Lecture 2 Some Popular Continuous Random Variables

In the last lecture we addressed a number of popular discrete random variables. In these notes we will address sompe popular continuous random variables. Recall how these types of random variables are defined.

Definition 1 Let X denote a random variable with sample space S_X. If the number of elements in S_X is finite or countably infinite, then X is said to be a ***discrete random variable***. If S_X is a continuum on the real line, then X is said to be a ***continuous random variable***.

The Uniform Random Variable- A random variable X with $S_X = [a,b]$, $[a,b)$, $(a,b]$ *or* (a,b) and with probability density function (*pdf*) $f_X(x) = 1/(b-a)$ is said to be a ***uniform random variable***.
[Related *Matlab* commands are: unifpdf, unifcdf, unifrnd]
[https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)]

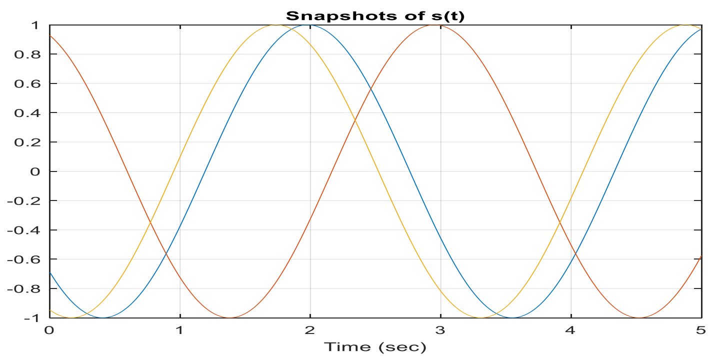
Example 1 Calibration of many instruments and modeling of many signals entail the use of sinusoids. Recall that a sinusoid signal has the form $s(t) = A\sin(\omega t + \theta)$. In particular, $s(0) = A\sin(\theta)$. Hence, the amplitude at $t = 0$ depends on the *phase* variable θ. By taking repeated snapshots of this sinusoid at randomly spaced intervals (e.g. using an oscilloscope in the free-run mode), one can assume that Θ = *the act of recording the phase* is a random variable that has a *uniform* distribution over $S_\Theta = [-\pi, \pi)$. The *pdf* is, therefore, $f_\Theta(\theta) = 1/2\pi$ *for* $-\pi \le \theta < \pi$.

(a) Compute the corresponding cumulative distribution function (*cdf*). $\Pr[\Theta \le \theta] \overset{\Delta}{=} F_\Theta(\theta) = \int_{-\pi}^{\theta} f_\Theta(\phi)\, d\phi = \int_{-\pi}^{\theta} \frac{1}{2\pi}\, d\phi = \frac{\theta + \pi}{2\pi}$.

(b) Compute $\Pr[-0.1 < \Theta < 0.1]$. $\Pr[-0.1 < \Theta < 0.1]$ = unifcdf(0.1,-pi,pi)-unifcdf(-0.1,-pi,pi) = 0.0318.

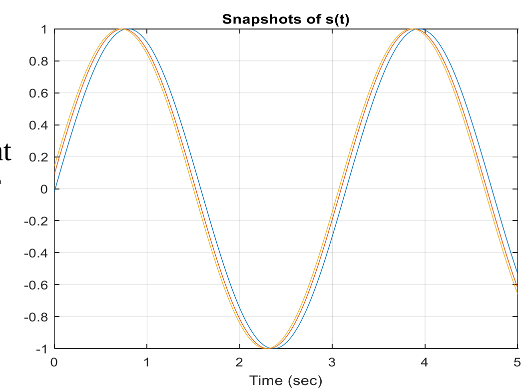
(c) Simulate 3 snapshots of $s(t) = \sin(2t + \Theta)$ over a 5-second window.

```
%Example 1
t=0:.001:5;
nt=length(t);
nsim=3;
th=unifrnd(-pi,pi,nsim,1);
s=zeros(nsim,nt);
for k=1:nsim
    s(k,:)=sin(2*t + th(k));
end
plot(t,s)
title('Snapshots of s(t)')
xlabel('Time (sec)')
grid
```



(d) Suppose that the scope is switched from the free-run mode to the trigger mode, and that the trigger is set to take a snapshot when $s(t)$ crosses zero with a positive slope. There will always be a slight amount of trigger 'jitter'. Assume that $\Theta \sim Uniform(-0.2, 0.2)$. Repeat (c) for this case.

The only needed code change is: `th=unifrnd(-.2,.2,nsim,1);`



The Exponential Random Variable- The *pdf* is $f_X(x) = \lambda e^{-\lambda x}$ $with\, S_X = [0, \infty)$.

[Related *Matlab* commands are: exppdf, expcdf, exprnd]
[https://en.wikipedia.org/wiki/Exponential_distribution]

(a) Compute the corresponding cumulative distribution function (*cdf*). $F_X(x) = \int_0^x \lambda e^{-\lambda u}\, du = 1 - e^{-\lambda x}$.

(b)*Given* the event $[X > x_0]$, repeat (a).

$$F_{X|X>x_0}(x) = \Pr[X \le x \mid X > x_0] = \frac{\Pr[X \le x \cap X > x_0]}{\Pr[X > x_0]} = \frac{\Pr[x_0 < X \le x]}{\Pr[X > x_0]} =$$

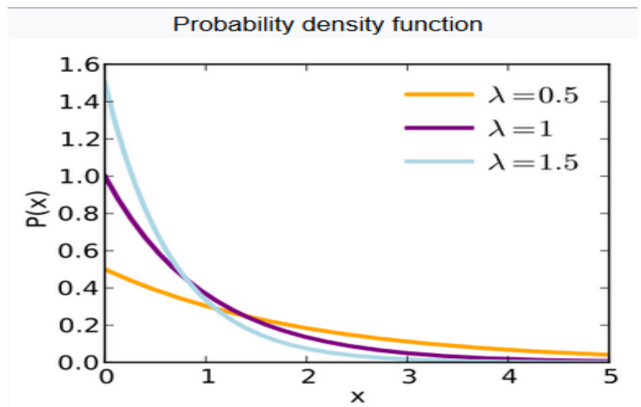
$$\frac{F_X(x) - F_X(x_0)}{1 - F_X(x_0)} = \frac{(1-e^{-\lambda x}) - (1-e^{-\lambda x_0})}{1-(1-e^{-\lambda x_0})} = \frac{e^{-\lambda x_0} - e^{-\lambda x}}{e^{-\lambda x_0}} = 1 - e^{-\lambda(x-x_0)} = F_X(x - x_0)$$

Notice that this *cdf* is <u>exactly</u> the same as that in (a), except that now the sample space is $S_X = [x_0, \infty)$. Because of this, the exponential distribution is said to be *memoryless* (e.g. Given that a part has survived an amount of time x_0, its failure probability is exactly the same as it was when it was new).

(c) From https://en.wikipedia.org/wiki/Exponential_distribution copy/paste the following: (i) μ_X, (ii) σ_X^2, and (iii) plots of the *pdf* for various values of λ.

Mean	λ^{-1}

Variance	λ^{-2}



The Normal Random Variable- The *pdf* is $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ $with\, S_X = (-\infty, \infty)$.

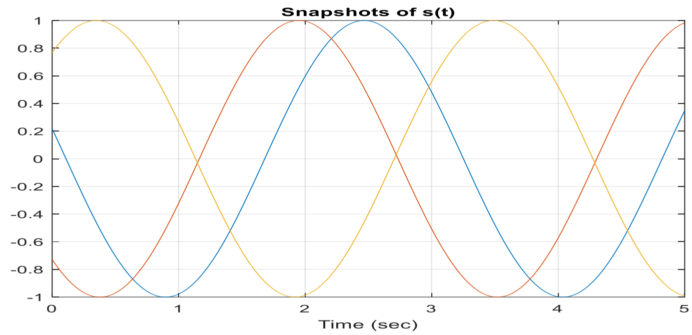
[Related *Matlab* commands are: normpdf, normcdf, normrnd]
[https://en.wikipedia.org/wiki/Normal_distribution]

Example 1 continued. Many periodic signals can be modeled as a sum of sinusoids. Suppose that for a chosen sinusoid we have $\Theta \sim N(\mu = 45^o, \sigma = 3^o)$. Note the now our units are *degrees.*

(e) Compute $\Pr[40^o < \Theta < 50^o]$. $\Pr[40^o < \Theta < 50^o]$ = normcdf(50,45,3) - normcdf(40,45,3) = 0.9044

(f) Repeat (c) for this case.

The needed code change is: `th=normrnd(45,3,nsim,1);`



Example 2 Suppose that when a lathe cutting tool is good the diameter of any turned shaft is $D_g \sim N(\mu = 2, \sigma = .005)$, and that when the tool is bad it is $D_b \sim N(\mu = 2.01, \sigma = .005)$.

(a) For each condition compute $\Pr[1.99 < D < 2.01]$.

$\Pr[1.99 < D_g < 2.01] =$ normcdf(2.01,2,.005)-normcdf(1.99,2,.005) = 0.9545

$\Pr[1.99 < D_b < 2.01] =$ normcdf(2.01,2.01,.005)-normcdf(1.99,2.01,.005) = 0.5000

(b) Suppose that any shaft with diameter not in the range $[1.99 < d < 2.01]$ is considered as waste, and that the expense associated with each shaft is $20. Compute the total waste cost for every 1000 shafts.

For a good cutting tool, we can expect that ~955 will be good and 45 will be waste. So the cost is ~$900.

For a bad cutting tool, we can expect that ~500 will be good and 500 will be waste. So the cost is ~$10,000.

(c) To reduce the cost associated with a bad cutting tool, you have incorporated a protocol that is: Whenever a shaft diameter exceeds 2.015 the cutting tool shall be replaced. Compute the probability that you will erroneously replace a good cutting tool.

$\Pr[D_g > 2.015]$ = 1-normcdf(2.015,2,.005) = 0.0013

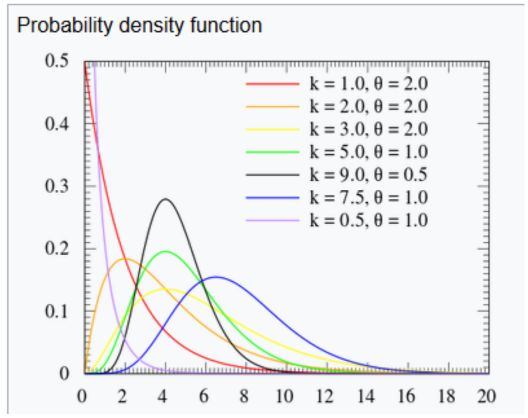
(d) Compute the probability that you will not replace a bad cutting tool.

$\Pr[D_b < 2.015]$ = normcdf(2.015,2.01,.005) = 0.8413

(e) In view of (c-d), comment on the effectiveness of the protocol.

The Gamma Random Variable- The *pdf* is too ugly to give here ☹

[Related *Matlab* commands are: gampdf, gamcdf, gamrnd]
[https://en.wikipedia.org/wiki/Gamma_distribution]



(a) From https://en.wikipedia.org/wiki/Gamma_distribution copy/paste the following: (i) μ_X, (ii) σ_X^2, and (iii) plots of the *pdf* for various values of the *shape* parameter *k*, and the *scale* parameter θ.

Mean	$\mathrm{E}[X] = k\theta$	Variance	$\mathrm{Var}[X] = k\theta^2$

(b) From the plots in (a) identify the one that most resembles a bell shape (i.e. normal) *pdf*. BLUE $k = 7.5 ; \theta = 1$

(c) Compute expressions for *k* and θ as functions of μ_X and σ_X^2. $\theta = \frac{\sigma_X^2}{\mu_X}$, hence $k = \frac{\mu_X^2}{\sigma_X^2}$

Example 3 A collection of 100 surface roughness measurements yielded the estimates $\hat{\mu}_X = 48.89 ; \hat{\sigma}_X = 14.85$. The skewed nature of the histogram suggested that *X* could be modeled by a *gamma* distribution.

(a) Compute the values of the *shape* parameter *k*, and the *scale* parameter θ.

$\theta = \frac{\sigma_X^2}{\mu_X}$ = 14.85^2/48.89 = 4.5106 ; $k = \frac{\mu_X^2}{\sigma_X^2}$ = (48.89/14.85)^2 = 10.8389

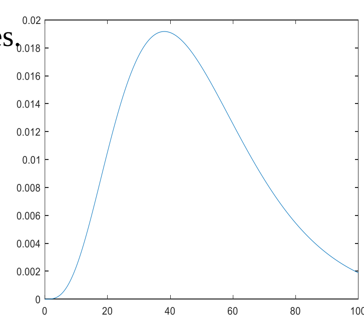
(b) Plot the *pdf* over an appropriate range of *x*-values.

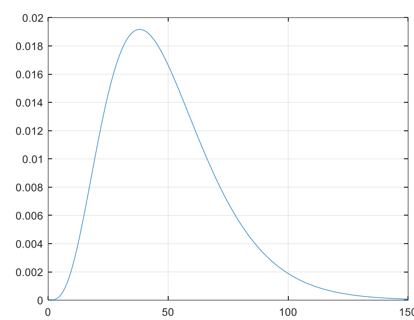
Since $\hat{\mu}_X + 3\hat{\mu}_X = 48.89 + 3(14.85) \cong 94$,

I will try a plot over 0-100.

```
>> x=0:.01:100;
>> f=gampdf(x,4.516,10.8389);
>> plot(x,f)
```

Not so good ☹. So try 0-150. *Sweet* ☺





(c) Unlike yourself, your colleague did not look at the data histogram. He simply assumed that *X* was *normally* distributed. Compute $\Pr[X > \hat{\mu}_X + 3\hat{\sigma}_X]$ for his model, and compare it to the value given by yours.

HIS: 1-normcdf(48.89+3*14.85,48.89,14.85) = 0.0013 ; MINE: 1-gamcdf(48.89+3*14.85,10.8389,4.5106) = 0.0065

My probability is **5** times greater than his.