연차 보고서 (2년차)

사업명	KAIST Grand Challenge 30 Project		
과제명	(국문) 지식의 물리적 토대		
	(영문) The Physical Basis of Knowledge		
연구책임자	C. D. Fiorillo	소 속	KAIST 바이오및뇌공학
총수행기간 (1단계)	2017. 01. 01. ~ 2021. 12. 31. (5년)		
당해연도 협약기간	2018. 01. 01. ~ 2018. 12. 31. (1년)		
당해년도 사업비(원)			

자체연구협약서(KAIST Grand Challenge 30 Project)제5조에 의거하여
연차보고서 2부를 제출합니다.

2019년 1 월 15일

연구책임자: C. D. Fiorillo (인)

한국과학기술원 총장 귀하

The Physical Basis of Knowledge
2018 Year-End Report and 2019 Plan for KC30 Grant

Christopher D. Fiorillo

January 14, 2019

Project Summary

Our goal is to describe the knowledge in a physical system with probabilities, in analogy to the way that calculus can describe the motion of a physical system. We plan to first describe the knowledge in a geometrical configuration of points using probabilities, before using the same methods to describe the knowledge in a physical system (such as the position and velocity of a particle, or the voltage and current across a neuron's membrane).

Specific Aims

Below are the original aims of the whole project.

1. Solve the 'N-point problem.'
2. Given the position, velocity, and acceleration of one particle, where is another particle? 'Position and velocity' is similar to (or the same as) the special case of N=2 in the 'N-point problem.'
3. Given a neuron's membrane voltage and current, what is the external state of the world? We suspect that this may be similar to the case of the position and velocity of a particle (above), but on a macroscopic level in which many particles move in a highly coherent manner.

Ⅰ. 해당 연도 추진 현황

Ⅰ-1 기술개발 추진 내용

1.1.1. Sunil Kim and I have found a partial solution to the N-point problem (specific aim 1) by proving than a known size x (a positive number representing length, area, volume, etc.) is the median of the probability distribution over an a unknown size y. We have a manuscript entitled "*The Probability that an Unknown Positive Real Variable is Larger than a Known is $p(y>x \mid x) = p(y<x \mid x) = 1/2$*". We spent a lot of time in 2018 revising the manuscript and trying to publish it. We encountered difficulties with reviewers. Our most general obstacle is that the question that we ask (given one size, what is another?) is unfamiliar to most statisticians. In particular, reviewers became confused over our use of the unnormalizable distribution $p(x) \propto 1/x$, known as "Jeffreys prior." This caused confusion even though our

non-trivial results concern normalizable distributions. We will revise the manuscript so that this issue does not arise, and we will then submit the manuscript to a new journal.

1.1.2. We have additional observations beyond the 'median property' that we plan to publish separately in a manuscript titled "*Probability Distributions Over One Positive Real Number Given Another, p(y / x).*" The manuscript will characterize the set of probability distributions conditional on one known size (Figure 1). We will show that the entire family of logistic distributions is conditional on one known size (one positive number). We will also show that probabilities conditional on one known size have the same form as distributions over an unknown ratio, p(y | x=1) = p(y/x). These distributions are important because they are candidate distributions to describe the information in the simplest models of a physical observer (specific aim 3). For example, in a model of a neural observer, y could be the unknown and external energy of a sensory stimulus, and x could be the known internal energy in a neuron's membrane voltage. In a practical problem confronting a scientist, x could the known distance of 1 meter (the approximate length of a human) and y could be a distance to be measured. Knowledge of standard units is information that scientists routinely use, but without knowing how it relates to probabilities (we were not aware of this issue when we applied for the grant in 2016).

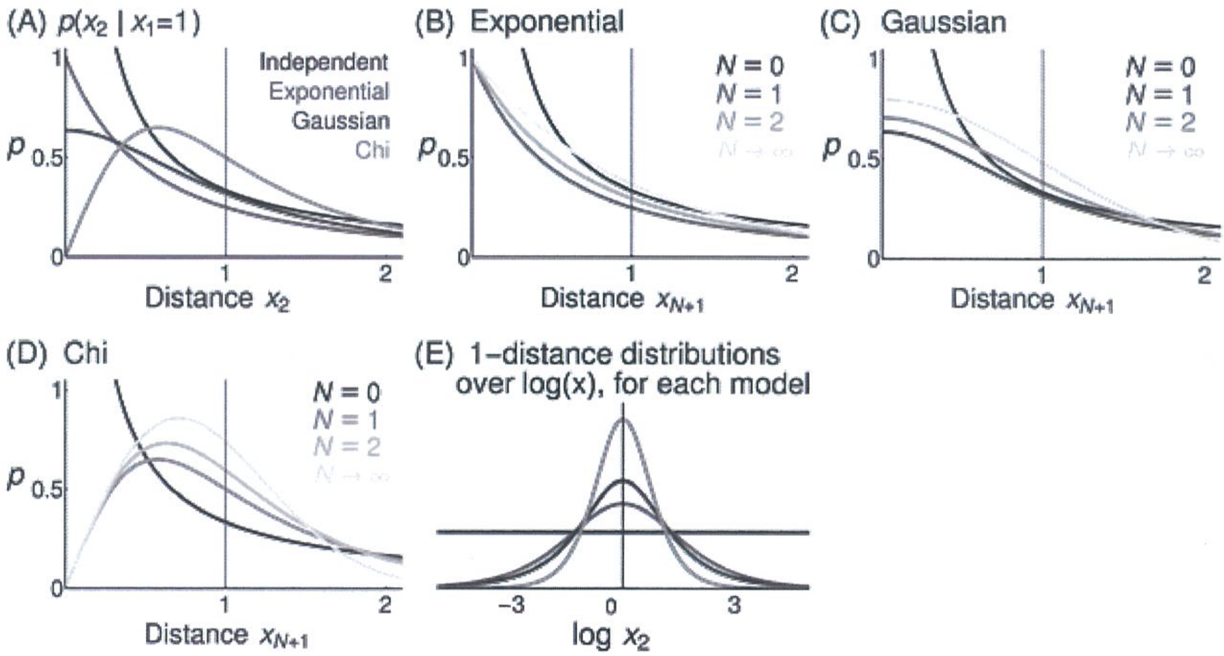

Figure 1. Conditional distributions over distance. A. Conditional 1-distance distributions 'p(x2|x1)' for our four exemplary size metrics. 1-distance distributions for exponential and chi

metrics are log-logistic. Unlike the other cases, the distribution given independence (Jeffreys's distribution) cannot be normalized, and therefore its position on the identity line was chosen only for convenience [$p(x2|x1) = c/x2$, where 'c' was arbitrarily chosen]. B-D. Distributions conditional on 'N' of 0, 1, 2, and infinite distances, given that all distances equal '1,' based on exponential (B), Gaussian (C), and chi (D) metrics. As N approaches infinity, the distributions approach exponential, half-Gaussian (log-normal), and chi distributions, respectively. E. If we transform distance to logarithm of distance, the distributions have symmetry around the median, which becomes the mean and mode.

1.1.3 We continued work on the "n-point problem" (specific aim 1). We still have not determined a solution, but we have made substantial progress. Our best candidate solution for n=3 (the distribution over triangles) is

p(a,b,c) proportional to '1 / (a^2 + b^2 + c^2)

where 'a,b,c' are the lengths of sides of a triangle.

1.1.4 We have developed ideas that may form an improved foundation for probability theory. This was not part of our plan, since we did not have the ideas until October 2017. We continued to make progress on these ideas in 2018.

We had two major advances in 2018. First, we found a way to derive $p(x) \propto 1/x$ (known as "Jeffreys prior"), where x is the size of some thing, from a uniform distribution over space. The advantage is that uniform distributions are very easily understood, whereas we have encountered skepticism about $p(x) \propto 1/x$, because its justification is not obvious.

Second, we found a general way to derive probabilities from by formulating propositions that are functions of variables. Jeffreys prior over positive numbers is usually written $p(x) \propto 1/x$. We can replace 'x' with a function of x and y and any number of additional variables to give $p(f(x,y)) \propto 1 / f(x,y)$. For example, we could compare the probabilities of two proposition, (x+y)^2 and x^2+y^2.
p((x+y)^2) / p(x^2+y^2) = x^2+y^2 / (x+y)^2

If we have "data" x=1 and y=2, we have

$p((1+2)^2) / p(1^2+2^2) = 1^2+2^2 / (1+2)^2 = 5 / 9$

The first of these propositions is an "exponential model" (since it leads to the exponential distribution under certain conditions), and the second is a Gaussian model. The equation above indicates that if we have only data x=1,y=2, the evidence favors the Gaussian model by a factor of 9/5.

Although it was not an original goal to develop a method like this, it could help us to solve the n-point problem. We may be able to prove that our candidate solution (see 1.1.3 above) is favored over all alternatives in the absence of any data.

1.1.5 Some formulations of specific aim 2 could be solved given our ideas in 1.1.2 (e.g, given one velocity of one particle, what is another?) However, the most important formulation of aim 2 involves 'time.' Although this has not been a major focus in 2018, I did continue to work on it.

1.1.6. I maintained my collaborative relationship with Professor Jaime Gomez-Ramirez of Universidad Computense and the Queen Sofia Foundation in Madrid, Spain. He has been a useful resource and we plan to continue our interaction in 2019.

I-2 해당 연도 추진 실적

We were disappointed that our first manuscript on the median was not published in 2018 (see 1.1.1.). However, we were encouraged to have made significant progress on the conceptual foundations (see 1.1.3 and 1.1.4). I would grade our progress in 2017 to be 80/100.

II. 기술개발결과

2.1 We introduced the problem in a publication in 2016.
Kim SL, Fiorillo CD (2016) Describing realistic states of knowledge with exact probabilities. *American Institute of Physics Conf. Proc.* 1757, 060008-1–060008-8.

2.2 We presented our result for the median at 'Entropy 2018: From physics to information sciences and geometry' in Barcelona, Spain in May 2018.

The Probability that an Unknown Number is Greater than a Known Number is 1/2

2.3 We will resubmit the following manuscript in the next few weeks.

The Probability that an Unknown Positive Real Variable is Larger than a Known is $p(y>x \mid x) = p(y<x \mid x) = 1/2$

Sunil L. Kim and Christopher D. Fiorillo

If there are two positive real variables x and y, and only x is known, what is the probability that y is larger versus smaller than x? This simple question has not had an established answer. Reason compels us to treat two possibilities as equally probable in the absence of evidence to the contrary (the 'principle of indifference'). Therefore it is obvious that $p(y<x) = p(y>x)$, since we must treat x and y equally, and the spaces $y<x$ and $y>x$ have equal size. However, the symmetry of spaces is absent if x is known, since nearly all numbers y in the infinite set $\mathbb{R}^+$ are larger than any finite number x. Here we prove that despite this asymmetry, $p(y<x \mid x) = p(y>x \mid x)$, and thus x is the median of the predictive distribution $p(y|x)$. This holds true for all cases in which the joint distribution $p(x,y)$ has no information discriminating either scales or variables, and thus exhibits scale invariance and exchangeability of variables, respectively. These two symmetries define a set of joint distributions, each member of which exhibits a distinct dependence between variables. This set includes independence, as well as every familiar form of dependence that has a scale parameter, including gamma, Gaussian, chi and logistic dependencies. Our results strengthen the objective Bayesian theory of probability, which should ideally characterize cases of minimal information before advancing systematically towards greater and more complex information.

2.4 We have partially prepared the following manuscript and hope to submit it in the next few months.

Probability Distributions Over One Positive Real Number Given Another, $p(y \mid x)$

Sunil L. Kim and Christopher D. Fiorillo

We characterize a set of probability distributions over one unknown positive real number y (e.g., a size) conditional on one known number x. Each member of the set exhibits a unique dependence between x and y, such as x^2+y^2 (the "Gaussian" dependence). We define the set of joint distribution $p(x,y)$ to be those that contain no information discriminating either variables or scales, and thus all distributions in the set exhibit the symmetries of exchangeability of variables and scale invariance. We demonstrate that the entire family of log-logistic and logistic distributions are members of this set and are thus conditional on a single known number. We further demonstrate that a distribution given one known $p(y|x)$ has the same information content as a distribution over an unknown ratio, $p(y/x)$, since $p(y \mid x=1) = p(y/x)$. These distributions are appropriate as a 'prior' given only knowledge of the human scale, as reflected in standard units such as meters, kilograms, and seconds.

Ⅲ. 결론 및 차년도 계획

3.1 Publish manuscript on 'the median.' (see 1.1.1)
3.2 Publish manuscript characterizing distributions p(y|x) (see 1.1.2)
3.3 Complete formulation on the foundations of probability theory, and begin manuscript (see 1.1.4)
3.4 Solve the general n-point problem (see 1.1.3).

3.5 Work towards specific aim 2 (see 1.1.5)
3.6 Continue collaborative interaction with Professor Gomez-Ramirez (see 1.1.6).

Ⅳ. 기타

Publication of our work on 'the median' has been delayed beyond our expectation of one year ago (see 1.1.1) The only major change from the original plan of 2016 is the introduction of a new project on the foundations of probability theory (see 1.1.4). Otherwise there have been only minor changes to the general project and plans for the future.