IND522
Advanced Statistical Modelling

Fall 2025

Professor Orhan FEYZİOĞLU

Assignment ♯3: Due October 10

1) Write a function using MATLAB's functions for numerical integration such as `integral` that will find $P(X \leq x)$ when the random variable is exponentially distributed with parameter λ. See help for information on how to use these functions.

2) When a random variable is equally likely to be either positive or negative, then the Laplacian or the double exponential distribution can be used to model it. The Laplacian probability density function for $\lambda > 0$ is given by

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}; \qquad -\infty < x < \infty$$

- **a.** Derive the cumulative distribution function for the Laplacian.
- **b.** Write a MATLAB function that will evaluate the Laplacian probability density function for given values in the domain.
- **c.** Write a MATLAB function that will evaluate the Laplacian cumulative distribution function.
- **d.** Plot the probability density function when $\lambda = 1$.

3) Suppose X follows the exponential distribution with parameter λ. Show that for $s \geq 0$ and $t \geq 0$,

$$P(X > s + t \mid X > s) = P(X > t).$$

4) The time to failure for a widget follows a Weibull distribution, with $\nu = 0$, $\beta = 1/2$ and $\alpha = 750$ hours.

- **a.** What is the mean time to failure of the widget?
- **b.** What percentage of the widgets will fail by 2500 hours of operation? That is, what is the probability that a widget will fail within 2500 hours?

5) Using the functions `fminbnd` (available in the standard MATLAB package), find the value for x where the maximum of the $N(3, 1)$ probability density occurs. Note that you have to find the minimum of $-f(x)$ to find the maximum of $f(x)$ using these functions. Refer to the `help` files on these functions for more information on how to use them.

6) Generate 500 random samples from the standard normal distribution for sample sizes of $n = 2$, 15, and 45. At each sample size, calculate the sample mean for all 500 samples. How are the means distributed as n gets large? Look at a histogram of the sample means to help answer this question. What is the mean and variance of the sample means for each n? Is this what you would expect from the Central Limit Theorem? Here is some MATLAB code to get you started.

For each n:

```
% Generate 500 random samples of size n:
x = randn(n, 500);
% Get the mean of each sample:
xbar = mean(x);
% Do a histogram with superimposed normal density.
% This function is in the MATLAB Statistics Toolbox.
% If you do not have this, then just use the
% function hist instead of histfit.
histfit(xbar);
```

7) Generate a random sample that is uniformly distributed over the interval $(0, 1)$. Plot the empirical distribution function over the interval $(-0.5, 1.5)$. There is also a function in the Statistics Toolbox called `cdfplot` that will do this.

8) Generate a random sample of size 100 from a normal distribution with mean 10 and variance of 2. Use the following:

```
randn(1,100)*sqrt(2)+10
```

Plot the empirical cumulative distribution function. What is the value of the empirical distribution function evaluated at a point less than the smallest observation in your random sample? What is the value of the empirical cumulative distribution function evaluated at a point that is greater than the largest observation in your random sample?

9) Another measure of skewness, called the *quartile coefficient of skewness*, for a sample is given by

$$\hat{\gamma}_{1_q} = \frac{\hat{q}_{0.75} - 2\hat{q}_{0.5} + \hat{q}_{0.25}}{\hat{q}_{0.75} - \hat{q}_{0.25}}$$

Write a MATLAB function that returns this statistic.

10) Investigate the bias in the maximum likelihood estimate of the variance that is given by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 . \tag{1}$$

Generate a random sample from the standard normal distribution. You can use the `randn` function that is available in the standard MATLAB package. Calculate $\hat{\sigma}^2$ using Equation (1) and record the value in a vector. Repeat this process (generate a random sample from the standard normal distribution, estimate the variance, save the value) many times. Once you are done with this procedure, you should have many estimates for the variance. Take the mean of these estimates to get an estimate of the expected value of $\hat{\sigma}^2$. How does this compare with the known value of $\hat{\sigma}^2 = 1$? Does this indicate that the maximum likelihood estimate for the variance is biased? What is the estimated bias from this procedure?