

Technical Note: The Generalized Likelihood Ratio Test

Paul F. Roysdon[†]

Jay A. Farrell[‡]

I. DEFINITION OF THE GLRT

Consider a test for a signal present in Gaussian additive noise with non-zero mean [1]. A binary test can be performed for a random sample from a population that is normally distributed and has known variance. Based on the Neyman-Pearson (N-P) Lemma for binary hypothesis testing [2], [3], consider

$$\begin{aligned}\mathcal{H}_0 : \mathbf{y} &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \\ \mathcal{H}_1 : \mathbf{y} &\sim \mathcal{N}(\mathbf{H}\boldsymbol{\theta}, \sigma^2 \mathbf{I})\end{aligned}\quad (1)$$

for measurement $\mathbf{y} \in \mathbb{R}^{m \times 1}$, where $\sigma^2 > 0$ is known, $\mathbf{H} \in \mathbb{R}^{m \times n}$ is known, and the unknown $\boldsymbol{\theta} \in \mathbb{R}^{n \times 1}$. The standard, or *null-hypothesis*, with known mean is defined as \mathcal{H}_0 , and the *alternate-hypothesis* with unknown mean is defined as \mathcal{H}_1 .

The Likelihood Ratio Test (LRT) [4] compares the model in \mathcal{H}_1 to the model in \mathcal{H}_0 , for threshold γ , such that

$$\frac{p(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta})}{p(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \quad (3)$$

where

$$p(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\top(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})} \quad (4)$$

$$p(\mathbf{y}|\mathcal{H}_0) = \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}^\top\mathbf{y})}. \quad (5)$$

When \mathcal{H}_1 is decided:

- if \mathcal{H}_1 is valid, this is a *correct detection*,
- if \mathcal{H}_1 not valid, this is a *false alarm*.

When \mathcal{H}_0 is decided:

- if \mathcal{H}_1 is valid, this is a *missed detection*,
- if \mathcal{H}_1 not valid, this is a *correct rejection*.

The log likelihood ratio test is

$$\ln(\Lambda(\mathbf{y})) = \ln\left(\frac{p(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}_1)}{p(\mathbf{y}|\mathcal{H}_0)}\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma' \quad (6)$$

where $\gamma' = \ln(\gamma)$.

Defining eqn. (6) in terms of eqn. (1) & (2) yields

$$\ln(\Lambda(\mathbf{y})) = -\frac{1}{2\sigma^2} \left((\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^\top(\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) - \mathbf{y}^\top\mathbf{y} \right) \quad (7)$$

$$= -\frac{1}{2\sigma^2} (-\mathbf{y}^\top\mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{y} + \boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{H}\boldsymbol{\theta}) \quad (8)$$

$$= -\frac{1}{2\sigma^2} (-2\boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{y} + \boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{H}\boldsymbol{\theta}). \quad (9)$$

The simplification in eqn. (9) is possible because: $\mathbf{y}^\top\mathbf{H}\boldsymbol{\theta} = \mathbf{y} \bullet (\mathbf{H}\boldsymbol{\theta}) = (\mathbf{H}\boldsymbol{\theta})^\top\mathbf{y}$. Because $\boldsymbol{\theta}$ is unknown, eqn. (9) cannot be evaluated to implement a test.

The Generalized Likelihood Ratio Test (GLRT) [4] compares the *most likely* model in \mathcal{H}_1 to the *most likely* model in \mathcal{H}_0 , for threshold γ , such that

$$\frac{\max_{\boldsymbol{\theta}} p(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta})}{\max_{\boldsymbol{\theta}} p(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \quad (10)$$

The GLRT is determined by finding the Maximum Likelihood Estimate (MLE) of $\boldsymbol{\theta}$. The MLE estimates $\hat{\boldsymbol{\theta}}$ by finding the value of $\boldsymbol{\theta}$ that maximizes $\hat{\Lambda}(\boldsymbol{\theta}; \mathbf{y})$ [3], for $i = \{0, 1\}$:

$$\hat{\boldsymbol{\theta}}_i \triangleq \arg \max_{\boldsymbol{\theta}} p(\mathbf{y}|\mathcal{H}_i, \boldsymbol{\theta}). \quad (11)$$

For the alternate-hypothesis, the $\boldsymbol{\theta}$ that makes \mathbf{y} most likely is

$$\hat{\boldsymbol{\theta}}_1 = \arg \max_{\boldsymbol{\theta}} p(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) \quad (12)$$

$$= \arg \max_{\boldsymbol{\theta}} \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\top(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})} \quad (13)$$

$$= \arg \max_{\boldsymbol{\theta}} -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^\top(\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) \quad (14)$$

$$= \arg \min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^\top(\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) \quad (15)$$

$$= \arg \min_{\boldsymbol{\theta}} (\mathbf{y}^\top\mathbf{y} - 2\boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{y} + \boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{H}\boldsymbol{\theta}). \quad (16)$$

The exponential function of $\boldsymbol{\theta}$ is an increasing function. Eqn. (13) can be reduced to eqn. (14) because the log of the exponent does not change the maximization of the exponent over $\boldsymbol{\theta}$. Eqn. (14) can be reduced to eqn. (15) because $\frac{1}{2\sigma^2}$ is independent of $\boldsymbol{\theta}$, which will not change the maximum relative to $\boldsymbol{\theta}$. Accounting for the negative value in eqn. (14) changes the problem from a maximization over $\boldsymbol{\theta}$, to an equivalent minimization over $\boldsymbol{\theta}$, in eqn. (15). Finally eqn. (16) is simply algebra.

To find $\hat{\boldsymbol{\theta}}_1$, take the partial derivative of eqn. (16) and set it equal to zero:

$$\frac{\partial}{\partial \boldsymbol{\theta}} (\mathbf{y}^\top\mathbf{y} - \boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{y} + \boldsymbol{\theta}^\top\mathbf{H}^\top\mathbf{H}\boldsymbol{\theta}) = 0 \quad (17)$$

$$0 - 2\mathbf{H}^\top\mathbf{y} + 2\mathbf{H}^\top\mathbf{H}\boldsymbol{\theta} = 0 \quad (18)$$

$$\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{y}. \quad (19)$$

Substituting eqn. (19) into eqn. (9) yields the analytical form

[†]Ph.D. graduate, [‡]Professor at the Dept. of Electrical & Computer Engineering, UC Riverside. {proysdon, farrell}@ece.ucr.edu.

of the GLRT

$$\ln(\hat{\Lambda}(\mathbf{y})) = -\frac{1}{2\sigma^2} \left(-2\mathbf{y}^\top \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{y} + \mathbf{y}^\top \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{y} \right) \quad (20)$$

$$= -\frac{1}{2\sigma^2} \left(-2\mathbf{y}^\top \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{y} + \mathbf{y}^\top \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{y} \right) \quad (21)$$

$$= -\frac{1}{2\sigma^2} \left(-2\mathbf{y}^\top \mathbf{P} \mathbf{y} + \mathbf{y}^\top \mathbf{P} \mathbf{y} \right) \quad (22)$$

$$= \frac{1}{\sigma^2} \left(\mathbf{y}^\top \mathbf{P} \mathbf{y} - \frac{1}{2} \mathbf{y}^\top \mathbf{P} \mathbf{y} \right) \quad (23)$$

$$= \frac{1}{2\sigma^2} \mathbf{y}^\top \mathbf{P} \mathbf{y} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma', \quad (24)$$

where $\mathbf{P} \triangleq \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top$.

From the result in eqn. (24), we can now determine the relation of the GLRT to the Probability of False Alarm (P_{FA}) and the Chi-square distribution.

II. GLRT RELATION TO P_{FA} AND χ^2

The objective is to choose γ' for the desired P_{FA} by evaluating eqn. (24) for the binary hypothesis. First, consider $\mathbf{y}^\top \mathbf{P} \mathbf{y}$ under \mathcal{H}_0 . Define \mathbf{H} in terms of the “thin” QR factorization [5], e.g. $\mathbf{H} = \mathbf{Q}_1 \mathbf{R}_1$:

$$\mathbf{H} = \mathbf{Q} \mathbf{R} \quad (25)$$

$$= [\mathbf{Q}_1 \ \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \quad (26)$$

$$= \mathbf{Q}_1 \mathbf{R}_1 \quad (27)$$

where $\mathbf{Q} \in \mathbb{R}^{m \times m}$ is a basis for the column space of \mathbf{H} , and $\mathbf{R} \in \mathbb{R}^{m \times n}$ with $m > n$. Both \mathbf{Q}_1 and \mathbf{Q}_2 have orthogonal columns, where $\mathbf{Q}_1 \in \mathbb{R}^{m \times n}$, $\mathbf{Q}_2 \in \mathbb{R}^{m \times (m-n)}$. The parameter $\mathbf{R}_1 \in \mathbb{R}^{n \times n}$ is an invertible upper triangular matrix, and the zeros matrix, $\mathbf{0} \in \mathbb{R}^{(m-n) \times n}$. For full column-rank \mathbf{H} , i.e. $\text{rank}(\mathbf{H}) = n$, then both \mathbf{Q}_1 and \mathbf{R}_1 are unique.

Using the QR factorization of \mathbf{H} allows analysis of \mathbf{P} as

$$\mathbf{P} = \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \quad (28)$$

$$= \mathbf{Q}_1 \mathbf{R}_1 (\mathbf{R}_1^\top \mathbf{Q}_1^\top \mathbf{Q}_1 \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{Q}_1^\top \quad (29)$$

$$= \mathbf{Q}_1 \mathbf{R}_1 (\mathbf{R}_1^\top \mathbf{I} \mathbf{R}_1)^{-1} \mathbf{R}_1^\top \mathbf{Q}_1^\top \quad (30)$$

$$= \mathbf{Q}_1 \mathbf{R}_1 \mathbf{R}_1^{-1} (\mathbf{R}_1^\top)^{-1} \mathbf{R}_1^\top \mathbf{Q}_1^\top \quad (31)$$

$$= \mathbf{Q}_1 \mathbf{Q}_1^\top \quad (32)$$

where $\mathbf{Q}_1^\top \mathbf{Q}_1 = \mathbf{I}$. Substituting eqn. (32) into eqn. (24), the decision statistic is

$$\ln(\hat{\Lambda}(\mathbf{y})) = \frac{1}{2\sigma^2} \mathbf{y}^\top \mathbf{P} \mathbf{y} \quad (33)$$

$$= \frac{1}{2\sigma^2} \mathbf{y}^\top \mathbf{Q}_1 \mathbf{Q}_1^\top \mathbf{y} \quad (34)$$

$$= \frac{1}{2\sigma^2} \mathbf{z}^\top \mathbf{z} \quad (35)$$

where $\mathbf{z} = \mathbf{Q}_1^\top \mathbf{y} \in \mathbb{R}^{n \times 1}$ is a Gaussian random variable with m degrees of freedom, and $\ln(\hat{\Lambda}(\mathbf{y}))$ is a $\chi_{(m-n)}^2$ random variable with $m-n$ degrees-of-freedom: the degrees-of-freedom of a Chi-square random variable is the number of measurements m , minus the number of parameters n .

Under the alternate hypothesis, $\mathcal{H}_1 : \mathbf{y}_1 \sim \mathcal{N}(\mathbf{H}\boldsymbol{\theta}, \sigma^2 \mathbf{I})$, the expected value of \mathbf{z}_1 is

$$E\langle \mathbf{z}_1 \rangle = E\langle \mathbf{Q}_1^\top \mathbf{y} \rangle \quad (36)$$

$$= \mathbf{Q}_1^\top E\langle \mathbf{y} \rangle \quad (37)$$

$$= \mathbf{Q}_1^\top \mathbf{H} \boldsymbol{\theta}, \quad (38)$$

and covariance of \mathbf{z}_1 is

$$E\langle \mathbf{z}_1 \mathbf{z}_1^\top \rangle = E\langle \mathbf{Q}_1^\top \mathbf{y} \mathbf{y}^\top \mathbf{Q}_1 \rangle \quad (39)$$

$$= \mathbf{Q}_1^\top (\sigma^2 \mathbf{I}_n) \mathbf{Q}_1 \quad (40)$$

$$= \sigma^2 \mathbf{I}_n. \quad (41)$$

Therefore, under \mathcal{H}_1 , $\mathbf{z}_1 \sim \mathcal{N}(\mathbf{H}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \in \mathbb{R}^{n \times 1}$.

Under the null hypothesis, $\mathcal{H}_0 : \mathbf{y}_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, the expected value of \mathbf{z}_0 is

$$E\langle \mathbf{z}_0 \rangle = E\langle \mathbf{Q}_1^\top \mathbf{y} \rangle \quad (42)$$

$$= \mathbf{Q}_1^\top E\langle \mathbf{y} \rangle \quad (43)$$

$$= \mathbf{0}, \quad (44)$$

and covariance of \mathbf{z}_0 is

$$E\langle \mathbf{z}_0 \mathbf{z}_0^\top \rangle = E\langle \mathbf{Q}_1^\top \mathbf{y} \mathbf{y}^\top \mathbf{Q}_1 \rangle \quad (45)$$

$$= \mathbf{Q}_1^\top (\sigma^2 \mathbf{I}_n) \mathbf{Q}_1 \quad (46)$$

$$= \sigma^2 \mathbf{I}_n. \quad (47)$$

Therefore, under \mathcal{H}_0 , $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n) \in \mathbb{R}^{n \times 1}$.

From eqn. (24), the test statistic is

$$\frac{\mathbf{z}^\top \mathbf{z}}{2\sigma^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma'. \quad (48)$$

Given the P_{FA} constraint, the optimum decision threshold γ' is found by applying the inverse CDF of the Chi-square distribution with $m-n$ degrees-of-freedom. Thus, under \mathcal{H}_0 , we can define the P_{FA} in terms of the GLRT

$$P_{FA} = \text{p}(\chi_{(m-n)}^2 > \gamma). \quad (49)$$

In statistics literature, eqn. (48) is referred to as Wilks Theorem [6].

III. EXAMPLE

In practice, it is common to use the MatlabTM function `chi2inv(1/γ, m-n)` or `chi2cdf(y, m-n)`. For example, set the degrees-of-freedom $m-n = 10$, and the P_{FA} constraint $\gamma = 0.05$. Then $v = \text{chi2inv}(0.95, 10) = 18.3070$. Then decide \mathcal{H}_1 if $\mathbf{z}^\top \mathbf{z} > v\sigma^2$.

REFERENCES

- [1] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.
- [2] J. Neyman and E. S. Pearson, "The testing of statistical hypotheses in relation to probabilities a priori," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 29, pp. 492–510, 1933.
- [3] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I - Estimation Theory*. Prentice Hall PTR, 2013.
- [4] —, *Fundamentals of Statistical Signal Processing, Vol. II - Detection Theory*. Prentice Hall PTR, 1998.
- [5] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd ed.)*. Johns Hopkins, 1996.
- [6] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.