Propagation of Probabilistic Uncertainty: The Simplest Case (A Brief Pedagogical Introduction)

Olga Kosheleva[1] and Vladik Kreinovich[2]
[1]Department of Teacher Education
[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

The main objective of this text is to provide a brief introduction to formulas describing the simplest case of propagation of probabilistic uncertainty – for students who have not yet taken a probability course.

1 Need to Take Uncertainty into Account

Measurements and estimates are practically never absolutely accurate: the measurement result $\widetilde{x}$ is, in general, different from the actual (unknown) value x of the corresponding quantity. To increase the accuracy, a natural idea is to perform several measurements and to combine the corresponding measurement results. Thus, we need to handle the following problems:

- Suppose that we have N results $\widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(N)}$ of measuring or estimating the same quantity. We would like to combine these N numbers into a single – ideally more accurate – estimate $\widehat{x}$. What is the best way to combine these estimates? What is the accuracy of the resulting combination?
- Suppose that a difficult-to-directly-estimate quantity y is related to n easier-to-estimate quantities $x_1, \ldots, x_n$ by a relation $y = f(x_1, \ldots, x_n)$, where f is a known algorithm. In this case, if we have estimates $\widehat{x}_1$, $\ldots$, $\widehat{x}_n$ of $x_1, \ldots, x_n$, it is reasonable to estimate y as $\widehat{y} = f(\widehat{x}_1, \ldots, \widehat{x}_n)$. What is the accuracy of this estimate?

2 How to Combine Different Estimates

How to combine: main idea. We have a vector $\widetilde{x} = \left(\widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(N)}\right)$ formed by different estimates $\widetilde{x}^{(i)}$. We know that these estimates estimate the same

values x. If all the measurements were absolutely accurate, we would get a vector $(x, \ldots, x)$ in which all the components are equal to each other.

We want to find a single estimate $\widehat{x}$, i.e., we want to approximate the original vector $\widetilde{x}$ by a vector $(\widehat{x}, \ldots, \widehat{x})$. Out of all such vectors – corresponding to different combined values $\widehat{x}$ – it is reason able to select a vector which is the closest to the original vector $\widetilde{x}$, i.e., for which the distance

$$d((\widehat{x}, \ldots, \widehat{x}), \widetilde{x}) = \sqrt{\sum_{i=1}^{N} \left(\widetilde{x}^{(i)} - \widehat{x}\right)^2}$$

takes the smallest possible value.

From idea to a formula. Minimizing the distance is the same as minimizing its square

$$d^2((\widehat{x}, \ldots, \widehat{x}), \widetilde{x}) = \sum_{i=1}^{N} \left(\widetilde{x}^{(i)} - \widehat{x}\right)^2.$$

To find the value $\widehat{x}$ for which this expression is the smallest possible, we differentiate this expression with respect to $\widehat{x}$ and equate the resulting derivative to 0:

$$\sum_{i=1}^{N} 2 \cdot \left(\widetilde{x}^{(i)} - \widehat{x}\right) \cdot (-1) = 0.$$

If we divide both sides by -2, we get

$$\sum_{i=1}^{N} \left(\widetilde{x}^{(i)} - \widehat{x}\right) = 0.$$

If we move all the terms containing the unknown $\widehat{x}$ to the right-hand side, we get

$$\sum_{i=1}^{N} \widetilde{x}^{(i)} = N \cdot \widehat{x},$$

thus

$$\widehat{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} \widetilde{x}^{(i)} = \frac{\widetilde{x}^{(1)} + \ldots + \widetilde{x}^{(N)}}{N}. \tag{1}$$

So, *a reasonable combination is the arithmetic average of the estimates.* The arithmetic average is also known as the *sample mean* and is denoted by $E\,[\widetilde{x}]$.

Simple properties of sample mean. For a constant $\widetilde{x}^{(i)} = c$, we clearly have

$$E[c] = \frac{c + \ldots + c}{N} = c.$$

If we change the measuring unit to a new one which is λ times smaller – e.g., replace meters with centimeters, in which case $\lambda = 100$ – then all the measured

values multiply by λ, i.e., instead of the original values $\widetilde{x}^{(i)}$, we will get new values $\lambda \cdot \widetilde{x}^{(i)}$. The sample mean of the new values is

$$E\left[\lambda \cdot \widetilde{x}\right] = \frac{1}{N} \cdot \sum_{i=1}^{N} (\lambda \cdot \widetilde{x}^{(i)}).$$

All the terms in this sum have the same factor λ, so we can move λ outside the sum, and thus get

$$E\left[\lambda \cdot \widetilde{x}\right] = \frac{1}{N} \cdot \lambda \cdot \sum_{i=1}^{N} \widetilde{x}^{(i)}.$$

Thus,

$$E\left[\lambda \cdot \widetilde{x}\right] = \lambda \cdot E\left[\widetilde{x}\right]. \tag{2}$$

In some cases, the quantity z of interest is the sum $z = x + y$ of two components x and y: e.g., to find the overall income of a husband-and-wife nuclear family, we can simply add the incomes of both spouses. In this case, if we estimate each of the components N times and add the resulting estimates $\widetilde{x}^{(i)}$ and $\widetilde{y}^{(i)}$, we get N estimates for z: $\widetilde{z}^{(i)} = \widetilde{x}^{(i)} + \widetilde{y}^{(i)}$. The sample mean of $\widetilde{z} = \widetilde{x} + \widetilde{y}$ is equal to

$$E\left[\widetilde{x} + \widetilde{y}\right] = \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\widetilde{x}^{(i)} + \widetilde{y}^{(i)}\right).$$

If we change the order of addition, we get

$$E\left[\widetilde{x} + \widetilde{y}\right] = \frac{1}{N} \cdot \left(\sum_{i=1}^{N} \widetilde{x}^{(i)} + \sum_{i=1}^{N} \widetilde{y}^{(i)}\right),$$

i.e., equivalently,

$$E\left[\widetilde{x} + \widetilde{y}\right] = \frac{1}{N} \cdot \sum_{i=1}^{N} \widetilde{x}^{(i)} + \frac{1}{N} \cdot \sum_{i=1}^{N} \widetilde{y}^{(i)}.$$

Thus, we get

$$E\left[\widetilde{x} + \widetilde{y}\right] = E\left[\widetilde{x}\right] + E\left[\widetilde{y}\right]. \tag{3}$$

Based on (2) and (3), we can find the formula for any linear combination $\widetilde{y} = \sum_{i=1}^{n} c_i \cdot \widetilde{x}_i$: indeed, by (3),

$$E\left[\sum_{i=1}^{n} c_i \cdot \widetilde{x}_i\right] = \sum_{i=1}^{n} E\left[c_i \cdot \widetilde{x}_i\right],$$

and by (2),

$$E\left[c_i \cdot \widetilde{x}_i\right] = c_i \cdot E\left[\widetilde{x}_i\right].$$

Thus,

$$E\left[\sum_{i=1}^{n} c_i \cdot \widetilde{x}_i\right] = \sum_{i=1}^{n} c_i \cdot E\left[\widetilde{x}_i\right]. \tag{4}$$

Case of discrete estimates. In some cases, estimates come from a limited list – e.g., they are student evaluations of faculty, when a student marks one of the four possible grades: 1, 2, 3, and 4 for each of the questions. In general, let us denote possible values of $\widetilde{x}$ by $v_1, \ldots, v_m$. When N is large, we have a large number of equal estimates. Adding several identical estimates is not the most efficient thing to do: e.g., instead of adding 4 to itself 25 times, it makes more sense to multiply 4 by 25 – this is what multiplication is about.

If we re-order the sum in the formula (1) by first placing all values equal to v_1, then all the values equal to v_2, etc., we get

$$E\left[\widetilde{x}\right] = \frac{1}{N} \cdot ((v_1 + \ldots + v_1)(N_1 \text{ times}) + \ldots + (v_m + \ldots + v_m)(N_m \text{ times})) =$$

$$\frac{N_1 \cdot v_1 + \ldots + N_m \cdot v_m}{N},$$

where N_j is the number of times the value v_j appeared. This formula can be re-written as

$$E\left[\widetilde{x}\right] = p_1 \cdot v_1 + \ldots + p_m \cdot v_m, \tag{5}$$

where we denoted

$$p_j = \frac{N_j}{N}. \tag{6}$$

The ratio p_j is the frequency with which we encounter the value v_j – i.e., in effect, the probability of encountering this value.

3 What Is the Accuracy of the Sample Mean?

A seemingly natural idea. When we looked for an appropriate combination, we look for the value $\widehat{x}$ for which the distance between the original vector $\widetilde{x}$ and the approximating vector $(\widehat{x}, \ldots, \widehat{x})$ is the smallest possible. It is therefore reasonable to take the resulting smallest distance as the desired measure of accuracy of the combined estimate $\widehat{x}$.

Limitations of the seemingly natural idea. The problem with the above idea is that the resulting estimate depends not only on accurate the measurements are, but also on how many estimates we had. Indeed, if we repeat the same N estimates again and get the same values, we get the same sample mean with the same accuracy, but the square of the distance increases by the factor of 2 – since each term $\left(\widetilde{x}^{(i)} - \widehat{x}\right)^2$ is now repeated twice. If we repeat all estimates k times, the square of the distance increases by a factor of k.

Modified idea and the resulting formulas. To eliminate the above-mentioned undesired dependence, it is reasonable to divide the square of the

distance by N – or, equivalently, to divide the distance itself by $\sqrt{N}$. Thus divided distance is known as the *sample standard deviation* $\sigma\,[\widetilde{x}]$, and its square as the *sample variance* $V\,[\widetilde{x}]$:

$$V\,[\widetilde{x}] = \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\widetilde{x}^{(i)} - E\,[\widetilde{x}]\right)^2; \tag{7}$$

$$\sigma\,[\widetilde{x}] = \sqrt{V\,[\widetilde{x}]}. \tag{8}$$

By using the notation for sample mean, we equivalently describe the formula (7) as

$$V\,[\widetilde{x}] = E\left[(\widetilde{x} - E\,[\widetilde{y}])^2\right]. \tag{7a}$$

Comment. Sometimes, practitioners use similar formulas with $N - 1$ instead of N. For large N, this practically does not affect the numerical values, but it helps to avoid a counter-intuitive conclusion that one can make based on the formula (7) – that when we have only one estimate $N = 1$, it leads to the perfect accuracy $V = \sigma = 0$.

Simple properties of sample standard deviation and sample variance. If we multiply all the estimates $\widetilde{x}^{(i)}$ by λ, then, as we have shown, the sample mean $E\,[\widetilde{x}]$ also gets multiplied by λ. Thus, all the differences $\widetilde{x}^{(i)} - E\,[\widetilde{x}]$ in the formula (7) are multiplied by λ, hence their squares are multiplied by λ^2, and we get

$$V\,[\lambda \cdot \widetilde{x}] = \lambda^2 \cdot V\,[\widetilde{x}]; \tag{9}$$

$$\sigma\,[\lambda \cdot \widetilde{x}] = \lambda \cdot \sigma\,[\widetilde{x}]. \tag{10}$$

Comment. To deal with the sample variance of the sum, we need to recall the notion of independence.

4 The Notion of Independence and the Resulting Formulas for the Variance of the Sum and of a Linear Combination

What is independence? From the commonsense viewpoint, independence of two events A and B means that the frequency of satisfying the property A in the general population is the same as the frequency of satisfying the property A among all the objects that satisfy the property B. For example, independence of good grades on student gender means that the proportion of, e.g., straight-A students in the general student population is the same as the proportion of straight-A students among all female students.

An example where there is no independence is if we consider A to be the property to be a basketball player and B to be tall (to be more precise, taller

than a certain amount). In this case, clearly, if we limit ourselves to objects that satisfy the property B, the proportion of objects that satisfy the property A increases.

Let us describe independence in precise terms.

Towards a precise definition of independence. Let N be the overall number of objects, let $N(A)$ denote the number of objects that satisfy the property A, $N(B)$ the number of objects that satisfy the property B, and $N(A\,\&\,B)$ the number of objects that satisfy both properties. We can then estimate the probabilities $P(A)$, $P(B)$, and $P(A\,\&\,B)$ of the events A, B, and $A\,\&\,B$ as the corresponding ratios:

$$P(A) = \frac{N(A)}{N}, \quad P(B) = \frac{N(B)}{N}, \quad P(A\,\&\,B) = \frac{N(A\,\&\,B)}{N}. \tag{11}$$

We can also define the *conditional probability* of the event A under the condition B as

$$P(A\,|\,B) \stackrel{\text{def}}{=} \frac{N(A\,\&\,B)}{N(B)}. \tag{12}$$

If we divide both numerator and denominator in the formula (12) by N, we conclude that

$$P(A\,|\,B) \stackrel{\text{def}}{=} \frac{P(A\,\&\,B)}{P(B)}. \tag{13}$$

In these terms, independence means that

$$P(A) = P(A\,|\,B). \tag{14}$$

How independence affects the probability of a joint event. Substituting the expression (13) into the definition (14) of independence, we get $P(A) = \dfrac{P(A\,\&\,B)}{P(B)}$. Multiplying both sides by $P(B)$, we get

$$P(A\,\&\,B) = P(A) \cdot P(B). \tag{15}$$

This makes perfect sense: this is how we estimate, e.g., the probability 0.25 that a coin falls heads twice in a row – by multiplying the probability 0.5 that it falls head the first time and the probability 0.5 that it falls head the second time.

What is the sample mean of the product of two independent quantities? Suppose that we have a quantity x with possible values $v_1, \ldots, v_m$ and a quantity y with possible values $w_1, \ldots, w_\ell$. For the first quantity, we have m possible events $x = v_j$; let $p_j = P(x = v_j)$ denote their probabilities. For the second quantity, we have ℓ possible events $y = w_k$; let $q_k = P(y = w_k)$ denote the corresponding probabilities.

If the quantities are independent, this means that all these events are independent. Thus, for all possible j and k, we have

$$P(x = v_j\,\&\,y = w_k) = P(x = v_j) \cdot P(y = w_k) = p_j \cdot q_k.$$

The product $z = x \cdot y$ gets all possible values $v_j \cdot w_k$ with probability $p_j \cdot q_k$, so

$$E[x \cdot y] = \sum_{j=1}^{m} \sum_{k=1}^{\ell} P(v_j \cdot w_k) \cdot v_j \cdot w_k = \sum_{j=1}^{m} \sum_{k=1}^{\ell} p_j \cdot q_k \cdot v_j \cdot w_k.$$

One can easily see that this product is the product of two products:

$$E[x \cdot y] = \left(\sum_{j=1}^{m} p_j \cdot v_j \right) \cdot \left(\sum_{k=1}^{\ell} q_k \cdot w_k \right),$$

hence

$$E[x \cdot y] = E(x] \cdot E[y]. \tag{16}$$

What is the variance of the sum of two independent random variables? Let us assume that the variables x and y are independent. We want to estimate the variance of $z = x + y$. By definition of the variance (formula (7a)), we get

$$V[x + y] = E\left[((x + y) - E[x + y])^2 \right].$$

We already know that $E[x + y] = E[x] + E[y]$, thus

$$(x + y) - E[x + y] = x + y - (E[x] + E[y]) = (x - E[x]) + (y - E[y]).$$

Thus,

$$((x + y) - E[x + y])^2 = ((x - E[x]) + (y - E[y]))^2 =$$
$$(x - E[x])^2 + (y - E[y])^2 + 2\,(x - E[x]) \cdot (y - E[y]).$$

We know that the mean of the sum is equal to the sum of the means, so

$$V[x + y] = E\left[((x + y) - E[x + y])^2 \right] =$$

$$E\left[(x - E[x])^2 \right] + E\left[(y - E[y])^2 \right] + 2 \cdot E\left[(x - E[x]) \cdot (y - E[y]) \right].$$

The first two terms in the right-hand side are simply $V[x]$ and $V[y]$. To compute the last term, we take into account that x and y are independent, thus

$$E\left[(x - E[x]) \cdot (y - E[y]) \right] = E\left[x - E[x] \right] \cdot E\left[y - E[y] \right].$$

Here,

$$E\left[x - E[x] \right] = E[x] - E[x] = 0,$$

thus the whole product is 0, and we conclude that

$$V[x + y] = V[x] + V[y]. \tag{17}$$

What is the variance of a linear combination of independent random variables? For $y = \sum\limits_{i=1}^{n} c_i \cdot x_i$, we get $V[c_i \cdot x_i] = c_i^2 \cdot V[x_i]$, hence, by (17),

$$V[y] = V\left[\sum_{i=1}^{n} c_i \cdot x_i\right] = \sum_{i=1}^{n} c_i^2 \cdot V[x_i] = \sum_{i=1}^{n} c_i^2 \cdot \sigma_i^2, \tag{18}$$

where we denoted $\sigma_i \stackrel{\text{def}}{=} \sigma[x_i]$. This implies that

$$\sigma[y] = \sqrt{\sum_{i=1}^{n} c_i^2 \cdot \sigma_i^2}. \tag{19}$$

Comment 1. What if we have $y = f(x_1, \ldots, x_n)$ for some non-linear function? We have estimates $\widehat{x}_i$ for the corresponding quantities x_i. Based on these estimates, we can compute the estimate for y: $\widehat{y} = f(\widehat{x}_1, \ldots, \widehat{x}_n)$. How accurate is this estimate? In other words, how big is the difference

$$\Delta y \stackrel{\text{def}}{=} \widehat{y} - y = f(\widehat{x}_1, \ldots, \widehat{x}_n) - f(x_1, \ldots, x_n).$$

We know that that the estimates $\widehat{x}_i$ are not absolutely exact, so there is an estimation error $\Delta x_i \stackrel{\text{def}}{=} \widehat{x}_i - x_i$. Because of this definition, we have $x_i = \widehat{x}_i - \Delta x_i$ and thus,

$$\Delta y = f(\widehat{x}_1, \ldots, \widehat{x}_n) - f(\widehat{x}_1 - \Delta x_1, \ldots, \widehat{x}_n - \Delta x_n).$$

Usually, estimates $\widetilde{x}_i$ for the quantities x_i are reasonable accurate, so the differences $\Delta x_i = \widehat{x}_i - x_i$ is reasonably small. Thus, if we expand the above expression in Taylor series, we can keep only terms which are linear in Δx_i and safely ignore terms which are quadratic or higher order; as a result, we get the following expression:

$$\Delta y = \sum_{i=1}^{n} c_i \cdot \Delta x_i, \tag{20}$$

where

$$c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}. \tag{21}$$

So, if we know the sample standard deviation σ_i os each of the estimates, we can use the formula (19) to estimate the standard deviation $\sigma[\Delta y]$ as

$$\sigma[\Delta y] = \sqrt{\sum_{i=1}^{n} c_i^2 \cdot \sigma_i^2} = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \cdot \sigma_i^2}. \tag{22}$$

Comment 2. What is the variables x_i are not independent? In this case, similar arguments leads to

$$\sigma[y] = \sqrt{\sum_{i=1}^{n} c_j^2 \cdot \sigma_j^2 + \sum_{j \neq k} c_j \cdot c_k \cdot \mathrm{cov}(x_j, x_k) \cdot \sigma_j \cdot \sigma_k}, \tag{23}$$

where we denoted

$$\mathrm{cov}(x_j, x_k) \stackrel{\mathrm{def}}{=} E[(x_j - E[x_j]) \cdot (x_k - E[x_k])] =$$

$$\frac{1}{N} \cdot \sum_{k=1}^{N} \left(x_j^{(i)} - E[x_j] \right) \cdot \left(x_j^{(i)} - E[x_j] \right). \tag{24}$$

This expression $\mathrm{cov}(x_j, x_k)$ is known as *sample covariance.*

Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.