The Generalized Likelihood Ratio Test: Derivations & Proofs

Paul F. Roysdon, Ph.D.

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

$$\mathcal{H}_0 : \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}) \tag{1}$$

$$\mathcal{H}_1 : \mathbf{y} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\theta}, \boldsymbol{\sigma}^2\mathbf{I}) \tag{2}$$

for measurement $\mathbf{y} \in \mathbb{R}^{m\times 1}$, where $\boldsymbol{\sigma}^2 > 0$ 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 $\boldsymbol{\gamma}$, such that

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

where

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

$$\mathrm{p}(\mathbf{y}|\mathcal{H}_0) = \frac{1}{(2\pi\boldsymbol{\sigma}^2)^{k/2}} e^{\left(-\frac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}^\intercal\mathbf{y})\right)}. \tag{5}$$

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

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

where $\boldsymbol{\gamma}' = \ln(\boldsymbol{\gamma})$.

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

$$\ln(\Lambda(\mathbf{y})) = -\frac{1}{2\boldsymbol{\sigma}^2}\Big((\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta}) - \mathbf{y}^\intercal\mathbf{y}\Big) \tag{7}$$

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

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

The simplification in eqn. (9) is possible because: $\mathbf{y}^\intercal\mathbf{H}\boldsymbol{\theta} = \mathbf{y}\bullet(\mathbf{H}\boldsymbol{\theta}) = (\mathbf{H}\boldsymbol{\theta})^\intercal\mathbf{y}$. 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 $\boldsymbol{\gamma}$, such that

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

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 $i = \{0, 1\}$:

$$\hat{\boldsymbol{\theta}}_i \triangleq \underset{\boldsymbol{\theta}}{\mathit{arg\,max}}\ \mathrm{p}(\mathbf{y}|\mathcal{H}_i, \boldsymbol{\theta}). \tag{11}$$

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

$$\hat{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\theta}}{\mathit{arg\,max}}\ \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) \tag{12}$$

$$= \underset{\boldsymbol{\theta}}{\mathit{arg\,max}}\ \frac{1}{(2\pi\boldsymbol{\sigma}^2)^{k/2}} e^{-\frac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})} \tag{13}$$

$$= \underset{\boldsymbol{\theta}}{\mathit{arg\,max}}\ -\frac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta}) \tag{14}$$

$$= \underset{\boldsymbol{\theta}}{\mathit{arg\,min}}\ (\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta}) \tag{15}$$

$$= \underset{\boldsymbol{\theta}}{\mathit{arg\,min}}\ (\mathbf{y}^\intercal\mathbf{y} - 2\boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{y} + \boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{H}\boldsymbol{\theta}). \tag{16}$$

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 $\frac{1}{2\boldsymbol{\sigma}^2}$ 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:

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

$$0 - 2\mathbf{H}^\intercal\mathbf{y} + 2\mathbf{H}^\intercal\mathbf{H}\boldsymbol{\theta} = 0 \tag{18}$$

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

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

$$\ln\left(\hat{\Lambda}(\mathbf{y})\right) = -\frac{1}{2\boldsymbol{\sigma}^2}\bigg(-2\mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y} + \mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}\bigg) \tag{20}$$

$$= -\frac{1}{2\boldsymbol{\sigma}^2}\bigg(-2\mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y} + \mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}\bigg) \tag{21}$$

$$= -\frac{1}{2\boldsymbol{\sigma}^2}\bigg(-2\mathbf{y}^\intercal\mathbf{P}\mathbf{y} + \mathbf{y}^\intercal\mathbf{P}\mathbf{y}\bigg) \tag{22}$$

$$= \frac{1}{\boldsymbol{\sigma}^2}\bigg(\mathbf{y}^\intercal\mathbf{P}\mathbf{y} - \frac{1}{2}\mathbf{y}^\intercal\mathbf{P}\mathbf{y}\bigg) \tag{23}$$

$$= \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{y}^\intercal\mathbf{P}\mathbf{y} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma', \tag{24}$$

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

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 $\mathbf{y}^\intercal\mathbf{P}\mathbf{y}$ under $\mathcal{H}_0$. Define $\mathbf{H}$ in terms of the "thin" *QR* factorization [5], e.g. $\mathbf{H} = \mathbf{Q}_1\mathbf{R}_1$:

$$\mathbf{H} = \mathbf{Q}\mathbf{R} \tag{25}$$

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

$$= \mathbf{Q}_1\mathbf{R}_1 \tag{27}$$

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. 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

$$\mathbf{P} = \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal \tag{28}$$

$$= \mathbf{Q}_1\mathbf{R}_1(\mathbf{R}_1^\intercal\mathbf{Q}_1^\intercal\mathbf{Q}_1\mathbf{R}_1)^{-1}\mathbf{R}_1^\intercal\mathbf{Q}_1^\intercal \tag{29}$$

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

$$= \mathbf{Q}_1\mathbf{R}_1\mathbf{R}_1^{-1}(\mathbf{R}_1^\intercal)^{-1}\mathbf{R}_1^\intercal\mathbf{Q}_1^\intercal \tag{31}$$

$$= \mathbf{Q}_1\mathbf{Q}_1^\intercal \tag{32}$$

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

$$\ln\left(\hat{\Lambda}(\mathbf{y})\right) = \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{y}^\intercal\mathbf{P}\mathbf{y} \tag{33}$$

$$= \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{y}^\intercal\mathbf{Q}_1\mathbf{Q}_1^\intercal\mathbf{y} \tag{34}$$

$$= \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{z}^\intercal\mathbf{z} \tag{35}$$

where $\mathbf{z} = \mathbf{Q}_1^\intercal\mathbf{y} \in \mathbb{R}^{n\times 1}$ is a Gaussian random variable with m degrees of freedom, and $\ln\left(\hat{\Lambda}(\mathbf{y})\right)$ is a $\chi^2_{(m-n)}$ 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, $\mathcal{H}_1 : \mathbf{y}_1 \sim \mathcal{N}(\mathbf{H}\boldsymbol{\theta}, \boldsymbol{\sigma}^2\mathbf{I})$, the expected value of $\mathbf{z}_1$ is

$$E\langle\mathbf{z}_1\rangle = E\langle\mathbf{Q}_1^\intercal\mathbf{y}\rangle \tag{36}$$

$$= \mathbf{Q}_1^\intercal E\langle\mathbf{y}\rangle \tag{37}$$

$$= \mathbf{Q}_1^\intercal\mathbf{H}\boldsymbol{\theta}, \tag{38}$$

and covariance of $\mathbf{z}_1$ is

$$E\langle\mathbf{z}_1\mathbf{z}_1^\intercal\rangle = E\langle\mathbf{Q}_1^\intercal\mathbf{y}\mathbf{y}^\intercal\mathbf{Q}_1\rangle \tag{39}$$

$$= \mathbf{Q}_1^\intercal(\boldsymbol{\sigma}^2\mathbf{I}_n)\mathbf{Q}_1 \tag{40}$$

$$= \sigma^2\mathbf{I}_n. \tag{41}$$

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

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

$$E\langle\mathbf{z}_0\rangle = E\langle\mathbf{Q}_1^\intercal\mathbf{y}\rangle \tag{42}$$

$$= \mathbf{Q}_1^\intercal E\langle\mathbf{y}\rangle \tag{43}$$

$$= \mathbf{0}, \tag{44}$$

and covariance of $\mathbf{z}_0$ is

$$E\langle\mathbf{z}_0\mathbf{z}_0^\intercal\rangle = E\langle\mathbf{Q}_1^\intercal\mathbf{y}\mathbf{y}^\intercal\mathbf{Q}_1\rangle \tag{45}$$

$$= \mathbf{Q}_1^\intercal(\boldsymbol{\sigma}^2\mathbf{I}_n)\mathbf{Q}_1 \tag{46}$$

$$= \sigma^2\mathbf{I}_n. \tag{47}$$

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

From eqn. (24), the test statistic is

$$\frac{\mathbf{z}^\intercal\mathbf{z}}{2\boldsymbol{\sigma}^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma'. \tag{48}$$

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

$$P_{FA} = \mathrm{p}(\chi^2_{(m-n)} > \gamma). \tag{49}$$

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

III. Example

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

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.