AOE 5984 Hypothesis Testing Example

Consider two hypotheses about the current deflection of Prof. Psiaki's ear drums due to noise in his house:

$$p(x|H_0) = \frac{1}{\sqrt{2\pi}\sigma_A} exp(-\frac{x^2}{2\sigma_A^2})$$

$$p(x|H_1) = \frac{1}{\sqrt{2\pi}\sigma_B} exp(-\frac{x^2}{2\sigma_B^2})$$

with $\sigma_B > \sigma_A$. These are both Gaussian distributions, but one has a higher standard deviation. Hypothesis H_0, the null hypothesis, is that his son Tim is not home. Hypothesis H_1 is that Tim is home and that he is playing his guitar loudly, which is why σ_B is larger than σ_A.

A Neyman-Pearson optimal hypothesis test based on a single measurement takes the form:

Accept H_1 if $\frac{p(x|H_1)}{p(x|H_0)} \geq \Lambda$; otherwise, accept H_0.

One can use the probability density function definitions to re-write the hypothesis test's H_1 acceptance condition in the form:

$$\frac{\sigma_A}{\sigma_B} exp[-\tfrac{1}{2}(\frac{1}{\sigma_B^2} - \frac{1}{\sigma_A^2})x^2] \geq \Lambda$$

Suppose that one next multiplies both sides of this equation by σ_B/σ_A, takes the natural logarithm of both sides, which does not affect the inequality, and further multiplies both sides by $2\sigma_A^2\sigma_B^2/(\sigma_B^2-\sigma_A^2)$. The resulting H_1 acceptance condition then becomes:

$$x^2 \geq \frac{2\,ln(\Lambda\sigma_B/\sigma_A)\sigma_A^2\sigma_B^2}{\sigma_B^2 - \sigma_A^2} = x_0^2$$

where x_0 is a new unknown constant that replaces the unknown constant Λ in the design of the test. One determines the value of x_0 that constrains the false alarm probability to take on a certain value. This value of x_0, in effect, determines Λ:

$$\Lambda = \frac{\sigma_A}{\sigma_B} exp[-\tfrac{1}{2}(\frac{1}{\sigma_B^2} - \frac{1}{\sigma_A^2})x_0^2]$$

One determines x_0 (and by extension Λ), by finding the value which satisfies the false alarm bound condition:

$$P\{x^2 \geq x_0^2 \mid H_0\} = \alpha$$

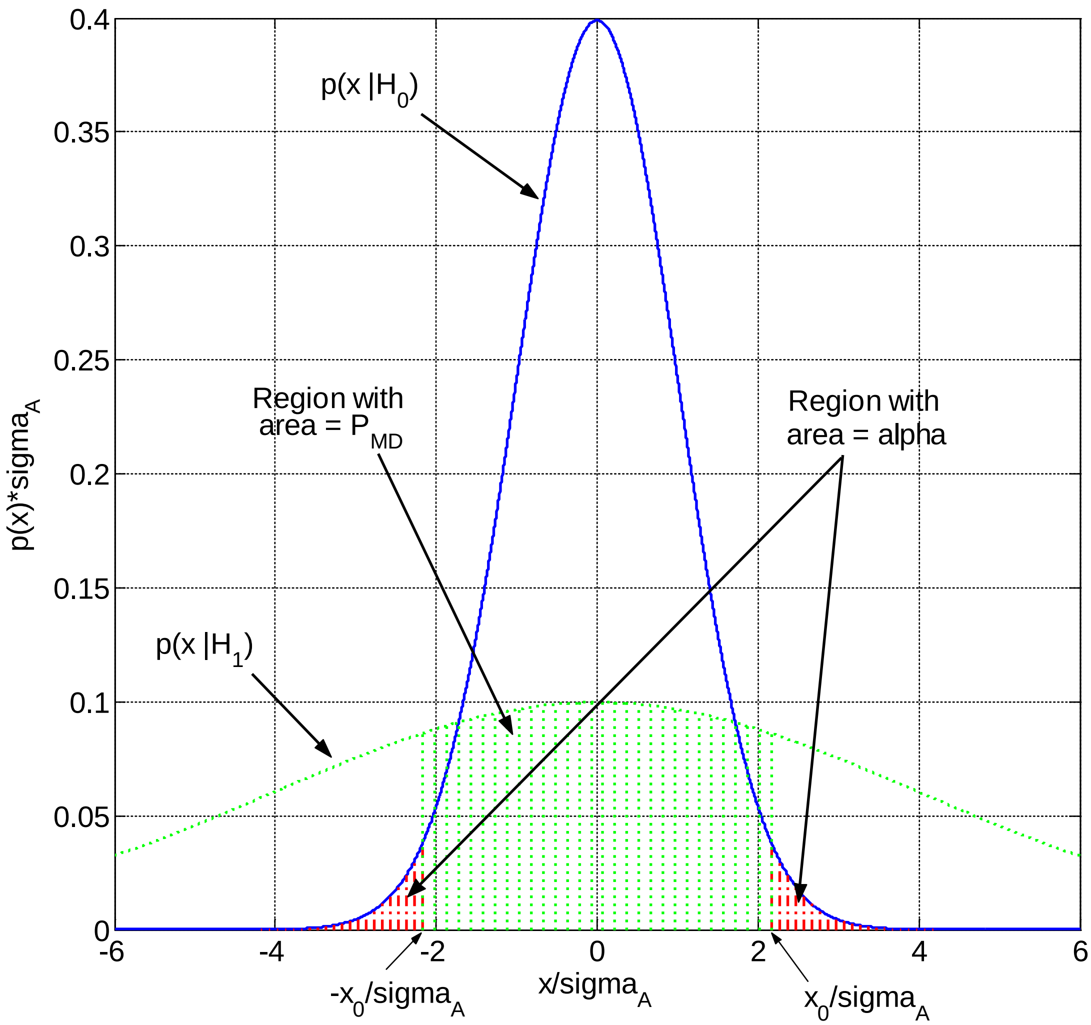
which is equivalent to

$$\alpha = P\{|x| \geq x_0 \mid H_0\} = \int_{-\infty}^{-x_0} \frac{1}{\sqrt{2\pi}\sigma_A} exp(-\frac{x^2}{2\sigma_A^2})dx + \int_{x_0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_A} exp(-\frac{x^2}{2\sigma_A^2})dx$$

or

$$\alpha = 2\int_{-\infty}^{-x_0} \frac{1}{\sqrt{2\pi}\sigma_A} exp(-\frac{x^2}{2\sigma_A^2})dx$$

This is a condition on the tail areas of the H_0 Gaussian distribution as in the following figure:



The corresponding value of x_0 can be determined by using MATLAB's inverse cumulative normal (Gaussian) distribution function: $x_0 = -\,norminv(\alpha/2,0,\sigma_A) = -\,\sigma_A norminv(\alpha/2,0,1)$, where $norminv(P,\mu,\sigma)$ gives the value of x_P such that P is the probability that $-\infty < x \leq x_P$ if x is sampled from a Gaussian (normal) distribution with mean μ and standard deviation σ. Other helpful MATLAB functions are *normpdf*() and *normcdf*(), which are, respectively, the probability density function and the cumulative distribution function for a Gaussian distribution. The function *norminv*(,μ,σ) is the inverse of the function *normcdf*(,μ,σ). One can learn how to use these functions (or any MATLAB function) if one types >>*help functionname* after the MATLAB prompt.

If one chooses $\alpha = 0.03$ for a 3% false alarm probability, then $x_0 = 2.1701\sigma_A$.

One can use this value to compute the probability of a missed detection according to the formula:

$$P_{MD} = P\{|x| < x_0 \mid H_1\} = \int_{-x_0}^{x_0} \frac{1}{\sqrt{2\pi}\sigma_B} exp(-\frac{x^2}{2\sigma_B^2})dx$$

This area is shown in the figure given above for the case where $\sigma_B = 4\sigma_A$. One can also compute the power of the hypothesis test:

$$Power = 1 - P_{MD} = 1 - \int_{-x_0}^{x_0} \frac{1}{\sqrt{2\pi}\sigma_B} exp(-\frac{x^2}{2\sigma_B^2})dx$$

$$= \int_{-\infty}^{-x_0} \frac{1}{\sqrt{2\pi}\sigma_B} exp(-\frac{x^2}{2\sigma_B^2})dx + \int_{x_0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_B} exp(-\frac{x^2}{2\sigma_B^2})dx$$

$$= 2\int_{-\infty}^{-x_0} \frac{1}{\sqrt{2\pi}\sigma_B} exp(-\frac{x^2}{2\sigma_B^2})dx$$

The *Power* area integral for this example is similar to the false alarm probability area integral shown in the first figure, except that $p(x|H_1)$ is used in the integral in place of $p(x|H_0)$. One can use the MATLAB function *normcdf*() to calculate the probability of a missed detection and the power: P_{MD} = *normcdf*(x_0,0,σ_B) - *normcdf*(-x_0,0,σ_B) = *normcdf*(x_0/σ_B,0,1) - *normcdf*(-x_0/σ_B,0,1) and *Power* = 2 *normcdf*(-x_0,0,σ_B) = 2 *normcdf*(-x_0/σ_B,0,1).

According to the Neyman-Pearson Lemma, the hypothesis test rule that accepts H_1 if $|x| \geq x_0$ is the most powerful test (i.e., has the largest power and therefore the lowest probability of a missed detection) for a given false alarm rate. The following P_{MD} and *Power* values apply to this test when the false alarm rate is $\alpha = 0.03$ and the corresponding test threshold is $x_0 = 2.1701\sigma_A$:

σ_B/σ_A	P_{MD}	*Power*
4	0.4125	0.5875
6	0.2824	0.7176
8	0.2138	0.7862
10	0.1718	0.8282

An Improved Hypothesis Test Based on Multiple Measurements

The hypothesis test defined above is not very powerful. When $\sigma_B = 4\sigma_A$ there is a 41.25% probability of missed detection even though the false alarm rate is relatively high, 3%. The theory of hypothesis testing provides a method for decreasing both the false alarm rate and the probability of a missed detection. The probabilities of both types of errors can be reduced if one uses more measurements. The above example will be used in order to illustrate this point.

Suppose that Prof. Psiaki measures his ear drum deflection n times rather than just once. Suppose, also, that each of the measurements is spaced far enough apart in time from the others to be independent of all the others. In this case, the following joint probability density functions apply to the n measurements under the two competing hypotheses:

$$p(x_1, x_2, x_3, ..., x_n|H_0) = \frac{1}{(2\pi)^{n/2}\sigma_A^n} exp(-\frac{1}{2\sigma_A^2}\sum_{i=1}^{n} x_i^2)$$

$$p(x_1, x_2, x_3, ..., x_n|H_1) = \frac{1}{(2\pi)^{n/2}\sigma_B^n} exp(-\frac{1}{2\sigma_B^2}\sum_{i=1}^{n} x_i^2)$$

The Neyman-Pearson optimal hypothesis test then becomes:

Accept H_1 if $\frac{p(x_1, x_2, x_3, ..., x_n\,|H_1)}{p(x_1, x_2, x_3, ..., x_n\,|H_0)} \geq \Lambda$; otherwise, accept H_0.

Substitution of the two multi-variable probability density functions into this inequality yields the test condition:

$$\left(\frac{\sigma_A}{\sigma_B}\right)^n exp[-\tfrac{1}{2}(\frac{1}{\sigma_B^2} - \frac{1}{\sigma_A^2})\sum_{i=1}^{n} x_i^2] \geq \Lambda$$

which can be manipulated into the form:

$$q = \sum_{i=1}^{n} x_i^2 \geq \frac{2\,ln(\Lambda[\sigma_B/\sigma_A]^{\mathrm{n}})\sigma_A^2\sigma_B^2}{\sigma_B^2 - \sigma_A^2} = q_0$$

In other words, the test statistic q, which equals the sum of the squares of the measurements, must be greater than or equal to some lower bound q_0 in order for hypothesis H_1 to be accepted. Note how the undetermined bound q_0 substitutes for the undetermined constant Λ. As in the single measurement case, one can use q_0 to determine Λ:

$$\Lambda = \left(\frac{\sigma_A}{\sigma_B}\right)^n exp[-\tfrac{1}{2}(\frac{1}{\sigma_B^2} - \frac{1}{\sigma_A^2})q_0]$$

although there is never any need to determine Λ.

It is necessary to determine the two probability distributions for the test statistic q under the two hypotheses, $p(q|H_0)$ and $p(q|H_1)$. The former function is needed in order to determine the q_0

value that yields a false alarm probability equal to α, and the latter function is needed in order to calculate the probability of a missed detection and the power of the hypothesis test criterion. These two probability densities can be calculated by recognizing that $q/\sigma_A^2 \sim \chi_n^2$ under hypothesis H_0 and that $q/\sigma_B^2 \sim \chi_n^2$ under hypothesis H_1. Therefore, the requisite probability density functions are:

$$p(q|H_0) = \begin{cases} 0 & \text{if } q < 0 \\ \dfrac{1}{(2)^{n/2}\Gamma(n/2)\sigma_A^2}\left(\dfrac{q}{\sigma_A^2}\right)^{\frac{(n-2)}{2}} exp(-\dfrac{q}{2\sigma_A^2}) & \text{if } 0 \le q \end{cases}$$

$$p(q|H_1) = \begin{cases} 0 & \text{if } q < 0 \\ \dfrac{1}{(2)^{n/2}\Gamma(n/2)\sigma_B^2}\left(\dfrac{q}{\sigma_B^2}\right)^{\frac{(n-2)}{2}} exp(-\dfrac{q}{2\sigma_B^2}) & \text{if } 0 \le q \end{cases}$$

One can use the first of these two probability density functions to determine q_0 by solving the following equation:

$$\begin{aligned} \alpha = P\{q \ge q_0 |H_0\} &= \int_{q_0}^{\infty} \frac{1}{(2)^{n/2}\Gamma(n/2)\sigma_A^2}\left(\frac{q}{\sigma_A^2}\right)^{\frac{(n-2)}{2}} exp(-\frac{q}{2\sigma_A^2})dq \\ &= \int_{q_0/\sigma_A^2}^{\infty} \frac{1}{(2)^{n/2}\Gamma(n/2)}\beta^{\frac{(n-2)}{2}} exp(-\frac{\beta}{2})d\beta \\ &= 1 - \int_{0}^{q_0/\sigma_A^2} \frac{1}{(2)^{n/2}\Gamma(n/2)}\beta^{\frac{(n-2)}{2}} exp(-\frac{\beta}{2})d\beta \end{aligned}$$

where the change of integration variable $\beta = q/\sigma_A^2$ has been used to go from the first line to the second line in this equation. This following MATLAB calculation solves this equation: $q_0 = \sigma_A^2 chi2inv(1-\alpha, n)$. The function $chi2inv(P, n)$ gives a gives a threshold value q_{max} such that P is the probability that q is less then or equal to this threshold if q is a sample from a chi-squared distribution of degree n. The two related functions $chi2pdf(q, n)$ and $chi2cdf(q, n)$ calculate, respectively, the corresponding probability density function and cumulative distribution for a degree-n chi-squared distribution. The function $chi2inv(, n)$ is the mathematical inverse of the function $chi2cdf(, n)$.

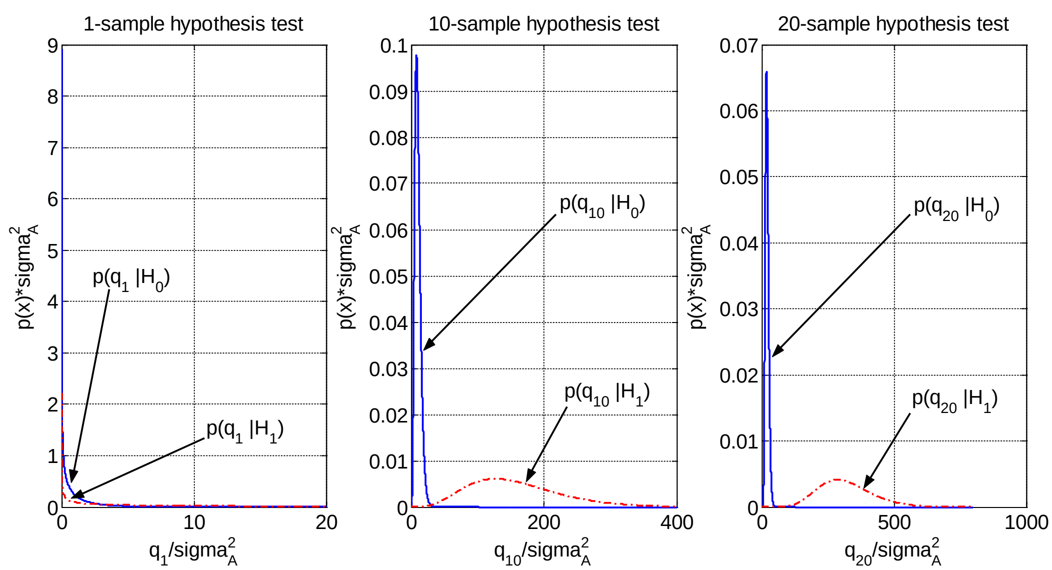
Given q_0, one can use $p(q|H_1)$ to calculate the probability of a missed detection and the power of the test:

$$
\begin{aligned}
P_{MD} = P\{q < q_0 \,|H_1\} &= \int_0^{q_0} \frac{1}{(2)^{n/2}\Gamma(n/2)\sigma_B^2}\left(\frac{q}{\sigma_B^2}\right)^{\frac{(n-2)}{2}} exp(-\frac{q}{2\sigma_B^2})dq \\
&= \int_0^{q_0/\sigma_B^2} \frac{1}{(2)^{n/2}\Gamma(n/2)}\beta^{\frac{(n-2)}{2}} exp(-\frac{\beta}{2})d\beta \\
&= 1 - Power
\end{aligned}
$$

where the change of integration variable $\beta = q/\sigma_B^2$ has been used to go from the first line to the second line in this equation. The following MATLAB calculation gives the missed detection probability $P_{MD} = chi2cdf(q_0/\sigma_B^2, n)$.

The improvement of the test due to the use of more data is easily illustrated by an example. Suppose that Prof. Psiaki takes $n = 10$ samples and suppose that he wants to limit his false alarm probability to $\alpha = 0.001$. Suppose, also, that $\sigma_B = 4\sigma_A$. In this case, the optimal test uses the threshold $q_0 = 29.5883\sigma_A^2$, the probability of a missed detection is $P_{MD} = 0.0026$, and the test power is $Power = 0.9974$. This false alarm probability is 30 times smaller than the false alarm probability used with the $n = 1$ test in the first example, and the probability of a missed detection is 157 times smaller than the corresponding P_{MD} value from $n = 1$ example -- review the first line of the table shown above.

The value of using more data to construct a hypothesis test is further illustrated in the following figure:



This figure compares $p(q|H_0)$ and $p(q|H_1)$ for 3 different n values. In all three cases $p(q|H_0)$ is the solid blue curve, and $p(q|H_1)$ is the dash-dotted red curve. The left-most plot uses $n = 1$ data point, as in the original example of this write-up. (This alternate analysis of that example in terms of a degree-1 chi-squared distribution is equally as valid as the original analysis that considers the 2 tails of the corresponding Gaussian distribution). The center plot corresponds to $n = 10$ data points, as in the second example, and the right-hand plot corresponds to $n = 20$ data points. The important feature to notice is the separation between the two probability density functions. The goal in designing a hypothesis test is to define a test static whose probability density functions under the two competing hypotheses are well separated. If these probability density functions are well separated, then it will be possible to achieve both a low probability of false alarm and a low probability of missed detection. As one can see from the figure, these goals becomes easier and easier to achieve simultaneously as one uses more data -- note how the peak in $p(q|H_1)$ curve moves to the right faster than does the peak in the $p(q|H_0)$ curve as n increases.

Another important point of the Neyman-Pearson lemma is that it gives the optimal test statistic. One might decide to use the following alternate hypothesis test criterion for the example problem:

$$\text{Accept } H_1 \text{ if } \quad r = \sum_{i=1}^{n} | x_i | \geq r_0 \text{; otherwise, accept } H_0.$$

This test statistic is the sum of the absolute values of the x_i rather than the sum of their squares. This seems like a reasonable hypothesis test because it evaluates the cumulative magnitude of the x_i and decides in favor of hypothesis H_1 if this cumulative magnitude is too large. One could derive probability density functions for this test statistic, $p(r|H_0)$ and $p(r|H_1)$, and one could use $p(r|H_0)$ to define an integral relationship which could be used to determine r_0 as a function of α. Thus, the test would have the desired false alarm rate. Unfortunately, this test would not have as low a probability of missed detection as the optimal Neyman-Pearson test, which considers the sum of the squares of the x_i values rather than the sum of their absolute values.