Distribution function of generalized concentration

V. HERLES

Institute of Information Theory and Automation,
Czechoslovak Academy of Sciences, 182 08 Prague

Received 29 May 1980

The distribution function and probability density of the function $z = x/(x+y)$ where x, y are independent random variables are derived. As two special cases the normal and uniform distributions of x, y are analyzed. In both cases the programs for TI 59 are available.

Была получена функция распределения и плотности вероятности для функции $z = x/x + y$, где x и y независимые случайные величины. В качестве двух частных случаев анализируется нормальное и равномерное распределение x, y. Для обоих случаев были разработаны программы для TI 59.

The large spread of automation and optimization techniques in modern chemical industry necessarily needs more information about the actual state of technological processes. The implementation of these tendencies necessitates not only a suitable hardware equipment but sometimes a "complex" theoretical analysis of the "whole" problem may be much more important and effective.

Bearing in mind the expected aim of the solved problem, this analysis should include a critical comparison and individual choice of a suitable theoretical solution taking into account special technological conditions and constrains of either technical or economical nature (*e.g.* limited precision of measured information and its influence upon a reliability of a given theoretical algorithm, *etc.*). Not always the higher instrumentation level or application of a promising theoretical method must necessarily yield a better practical result. But on the other hand, an unusual solution may sometimes exhibit very interesting advantages.

As an example there will be briefly described an actual problem. The results of the complex technological analysis pointed out that the quality and homogeneity of the output chemical product is closely dependent on the concentration of the entering gas mixture. Fluctuations of this value have been depending mainly on randomly varying flow rates of both (pure) gases at the input of the mixing element.

The recently provided laboratory analysis of samples of the mixture did not assure the promptness sufficient for a control. Because of technical difficulties in installation of a direct measuring equipment, the problem was analyzed at first theoretically. The aim of the analysis was to specify the band limits of input flow rates assuring the acceptable tolerances of output concentration.

Solution

In the mathematical formulation the problem necessitates to find out the distribution function of a random variable

$$z = \frac{x}{x+y} \tag{1}$$

where x, y are independent random variables with given distribution functions or probability densities.

The usual procedure for finding the distribution function of a function of n variables

$$z = f(x_1, \dots x_n) \tag{2}$$

consists in solving the n-dimensional integral

$$\mathrm{F}(z) = \iint_D \dots \int p_{1,\dots n}(x_1, \dots x_n)\, \mathrm{d}x_1, \dots \mathrm{d}x_n \tag{3}$$

where $p(.)$ is the n-dimensional probability density of random variables $x_1, \dots x_n$ and the region of integration

$$D = \{\zeta : \zeta < z\} \tag{4}$$

In our case with two variables

$$\mathrm{F}(z) = \iint_D p_{xy}(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{5}$$

The probability density is given

$$p(z) = \frac{\partial}{\partial z} \iint_D p_{xy}(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{6}$$

The integration region D (Fig. 1) is a sector in the plane (x, y) between the straight lines $x = -y$ $(z = -\infty)$ and $(z = \text{const})$

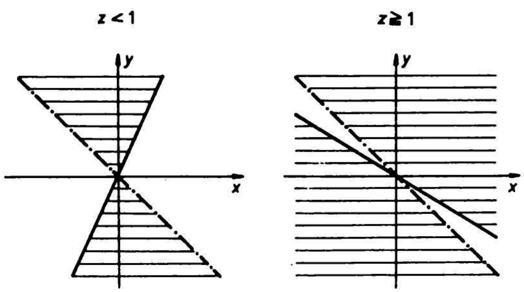

Fig. 1. Integration region D.

$$x = \frac{z}{1-z}\, y \tag{7}$$

The physical origin of the problem implies the restrictions $x, y \geqq 0$; $0 \leqq z \leqq 1$, but the theoretical solution can be easily extended to all real values of x, y, z.

For independent variables x, y the integral (*5*) equals

$z<1$

$$\mathrm{F}(z) = \int_0^\infty p_y(y) \int_{-y}^{(z/1-z)\,y} p_x(x)\,\mathrm{d}x\,\mathrm{d}y + \\ + \int_{-\infty}^{0} p_y(y) \int_{(z/1-z)\,y}^{-y} p_x(x)\,\mathrm{d}x\,\mathrm{d}y \tag{8}$$

$z \geqq 1$

$$\mathrm{F}(z) = \int_0^\infty p_y(y) \left\{ \int_{-\infty}^{(z/1-z)\,y} p_x(x)\,\mathrm{d}x + \int_{-y}^{\infty} p_x(x)\,\mathrm{d}x \right\} \mathrm{d}y + \\ + \int_{-\infty}^{0} p_y(y) \left\{ \int_{-\infty}^{-y} p_x(x)\,\mathrm{d}x + \int_{(z/1-z)\,y}^{\infty} p_x(x)\,\mathrm{d}x \right\} \mathrm{d}y \tag{9}$$

If $p(.)$ are continuous and limited in the whole integration region, the probability density (*6*) for real z in both cases (*8*) and (*9*) equals

$$p_z(z) = \frac{1}{(1-z)^2} \int_0^\infty y \left[p_x\left(\frac{yz}{1-z}\right) p_y(y) + \\ + p_x\left(-\frac{yz}{1-z}\right) p_y(-y) \right] \mathrm{d}y \tag{10}$$

Special cases

Normal distribution

If x, y are normal independent random variables with parameters $(\bar{x}, \sigma_x)$; $(\bar{y}, \sigma_y)$ (or may be so approximated), eqn (*10*) can be solved analytically

$$p_z(z) = \frac{1}{2\pi\sigma_x\sigma_y(1-z)^2} \int_0^\infty y\left(e^{E_1} + e^{E_2}\right) \mathrm{d}y \tag{11}$$

where

$$E_{1,2} = -\frac{1}{2\sigma_x^2}\left(\frac{yz}{1-z} \mp \bar{x}\right)^2 - \frac{1}{2\sigma_y^2}(y \mp \bar{y})^2 \tag{12}$$

$$p_z(z) = \frac{1}{2\pi\sigma_x\sigma_y(1-z)^2} \exp\left\{ -\frac{[z\bar{y} - (1-z)\,\bar{x}]^2}{2[\sigma_x^2(1-z)^2 + \sigma_y^2 z^2]} \right\}.$$

$$\cdot\int_0^{\infty} y\,[e^{-\frac{(y-B)^2}{2A^2}} + e^{-\frac{(y+B)^2}{2A^2}}]\,\mathrm{d}y \tag{13}$$

where

$$A^2 = \frac{\sigma_x^2\sigma_y^2(1-z)^2}{\sigma_x^2(1-z)^2+\sigma_y^2z^2} \tag{14}$$

$$B = (1-z)\,\frac{[\sigma_x^2\bar{y}(1-z)+\sigma_y^2\bar{x}z]}{[\sigma_x^2(1-z)^2+\sigma_y^2z^2]} \tag{15}$$

Integrals of the type

$$I_{1,2} = \int_0^{\infty} y\,e^{\frac{(y\mp B)^2}{2A^2}}\,\mathrm{d}y \tag{16}$$

can be solved by substitutions $u = y \mp B$.

$$I_{1,2} = \int_{\mp B}^{\infty} u\,e^{-\frac{u^2}{2A^2}}\,\mathrm{d}u \pm B\int_{\mp B}^{\infty} e^{-\frac{u^2}{2A^2}}\,\mathrm{d}u \tag{17}$$

and then $w = u^2$ in the first and $u = Av$ in the second integral

$$I_{1,2} = A^2e^{-\frac{B^2}{2A^2}} \pm AB\int_{\mp(B/A)}^{\infty} e^{-\frac{v^2}{2}}\,\mathrm{d}v \tag{18}$$

The expression (*13*) contains the sum of I_1 and I_2

$$I_1 + I_2 = 2A^2e^{-\frac{B^2}{2A^2}} + 2AB\int_0^{B/A} e^{-\frac{v^2}{2}}\,\mathrm{d}v \tag{19}$$

The final probability density is then equal

$$p_z(z) = \frac{\sigma_x\sigma_y}{\pi[\sigma_x^2(1-z)^2+\sigma_y^2z^2]}\left[e^{-\frac{C^2}{2}} + C\int_0^{C} e^{-\frac{v^2}{2}}\,\mathrm{d}v\right]\cdot$$

$$\cdot\exp\left\{-\frac{[z\bar{y}-(1-z)\,\bar{x}]^2}{2[\sigma_x^2(1-z)^2+\sigma_y^2z^2]}\right\} \tag{20}$$

where

$$C = \frac{[\sigma_x^2\bar{y}(1-z)+\sigma_y^2\bar{x}z]}{\sigma_x\sigma_y\sqrt{\sigma_x^2(1-z)^2+\sigma_y^2z^2}} \tag{21}$$

As it has been shown in [1] the mean and standard deviation of random variable z equals

$$\bar{z} \doteq \frac{\bar{x}}{\bar{x}+\bar{y}} \tag{22}$$

$$\sigma_z \doteq \bar{z}(1-\bar{z}) \sqrt{\left(\frac{\sigma_x}{\bar{x}}\right)^2 + \left(\frac{\sigma_y}{\bar{y}}\right)^2} \tag{23}$$

Fig. 2 shows three examples with different mean values $\bar{z}$ (Table 1) but all with the same relative standard deviations of x and y

$$\frac{\sigma_x}{\bar{x}} = \frac{\sigma_y}{\bar{y}} = 0.1$$

Table 1

Means and standard deviations of examples in Figs. 2 and 3

$\bar{z}$	0.1	0.5	0.9
σ_z	0.0127	0.0354	0.0127
$\sigma_z/\bar{z}$	0.125	0.0707	0.0141

The distribution functions (Fig. 3) for the same cases are plotted on the probabilistic paper for normal distribution. The curve for $\bar{z} = 0.5$ approaches a straight line. In this special case the distribution function can be approximated by the normal law with parameters (*22*) and (*23*).

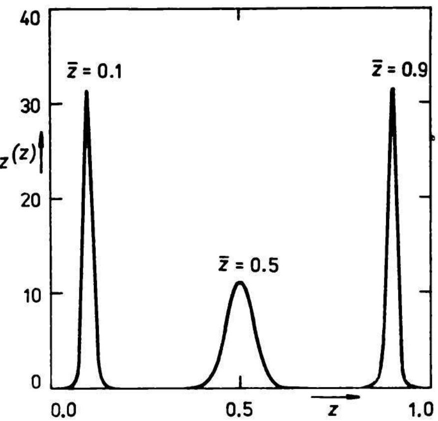

Fig. 2. Probability densities of concentration $z = x/(x+y)$ for independent normal random variables x, y with relative standard deviations $\sigma_x/\bar{x} = \sigma_y/\bar{y} = 0.1$.

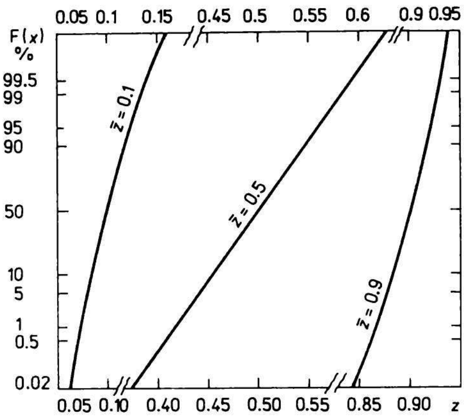

Fig. 3. Distribution functions of concentration for the same cases as in Fig. 2 (plotted on probabilistic paper for normal distribution).

The complete program for computing the probability density and distribution function on calculator TI 59 is available at the author.

Uniform distribution

If the independent random variables x, y have uniform joint distribution in the region $\bar{x} \pm b_x$; $\bar{y} \pm b_y$, the distribution function of $F_z(z)$ can be obtained by the integration of the truncated area of nonzero region of $p_{xy}(x, y)$ within the integration region according to Fig. 1.

Program for the numerical solution consists in finding all vertices of that polygon for the given z and in computing its area according to the formula

$$P = \frac{1}{2}\left(\begin{vmatrix} x_n & x_1 \\ y_n & y_1 \end{vmatrix} + \sum_{i=1}^{n-1} \begin{vmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{vmatrix}\right) \tag{24}$$

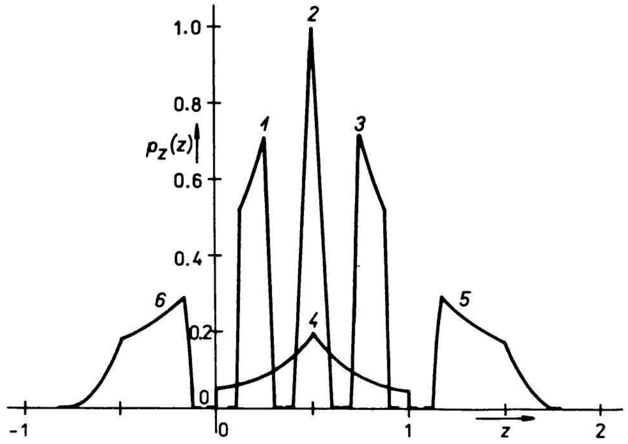

Fig. 4. Probability densities of random variable $z = x/(x+y)$ for independent random variables x, y with uniform distribution and following parameters

Example	*1*	*2*	*3*	*4a*	*4b*	*5*	*6*
$\bar{x}$	0.2	0.5	0.8	0.1	0.5	0.8	−0.2
$\bar{y}$	0.8	0.5	0.2	0.1	0.5	−0.2	0.8
$b_x = b_y$	0.1	0.1	0.1	0.1	0.5	0.1	0.1
$\bar{z}$	0.2	0.5	0.8	0.5	0.5	1.3	−0.3

$$F_z(z)=\frac{1}{4b_xb_y}|P| \qquad (25)$$

The vertices (x_i y_i) must be arranged successively according to the circulation along the perimeter of the polygon. The detailed program for calculators Texas Instruments TI 59, is available at the author.

Fig. 4 demonstrates the probability densities of random variable z for several combinations of x and y. In almost all cases (with only one exception)

$$b_x=b_y=0.1$$

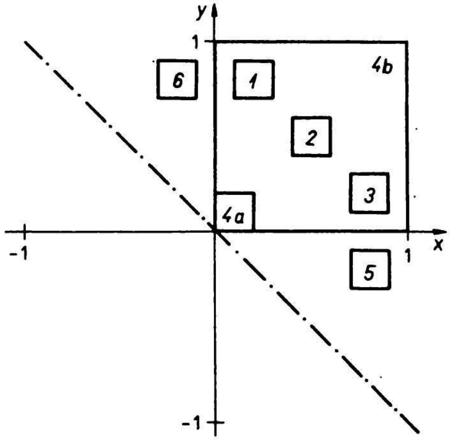

Fig. 5. Regions of uniform joint probability densities for examples in Fig. 4.

The regions of nonzero joint probability densities of $p_{xy}(x, y)$ for the demonstrated cases are shown in Fig. 5.

References

1. Herles, V., *Chem. Listy 75* (1981).

Translated by V. Herles