

A Probabilistic Time Reversal Theorem

Kenneth Baclawski

Abstract

Combining independent observations is commonly performed by using a least squares technique, as it is thought that this is necessary to achieve an optimal solution. The purpose of this article is to show that this is not always the case. The particular example combines observations that are exponentially distributed. One application of this technique is to determine the time of a singular event which initiated a set of decay processes having known half-lives. The time of the singular event decays backwards in time with an exponential distribution. We find that the accuracy of this method is significantly better than the accuracy of a least squares technique.

1 Introduction

The combination (fusion) of independent observations is a fundamental mechanism of probability theory that is important in many domains, especially in sensor networks. For an introduction to information combination see [1]. Since the early nineteenth century when the method of least squares was developed by Gauss [6], it has been a common practice to use a least squares technique to determine the combination of a set of independent observations. While such a method is well-justified and optimal if observations are normally distributed, it is not optimal for all distributions. To illustrate this phenomenon, we consider the case of a set of independent observations that are exponentially distributed, and we show that the additional information about the distributions can have a significant impact on accuracy.

The Gauss-Markov theorem states that in a linear stochastic model whose errors have expectation zero, equal variances and no correlations, the best linear unbiased estimators of the coefficients of the model are the least-squares estimators. [7, 8] This result makes no assumptions about the distributions beyond their

means and variances, and so it can only yield simple estimation values, not distributions. When one knows the distributions of the errors, one can not only make an estimate, one can also determine a probability distribution. In other words, both the inputs and the outputs of the information combination process are probability distributions. The normal distribution is the best known example of this process. The distribution of the combination of a set of independent normally distributed observations is also normally distributed. Moreover, this result does not depend on any other assumptions or results, such as a least squares technique or Bayes' Law [2, 9]. The fact that the normal distribution is closed under combination makes it an especially useful distribution for combining independent observations. In this article, it is shown that the class of exponential distributions is also closed under combination. Whether there are other classes of distributions have this property is an interesting question.

Time reversal is a useful technique for audio and radio signaling and detection applications.[3, 4, 5] In this technique, a received signal is recorded, reversed, and sent back to the original source. The time reversal theorem in this paper differs in its concern with stochastic phenomena in which the time of an event (or more precisely knowledge of the time of an event) decays exponentially backwards in time. The usefulness of the theorem is due to the improved accuracy of the estimation, not the time reversal itself.

2 The Class of Exponential Distributions

An exponential distribution models the behavior of a continuous, memoryless waiting time. In other words, if one has been waiting a period of time, then the conditional distribution starting at the end of the period is the same as the distribution starting at the initial time. The behavior is determined by just two constants: the starting point s and the expected duration $\tau > 0$ after the starting point s . Let $ED(s, \tau)$ denote the distribution in this case. If X is a random variable whose distribution is $ED(s, \tau)$, then the probability density of X is given by:

$$dens(X = x) = \begin{cases} 0, & \text{if } x < s, \\ \frac{1}{\tau}e^{-\frac{x-s}{\tau}}, & \text{if } x \geq s. \end{cases}$$

The expectation of X is easily seen to be $s + \tau$, and the variance is τ^2 . The exponential distribution models the time between independent occurrences in a sequence that occurs at a constant rate. The rate is given by $\lambda = \frac{1}{\tau}$. One can also

regard the exponential distribution as modeling a quantity that decays at the rate λ . When viewed in terms of a decay, one usually takes the base of the exponential to be 2 rather than e , and the parameter analogous to the mean duration τ is the *half-life*, which is commonly written $t_{1/2}$. The half-life is related to the mean duration by the formula: $t_{1/2} = \ln(2)\tau$. The half-life is the median duration, so the median of the distribution is $s + t_{1/2}$.

The constraint that τ be positive is not essential, and one can easily generalize the class of exponential distributions so that τ can be any real number. If $\tau = 0$, then the distribution is a discrete distribution with all probability concentrated at s (i.e., the density is a delta function). Having the distribution $ED(s, 0)$ means that an observation has value s with probability 1. When $\tau < 0$, the observation decays in the reverse direction. The probability density of a random variable X whose distribution is $ED(s, \tau)$ for a negative τ is given by:

$$\text{dens}(X = x) = \begin{cases} \frac{-1}{\tau} e^{-\frac{(x-s)}{\tau}}, & \text{if } x \leq s, \\ 0, & \text{if } x > s. \end{cases}$$

Given a random variable X whose distribution is $ED(s, \tau)$, the random variable $X - s$ will have distribution $ED(0, \tau)$, and the mean of $X - s$ is τ . If τ is nonzero, then $\frac{X-s}{\tau}$ has distribution $ED(0, 1)$. As already discussed above, the median of $ED(s, \tau)$ is $s + \ln(2)\tau$. Because this distribution is so highly asymmetric, it is more reasonable to use the median rather than the mean when it is necessary to reduce the distribution to an estimate consisting of a single number.

3 Information Combination of Exponential Distributions

We now compute the distribution of the combination of a set of N independent exponentially distributed random variables. The following is the general case:

Theorem 3.1 *Let X_i be a set of independent random variables whose distributions are $ED(s_i, \tau_i)$, for $i = 1, \dots, N$. Write Y for the combination of the random variables $\{X_i\}$.*

1. *If $\tau_i > 0$ for every i , then Y has the distribution $ED(\max_{i=1}^N(s_i), (\sum(\frac{1}{\tau_i}))^{-1})$.*
2. *If $\tau_i < 0$ for every i , then Y has the distribution $ED(\min_{i=1}^N(s_i), (\sum(\frac{1}{\tau_i}))^{-1})$.*

Proof. We first consider the case in which all the τ_i are positive. By the Continuous Information Combination Theorem[1], the probability density of Y is given by normalizing the product of the densities of the random variables X_i as follows:

$$\text{dens}(Y = y) = \begin{cases} 0, & \text{if } y < s_i \text{ for some } i, \\ C \prod_{i=1}^N e^{-\left(\frac{y-s_i}{\tau_i}\right)}, & \text{if } y \geq s_i \text{ for all } i, \end{cases}$$

where C is the normalization constant such that this formula defines a probability density function. Now $y \geq s_i$ for all i if and only if $y \geq \max_i(s_i)$. Let $s = \max_i(s_i)$. The formula above may then be written as follows:

$$\text{dens}(Y = y) = \begin{cases} 0, & \text{if } y < s, \\ C e^{-\sum_{i=1}^N \left(\frac{y-s_i}{\tau_i}\right)}, & \text{if } y \geq s. \end{cases}$$

The exponent in the formula above may be written as follows:

$$\begin{aligned} -\sum_{i=1}^N \left(\frac{y-s_i}{\tau_i}\right) &= -\sum_{i=1}^N \left(\frac{y}{\tau_i} - \frac{s_i}{\tau_i}\right) \\ &= -\sum_{i=1}^N \left(\frac{y}{\tau_i}\right) + \sum_{i=1}^N \left(\frac{s_i}{\tau_i}\right) \\ &= -y \sum_{i=1}^N \left(\frac{1}{\tau_i}\right) + \sum_{i=1}^N \left(\frac{s_i}{\tau_i}\right) \end{aligned}$$

Let

$$\tau = \frac{1}{\sum_{i=1}^N \left(\frac{1}{\tau_i}\right)},$$

or in terms of the rate parameters, $\lambda = \sum_{i=1}^N \lambda_i$. Also let

$$D = \frac{-s}{\tau} + \sum_{i=1}^N \left(\frac{s_i}{\tau_i}\right).$$

Note that D does not depend on y . One can then write the sum above as follows:

$$\begin{aligned}
-\sum_{i=1}^N \left(\frac{y - s_i}{\tau_i} \right) &= \frac{-y}{\tau} + \sum_{i=1}^N \left(\frac{s_i}{\tau_i} \right) \\
&= \frac{-y}{\tau} + \frac{s}{\tau} - \frac{s}{\tau} + \sum_{i=1}^N \left(\frac{s_i}{\tau_i} \right) \\
&= \frac{-y + s}{\tau} + D
\end{aligned}$$

The probability density when $y \geq s$ may then be computed as follows:

$$\begin{aligned}
C e^{-\sum \left(\frac{y-s_i}{\tau_i} \right)} &= C e^{\frac{-y+s}{\tau} + D} \\
&= C e^{-\left(\frac{y-s}{\tau} \right)} e^D \\
&= C e^D e^{-\left(\frac{y-s}{\tau} \right)}
\end{aligned}$$

Since $C e^D$ is a constant independent of y , it follows that Y is exponentially distributed with distribution $ED(s, \tau)$. In the same way, one can also compute the distribution of the combination when all of the τ_i are negative. The only difference is that the maximum of the $\{s_i\}$ is replaced by the minimum. The result then follows.

In the special case where all of the τ_i are the same, the formula simplifies to the following:

Corollary 3.2 *Let X_i be a set of independent random variables whose distributions are $ED(s_i, \tau)$, for $i = 1, \dots, N$. Write Y for the combination of the random variables $\{X_i\}$.*

1. *If $\tau > 0$, then Y has the distribution $ED(\max_{i=1}^N(s_i), \frac{\tau}{N})$.*
2. *If $\tau < 0$, then Y has the distribution $ED(\min_{i=1}^N(s_i), \frac{\tau}{N})$.*

The case in which some of the τ_i are positive and some are negative does not result in an exponential distribution, so it is not considered.

4 Experimental Observations

We now consider an experiment in which a particle splits into a set of N particles, each of which subsequently decays independently. Suppose that the half-lives of

the products are known to be $\{h_i | 1 \leq i \leq N\}$ and that one can observe the time when each of the products decays. The splitting of the original particle is called the *singular event*. The problem is to determine the time when the singular event occurred.

Theorem 4.1 *If a singular event results in N independent exponential decay processes with half lives $\{h_i | 1 \leq i \leq N\}$, then the time of the singular event decays backwards in time from the time when the first decay product is observed, with a half-life equal to $(\sum(\frac{1}{h_i}))^{-1}$.*

Proof. Let t be the time when the original particle split, and let $\{t_i\}$ be the times when the product particles are observed to decay. Then the duration from t to t_i has the distribution $ED(t, h_i/\ln(2))$, and the duration from t_i to t has the distribution $ED(t_i, -h_i/\ln(2))$. The latter set of distributions are all observing the same time t , and they are assumed to be independent, so the Information Combination Theorem applies. By Theorem 3.1, it follows that the distribution of the time t is

$$ED\left(\min_{i=1}^N(t_i), \frac{1}{\sum(\frac{\ln(2)}{h_i})}\right) = ED\left(\min_{i=1}^N(t_i), \frac{1}{\sum(\frac{1}{h_i})/\ln(2)}\right).$$

The result then follows.

In the special case where all of the half lives are the same, the formula simplifies as follows:

Corollary 4.2 *If a singular event results in N independent identically distributed exponential decay processes with half-life $t_{1/2}$, then the time of the singular event decays backwards in time from the time when the first decay product is observed, with a half-life equal to $t_{1/2}/N$.*

5 The Least Squares Method

We now apply the least squares method to the scenario of Section /refsec:experiment above. For simplicity we consider only the case where the N exponential distributions have the same half-life $t_{1/2}$. The mean lifetime is $\tau = t_{1/2}/\ln(2)$, and the variance is τ^2 . Let x_i be the actual decay observation for the i^{th} product particle. Each decay observation X_i may be written in the form $t + \tau + e_i$, where e_i is a random “error” having mean 0 and variance τ^2 . Therefore each $X_i - \tau$

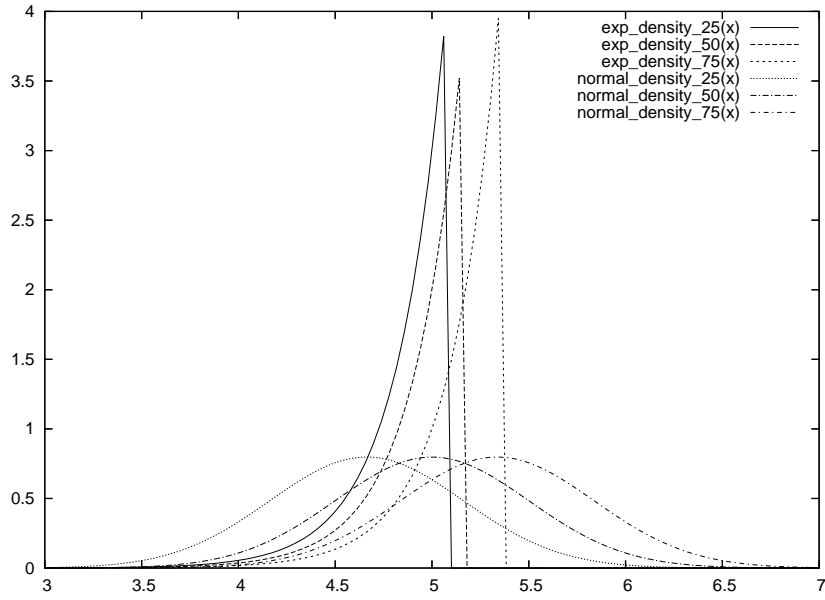


Figure 1: Information Combination versus Least Squares Methods for $N=4$

has mean t and variance τ^2 . The least squares estimate for t is the sample average $\frac{1}{N} \sum_{i=1}^N (x_i - \tau) = (\frac{1}{N} \sum_{i=1}^N x_i) - \tau$. This estimate has variance τ^2/N and standard deviation τ/\sqrt{N} . The information combination method with the least squares method are compared in Figure 2.

Method	Estimate	Variance	Standard deviation
Least Squares	$(\frac{1}{N} \sum_{i=1}^N x_i) - \tau$	τ^2/N	τ/\sqrt{N}
Information Combination	$\min_{i=1}^N (x_i) - \frac{\ln(2)\tau}{N}$	τ^2/N^2	τ/N

Figure 2: Comparison of Information Combination and Least Squares Methods

It is apparent that information combination will be more accurate than the least squares technique, especially for large N . For example, consider the case in which $t = 5$, $\tau = 1$ and $N = 4$. Information combination yields the distribution $ED(\min(t_i), -0.25)$. The median value for $\min(t_i)$ is $t + \ln(2)/4 = 5.17329$. If this experiment is run many times, the values for $\min(t_i)$ will vary. To show a representative sample of the behavior, the quartiles are shown for the two methods in Figure 1, and in the later graphs. The peaks of the exponential distribution

should all be the same, but the graphs do not show this due to round-off errors.

The least squares method yields the estimate $(\frac{1}{4} \sum t_i) - 1$. This estimate is normally distributed with mean 5 and standard deviation 0.5. Accordingly it will be within 0.3372 of 5 about half of the time. The quartiles are therefore at 4.6628, 5 and 5.3372. These are shown in Figure 1. The information combination distribution is approximately 4 times more accurate than the least squares distribution. Half of the improvement is due to the fact that the normal distribution is a two-tailed distribution while the exponential distribution is a single-tailed distribution. The other half of the improvement is the result of the improvement in the variance.

In Figure 3 the density functions are compared for the case in which $N = 16$. In this case the information combination distribution is approximately 8 times more accurate than the least squares distribution. In general the information combination distribution will be approximately $2\sqrt{N}$ times more accurate. The factor of 2 in this approximation is somewhat arbitrary, and represents the fact that the exponential distribution is single-tailed, while the normal distribution is double-tailed. Based on the estimates alone, one might take the factor to be 1.4. See Figure 5 which compares the estimates for various values of N .

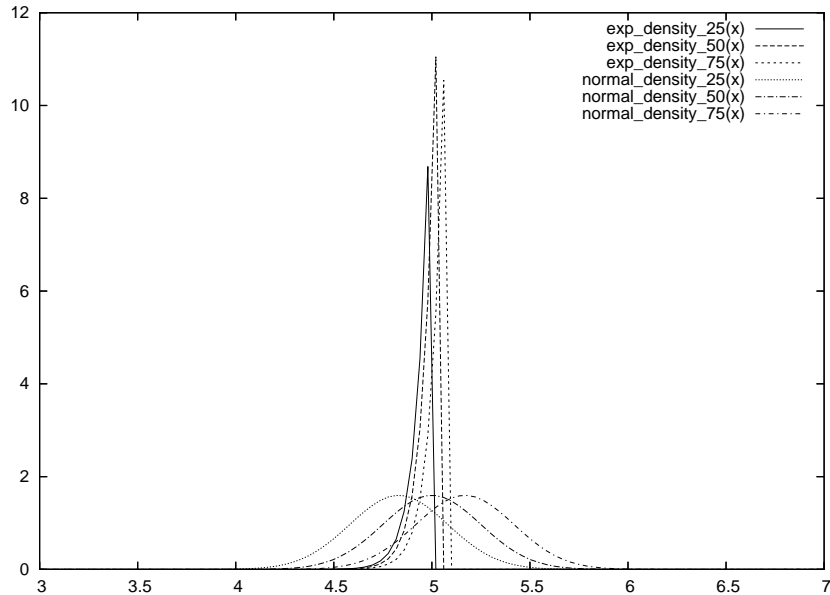


Figure 3: Comparison of Information Combination and Least Squares Methods for $N=16$

In Figure 4 the density functions are compared for the case in which $N = 1$. Here there is no information combination, so it is only comparing the exponential

distribution with the normal distribution.

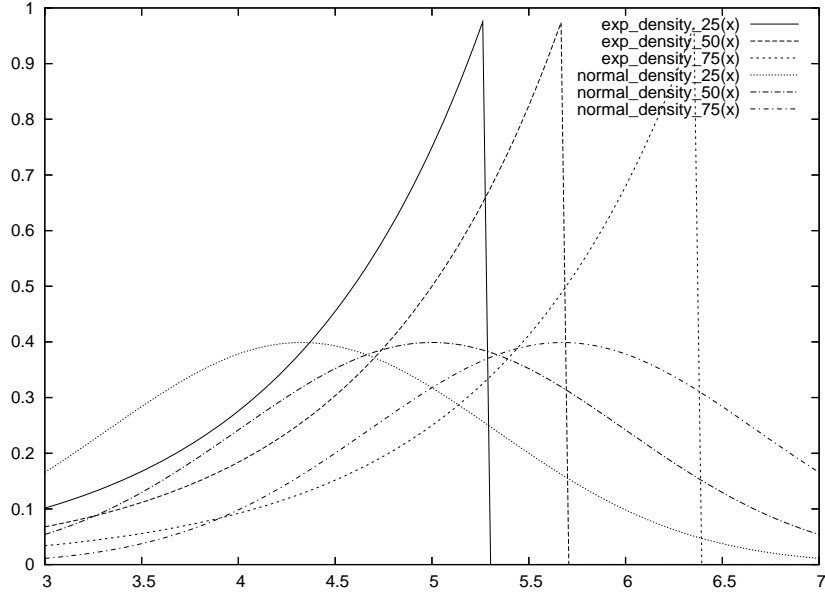


Figure 4: Comparison of Information Combination and Least Squares Methods for $N=1$

6 Conclusion

We have shown an example for which one can significantly improve estimation based on observations whose probability distributions are known, compared with techniques such as a least squares technique that do not require specific probability distributions. Using information combination rather than estimation has the further advantage that the result of the method is a probability distribution rather than a simple estimated value. Consequently the result may be used as input for further computations.

The information combination and least squares methods were compared for an example in which a singular event initiates an independent set of decay processes having known half-lives. The result can be viewed as reversing time, because our knowledge of the time of the singular event decays backwards in time with an exponential distribution whose parameters were computed. The accuracy of this technique was found to be significantly better than the accuracy obtained by using a least squares method.

Method	N	(25 th percentile)-5	(75 th percentile)-5
Information Combination	1	-0.693	0.405
Least Squares	1	-0.674	0.674
Information Combination	4	-0.173	0.101
Least Squares	4	-0.337	0.337
Information Combination	16	-0.043	0.025
Least Squares	16	-0.169	0.169

Figure 5: Estimates for Information Combination and Least Squares Methods

The information combination technique has many advantages, but there are only a few examples of classes of probability distributions to which the technique is currently being applied. Introducing additional classes of distributions that are closed under combination could have a significant impact on the accuracy of estimation in a variety of domains.

7 Acknowledgments

This work was performed while the author was on a sabbatical at the Division of Preventive Medicine, Brigham and Women’s Hospital, and the author wishes to thank Tianhua Niu, JoAnn Manson and Simin Liu for their help and encouragement. I also appreciate the contributions by my many colleagues at Northeastern University, Versatile Information Systems, and Composable Logic, including Mitch Kokar and Jeff Smith. I especially appreciate Jeff Keller for carefully proof-reading the manuscript.

References

- [1] K. Baclawski and T. Niu. *Ontologies for Bioinformatics*. MIT Press, Cambridge, MA, October 2005.

- [2] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, 53:370–418, 1763. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.
- [3] M. Fink. Time reversal of ultrasonic fields—part 1: Basic principles. *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, 39(5):555–566, September 1992.
- [4] M. Fink. Time-reversed acoustics. *Scientific American*, pages 91–97, November 1999.
- [5] M. Fink. Acoustic time-reversal mirrors. *Topics Appl. Phys.*, 84:17–43, 2002.
- [6] C.F. Gauss. *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*. F. Perthes and I.H. Besser, Hamburg, 1809.
- [7] C.F. Gauss. *Theoria combinationis observationum erroribus minimis obnoxiae: Pars prior*. Gottingische gelehrte Anzeigen, 1821.
- [8] C.F. Gauss. *Theoria combinationis observationum erroribus minimis obnoxiae: Pars posterior*. Gottingische gelehrte Anzeigen, 1823.
- [9] Pierre-Simon Laplace. Mémoire sur la probabilité des causes par les événements. *Savants Étranges*, 6:621–656, 1774.