Risk Analysis, Vol. 7, No. 1, 1987

An Improved Condensation Procedure in Discrete Probability Distribution Calculations

Stan Kaplan[1] **and James C. Lin**[1]

Received December 10, 1985; revised April 16, 1986

A "vertical" condensation scheme for discrete probability distribution (DPD) calculations is presented as an alternative to the earlier "horizontal" scheme, an example of which was presented recently by Kurth and Cox. When applied to DPDs over a space of curves, the vertical condensation results in a "regularization" of the "spaghetti" of curves that results from combination operations on such DPDs.

KEY WORDS: Probability; discrete probability distributions; condensation.

1. INTRODUCTION

A recent paper by Kurth and Cox[1] compared the use of DPD and Monte Carlo methods in computing the size of a crack growing according to the formula

$$\frac{da}{dN} = C\sigma^4\Pi^2a^2 \tag{1}$$

where a = the crack size,
N = the number of stress cycles,
σ = the far field stress, and,
C = an empirical constant.

In their calculations, Kurth and Cox included probabilistic representations of their uncertainty in C and in the initial size, $a^{(0)}$, of the crack. The σ was taken to be a true random or fluctuating variable whose value changes from cycle to cycle. Because of this, the numer of doublets in their DPD increased every cycle, and they required a condensation procedure to reduce this number at each step.

Thus, for fixed C, if the DPDs for σ

$$\{\langle r_j, \sigma_j\rangle\} \tag{2}$$

and for $a^{(n)}$

$$\{\langle p_i, a_i^{(n)}\rangle\} \tag{3}$$

each had 20 doublets, the DPD for $a^{(n+1)}$

$$\{\langle p_{ij}, a_{ij}^{(n+1)}\rangle\} \tag{4a}$$

$$p_{ij} = p_i r_j \tag{4b}$$

$$a_{ij}^{(n+1)} = a_i^{(n)} + C\sigma_j^4\Pi^2\left[a_i^{(n)}\right]^2 \tag{4c}$$

would have 400 doublets. Kurth and Cox therefore condensed these 400 down to 20 by establishing 20 bins on the a axis. An interesting feature was that these bins were calculated anew after each cycle to keep up with the growing crack.

This method of condensation is an example of what could be called "condensation along the horizontal axis"; that is, establishing bins along the axis of the unknown variable, in this case crack size a. This is the kind of condensation we described in Ref.

[1] Pickard, Lowe and Garrick, Inc., 2260 University Drive, Newport Beach, California 92660.

0272-4332/87/0300-0015$05.00/1 © 1987 Society for Risk Analysis

2. Shortly after that paper was written, we began using a condensation procedure based on the vertical axis (i.e., the probability axis). We have found this procedure to be superior in most cases and the purpose of this paper is to make it more widely known.

2. VERTICAL CONDENSATION

The best way to explain the vertical condensation procedure is with an example. Suppose, then, we have a 15-doublet DPD

$$\{\langle p_i, x_i \rangle\} \tag{5}$$

that we wish to condensate down to 6 doublets.

Let us write the original DPD in a "fishbone diagram" (Scheme I) to which we have also added the cumulative probability, P_i.

These cumulative probabilities are plotted against the x_i in Fig. 1.

Now imagine that we have defined six bins against the vertical axis as shown. The first of these extends from $P = 0$ to $P = .10$, the second from $P = .10$ to $P = .20$, the third from $P = .20$ to $P = .40$, etc. The first two bins, therefore, contain 10% probability, and the remaining four each contain 20% probability.

To obtain a condensed DPD according to this vertical binning, we simply proceed as follows. Let the condensed DPD be denoted

$$\{\langle \tilde{p}_j, \tilde{x}_j \rangle\} \tag{6}$$

Then, since the first three p_i fit exactly the first vertical bin, we set

$$\tilde{p}_i = p_1 + p_2 + p_3 = 0.1 \tag{7a}$$

$$\tilde{x}_1 = \frac{1}{\tilde{p}_1}(p_1x_1 + p_2x_2 + p_3x_3) = \frac{1}{0.1}(.23) = 2.3 \tag{7b}$$

x_i	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0
p_i	.02	.03	.05	.06	.10	.09	.10	.05	.20	.05	.10	.05	.06	.02	.02

P_i 0 .02 .05 .10 .16 .26 .35 .45 .50 .70 .75 .85 .90 .96 .98 1.0

Scheme I.

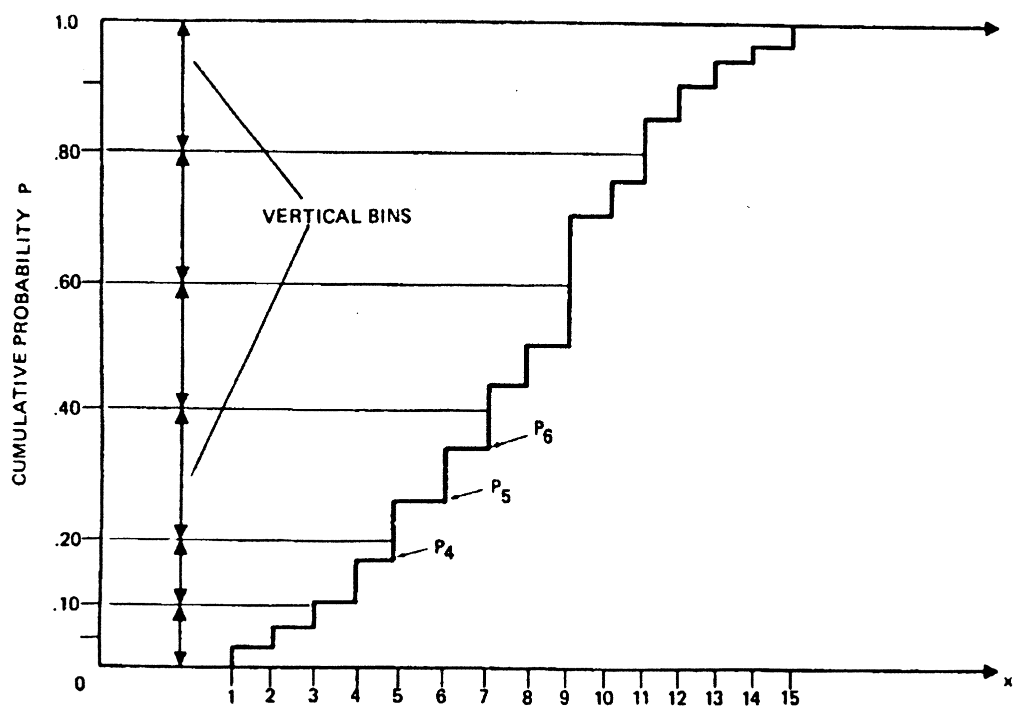

Fig. 1. Cumulative DPD.

The doublet $\langle p_4, x_4 \rangle$ fits entirely within the second vertical bin, but the point $\langle p_5, x_5 \rangle$ must be apportioned between the second and third vertical bins according to the ratio

$$\frac{.20 - P_4}{P_5 - P_4} = \frac{.20 - .16}{.26 - .16} = .40 \tag{8a}$$

Thus,

$$\tilde{p}_2 = p_4 + .40 p_5 = .06 + (.40)(.10) = .10 \tag{8b}$$

$$\tilde{x}_2 = \frac{1}{\tilde{p}_2}[p_4 x_4 + (.40)(.10) x_5]$$

$$= \frac{1}{.10}[(.06)(4.0) + (.40)(.10)(5.0)] = 4.4 \tag{8c}$$

$$\tilde{p}_3 = (0.6) p_5 + p_6 + (0.5) p_7 = .06 + .09 + .05 = .20 \tag{9a}$$

$$\tilde{x}_3 = \frac{1}{.20}[(0.6) p_5 x_5 + p_6 x_6 + (0.5) p_7 x_7]$$

$$= \frac{1}{.20}[(.06)5 + (.09)6 + (.05)7] = 5.95 \tag{9b}$$

Proceeding similarly,

$$\tilde{x}_4 = \frac{1}{.20}[(.05)7 + (.05)8 + (.10)9] = 8.25 \tag{10}$$

$$\tilde{x}_5 = \frac{1}{.20}[(.10)9 + (.05)10 + (.05)11] = 9.75 \tag{11}$$

$$\tilde{x}_6 = \frac{1}{.20}[(.05)11 + (.05)12 + (.06)13 + (.02)14 + (.02)15] = 12.55 \tag{12}$$

Thus, we have the new DPD (Scheme II)

$\tilde{x}_i$	2.3	4.4	5.95	8.25	9.75	12.55
$\tilde{p}_i$	0.1	0.1	.20	.20	.20	.20

(13)

Scheme II.

Observe that the mean value of both the new and old DPD is

$$\bar{x} = 7.97 \tag{14}$$

Thus, this condensation procedure has the virtue of being "mean preserving." By selecting the vertical bins appropriately, it is always possible to retain the features of interest in the condensed curve. For example, in the crack growth example most interest centers on the high side tail of the $p(a^N)$ curve. We could thus choose small bins in the upper probability range to get fine detail in the high tail. Similarly, small bins at both ends (e.g., .99 to 1.0 and 0 to 0.01) would retain the "spread" of the distribution, and so on.

Thus, the vertical condensation provides better control of the process and avoids the problems of selecting horizontal bins discussed by Kurth and Cox in their Sec. 3.

Applied to the crack growth problem with a fixed set of vertical bins, this condensation method yields a set of discrete crack sizes, a_i, moving smoothly in time. That is, it yields a DPD of the form

$$\{\langle p_i, a_i(N) \rangle\}$$

which can be plotted conveniently as shown in Fig. 2 (see Ref. 3 for similar crack growth curves in turbine rotors).

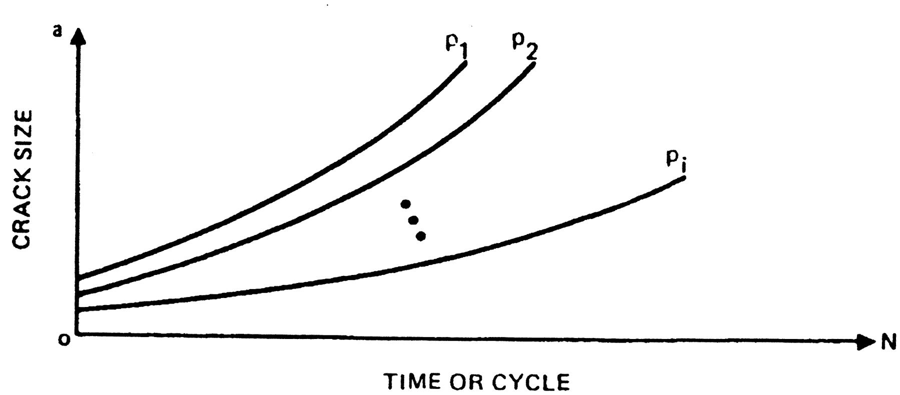

Fig. 2. Plot of the DPD $\{\langle p_i, a_i(N) \rangle\}$ showing crack growth vs. time.

3. APPLICATION TO DPDs ON A SPACE OF CURVES—DEFINITION OF "SPAGHETTIS" OF CURVES

In the above, we described the application of the vertical condensation procedure to DPDs over a "scalar" space (i.e., where the unknown x is an ordinary scalar variable). In many applications, we find it necessary to do operations with probability distributions over a space of curves. For example, in seismic risk assessment,[4] we deal with DPDs of the type

$$F^1 = \{\langle p_i, F_i^1 \rangle\} \tag{15}$$

$$F^2 = \{\langle q_j, F_j^2 \rangle\} \tag{16}$$

where the F_i^1 and F_j^2 are fragility curves against the variable a, representing ground acceleration. The sets F^1 and F^2, thus, might represent fragility families for two different components, 1 and 2. These might appear as shown in Fig. 3 for example.

Now, suppose we combine the two components through an "OR" gate.

$$\text{Ⓢ} = \text{①} \vee \text{②} \tag{17}$$

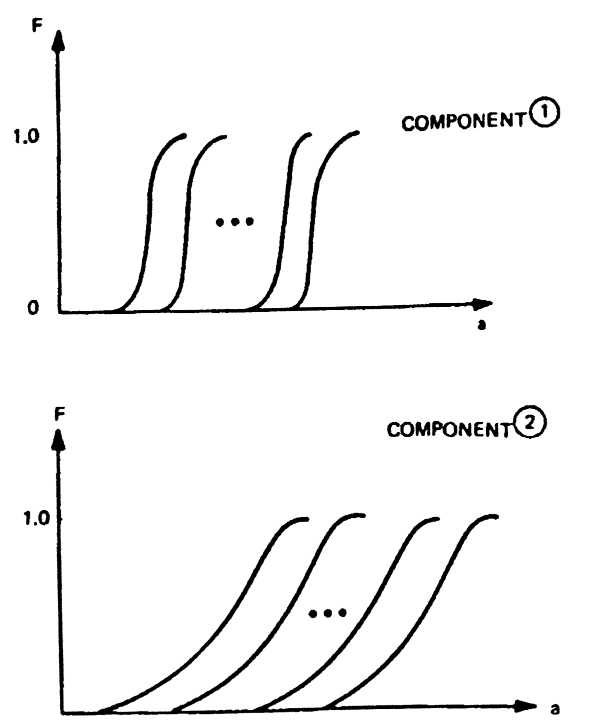

Fig. 3. Fragility families for two components.

then,

$$F^s = \{\langle p_{ij}^s, F_{ij}^s \rangle\} \tag{18a}$$

where

$$p_{ij}^s = p_i q_j \tag{18b}$$

$$\begin{aligned} F_{ij}^s &= F_i^1(a) \vee F_j^2(a) \\ &= F_i^1(a) + [1 - F_i^1(a)] F_j^2(a) \end{aligned} \tag{18c}$$

We would then obtain a DPD looking like Fig. 4. We call such a family of curves a "spaghetti" of curves for obvious reasons. If we now wanted to combine this family with further components through further "AND" gates and "OR" gates, the spaghetti would get more and more complicated.

Obviously, then a condensation process is needed. To do this simply, imagine that we have discretized the a axis into discrete accelerations, a_k. Now, imagine cutting the family in Fig. 4 with a vertical line at a_k. We would obtain a set of ordinates

$$F_{ij}^s(a_k) \tag{19}$$

Each of these would come from a curve with probability p_{ij}^s. Thus, at a_k, we have the ordinary scalar DPD

$$\{\langle p_{ij}^s, F_{ij}^s(a_k) \rangle\} \tag{20}$$

Suppose we now put this DPD through the vertical condensation process. We then obtain

$$\{\langle \tilde{p}_i^s, \tilde{F}_i^s(a_k) \rangle\} \tag{21}$$

Fig. 4. Fragility family for the combination ① ∨ ② = Ⓢ.

If we do this for each a_k, keeping the same vertical binning structure, we may then connect the points at different a_k for the same i. This leads us to a family having the appearance of Fig. 5, which we may refer to as a "regularized" spaghetti. Thus, the vertical condensation process applied to a spaghetti of curves regularizes the spaghetti. This condensation process can then be applied after each combination operation for curves, just as it is applied for scalars.

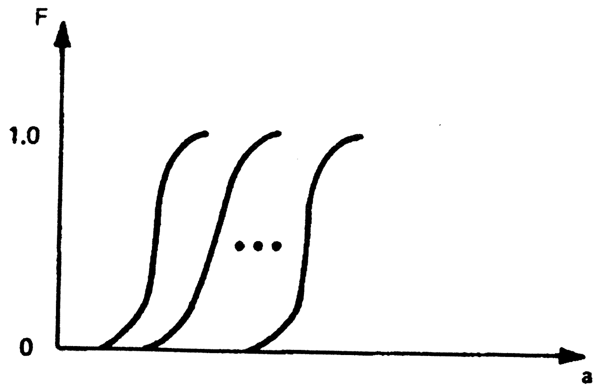

Fig. 5. Regularized spaghetti for the combination Ⓢ = ① ∨ ②.

The set of curves shown in Fig. 2 can now also be recognized as the regularization of the very complicated spaghetti of time-dependent curves, $a_i(n)$, that would result from a straight application of DPD marching forward in time.

REFERENCES

1. R. E. Kurth and D. C. Cox, "Discrete Probability Distributions for Probabilistic Fracture Mechanics," *Risk Anal.* **5**, 235–241 (1985).
2. S. Kaplan, "On the Method of Discrete Probability Distributions in Risk and Reliability Calculations—Application to Seismic Risk Assessment," *Risk Anal.* **1**, (1981).
3. S. Kaplan, *On Determining Inspection Times for Turbine Discs Subject to Stress Corrosion Cracking: An Example of the Use of PRA in Structural Mechanics* (presented at the 8th International Conference on Structural Mechanics in Reactor Technology, Brussels, Belgium, August 19–23, 1985.
4. S. Kaplan, H. F. Perla, and D. C. Bley, "A Methodology for Seismic Risk Analysis of Nuclear Power Plants," *Risk Anal.* **3**, 169–180 (1980).