IT 50700 Measurement and Evaluation of Industry and Technology

Lecture Note (sections 4-7 to 4-12)

Chapter 4 Continuous Random Variables and Probability Distributions

Based on the text book: Applied Statistics and Probability for Engineers, 6th Ed, by D. C. Montegomery and G. C. Runger, published by Wiley

References

- Intro to Probability, Statistics and Random Process, by A. Rankhshan and H. Pishro-Nik, https://www.probabilitycourse.com/
    -
    - Ch 12. Introduction to Probability Using MATLAB, by A. Rankhshan and H. Pishro-Nik, https://www.probabilitycourse.com/chapter12/Chapter_12.pdf
    - Ch 13. Intro to Simulation Using R, https://www.probabilitycourse.com/chapter13/chapter13.php
    - Ch 14. Recursive Methods, https://www.probabilitycourse.com/chapter14/chapter14.php
- Exponential Distribution, http://www.unc.edu/~jbhill/Sample_Matlab_Code_770.pdf
- M-Lab 6: Densities of Random Variables, http://www4.stat.ncsu.edu/~boos/mlab/mlab6.html
- Exponential and Poisson Distribution, Poisson Process, http://www.winlab.rutgers.edu/~zhibinwu/html/Exponentail_and_Poisson.html

Example 3—31. Calculations for Wire Flaws.

- Thin copper wire
- The no of flaws follows a Poisson distribution with a means of 2.3 Flaws per mm.
- (A) Find the probability of exactly two flaws in 1 mm of wire
- (B) Find the probability of 10 flws in 5 mm of wire.
- (C) Determine the probability of at least one flaw in 2 mm of wire.

Sol:

(A)

f = λT = 2.3

P(X = 2) = $\frac{e^{-\lambda T}(\lambda T)^x}{x!} = \frac{e^{-2.3}(2.3)^2}{2!} = \frac{0.1*5.29}{2} = 0.265$

Excel => POSSION(2, 2.3, FALSE) = 0.26518

MATLAB

```
function f = f_poisson_dist(x, f)
% f_poisson_dist.m
%
% lambda - mean number of events per unit length
% T - a lenght of T millimeter, millisecond, etc
% f = lambda*T (number of events; frequency, etc)
% x = Test frequency
% log() = Natual Log (LN)
x_fac = factorial(x);
b = exp(-f)*(f^x);
```

```
f = b/x_fac;
```

Call f_poisson_dist

```
>> f = 2.3; x = 2;
>> PX = f_poisson_dist(x, f)
PX =
  265.1846e-003
```

(B)

f = λT = 2.3 flaws/mm x 5 mm = 11.5 flaws

x = 10

P(X = 10) = $\frac{e^{-\lambda T}(\lambda T)^x}{x!} = \frac{e^{-11.5}(11.5)^{10}}{10!} = \frac{10.13e{-6}*40.455e9}{3.6288e{+6}} = 0.113$

MATLAB Solution

```
>> f = 11.5; x = 10;
>> PX = f_poisson_dist(x, f)
PX =
  112.9351e-003
```

(C)

f = λT = 2.3 flaws/mm x 2 mm = 4.6 flaws

x = 0

P(X ≥ 1) = 1 – P(X =0) = 1 – $e^{-4.6}$ = 0.9899

```
>> f = 4.6; x = 0;
>> PX = f_poisson_dist(x, f)
>> 1- f_poisson_dist(x, f)
ans =
  989.9482e-003
```

Example 3-32. Magnetic Storage and Contamination

- Assume that the number of particle of contamination that occur on a disk surface has a Poisson distribution.
- The average no of particles per square cm of media surface is 0.1.
- The area of a disk under study is 100 square cm.
- (A) Determine the probability that 12 particles occur in the area of a disk under study.
- (B) Find the probability that zero particle occur in the area of the disk under study
- (C) Determine the probability that 12 or fewer particles occur in the area of the disk under study.

(A)

λ = 0.1, T = 100 cm^2

λT = 0.1(100) = 10 particles

x = 12

P(X = 12) = $\frac{e^{-\lambda T}(\lambda T)^x}{x!} = \frac{e^{-105}(10)^{12}}{12!}$ = 0.095

MATLAB Solution
>> f = 10, x = 12;
>> PX = f_poisson_dist(x, f)
PX = 94.7803e-003

(B)
$P(X =0) = e^{-10} = 4.54x10^{-5}$

MATLAB Solution
x = 0; f = 10;
>> PX = f_poisson_dist(x, f)
PX = 45.3999e-006

(C)

$P(X \leq 12) = P(X = 0) + P(X =1) +\ldots + P(X = 12)$

$= \sum_{x=0}^{12} \frac{e^{-10}(10)^x}{x!} = 0.792$

MATLAB Solution

```
%ex3_32.m
x = 12;
lambda = 0.1;  % 0.1 partciles per cm
T = 100;       % Area 100 cm^2
f = lambda * T;
Psum = 0.0;

for R = 1:x
        Psum = Psum + f_poisson_dist(R, f);
end
Psum

>> ex3_31
Psum =
   791.5111e-003
```

4-7 Normal Approximation to the Binomial and Poisson Distributions

Binomial Distribution:
μ = E(X) = np
σ2 = V(X) = np(1-p)
$\sigma = \sqrt{np(1-np)}$

Normal Approximation to the Binomial Distribution
Note: from page 122, $Z = \frac{X-\mu}{\sigma}$

If X is a binomial RV with parameter n and p,

$$Z = \frac{X-np}{\sqrt{np(1-p)}}$$

is approximately a standard normal RV. To approximate a binomial probability with a normal distribution, a continuity correction is applied as follows:

$$P(X \le x) = P(X \le x + 0.5) = \text{P}\left(Z \le \frac{x+0.5-np}{\sqrt{np(1-p)}}\right)$$

and

$$P(x \le X) = P(x - 0.5 \le X) = \text{P}\left(\frac{x-0.5-np}{\sqrt{np(1-p)}} \le Z\right)$$

The approximation is good for np > 5 and n(1-p) >5.

Normal Approximation to the Poisson Distribution
If X is a Poisson RV with E(X) = λ and V(X) = λ,

$$Z = \frac{X-\lambda}{\sqrt{\lambda}}$$

is approximately a standard normal RV. The same continuity correction used for the binomial distribution can also be applied. The approximation is good for λ > 5.

Example 4-20. Normal Approximation to Poisson

- The number of asbestos particles in a square meter of dust on a surface follows a Poisson distribution with a mean of 1000.
- (A) If a square meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

(A)
Due to the large λT = 1000, and x = 950, the computation complexity is clear.

$$\text{P}(X \le 950) = \sum_{x=0}^{950} \frac{e^{-1000}(1000)^x}{x!}$$

```
>> PX = f_poisson_dist(950, 1000)
PX =  NaN
```

The probability can be approximated as

P(X ≤ 950) = P(X ≤ 950.5) = P(Z ≤ (X – λ)/$\sqrt{\lambda}$) = P(Z ≤ (950.5 – 1000)/ $\sqrt{1000}) = P(-1.57) = 0.058$
(Table III, on page 742)

Other References

- POISSON-SIMULATION, https://people.sc.fsu.edu/~jburkardt/m_src/poisson_simulation/poisson_simulation.html
- Poisson Tutorial, http://www.hms.harvard.edu/bss/neuro/bornlab/nb204/statistics/poissonTutorial.txt
- MATLAB Geeks, by Eric Verner, June 16, 2014, http://matlabgeeks.com/tips-tutorials/random-numbers-in-matlab-part-3/
- http://mathworld.wolfram.com/ExponentialDistribution.html
- https://en.wikipedia.org/wiki/Exponential_distribution
- http://www.mast.queensu.ca/~stat455/lecturenotes/set4.pdf
- Keisan Online Calculator, http://keisan.casio.com/exec/system/1180573222
- Computational Statistics with Matlab, Mark Steyvers, May, 13, 2011, http://psiexp.ss.uci.edu/research/teachingP205C/205C.pdf
- http://pages.cs.wisc.edu/~dsmyers/cs547/lecture_8_continuous_random_variables.pdf

4-8 Exponential Distribution

Applications

- Network usage, Computer usage
- Computer usage – user log-on
- Reliability studies - a model for the time until failure of a device
  - For example, the life of a semiconductor chip might be modeled as an exponential RV with a mean of 40,000 hours.
  - The lack of memory property of the exponential distribution implies that the device does not wear out.
  - That is, regardless of how long the device has been operating, the probability of a failure in the next 1000 hours is the same as the probability of a failure in the first 1000 hrs of operation.
  - The lifetime L of a device with failures caused by random shocks might be appropriately modeled as an exponential RV.

The RV X that equals the distance between successive events form a Poisson process with mean number of events λ > 0 per unit interval is an exponential RV with parameter λ. The probability density function X is

$$f(x) = \lambda e^{\lambda x} \;\; for\; 0 \le x < \infty$$

Mean $\mu = E(X) = \frac{1}{\lambda}$

Variance $\sigma^2 = V(X) = \frac{1}{\lambda^2}$

Or, https://en.wikipedia.org/wiki/Exponential_distribution

$$f(x;\, \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

It can be verify that its total probability = 1:

$$\int_0^{\infty} \lambda e^{-\lambda x} dx = 1$$

Cumulative Distribution Function (CDF): F(x) = P(X ≤ x)

We can derive the cumulative distribution F(a) by integrating from 0 to a,

$$F(a) = P(X \leq a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}$$

Cumulative density function

$$F(x) = 1 - e^{-\lambda x}, 0 \ \leq x < \infty$$

Mean:

The expected value of the exponential RV is given by

$$E[X] = \int_0^\infty x \cdot (\lambda e^{-\lambda x}) dx$$

Using integration by parts ($u = x, and\ dv = e^{-\lambda x} dx$) to find the integral

E[X] = 1/λ

Generating Random Variables

- Cumulative Distribution Function (CDF): F(x) = P(X ≤ x), which is a number between 0 and 1.
- Generate a RV having the desired CDF using
  - Choose a random probability u between 0 and 1
  - Solve the unique value of x such that F(x) = u

    $F(x) = 1 - e^{-\lambda x} = u$

  Find

  $$x = \frac{\ln(1-u)}{-\lambda}$$

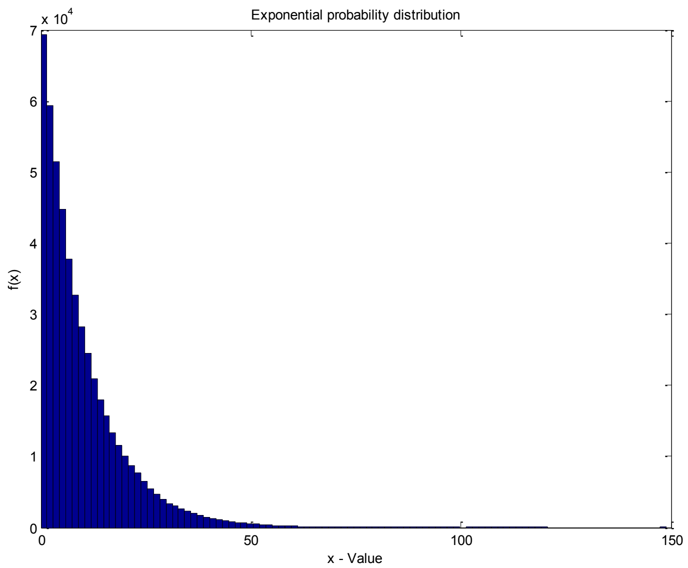
Natual Log – LN – MATLAB function log()

x = -log(1-u)/lambda

```
function f = f_exp_dist(x, lambda)
% lambda - mean number of events per unit length
%
% x = Test frequency
f = -log(x)/lambda;
```

```
% exp_dis_plot.m
lambda = 0.1;       %Rate parameter = 1/mean = 1/stdev
N = 500000;
nbins = 100;
X = rand(N,1);
% EXP = -log(X)/lambda;
```

```
EXP = f_exp_dist(X, lambda);
hist(EXP, nbins);
xlabel('x - Value'); ylabel('f(x)');
title('Exponential probability distribution')
```



Example 4-21. Computer Usage

- A large computer network, user log-on to the system can be modeled as a Poisson process with a mean λ = 25 log-ons per hour
- (A) What is the probability that there are zero log-ons (no logs-on) in an interval of 6 minutes?
- (B) What is the probability that the time until the next log-on is between 2 and 3 minutes? Upon converting all units to hours.
- (C) Determine the interval time such that the probability that no log-on occurs in the interval is 0.90.

(A)

- Let X denote the time in hours from the start of the interval until the first log-on

λ = 25 log-ons/hr

6 min = 0.1 hr

$$P(X > 0.1) = \int_{0.1}^{\infty} \lambda e^{-\lambda x} dx = \int_{0.1}^{\infty} 25e^{-25x}\, dx = e^{-25(0.1)} = 0.082$$

Or

$$F(x) = 1 - e^{-\lambda x}$$

$$P(X > 0.1) = 1 - F(x = 0.1) = 1 - \left(1 - e^{-25(0.1)}\right) = e^{-2.5}$$

MATLAB Solution:

>> lambda =25;

```
>> exp(-lambda*0.1)
ans =     82.0850e-003
```

(B) What is the probability that the time until the next log-on is between 2 and 3 minutes? Upon converting all units to hours.
2 min = 2/60 = 0.033 hr
3 min = 3/60 = 0.05 hr

$$P(0.033 < X < 0.05) = \int_{0.033}^{0.05} \lambda e^{-\lambda x} dx = \int_{0.033}^{0.05} 25e^{-25x}\, dx = -e^{-25x}\Big|_{0.033}^{0.05} = 0.152$$

Or from $F(x) = 1 - e^{-\lambda x}$

$$P(0.033 < X < 0.05) = F(0.05) - F(0.033) = 0.152$$

MATLAB Solution:

```
>> F1=1-exp(-lambda*0.05)
F1 =  713.4952e-003
>> F2 =1-exp(-lambda*0.033)
F2 =   561.7650e-003
>> F1-F2
ans =  151.7302e-003
```

(C) Determine the interval time such that the probability that no log-on occurs in the interval is 0.90.

- The question asks for the length of the time x such that P(X > x) = e^{-25x} = 0.9.
- Solve the unique value of x such that F(x) = u, where u = 0.9

Find

$$x = \frac{\ln(u)}{-\lambda}$$

```
>> u = 0.9;
>> x =log(u)/-lambda
x =    4.2144e-003
```

% convert Hour => min: 0.00421*60 = 0.25 min
The mean time until the next log-on = μ = 1/25 = 1/λ = 2.4 minutes
The standard deviation σ = 1/λ = 2.4 minutes

4-9 Erlang and Gamma Distributions

4-10 Weibull Distribution

4-11 Log-Normal Distribution

4-12 Beta Distribution