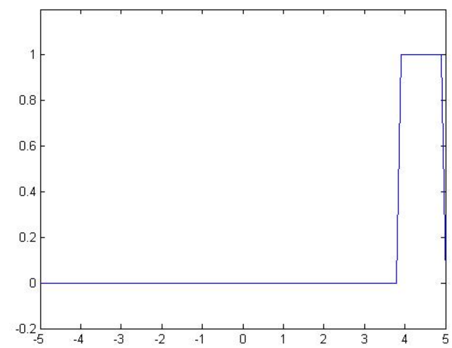
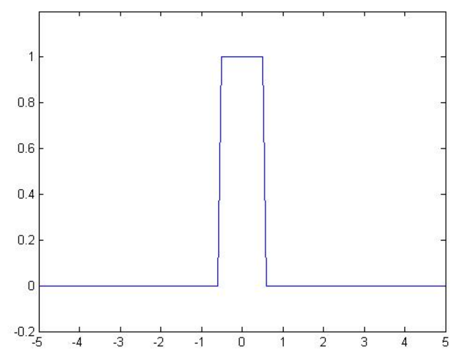
Exam 1 – Reworked

ONE –

Given the following probability distributions present a sketch of the error P(E) for a maximum likelihood classifier as a function of α and $P(\omega_1)$.



$p(x|\ \omega_1) = 1$ while $\alpha - \frac{1}{2} \leq x \leq \alpha + 1/2$, 0 otherwise

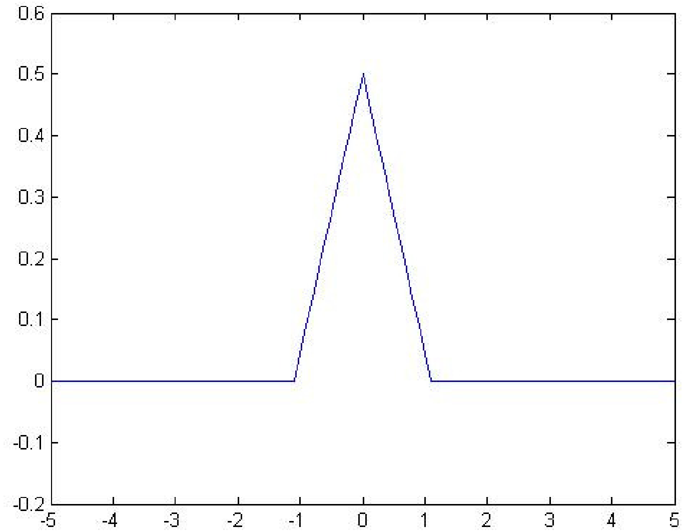


$p(x|\ \omega_2) = 1$ while $-1/2 \leq x \leq \frac{1}{2}$, 0 otherwise

The error of the functions can be found by shifting the one with variable α along the x axis and multiplying by the value of the static function. In essence, if the functions are convolved the resultant area will indicate the region and strength of the errors. It should be noted that there can be no error until the functions overlap and then when they fully overlap each other the error should peak. The maximum error seen should be 50%, verified from the equation below:

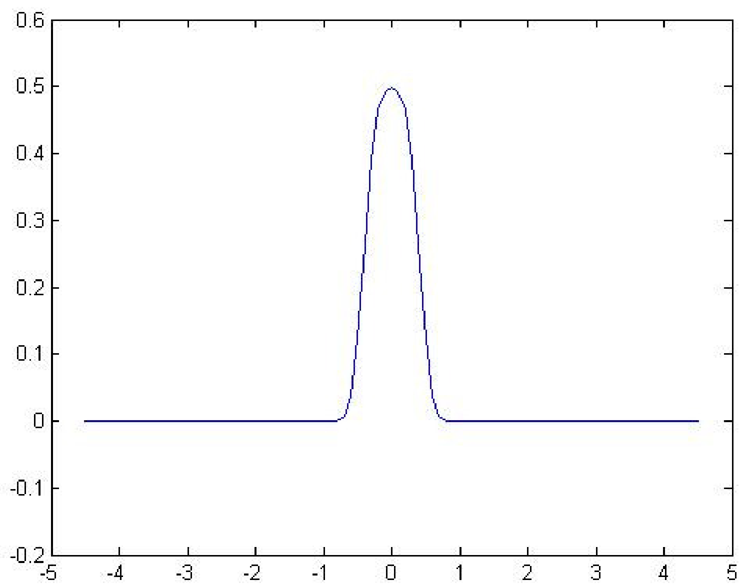
$P(E) = \frac{1}{2} * \int p(x|\omega_2)p(x|\ \omega_1)\, dx$; over the range $-\frac{1}{2} + \alpha$ to $\alpha + \frac{1}{2}$ window

This results in zero error where the two functions do not overlap, and produces error only once they have being to overlap. Two squares convolved together should result in a triangle, and this is what is found from the P(E) plot.



If the distributions were instead Gaussians, how would this later the probability of error?

Building two Gaussian functions to mimic the uniform density functions requires a sigma of 1/6 to ensure 99% of the function falls in an area of 1 on the x axis. Again, the error rate does not exceed 50% and can be found using the same equation as before, but with Gaussian distributions that have a sigma of 1/6.



TWO -

Given the following uniform density function: $p(x|\Theta) = 1/\Theta$ when $0 \leq x \leq \Theta$ and 0 otherwise. If *n* samples are drawn to create a set of $D_n = \{ x_1, x_2, \ldots, x_n \}$ what is the expression for the likelihood estimate of Θ. That is, what is the probability of Θ given various sample sizes of data and what happens to this as *n* beings to approach infinity?

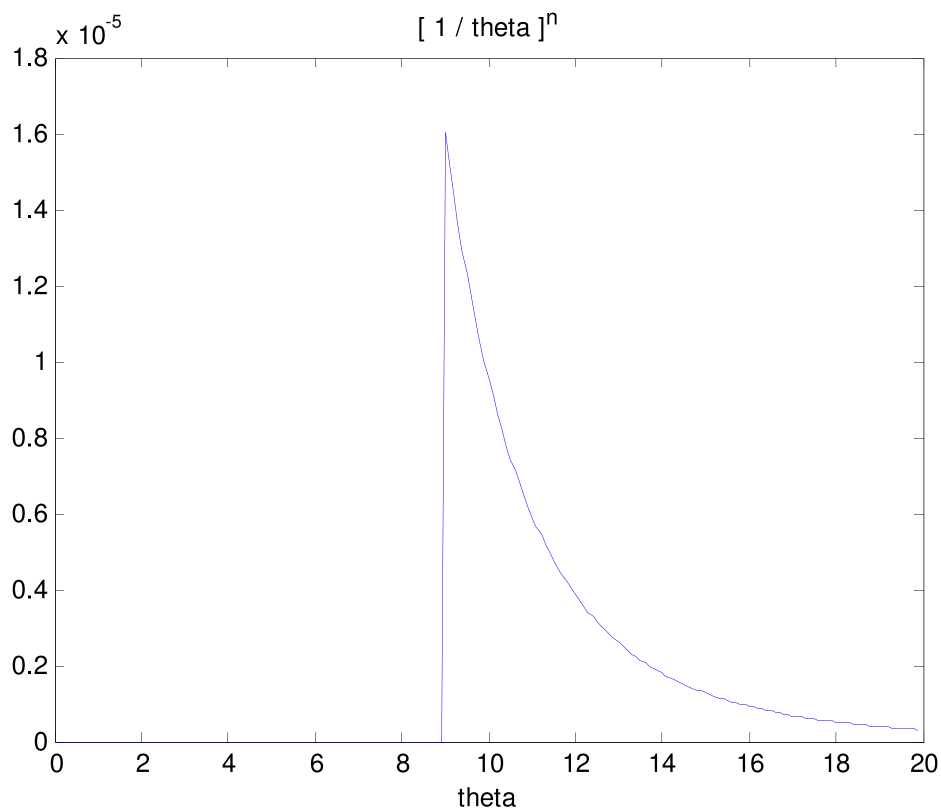
Pulling from Bayesian conditional probability, the likelihood of the data given Θ can be expressed in the following manner:

$$p(D_n \mid \Theta) = p(\Theta \mid D_n)p(D_n) / p(\Theta)$$

This equation is troublesome in that we are not presented with any of the required probabilities to solve for the likelihood of the data given Θ. If the data set was small, the solution would be to find the probability of each data point in the data set and build it into the numerator. While not practical, it shows that our true goal is the total likelihood of all of the variables chosen with D_n. This is an equation we can produce and looks like the following:

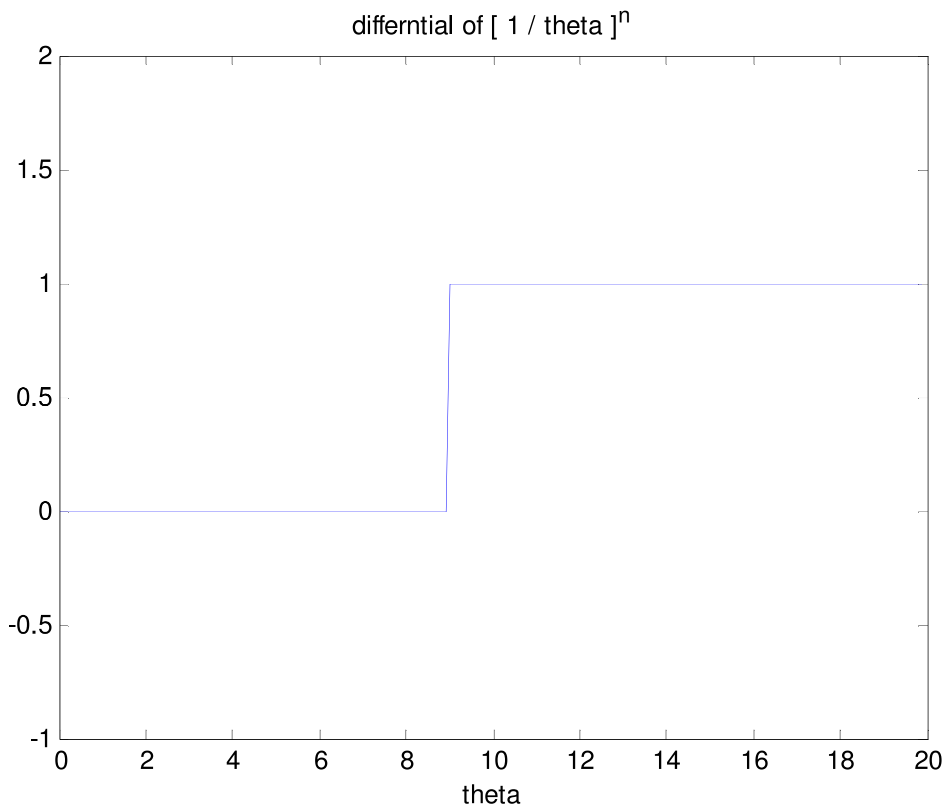
$$p(D_n \mid \Theta) = \Pi\, p(x_k \mid \Theta) \text{ over the range of } k=1 \text{ to } n$$

$$p(D_n \mid \Theta) = [1/\Theta]^n \text{ where } 0 \leq x \leq \Theta$$



For a given set of 5 random values of theta the results look like this when the maximum random value selected is nine. This also results in the limits on the function changing to enable a step response only when the min of x_k to be greater than or equal to 0 and the max of x_k to be greater than or equal to Θ. This only allows the function to exist for values greater than 0, where the function will be zero, until Θ equals the max of x_k after which the function will trail off in terms of $1/\Theta^n$. This is seen in the above plot.

The derivative of the function becomes $-n * (1/\Theta)^{(n+1)}$ and results in a function that returns one once the ideal value of theta has been reached.

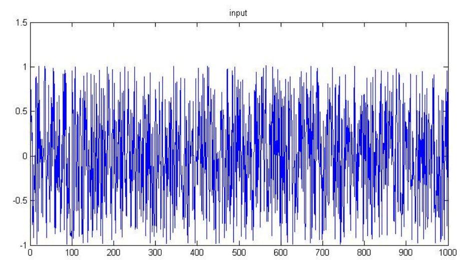


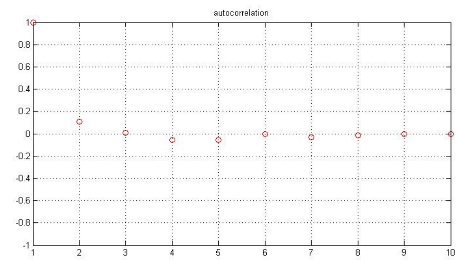
It can be seen that as *n* increases the likelihood of the function approaches very small values of probability as Θ grows. For small values of *n* and small values of Θ the likelihood is spread further out as each additional selection for D_n increments Θ and thus reduces the likelihood of all possible outcomes. This manifests once the value of Θ is exceeded as the function only returns values starting at Θ equals max x_k.

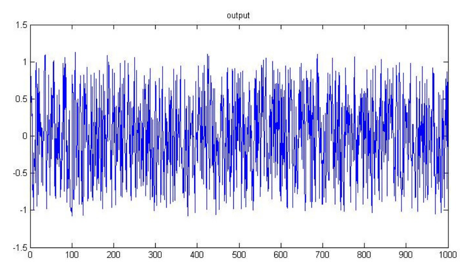
THREE –

A zero-mean unit variance discrete-time Gaussian white noise single, $x[n]$, is applied to a digital filter: $H(z) = 1\ /\ 1(\ 1 - \alpha z^{-1}\ )$. Assume you only have access to the output of this filter, but you do know the form of the filter (you just don't have the specific value of α), and you can assume the input is a zero-mean Gaussian white noise. Derive or explain how you would construct a maximum likelihood estimate of the filter coefficient. Hint: think about the pdf for the difference of two random variables. Second hint: think about the role correlation can play in this estimate.

For a given α value the input, output and first few terms of autocorrelation should look like the following plots.







As stated in the problem, access to the initial signal, $x[n]$, is not given so the only data that can be worked with is the output data. When given the transfer function of the system, the equation for the output can be used to find the minimum error between input and output. Starting with the following equation:

$$\Sigma\ x[n] = \Sigma\ (\ y[n] + \alpha y[n-1]\ )$$

Where the idea is to minimize the difference between the left and right hand sides of the equation. Using a least squares regression approach it will be found that:

$$\Sigma\ e^2[n] = \Sigma\ (\ y[n] - x[n] - \alpha y[n-1]\ )^2$$

$$\Sigma\ e^2[n] = \Sigma\ \ (\ y^2[n] + 2\alpha x[n]y[n-1] + \alpha^2 y^2[n-1]\ )$$

Differentiating with respect to alpha and setting the result equal to zero to find the inflection of the resultant curve allows one to solve for α.

$$0 = \Sigma \left(2^*x[n]y[n-1] + 2\alpha y^2[n-1] \right)$$

Isolating α produces a division of the two summations, which will resemble terms from the autocorrelation equations. Where R(k) is the result of the autocorrelation function for various intervals of k. Given that the input is a zero mean Gaussian, Σ x[n] is concluded to be one, just like Σy[n] would be which allows one to replace the numerator with the autocorrelation value of R(1).

$$-\alpha = \Sigma \left(x[n]y[n-1] \right) / \Sigma \left(y[n-1]y[n-1] \right)$$

$$-\alpha = R(1) / R(0)$$

This result should be mirrored when applying Maximum Likelihood Estimation to the coefficient of alpha. Based out of the difference equation $y[n] = \alpha y[n-1] + x[n]$ and knowing that the regression equation to find the output in terms of the previous value is $y[n] = ß_0 + ß_1 y[n-1]$. The goal should be to find that the beta terms result in those present in the different equation.

The initial step is to work out what the Gaussian distribution of y[n] would be so the regression equation can be solved for partial derivatives of the betas and sigma.

$$Y_i = N(ß_0+ß_1Y_{i-1},\sigma^2)$$

$$\text{Pdf of } Y_i \rightarrow f_i = 1 / \left(\text{sqrt}(2^*\text{pi})^*\sigma \right) * \exp\left(-½ * \left(\left(Y_i - ß_0 - ß_1 Y_{i-1}\right)/\sigma \right)^2 \right)$$

The two equations above provide us with a way to generate a distribution for each point of the output as if they were an unique normal distribution. This works because the input is known to be a Gaussian normal distribution with zero mean and unit variance. This equation provides the likelihood of the current value, but ideally the log of this equation should be taken to make for easier partial derivatives of the unknown terms.

$$\log(L) = -n/2\log((2^*\text{pi}) -n/2\log(\sigma^2) -1/(2\ \sigma^2)\ \Sigma(Y_i - ß_0 - ß_1 Y_{i-1})^2$$

The partial derivatives with respect to $ß_0$, $ß_1$, and σ when set to 0 for maximization result in the following three equations.

$$\Sigma Y_i = nß_0 + ß_1\ \Sigma Y_{i-1}$$

$$\Sigma Y_i Y_{i-1} = ß_0\ \Sigma Y_i + ß_1\ \Sigma(Y_{i-1})^2$$

$$n/\ \sigma^2 = 1/\ \sigma^4\ \Sigma(Y_i - ß_0 - ß_1 Y_{i-1})^2$$

Solving for the $ß_0$ and $ß_1$ result in terms that are identical to those found with the Bayesian approach earlier in the problem.

$$ß_1 = R(1)/R(0)$$

$$ß_0 = x[n]$$