ISyE 3770
Chapter 4: Continuous Random Variables

Instructor: Dan Li
Slides by Prof. Nagi Gebraeel

School of Industrial and Systems Engineering
Georgia Tech

Spring 2021

1. Continuous Random Variables
   - pdf
   - cdf
   - Mean and Variance

2. Continuous Distributions
   - Continuous Uniform Distribution
   - Normal Distribution
   - Exponential Distribution
   - Erlang & Gamma Distribution
   - Weibull, Lognormal, & Beta Distributions
     - Weibull Distribution
     - Lognormal Distribution
     - Beta Distribution

Continuous Random Variables

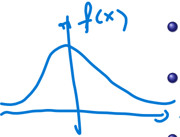
- In this section, we focus on the analysis of several common continuous random variables that frequently arise in applications.
- A continuous random variable is one which takes values in an uncountable set.
- They are used to measure physical characteristics such as height, weight, time, volume, position, etc...

Probability Density Function

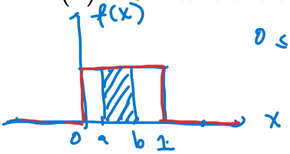
$$P(X \le a) = \int_{-\infty}^{a} f(x)\,dx = F(a)$$

$$P(X \ge a) = 1 - \int_{-\infty}^{a} f(x)\,dx = \int_{a}^{\infty} f(x)\,dx$$

- For a continuous random variable X, a probability density function (pdf) is a function such that



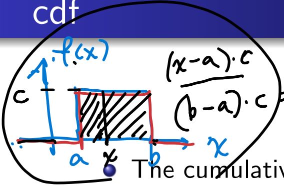
  - $f(x) \geq 0$ means that the function is always non-negative.
  - $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ $\quad P(X=x) = P(x \le X \le x) = \int_x^x f(x)\,dx = 0$
  - $P(a \leq X \leq b) = \int_a^b f(x)\mathrm{d}x =$ area under $f(x)$ from a to b
  - $f(x) = 0$ means there is no area exactly at x

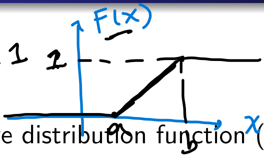


$0 \le X \le 1 \Rightarrow$

$x > 1$ or $X < 0$ $\quad f(x) = 0$

cdf

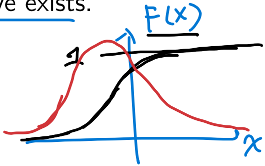




- The cumulative distribution function (cdf) of a continuous random variable X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\mathrm{d}u \text{ for } -\infty < x < \infty$$

- The cumulative distribution function is defined for all real numbers
- Given $F(x)$, $f(x) = \dfrac{dF(x)}{dx}$ as long as the derivative exists.



Mean and Variance

Suppose X is a continuous random variable with probability density function $f(x)$.

$\mu = \sum_{x \in S} x f(x)$ $\quad V(x) = \sum_{x \in S} (x-\mu)^2 f(x)$

- The mean or expected value of X, denoted as μ or $E(x)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

- The variance of X, denoted as $V(X)$ or σ^2, is $\quad V(x) = E(x^2) - \mu^2$

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \mathrm{d}x = \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x - \mu^2$$

$= E[(x-\mu)^2]$

- The standard deviation of X is $\sigma = \sqrt{\sigma^2}$
- The mean (expectation) of a function of X, $h(X)$, is $\quad E[h(x)] = \sum_{x \in S} h(x) f(x)$.

$h(x) = x^3$

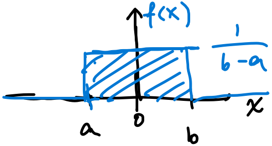
$= e^x$, ...

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) \mathrm{d}x$$

Continuous Uniform Distribution slide 14-15

- This is the simplest continuous distribution and analogous to its discrete counterpart.
- A continuous random variable X with probability density function.

$$f(x) = \frac{1}{(b-a)} \quad \text{for} \quad a \leq x \leq b$$



$F(x) = 0 \quad x < a$

$F(x) = 1 \quad x > b$

$F(x) = \int_{-\infty}^{x} f(x)\,dx \quad a \leq x \leq b$

$= \int_{a}^{x} f(x)\,dx = \int_{a}^{x} \frac{1}{b-a}\,dx = \frac{x-a}{b-a}$

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b \end{cases}$$

Example of Continuous Uniform Distribution *slide 16*

- The random variable X has a continuous uniform distribution on $[4.9, 5.1]$. The probability density function of X is $f(x) = 5$, $4.9 \le x \le 5.1$. What is the probability that a measurement of current is between 4.95 & 5.0 mA?

$$P(4.95 \le x \le 5) = \int_{4.95}^{5} f(x)\,dx = 5 \times (5 - 4.95)$$

$$= 0.25$$

CDF of the Continuous Uniform Random variable slide 17

$$F(x) = \int_a^x \frac{1}{(b-a)} du = \frac{x-a}{b-a}$$

The Cumulative distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \le x < b \\ 1 & b \le x \end{cases}$$

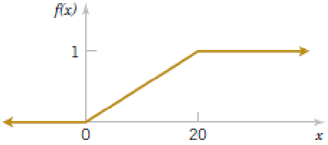


Figure 4-6 Cumulative distribution function

Normal Distribution

- A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

- is a **normal random variable** with parameter μ where $-\infty < \mu < \infty$, and $\sigma > 0$.
- Also

$$E[X] = \mu \quad \text{and} \quad V(X) = \sigma^2$$

- And the notation $N(\mu, \sigma^2)$ is used to denote the distribution.

Normal Distribution

For any normal random variable,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

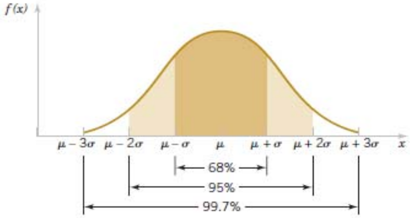


Figure 4-12 Probabilities associated with a normal distribution

Standard Normal Distribution

- A normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is called a standard normal random variable and is denoted as Z. The cumulative distribution function of a standard normal random variable is denoted as:

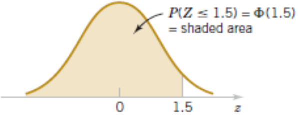
$$\Phi(z) = P(Z \leq z)$$

- Values are found in Appendix Table III, Excel, Minitab, R and other statistical packages.

How to Read a Standard Normal Table

Assume Z is a standard normal random variable.

Find $P(Z \leq 1.50)$. Answer: 0.93319



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Figure 4-13 Standard normal Probability density function

Find $P(Z \leq 1.53)$. Answer: 0.93699

Find $P(Z \leq 0.02)$. Answer: 0.50398

NOTE : The column headings refer to the hundredths digit of the value of z in $P(Z \leq z)$.
For example, $P(Z \leq 1.53)$ is found by reading down the z column to the row 1.5 and then selecting the probability from the column labeled 0.03 to be 0.93699.

Standardizing a Normal Random Variable

Suppose X is a normal random variable with mean μ and variance σ^2, the random variable

$$Z = \frac{(X - \mu)}{\sigma}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$.

The probability is obtained by using Appendix Table III with $z = \frac{(x-\mu)}{\sigma}$.

Standardizing a Normal Random Variable

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with $\mu = 10$ and $\sigma = 2$ mA, what is the probability that the current measurement is between 9 and 11 mA?

$$P(9 < X < 11) = P\left(\frac{9-10}{2} < \frac{x-10}{2} < \frac{11-10}{2}\right)$$

$$= P(-0.5 < z < 0.5)$$

$$= P(z < 0.5) - P(z < -0.5)$$

$$= 0.69146 - 0.30854 = 0.38292$$

Using Excel	
0.38292	= NORMDIST(11,10,2,TRUE) - NORMDIST(9,10,2,TRUE)

Standardizing a Normal Random Variable

Determine the value for which the probability that a current measurement is below 0.98.

Answer:

$$P(X < x) = P\left(\frac{X-10}{2} < \frac{x-10}{2}\right)$$

$$= P\left(Z < \frac{x-10}{2}\right) = 0.98$$

$$z = 2.05 \text{ is the closest value.}$$

$$z = 2(2.05) + 10 = 14.1 \text{ mA.}$$

Using Excel	
14.107	= NORMINV(0.98,10,2)

Normal Approximations

- The binomial and Poisson distributions become more bell-shaped and symmetric as their mean value increase.
- For manual calculations, the normal approximation is practical - exact probabilities of the binomial and Poisson, with large means, require technology (Minitab, Excel).
- The normal distribution is a good approximation for:
  - Binomial if $np > 5$ and $n(1-p) > 5$.
  - Poisson if $\lambda > 5$

Normal Approximations for Binomial

- If X is a binomial random variable with parameters n and p, then

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

- Is approximately a standard normal random variable.
- To approximate the binomial probability with a normal distribution, a *continuity correction*, is applied as follows:

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

and

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

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

Applying the Approximation

In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable. The probability that a bit is received in error is 10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur?

$$P(X \le 150) = P(X \le 150.5)$$

$$= P\left(\frac{X-160}{\sqrt{160(1-10^{-5})}} \le \frac{150.5-160}{\sqrt{160(1-10^{-5})}}\right)$$

$$= P\left(Z \le \frac{-9.5}{12.6491}\right) = P(Z \le -0.75104) = 0.2263$$

Using Excel	
0.2263	= NORMDIST(150.5, 160, SQRT(160*(1-0.00001)), TRUE)
-0.7%	= (0.2263-0.228)/0.228 = percent error in the approximation

Normal Approximation for Poisson

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

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

is approximately a standard normal random variable. The same continuity correction used for the binomial distribution can also be applied. The approximation is good for $\lambda \geq 5$

Normal Approximation for Poisson

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

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

The probability can be approximated as

$$\begin{aligned} P(X \le 950) &= P(X \le 950.5) \\ &\approx P\left(Z \le \frac{950.5 - 1000}{\sqrt{1000}}\right) \\ &= P(Z \le -1.57) = 0.058 \end{aligned}$$

Using Excel	
0.0578	= POISSON(950,1000,TRUE)
0.0588	= NORMDIST(950.5, 1000, SQRT(1000), TRUE)
1.6%	= (0.0588 - 0.0578) / 0.0578 = percent error

Exponential Distribution

- The random variable X that equals the distance between successive events of a Poisson process with mean number of events $\lambda > 0$ per unit interval is an exponential random variable with parameter λ. The probability density function of X is:

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \leq x \leq \infty$$

- Mean and variance of an exponential Random variable are

$$\mu = E[x] = 1/\lambda \quad \text{and} \quad \sigma^2 = V(X) = 1/\lambda^2$$

Example of Exponential Distribution

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in the next 6 minutes (0.1 hour)?

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

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

The cumulative distribution function also can be used to obtain the same result as follows

$$P(X > 0.1) = 1 - F(0.1) = 0.082$$

Using Excel	
0.0821	= 1 - EXPONDIST(0.1,25,TRUE)

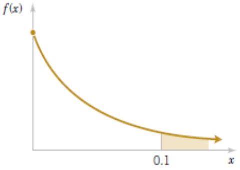


Figure 4-23 Desired probability

Example of Exponential Distribution

Continuing, what is the probability that the time until the next log-on is between 2 and 3 minutes (0.033 & 0.05 hours)?

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

$$= -e^{-25x}\Big|_{0.033}^{0.05} = 0.152$$

An alternative solution is

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

Using Excel	
0.148	= EXPONDIST(3/60, 25, TRUE) - EXPONDIST(2/60, 25, TRUE)
	(difference due to round-off error)

Example of Exponential Distribution

- Continuing, what is the interval of time such that the probability that no log-on occurs during the interval is 0.90?

$$P(X > x) = e^{-25x} = 0.90, \quad -25x = \ln(0.90)$$

$$x = \frac{-0.10536}{-25} = 0.00421 \text{ hour} = 0.25 \text{ minute}$$

- What is the mean and standard deviation of the time until the next log-in?

$$\mu = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hour } = 2.4 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hour } = 2.4 \text{ minutes}$$

Example of Exponential Distribution

An interesting property of an exponential random variable concerns conditional probabilities.

For an exponential random variable *X*,

$$P(X<t_1+t_2|X>t_1)= P(X < t_2)$$

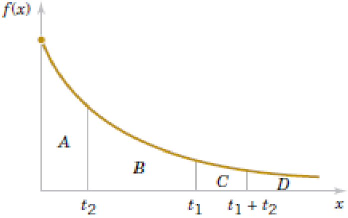


Figure 4-24 Lack of memory property of an exponential distribution.

Example of Exponential Distribution

Let X denote the time between detections of a particle with a Geiger counter. Assume X has an exponential distribution with E(X) = 1.4 minutes. What is the probability that a particle is detected in the next 30 seconds?

$$P(X < 0.5) = F(0.5) = 1 - e^{-0.5/1.4} = 0.30$$

Using Excel	
0.300	= EXPONDIST(0.5, 1/1.4, TRUE)

No particle has been detected in the last 3 minutes. Will the probability increase since it is "due"?

$$P(X < 3.5 | X > 3) = \frac{P(3 < X < 3.5)}{P(X > 3)} = \frac{F(3.5) - F(3)}{1 - F(3)} = \frac{0.035}{0.117} = 0.30$$

No, the probability that a particle will be detected depends only on the interval of time, not its detection history.

Erlang & Gamma Distribution

- The Erlang distribution is a generalization of the exponential distribution.
- The exponential distribution models the interval to the 1*st* event, while the Erlang distribution models the interval to the *rth* event, i.e., a sum of exponentials.
- If r is not required to be an integer, then the distribution is called gamma.
- The exponential, as well as its Erlang and gamma generalizations, is based on the Poisson process.

Example Process Failure slides 39

- The failures of CPUs of large computer systems are often modeled as a Poisson process. Assume that units that fail are repaired immediately and the mean number of failures per hour is 0.0001. Let X denote the time until 4 failures occur. What is the probability that X exceed $40,000$hours?

Erlang Distribution

Generalizing from the prior exercise:

$$P(X > x) = \sum_{k=0}^{r-1} \frac{e^{\lambda x} (\lambda x)^k}{k!} = 1 - F(x)$$

Now differentiating $F(x)$:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \quad \text{for } x > 0 \;\; \text{and} \;\; r = 1, 2, \ldots$$

Gamma Distribution

The gamma function is the generalization of the factorial function for r > 0, not just non-negative integers.

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx, \quad \text{for } r > 0$$

Properties of the gamma function

$$\Gamma(r) = (r-1)\Gamma(r-1) \quad \text{recursive property}$$

$$\Gamma(r) = (r-1)! \quad \text{factorial function}$$

$$\Gamma(1) = 0! = 1$$

$$\Gamma(1/2) = \pi^{1/2} = 1.77 \quad \text{useful if manual}$$

Gamma Distribution

The random variable X with probability density function:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \text{ for } x > 0$$

is a gamma random variable with parameters $\lambda > 0$ and $r > 0$. If r is an integer, then X has an Erlang distribution.

Gamma Distribution

If X is a gamma random variable with parameters λ and r,

$$\mu = E(X) = r / \lambda$$

and

$$\sigma^2 = V(X) = r / \lambda^2$$

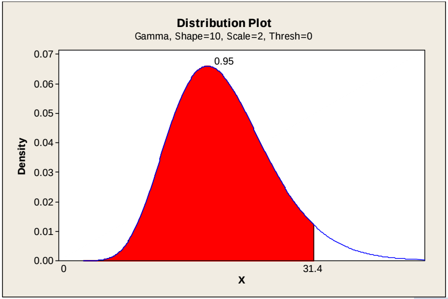
Example Gamma Application slides 44-45

- The time to prepare a micro-array slide for high-output genomics is a Poisson process with a mean of 2 hours per slide. What is the probability that 10 slides require more than 25 hours?
- Let X denote the time to prepare 10 slides. Because of the assumption of a Poisson process, X has a gamma distribution with $\lambda = 1/2$, $r = 10$, and the requested probability is $P(X > 25)$.

Example Gamma Application slides 44-45

Example Gamma Application

The slides will be completed by what length of time with 95% probability? That is: $P(X \leq x) = 0.95$



Minitab: Graph > Probability Distribution Plot > View Probability

Using Excel	
31.41	= GAMMAINV(0.95, 10, 2)

Weibull Distribution

The random variable X with probability density function

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-(x/\delta)^{\beta}}, \text{ for } x > 0$$

is a Weibull random variable with
scale parameter $\delta > 0$ and shape parameter $\beta > 0$.

The cumulative distribution function is:

$$F(x) = 1 - e^{-(x/\delta)^{\beta}}$$

The mean and variance is given by

$$\mu = E(X) = \delta \cdot \Gamma\left(1 + \frac{1}{\beta}\right) \quad \text{and}$$

$$\sigma^2 = V(X) = \delta^2\left[\Gamma\left(1 + \frac{2}{\beta}\right)\right] - \delta^2\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2$$

Example Weibull Distribution slides 48

- The time to failure (in hours) of a bearing in a mechanical shaft is modeled as a Weibull random variable with $\beta = 1/2$ and $\delta = 5,000$ hours.
- What is the mean time until failure?
- What is the probability that a bearing will last at least 6,000 hours?

Lognormal Distribution

Let W denote a normal random variable with mean θ and variance ω^2, then $X = \exp(W)$ is a lognormal random variable with probability density function

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} e^{-\left[\frac{(\ln(x)-\theta)^2}{2\omega^2}\right]} \qquad 0 < x < \infty$$

The mean and variance of X are

$$E(X) = e^{\theta+\omega^2/2} \quad \text{and}$$

$$V(X) = e^{2\theta+\omega^2}\left(e^{\omega^2} - 1\right)$$

Example Lognormal Distribution slides 50

- The lifetime of a semiconductor laser has a lognormal distribution with $\theta = 10$ and $\omega = 1.5$ hours. What is the probability that the lifetime exceeds 10,000 hours?

Example Lognormal Distribution slides 50

- What lifetime is exceeded by 99% of lasers?

Example Lognormal Distribution slides 50

- What is the mean and variance of the lifetime?

Beta Distribution

- The random variable X with probability density function

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for} \quad x \in [0,1]$$

- is a beta random variable with parameters $\alpha > 0$ and $\beta > 0$
- If X has a beta distribution with parameters α and β,

$$\mu = E[X] = \frac{\alpha}{\alpha+\beta}$$

$$\sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Example Beta Distribution slides 53

- Consider the completion time of a large commercial real estate development. The proportion of the maximum allowed time to complete a task is a beta random variable with $\alpha = 2.5$ and $\beta = 1$. What is the probability that the proportion of the maximum time exceeds 0.7?

Beta Distribution slides 53

- Consider the completion time of a large commercial real estate development. The proportion of the maximum allowed time to complete a task is a beta random variable with $\alpha = 2.5$ and $\beta = 1$. What is the probability that the proportion of the maximum time exceeds 0.7?