QUESTION # 1

DESCRIPTION OF PROBLEM AND DEVELOPMENT OF ALGORITHM:

This problem is the continuation of Question # 2 from Assignment # 4 and we implement Likelihood Ratio Test to evaluate the probability of hypothesis H_o that $\tau = \tau_o$ = 1sec being true given we perform 100 experiments. We have developed two separate pieces of code, QUESTION1.cpp and QUESTION1_DISTRIBUTION.cpp; former code simulates Monte Carlo Data for 200 events per experiment, uses them to calculate log(Likelihood) functions for Null hypothesis H_o and Alternate Hypothesis H_1 for 100 experiments and later code histograms the produced data, compares it with χ^2 distribution with one degree of freedom, calculates the rejection region and finally reports number of experiments that fail likelihood ratio test for H_o.

$$H_o : \tau = \tau_o, 0 < \sigma_t < \infty$$
$$H_1 : 0 < \tau < \infty, 0 < \sigma_t < \infty$$

GENERATION OF 200 EVENTS USING INVERSION TECHNIQUE AND GAUSSIAN SMEARING:

The expected pdf is a convolution of exponential decay and Gaussian resolution. The evaluated convolution integral was provided to us.

$$f(t;\tau,\sigma_t) = \frac{1}{2\tau}\exp\left(\frac{\sigma_t^2}{2\tau^2} - \frac{t}{\tau}\right) erfc\left(\frac{\sigma_t}{\sqrt{2}\tau} - \frac{t}{\sqrt{2}\sigma_t}\right)$$

We note that this function cannot be analytically integrated and therefore, we adopt sample importance method to generate 200 events. We declare an approximate function as followed:

$$f^{approximate} = \frac{1}{2\tau}\exp\left(\frac{\sigma_t^2}{2\tau^2} - \frac{t}{\tau}\right) \Rightarrow f = w \times f^{approximate}$$

And we can now adopt Inversion Technique on approximate function. We note that it is a decaying exponential function and therefore select [-5.0sec, 100sec] as our domain, realizing that approximate function falls to zero outside this interval. Mathematics of Inversion Technique described in following few steps.

$$\lambda = \frac{\int_{-5}^{t} \frac{1}{2\tau}\exp\left(\frac{\sigma_t^2}{2\tau^2} - \frac{t}{\tau}\right)dt}{\int_{-5}^{100} \frac{1}{2\tau}\exp\left(\frac{\sigma_t^2}{2\tau^2} - \frac{t}{\tau}\right)dt}$$

Where λ is a random number generated using TRandom3 RNG. Solving for t yields:

$$t = -\tau \ln\left[\lambda\left\{\exp\left(\frac{-100}{\tau}\right) - \exp\left(\frac{5}{\tau}\right)\right\} + \exp\left(\frac{5}{\tau}\right)\right]$$

A user-defined function Solve_For_t() included in QUESTION1.cpp solves above equation when this function is called in main program within do-while() loop. We realize that weight w is the complimentary error function provided in pdf. The *erfc()* provided in TMath library makes our life easier:

$$w = erfc\left(\frac{\sigma_t}{\sqrt{2}\tau} - \frac{t}{\sqrt{2}\sigma_t}\right)$$

And hence, we select a particular t if w evaluated at that t is greater than product of a new random number λ_2 and maximum value of w.

EVALUATION OF LOG(LIKELIHOOD) FUNCTIONS FOR H_o AND H_1:

The Log(Likelihood) function is evaluated for H_o and H_1 using User-Defined-Functions LOG_Likelihood_Function_H0() and LOG_Likelihood_Function_H1(). The governing equation is identically the same in both functions with only difference that we vary only σ_t from 0.0 to 10.0 with an increment of 0.01 for H_o to produce 1000 values of ln(L_H$_o$) but in case of H_1 we vary σ_t from 0.0 to 10.0 with an increment of 0.01 and for each value of σ_t, we vary τ from 0.0 to 10.0 with an increment of 0.01 to produce 1000x1000 values of ln(L_H$_1$).

$$\sum_{i=1}^{200} \ln f(t_i;\tau,\sigma_t) = \sum_{i=1}^{200}\left[\ln\left(\frac{1}{2\tau}\right) + \left(\frac{\sigma_t^2}{2\tau^2} - \frac{t_i}{\tau}\right) + \ln\left(erfc\left(\frac{\sigma_t}{\sqrt{2}\tau} - \frac{t_i}{\sqrt{2}\sigma_t}\right)\right)\right]$$

We declare maximum among 1000 values of ln(L_H$_o$) as $\ln L(\hat{\omega})$ and maximum among 1000x1000 values of ln(L_H$_1$) as $\ln L(\hat{\Omega})$ for each experiment assuming that we chose very fine meshing thus rectifying need of interpolation between points and hence evaluate $\ln\lambda$ using following equation:

$$\ln\lambda = \ln L(\hat{\omega}) - \ln L(\hat{\Omega})$$

QUESTION1.cpp writes 100 values of $\ln\lambda$ for 100 experiments to QUESTION1.txt in C++ Array Syntax Format. We copy these values from QUESTION.txt and paste them to QUESTION1_DISTRIBUTION.cpp for further analysis. We couldn't find out a better way out to read in these values in ROOT session perhaps because of apparent bug.

-2lnλ DISTRIBUTION AND ITS COMPARISON WITH χ^2 DISTRIBUTION:

QUESTION1_DISTRIBUTION.cpp shows distribution of $-2\ln\lambda$ and compares it with χ^2 distribution with one degree of freedom using user-defined TF1 CHI_SQUARE_FIT(). The governing equation for CHI_SQUARE_FIT() is given below:

$$f(\chi^2;1) = P\frac{\exp\left(\frac{-\chi^2}{2}\right)}{\sqrt{\chi^2}}$$

Where P=3.99532 is the scaling factor and is calculated iteratively by ROOT's curve fitting module to best fit our histogrammed data.

CUMULATIVE χ^2 DISTRIBUTION:

We integrate above equation for χ^2 distribution with one degree of freedom from 0 to ∞ to find analytical expression for cumulative χ^2 distribution.

$$\int_0^x f(\chi^2;1)d\chi^2 = \int_0^x P\frac{\exp\left(\frac{-\chi^2}{2}\right)}{\sqrt{\chi^2}}d\chi^2$$

$$\Rightarrow \int_0^x f(\chi^2;1)d\chi^2 = \sqrt{2}P\gamma\left(\frac{1}{2},\frac{x}{2}\right)$$

Where $\gamma\left(\frac{1}{2},\frac{x}{2}\right)$ is Lower Incomplete Gamma Function. It can be represented by Error Function as followed:

$$\gamma\left(\frac{1}{2},\frac{x}{2}\right) = \sqrt{\pi}erf\left(\sqrt{\frac{x}{2}}\right)$$

And hence, cumulative χ^2 distribution can be finally written as:

$$\int_0^x f(\chi^2;1)d\chi^2 = P\sqrt{2\pi}erf\left(\sqrt{\frac{x}{2}}\right)$$

We declare a user-defined function CUMULATIVE_CHI_SQUARE() and make use of above equation to show cumulative χ^2 distribution in Figure1.1.

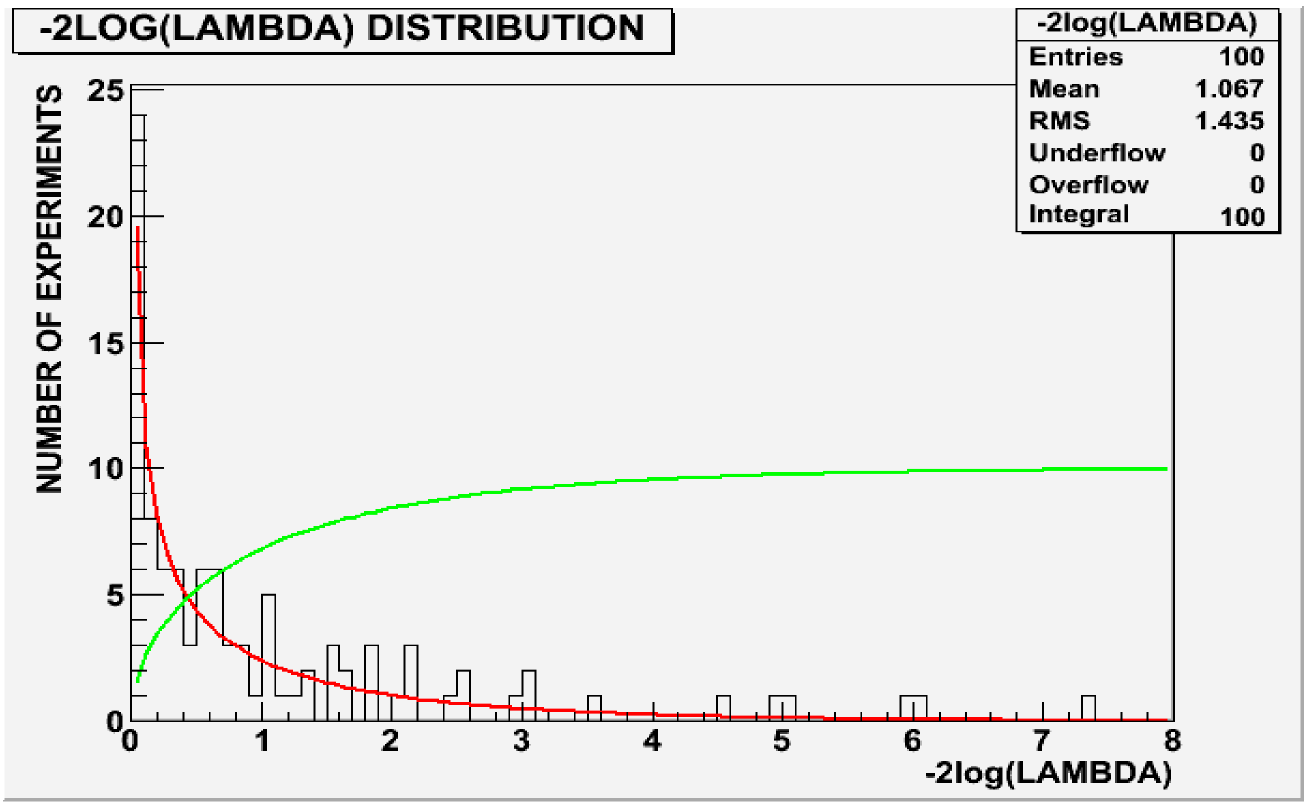

FIGURE 1.1: Histogram showing distribution of -2lnλ for 100 experiments. Red curve is the χ^2 distribution for comparison and green curve represents corresponding cumulative χ^2 distribution.

10% REJECTION REGION:

We note that our bin size is 0.1 and we collected data for 100 experiments. This means that one event in our histogram represents 0.1 $unit^2$ area and 100 events represent a total area of 10.0 $unit^2$. To define 90% confidence region and hence 10% rejection region, we need to find such value of $x=-2\ln\lambda$ for which value of green curve starting from 0.0 equals 9.0. We write a small block of instructions which evaluate functional value for TF1 CUMULATIVE_CHI_SQUARE() and hence gives us rejection region.

$$\text{For 10\% Rejection Region: } -2\ln\left(\lambda_{cut}\right) > 2.68452$$

And hence:

$$\lambda_{cut}=0.261255$$

Which means $\lambda<0.261255$ for any particular trial which fails Likelihood Ratio Test based on 10% Rejection Region.

NUMBER OF EXPERIMENTS FAILING LIKELIHOOD RATIO TEST:

Based on evaluated value of $-2\ln(\lambda_{cut})$, we count number of trial experiments that occur on right side of this particular point. In this specific set of 100 experiments, we find that 12 trials happen to fail Likelihood Ratio Test. These 12 trials lie from BIN # 26 to BIN # 80.

This concludes short description of problem, development of algorithm and presentation of results.