Preprint of paper to be presented at the Seventh Annual Pittsburgh Conference on Modeling and Simulation, Univ. of Pittsburgh, Pennsylvania, April 26-27, 1976.

THE STATISTICAL SIGNIFICANCE OF ERROR PROBABILITY AS DETERMINED FROM DECODING SIMULATIONS FOR LONG CODES

James L. Massey
Freimann Professor of Electrical Engineering
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

ABSTRACT

The very low error probability obtained with long error-correcting codes results in a very small number of observed errors in simulation studies of practical size and renders the usual confidence interval techniques inapplicable to the observed error probability. A natural extension of the notion of a "confidence interval" is made and applied to such determinations of error probability by simulation. An example is included to show the surprisingly great significance of as few as two decoding errors in a very large number of decoding trials.

INTRODUCTION

The whole point of using error-correcting codes in a communication system is to make the decoding error probability, P_e, very small. It has been well known since the work of Shannon [1] that P_e can be made arbitrarily small provided only that the code rate is less than the capacity of the channel. Shannon proved this fact by showing that the average of P_e, taken over the ensemble of all suitably long codes used with maximum likelihood decoders, could be made arbitrarily small. For a particular long code, however, it is well-nigh impossible to calculate P_e even for a maximum-likelihood decoder. This fact places the coding engineer on the horns of dilemma. How can he be sure, once he has chosen a particular long code and a "good" decoder for that code, that he does not have one of the "bad" codes in the ensemble for which P_e is not small at all? The obvious answer is to determine P_e for his system by a simulation in which enough decoding decisions are made so that P_e can be accurately inferred. The catch is that, if P_e is indeed very small, a very large number of decodings will be required; it becomes a practical necessity that the simulation experiment be carefully designed to hold this number to a minimum.

A specific example from our own recent experience may sharpen the issues involved in determining P_e by such decoding simulations. Our task was to compare P_e for several constraint length K = 24, rate R = 1/2, binary convolutional codes when used with sequential decoding [2] on a 3-bit quantized Gaussian channel (the "deep-space channel") with an energy-per-information-bit to one-sided-noise power-spectral-density ratio E_b/N_o of 3 db. The decoding was done on "frames" of 256 information bits. P_e is the frame error probability, i.e., the probability that not all 256 information bits are correctly decoded. For code A, 2 frame errors were observed among 10,000 frames decoded on the simulated deep-space channel. For code B, no frame errors were observed among 10,000 decoded frames; we then decoded 40,000 further frames for code B with the same result -- no errors. Could we now with any reasonable confidence assert that the P_e for code B (no errors in 50,000 trials) was actually greater than that for code A (2 errors in 10,000 trials)? [The fact that we had already made 256 x 50,000 ≈ 10^7 bit decoding decisions for code B may indicate why we were not particularly eager to perform further decodings, even if we could have found the money for the necessary computer time.] The analysis given below was motivated by our very real need to answer that question.

CONFIDENCE INTERVALS FOR POISSON RANDOM VARIABLES WITH SMALL MEANS

Because the true error probability P_e of a decoding error is very small (say 10^{-4} or less) and the number of decoding trials N is very large (say 10^4 or more) in decoding simulation experiments for very long codes, the actual number of decoding errors X in the N trials is a Poisson-distributed random variable whose mean and variance (which always coincide for Poisson random variables) is

$$\lambda = NP_e. \tag{1}$$

Suppose that X = x is observed where x is small, say x = 2 for specificity. Then x is our best estimate of both E(X) and Var(X), and hence x/N is our best estimate of P_e. But we are stymied in our attempts to use the usual confidence intervals of statistics to determine the "confidence" that we should place in x/N because our estimate of the standard deviation of X is of the same order as x -- $\sqrt{2}$ in our example -- and nearly as unreliable as our estimate of x itself. We are sharply brought to realize that the use of the usual confidence intervals in statistics to quantify the closeness of a sample mean to the average of a random variable requires that the sample standard deviation itself have a standard deviation that is small. But this is glaringly not the case for a Poisson random variable when the observation X = x is small. We now proceed to define confidence intervals for such Poisson random variables in a way that both seems natural and proves simple to use.

Suppose we observe $X = x$ for a random variable X which we know to be Poisson. Because X is Poisson, its complete statistical description is contained in the parameter λ. We now define the $100\beta\%$ confidence interval for λ, given the observation $X = x$, as the interval (λ_L, λ_U) such that (1) λ_U is the smallest number such that the probability that X would exceed x is at least β for all λ such that $\lambda \geq \lambda_U$, and (2) λ_L is the smallest number such that the probability that X would fall beneath x is at least β for all λ such that $\lambda \leq \lambda_L$. (By way of convention, we take $\lambda_L = 0$ if $x = 0$.) In other words, the probability is at least β that we would not have made an observation $X \leq x$ given that X had a true λ greater than λ_U, and at least β that we would not have made an observation $X \geq x$ given that X had a true λ less than λ_L. This definition seems to us to be in accord with the intuitive notion of a "confidence interval."

The cumulative distribution function $F(x) = P(X \leq x)$ for a Poisson random variable may be written

$$F(x) = \sum_{i=0}^{x} \frac{\lambda^i}{i!} e^{-\lambda}. \tag{2}$$

It will be convenient to consider the right side of (2) as a function of both x and λ which, in a slight abuse of notation, we denote as $F(x, \lambda)$. It follows immediately from our definition of the $100\beta\%$ confidence interval (λ_L, λ_U) for the observation $X = x$ that

$$F(x-1, \lambda_L) = \beta \tag{3}$$

and

$$F(x, \lambda_U) = 1-\beta. \tag{4}$$

Equations (3) and (4) are transcendental equations for λ_L and λ_U which, upon using the definition of $F(x, \lambda)$ as the right side of (2), may be solved by any convenient numerical method for a given value of x.

In the Table, we give the 90% confidence intervals ($\beta = .9$) as determined by the solution of (3) and (4) for $x = 0, 1, 2, 3, 4, 5$. The remarkable feature of this table is the extremely rapid contraction of the confidence interval as x increases.

x	λ_L	λ_U
0	0	2.30
1	.105	3.88
2	.532	5.29
3	1.107	6.65
4	1.76	7.99
5	2.52	9.06

Table. 90% Confidence Intervals (λ_L, λ_U) for $X = x$.

AN EXAMPLE

We now apply the techniques of the previous section to the example described in the Introduction.

For Code A, the observation $X = 2$ gives the 90% confidence interval (.532, 5.29) for λ according to the Table. Hence, we are 90% confident that the error probability $P_e = \lambda/N = \lambda/10{,}000$ lies in the range

$$5.32 \times 10^{-5} \leq (P_e)_A \leq 5.29 \times 10^{-4}. \tag{5}$$

For Code B, the observation $X = 0$ gives the 90% confidence interval (0, 2.30) for λ according to the Table. Hence, we are 90% confident that the error probability $P_e = \lambda/N = \lambda/50{,}000$ lies in the range

$$0 \leq (P_e)_B \leq 4.60 \times 10^{-5}. \tag{6}$$

Upon comparing (5) and (6), we see that we can with confidence slightly exceeding 90% assert that the decoding error probability for code B is indeed less than that for code A -- even though only two decoding errors were observed during the entire simulation in which 60,000 frames were decoded!

It should be apparent to the reader how the results of this paper can be used for the design of efficient simulation experiments for determining the decoding error probabilities for long codes at some specified confidence level.

ACKNOWLEDGMENT

The support of this research by the National Aeronautics and Space Administration, under NASA Grant NSG 5025 at the University of Notre Dame in liaison with the Goddard Space Flight Center, is gratefully acknowledged.

REFERENCES

(1) Shannon, C. E., "A Mathematical Theory of Communication," BELL SYSTEM TECHNICAL JOURNAL, Vol 27 (1948), 379-423, 623-656.

(2) Fano, R. M., "A Heuristic Discussion of Probabilistic Decoding," IEEE TRANS. ON INFORMATION THEORY, Vol IT-19 (1963), 64-73.