Fusion of Probabilistic Knowledge as Foundation for Sliced-Normal Approach

Michael Beer[1], Olga Kosheleva[2], and Vladik Kreinovich[2]

[1]Institute for Risk and Reliability, Leibniz University Hannover
30167 Hannover, Germany, beer@irz.uni-hannover.de
[2]University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA
sraguilar4@miners.utep.edu, olgak@utep.edu, vladik@utep.edu

1. Sliced-Normal Distributions Are Efficient

- In many practical applications, it turns out to be efficient to use Sliced-Normal multi-D distributions.
- These are distributions for which log of probability density function (pdf) $f(x_1, \ldots, x_n)$ is a polynomial:

$$\ln(f(x_1, \ldots, x_n)) = P(x_1, \ldots, x_n), \text{ so}$$

$$f(x_1, \ldots, x_n) = \exp(P(x_1, \ldots, x_n)).$$

- To be more precise, $\log(f)$ is a sum of squares of several polynomials
- This class is a natural extension of:
  - normal distributions, i.e.,
  - distributions for which the logarithm of the pdf is a quadratic polynomial.

2. But Why?

- The sliced-normal distributions have been empirically successful.
- However, there seems to be no convincing theoretical explanation for their empirical success.
- The main goal of this paper is to provide such an explanation.

3. Let Us Formulate This Problem in Precise Terms

- In principle, we can have many different probability density functions.
- The class of all the functions is infinite-dimensional.
- This means that:
  - to select a single probability density function out of all possible such functions,
  - we need to know the values of infinitely many parameters,
  - e.g., values of the pdf at points with rational coordinates.

4. Let Us Formulate This Problem in Precise Terms (cont-d)

- In practice, however, at any given moment of time, we only have finitely many observations.
- Based on these observations, we can determine only finitely many parameters.
- Thus, it makes sense to looks for families F of probability density functions:
  - that depend on finitely many parameters $c_1, \ldots, c_m$,
  - i.e., on families of the type

$$F = \{f(x_1, \ldots, x_n, c_1, \ldots, c_m)\}_{c_1,\ldots,c_m}.$$

5. The Dependence Should Be Continuous

- All our information about the physical world comes from measurements and from expert estimates.
- Measurements are never 100% accurate, expert estimates are even less accurate.
- So, we can only determine the values x_i and c_j with some accuracy:
  - based on these approximate values of x_i and c_j,
  - we estimate of the value $f(x_1, \ldots, x_n, c_1, \ldots, c_m)$ of the pdf.
- The more accurately we perform measurements, the more accurate should be our estimates.
- So, the dependence of $f(x_1, \ldots, x_n, c_1, \ldots, c_m)$ on all its inputs x_i and c_j should be continuous.

6. The Dependence Should Be Differentiable

- Moreover, small inaccuracy in x_i and c_j should lead to proportionally small inaccuracy in the value of

$$f(x_1, \ldots, x_n, c_1, \ldots, c_m).$$

- Thus, the function $f(x_1, \ldots, x_n, c_1, \ldots, c_m)$ should be *differentiable*.

7. The Class of Distributions Should Be Closed Under Fusion

- We only know the *probability* of different tuples

$$x = (x_1, \ldots, x_n).$$

- This means that we do not know which of the tuples describe the corresponding real-life situation.
- In other words, the fact that we have a probabilistic knowledge means that our knowledge is incomplete.
- It is therefore desirable to gain additional knowledge about the situation:
  - either by performing additional measurements,
  - or by requesting additional expert estimates.

8. Closed Under Fusion (cont-d)

- This additional knowledge usually comes in the form of a probability distribution; once we have it:
  - we need to fuse it
  - with the distribution describing our original knowledge.
- The selected family should describe reasonably well all possible states of our knowledge.
- From this viewpoint, it is reasonable to require that:
  - if both fused pieces of knowledge are described by distributions from our family,
  - then the result of fusing these two pieces of knowledge should also belong to our family.

9. Closed Under Fusion (cont-d)

- This means that the desired family of probability distributions should be *closed* under fusion.
- To describe this requirement in precise terms, let us describe fusion in precise terms.

10. How to Describe Fusion: A Natural Idea

- In probability theory:
  - if we have two independent events with probabilities p_1 and p_2,
  - then the probability that both events will happen is equal to the product of these probabilities.
- Similarly, suppose that we have two independent sources of information, so that.
- Based on information from Source 1, we assign:
  - to each of N alternatives $a_1, \ldots, a_N$,
  - the probabilities $p_{1,1}, \ldots, p_{1,N}$.
- Based on information from Source 2, we assign:
  - to each of M alternatives $b_1, \ldots, b_M$,
  - the probabilities $p_{2,1}, \ldots, p_{2,M}$.

11. How to Describe Fusion (cont-d)

- Then:
  - the probability that we have alternative a_i in the first case and alternative b_j in the second case
  - is equal to the product of the corresponding probabilities $p_{1,i} \cdot p_{2,j}$.
- Sometimes, in both cases, we have the exact same set of alternatives.
- Then we need to consider *conditional* probabilities, namely probabilities under the condition that $i = j$.
- In general, the conditional probability $P(A\,|\,B)$ of an event A under the condition B can be obtained by:
  - dividing the probability $P(A\,\&\,B)$ of $A\,\&\,B$
  - by the probability $P(B)$ that the condition B is satisfied.

12. How to Describe Fusion (cont-d)

- In our case, this means that after the fusion, the probability of the i-th alternative is equal to $p_i = C \cdot p_{1,i} \cdot p_{2,i}$.
- The coefficient $C \stackrel{\text{def}}{=} \dfrac{1}{P(B)}$ can be obtained from the requirement:
  - that the resulting probabilities add up to 1,
  - i.e., that $\sum\limits_{i=1}^{N} p_i = C \cdot \sum\limits_{i=1}^{N} p_{1,i} \cdot p_{2,i} = 1$, so that

$$C = \frac{1}{\sum\limits_{i=1}^{N} p_{1,i} \cdot p_{2,i}}.$$

13. How to Describe Fusion (cont-d)

- Similar formulas can be obtained for continuous distributions:
  - if we have two independent sources of information that lead to distributions $f_1(x)$ and $f_2(x)$,
  - then the fusion of these two pieces of information is a probability distribution $f(x) = C \cdot f_1(x) \cdot f_2(x)$.
- Here C is a normalization constant.
- Similarly, we can define the result of fusing several probability distributions as $f(x) = C \cdot f_1(x) \cdot \ldots \cdot f_k(x)$.

14. Every Piece of Knowledge Can Be Obtained by Fusing "Smaller" Pieces of Information

- Sometimes, knowledge comes in one big step.
- However, more typically, to gain the knowledge, we must acquire it piece by piece.
- Sometimes it comes in two steps, sometimes in three steps, sometimes in four steps, etc.
- So, it is natural to come up with the following definition.
- We say that in a family $f(x,c)$, *every piece of knowledge can be obtained by fusing* if
  - for every pdf $f(x,c)$ from this family and for every integer $M \geq 2$,
  - there exists another pdf $f(x,c')$ from this family so that fusing M copies of $f(x,c')$ leads to $f(x,c)$.

15. The Family Should Not Depend on the Starting Point and Measuring Unit

- We want to deal with physical quantities, but in reality, we deal with their numerical values.
- These numerical values depend on:
  - what measuring unit we use for measuring the quantity, and
  - what starting point we select for this measurement.
- When we change the measuring unit and/or the starting point, the numerical values change; e.g.:
  - if we change the measuring unit from meters to centimeters,
  - all the numerical values are multiplied by 100, so that, 2 m becomes 200 cm.

16. The Family Should Not Depend on the Starting Point and Measuring Unit (cont-d)

- In general:
  - if we change from the original measuring unit to a new one which is a times smaller,
  - then all the numerical values are multiplied by a:

$$x \to a \cdot x.$$

- This transformation is known as *scaling*.
- Similarly, we can change the starting point to the one which is b units before.
- We can do it for time, temperature, and many other quantities.
- Then b is added to all the numerical values $x \to x + b$.
- This transformation is known as *shift*.

- A shift can also be viewed as a kind of re-scaling.

17. The Family Should Not Depend on the Starting Point and Measuring Unit (cont-d)

- If we change both the measuring unit and the starting point, then numerical values change as $x \to a \cdot x + b$.
- These transformations change the pdf $f(x_1, \ldots, x_n)$:
  - if we apply such transformation $x_i \to x'_i \stackrel{\text{def}}{=} a_i \cdot x_i + b_i$ to each of inputs,
  - then in terms of the new numerical values $x'_1, \ldots, x'_n$, the pdf takes a different form:

$$f'(x'_1, \ldots, x'_n) = \frac{1}{\prod\limits_{i=1}^{n} a_i} \cdot f\left(\frac{x'_1 - b_1}{a_1}, \ldots, \frac{x'_n - b_n}{a_n}\right).$$

18. The Family Should Not Depend on the Starting Point and Measuring Unit (cont-d)

- We want to come up with a universal family of probability distributions.
- This family should be applicable no matter what measuring units and what starting points we select.
- Thus, it is reasonable to require that our family is invariant w.r.t. the corresponding transformations.
- We say that a family F is *scale- and shift-invariant* if:
  - every pdf $f(x, c)$ from this family and for every two tuples a and b,
  - the (a, b)-re-scaling of the pdf $f(x, c)$ also belongs to the family F.

19. Main Result

- Let F be a differentiable family $f(x, c)$:
  - which is closed under fusion,
  - for which every piece of knowledge can be obtained by fusing small pieces of information, and
  - which is scale- and shift-invariant.
- Then, there exists an integer $d \leq m + 1$ such that:
  - every probability density function from this family has the form $f(x, c) = \exp(P(x_1, \ldots, x_n))$,
  - where $P(x_1, \ldots, x_n)$ is a polynomial of degree $\leq d$ with respect to each of its variables.
- This result explains the empirical success of sliced-normal distributions.

20. Proof: Part 1

- Let F be the family that satisfies all the conditions described in the formulation of our result.
- By a *log-function*, we will mean a function of the type $L(x, c, s) = \ln(f(x, c)) + s$ for some c and s.
- Let us denote the class of all log-functions by $\mathcal{L}$.
- Let us prove that the class of all log-functions is closed under addition, i.e., that:
  - for every two log-functions

$$L(x, c;, s') \text{ and } L(x, c'', s''),$$

  - their sum is also a log-function.
- Indeed, by definition, $L(x, c', s') = \ln(f(x, c')) + s'$ and $L(x, c'', s'') = \ln(f(x, c'')) + s''$.

21. Proof: Part 1 (cont-d)

- Since the family F is closed under fusion, the result of fusing the corresponding pdfs is also a pdf from F:

$$C\cdot f(x,c')\cdot f(x,c'')=f(x,c) \text{ for some tuple } c.$$

- By taking logarithms of both sides of this equality, we conclude that

$$\ln(C)+\ln(f(x,c'))+\ln(f(x,c''))=\ln(f(x,c)).$$

- If we add $s'+s''-\ln(C)$ to both sides of the resulting equality, we conclude that

$$(\ln(f(x,c'))+s')+(\ln(f(x,c''))+s'')=$$

$$\ln(f(x,c))+(s'+s''-\ln(C)).$$

22. Proof: Part 1 (cont-d)

- So, the sum of the two given log-functions is indeed a log-function:

$$L(x, c', s') + L(x, c'', s'') = L(x, c, s' + s'' - \ln(C)).$$

- The statement is proven.

23. Proof: Part 2

- Let us now prove that:
  - for each log-function $L(x, c, s)$ and for every integer $M \geq 2$,
  - the function $M^{-1} \cdot L(x, c, s)$ is also a log-function.
- By definition, $L(x, c, s) = \ln(f(x, c)) + s$.
- For the family F, every piece of knowledge can be obtained by fusing small pieces of information.
- So, the pdf $f(x, c)$ can be obtained by fusing M instances of some other pdf $f(x, c')$:

$$f(x, c) = C \cdot (f(x, c'))^M.$$

24. Proof: Part 2 (cont-d)

- We have $f(x,c) = C \cdot (f(x,c'))^M$.
- By taking logarithms of both sides of this equality, we get $\ln(f(x,c)) = M \cdot \ln(f(x,c')) + \ln(C)$, thus

$$M^{-1} \cdot \ln(f(x,c)) = \ln(f(x,c')) + M^{-1} \cdot \ln(C).$$

- By adding $M^{-1} \cdot s$ to both sides, we get

$$M^{-1} \cdot (\ln(f(x,c)) + s) =$$

$$\ln(f(x,c')) + M^{-1} \cdot (\ln(C) + s).$$

- The left-hand side of this formula is exactly

$$M^{-1} \cdot L(x,c,s).$$

- So, it is indeed a log-function.
- The statement is proven.

25. Proof: Part 3

- Let us now consider the closure $\mathcal{C}$ of the set $\mathcal{L}$ of all log-functions:
  - the closure in the usual topological sense, i.e.,
  - the set of all limit functions with respect to some natural topology on the class of all functions.
- Since the set $\mathcal{L}$ is closed under addition, its closure $\mathcal{C}$ is also closed under addition.
- Let us prove that this closure is closed under multiplication by positive numbers.
- In other words, let us prove that:
  - for each function $f(x) \in \mathcal{C}$ and for every positive real number $r > 0$,
  - the function $r \cdot f(x)$ also belongs to $\mathcal{C}$.

26. Proof: Part 3 (cont-d)

- Since $\mathcal{C}$ is the closure of the set of all log-functions, it is sufficient to prove that:
  - for each log-function $L(x, c, s)$ and for every positive real number $r > 0$,
  - the function $r \cdot L(x, c, s)$ is a limit of log-functions.
- Indeed, for every possible accuracy $\varepsilon > 0$:
  - we can approximate, with this accuracy,
  - the real number r by a rational number $\dfrac{N}{M}$.
- By Part 2 of this proof, the function $M^{-1} \cdot L(x, c, s)$ is also a log-function.

27. Proof: Part 3 (cont-d)

- Now, by Part 1 of this proof:
  - the function $\dfrac{N}{M}\cdot L(x,c,s)$ is also a log-function,
  - as the sum of N log-functions $M^{-1}\cdot L(x,c,s)$.
- When $\dfrac{N}{M}$ tends to r, the corresponding function $\dfrac{N}{M}\cdot L(x,c,s)$ tends to $r\cdot L(x,c,s)$.
- Thus, the function $r\cdot L(x,c,s)$ is indeed a limit of log-functions.
- The statement is proven.

28. Proof: Part 4

- By combining Parts 1 and 3, we conclude that:
  - for every finite set of functions $C_1(x), \ldots, C_k(x)$ from the set $\mathcal{C}$, and
  - for every tuple of positive numbers $r_1, \ldots, r_k$,
  - the linear combination $r_1 \cdot C_1(x) + \ldots + r_k \cdot C_k(x)$ also belongs to $\mathcal{C}$.
- Let us prove that the set $\mathcal{C}$ cannot contain more than $m+1$ linearly independent functions; indeed:
  - if this was the case, and we would have more than $m+1$ linearly independent functions,
  - then we would have at least $m+2$ of them $C_1(x), \ldots, C_{m+2}(x)$ in the class $\mathcal{C}$.
- Then, the class $\mathcal{C}$ will contain a $(m+2)$-parametric family of functions $r_1 \cdot C_1(x) + \ldots + r_{m+2} \cdot C_{m+2}(x)$.

29. Proof: Part 4 (cont-d)

- However, the class $\mathcal{C}$ is the closure of the class $\mathcal{L}$ of functions of the type $\ln(f(x,c)) + s$.
- This class depends on $m+1$ parameters:
  - we have m parameters $c_1, \ldots, c_m$ and
  - we have an additional parameter s.
- So, the closure of this set is also of dimension $m+1$ (or less).
- Thus, cannot contain more-dimensional subfamilies.
- The statement is proven.

30. Proof: Part 5

- Let us denote by $\mathcal{S}$ the class of all linear combinations of functions from the class $\mathcal{C}$.
- Clearly, $\mathcal{C} \subseteq \mathcal{S}$.
- Due to Part 4, the dimension d of the linear space $\mathcal{S}$ cannot exceed $m+1$; so:
  - if we pick any basis $e_1(x), \ldots, e_d(x)$ in this class,
  - then each function $f(x) \in \mathcal{S}$ can be represented as

$$f(x) = C_1 \cdot e_1(x) + \ldots + C_d \cdot e_d(x).$$

- We can pick the basis from the set $\mathcal{C}$; moreover:
  - since the closure does not change the dimension,
  - we can pick this basis from the original class $\mathcal{L}$ of log-functions.

31. Proof: Part 5 (cont-d)

- All the pdf functions from the family F are, by our assumption, differentiable.
- Thus, every log-function is also differentiable.
- Hence, we can choose the basis of differentiable functions.

32. Proof: Part 6

- Let us prove that the class $\mathcal{L}$ is closed under arbitrary re-scalings, i.e.:
  - if a function $f(x_1, \ldots, x_n)$ is in this class,
  - then for each tuple $a = (a_1, \ldots, a_n)$ of positive numbers and for each tuple $b = (b_1, \ldots, b_n)$ of real numbers, the function $f(a \cdot x_1 + b_1, \ldots, a_n \cdot x_n + b_n)$ also belongs to the class $\mathcal{L}$.
- This follows:
  - from the requirement that the family F is scale- and shift-invariant,
  - if we take logarithms of both sides and add appropriate constants to both sides.
- From this, we can conclude that the closure class $\mathcal{C}$ is also invariant with respect to arbitrary re-scalings.

33. Proof: Part 6 (cont-d)

- So, the class $\mathcal{S}$ of all linear combinations of functions from $\mathcal{C}$ is also thus invariant.
- Let us first study the consequences of shift-invariance of the class $\mathcal{S}$ with respect to the first variable.
- This shift-invariance implies, in particular, that:
  - for each basis function $e_i(x_1, x_2, \ldots, x_n)$,
  - the result of its shift $e_i(x_1 + b_1, x_2, \ldots, x_n)$ is also a function from $\mathcal{S}$, i.e., for some C_{ij}:

$$e_i(x_1+b_1, x_2, \ldots, x_n) = \sum_{j=1}^{d} C_{ij}(b_1)\cdot e_j(x_1, x_2, \ldots, x_n).$$

- For a while, let us fix the values $x_2, \ldots, x_n$ and only consider the dependence on x_1.

34. Proof: Part 6 (cont-d)

- In other words, let us consider auxiliary functions

$$E_i(x_1) \stackrel{\text{def}}{=} e_i(x_1, x_2, \ldots, x_n).$$

- For these auxiliary functions, the above formula takes the form

$$E_1(x_1 + b_1) = C_{11}(b_1) \cdot E_1(x_1) + \ldots + C_{1d}(b_1) \cdot E_d(x_1);$$

$$\ldots$$

$$E_d(x_1 + b_1) = C_{d1}(b_1) \cdot E_1(x_1) + \ldots + C_{dd}(b_1) \cdot E_d(x_1).$$

- Here, all the functions $E_1(x_1) \ldots, E_d(x_1)$ are differentiable – since:
  - they come by fixing some values from the basis functions $e_i(x_1, \ldots, x_n)$, and
  - the basis functions are differentiable.

35. Proof: Part 6 (cont-d)

- Let us prove that the dependencies $C_{ij}(b_1)$ are also differentiable.
- Indeed, for each i, let us pick d different values $x_{1,1},\ldots,x_{1,d}$ of x_1.
- Then we get the following d linear equations for d unknowns $C_{i1}(b_1),\ldots,C_{in}(b_1)$:

$$E_i(x_{1,1}+b_1)=C_{i1}(b_1)\cdot E_1(x_{1,1})+\ldots+C_{id}(b_1)\cdot E_d(x_{1,1});$$

$$\ldots$$

$$E_i(x_{1,d}+b_1)=C_{i1}(b_1)\cdot E_1(x_{1,d})+\ldots+C_{id}(b_1)\cdot E_d(x_{1,d}).$$

- Each element $C_{ij}(b_1)$ of the solution to a system of linear equations can be described, by the Cramer rule:
  - as the ratio of two determinants, i.e.,
  - as a smooth function of all the coefficients.

36. Proof: Part 6 (cont-d)

- The coefficients $E_i(x_{1,k}+b_1)$ smoothly depend on b_1.
- So, we conclude that the solutions $C_{ij}(b_1)$ are also differentiable functions of b_1.
- Since all the functions $E_i(x_1)$ and $C_{ij}(b_1)$ are differentiable, we can:
  - differentiate both sides of all equalities describing $E_i(x_1+b_1)$ with respect to b_1,
  - and take $b_1 = 0$.
- Then, we get the following system of equations:

$$E_1'(x_1) = c_{11}\cdot E_1(x_1)+\ldots+c_{1d}\cdot E_d(x_1);$$

$$\ldots$$

$$E_d'(x_1) = c_{d1}\cdot E_1(x_1)+\ldots+c_{dd}\cdot E_d(x_1).$$

- Here, $E_i'(x_1)$ denotes the derivative, and $c_{ij} \stackrel{\text{def}}{=} C_{ij}'(0)$.

37. Proof: Part 6 (cont-d)

- In other words:
  - for the functions $E_1(x), \ldots, E_d(x)$,
  - we get a system of linear differential equations with constant coefficients.
- It is known that a general solution to such system of equations is a linear combination
  - of functions of the type $x_1^k \cdot \exp((p + \mathrm{i} \cdot q) \cdot x_1)$, i.e.,
  - functions of the type $x_1^k \cdot \exp(p \cdot x_1) \cdot \cos(q \cdot x_1)$ and $x_1^k \cdot \exp(p \cdot x_1) \cdot \sin(q \cdot x_1)$.

38. Proof: Part 6 (cont-d)

- Here:
  - the values $p + \mathrm{i} \cdot q$ are eigenvalues of the matrix c_{ij},
  - and k is a non-negative integer corresponding to duplicate eigenvalues.
- For a $d \times d$ matrix, the multiplicity of an eigenvalue cannot exceed d, so $k \leq d$.

39. Proof: Part 7

- Let us now study the consequences of *scale*-invariance of the class $\mathcal{S}$ with respect to the first variable.
- This scale-invariance implies, in particular, that:
  - for each basis function $e_i(x_1, x_2, \ldots, x_n)$,
  - the result of its re-scaling $e_i(a_1 \cdot x_1, x_2, \ldots, x_n)$ is also a function from $\mathcal{S}$, i.e., for some D_{ij}:

$$e_i(a_1 \cdot x_1, x_2, \ldots, x_n) = \sum_{j=1}^{d} D_{ij}(a_1) \cdot e_j(x_1, x_2, \ldots, x_n).$$

- Thus:

$$E_1(a_1 \cdot x_1) = D_{11}(a_1) \cdot E_1(x_1) + \ldots + D_{1d}(a_1) \cdot E_d(x_1);$$

$$\ldots$$

$$E_d(a_1 \cdot x_1) = D_{d1}(a_1) \cdot E_1(x_1) + \ldots + D_{dd}(a_1) \cdot E_d(x_1).$$

40. Proof: Part 7 (cont-d)

- Similarly to Part 6, we can prove that the dependencies $D_{ij}(a_1)$ are also differentiable.
- By differentiating both sides of the above equations with respect to a_1 and taking $a_1 = 1$, we conclude that

$$x_1 \cdot E_1'(x_1) = d_{11} \cdot E_1(x_1) + \ldots + d_{1d} \cdot E_d(x_1);$$

$$\ldots$$

$$x_1 \cdot E_d'(x_1) = d_{d1} \cdot E_1(x_1) + \ldots + d_{dd} \cdot E_d(x_1).$$

- In each equation, the left-hand side $x_1 \cdot \dfrac{dE_i}{dx_1}$ can be reformulated as $\dfrac{dE_i}{dx_1/x_1} = \dfrac{dE_i}{d(\ln(x_1))}$.

41. Proof: Part 7 (cont-d)

- Thus, for $X_1 \stackrel{\text{def}}{=} \ln(x_1)$, we get the system of linear differential equations with constant coefficients:

$$\frac{dE_1}{dX_1} = d_{11} \cdot E_1(X_1) + \ldots + d_{1d} \cdot E_d(X_1);$$

$$\ldots$$

$$\frac{dE_d}{dX_1} = d_{d1} \cdot E_1(X_1) + \ldots + d_{dd} \cdot E_d(X_1).$$

- We already know that a general solution to this equation is a linear combination of functions

$$X_1^k \cdot \exp(p \cdot X_1) \cdot \cos(q \cdot X_1) \text{ and}$$

$$x_1^k \cdot \exp(p \cdot X_1) \cdot \sin(q \cdot X_1).$$

42. Proof: Part 7 (cont-d)

- Let us substitute $X_1 = \ln(x_1)$ into these formulas and take into account that

$$\exp(p \cdot \ln(x_1)) = (\exp(\ln(x_1))^p = x_1^p.$$

- We conclude that a general solution is a linear combination of functions

$$(\ln(x_1))^k \cdot x_1^p \cdot \cos(q \cdot \ln(x_1)) \text{ and}$$

$$(\ln(x_1))^k \cdot x_1^p \cdot \sin(q \cdot \ln(x_1)).$$

43. Proof: Part 8

- From Parts 7 and 8, we get two different expressions for the functions $E_i(x_1)$.
- By comparing these expressions, one can easily see that:
  - the only functions that can be described in both forms
  - are functions of the form x^k for some non-negative integer $k \le d$
  - or their linear combinations.
- So, each function $E_i(x_1)$ is a linear combination of such functions – i.e., a polynomial.

44. Proof: Part 8 (cont-d)

- We have shown that for each combination of values of $x_2, \ldots, x_n$:
  - the dependence of each function $e_i(x_1, x_2, \ldots, x_n)$ on x_1
  - can be described by a polynomial of degree $\le d$.
- Similarly, we can prove that:
  - for each combination of values $x_1, x_3, \ldots, x_n$,
  - the dependence on x_2 is also described by a polynomial.
- Let us combine these two conclusions and prove that for all possible values of $x_3, \ldots, x_n$:
  - the dependence of $e_i(x_1, x_2, x_2, \ldots, x_n)$ on x_1 *and* on x_2
  - can be described by a polynomial of two variables.

45. Proof: Part 8 (cont-d)

- Indeed, let us denote $T(x_1, x_2) \stackrel{\text{def}}{=} e_i(x_1, x_2, x_3, \ldots, x_n)$.
- We know that:
  - for each x_2, this expression is a polynomial is x_1, and
  - for each x_1, this expression is a polynomial is x_2.
- Let us prove that $T(x_1, x_2)$ is a polynomial of two variables.
- Indeed:
  - the fact that the dependence of e_i on x_1 can be described by a polynomial of order $\le d$
  - can be rewritten, in terms of $T(x_1, x_2)$, as:

$$T(x_1, x_2) = a_0(x_2) + a_1(x_2) \cdot x_1 + \ldots + a_d(x_2) \cdot x_1^d.$$

46. Proof: Part 8 (cont-d)

- In writing this expression, we took into account that, in general:
  - for different values of x_2,
  - the coefficients $a_0, \ldots, a_d$ of this polynomial may be different.
- Let us substitute d_1 different values $x_{1,0}, \ldots, x_{1,d}$ of x_1 into this formula.
- As a result, we have $d+1$ linear equations for $d+1$ unknowns $a_0(x_2), \ldots, a_d(x_2)$, with constant coefficients:

$$T(x_{1,0}, x_2) = a_0(x_2) + a_1(x_2) \cdot x_{1,0} + \ldots + a_d(x_2) \cdot x_{1,0}^d;$$

$$\ldots$$

$$T(x_{1,d}, x_2) = a_0(x_2) + a_1(x_2) \cdot x_{1,d} + \ldots + a_d(x_2) \cdot x_{1,d}^d.$$

47. Proof: Part 8 (cont-d)

- In general:
  - each component in a solution to a system of linear equations
  - is a linear combination of the right-hand sides.
- The right-hand sides $T(x_{1,i}, x_2)$ are polynomials of x_2.
- Thus, each coefficient $a_i(x_2)$ is a linear combination of polynomials – thus, a polynomial itself.
- All the expressions $a_i(x_2)$ are polynomials.
- Thus, the whole above expression for $T(x_1, x_2)$ becomes a polynomial in two variables x_1 and x_2.

48. Proof: Part 8 (cont-d)

- By adding variables one by one, we can prove that the dependence of each basis function $e_i(x_1, \ldots, x_n)$:
  - on x_1, x_2, and x_3 is a polynomial,
  - ...
  - on all n variables $x_1, \ldots, x_n$ is a polynomial.
- Thus, each element of $\mathcal{S}$ – which is a linear combination of the basis functions – is also a polynomial.
- For each tuple c, the function $\ln(f(x, c))$ belongs to the class $\mathcal{L} \subseteq \mathcal{S}$ and is, thus, also a polynomial.
- So, indeed, each pdf $f(x, c)$ from the family F has the form $\exp(P(x_1, \ldots, x_n))$ for some polynomial

$$P(x_1, \ldots, x_n).$$

- Our main result is proven.

49. Acknowledgments

This work was supported in part by the National Science Foundation grants:

- 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science);
- HRD-1834620 and HRD-2034030 (CAHSI Includes).

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.