Modeling error propagation in a measurement system

Authors: Pierre-Augustin Grivelet And Luc Thomas
RMS S.A.S
355 av. Albert Einstein
BP 40054
13792 Aix En Provence

Abstract

We outline in this paper a hybrid method to propagate uncertainty in systems described by trees. Given a system described by a tree where nodes are basic operations and leaves are random variables described by their probability density functions (PDFs), one show that by a combination of exact computations with sampled versions of PDFs one can obtain the PDFs of output variables with great efficiency. This can be applied to a large variety of problems including uncertainty propagation in measurement systems, combination of PDFs in Fault trees or junction trees for expert systems. On top of that, this calculus is fast enough to cope with sensitivity studies, where one seeks to optimize one of the system parameter to achieve some requested features on the output PDFs. This paper focuses on the calculation of output random variables PDFs with sampled versions of the input PDFs.

Keywords: uncertainty, propagation, sampled, PDF.

1. Basic hypothesis

In numerous systems, mastering the propagation of errors or random variables can be or prime importance. Some system optimization may rely for example on a proper handling of measurement accuracy and errors, which will then be very important for productivity gains. In some other situations, transducers may be used by a control command system to assess critical situations and take high consequences decisions. The estimation of uncertainty bounds is therefore of great interest. However, uncertainty calculations are up to now mainly based on crude Gaussian assumptions, and complex processing of the raw measurements tend to prevent the use of simple variance combination rules to estimate uncertainty. That is why modern simulation tries to carry on through entire systems the best estimate of random variables probability densities. We describe here one technical solution that we think is able to cope with this not so easy task.
We consider trees or acyclic graphs describing systems where one sought to estimate output variables PDFs given sampled or parametric descriptions of input variables PDFs. Graphs nodes are deterministic operations on variables and at this stage we consider the following operators:

- Linear combination
- Multiplication and division
- Transformation by a monotonic differentiable function
- Mixing
- Saturation and truncatures
- Median
- Filtering by a transverse (MA) filter.

There are no particular assumptions made on the random variables which can have discrete or continuous PDFs (though we here talk about continuous variables only). The basis of this propagation method is to distinguish two distinct cases when combining variables:

- Operation on independent random variables
- Operation on correlated random variables

In the second case, our method relies on efficient Monte Carlo methods such as Markov chain [1],[2] Monte Carlo or importance sampling [3]. We focus in this paper on the first case where variables are independent and consequently there exists a computable formulation for the resulting variable. We will show how sampled versions of the PDFs may be used in numerical implementations of these analytic formulae, and how computation parameters might be tailored to match requested bounding errors. Examples are shown comparing theoretical and true results.

2 Sampling of random variables PDFs

a. Definitions

Let \( p_X \) be the PDF of the real and continuous random variable \( X \). The sampled version of \( p \) is defined by the 5-uplet \( \theta_X = \{ x_0, \Delta, p, E_0, E_1 \} \)

Where \( \theta_X \) components are defined as:

- \( x_0 \) lower bound of the sampled definition of \( p_X \)
- \( \Delta \) sampling rate of \( p_X \)
- \( p= \{ p_i, i=0 \text{ to } n-1 \} \) sample values
- \( x_1 \) higher bound \( (x_1=x_0+(n-1)\Delta) \)
- \( E_0 \) upper bound for the left missing probability \( e_0 \)
- \( E_1 \) upper bound for the right missing probability \( e_1 \)

Relations between those values and the continuous PDF are:
\[ p_i = p (x_0 + i\Delta) \]

The total integral error \( \varepsilon \) is defined by:

\[
\left| \int_{x_0}^{x_0+n\Delta} p(x)dx - I(p) \right| = \varepsilon
\]

where \( I(p) \) is a numerical quadrature of \( p \) defined by

\[
I(p) = \Delta \sum_{i=0}^{n} p_i - \frac{\Delta}{2} \left( p_0 + p_{n-1} \right)
\]

This quadrature method is far from optimal, but very fast. It underestimates probability in the vicinity of the distribution modes and overestimates them around troughs. If this sampling holds with the Nyquist criteria, the continuous PDF can be restored from its sampled version between lower and upper bounds and \( \varepsilon \) can be made arbitrarily small. Therefore one should distinguish between requirements on \( \Delta \) due to Nyquist criterion and due to requested accuracy of the computation of probability.

\( E_0 \) and \( E_1 \) are such that:

\[
E_i > e_i = \int_{x_0}^{x_0+n\Delta} p(x)dx
\]

\[
E_0 > e_0 = \int_{-\infty}^{x_0} p(x)dx
\]

As \( \varepsilon \) might be arbitrarily minimized, one has:

\[
\int_{-\infty}^{x_0} p(x)dx < 1 - I(p)
\]

\[
\int_{x_0+n\Delta}^{+\infty} p(x)dx < 1 - I(p)
\]

We therefore choose \( E_0 = E_1 = 1-I(p) \) which means that any probability is defined with an accuracy of \( \varepsilon+2(1-I(p)) \).

One furthermore seeks that distribution tails might be known with an error in the same bounds as the quadrature error. Then when one computes a probability with \( \theta_X \) the error is bounded by \( 3E_0 \).

The choice of \( D \) depends of the regularity of \( p_X \). One will see for example that for the Laplace-Gauss distribution, \( \Delta=\sigma/30 \) is fairly enough to achieve a \( 10^{-10} \) accuracy.

\textbf{b. Sampling from histograms or parametric densities}
From a practical point of view, input random variables may be known as:

- Parametric probability density functions, such as Laplace-Gauss
- Observation histograms, computed from a particular set of realizations.

The first case is obvious, as the sampled values are just computed from the analytic formulation of the PDF. In the second case, there is a standard computing scheme called “kernel methods”. One can find a good reference to kernel methods in [2]. However histograms obtained by counting generally lack substantial covering of the distribution tails. If the considered industrial application calls for an estimation of the probability of rare events, this could be a problem which needs to be addressed by incorporating some additional à priori information on the behaviour of distribution tails, via for example a generalized Pareto model. The Pareto model can be ultimately sampled, once its parameters are chosen appropriately.

Let’s take here the example of the Laplace-Gauss parametric distribution $N(\mu, \sigma)$ with $\mu=0$ and $\sigma=1$.

If $x_0$ is set to -5, and $\Delta$ to 0.01 with $n = 1000$, one has:

$\varepsilon = 1.5 \times 10^{-10}$
$e_0 = e_1 = 2.8 \times 10^{-7}$
$E_0 = E_1 = 5.7 \times 10^{-7}$.

A probability such as $P(x>A)$ is therefore known with a $10^{-6}$ accuracy. There is no reason to choose a thinner sampling rate as $\varepsilon$ is already lower than $E_0$. If we now deal with rare events such as a $10^{-10}$ probability, we will have to choose $x_0=7$ and $\Delta=0.01$ in order to get:

$\varepsilon = 2.5 \times 10^{-14}$
$e_0 = e_1 = 1.0 \times 10^{-15}$
$E_0=E_1 = 2.5 \times 10^{-12}$.

In this case one can expect to compute this probability with an accuracy of 5% by sampling $p_X$ up to 7 standard deviates.

c. Sampling parameters selection method

The choice of $\Delta$ depends on $p_X$ smoothness. The minimal request if for $\Delta$ to match the Nyquist criterion, which is linked to the spectral support of $p_X$ considered as a sampled signal. Furthermore, in order to use a fast and simple numerical quadrature $I(p)$, it is necessary to outweigh this criterion. As a rule of thumb we use the following formula:

$$\Delta = \frac{1}{2kB}$$

Where B is the bandwidth of $p_X$ and $k$ is the oversampling factor.
B is computed from the modulus of the Fourier transform of \( p_X \), which is the characteristic function \( g_X \) of the random variable \( X \), by the following inertia computation:

\[
B^2 = \frac{1}{E_p} \left| \int_{-\nu}^{\nu} v^2 |g_X(v)|^2 dv \right|
\]

where 

\[
E_p = \int p_X^2(x) dx
\]

\( k \) is heuristically chosen to be 100.

The choice of \( x_0 \) and \( x_1 \) depends on the sought accuracy. One practically limits to the choice of the half domain \( D \) by setting \( x_0 = \mu - D \) and \( x_1 = \mu + D \). \( D \) is specified as a number \( m \) of standard deviates: \( D = m \sigma \).

If \( u \) is the requested accuracy, one choose \( m \) such that \( u < |1 - I(m)| \), knowing that \( p \) is defined by the bounds \( x_0 = \mu - m \sigma \) and \( x_1 = \mu + m \sigma \). \( I \) is fast\(^1\) to compute and can be achieved through incremental computation for integer values of \( m \), \( m = 3, 4, 5, \ldots \).

Applied to the \( N(0,1) \) Laplace-Gauss distribution this method yield the following results:

The characteristic function of \( N(0,1) \) is the Gaussian
\[
\frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]
which yields \( B = 0.16 \). One deduce from that the value \( \Delta = 0.03 \).

We wish to achieve an accuracy \( u = 10^{-9} \) and the table of \( 1 - I(m) \) is the following:

<table>
<thead>
<tr>
<th>( m )</th>
<th>( u )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.0027</td>
</tr>
<tr>
<td>4</td>
<td>6.0 ( \times ) 10(^{-3} )</td>
</tr>
<tr>
<td>5</td>
<td>5.7 ( \times ) 10(^{-4} )</td>
</tr>
<tr>
<td>6</td>
<td>2.0 ( \times ) 10(^{-5} )</td>
</tr>
<tr>
<td>7</td>
<td>2.5 ( \times ) 10(^{-7} )</td>
</tr>
<tr>
<td>8</td>
<td>7.5 ( \times ) 10(^{-8} )</td>
</tr>
</tbody>
</table>

One has therefore to take \( m = 7 \) to guarantee in any case an accuracy of \( 10^{-9} \) on the computation of probabilities. The necessary number of samples is then \( n = 14/0.03 = 466 \).

### 3 Operations on sampled PDFs

#### 3.1 Addition of two independent random variables

We will review in this paragraph some of the operations listed in paragraph 1. However to start with a simple and thoroughly detailed case, we will examine the case of the addition of two independent random variables with a focus on the control

\(^1\) Computation of \( I \) may be prepared by a suitable upsampling stage.
of errors and cross check the results with a theoretically mastered situation, namely the addition of two Gaussian variables.

The addition of two variables $X_1$ and $X_2$ yields a random variable $Z$ such that:

$$P(Z = z_0) = P(X_1 + X_2 = z_0) = \int p_{X_1}(u) p_{X_2}(z_0 - u) du$$

Where we see that $Z$’s PDF is the convolution of the PDF of $X_1$ by the PDF of $X_2$. Practically this convolution is implemented in the Fourier domain, that is to say by multiplying the characteristic functions of $X_1$ et $X_2$: $g_z = g_{X_1} \cdot g_{X_2}$.

The errors on the resulting sampled distribution solely results from the approximate representation of the sources PDF, as the Fourier product by itself do not carry any error. The bulk of error is due to the tail errors $e_0$ and $e_1$, as the supports of $X_1$ and $X_2$ are usually unbounded. To evaluate this error, each PDF is divided in two functions, one representing the main lobe and the other representing the distribution tails.

Let $M_1$ and $M_2$ be the limitation of $X_1$ and $X_2$ to the interval $[a_1, b_1]$ and $[a_2, b_2]$. Let $R_1$ and $R_2$ be the limitations to the corresponding distribution tails. Let’s estimate the error made on a given probability on $Z$ by assuming that $p_{X_1} \ast p_{X_2}$ is approximated by $M_1 \ast M_2$:

$$p_{X_1} \ast p_{X_2} = (M_1+R_1) \ast (M_2+R_2) = M_1 \ast M_2 + M_1 \ast R_2 + M_2 \ast R_1 + R_1 \ast R_2$$

Let $H$ be a given hypothesis on $Z$ (for example $Z > c$). We have:

$$P(H) = P(H \mid x_1 \in [a_1, b_1]) + P(H \mid x_1 \notin [a_1, b_1])$$

and:

$$P(H) = P(H \mid x_1 \in [a_1, b_1], x_2 \in [a_2, b_2]) + P(H \mid x_1 \in [a_1, b_1], x_2 \notin [a_2, b_2])$$

$$+ P(H \mid x_1 \notin [a_1, b_1], x_2 \in [a_2, b_2]) + P(H \mid x_1 \notin [a_1, b_1], x_2 \notin [a_2, b_2])$$

That is to say:

$$P(H) = P(H \mid x_1 \in [a_1, b_1], x_2 \in [a_2, b_2]) + E$$

But for any domains $I_1$ et $I_2$, $P(H \mid x_1 \in I_1, x_2 \in I_2) < P(x_1 \in I_1).P(x_2 \in I_2)$

Therefore

$$E < P(x_1 \in [a_1, b_1], x_2 \notin [a_2, b_2]) + P(x_1 \notin [a_1, b_1], x_2 \in [a_2, b_2]) + P(x_1 \notin [a_1, b_1], x_2 \notin [a_2, b_2])$$

and:

$$E < P(x_1 \notin [a_2, b_2]) + P(x_1 \notin [a_1, b_1], x_2 \notin [a_2, b_2])$$

Actually we have:

$$P(H \mid x_1 \in [a_1, b_1], x_2 \in [a_2, b_2]) = \int_{\mathbb{R}^H} (M_1 \ast M_2)(z) dz$$

Therefore if we approximate the PDF of $Z$ by the finite support convolution $M_1 \ast M_2$, every probability computed on $Z$ can be deduced from the matching summation on $M_1 \ast M_2$ with an error bounded by the cumulated probabilities of the distribution tails of $X_1$ and $X_2$. 
3.1.1 Example of two independent Gaussian variables

Addition of two independent Laplace-Gauss distributions being itself Laplace-Gauss, we compute the addition with the previous method and we vary the domain used for the convolution between +/- 3 \( \sigma \) and +/- 10 \( \sigma \). For each computation, we write in the following table the true and estimated probabilities \( P(0 \leq z < 3\sigma) \) and \( P(z \geq 3\sigma) \).

Table 1: error (expressed in thousands) on the addition of the two random variables in function of the domain width around the mean (expressed in standard deviates).

<table>
<thead>
<tr>
<th>Domain ([-k\sigma, +k\sigma])</th>
<th>Probability error on ([0, 2\sigma])</th>
<th>Probability error on ([2\sigma, +\infty])</th>
<th>(P(Z \geq 2\sigma)) theoretical</th>
<th>Cumulative sum distribution tails</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-29.9495</td>
<td>-14.5105</td>
<td>78.6496</td>
<td>91.0005</td>
</tr>
<tr>
<td>3</td>
<td>-0.8895</td>
<td>-1.8062</td>
<td>16.9474</td>
<td>5.3996</td>
</tr>
<tr>
<td>4</td>
<td>-0.0026</td>
<td>-0.0607</td>
<td>2.3389</td>
<td>0.1267</td>
</tr>
<tr>
<td>5</td>
<td>0.0028</td>
<td>-0.0033</td>
<td>0.2035</td>
<td>0.0011</td>
</tr>
<tr>
<td>6</td>
<td>0.0028</td>
<td>-0.0028</td>
<td>0.011</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0.0028</td>
<td>-0.0028</td>
<td>0.0004</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0.0028</td>
<td>-0.0028</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0.0028</td>
<td>-0.0028</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.0028</td>
<td>-0.0028</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

We see from this table that the tail sum is far superior to errors for \( k=2,3,4 \) but then this order reverse and this relation holds no longer. When we compute the difference between the theoretical Laplace Gauss output distribution and the sampled distribution computed with different domains we got the following curves:
Figure 1: addition of two random variables: differences between theoretical and computed PDF, for increasing domains. The error rapidly decreases with domain width and becomes negligible at 10 standard deviations.

The two bumps on these error curves are the image of the convolution of the main lobe by the missing distribution tails. The max value for each curve decreases rapidly with domain extension. We observe that for a domain of $8 \sigma$, a high frequency noise caused by finite arithmetic errors shows up. This proves that from a domain of $6 \sigma$ errors are dominated by the quadrature term. Thus the quadrature method used to compute a probability is very important, and we have devised an adequate upsampling scheme to realize this task.

3.2 Transformation by a monotonic function

3.2.1 Basics of monotonic transforms

Monotonic transformations encompasses following operations:

- Affine transform $Y = aX + b$, and by association with addition, any linear transforms
- Application of analytic functions: exp, log, powers of $x$ (when $x$ is restricted to a domain where $x^y$ is monotonic).

Let $\phi(x)$ be the considered transformation. The PDF of the transformed random variable may be defined through the cumulative distribution function (CDF):

$$F_y(y) = P(Y \leq y) = P(\phi(x) \leq y) = P\left(x \in \phi^{-1}\left(]-\infty, y]\right)\right)$$
That is \( F_y(y) = F_x(\varphi^{-1}(y)) \)

This yields by derivation:
\[
D_y = D_x(\varphi^{-1}(y)) \cdot \frac{d\varphi^{-1}(y)}{dy}
\]

Thus in all these transforms we will find:

- The mapping of the variable domain by the inverse function.
- The normalization of the result by the function derivative.

One can note that this particular transform is also used by some software to produce random generators with arbitrary laws, by transforming a very common constant density random generator. On top of that, importance sampling is also in line with this same concept.

1.1.2 Example of the affine transform

\[
\varphi^{-1}(y) = \frac{y-b}{a} \quad \text{and} \quad \frac{d\varphi^{-1}(y)}{dy} = \frac{1}{a} \quad \text{therefore} \quad D_y = \frac{1}{a} D_x \left( \frac{y-b}{a} \right)
\]

We get \( D_y \) from \( D_x \) by a translation of \( b \) followed by a dilation of \( a \) and then by a scaling by \( 1/a \). Constraints on \( D_x \) are exactly the same than for \( D_y \) except for the dilatation factor, that is to say:

- If \( B_y \) is the bandwidth of \( Y \) the bandwidth of \( X \) is \( B_x = a B_y \)
- If \( m_y \) is the full range (expressed in standard deviates) sought for \( Y \) then \( m_x = m_y \).

If this range is expressed in crude values, then \( e_x = \frac{e_y}{a} \).

1.1.3 Example of the exponential function

We find by using Eq (1) that \( D_y = \frac{1}{y} D_x(\ln(y)) \)

As there is no straightforward relation between bandwidth, one should assert the bandwidth of the transformed PDF in the Fourier domain and loop back if necessary.

We give here an example of this transformation on a Laplace-Gauss distribution, which we know maps onto a Log-normal distribution. Figure 2 show an error peaking at \( 10^{-4} \) which could be reduced by further refining the calculus.
3.3 Product of two independent random variables

Rather than using the standard analytical formulation, which states that the random variable \( Z = X \cdot Y \) has for PDF the function:

\[
\int_\infty^\infty \int_{-\infty}^\infty f_X(t)f_Y\left(\frac{z}{t}\right) \, dt
\]

We prefer to use a log-based scheme, that is to say \( f_z \) is the exponentially transformed PDF of \( \log(X) + \log(Y) \). However it assumes \( X \) and \( Y \) to be strictly positive variables. To avoid such a shortcoming, we implemented the transform for any real valued variables by cutting the variable ranges in three domains:

\([-\infty, 0], [0], \infty\]

Nevertheless, getting a grip on the error implies to use a very dense sampling around zero. This is for now achieved with constant sampling of PDF, at the extent of very large buffers. We are currently investigating the use of variable sampling schemes to avoid this drawback.
3.4 Division of two independent random variables

The division of two independent variables X and Y is the product of X by 1/Y. It is therefore implemented as the sequence of a monotonic transform (1/x) followed by the product of the transformed Y by X. The alternative implementation goes through Log(X) – Log(Y) and is similar to the implementation of the product.

The division brings constraints similar to the product constraints (very thin sampling near zero). The sampling near zero of the divider variable controls directly the tails of the resulting distribution. This show in a certain way that variable sampling for this kind of computation is the counterpart of inverse shaping of the PDFs in importance sampling Monte Carlo schemes: if we would realize the division by importance sampling, the normalized divider would have to exhibit a strong mode at zero.

We validated the division computation by calculating the ratio of two independent Laplace-Gauss distributions N(0,1) and comparing it with the theoretical result, a Cauchy distribution \( f_c \):

\[
f(x) = \frac{1}{\pi(1 + x^2)}
\]

In that case the variable \( Z = X/Y \) is heavy-tailed and has an infinite variance and no mean. Figure 3 show the result of the calculus. This is still a work in progress as the level of error is too high and biased, particularly around the value 0. However, it is a valuable test as it shows that computation of sampled PDF can be achieved even for undefined mean/infinite variance distributions.
Figure 3: Comparison of Cauchy distribution with the calculus of the ration of two \(N(0,1)\) variables.

4. Case study

This computation capability was originally designed to gain insights on the performances of a future optical detector for the French army. The “centre de l’électronique pour l’armement” has developed a simulator for optical systems.

We model an optical system by a series of transformations taking into account physical or electronic interactions with the input signal. From a stochastic point of view, the input signal has a given PDF, theoretical or modeled. Along its path across the system components, there are transforms and combinations with external sources of noise.
The system is build with the following components:

- **MIRE**: creation of a test pattern in visible or near infra red spectra. It can be made of several spectral planes.
- **PUPILLE**: Transformation of luminance \((W.m^{-2}.sr^{-1})\) acquired by the pupil in \((W.m^{-2})\) on the transducer.
- **DIFFRACTION**: diffraction by a round hole (pupil);
- **RESPONSE SPECTRALE**: This procedure apply the transfer gain of the transducer.
- **DETECTEUR**: This stage simulates the geometry of the matrix transducer: projection and resampling of the scene picture on the detector matrix.
- **SENSITIVITY**: this stage transforms the input flow expressed as \(\mu\)Joules in output expressed in e-.
- **PHOTON NOISE**: This stage quantify the noise associated with the distribution of the incident flow of photons. It is a poisson process.
- **DARKNESS NOISE**: Simulates the time noise created by the darkness signal for a given operating temperature. Noise is added to each pixel of the image.
- **CMOS**: The CMOS stage simulates the CMOS conversion to Volts.
- **SCAN**: This stage is the bits quantification stage (digital picture).

The first component MIRE generates a signal with a LogNormal distribution. This is the input signal. The whole process has been simulated with our method on one hand, and with Monte Carlo simulation on the other hand \((2000 \times 512 \times 384)\) realizations. Next plot shows the two results obtained at the output of the whole process:
5. Conclusion

This paper shows that sampled representations of PDFs of independent variables can be propagated through a system graph. Applications range from uncertainty calculation in measurement systems to reliability and to estimation of performances for various types of detectors. Though the current implementation is far from perfect, we have shown that a careful study of each operation will give guidelines for parameters selection, and that appropriate parameters might be selected to match requested error bounds.

This PDF calculus method is implemented in the SignalLAB© software together with graph-oriented editors so as to be able to quickly simulate systems. Links to this software and related materials can be found in [4].

Acknowledgements

We thank Mr Deshors (CElar) for his support during the development of this computation method.

References