MULTILEVEL MINIMAX HYPOTHESIS TESTING

Kush R. Varshney[1] *and Lav R. Varshney*[2]

IBM Thomas J. Watson Research Center
[1]1101 Kitchawan Rd., Yorktown Heights, NY 10598, USA
[2]19 Skyline Dr., Hawthorne, NY 10532, USA

ABSTRACT

In signal detection, Bayesian hypothesis testing and minimax hypothesis testing represent two extremes in the knowledge of the prior probabilities of the hypotheses: full information and no information. We propose an intermediate formulation, also based on the likelihood ratio test, to allow for partial information. We partition the space of prior probabilities into a set of levels using a quantization-theoretic approach with a minimax Bayes risk error criterion. Within each prior probability level, an optimal representative probability value is found, which is used to set the threshold of the likelihood ratio test. The formulation is demonstrated on signals with additive Gaussian noise.

Index Terms— quantization, categorization, hypothesis testing, signal detection, Bayes risk error

1. INTRODUCTION

Signal detection or hypothesis testing is one of the basic endeavors in statistical signal processing. The likelihood ratio test is the optimal decision rule for most variations of this task [1]. Different variations, including Bayesian, Neyman–Pearson and minimax, induce different thresholds to which the likelihood ratio is compared. Thus, the setting of the threshold is the key component of the hypothesis test. In this paper, we devise a new formulation for setting the threshold that interpolates between Bayesian and minimax hypothesis testing using an application of quantization theory [2]. We term this formulation *multilevel minimax hypothesis testing*.

In the binary Bayesian formulation, the decision rule is chosen to minimize the Bayes risk. The ratio of the prior probabilities of the two hypotheses is one factor in the threshold of the optimal likelihood ratio test. Thus the Bayesian threshold requires prior knowledge of the prior probabilities. In many real-world scenarios, it is difficult to obtain the exact prior probabilities before a signal detection system is deployed. This difficulty motivates minimax hypothesis testing. In minimax hypothesis testing, the threshold is chosen to minimize the Bayes risk under the worst-case prior probabilities of the hypotheses. However, the worst-case formulation ignores any possible partial information about the prior probabilities that may be available in advance.

We formulate multilevel minimax hypothesis testing to incorporate partial knowledge of the prior probability in the following form. An interval or cell of the probability simplex where the true prior probability lies is known and the threshold is set to minimize the worst-case error between the Bayes risk over the cell and the Bayes risk had the true prior probability been known exactly. This formulation appeals to quantization theory and the Bayes risk error distortion measure that we introduced previously in [3, 4].

Our previous work reduced to quantization that minimizes *expected* distortion [3, 4]. Here the problem is quantization to minimize *maximum* distortion; the formulation here is minimax Bayes risk error quantization of prior probabilities rather than minimum mean Bayes risk error quantization. To the best of our knowledge, there has been no other previous work on the quantization of prior probabilities for hypothesis testing. Studies and results in quantization theory typically focus on expected distortion but maximum distortion does also appear occasionally, e.g. [5, 6, 7], and has connections to ϵ-covering and ϵ-entropy [8]. Our work, providing a means to consider intervals of prior belief rather than exact prior belief, is similar in spirit but differs in details to decision making based on interval-valued probability described in [9]. There are also connections to representative prior distributions [10] and the robust Bayesian viewpoint [11, 12].

The remainder of the paper is organized in the following manner. In Sec. 2, we describe the signal detection setup under consideration and the minimax quantization optimization problem that follows. In Sec. 3, we derive the Lloyd–Max optimality conditions for minimax Bayes risk error quantization. We provide examples of optimal quantizers and corresponding maximum distortions in Sec. 4. Finally, Sec. 5 concludes with discussion.

2. MATHEMATICAL FORMULATION

Consider the binary hypothesis testing problem. There are two hypotheses h_0 and h_1 with prior probabilities $p_0 = \Pr[H = h_0]$ and $p_1 = \Pr[H = h_1] = 1 - p_0$, a noisy observation Y, and likelihoods $f_{Y|H}(y|H = h_0)$ and $f_{Y|H}(y|H = h_1)$. A decision rule $\hat{h}(y)$ that uniquely maps every possible y to either h_0 or h_1 is to be determined. There are two types of error probabilities: $p_E^{\mathrm{I}} = \Pr[\hat{h}(Y) = h_1|H = h_0]$ and $p_E^{\mathrm{II}} = \Pr[\hat{h}(Y) = h_0|H = h_1]$.

The optimal decision rule $\hat{h}(y)$ minimizes the Bayes risk function $J(p_0)$:

$$J(p_0) = c_{10}p_0p_E^{\mathrm{I}}(p_0) + c_{01}(1 - p_0)p_E^{\mathrm{II}}(p_0), \tag{1}$$

where c_{10} is the cost of the first type of error and c_{01} is the cost of the second type of error. The error probabilities are functions of the prior probability p_0 via the threshold on the right side of the Bayes risk optimal decision rule:

$$\frac{f_{Y|H}(y|H = h_1)}{f_{Y|H}(y|H = h_0)} \underset{\hat{h}(y)=h_0}{\overset{\hat{h}(y)=h_1}{\gtrless}} \frac{p_0c_{10}}{(1 - p_0)c_{01}}. \tag{2}$$

The function $J(p_0)$ is zero at the points $p_0 = 0$ and $p_0 = 1$ and is positive-valued, strictly concave, and continuous in the interval $(0, 1)$ [13].

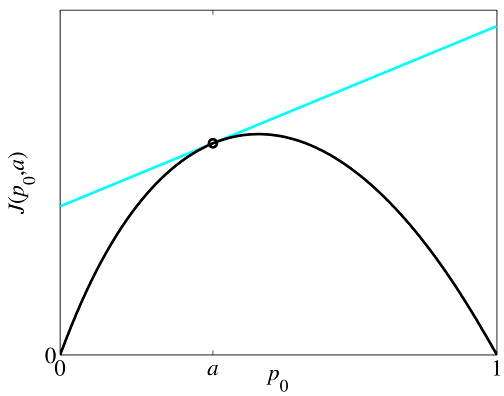

Fig. 1. Example $J(p_0)$ (black curve) and $J(p_0, a)$ (cyan line).

The Bayesian hypothesis testing threshold on the right side of (2) relies on the true prior probability p_0. As discussed in Sec. 1, this value may not be known in advance. When the true prior probability is p_0, but the threshold in $\hat{h}(y)$ uses some other value a, there is mismatch. The Bayes risk of the decision rule with threshold $\frac{ac_{10}}{(1-a)c_{01}}$ is:

$$J(p_0, a) = c_{10} p_0 p_E^{\mathrm{I}}(a) + c_{01}(1 - p_0) p_E^{\mathrm{II}}(a). \quad (3)$$

$J(p_0, a)$ is a linear function of p_0 with slope $(c_{10} p_E^{\mathrm{I}}(a) - c_{01} p_E^{\mathrm{II}}(a))$ and intercept $c_{01} p_E^{\mathrm{II}}(a)$. $J(p_0, a)$ is tangent to $J(p_0)$ at a and $J(p_0, p_0) = J(p_0)$. Example $J(p_0)$ and $J(p_0, a)$ are shown in Fig. 1.

The minimax hypothesis testing threshold is determined by finding the a that minimizes the worst-case $J(p_0, a)$, that is:

$$a^*_{\text{minimax}} = \arg\min_{a} \max_{p_0} J(p_0, a). \quad (4)$$

In such notation, $a^*_{\text{Bayesian}} = p_0$. Thus in Bayesian hypothesis testing, a continually changes with p_0, whereas in minimax hypothesis testing, there is a single value of a for all p_0. In the multilevel minimax hypothesis testing that we propose, there are K different possible values of a that depend discontinuously on p_0 through the quantizer function $q_K(p_0)$.

The K-level regular quantizer function $q_K(p_0)$ is defined as follows. The probability simplex $[0, 1]$ is partitioned into K intervals $\mathcal{Q}_1 = [0, b_1]$, $\mathcal{Q}_2 = (b_1, b_2]$, $\mathcal{Q}_3 = (b_2, b_3]$, ..., $\mathcal{Q}_K = (b_{K-1}, 1]$. There are also K representation points $a_k \in \mathcal{Q}_k$, $k = 1, \ldots, K$. The quantizer function is $q_K(p_0) = a_k$ for $p_0 \in \mathcal{Q}_k$. The quantizer design problem is to determine $\{b_1, \ldots, b_{K-1}\}$ and $\{a_1, \ldots, a_K\}$ to optimize some objective.

The multilevel minimax hypothesis testing formulation that we propose is captured by the following quantizer design problem:

$$q^*_K = \arg\min_{q_K} \max_{p_0} d(p_0, q_K(p_0)), \quad (5)$$

where as in [3, 4], the *Bayes risk error* $d(p_0, a)$ is:

$$d(p_0, a) = J(p_0, a) - J(p_0). \quad (6)$$

Operationally, knowing in advance that the true prior probability p_0 falls in level $\mathcal{Q}_k$, multilevel minimax hypothesis testing indicates that the representation point a_k be used in setting the threshold.

In the $K = 1$ case, it is straightforward to show that the representation point of q^*_1, a^*_1, equals the minimax hypothesis testing value a^*_{minimax}, and occurs at the peak of $J(p_0)$. However for $K > 1$, the representation point a_k within an interval $\mathcal{Q}_k$ is not the point that minimizes the maximum Bayes risk $J(p_0, a)$; rather it is the point that minimizes the maximum Bayes risk error $d(p_0, a)$.

In this section, we have defined an intermediary formulation between Bayesian and minimax hypothesis testing that we call multilevel minimax hypothesis testing. At its extremes, this formulation reduces to the two well-established hypothesis testing methodologies: minimax at $K = 1$ and Bayesian at $K = \infty$. Toward this end, we have formulated a minimax quantization problem. The next section discusses how to find the optimal quantizer q^*_K.

3. OPTIMALITY CONDITIONS

This section develops necessary conditions for local optimality of a quantizer for the probability simplex under the multilevel minimax criterion defined above. In particular, first we find a centroid condition to locally optimize a representation point a_k when its corresponding interval boundaries b_{k-1} and b_k are fixed. Then we find a nearest neighbor condition to locally optimize a boundary point b_k when its adjacent representation points a_k and a_{k+1} are fixed. Optimal quantizers can be found by alternately applying the nearest neighbor and centroid conditions through a version of the iterative Lloyd–Max algorithm [2, 5].

3.1. Centroid Condition

Within a fixed quantization interval $\mathcal{Q}_k$ with boundaries b_{k-1} and b_k, we would like to derive an expression for the optimal representation point a_k. The optimization problem is:

$$a_k = \arg\min_{a \in \mathcal{Q}_k} \max_{p_0 \in \mathcal{Q}_k} d(p_0, a). \quad (7)$$

Let us first focus on the inner maximization. As an initial step, we write the equation for the Bayes risk error in terms of the derivative of the Bayes risk function, which we denote as $J'(p_0)$.

$$d(p_0, a) = (p_0 - a)J'(a) + J(a) - J(p_0). \quad (8)$$

From this form, we see that the second derivative of $d(p_0, a)$ with respect to p_0 is $-J''(p_0)$, which is greater than zero due to the strict concavity of $J(p_0)$. Thus, $d(p_0, a)$ has no local maxima in the interior of $\mathcal{Q}_k$; the maximum occurs at an endpoint: b_k or b_{k-1}. Consequently,

$$\max_{p_0 \in \mathcal{Q}_k} d(p_0, a) = \max\{d(b_k, a), d(b_{k-1}, a)\} = \frac{d(b_{k-1}, a) + d(b_k, a) + |d(b_{k-1}, a) - d(b_k, a)|}{2}. \quad (9)$$

Substituting (8) into (9) and simplifying, we find that (9) equals

$$\frac{(b_{k-1} + b_k - 2a)J'(a) - J(b_{k-1}) - J(b_k) + 2J(a)}{2} + \frac{|(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k)|}{2}, \quad (10)$$

which is to be minimized with respect to $a \in \mathcal{Q}_k$.

Due to the absolute value function, there are two cases to consider:

1. $(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k) \leq 0$ and
2. $(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k) > 0$.

Due to the concavity of the Bayes risk function, $J'(a)$ is monotonically decreasing. Therefore, since $(b_{k-1} - b_k)$ is negative, $(b_{k-1} - b_k)J'(a) - J(b_{k-1}) + J(b_k)$ is a monotonically increasing function of a. Consequently the two cases of the absolute value correspond to the intervals $(b_{k-1}, a^\dagger]$ for case 1 and $(a^\dagger, b_k]$ for case 2, where $a^\dagger$ satisfies:

$$(b_{k-1} - b_k)J'(a^\dagger) - J(b_{k-1}) + J(b_k) = 0. \tag{11}$$

In the first case, (10) simplifies to:

$$(b_k - a)J'(a) + J(a) - J(b_k)$$

with derivative with respect to a:

$$(b_k - a)J''(a),$$

which is less than zero because $(b_k - a) > 0$ and $J''(a) < 0$ due to Bayes risk concavity. Thus the minimization objective is monotonically decreasing in the first case.

In the second case, (10) simplifies to:

$$(b_{k-1} - a)J'(a) + J(a) - J(b_{k-1}),$$

which has derivative with respect to a:

$$(b_{k-1} - a)J''(a),$$

which is greater than zero because $(b_{k-1} - a) < 0$ and $J''(a) < 0$. In the second case, the minimization objective is monotonically increasing.

Since (10) is decreasing over $(b_{k-1}, a^\dagger]$ and increasing over $(a^\dagger, b_k]$, it is minimized at $a^\dagger$. Therefore $a_k = a^\dagger$. The representation point satisfies (11). This is equivalently a slope matching condition:

$$J'(a_k) = \frac{J(b_k) - J(b_{k-1})}{b_k - b_{k-1}}. \tag{12}$$

At the optimal representation point a_k, the slope of the Bayes risk function equals the slope of the line connecting the Bayes risk function evaluated at the endpoints of the interval $\mathcal{Q}_k$.

3.2. Nearest Neighbor Condition

In the nearest neighbor condition, we would like to find the interval boundary b_k given the representation points a_k and a_{k+1}. As discussed in Sec. 3.1, the maximum Bayes risk error within an interval occurs at the interval boundary. Therefore, we would like to minimize the Bayes risk error at the interval boundary.

Specifically, b_k should be chosen to minimize the maximum of $d(b_k, a_k)$ and $d(b_k, a_{k+1})$. At a given potential boundary point b, the $J(b)$ term is the same in both $d(b, a_k)$ and $d(b, a_{k+1})$, so only $J(b, a_k)$ and $J(b, a_{k+1})$ need be considered. Due to the geometry of the problem, b_k should be the abscissa of the point at which the lines $J(p_0, a_k)$ and $J(p_0, a_{k+1})$ intersect. Working with the definitions of $J(p_0, a_k)$ and $J(p_0, a_{k+1})$, we find the point of intersection to be:

$$b_k = \frac{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1}) - p_E^{\mathrm{II}}(a_k)\right)}{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1}) - p_E^{\mathrm{II}}(a_k)\right) - c_{10}\left(p_E^{\mathrm{I}}(a_{k+1}) - p_E^{\mathrm{I}}(a_k)\right)}. \tag{13}$$

The nearest neighbor condition for minimax Bayes risk error quantization is the same as that for minimum mean Bayes risk error quantization [3, 4].

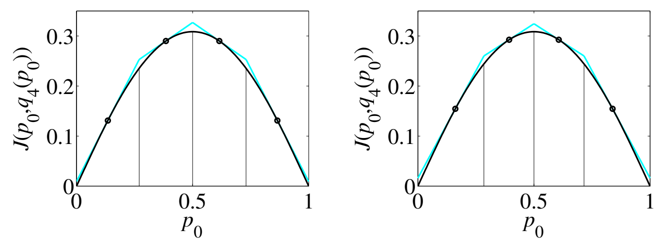

Fig. 2. (a) Minimum mean and (b) minimax Bayes risk error quantizers for $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 1$, $c_{01} = 1$.

4. EXAMPLES

As an example, we consider the following signal and measurement model:

$$Y = s_m + W, \quad m \in \{0, 1\}, \tag{14}$$

where $s_0 = 0$ and $s_1 = \mu$, and W is a zero-mean, Gaussian random variable with variance σ^2. The parameters μ and σ^2 are known, deterministic quantities. The error probabilities for this signal model are:

$$p_E^{\mathrm{I}}(p_0) = Q\left(\frac{\mu}{2\sigma} + \frac{\sigma}{\mu}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right), \text{ and}$$
$$p_E^{\mathrm{II}}(p_0) = Q\left(\frac{\mu}{2\sigma} - \frac{\sigma}{\mu}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right),$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{-x^2/2}dx$.

We use the Lloyd–Max algorithm to design quantizers for the proposed criterion using the centroid and nearest neighbor conditions derived in Sec. 3. We show such quantizers for $K = 4$ and different ratios of the Bayes costs c_{10} and c_{01} along with different ratios of μ and σ^2. We also compare these quantizers with those designed to minimize mean Bayes risk error.

Fig. 2 shows quantizers for equal Bayes costs and equal mean and standard deviation. In the plots, the black curve is $J(p_0)$ and the cyan line is $J(p_0, q_4(p_0))$, with their difference being $d(p_0, q_4(p_0))$. The circle markers are the representation points and the vertical lines indicate the interval boundaries of the different prior probability levels. The minimax levels and representation points are more clustered in the middle of the probability simplex and around the peak of $J(p_0)$ than the minimum mean levels and representation points. This is more apparent in the quantizers for the noisier measurement model with $\mu = 1$ and $\sigma^2 = 2$ shown in Fig. 3, and the quantizers for unequal Bayes costs $c_{10} = 10$ and $c_{01} = 1$ shown in Fig. 4.

The scaling behavior of the minimax Bayes risk error as a function of K is shown in Fig. 5 on both linear and logarithmic scales. The convergence from the edge case of minimax hypothesis testing to the other edge case of Bayesian hypothesis testing is linear on the plot with logarithmic axes (with approximately the same slope for different Bayes cost and noise settings). This indicates that the minimax distortion converges proportional to $K^{-\beta}$, where β is a positive constant. This minimax error scaling can also be viewed as the asymptotic behavior of the minimum covering radius with respect to Bayes risk error distortion.

5. CONCLUSION

The minimax hypothesis testing detection rule is employed for robustness to uncertainty in the prior probabilities, but can give poor

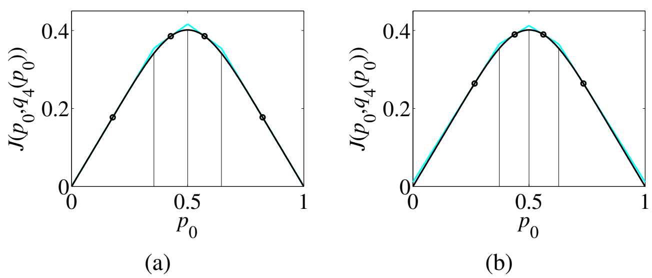

Fig. 3. (a) Minimum mean and (b) minimax Bayes risk error quantizers for $\mu = 1$, $\sigma^2 = 2$, $c_{10} = 1$, $c_{01} = 1$.

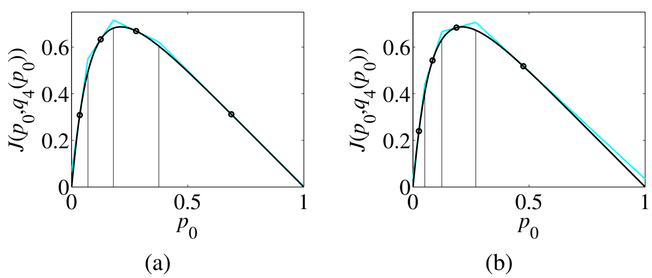

Fig. 4. (a) Minimum mean and (b) minimax Bayes risk error quantizers for $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 10$, $c_{01} = 1$.

detection performance in terms of Bayes risk. The Bayesian hypothesis testing detection rule gives the best possible detection performance in terms of Bayes risk, but makes use of completely certain prior probabilities. It is not difficult to imagine scenarios in which there is uncertainty in the priors, but this uncertainty is not complete. It is for these scenarios that we have developed multilevel minimax hypothesis testing. As can be seen in Fig. 5, even $K = 2$ or $K = 3$ levels provide a significant gain in performance.

Our formulation of multilevel minimax hypothesis testing is an interesting variation of quantization to minimize Bayes risk error, a problem we studied in the expected distortion case in [3, 4]. For expected distortion, a probability distribution over the prior probabilities is needed, i.e. a distribution over the population of events whose detection we are considering. This probability distribution over prior probabilities may be as difficult to obtain as the prior probability of a single event of interest in real-world settings; examination of the minimax criterion, which does not require this distribution of priors, is therefore of importance.

In [3, 4], we examined the implications of quantized prior hypothesis testing through the lens of human decision making, specifically looking at members of different racial populations whose actions were to be judged, e.g. was a foul or crime committed. In that work, we assume that the different racial groups have the same prior probability distributions of committing fouls or crimes so that the model is fair across racial groups. With the minimax criterion, no assumption on racial population prior probability distributions is required, and thus has fairness as an intrinsic quality.

The minimax criterion is one way to provide robustness, but segment method hypothesis testing, proposed as an alternative to minimax hypothesis testing in [14], is another way. In future work, multilevel segment method hypothesis testing could be developed.

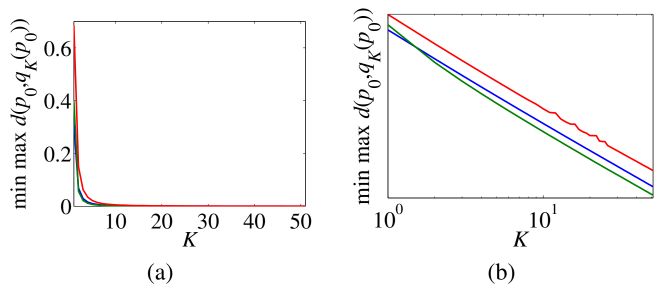

Fig. 5. Minimax Bayes risk error for $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 1$, $c_{01} = 1$ (blue line), $\mu = 1$, $\sigma^2 = 2$, $c_{10} = 1$, $c_{01} = 1$ (green line), and $\mu = 1$, $\sigma^2 = 1$, $c_{10} = 10$, $c_{01} = 1$ (red line), on (a) linear and (b) logarithmic scales.

6. REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968.

[2] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic Publishers, 1992.

[3] K. R. Varshney and L. R. Varshney, "Minimum mean Bayes risk error quantization of prior probabilities," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Las Vegas, Nevada, Apr. 2008, pp. 3445–3448.

[4] ——, "Quantization of prior probabilities for hypothesis testing," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4553–4562, Oct. 2008.

[5] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Berlin: Springer-Verlag, 2000.

[6] N. Sarshar and X. Wu, "Minimax multiresolution scalar quantization," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 2004, pp. 52–61.

[7] Y. A. Reznik, "An algorithm for quantization of discrete probability distributions," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 2011.

[8] A. N. Kolmogorov and V. M. Tihomirov, "ϵ-entropy and ϵ-capacity of sets in functional spaces," *Am. Math. Soc. Translations Series 2*, vol. 17, pp. 277–364, 1961.

[9] M. Wolfenson and T. L. Fine, "Bayes-like decision making with upper and lower probabilities," *J. Am. Stat. Assoc.*, vol. 77, no. 377, pp. 80–88, Mar. 1982.

[10] C. Hildreth, "Bayesian statisticians and remote clients," *Econometrica*, vol. 31, no. 3, pp. 422–438, Jul. 1963.

[11] J. O. Berger, "The robust Bayesian viewpoint," Purdue Univ., Tech. Rep. 82-9, Apr. 1982.

[12] L. R. Pericchi and P. Walley, "Robust Bayesian credible intervals and prior ignorance," *Int. Stat. Rev.*, vol. 59, no. 1, pp. 1–23, Apr. 1991.

[13] R. A. Wijsman, "Continuity of the Bayes risk," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 1083–1085, Jun. 1970.

[14] B. H. Krogh and H. V. Poor, "The segment method as an alternative to minimax in hypothesis testing," *Inform. Sciences*, vol. 27, no. 1, pp. 9–37, Jun. 1982.