-13. (A) Realization of $x_i = ax_{i-1} + u_i$ ($0 < a < 1$). (B) Realization of $y_i = ay_{i-1} + v_i$ ($0 < a < 1$).

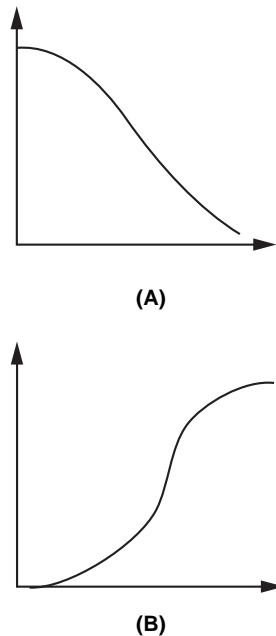


Fig. Gweb-14. (A) Power Spectra of $x_i = ax_{i-1} + u_i$. (B) Power spectra $y_i = ay_{i-1} + v_i$.

If we sum x and y we obtain a process with ostensibly the sum of their variances. This is most easily seen by transforming these autoregressive processes into moving average processes. We have

$$(1 - aB)x_i = u_i \tag{Gweb-106}$$

where B is the back-shift operator,

$Hence$

$$x_i = u_i (1 + aB + [aB]^2 + \dots) \tag{Gweb-108}$$

and

$$\begin{aligned} \text{Var}(x_i) &= \text{Var}(u)(1 + a^2 + a^4 + \dots) \\ &= \sigma^2 / (1 - a^2) \end{aligned} \tag{Gweb-109}$$

Similarly,

$$y_i = v_i (1 - aB + [aB]^2 - \dots) \tag{Gweb-110}$$

and

$$\begin{aligned} \text{Var}(y_i) &= \text{Var}(v)(1 + a^2 + a^4 + \dots) \\ &= \sigma^2 / (1 - a^2) \end{aligned} \tag{Gweb-111}$$

If we call

$$z = x + y \tag{Gweb-112}$$

we have

$$\begin{aligned} z_i &= u_i (1 + aB + [aB]^2 + \dots) \\ &\quad + v_i (1 - aB + [aB]^2 - \dots) \end{aligned} \tag{Gweb-113}$$

and

$$\begin{aligned} \text{Var}(z_i) &= \text{Var}(u) / (1 - a^2) + \text{Var}(v) \\ &\quad / (1 - a^2) = \sigma_{x_i}^2 + \sigma_{y_i}^2 \end{aligned} \tag{Gweb-114}$$

This corresponds to the unshaded area alluded to in the introduction to this section. To see quantitatively that the short term variance is lower we can proceed in various ways. First, we can simply add the two plots, obtaining the results shown in Fig. Gweb-15.

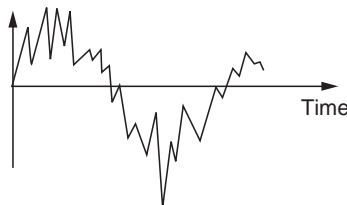


Fig. Gweb-15. Sum of the Two Processes $x_i = ax_{i-1} + u_i$ and $y_i = -ay_{i-1} + v_i$

We essentially get y modulated by x . For short time periods, the local variation is predominantly that of y . Alternatively, we can note that the spectrum at high frequencies is essentially that of y as shown in Fig. Gweb-16.

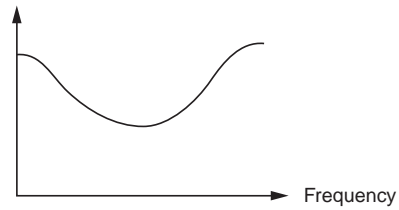


Fig. Gweb-16. Spectrum of Combined Process.

This again implies that the short term fluctuations are mainly those due to y . Finally, we can look at the autocorrelation function, as seen in Fig. Gweb-17, and note that for short time differences, deviations of the autocorrelation function from unity (i.e., “smoothness”) are due to y :

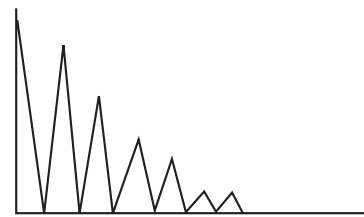


Fig. Gweb-17. Autocorrelation Function for Combined Process.

In summary, the basic point is that we can “achieve” σ_y^2 even though the net variance of the system is $\sigma_x^2 + \sigma_y^2$, over time scales short compared with the main frequency content of “ x ”. Hence, in coming up with an error budget (for short time scales) we can reduce the variance from the “nominal value”.

For example, if we return to Table 5-18 in OCDM by Wertz [2009], we would expect errors (2) and (7) to vary much more slowly than the others. On the shorter time scale of the other errors we would thus expect to be able to estimate these “biases” and correspondingly reduce the system error. The effect of time scale on variance can be conveniently summarized in terms of the Allen Variance $\sigma_A(t)$, which gives the variance in output measurements when those measurements are averaged over a time scale, t . If the output is a sum of inputs with different spectra, these can be identified in a plot of $\sigma_A(t)$. At different time scales, different spectra become dominant. There is an optimum time scale that involves a minimum variance. From the standpoint of error budgeting, we need to identify an appropriate frequency band. This, in turn will yield the corresponding output variance.

G.4 Example: Minimizing Cost

Up to this point, we have not discussed how the error budget is distributed among the various components. We have only shown how a given set of errors contributed by the components can be added. Normally the flowdown of requirements to the components is accomplished by sim-

ply spreading the pain as evenly as possible. In this section we briefly describe how the budget can be optimized in terms of cost.

An error budget should be formulated with an eye toward the sensitivity of the cost of each component to its accuracy requirement. However, cost figures are hard to come by. If an item can be bought off the shelf, there should be no problem. If a new item has to be developed, the risk and cost go up. The more challenging the requirement, the more surprises can arise which tend to add to the cost.

For costing analyses, these distributions are often approximated by triangular distributions as shown in Fig. Gweb-18. As the annotations in the figure indicate, the shape of the triangle will correspond to the amount of risk, and the location to the expected cost. As the requirements on a given object are made more stringent, its triangle will start looking more risky. Fig. Gweb-19 shows a hypothetical family of distributions for a given type of instrument, as the error bounds are tightened. Note that the highest possible cost grows much more rapidly than the “expected cost”, and the corresponding distributions become more and more skewed. The idea is to come up with a family of such curves for each of the components. Having compiled all this information, we now must come up with our error budget—i.e., we must apportion out the total allowable system error to the various components in such a way as to minimize the cost. Conceptually, we want to superimpose curves of constant system cost onto a plot of the error budget constraint (Fig. Gweb-20), and find the lowest cost consistent with the constraint, i.e., the point of tangency. From this we can read off the optimal error budget process as follows:

1. For each component, we can translate the family of cost curves into a curve of accuracy vs cost. We then sum these to obtain curves of system cost. The summation procedure will require care, since it should account for correlations, and, perhaps more important, account for risk. This amounts to deciding which curves to add. A conservative approach would be to use the upper bound curves. This is usually too conservative. Note, for example, that if we push the performance of one item, we may be able to back off on another—i.e., the costs tend to be correlated. We have seen that it is simple to add mean values, and standard deviations. Hence, we can find “mean plus n -sigma” curves for the system cost.
2. Finding the error budget constraint curve, i.e., combinations of component errors which are compatible with the allowed system error, is straightforward assuming we characterize the errors via their standard deviations. For example, if we have n components, then the surface of all combinations of sigmas which will RSS to a given system sigma is an n -sphere. For example, in 2-dimensions we have a circle. This can be

compared with the more conservative sum, given by the straight line in Fig. Gweb-21. Other correlation values will give intermediate curves. These can now be overlaid on the cost curves to find the optimum vector of sigmas to create the error budget.

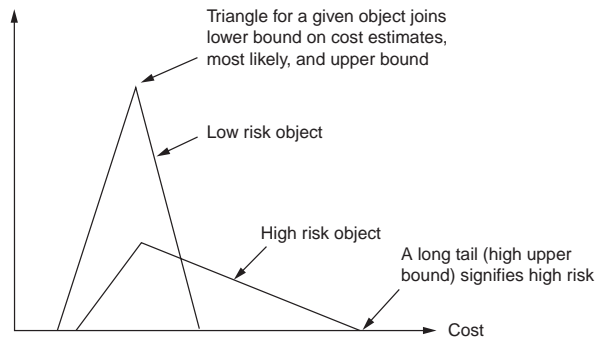


Fig. Gweb-18. Idealized Probability Density Functions for Low and High Risk Items.

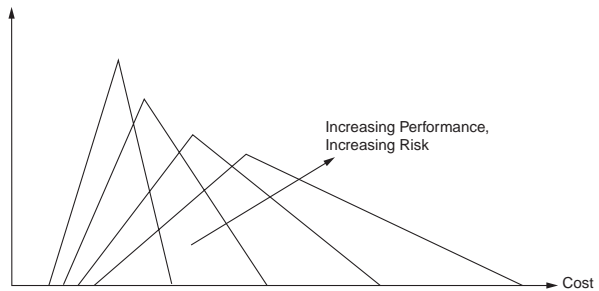


Fig. Gweb-19. Family of Probability Density Functions Parametrized Accordingly to Risk (or Performance).

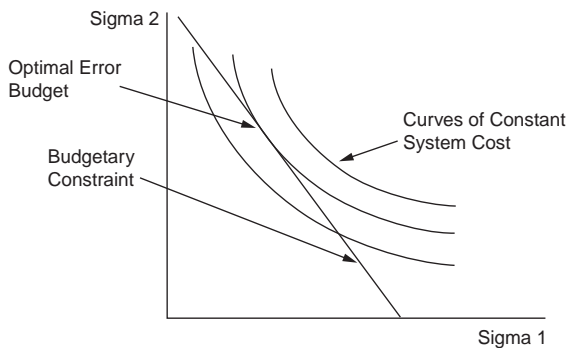


Fig. Gweb-20. Curves of Constant System Cost as a Function of Component Performance Requirements—and Hence Risk—as Derived from the Cost-Risk Curves in Fig. Gweb-19. Superimposed on a Curve of the Error Budget Constraint (Combination of Component Performances which Yield a Given System Performance). The optimal error budget can be read off from this plot.

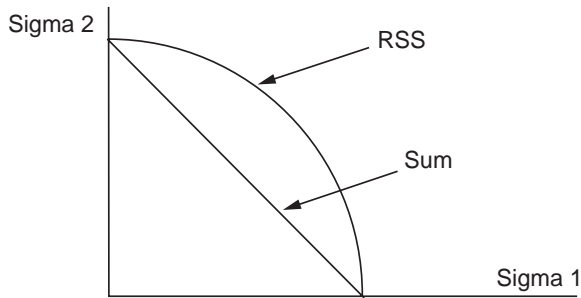


Fig. Gweb-21. Curves of Constant System Error Budget (Sum and RSS).

In a cost setting it is common to have to provide a probability distribution of system cost—i.e., not simply an expected value or a mean plus n -sigma. Such distributions are usually generated using Monte Carlo techniques based on the triangular distributions described above. The cost portion of the error budgeting problem described in this example can also be handled directly via Monte Carlo methods.

It should be noted that error budgets are not typically developed as outlined in this example. The error is often uniformly distributed among the components, or else is allocated based on the known capabilities of existing components. We will generally have some awareness of the state of the art, and will avoid setting unreasonable and, hence, expensive requirements. The error budgeting process usually involves some iterations and negotiation of “terrible injustices” [Williams, 1992].

References

- Feller, W. 1957, 1976. *An Introduction to Probability Theory and its Applications*. Vol. 1 (1957), Vol. 2 (1966), Wiley.
- Meyer, P. L. 1972. *Introductory Probability and Statistical Applications*. Addison-Wesley.
- Papoulis, A. 1965. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill.
- Parzen, E. 1960. *Modern Probability Theory and its Applications*. Wiley .
- Kac, M. 1989. *Statistical Independence in Probability, Analysis and Number Theory*. MAA.
- Box, G. and G. Jenkins. 1970. *Time Series Analysis*. Holden-Day.
- Karatzas, I. and S. Shreve. 1997. *Brownian Motion and Stochastic Calculus*. Springer.
- Oksendal, B. 1998. *Stochastic Differential Equations*. Springer.
- Williams, Michael. 1992. “Requirements Definition.” In Larson, Wiley J. and Wertz, James R. (eds.). *Space Mission Analysis and Design*, 2nd ed. Dordrecht, The Netherlands and Torrance, CA: Kluwer Academic and Microcosm, Inc.
- Wertz, James R. 2009. *Orbit & Constellation Design and Management: Spacecraft Orbit and Attitude Systems*. Hawthorne, CA: Microcosm Press.