

Supplementary Lecture Note on Information Content

2.882
Spring 2005
by Taesik Lee

1. Information content for one-FR problem

We all know the definition of the information contents in Axiomatic Design. For a one-FR design,

$$I = -\log_2 P = -\log_2 P(FR) = -\log_2 \int_{dr^l}^{dr^u} f(FR) dFR \quad (1)$$

$P(FR)$ is a probability of success for the functional requirement, i.e. the probability that a value of FR output by the designed system falls within the acceptable range. $f(FR)$ is a probability density function (or probability mass function in the case that the FR is a discrete variable) of the FR. Given $f(FR)$, integrating the function over the acceptable range, from dr^l to dr^u , gives the probability of success. A visual illustration of this is the following graph that we have seen many times in the class and the textbook.

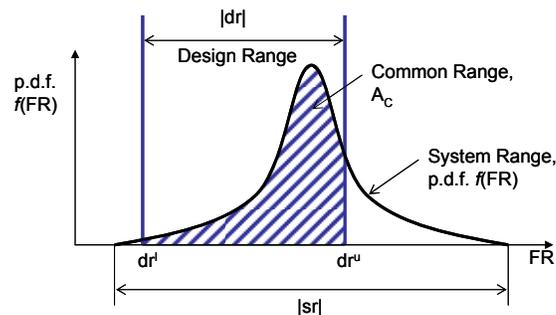


Figure 1. Probability of success is obtained by integrating the probability density function of the FR, $f(FR)$, over the Design Range. The area of the hatched section is equal to the value of the integration.

Example: Computing probability of success for an $FR \sim N(\mu, \sigma^2)$

Normal distribution (a.k.a. Gaussian distribution) is probably the most popular distribution among many engineers. It is a bell-shaped curve centered on a mean, μ , with variance, σ^2 . It is common to see outputs of a manufacturing process/system being assumed to follow a normal distribution. Use of normal distribution in statistical modeling is not an arbitrary choice of a modeler. It is justified by the fact that when many independent random factors act in an additive manner to create

variability, data will follow the normal distribution. In statistics, this is formalized as the Central Limit Theorem.

If a random variable, X , follows the normal distribution, $N(\mu, \sigma^2)$, the probability of X being between a and b is the area, A , of the hatched section in Figure 2. A can be computed by subtracting $A2$ from $A1$ as shown in the figure.

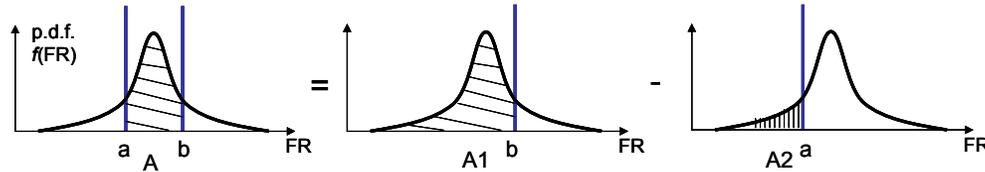


Figure 2. Computing probability of X being between a and b by considering two integrations. The area, A equals to $(A1 - A2)$.

$A1$ and $A2$ can be easily obtained by using a standard normal distribution table found in most probability and statistics textbooks. Standard normal distribution is a normal distribution where a random variable, Z , follows normal distribution with its mean at 0 and a variance of 1, i.e. $Z \sim N(0,1)$. The standard normal distribution table shows the values (numerical approximations) for integrals of the normal distribution function.

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z f(z) dz \quad (2)$$

If we imagine that Figure 2 is drawn for Z , the probability that $a \leq Z \leq b$ is

$$P(a \leq Z \leq b) = A1 - A2 = \Phi(b) - \Phi(a) \quad (3)$$

and the value $\Phi(a)$ and $\Phi(b)$ can be found from the standard normal distribution table.

In order to use the standard normal distribution table, the random variable $X \sim N(\mu, \sigma^2)$ has to be transformed to $Z \sim N(0,1)$. That can be done by the following transformation:

$$Z = \frac{X - \mu}{\sigma} \quad (4)$$

Thus, for any $FR \sim N(\mu, \sigma^2)$, we can use the standard normal distribution table to compute the probability of success.

For example, suppose we have designed a system that yields $FR \sim N(6,1)$. The design range is given as $4 \leq FR \leq 7$. See Figure 3.

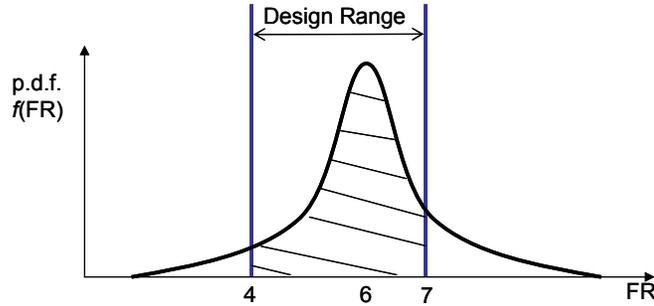


Figure 3. $FR \sim N(6,1)$. The design range is indicated by two vertical lines.

Substituting the given values into equation (4), we get $Z^+ = 1$ and $Z^- = -2$. Thus, the probability of success (the hatched area in Figure 3) can be computed as

$$\begin{aligned} P(4 \leq FR \leq 7) &= P\left(\frac{4-6}{1} \leq Z \leq \frac{7-6}{1}\right) \\ &= P(-2 \leq Z \leq 1) = \Phi(1) - \Phi(-2) \end{aligned} \quad (5)$$

Note that $\Phi(-2)$ equals to $1-\Phi(2)$ from the symmetry of the normal distribution. Equation (5) becomes,

$$P(4 \leq FR \leq 7) = P(-2 \leq Z \leq 1) = \Phi(1) - (1 - \Phi(2)) \quad (6)$$

Reading from the standard normal distribution table, we get $\Phi(1) = 0.8413$ and $\Phi(2) = 0.9772$. Therefore, $P(4 \leq FR \leq 7) = 0.8185$. \blacksquare

As you may have noticed, transforming $X = x$ to $Z = z$ is equivalent to identifying how far x is from the mean of X in the unit of its standard deviation. In the above example, what we saw was that $FR=7$ is $(+1\sigma)$ apart from its mean and $FR=4$ is (-2σ) apart, thus $Z^+=1$ and $Z^-=-2$.

It is common to hear someone saying “our system’s quality performance is at 3 sigma level.” This means (with little possibility of different interpretation) that the probability of getting a ‘good’ output from the system is 99.73%. In other words, the system range follows a distribution such that the design range corresponds to $\pm 3\sigma$ of the distribution. You can see from Table 1 that $\pm 3\sigma$ level performance would yield about 2700 ‘bad’ outputs per 1 million outputs (or ~ 3 bad outputs per 1,000 outputs). As the quality level goes up, say to $\pm 6\sigma$ level, it becomes ~ 2 bad outputs per billion outputs. That is why the term ‘six sigma’ has become a catchphrase for the popular quality improvement initiative.

Table 1. Sigma value and its quality level. For 3σ level, it is on the order of one bad output per thousand; for 6σ , it is on the order of one ppb(part per billion).

z	phi(z)	1-phi(z)	prob between +z, -z	ppm	ppb
1	0.84134474	0.15865526	0.68268948	317310.5	
2	0.977249938	0.022750062	0.954499876	45500.12	
3	0.998650033	0.001349967	0.997300066	2699.934	
4	0.999968314	3.1686E-05	0.999936628	63.37207	
5	0.999999713	2.87105E-07	0.999999426	0.57421	574.21
6	0.999999999	9.90122E-10	0.999999998	0.00198	1.980244

It should be noted that equation (1) does not include any information ('information' as in common use of the term) on the design parameter at all. The information content is only a function of the distribution of the functional requirement and the design range. As the design range is given also in a functional domain, you can see that the information content is defined solely in a functional domain. 'How we achieve the FR' is behind the scene, and the information content addresses the *resulting performance*, i.e. how well the functional requirement is satisfied, of the *complete* system. Therefore, in the strictest sense, the true information content can only be measured by observing the output of a complete system. This, in some sense, makes the information content less useful in practice: if we can measure the information content only after the system design is complete, how does it help us in the design phase? Although computing the true value of information content is impossible before the design is finished, some qualitative conclusions derived from the concept turn out to be useful in a design process. For example, means to reduce the information content or the relationship between the information content and coupling hold true regardless of our ability to compute the value of the information content. Also, estimating information content from the knowledge about design parameters and design matrix can be useful in comparing potential solutions.

If we assume that the variation of FRs is entirely attributed to the variation of DPs, then we can estimate the information content of the FR given the knowledge about the DP. A design equation for one-FR design problem is,

$$FR = [A]DP = f(DP) \quad (7)$$

Given the distribution of DP and the mapping relationship from DP to FR (A or f), we can derive the distribution of FR, a *system range* (Figure 4). This can be done by either the distribution function technique or the change-of-variable technique [refer to any statistics textbook such as the one by Hogg & Tanis]. Although it is particularly easy when the mapping is linear or strictly in(de)creasing, it still takes some effort to do the derivation. Instead, it is much easier to map the *design range* to physical domain (Figure 5) because the design range is simply an interval rather than a function. For example, in linear mapping, it is simply either stretching or shrinking the design range. This turns out to be particularly convenient for multi-FR problems, which is discussed in the next section.

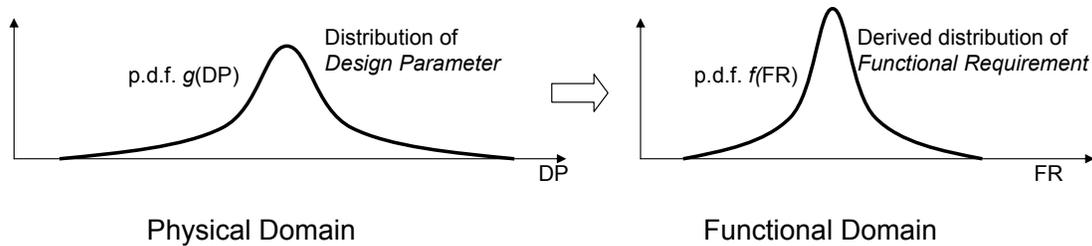


Figure 4. System range can be derived from the distribution of DP. The task may become non-trivial depending on the mapping relationship and DP distribution.

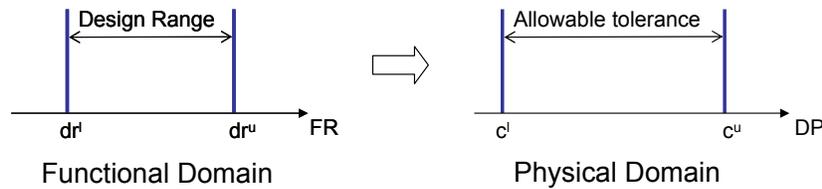


Figure 5. An alternative is mapping the design range onto physical domain, which can be easily done in most cases. The mapped range in DP domain is called ‘allowable tolerance.’

The range in DP domain, mapped from the design range, is called the *allowable tolerance* to avoid confusion with the design range while denoting the fact that once a DP value is within this range, the FR values will be within the design range. The probability of success for FR is then equal to the probability of DP's falling within the allowable tolerance. This probability is obtained by integrating the p.d.f. of DP, $g(DP)$, over the allowable tolerance. (Figure 6)

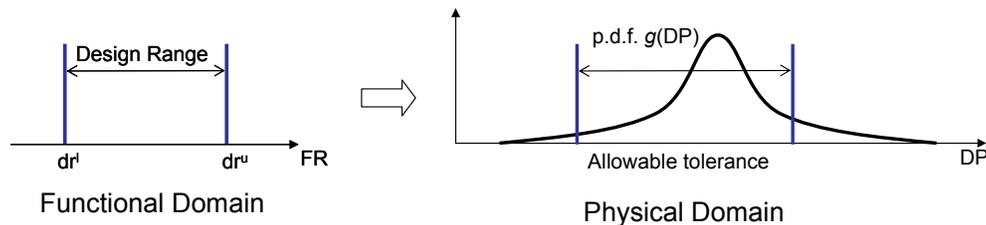


Figure 6. Probability of success for FR is equal to the probability of DP's falling within the allowable tolerance.

2. Information content for multi-FR problem

We can easily extend the definition of information content, equation (1), to multi-FR case:

$$I(FR_1, FR_2, \dots, FR_n) = -\log_2 p_{1,2,\dots,n} \quad (8)$$

where $p_{1,2,\dots,n}$ denotes the probability of success for all of the FRs. $p_{1,2,\dots,n}$ is the probability that the value of FR1 output by the designed system falls within FR1 design range, the value of FR2 within FR2 design range, etc. $p_{1,2,\dots,n}$ is defined as

$$P_{1,2,\dots,n} = \int_{\text{design hyperspace}} f(FR_1, FR_2, \dots, FR_n) dFR_1 dFR_2 \dots dFR_n \quad (9)$$

As explained in the textbook, it should be emphasized that, unless the design is uncoupled, we have to make sure the probability of success is properly computed/estimated using the joint probability.

In computing the information content, the same arguments presented in section 1 remains valid: 1) under the assumption that the only source of variation is DP, we can calculate the probability of success given DP variation and [DM], and 2) in doing so, it is much easier to map the design range onto DP domain than deriving the *joint* system range from the DP distributions.

Example: Computing probability of success for two-FR problem

Consider the following two-FR design problem.

$$\begin{cases} FR1 \\ FR2 \end{cases} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{cases} DP1 \\ DP2 \end{cases} \quad (10)$$

The design ranges for FRs are given as $-0.5 \leq FR1 \leq 0.5$ and $-2 \leq FR2 \leq 2$. The variation of DPs is given by two uniform distribution: $DP1 \sim U[-1,1]$ and $DP2 \sim U[0,1.5]$ where $U[a,b]$ denotes a uniform distribution between a and b . Let g_1 denote the probability density function of DP1 and g_2 be the p.d.f. of DP2. Then, g_1 and g_2 are constant within the two bounds: $g_1 = 0.5$ for $-1 \leq DP1 \leq 1$, and $g_2 = 2/3$ for $0 \leq DP2 \leq 1.5$. (see Figure 7)

First, we need to assume that DPs are statistically independent to each other to simplify the problem. Under the assumption, the joint probability density function of (DP1, DP2), $g_{1,2}$, is simply a product of two uniform probabilities. This, again, yields a uniform p.d.f. in DP1-DP2 plane:

$$g_{1,2}(DP1, DP2) = \begin{cases} 1/3 & \text{if, } -1 \leq DP1 \leq 1 \text{ and } 0 \leq DP2 \leq 1.5 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Figure 7 is a graphical illustration of obtaining $g_{1,2}$.

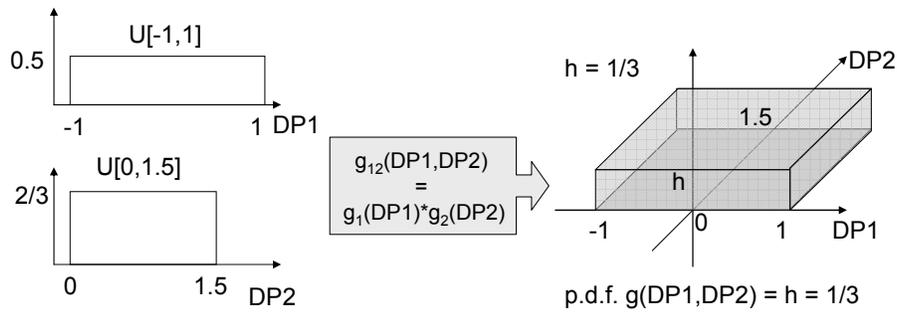


Figure 7. Assuming DP1 and DP2 are statistically independent to each other, the joint p.d.f. $g_{1,2}(DP1,DP2)$ is a product of g_1 and g_2 . $g_{1,2}$ is another uniform density function, i.e. constant $1/3$.

Now, we can map the design range onto DP domain to obtain the allowable tolerance (in this case, an area). The design range is represented by a rectangle on the FR-plane (Figure 8a). From the design matrix,

$$FR1 = DP1$$

$$FR2 = DP1 + DP2$$

If we rearrange these equations such that DPs are expressed in terms of FRs,

$$DP1 = FR1$$

$$DP2 = -DP1 + FR2$$

FR1 design range is given as $-0.5 \leq FR1 \leq 0.5$. Thus, the first equation of the above represents a vertical line in the area between $-0.5 \leq DP1 \leq 0.5$ as shown in Figure 8b. The second equation represents a set of straight line with the slope of -1, and its intersection with DP2 axis ranges from -2 to 2 (given by FR2 design range). See Figure 8c. Finally the allowable tolerance for DP1 and DP2 is obtained by taking intersection of 8b and 8c.

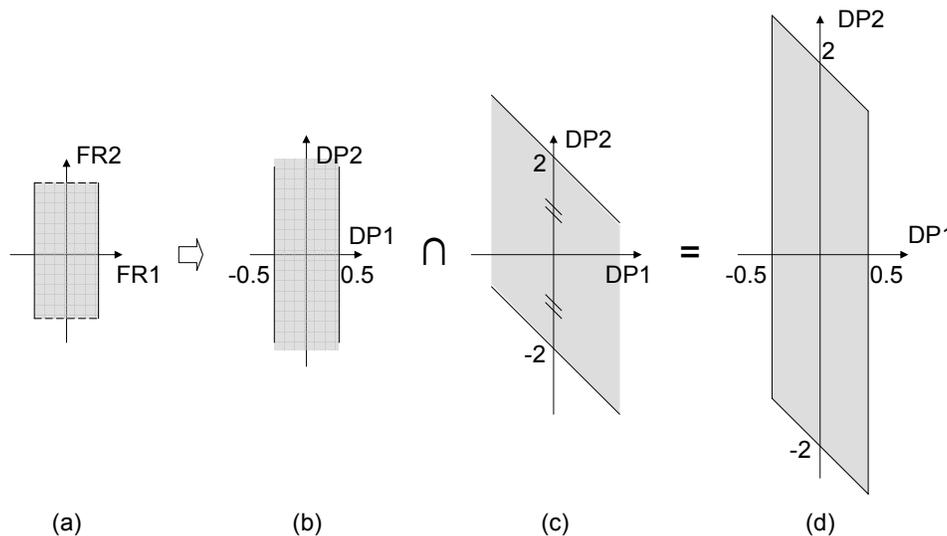


Figure 8. Mapping the design range onto DP space to obtain the allowable tolerance.

Alternatively, since the design matrix is constant (linear mapping), we know that four vertices of the design range will become the vertices of the allowable tolerance. These vertices can be easily calculated by multiplying the inverse of the design matrix by the vertices of the design range.

Now that we have the allowable tolerance and the joint probability density function of (DP1, DP2), the probability of DPs' being within the allowable tolerance is readily computed by integrating $g_{1,2}$ over the allowable tolerance. Since $g_{1,2}$ is constant, the integral is simply the product of $g_{1,2}$ and the overlapping area between the allowable tolerance and the projection of $g_{1,2}$:

$$\text{Probability} = 1/3 * (1 * 1.5) = 1/2$$

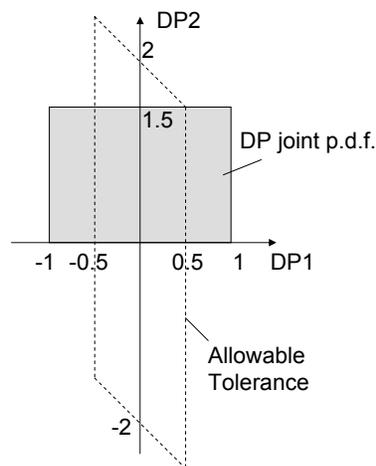


Figure 9. Probability of success is the integral of $g_{1,2}$ over the allowable tolerance

Assuming DPs are the only source of FR variation, we can say the probability of success of FR1 and FR2 equals to the probability we computed above. ■

We know that, if the design matrix is coupled or decoupled, the probability of success for FRs is not equal to the product of the probability of success for individual FRs: $p_{1,2,\dots,N} \neq p_1 p_2 \dots p_N$. This individual probability, p_i , is called the marginal probability (of success) of FR_i . The marginal probability is defined based on the joint probability. For example, in a 2-FR case:

$$p_1 = \int_{dr^{l,FR1}}^{dr^{u,FR1}} f_1(FR1) dFR1 \quad (12)$$

$$f_1(FR1) = \int_{-\infty}^{\infty} f_{1,2}(FR1, FR2) dFR2$$

$f_1(FR1)$ is called the marginal p.d.f. for FR1. Suppose we have a system that has FR1 and FR2. Let us say we only care about FR1 and measure the FR1 samples to obtain its distribution. Then we can compute the probability of success for FR1 by examining the

distribution with the FR1 design range. This probability is the marginal probability of success for FR1.

It should be noted that the probability of success computed by multiplying the marginal probabilities is not only inaccurate, but potentially misleading in drawing a conclusion from the observation. The following example illustrates the point.

Example: Marginal probability vs. Joint probability

The two-FR design problem we saw in the earlier example is revisited.

$$\begin{Bmatrix} FR1 \\ FR2 \end{Bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{Bmatrix} DP1 \\ DP2 \end{Bmatrix} \tag{10}$$

The probability of success was obtained by the joint probability calculation, and it was 0.5.

In order to compute the marginal probability of FR1 and FR2, we must first derive the joint p.d.f. for (FR1,FR2). Since the DPs follow uniform distribution and the mapping is linear, the derivation is easy: the derived joint p.d.f. will also be that of a uniform distribution. We can map the *projection* of the DP's joint p.d.f. onto the FR domain, and the joint p.d.f. of FR will simply be the uniform p.d.f. over the mapped area. As the integral of any p.d.f. must be 1, the height of the uniform p.d.f. of (FR1,FR2) is 1/(projection area) = 1/3.

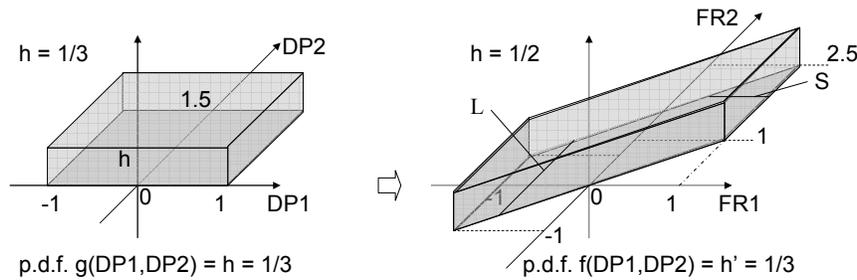


Figure 10. Deriving $f_{1,2}$ from $g_{1,2}$. DP's having the uniform distribution and its linear mapping make it simple.

Now, the second equation of equation (12) becomes,

$$f_1(FR1) = \int_{-\infty}^{\infty} f_{1,2}(FR1, FR2) dFR2 = \int_{-\infty}^{\infty} \frac{1}{3} dFR2 \quad \text{for } -1 \leq FR1 \leq 1 \tag{13}$$

You may notice that the integral $\int_{-\infty}^{\infty} dFR2$ is the length of the projected area along FR2 axis (L in Figure 10). The figure suggests that the length is 1.5 throughout the $-1 \leq FR1 \leq 1$ range. Thus,

$$f_1(FR1) = \begin{cases} 1/3 * 1.5 = 0.5 & \text{when } -1 \leq FR1 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

For FR2,

$$f_2(FR2) = \int_{-\infty}^{\infty} f_{1,2}(FR1, FR2) dFR1 = \int_{-\infty}^{\infty} \frac{1}{3} dFR1 \quad \text{for } -1 \leq FR2 \leq 2.5 \quad (14)$$

$\int_{-\infty}^{\infty} dFR1$ is the length of the projected area along FR1 axis (S in Figure 10), and, unlike L, S is a function of FR2. By inspecting Figure 10, you can see that S linearly increases from 0 to 1.5 as FR2 increases from -1 to 0.5. Then, for $0.5 < FR2 < 1.0$, S is constant at 1.5. From $FR2 = 1.0$ to $FR2 = 2.5$, S decreases from 1.5 back to zero. S, a function of FR2, multiplied by 1/3 yields the marginal p.d.f. f_2 that has a trapezoidal shape. Figure 11 summarizes the above discussion with the design range also shown.

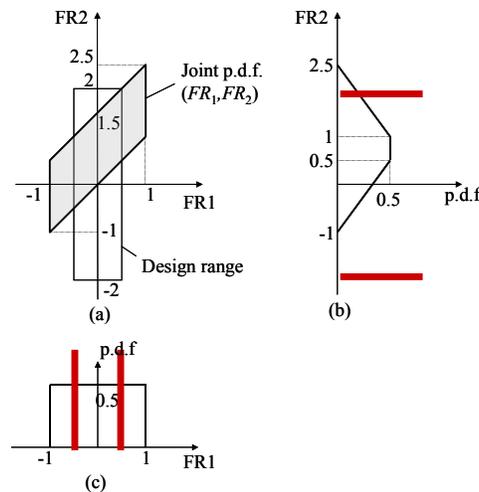


Figure 11. The marginal p.d.f. for FR1 and FR2 can be computed from the joint p.d.f. of (FR1,FR2). Each of the two marginal p.d.f. is integrated within the design range to give the marginal probability of success for FR1 and FR2, respectively.

Integrating each of the two marginal p.d.f. over its relevant design range yields the marginal probability of success. From Figure 11, you can figure out that $p_1 = 0.5$ and $p_2 = 0.9583$. Product of these two marginal probabilities is 0.4792. As expected, this value is a bit off from the correct value, $p_{1,2} = 0.5$.

Now suppose that DP2 variation has changed somehow (e.g. change of a supplier) from $U[0,1.5]$ to $U[-1,1.6]$. Following the same steps described above, we can obtain $p'_{1,2}$, p'_1 , and p'_2 . It turns out that $p'_{1,2} = 0.499$; $p'_1 = 0.5$; $p'_2 = 0.9654$. Again, $p'_1 * p'_2 (= 0.4827)$ is slightly off from the correct value.

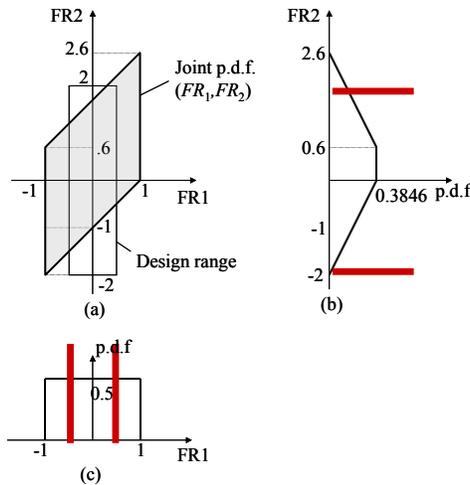


Figure 12. Joint p.d.f. of (FR1,FR2) and the marginal p.d.f.'s for FRs.

What is important here is that even though the probability of success by using the joint p.d.f. shows a hint of performance degradation ($p'_{1,2} < p_{1,2}$), the other probability of success by using two marginal probabilities indicates the opposite trend ($p'_1 * p'_2 > p_1 * p_2$). This can be a significant problem when detecting mal-functioning is time-critical for recovering the system. Therefore, as emphasized earlier, it is very important to consider the joint p.d.f. when there are multiple functional requirements that are not completely uncoupled. ■

This section explained how to compute information contents for multiple-FR problem using 2-FR uniform distribution examples. Applying the same technique to three or more-FR problems with general distribution will not be effective. Indeed, when it is not the simplest case, computing information contents is perhaps best handled by a probabilistic simulation technique such as Monte Carlo simulation.