# LEAST-FAVORABLE NOISE DISTRIBUTIONS FOR FIXED SIGNAL-TO-NOISE RATIOS AND BINARY SIGNALLING

Shlomo Shamai (Shitz)
Dept. Electrical Eng.
Technion
Haifa 32000, ISRAEL

Sergio Verdú
Dept. Electrical Eng.
Princeton University
Princeton, NJ 08544, USA

## ABSTRACT

Additive noise channels with binary-valued inputs and real-valued outputs are considered. We study both the maximum likelihood error probability and channel capacity, and we find the least-favorable noise distributions under both criteria with a constraint on the signal-to-noise ratio (SNR). The error probability decreases linearly with SNR (for SNR > 0 dB) rather than exponentially as in the Gaussian-noise case. The maximum difference between the Gaussian and worst-case capacity is 0.118 bit (for SNR = 7.2 dB).

## 1. INTRODUCTION

The error probability and the capacity of binary-input additive-noise channels are well-known if the noise is Gaussian. A basic problem in communication theory is to find the worst-case performance achievable by any noise distribution as a function of the signal-to-noise ratio. In addition to its applications to channels subject to jamming (where the least-favorable power-constrained noise distribution is of interest), the worst-case performance provides a baseline of comparison for any non-Gaussian channel in which the receiver knows the noise statistics. This paper gives a complete solution to this problem for the two major performance measures: error probability and capacity.

Consider the binary equiprobable hypothesis testing problem:

$$H_1: \quad \mathbf{Y} = +1 + \mathbf{N}$$

$$H_0: \quad \mathbf{Y} = -1 + \mathbf{N}$$

where $\mathbf{N}$ is a real-valued random variable constrained to satisfy an average-power limitation $E[\mathbf{N}^2] \leq \sigma^2$. In Section 2, we find an explicit expression for the least-favorable noise distribution and the worst-case error probability attained by the optimum (maximum-likelihood) detector, which knows the noise distribution. The worst-case error probability is depicted in Figure 1 along with its Gaussian counterpart $Q(\frac{1}{\sigma})$.

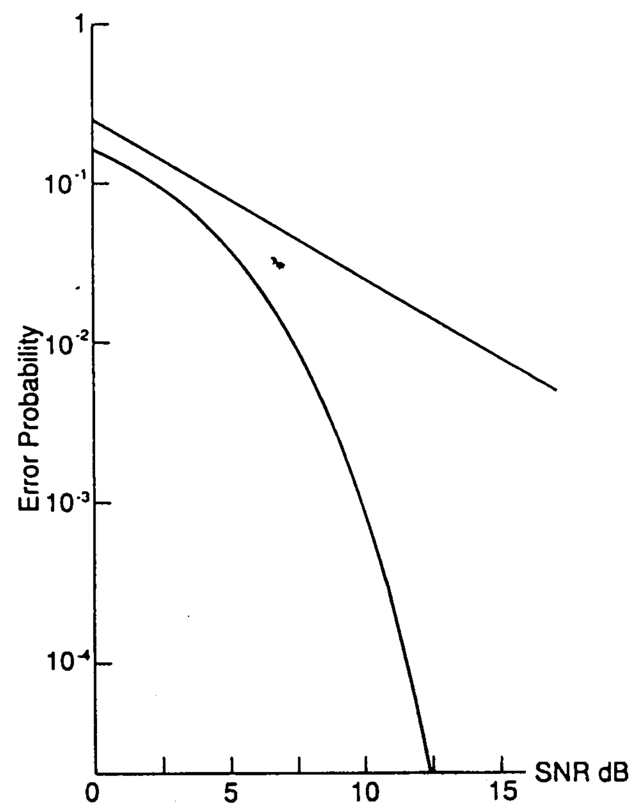


*Fig. 1. Gaussian and worst-case error probabilities.*

It is well-known [1,2] that the capacity of the additive memoryless channel

$$\mathbf{Y}_i = \mathbf{X}_i + \mathbf{N}_i$$



0-7803-0608-2/92/$3.00 © 1992 IEEE

is minimized, among all power-constrained noise distributions, by iid Gaussian noise, if the input is only constrained in power, in which case the optimal input is also iid Gaussian. If the input is binary valued (a ubiquitous constraint in many digital communication systems, such as direct-sequence spread spectrum) the worst-case capacity as well as the least-favorable noise distribution were previously unknown (beyond the fact shown in [3] that the worst-case noise is distributed on a lattice). The worst-case capacity of the binary-input channel is depicted in Figure 2 along with the Gaussian-noise capacity.

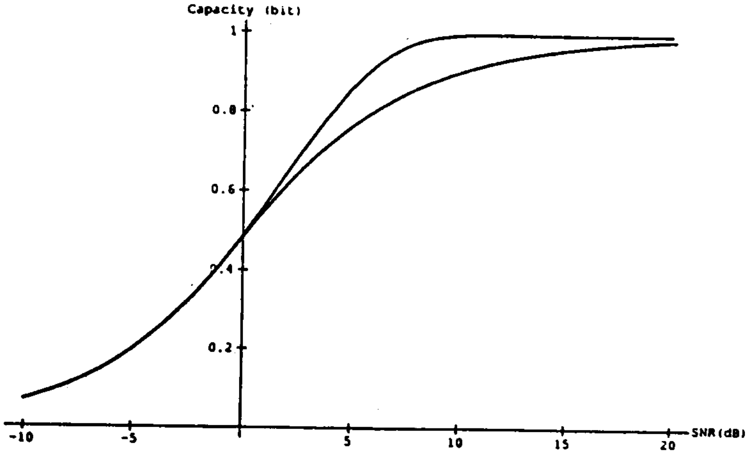


*Fig. 2. Gaussian and worst-case channel capacities.*

The foregoing results are obtained as an application of a general framework developed in [4] which applies to many other performance functionals of information-theoretic interest besides error probability and capacity, such as divergence, cutoff rate, random-coding error exponent, and Chernoff entropy. Those general results show that the worst-case performance functional is given by the convex hull of the functional obtained by minimizing only over power-constrained noise distributions which place all their mass on a lattice whose span is equal to the distance between the two inputs. This implies that the least-favorable distribution is, in general, the mixture of two lattice probability mass functions. This conclusion can actually be generalized to $m$-ary input constellations on finite-dimensional spaces, as long as the input constel ation puts its mass on a lattice. Then, the least-favorable .ioise distribution is a mixture of two distributions on lattices that are shifted versions of the input lattice [4].

The proof of the results presented in this paper can be found in its journal version [4].

## 2. ERROR PROBABILITY

**Theorem 1.** For $k = 1, 2, \cdots$ let

$$\sigma_k^2 \triangleq \frac{1}{3}(k^2 - 1)$$

The worst-case probability of error is

$$P_e(\sigma^2) = \frac{1}{2} - \frac{1}{2\sqrt{1 + 3\sigma_k^2}} + \frac{3}{2}\,\frac{\sigma^2 - \sigma_k^2}{k(k+1)(2k+1)} \qquad (1)$$

for $\sigma_k^2 \le \sigma^2 \le \sigma_{k+1}^2$.

Note that

$$\frac{1}{2} - \frac{1}{2\sqrt{3\sigma^2}} \le P_e(\sigma^2) \le \frac{1}{2} - \frac{1}{2\sqrt{3\sigma^2 + 1}}$$

where the lower bound is the error probability achieved when **N** is a zero-mean uniform random variable and the upper bound is a strictly concave function which coincides with $P_e(\sigma^2)$ at $\sigma^2 = \sigma_k^2$, $k = 1, 2, \cdots$.

We conclude from Theorem 1 that a single span-2 lattice achieves the maximum probability of error only when the allowed noise power is equal to $\sigma_k^2 = \frac{1}{3}(k^2 - 1)$, $k = 1, 2, \cdots$ Those worst-case distributions are symmetric and distribute their mass evenly on $k$ atoms. (Those atoms are located at $0, \pm 2, \pm 4, \cdots$ if $k$ is odd and at $\pm 1, \pm 3, \cdots$ if $k$ is even.) When the allowed noise power lies strictly between $\sigma_k^2 < \sigma^2 < \sigma_{k+1}^2$, then a single span-2 lattice is no longer least favorable. Instead, the worst-case distribution is the unique span-1 lattice which is a mixture of the span-2 lattices that are least-favorable for $\sigma_k^2$ and $\sigma_{k+1}^2$ with respective weights $(\sigma^2 - \sigma_{k+1}^2)/(\sigma_k^2 - \sigma_{k+1}^2)$ and $(\sigma_k^2 - \sigma^2)/(\sigma_k^2 - \sigma_{k+1}^2)$ (Figure 3). In particular, if $SNR > 0\ dB\,(\sigma^2 < 1)$, then the worst-case noise is symmetric with nonzero atoms at $-1, 0, +1$, i.e., the channel becomes a symmetric erasure channel. Thus, the noise distribution that maximizes error probability puts all its mass on the integers $\{-M, \cdots M\}$, where $M$ depends on the signal-to-noise ratio and the weight assigned to each of those integers depends (in addition to the signal-to-noise ratio) only on whether the integer is even or odd.

Note that for low signal-to-noise ratios, the worst-case noise cdf does not become asymptotically Gaussian, as might have been surmised from capacity considerations. In fact, for high $\sigma^2$, the Gaussian and worst-case error probabilities behave as

$$\frac{1}{2} - \frac{1}{\sqrt{2\pi}\,\sigma}$$

and

$$\frac{1}{2} - \frac{1}{\sqrt{12}\,\sigma},$$

respectively.

The nature of the least-favorable noise distribution implies that a sign decision (decide $H_1$ if $\mathbf{Y} > 0$, decide $H_0$ if $\mathbf{Y} < 0$ and arbitrary if $\mathbf{Y} = 0$) is a maximum-likelihood rule (yielding the minimum probability of error). It should be noted that the pair (*sign detector, worst-case noise*) achieves the maximin error probability solution, but not the minimax solution. For example, if $\sigma > 1$, the noise can make the error probability of the sign detector reach 1/2 by concentrating all its mass in the atoms $\sigma$ and $-\sigma$ in arbitrary proportion. Therefore, the game between the detection strategy and the power-constrained noise distribution has no saddle-point.

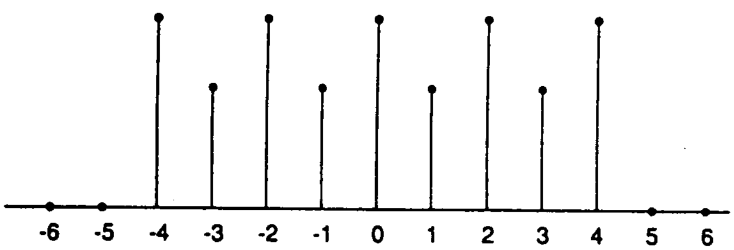


*Fig. 3. Least-favorable (error probability) noise distribution for $\sigma^2 = 7$ (SNR $= -8.45$ dB).*

As we remarked in the introduction, one of the main applications of the results in this paper is to communication in the presence of jamming. The nature of the solution found in this section may lead one to believe that both the worst-case probability curve and the least-favorable noise distribution are crucially dependent on the exact knowledge (from the jammer's viewpoint) of the input alphabet. Consider the case where the input values are $1+\delta_1$ and $-1+\delta_0$. If the jammer knew $\delta_1$ and $\delta_0$, then its best power-constrained strategy is to use the least-favorable discrete distribution found in this section on a span $1 + (\delta_1 - \delta_0)/2$ lattice in lieu of the integers. This would lead to a worst-case error probability curve given by $P_e((2\sigma/(2+\delta_1-\delta_0))^2)$. Thus, small perturbations of the input values lead to small deviations in the worst-case probability. But what if the jammer does not know the values of $\delta_1$ and $\delta_0$? If, for example, it assumes that $\delta_1 = \delta_0 = 0$, but in fact $\delta_1 \neq \delta_0$, then it is easy to see that the optimum detector for the least-favorable distribution found in this section achieves *zero* error probability. Since in practice small uncertainties are inevitable, the usefulness of our results for jamming problems would be seriously limited, unless we can show that the least-favorable distribution can be robustified so as to avoid the extreme sensitivity exhibited above. Choose an arbitrary $0 < \sigma_0 < \sigma$ and let the noise be

$$\mathbf{N} = \mathbf{N}^* + \mathbf{N}_0$$

where $\mathbf{N}^*$ and $\mathbf{N}_0$ are independent, $\mathbf{N}^*$ is the worst-case noise distribution with power $\sigma^2 - \sigma_0^2$ found in this section when the input is $\{+1, -1\}$ and $\mathbf{N}_0$ is zero-mean Gaussian with variance $\sigma_0^2$. It is shown in [4] that the error probability achieved by such a noise is

$$2\,Q\left(\frac{\delta_1 - \delta_0}{2\sigma_0}\right) P_e(\sigma^2 - \sigma_0^2).$$

The conclusion is that, $\hat{P}_e(\sigma^2)$, the worst-case error probability when the noise distribution is chosen without knowing the values of $\delta_1$ and $\delta_0$ is bounded by

$$2\,Q\left(\frac{\delta_1 - \delta_0}{2\sigma_0}\right) P_e(\sigma^2 - \sigma_0^2) \le \hat{P}_e(\sigma^2) \le P_e((2\sigma/(2+\delta_1-\delta_0))^2)$$

This implies that if $\delta_1 - \delta_0 \to 0$, then $\hat{P}_e(\sigma^2) \to P_e(\sigma^2)$, because $\sigma_0^2$ can be chosen arbitrarily close to 0. Thus, small deviations from the assumed input alphabet lead to small deviations in the worst-case error probability. Furthermore, we have shown how to robustify the least-favorable noise distribution against uncertainties in the input values by convolving it with a low-variance Gaussian distribution.

An important generalization of the model considered in this section is the case of *multisample* equiprobable hypothesis:

$$H_1\colon\ \mathbf{Y} = \mathbf{X}_1 + \mathbf{N}$$

$$H_0\colon\ \mathbf{Y} = \mathbf{X}_0 + \mathbf{N}$$

where all the quantities are $n$-dimensional vectors; $\mathbf{X}_1$, $\mathbf{X}_0$ are deterministic and $\mathbf{N}$ is a random vector satisfying

$$\frac{1}{n} E[\|\mathbf{N}\|^2] \le \sigma^2.$$

If $\mathbf{N}$ is zero-mean Gaussian with covariance matrix $\sigma^2 I$, then it is well known that the probability of error is $Q(\sqrt{SNR})$ with

$$SNR = \frac{\| X_1 - X_0 \|^2}{4\sigma^2}. \quad (2)$$

**Theorem 2.**

The maximum probability of error of the maximum-likelihood test between

$$H_1: \; Y = X_1 + N$$

$$H_0: \; Y = X_0 + N$$

over all $n$-dimensional random vectors $N$ satisfying $E[\|N\|^2] \le n\,\sigma^2$ is equal to $P_e(n/SNR)$ where $P_e$ and $SNR$ are given in (1) and (2) respectively.

It is interesting to note that the least-favorable noise distribution that achieves the worst-case error probability in Theorem 2 has the same memory structure as the worst-case Gaussian noise vector. To see this, it is straightforward to show (cf. [5], [6, Prop. 7]) that among all covariance matrices $\Sigma$ such that $tr(\Sigma) \le n\,\sigma^2$, $\Sigma^* = n\,\sigma^2\, AA^T / \|A\|^2$ is the one that minimizes the matched-filter signal-to-noise ratio

$$\min_{\substack{\Sigma \\ tr\,\Sigma \le n\sigma^2}} \max_{h} \frac{(h^T A)^2}{h^T \Sigma h}.$$

The probability of error achieved by the least-favorable Gaussian distribution is $Q(\sqrt{SNR/n})$, therefore the comparison between the worst-case and Gaussian curves is exactly as in the single-sample case (Figure 1).

## 3. CAPACITY

This section is devoted to the solution of the worst-case capacity of the binary-input memoryless channel

$$Y_i = X_i + N_i$$

where $X_i$ takes values on $\{-1,1\}$. Thus, we seek

$$C(\sigma^2) = \min_{\substack{N \\ E[N^2] \le \sigma^2}} \max_{X} I(X; X+N)) \quad (3)$$

$$= \max_{X} \min_{\substack{N \\ E[N^2] \le \sigma^2}} I(X; X+N)$$

where the maximum ranges over all distributions on $\{-1,1\}$ and the second equality follows from the concavity-convexity of mutual information in the respective arguments. Results on compound channels [7] indicate that $C(\sigma^2)$ remains the capacity of the channel even when the decoder is not informed of the noise statistics.

It is verified in [4] that the noise distribution that solves (3) puts its mass in the lattice $\{\cdots -4, -2, 0, +2, +4 \cdots\}$ with a probability mass function that satisfies

$$\log(1 + \frac{p_{-1}}{p_0}) + \log(1 + \frac{p_1}{p_0}) - \log(1 + \frac{p_{k+1}}{p_k}) - \log(1 + \frac{p_{k-1}}{p_k}) + \lambda k^2 = 0$$

where $p_k$ is the mass at $2k$. The least-favorable noise distribution is shown in Figure 4 for $SNR = 0\,dB$ and $SNR = -10\,dB$. For low SNRs the least-favorable cdf approaches a Gaussian shape (in agreement with [8]), whereas in the region $SNR > 0\,dB$ the least-favorable distribution is indistinguishable from a three-mass distribution with weights $(\sigma^2/8,\ 1 - \sigma^2/4,\ \sigma^2/8)$ at $(-2, 0, 2)$. This three-mass noise distribution achieves capacity equal to

$$\log 2 - (1 - \frac{\sigma^2}{8}) \log(\frac{8}{\sigma^2} - 1) + (1 - \frac{\sigma^2}{4}) \log(\frac{8}{\sigma^2} - 2) \quad (4)$$

in the interval $0 < \sigma^2 < 4$. The minimum of (4) and the capacity of the Gaussian noise channel is indistinguishable from the worst-case capacity.

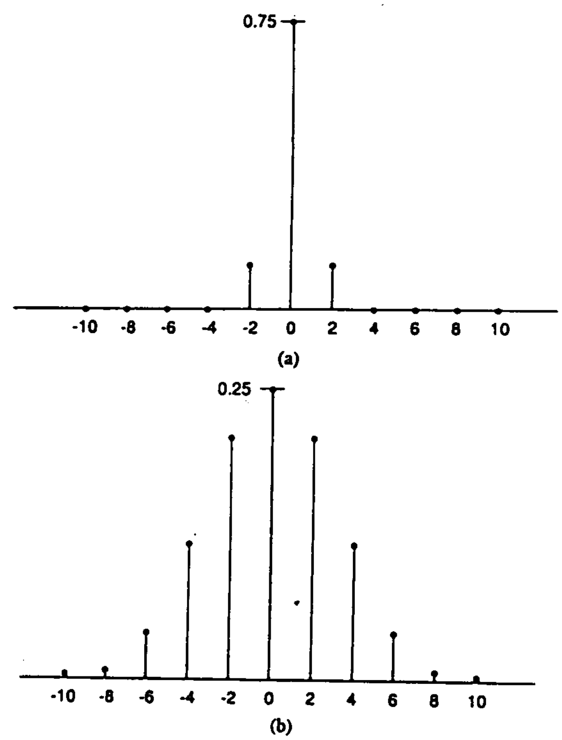


*Fig.4 Least-favorable (capacity) noise distributions for $SNR = 0\,db$ (a) and $-10\,dB$ (b).*

In the high SNR region the worst-case distributions for error probability and capacity differ in that in the former case, the masses are placed at $(-1, 0, 1)$ whereas in the latter, they are essentially distributed on $(-2, 0, 2)$. This implies that in the error probability problem the channel becomes an erasure channel, whereas in the capacity problem we obtain a binary symmetric channel to which "noiseless" outputs at $-3$ and $3$ are appended. In the single-sample hypothesis testing problem, zero outputs (erasures) carry no useful information, in contrast to the setting of encoded communication, in which mistaking $-1$ for $+1$ and viceversa is much more harmful than getting a zero output.

It is of interest to quantify the discrepancy between the worst-case capacity and the well-known capacity curve of the binary input Gaussian noise channel (e.g. [9, prob. 4.22]). The maximum difference between Gaussian capacity and worst-case capacity (Figure 2) is 0.118 bit and occurs at 7.2 dB, whereas the maximum relative decrease is 12.5% occuring at 6.7 dB. The relative power loss of worst-case capacity with respect to Gaussian capacity can be seen from Figure 2 to grow unbounded with the signal-to-noise ratio. However, in most applications this is not an important comparison, because in that range, a minimal increase in capacity requires a large increase in signal-to-noise ratio.

It is well-known (e.g [10, p. 103]) that for asymptotically low signal-to-noise ratio, the capacity of the input-power constrained Gaussian channel is reduced by a factor of $\pi/2$ (which translates into a power loss of about 2 dB) when the output is quantized to two levels (regardless of whether the power-limited input is further constrained to take two values only). In this context, both input and output quantization are assumed to be optimal, which corresponds to a zero threshold at the channel output and antipodal equiprobable inputs if the noise is Gaussian. It is interesting to see how this degradation factor from Gaussian capacity changes not only when the output is quantized but when the noise is not Gaussian. Of course, the answer depends on the noise distribution, but our previous results allow us to show that (at low signal-to-noise ratio) the degradation factor is at most 3 (or a power loss of 4.8 dB). This bound is achieved, among other noise distributions, by the uniform density. It should be emphasized that capacity degrades asymptotically by at most a factor of 3 when the binary output quantization is *optimal*. If the output quantization is forced to be a zero-threshold, then the worst-case (binary-input) capacity [3] is equal to $0$ if $\sigma > 1$ and $\log 2 - h(\sigma^2/2)$ if $\sigma \leq 1$. The latter expression should be compared to the worst-case capacity with optimal output quantization: $\log 2 - h(\sigma^2/4)$ if $\sigma \leq 1$. Thus, optimal output quantization buys 3 dB over straight zero-thresholding if the signal-to-noise ratio exceeds 0 dB. Below 0 dB, reliable communication is possible only with optimal quantization. It can be shown (from (4)) that for high signal-to-noise ratio, the worst-case binary-input capacity with unquantized outputs (Figure 2) behaves asymptotically as $\log 2 - h(\sigma^2/8)$, which means that optimal binary output quantization costs 3 dB at high signal-to-noise ratios, (and, thus, zero-thresholding costs 6 dB).

Analogously to the error probability problem, it is shown in [4] that the worst-case capacity does not break down when the noise distribution is chosen without exact knowledge of the input alphabet. As in the error probability problem, a small amount of Gaussian noise robustifies the choice of the least-favorable distribution.

## References


1. N. Blachman, "Communication as a Game," *IRE Wescon Rec.*, vol. 2, pp. 61-66, 1957.
2. R. Dobrushin, "Optimum Information Transmission through a Channel with unknown parameters," *Radiotech. Electron.*, vol. 4, pp. 1951-1956, 1959.
3. J. Borden, D. Mason, and R. McEliece, "Some Information Theoretic Saddlepoints," *SIAM J. Control and Optimization*, vol. 23, pp. 129-143, Jan. 1985.
4. S. Shamai (Shitz) and S. Verdu, "Worst-Case Power Constrained Noise for Binary-Input Channels," *IEEE Transactions on Information Theory*, Sep. 1992.
5. L. Zetterberg, "Signal Detection under Noise Interference in a Game Situation," *IRE Trans. Information Theory*, vol. IT-8, pp. 47-52, Sep. 1962.
6. S. Verdu and H. V. Poor, "Minimax Robust Discrete-Time Matched Filters," *IEEE Trans. Communications*, vol. COM-31, pp. 208-215, Feb. 1983.
7. J. Wolfowitz, *Coding Theorems of Information Theory*, Springer, New York, 1978. Third Edition
8. C. Cahn, "Worst Interference for Coherent Binary Channel," *IEEE Trans. Information Theory*, vol. 17, pp. 209-210, Mar. 1971.
9. R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
10. R.J. McEliece, *The Theory of Information and Coding: A mathematical framework for communication*, Addison-Wesley, Reading,Mass, 1977.



This work was supported in part by the Office of Naval Research under Grant N00014-90-J-1734.