Combining Variation on Inputs

In this module we consider simple probability-based methods of combining measures of input or component variability to predict output/system/overall variability. The basic tools available in this effort are

- exact formulas for means and variances of linear functions of several random variables,
- approximations for means and variances of general functions of several random variables (based on linear approximations to the functions and the above), and
- simple probabilistic simulations, easily done in any decent statistical package (or, heaven forbid, EXCEL).
Sometimes geometry or physical theory gives one an equation for a variable of interest in terms of more basic variables $X, Y, \ldots, Z$

$$U = g (X, Y, \ldots, Z)$$

and an issue of interest is how one might infer the level of variation to be seen in $U$ from given information on the levels of variation in the inputs and the form of $g$.

The context of this kind of study is usually design, where allowable levels of variation and the details that lead to a particular $g$ are being decided in a "What if?" mode of thinking.
Example 5.9 of *SQAME* is a nice "tolerance stack-up" example. A company wished to pack 4 units of product in a carton and was experiencing difficulty in packaging. The linear size of the carton interior ($Y$) and exterior linear sizes of four packages ($X$'s), together with the carton "head space" ($U$) are illustrated in the figure below. In this problem, simple geometry reveals that

$$U = Y - X_1 - X_2 - X_3 - X_4$$

**Figure:** Cartoon of a Packaging Problem (Example 5.9 of *SQAME*)
Example 5.8 of *SQAME* is a simple circuit example involving an assembly of 3 resistors. The figure below illustrates this situation. In this problem, elementary physics produce an equation for assembly resistance, $R$,

$$R = R_1 + \frac{R_2 R_3}{R_2 + R_3}$$

**Figure:** A Schematic for Example 5.8 of *SQAME*
Example

Example 19-3 Auto Door Tolerancing

The figure below concerns a problem faced by an engineer who must set tolerances on various geometric features of a car door assembly, with the end goal of creating a uniform gap between the door and the body of an automobile on which it is to be hung.

**Figure:** Cartoon of a Door Tolerancing Problem
Some plane geometry and trigonometry applied to this situation produce the following set of equations that in the end express the gaps $g_1$ and $g_2$ at the elevation of the top hinge and a distance $d$ below that hinge in terms of $x, w, \theta_1, y, \theta_2,$ and $\phi$ (which are all quantities on which a design engineer would need to set tolerances).

\[
\begin{align*}
p &= (-x \sin \phi, x \cos \phi) \\
q &= p + (y \cos (\phi + (\theta_1 - \frac{\pi}{2})), y \sin (\phi + (\theta_1 - \frac{\pi}{2}))) \\
s &= (q_1 + q_2 \tan(\phi + \theta_1 + \theta_2 - \pi), 0) \\
u &= (q_1 + (q_2 + d) \tan(\phi + \theta_1 + \theta_2 - \pi), -d) \\
g_1 &= w - s_1 \\
g_2 &= w - u_1
\end{align*}
\]

That is, though we have not written them out explicitly here, there are 2 functions of the inputs $x, w, \theta_1, y, \theta_2,$ and $\phi$ that produce the 2 gap values.
In the three examples, then, what are tools for predicting how variation in the inputs will be reflected in the outputs? If one is willing to model inputs $X, Y, \ldots, Z$ as independent random variables

- equations (5.23) and (5.24) of SQAME are straight from basic probability and say that for $g$ linear, i.e. where for constants $a_0, a_1, \ldots, a_k$,

$$U = a_0 + a_1 X + a_2 Y + \cdots + a_k Z$$

$U$ has mean

$$\mu_U = a_0 + a_1 \mu_X + a_2 \mu_Y + \cdots + a_k \mu_Z$$

and variance

$$\sigma^2_U = a_1^2 \sigma^2_X + a_2^2 \sigma^2_Y + \cdots + a_k^2 \sigma^2_Z$$
Tools

Approximate Methods for Nonlinear $g$

Still modeling inputs $X, Y, \ldots, Z$ as independent random variables

- equations (5.26) and (5.27) of *SQAME* are based on a (Taylor theorem) "linearization" of a general $g$ and the above relationship and say that roughly

$$
\mu_U \approx g(\mu_X, \mu_Y, \ldots, \mu_Z)
$$

and

$$
\sigma^2_U \approx \left( \frac{\partial g}{\partial x} \right)^2 \sigma_X^2 + \left( \frac{\partial g}{\partial y} \right)^2 \sigma_Y^2 + \cdots + \left( \frac{\partial g}{\partial z} \right)^2 \sigma_Z^2
$$

where the partial derivatives are evaluated at the point $(\mu_X, \mu_Y, \ldots, \mu_Z)$, and

- simple probabilistic simulations can be used to approximate the distribution of $U$ based on some choice of distributions for the inputs in very straightforward fashion for general $g$. 
Before proceeding to illustrate these 3 methods, there are several points to be made. In the first place, notice that since in the case of a linear $g$, the $a$’s are exactly the partial derivatives of $U$ with respect to the input variables, the general approximation produces the exact result in case $g$ is exactly linear. Second, note that while we will see that simulation is completely straightforward and indeed almost mindless to carry out, there will be occasions where it is desirable to use the formulas. In particular, it’s possible to look at a term like

$$
\left( \frac{\partial g}{\partial x} \right)^2 \sigma_X^2
$$

(in an approximation for $\sigma_U^2$) as the part of the variance of $U$ traceable to variation in the input $X$. Finally, we observe that the approximation for $\sigma_U^2$ is "qualitatively right." The variability in $U$ must depend upon both 1) how variable the inputs are and 2) what the rates of change of output with respect to inputs are. (These two are measured respectively by the variances and the derivatives.)
The figure below illustrates the importance of the derivatives in determining how variance is transmitted through $g$.

**Figure:** Cartoon Illustrating the Importance of Rates of Change in Determining Variance Transmission ($g$ a Function of One Input)
Returning to box packing, the exact form for the variance of $U$ for a linear $g$ implies

$$\sigma_U^2 = 1^2 \sigma_Y^2 + (-1)^2 \sigma_{X_1}^2 + (-1)^2 \sigma_{X_2}^2 + (-1)^2 \sigma_{X_3}^2 + (-1)^2 \sigma_{X_4}^2$$

The IE 361 team estimated that $\mu_Y \approx 9.556$ in and $\sigma_Y \approx .053$ in and $\mu_X \approx 2.577$ in and $\sigma_X = .061$ in. So

$$\sigma_U^2 \approx (.053)^2 + 4(.061)^2 = .0177 \text{in}^2 \quad \text{and} \quad \sigma_U = \sqrt{.0177} = .133 \text{in}$$

This together with the fact that $\mu_U \approx 9.556 - 4(2.577) = -.7520$ (the mean head space was about negative 3/4 in) created the packing problems. Some adjustment (downward) in $\mu_X$ and/or (upward) in $\mu_Y$ (potentially together with reductions in $\sigma_X$ and $\sigma_Y$) were necessary to eliminate packing problems. This kind of analysis allowed the company to project the impacts of various changes (coming with corresponding costs).
Returning to resistor assembly, with \( R = R_1 + (R_2 R_3) / (R_2 + R_3) \),

\[
\frac{\partial R}{R_1} = 1, \quad \frac{\partial R}{\partial R_2} = \frac{R_3^2}{(R_2 + R_3)^2}, \quad \text{and} \quad \frac{\partial R}{\partial R_3} = \frac{R_2^2}{(R_2 + R_3)^2}
\]

So, for example, making resistor assemblies using \( R_1 \) with mean 100\( \Omega \) and standard deviation 2\( \Omega \), and both \( R_2 \) and \( R_3 \) with mean 200\( \Omega \) and standard deviation 4\( \Omega \), the approximations for the mean and variance of \( R \) provide

\[
\mu_R \approx 100 + \frac{200 \cdot 200}{(200 + 200)^2} = 200\Omega
\]

and

\[
\sigma_R \approx \sqrt{1^2 (2)^2 + \left(\frac{1}{4}\right)^2 (4)^2 + \left(\frac{1}{4}\right)^2 (4)^2} = 2.45\Omega
\]

(Engineering mathematics software like MathCad can be used in problems like this, but this particular calculus is easy enough to do without resort to such a tool.)
To illustrate how easily simulations can be used to provide answers in problems of the type under discussion in this module, the next figures show what happens if one uses normal distributions with means and standard deviations as above to simulate 1000 values of $R$ and computes summary statistics from those (producing values in substantial agreement with the previous calculus-based figures).

**Figure:** JMP Dialogue Boxes for Setting Up Columns for Simulation of $R$
Figure: JMP Data Sheet and Report for Simulation of $R$

Vardeman and Morris (Iowa State University)
To illustrate how simply something as apparently complicated as the car door tolerancing problem can be handled using simulation, the following is a bit of Minitab code and corresponding output for one set of means and standard deviations for \( x = c_1 \), \( y = c_2 \), \( w = c_3 \), \( \phi = c_4 \), \( \theta_1 = c_5 \), \( \theta_2 = c_6 \), at \( d = 40 \) (in units of cm and rad).

MTB > Random 1000 c1;
MTB > Random 1000 c2;
MTB > Random 1000 c3;
MTB > Random 1000 c4;

SUBC> Normal 20 .01.
SUBC> Normal 90. .01.
SUBC> Normal 90.4 .01.
SUBC> Normal 0 .001.
Example 19-3 continued

MTB > Random 1000 c5 c6;
SUBC> Normal 1.570796 .001.
MTB > let c7=-c1*sin(c4)
MTB > let c8=c1*cos(c4)
MTB > let c9=c7+c2*cos(c4+(c5-1.570796))
MTB > let c10=c8+c2*sin(c4+(c5-1.570796))
MTB > let c11=c9+c10*tan(c4+c5+c6-3.141593)
MTB > let c12=c9+(c10+40)*tan(c4+c5+c6-3.141593)
MTB > let c13=c3-c11
MTB > let c14=c3-c12
MTB > Describe 'g1' 'g2'.

Vardeman and Morris (Iowa State University)
### Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>g1</td>
<td>1000</td>
<td>0.40053</td>
<td>0.40035</td>
<td>0.03078</td>
</tr>
<tr>
<td>g2</td>
<td>1000</td>
<td>0.40138</td>
<td>0.39775</td>
<td>0.09153</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>g1</td>
<td>0.31109</td>
<td>0.51216</td>
<td>0.37877</td>
<td>0.42168</td>
</tr>
<tr>
<td>g2</td>
<td>0.12174</td>
<td>0.75907</td>
<td>0.34043</td>
<td>0.46280</td>
</tr>
</tbody>
</table>

Where they are applicable, these methods of "statistical tolerancing" are powerful tools for understanding the consequences of various levels of input variation on the output variation of a system.