

[aprendtech.com](#) >> [blog](#) >> [this post](#)

If you have trouble viewing this, try [the pdf of this post](#). You can [download the code](#) used to produce the figures in this post.

Detection with multinormal data

In my [previous post](#), I showed that multivariate normal is a good model for x-ray measurements and in my [last post](#) I described the general properties of this distribution. In this post, I will discuss statistical detection theory with the normal model. I will show that the performance is characterized by a suitably defined signal to noise ratio. This will enable me to close the loop to the main topic of this series of posts, which is to explain the results in my paper, “*Near optimal energy selective x-ray imaging system performance with simple detectors*”[\[1\]](#), which is available for free download [here](#).

Section 3.3 of Kay[\[2\]](#) has an excellent discussion of classical statistical detection theory and Sec. 2.6 of Van Trees[\[3\]](#) describes its use with normal distributed data. I recommend that you study these references. Here, I will summarize and tie together their results. The main result will be that the probability of detection for a specified false alarm rate depends only on the signal to noise ratio (SNR) and the derivation of the SNR for vector data.

Binary decision making with mean-shifted data

Statistical detection theory is based on hypothesis testing. The simple binary hypothesis testing approach that I will use assumes that the probability distribution functions under both hypotheses are known. The problem is to decide between the hypotheses based on the random measurement(s). In my discussion, I will focus on the multinormal distribution with the same covariance under both hypotheses. With x-ray measurements, we know that the variance depends on the number of photons, which varies strongly with object thickness. This would seem to imply that the analysis is only applicable for thin features. In a future post, I will show that this is not necessarily true but the thin feature case is important since the objects significant for medical diagnosis like lung nodules or breast tumors are many times small. The equal covariance case is also widely used since it provides good insight to the factors affecting detection performance.

Suppose, we make just one measurement and the distributions P_{H_0} and P_{H_1} under the two hypotheses have the same variance σ with means m_0 and m_1 as shown in Fig. [1](#). To implement the detection algorithm, we divide the x-axis into two regions and decide H_0 if the measurement is in R_0 and H_1 in R_1 . Because of the random data, we cannot eliminate errors. We can measure the performance by two parameters, the probability of detection, $P_D = \text{Prob}(H_1; H_1)$ and the probability of false alarm, $P_{FA} = \text{Prob}(H_1; H_0)$, where the notation $\text{Prob}(H_i; H_j)$ means the probability of deciding H_i if H_j is true. Since m_0 is less than m_1 for the case shown in Fig. [1](#), a reasonable decision rule is to choose H_0 if x is less than a threshold γ and H_1 if it is larger. The equal case has zero probability so it can be assigned to either case. This leads to the decision regions R_0 and R_1 shown in the figure.

The Neyman-Pearson (NP) theorem

Given the data in Fig. [1](#), an immediate question is where to set the threshold. We cannot simultaneously minimize P_{FA} and maximize P_D but we can fix P_{FA} at some pre-determined value α and maximize P_D by using the Neyman-Pearson theorem

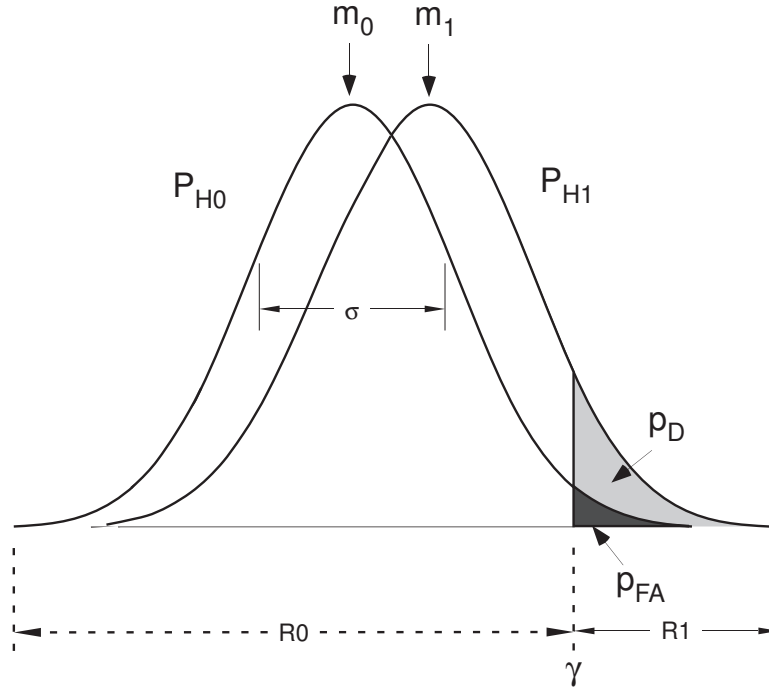


Figure 1: Hypothesis testing with a single measurement

To maximize P_D for a given $P_{FA} = \alpha$, decide H_1 if

$$R(\mathbf{x}) = \frac{p(\mathbf{x}; H_1)}{p(\mathbf{x}; H_0)} > \tau \quad (1)$$

where the parameter τ is chosen so

$$P_{FA} = \int_{\{\mathbf{x}; R(\mathbf{x}) > \tau\}} p(\mathbf{x}; H_0) d\mathbf{x} = \alpha$$

Note that the theorem is applicable to vector as well as scalar data.

Applying the NP theorem to scalar data

The probability distribution functions (PDF's) under the two hypotheses are

$$p(x; H_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{x - m_k}{\sigma} \right)^2 \right] \quad k = 1, 2 \quad (2)$$

Taking the logarithm of the likelihood ratio R in (1), substituting the PDF's (2) and changing the variable to $u = x - m_0$, the NP test is

$$\mathcal{L}(u) = \log R = \frac{u^2 - (u - \delta m)^2}{2\sigma^2} > \log \tau$$

Expanding the left side and rearranging terms, the NP test is $u > \gamma'$ where

$$\gamma' = \frac{\sigma^2 \log \tau}{\delta m} + \frac{\delta m}{2} \quad (3)$$

With these, we can compute the optimal probability of detection for a specified false alarm rate. First, I will define some notation. Let $F(x)$ be the cumulative distribution function of the $\mathcal{N}(0, 1)$ random variable

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

and $F_c = 1 - F$ be its complement. Since u has an $\mathcal{N}(0, \sigma^2)$ distribution

$$P_{FA} = F_c\left(\frac{\gamma'}{\sigma}\right)$$

We can invert this since F_c is a monotonically decreasing function so $\gamma' = \sigma F_c^{-1}(P_{FA})$. Since $p(x; H_1) \sim \mathcal{N}(\delta m, \sigma^2)$, the probability of detection is

$$P_D = F_c\left(\frac{\gamma' - \delta m}{\sigma}\right)$$

Substituting for γ' ,

$$P_D = F_c\left(F_c^{-1}(P_{FA}) - \frac{\delta m}{\sigma}\right) \quad (4)$$

This shows that the probability of detection for a given false alarm probability depends only on the signal to noise ratio (or its square)

$$SNR^2 = \frac{(\langle x; H_1 \rangle - \langle x; H_0 \rangle)^2}{\text{variance}(x; H_0)} \quad (5)$$

where $\langle x; H_k \rangle$ denotes the expected value of x with the $p(x; H_k)$ distribution.

We can evaluate Eq. 4 using Matlab functions to evaluate F_c , and F_c^{-1} . The Matlab function `normcdf` computes F and $F_c = 1 - \text{normcdf}$. Also, the Matlab `norminv` function computes F^{-1} and $F(F^{-1}(x)) = x = 1 - F_c(F^{-1}(x))$ so

$$\begin{aligned} F_c(F^{-1}(x)) &= 1 - x \\ F^{-1}(x) &= F_c^{-1}(1 - x) \end{aligned}$$

Setting $t = 1 - x$,

$$F_c^{-1}(t) = \text{norminv}(1 - t)$$

Fig. 2 shows the probability of detection P_D as a function of the SNR^2 in decibels (dB) for different values of the false alarm probability P_{FA} . Note that as expected, P_D always increases as the signal to noise ratio increases. The code to reproduce this figure is [available here](#). This figure should be compared with Fig. 3.5 of Kay[2].

Binary decisions with vector data: the equal covariance case

With multivariate normal data and assuming equal covariance under the two hypotheses, the probability distributions are (see my [last post](#)).

$$p(\mathbf{x}; H_k) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}|^{1/2}} \exp \left[-1/2 (\mathbf{x} - \mathbf{m}_k)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}_k) \right] \quad k = 1, 2$$

where \mathbf{C} is the covariance and n is the dimension of the vector data. Substituting in the definition of the likelihood ratio (1) and taking the logarithm of both sides, the log-likelihood is

$$\mathcal{L}(\mathbf{x}) = -\frac{(\mathbf{x} - \mathbf{m}_1)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}_1) - (\mathbf{x} - \mathbf{m}_0)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}_0)}{2}$$

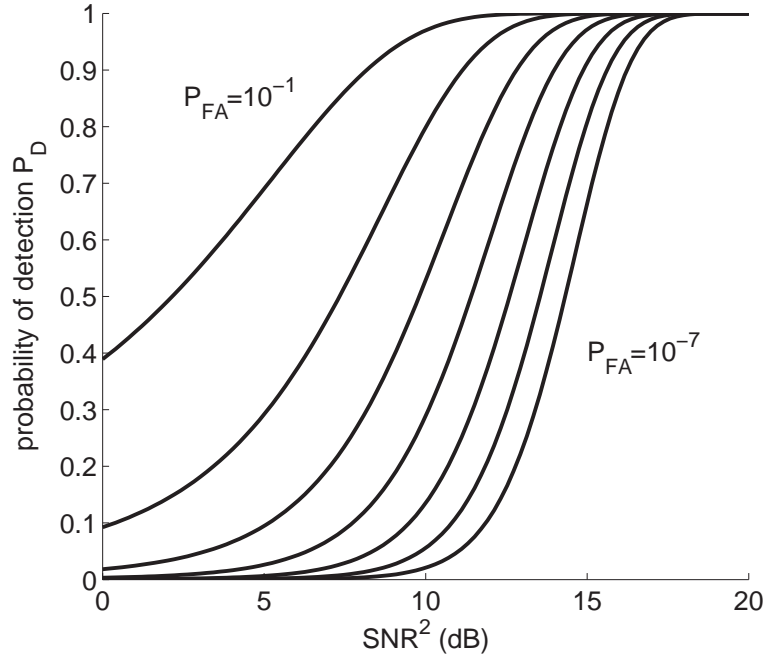


Figure 2: P_D as a function of signal to noise ratio squared.

Expanding the products, gathering terms, and defining $\delta \mathbf{m} = \mathbf{m}_1 - \mathbf{m}_0$

$$-2\mathcal{L}(\mathbf{x}) = \delta \mathbf{m}^T \mathbf{C}^{-1} \mathbf{x} + \mathbf{m}_1^T \mathbf{C}^{-1} \mathbf{m}_1 - \mathbf{m}_0^T \mathbf{C}^{-1} \mathbf{m}_0$$

The last two terms on the right hand side do not depend on the data and a multiplicative factor does not affect the results so the Neyman-Pearson test is

$$\mathcal{L}(\mathbf{x}) = \delta \mathbf{m}^T \mathbf{C}^{-1} \mathbf{x} > \gamma' \quad (6)$$

The log-likelihood $\mathcal{L}(\mathbf{x})$ is a linear combination of a multivariate normal random variable so it also has normal distribution. It is a scalar so we can apply the results of the previous section to see that the performance is determined by the SNR^2 from Eq. 5

$$SNR^2 = \frac{(\langle \mathcal{L}; H_1 \rangle - \langle \mathcal{L}; H_0 \rangle)^2}{\text{variance}(\mathcal{L}; H_0)} \quad (7)$$

From (6), the expected values are

$$\langle \mathcal{L}; H_k \rangle = \delta \mathbf{m}^T \mathbf{C}^{-1} \mathbf{m}_k \quad (8)$$

so the numerator of (7) is

$$(\langle \mathcal{L}; H_1 \rangle - \langle \mathcal{L}; H_0 \rangle)^2 = (\delta \mathbf{m}^T \mathbf{C}^{-1} \delta \mathbf{m})^2 \quad (9)$$

Using the definition of variance

$$\text{variance}(\mathcal{L}; H_0) = \left\langle (\mathcal{L}(\mathbf{x}) - \langle \mathcal{L}; H_0 \rangle)^T (\mathcal{L}(\mathbf{x}) - \langle \mathcal{L}; H_0 \rangle) \right\rangle \quad (10)$$

From the definition of \mathcal{L} (6) and the expected value (8),

$$\begin{aligned} \mathcal{L}(\mathbf{x}) - \langle \mathcal{L}; H_0 \rangle &= \delta \mathbf{m}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}_0) \\ &= (\mathbf{x} - \mathbf{m}_0)^T \mathbf{C}^{-1} \delta \mathbf{m} \end{aligned}$$

where the last step follows because \mathcal{L} is a scalar so it is equal to its transpose as is the covariance because it is symmetric.

$$\begin{aligned} \text{variance}(\mathcal{L}; H_0) &= \left\langle \delta \mathbf{m}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}_0) (\mathbf{x} - \mathbf{m}_0)^T \mathbf{C}^{-1} \delta \mathbf{m} \right\rangle \\ &= \delta \mathbf{m}^T \mathbf{C}^{-1} \mathbf{C} \mathbf{C}^{-1} \delta \mathbf{m} \\ &= \delta \mathbf{m}^T \mathbf{C}^{-1} \delta \mathbf{m} \end{aligned} \quad (11)$$

Substituting (9) and (11) in (7), the vector SNR^2 is

$$SNR^2 = \delta \mathbf{m}^T \mathbf{C}^{-1} \delta \mathbf{m} \quad (12)$$

Example: “white” data

Data with independent, equal variance provide good insight and are also important in practice since we can use the whitening matrix, Φ_w discussed in my [last post](#), to transform any multinormal to this case. In a future post, I will describe how we can use whitening with x-ray spectral data.

For whitened data, the covariance matrix is $\mathbf{C} = \sigma^2 \mathbf{I}$ and its inverse is $\mathbf{C}^{-1} = 1/\sigma^2 \mathbf{I}$. Substituting in (12), the signal to noise ratio is

$$SNR_W^2 = \frac{|\delta \mathbf{m}|^2}{\sigma^2}$$

This is the distance between the means squared divided by the variance. From (6), the NP test is

$$\mathcal{L}(\mathbf{x}) = \frac{1}{\sigma^2} \delta \mathbf{m}^T \mathbf{x} > \gamma'$$

This is the dot product of the data with the mean difference vector $\delta \mathbf{m}$.

Conclusion

I have now provided the statistical detection theory framework that I will use to analyze the Tapiovaara-Wagner imaging task[4] for x-ray data with spectral information. We can use this framework to introduce the basis set expansion of the attenuation coefficient as shown in my paper[1]. This will allow us to compare the performance of systems with limited energy resolution to the ideal case with full spectral information.

Bob Alvarez

Last edited January 20, 2012

©2012 by Aprend Technology and Robert E. Alvarez

Linking is allowed but reposting or mirroring is expressly forbidden.

References

- [1] R. E. Alvarez, “Near optimal energy selective x-ray imaging system performance with simple detectors,” *Med. Phys.* **37**, 822–841, (2010).
- [2] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, 1st ed., Prentice Hall PTR, 1998.

- [3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part I: Detection, Estimation, and Linear Modulation Theory*, John Wiley & Sons Inc, 2001.
- [4] M. J. Tapiovaara and R. Wagner, "SNR and DQE analysis of broad spectrum X-ray imaging," *Phys. Med. Biol.* **30**, 519–529, (1985).