QUARTERLY OF APPLIED MATHEMATHICS

Vol. XVIII October, 1960 No. 3

ON THE STRONG CONVERSE OF THE CODING THEOREM FOR SYMMETRIC CHANNELS WITHOUT MEMORY[1]

BY

LIONEL WEISS

Cornell University

1. Introduction. First we describe the channels we will discuss in this paper. The alphabet in which we send words contains just two letters, which we denote by 0, 1. Thus a sent word of length n is a sequence of n letters, each letter being a zero or a one. If we send the word $(x_1, \cdots, x_n)$, then the received word is $(Y_1, \cdots, Y_n)$, where $Y_1, \cdots, Y_n$ are independent chance variables, the probability distribution of Y_i depending only on the value of the parameter x_i. If there are just two possible values for Y_i, which we can assume are 0, 1, the channel is called a "binary channel." If Y_i has a probability density function whether x_i is zero or one, the channel is called a "semi-continuous channel." For a binary channel, $f_{x_i}(y)$ denotes $P(Y_i = y$, when the ith letter sent is $x_i)$. For a semi-continuous channel, $f_{x_i}(y)$ denotes the probability density function of Y_i when the ith letter sent is x_i.

A binary channel is called "symmetric" if $f_0(0) = f_1(1)$ (and therefore $f_0(1) = f_1(0)$). A semi-continuous channel is called "symmetric" if the probability distribution of $2[f_0(Y) + f_1(Y)]^{-1}f_0(Y)$ when Y has the probability density function $f_0(y)$ is the same as the probability distribution of $2[f_0(Y) + f_1(Y)]^{-1}f_1(Y)$ when Y has the probability density function $f_1(y)$.

A code of length L, word length n, and probability of error not greater than λ is a sequence of L pairs $(u_1, A_1), \cdots, (u_L, A_L)$ with the following properties:

(a) For each i, u_i is a sequence of zeros and ones, n symbols altogether;

(b) For each i, A_i is a collection of words of length n, each being a possible received word in the particular problem under consideration;

(c) For each i, when u_i is the sent word, $P((Y_1, \cdots, Y_n) \text{ is in } A_i) \geq 1 - \lambda$;

(d) The sets $A_1, \cdots, A_L$ are disjoint.

The use of such a code is well known. If the received word is in A_i, the receiver assumes that the word u_i was actually sent. Then, no matter which of the words $u_1, \cdots, u_L$ is sent, the probability that the receiver will be in error is not greater than λ.

A "converse to the coding theorem" is an inequality giving an upper bound for L, or, alternatively, for $\log_2 L$. (In this paper, each log is to the base e, unless another base is explicitly indicated.) Since several types of converse have appeared in the literature, it seems worthwhile to classify them carefully. The following classification seems useful.

[1]Received May 25, 1959. This research was performed under contract with the Office of Naval Research.

C denotes the capacity of the channel as defined in Feinstein [2]. A "weak converse" states the following. For any fixed $\epsilon > 0$, there is a positive value Z, such that there cannot exist a code of length $2^{n(C+\epsilon)}$ and probability of error not greater than Z, for large n. A "strong converse" states the following. For any fixed $\epsilon > 0$, and any fixed λ in the open interval (0, 1), there cannot exist a code of length $2^{n(C+\epsilon)}$ and probability of error not greater than λ, for n large. A "stronger converse" states the following. For any fixed λ in the open interval (0, 1), there is a constant K_λ such that there cannot exist a code of length $2^{nC+K_\lambda n^{1/2}}$ and probability of error not greater than λ, for large n.

Clearly, each of the converses listed is stronger than its predecessors. An interesting weak converse applicable to many types of channels is given in Theorem 5 of [3]. Wolfowitz [4] has proved the stronger converse for a binary channel, with K_λ a positive constant, and [5] has proved the strong converse for a semi-continuous channel. It is the purpose of the present paper to show that for both binary symmetric channels and semi-continuous symmetric channels, the stronger converse can be proved with K_λ a negative constant if $\lambda < \frac{1}{2}$. This makes the converse still stronger.

We define β_λ by the equation

$$\int_{-\infty}^{\beta_\lambda} (2\pi)^{-1/2} \exp\left(-\tfrac{1}{2}t^2\right) dt = 1 - \lambda.$$

Then if $\lambda < \frac{1}{2}$, $\beta_\lambda > 0$.

2. The binary symmetric channel. We denote the common value of $f_0(0)$ and $f_1(1)$ by q, and $1 - q$ by p. We assume that q is in the open interval $(\frac{1}{2}, 1)$. The capacity C for this channel is $1 + p \log_2 p + q \log_2 q$.

We have the following theorem. *For any $\delta > 0$, there is a number n_δ such that if $n > n_\delta$, then the length L of any code of word length n and probability of error not greater than λ must satisfy the inequality* $\log_2 L < nC - n^{1/2} [\beta_\lambda (pq)^{1/2} \log_2(q/p) - \delta]$.

Proof. If the word u_i is sent, the most probable received word is u_i itself, with probability q^n; the $\binom{n}{1}$ words differing from u_i in exactly one letter are tied for next most probable received word, each having probability pq^{n-1}; the $\binom{n}{2}$ words differing from u_i in exactly two letters are tied for next most probable received word, each having probability p^2q^{n-2}; etc.

We define K as the largest integer such that

$$\sum_{j=0}^{K} \binom{n}{j} p^j q^{n-j} \le 1 - \lambda. \tag{1}$$

Denote by M_i the number of words in A_i. Then we must have

$$M_i \ge \sum_{j=0}^{K} \binom{n}{j},$$

since even the $\sum_{j=0}^{K} \binom{n}{j}$ which are the most probably received words have a total probability no greater than $1 - \lambda$.

Since there are 2^n different words of length n, and since $A_1, \cdots, A_L$ are disjoint, we have

$$2^n \geq \sum_{i=1}^{L} M_i \geq L \sum_{j=0}^{K} \binom{n}{j}, \quad \text{or} \quad L \leq 2^n \Big/ \sum_{j=0}^{K} \binom{n}{j} < 2^n \Big/ \binom{n}{K}.$$

Since on the left-hand side of (1) we are summing binomial probabilities, and since the binomial distribution approaches the normal distribution as n increases, we have by the central-limit theorem that $K = np + \beta(n, \lambda)\,(npq)^{1/2}$, where for any given positive value Δ, a number n_Δ can be found such that if $n > n_\Delta$, then $\beta_\lambda - \Delta \leq \beta(n, \lambda) \leq \beta_\lambda + \Delta$. We denote $\beta(n, \lambda)\,(pq)^{1/2}$ by B. Then $K = np + Bn^{1/2}$, where for $n > n_\Delta$, $(\beta_\lambda - \Delta)\,(pq)^{1/2} \leq B \leq (\beta_\lambda + \Delta)\,(pq)^{1/2}$.

Now

$$1 \Big/ \binom{n}{K} = K!(n - K)!/n!\,,$$

and using Stirling's inequalities $r^r(2\pi r)^{1/2} \exp(-r) < r! < r^r(2\pi r)^{1/2} \exp(-r + 1/12r)$, we find

$$1 \Big/ \binom{n}{K} < \frac{K^K(2\pi K)^{1/2} \exp(-K+1/12K)(n-K)^{n-K}[2\pi(n-K)]^{1/2} \exp\{-n+K+1/12(n-K)\}}{n^n(2\pi n)^{1/2} \exp(-n)}. \tag{2}$$

Setting $K = np + Bn^{1/2}$, simplifying the right-hand side of (2), and taking logs, we get

$$-\log \binom{n}{K} < \frac{1}{2} \log(2\pi n) + \frac{1}{12[npq + Bn^{1/2}(q - p) - B^2]} + (np + Bn^{1/2} + \tfrac{1}{2}) \log(p + Bn^{-1/2}) + (nq - Bn^{1/2} + \tfrac{1}{2}) \log(q - Bn^{-1/2}). \tag{3}$$

Expanding $\log(p + Bn^{-1/2})$ and $\log(q - Bn^{-1/2})$ around p and q respectively, and substituting into (3), we get

$$-\log \binom{n}{K} < n(p \log p + q \log q) - n^{1/2}B \log(q/p) + \Omega \log n, \tag{4}$$

where Ω is a positive constant which does not depend on n. Converting the logarithms in (4) to the base 2, we find

$$-\log_2 \binom{n}{K} < n(p \log_2 p + q \log_2 q) - n^{1/2}[B \log_2(q/p) - \Omega n^{-1/2} \log_2 n].$$

Then

$$\log_2 L < n(1 + p \log_2 p + q \log_2 q) - n^{1/2}[B \log_2(q/p) - \Omega n^{-1/2} \log_2 n].$$

If

$$n > n_\Delta\,, \log_2 L < nC - n^{1/2}[\beta_\lambda(pq)^{1/2} \log_2(q/p) - \Delta(pq)^{1/2} \log_2(q/p) - \Omega n^{-1/2} \log_2 n].$$

This proves the theorem, since Δ can be taken as close to zero as desired, and $n^{-1/2} \log_2 n$ approaches zero as n increases.

3. A theorem of Cramér. Before discussing the semi-continuous channel, we quote a theorem of Cramér, [1], which will be used later.

$Z_1, Z_2, \cdots$ are independent, identically distributed chance variables, each with probability density function $v(z)$, expectation zero, and finite positive variance σ^2. There is a value $A > 0$ such that

$$R(h) = \int_{-\infty}^{\infty} \exp(hz)v(z)\,dz \quad \text{converges for} \quad |h| < A.$$

$m(h)$ denotes the first derivative of log $R(h)$ with respect to h, and $\sigma_1^2(h)$ denotes the second derivative of log $R(h)$ with respect to h. We define A_1 as sup $\{h \mid R(h) \text{ converges}\}$. From our assumption, A_1 is positive and may be infinite. M is defined as $\sigma^{-1} \lim_{h \to A_1 - 0} m(h)$. M is positive and may be infinite.

Theorem. For any value g in the open interval $(0, M)$, *the equation* $m(h) = \sigma g$ *has a unique root* $h(g)$, *which is positive.* $P(Z_1 + \cdots + Z_n \geq \sigma gn) =$

$$n^{-1/2}[\{h(g)\sigma_1(h(g))(2\pi)^{1/2}\}^{-1} + Q(n, g)] \exp\{-n[h(g)m(h(g)) - \log R(h(g))]\},$$

where for any values γ_1, γ_2 *with* $0 < \gamma_1 < \gamma_2 < M$, *there is a positive finite value* $B(\gamma_1, \gamma_2)$ *with* $n \mid Q(n, g) \mid < B(\gamma_1, \gamma_2)$ *for all g in the closed interval* $[\gamma_1, \gamma_2]$.

4. The semi-continuous symmetric channel. We assume that

$$\int_{-\infty}^{\infty} \tfrac{1}{2}(f_0(y) + f_1(y))[\log\{2[f_0(y) + f_1(y)]^{-1}f_0(y)\}]^2\,dy$$

exists. Also, for any value y for which $f_0(y) = f_1(y) = 0$, we consider $2[f_0(y) + f_1(y)]^{-1}f_i(y)$ as zero for $i = 0, 1$, and $0 \log^k 0$ is always to be considered as equal to zero, for any positive k. The capacity C for the semi-continuous symmetric channel is

$$\int_{-\infty}^{\infty} f_0(y) \log_2\{2[f_0(y) + f_1(y)]^{-1}f_0(y)\}\,dy.$$

For any given sequence $(x_1, \cdots, x_n)$ of zeros and ones, $W(x_1, \cdots, x_n)$ denotes a subset of $(Y_1, \cdots, Y_n)$ space such that

(a) $P((Y_1, \cdots, Y_n) \text{ is in } W(x_1, \cdots, x_n)) \geq 1 - \lambda$, when the joint probability density function of $Y_1, \cdots, Y_n$ is $f_{x_1}(y_1) f_{x_2}(y_2) \cdots f_{x_n}(y_n)$;

(b) $P((Y_1, \cdots, Y_n) \text{ is in } W(x_1, \cdots, x_n))$ is minimized when the joint probability density function of $Y_1, \cdots, Y_n$ is

$$\prod_{i=1}^{n} [\tfrac{1}{2}(f_0(y_i) + f_1(y_i))],$$

subject to (a). Then it is clear that the probability in (a) will be exactly $1 - \lambda$. We want to find the value of the probability in (b), which we denote by P^*.

By the familiar Neyman-Pearson lemma, the region $W(x_1, \cdots, x_n)$ is given by the set of points $(Y_1, \cdots, Y_n)$ with

$$\prod_{i=1}^{n} \{2[f_0(Y_i) + f_1(Y_i)]^{-1}f_{x_i}(Y_i)\} \geq k,$$

where k is a properly chosen constant. Alternatively, $W(x_1, \cdots, x_n)$ is the set of points $(Y_1, \cdots, Y_n)$ with

$$\sum_{i=1}^{n} \log\{2[f_0(Y_i) + f_1(Y_i)]^{-1}f_{x_i}(Y_i)\} \geq \log k.$$

By the definition of the symmetry of the channel, the distribution of

$$\sum_{i=1}^{n} \log \{2[f_0(Y_i) + f_1(Y_i)]^{-1} f_{x_i}(Y_i)\}$$

when the joint probability density function of $Y_1, \cdots, Y_n$ is $\prod_{i=1}^{n} f_{x_i}(y_i)$ does not depend on the sequence $(x_1, \cdots, x_n)$. Therefore the value of $\log k$ does not depend on the sequence $(x_1, \cdots, x_n)$.

Also, the probability P^* does not depend on the sequence $(x_1, \cdots, x_n)$. To show this, it suffices to show that the distribution of $2[f_0(Y) + f_1(Y)]^{-1} f_0(Y)$ when Y has the probability density function $\frac{1}{2}(f_0(y) + f_1(y))$ is the same as the distribution of $2[f_0(Y) + f_1(Y)]^{-1} f_1(Y)$ when Y has the probability density function $\frac{1}{2}(f_0(y) + f_1(y))$. The tth moments of these expressions are respectively

$$\int_{-\infty}^{\infty} 2^{t-1} [f_0(y) + f_1(y)]^{-t+1} f_0^t(y)\, dy, \qquad \int_{-\infty}^{\infty} 2^{t-1} [f_0(y) + f_1(y)]^{-t+1} f_1^t(y)\, dy.$$

But these integrals are respectively the $(t - 1)$st moment of the chance variable $2[f_0(Y) + f_1(Y)]^{-1} f_0(Y)$ when Y has the probability density function $f_0(y)$, and the $(t - 1)$st moment of the chance variable $2[f_0(Y) + f_1(Y)]^{-1} f_1(Y)$ when Y has the probability density function $f_1(y)$, and therefore are equal. Since the moments determine the distribution of a bounded chance variable, the demonstration is completed.

Therefore neither $\log k$ nor P^* depends on the sequence $(x_1, \cdots, x_n)$, and it is no loss of generality to assume that $x_1 = \cdots = x_n = 0$. We denote

$$\int_{-\infty}^{\infty} f_0(y) \log \{2[f_0(y) + f_1(y)]^{-1} f_0(y)\}\, dy \quad \text{by} \quad H,$$

and

$$\int_{-\infty}^{\infty} f_0(y)(\log \{2[f_0(y) + f_1(y)]^{-1} f_0(y)\})^2\, dy - H^2 \quad \text{by} \quad J^2,$$

J being taken as positive. Then, by the central-limit theorem, $\log k$ is equal to $nH - \beta(n, \lambda) n^{1/2} J$, where $\beta(n, \lambda)$ has the properties described in Sec. 2.

We denote

$$\int_{-\infty}^{\infty} \tfrac{1}{2}[f_0(y) + f_1(y)] \log \{2[f_0(y) + f_1(y)]^{-1} f_0(y)\}\, dy \quad \text{by} \quad S,$$

and

$$\int_{-\infty}^{\infty} \tfrac{1}{2}[f_0(y) + f_1(y)](\log \{2[f_0(y) + f_1(y)]^{-1} f_0(y)\})^2\, dy - S^2 \quad \text{by} \quad \sigma^2, \sigma$$

being taken as positive. To find P^*, we use Cramér's theorem of Sec. 3, with $Z_i = \log \{2[f_0(Y_i) + f_1(Y_i)]^{-1} f_0(Y_i)\} - S$, and $\sigma g n = nH - nS - \beta(n, \lambda) n^{1/2} J$, so that $g = (H - S)/\sigma - \beta(n, \lambda) J/\sigma n^{1/2}$. First we must verify that g is in the open interval $(0, M)$. It is easily shown that $m(h)$ is equal to

$$-S + \frac{\int_{-\infty}^{\infty} [\frac{1}{2}(f_0(y) + f_1(y))]^{-h+1} f_0^h(y) \log \{2[f_0(y) + f_1(y)]^{-1} f_0(y)\}\, dy}{\int_{-\infty}^{\infty} [\frac{1}{2}(f_0(y) + f_1(y))]^{-h+1} f_0^h(y)\, dy},$$

from which it follows that $m(1) = H - S$, and that $m(h)$ is a strictly increasing function of h for positive h, except in the trivial case (which we exclude) where $f_0(y) = f_1(y)$ almost everywhere. It is easily verified that $H - S$ is positive. This proves that g is in the open interval $(0, M)$, at least for large values of n.

Our next task is to find $h(g)$. We note that $dm(h)/dh$ is continuous in a neighborhood of $h = 1$, and is equal to J^2 at $h = 1$. Therefore we have $m(h) = m(1) + (h - 1)J^2 + \epsilon(h - 1)$, where $(1/r)\ \epsilon(r)$ approaches zero as r approaches zero. The equation $m(h(g)) = \sigma g$ becomes

$$H - S + (h(g) - 1)J^2 + \epsilon(h(g) - 1) = H - S - \beta(n, \lambda)Jn^{-1/2},$$

or

$$h(g) = 1 - \beta(n, \lambda)/(n^{1/2}J) + \delta_n ,$$

where $n^{1/2}\ \delta_n$ approaches zero as n increases.

Substituting this value of $h(g)$ in Cramér's theorem, we find that $\log P^* = -n[H - \beta(n, \lambda)\ Jn^{-1/2} + \Omega_n]$, where $n^{1/2}\ \Omega_n$ approaches zero as n increases. Denote by J' the quantity that J becomes when logarithms to the base 2 are used in the definition instead of logarithms to the base e. Then $\log_2 P^* = -n[C - \beta(n, \lambda)\ J'n^{-1/2} + D_n]$, where $n^{1/2}D_n$ approaches zero as n increases.

If we have a code $(u_1 , A_1), \cdots , (u_L , A_L)$ of length L and probability of error not greater than λ, it follows from our discussion that $P((Y_1 , \cdots , Y_n)$ is in $A_i) \geq P^*$, when the joint probability density function of $Y_1 , \cdots , Y_n$ is $\prod_{i=1}^{n} [\frac{1}{2}(f_0(y_i) + f_1(y_i))]$. Since $A_1 , \cdots , A_L$ are disjoint, it follows that $LP^* \leq 1$, or that $\log_2 L \leq n[C - \beta(n, \lambda)\ J'n^{-1/2} + D_n]$, where $n^{1/2}\ D_n$ approaches zero as n increases. Thus the stronger converse is proved for the semi-continuous symmetric channel.

5. Acknowledgment. The author would like to thank Professor J. Wolfowitz for many helpful discussions of coding theory.

References

1. H. Cramér, "Sur un nouveau théorème-limite de la théorie des probabilités," Colloque consacré à la théorie des probabilités, Hermann et Cie, Paris, 1938
2. A. Feinstein, "Foundations of information theory," McGraw-Hill, New York, 1958
3. C. E. Shannon, "Certain results in coding theory for noisy channels," Information and Control **1**, 6–25 (1957)
4. J. Wolfowitz, "The coding of messages subject to chance errors," Illinois Journal of Mathematics **1**, 591–606 (1957)
5. J. Wolfowitz, "Strong converse of the coding theorem for semi-continuous channels," to be published