# Lab Lecture 7

# Learning objectives

- Pseudo-random numbers
- Matlab functions for generating (pseudo-random) trial values from some standard univariate PDFs
- Draw trial values from PDFs with user-specified parameters
- Plotting histograms to estimate PDFs

# Pseudo-random numbers

- Pseudo-random numbers: When you generate a random number on a computer, it is actually following a deterministic algorithm to produce numbers that look random
- Pseudo-random number generators (PRNGs) require a starting seed value: when the seed is reset, the PRNG returns the same sequence of random numbers as before
- Homework reading: Find an online source to understand how random numbers are generated on a computer
- Example: [https://en.wikipedia.org/wiki/Pseudorandom_number_generator](https://en.wikipedia.org/wiki/Pseudorandom_number_generator)

- Run the following code:

```
rng('default') %Default seed
rand() % One trial value
rand() % Another trial value
disp('----reset seed-----')
rng('default')
rand() % One trial value
rand() % Another trial value
```

- You see that the sequence of trial values repeats after the seed is reset

# Standard PDFs

- The function `rand` draws a trial value from the uniform PDF: $U(x; 0,1)$
- The function `randn` draws a trial value from the Normal PDF: $N(x; 0,1)$
- To generate M trial values, use: `X=rand(1,M)` (or `X=rand(M,1)`)
  - Read the help for these functions

# Cumulative distribution function (cdf)

- $p_X(x)$ : probability density function (pdf)
- $F_X(a)$ : Cumulative distribution function (cdf)

$$F_X(a) = \int_{-\infty}^{a} p_X(x)dx = \Pr(X \le a)$$

$$\Pr(a \le X \le b) = \int_{a}^{b} p_X(x)dx = F_X(b) - F_X(a)$$

- The rigorous theory of probability for continuous random variables is based on defining the cdf first
- From the definition of $F_X(a)$, it follows that

$$\frac{d}{dx}F_X(x) = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{1}{\Delta x}\left(\int_{-\infty}^{x+\Delta x} p_X(x)dx - \int_{-\infty}^{x} p_X(x)dx\right)$$

$$= \lim_{\Delta x \to 0} \frac{1}{\Delta x}\left(\int_{x}^{x+\Delta x} p_X(x)dx\right) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} p_X(x)\Delta x = p_X(x)$$

# Generating trial value from $U(x; a, b)$

- Task: If $X$ has a PDF $U(x; 0,1)$, prove that the random variable

$$Y = (b - a)X + a$$

has a PDF that is $U(x; a, b)$

- Steps:

1. Compute the CDF of Y:

$$\mathrm{F_Y(x)} = \Pr(Y \le x) = \Pr\big((b-a)X + a \le x\big) = \Pr\left(X \le \frac{x-a}{b-a}\right) = \int_{-\infty}^{\frac{x-a}{b-a}} U(x; 0,1)dx$$

2. Obtain $p_Y(y) = \frac{d}{dy} F_Y(y)$

- Task: Use the above result to create your own function to generate trial values from $U(x; a, b)$ [Call the function ‘`customrand(a,b)`’]

# Histogram

- Given a set of $N_{trials}$ trial values of a random variable $X$ having a PDF $p_X(x)$, the histogram is generated by dividing the range of the trial values into equal sized intervals (‘bins) and counting the number of trial values that fell in each bin
- If the number of trial values in bin $[x, x + \Delta x]$ is $N_x$, the ratio $\frac{N_x}{N_{trials}}$ gives us an estimate (frequentist view) of

$$\Pr(X \in [x, x + \Delta x]) = \int_x^{x+\Delta x} p_X(x)dx \approx p_X(x)\Delta x$$

for sufficiently small $\Delta x$

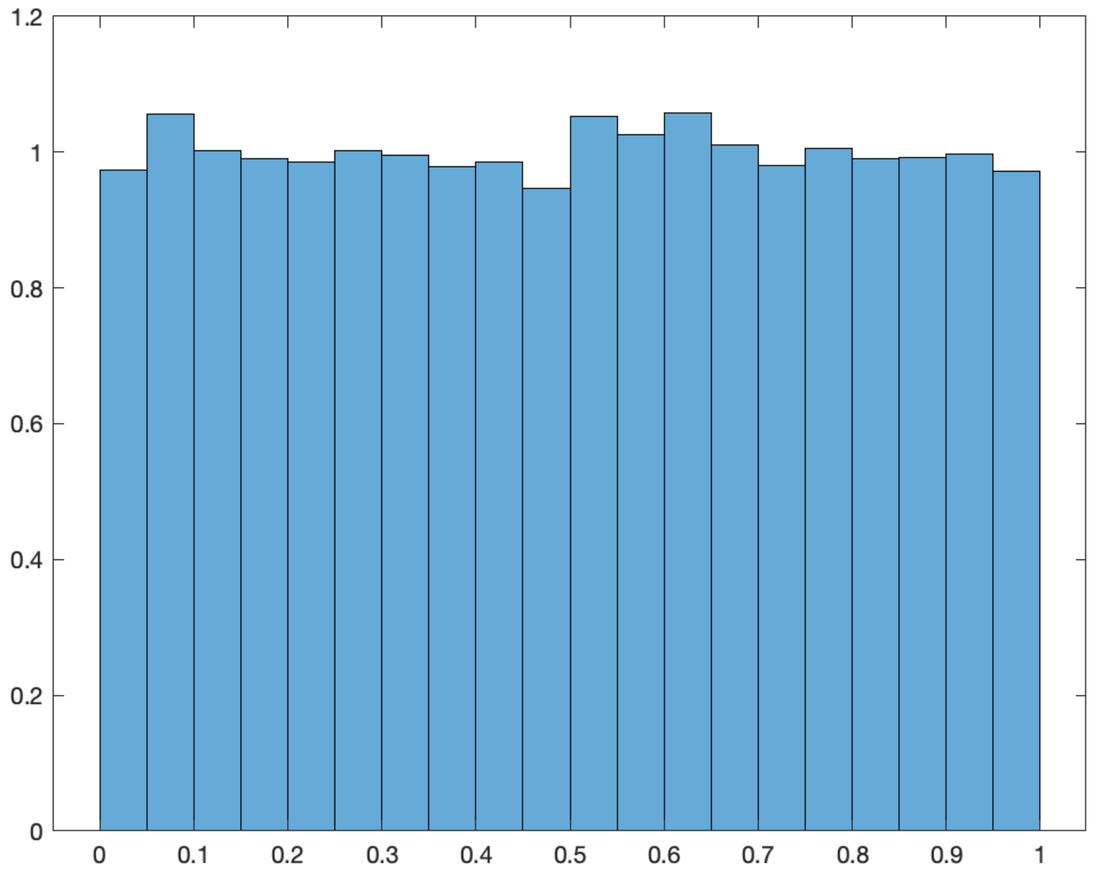
- Thus, dividing the counts in each bin of a histogram by the total number of trials and the bin size $\Delta x$, gives us an estimate of the PDF $p_X(x)$

# Histogram

- Run the code below

```
M=10000;
xVec = rand(1,M);
histogram(xVec,'normalization','pdf’)
```

- Read the help for `histogram` to learn how to use this function
- The option to normalize as pdf implements the normalization discussed earlier




Normalized histogram of trial values drawn from $U(x;0,1)$

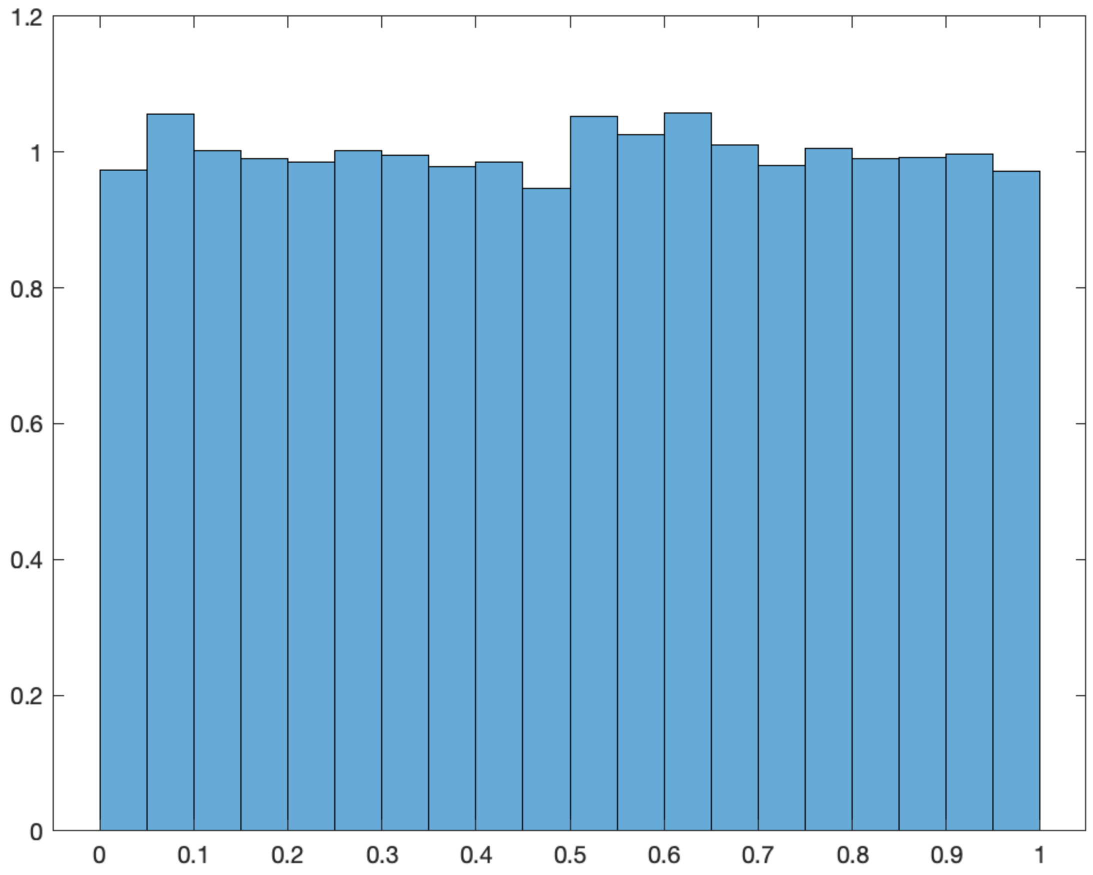
Food for thought: Is it acceptable that the estimated PDF exceeds the theoretical value of 1 for $U(x;0,1)$?

# Histogram

- Run the code below

```
M=10000;
xVec = rand(1,M);
histogram(xVec,'normalization','pdf')
```

- Read the help for `histogram` to learn how to use this function
- The option to normalize as pdf implements the normalization discussed earlier
- What happens to the estimated PDF when $M = 100{,}000$ and $1{,}000{,}000$ ?




Normalized histogram of trial values drawn from $U(x;0,1)$

Food for thought: Is it acceptable that the estimated PDF exceeds the theoretical value of 1 for $U(x;0,1)$?

# Generating trial values from $N(x; \mu, \sigma)$

- If $X$ has a PDF $N(x; 0,1)$, the random variable

$$Y = \sigma X + \mu$$

has a PDF $N(x; \mu, \sigma)$

- Exercise: Prove it!
- Task: Use the above result to create your own function to generate trial values from $N(x; a, b)$ [Call the function '`customrandn(a,b)`']

# Lab tasks

- Write a test script, `testcustomprng.m`, that:

1. Uses `customrand` and `customrandn` functions to generate 10,000 trials values each from
    1. $U(x; -2,1)$, and
    2. $N(x; 1.5,2.0)$
2. Make normalized histograms for each case
3. On top of each histogram, plot the respective PDFs above
4. Submit the codes and the plots