Technical Note: The Neyman-Pearson Lemma

Paul F. Roysdon† Jay A. Farrell‡

I. Definition of the Neyman-Pearson Lemma

Consider two densities $\mathrm{p}(\mathbf{y}|\mathcal{H}_0)$ and $\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}_1)$, for the parameter $\mathbf{y}$ and random variable $\boldsymbol{\theta}$, where $\mathcal{H}_0$ is the *null-hypothesis*, and $\mathcal{H}_1$ is the *alternate-hypothesis*. A constrained optimization problem can be formulated and solved by Lagrange multipliers, to maximize the probability of detection, P_D, of event $\mathcal{H}_1$, given the probability of false alarm, $P_{FA} = \alpha$.

For $\mathbf{y} \in \mathcal{X}$, the subspace $\mathcal{X}_1$, where $\mathcal{H}_1$ is decided, is found by maximizing

$$P_D = \int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) d\mathbf{y} \tag{1}$$

under the constraint

$$P_{FA} = \int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_0) d\mathbf{y} = \alpha \tag{2}$$

where $0 < \alpha < 1$.

Define an objective function (the Lagrangian) using the Lagrangian multiplier γ, such that

$$\begin{aligned} \mathcal{L} &= P_D - \gamma(P_{FA} - \alpha) & (3)\\ &= \int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) d\mathbf{y} - \gamma \left[\int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_0) d\mathbf{y} - \alpha \right] & (4)\\ &= \int_{\mathcal{X}_1} \left[\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) - \gamma \mathrm{p}(\mathbf{y}|\mathcal{H}_0) \right] d\mathbf{y} + \gamma\alpha. & (5) \end{aligned}$$

For any given value γ, the region $\mathcal{X}_1$ that maximizes $\mathcal{L}$, and hence P_D, under the constraint $P_{FA} = \alpha$, is given by

$$\mathcal{X}_1 = \{\mathbf{y} \in \mathcal{X} \mid \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) > \gamma \mathrm{p}(\mathbf{y}|\mathcal{H}_0)\}. \tag{6}$$

Equation (6) yields the likelihood ratio test (LRT) [1] which is the *uniformly most powerful* (UMP) test [1] with $P_{FA} = \alpha$,

$$\Lambda(\mathbf{y}) = \frac{\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta})}{\mathrm{p}(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{7}$$

where γ is determined from the constraint $P_{FA} \leq \alpha$ in eqn. (2).

II. Illustrative Example

In the following examples, we use modified figures from [2]. Consider the P_D versus the P_{FA} for given threshold γ for two signal distributions shown in Fig. 1. Distribution *a* is *noise-only*, while distribution *b* is *signal-plus-noise*. Both distributions in Fig. 1 have the same distance between the peaks. The height of the distributions represent how often a signal is present, and the spread of the distributions indicate the magnitude of noise present; less noise reduces the spread of the distributions, while more noise increases the spread.

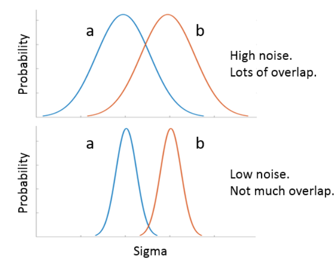

Fig. 1. Signal spread due to noise, and resulting overlap.

Consider the internal response to a signal detection system. In Fig. 2, the vertical line represents the threshold γ, which splits the figure into four regions: a *hit* is defined as the *signal-plus-noise* region greater than (to the right of) γ, and a *miss* is the *signal-plus-noise* region less than (to the left of) γ. *False alarms* represent the *noise-only* region greater than γ, while *correct rejection* represents the *noise-only* region less than γ.

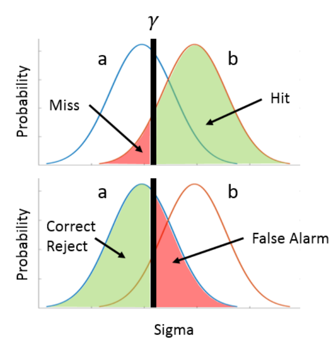

Fig. 2. The four regions of hypothesis testing, defined by two overlapping distributions (*a* & *b*) and the vertical line of the decision threshold γ.

Suppose that a low threshold is chosen, then the *signal-plus-noise* will likely be detected, and therefore the system will have a very high hit rate, at the cost of an increased number of false alarms. Conversely, if the threshold is chosen to be high, then the number of false alarms will be reduced, at the cost of an increased miss rate. This is demonstrated by an example of threshold-shifting, shown in Fig. 3.

Consider the following three conditions:

1) As the region $\mathcal{X}_1$ shrinks (γ tends toward infinity), both P_D and P_{FA} shrink toward zero.

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

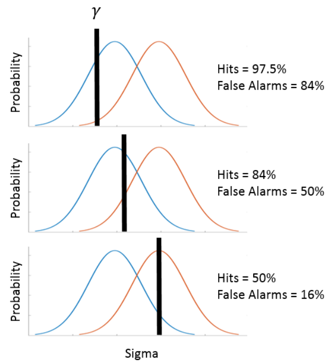

Fig. 3. An example result of *hits* versus *false alarms*, due to shifting the threshold γ.

2) As the region $\mathcal{X}_1$ grows (γ tends toward zero), both P_D and P_{FA} grow toward unity.
3) The case where $P_D = 1$ and $P_{FA} = 0$ will never occur if the conditional PDF's $\mathrm{p}(\boldsymbol{x}|\mathcal{H}_0)$ and $\mathrm{p}(\boldsymbol{x}|\mathcal{H}_1, \boldsymbol{\theta})$ overlap as in Fig. 1.

Item 3 in the list above represents the fundamental trade-off in hypothesis testing, and motivates the N-P Lemma.

III. Relation to ROC Plots

In the 1940's, the Allied forces in England sought to make sense of the signals received from new Radar technology [3]. Specifically, they needed a graphical way to represent and determine a good signal from random noise. This graphical method is called the Receiver Operating Characteristic (ROC) plot, with the P_D equal to one minus the Probability of Missed Detection (P_{MD}) on the vertical axis (shown as *Hits* in Fig. 4), versus the P_{FA} on the horizontal axis (shown as *False Alarms* in Fig. 4). The *Discriminability index* (d'), where $d' = 0$, is the *Line-of-No-Discrimination*, indicating that any signal along or below this line cannot be discerned from random noise. The curves above this line, $d' = 1 \cdots 4$, represent a detected signal with varying thresholds and/or discriminability. A perfectly detected signal lies on the vertical axis.

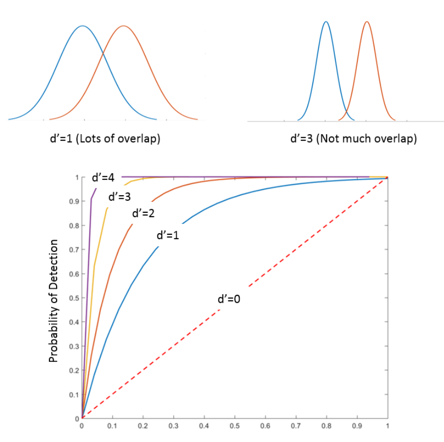

Fig. 4. ROC plot for Hits (P_D) versus False Alarms (P_{FA}).

The discriminability of a signal depends both on the separation and the spread of the *noise-only* and *signal-plus-noise* curves. Discriminability is made easier either by increasing the separation (stronger signal) or by decreasing the spread (less noise). The number, d', is often referred to as an estimate of the signal strength [3].

References

[1] 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.

[2] D. Heeger, "Signal Detection Theory," *Department of Psychology, New York University*, 2007.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. II - Detection Theory*. Prentice Hall PTR, 1998.