University of California, Los Angeles
Department of Statistics

Statistics 100B | Instructor: Nicolas Christou

Continuous probability distributions

- Let X be a continuous random variable, $-\infty < X < \infty$

- $f(x)$ is the so called probability density function (pdf) if

  $\int_{-\infty}^{\infty} f(x)dx = 1$

- Area under the pdf is equal to 1.

- How do we compute probabilities? Let X be a continuous r.v. with pdf $f(x)$. Then

$$P(X > a) = \int_a^{\infty} f(x)dx$$
$$P(X < a) = \int_{-\infty}^{a} f(x)dx$$
$$P(a < X < b) = \int_a^b f(x)dx$$

- Note that in continuous r.v. the following is true:

  $P(X \geq a) = P(X > a)$

  This is NOT true for discrete r.v.

- Cumulative distribution function (cdf):

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

- Therefore

$$f(x) = F(x)'$$

- Compute probabilities using cdf:

$$P(a < X < b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

- Example: Let the lifetime X of an electronic component in months be a continuous r.v. with $f(x) = \frac{10}{x^2}, x > 10$.

  a. Find $P(X > 20)$.

  b. Find the cdf.

  c. Use the cdf to compute $P(X > 20)$.

  d. Find the 75_{th} percentile of the distribution of X.

  e. Compute the probability that among 6 such electronic components, at least two will survive more than 15 months.

- Mean of a continuous r.v.

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

- Mean of a function of a continuous r.v.

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- Variance of continuous r.v.

$$\sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Or

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - [E(X)]^2$$

- Some properties: Let a, b constants and X, Y r.v.

$$\begin{aligned} E(X + a) &= a + E(X) \\ E(X + Y) &= E(X) + E(Y) \\ var(X + a) &= var(X) \\ var(aX + b) &= a^2 var(X) \end{aligned}$$

If X, Y are independent then

$$var(X + Y) = var(X) + var(Y)$$

- Example: Let X be a continuous r.v. with $f(x) = ax + bx^2$, and $0 < x < 1$.

  a. If $E(X) = 0.6$ find a, b.

  b. Find $var(X)$.

- **Uniform probability distribution:**
  A continuous r.v. X follows the uniform probability distribution on the interval a, b if its pdf function is given by

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

  - Find cdf of the uniform distribution.
  - Find the mean of the uniform distribution.
  - Find the variance of the uniform distribution.

- **The gamma distribution**

The gamma distribution is useful in modeling skewed distributions for variables that are not negative.

A random variable X is said to have a gamma distribution with parameters α, β if its probability density function is given by

$$f(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}, \quad \alpha, \beta > 0, x \geq 0.$$

$E(X) = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$.

A brief note on the *gamma* function:
The quantity $\Gamma(\alpha)$ is known as the gamma function and it is equal to:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx.$$

If $\alpha = 1$, $\Gamma(1) = \int_0^{\infty} e^{-x}dx = 1$.

With integration by parts we get $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ as follows:

$$\Gamma(\alpha+1) = \int_0^{\infty} x^{\alpha}e^{-x}dx =$$

Let, $v = x^{\alpha} \Rightarrow \frac{dv}{dx} = \alpha x^{\alpha-1}$
$\frac{du}{dx} = e^{-x} \Rightarrow u = -e^{-x}$

Therefore,

$$\Gamma(\alpha+1) = \int_0^{\infty} x^{\alpha}e^{-x}dx = -e^{-x}x^{\alpha}|_0^{\infty} - \int_0^{\infty} -e^{-x}\alpha x^{\alpha-1}dx = \alpha\int_0^{\infty} x^{\alpha-1}e^{-x}dx.$$

Or, $\Gamma(\alpha+1) = \alpha\Gamma(a)$.

Similarly, using integration by parts it can be shown that,
$\Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha+1) = (\alpha+1)\alpha\Gamma(\alpha)$, and,
$\Gamma(\alpha+3) = (\alpha+2)(\alpha+1)\alpha\Gamma(\alpha)$.

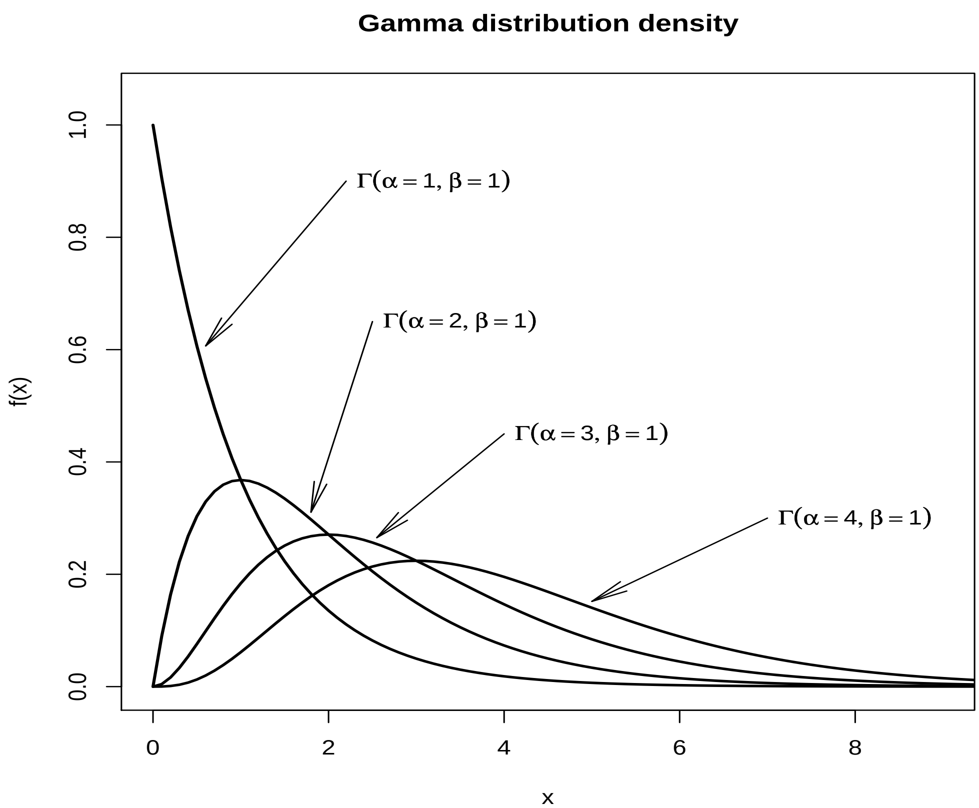
Therefore, using this result, when α is an integer we get $\Gamma(\alpha) = (\alpha-1)!$.

Example:
$\Gamma(5) = \Gamma(4+1) = 4\times\Gamma(4) = 4\times\Gamma(3+1) = 4\times 3\times\Gamma(3) = 4\times 3\times\Gamma(2+1) = 4\times 3\times 2\times\Gamma(1+1) = 4\times 3\times 2\times\times 1\times\Gamma(1) = 4!$.

Useful result:

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}.$$

The gamma density for $\alpha = 1, 2, 3, 4$ and $\beta = 1$.



- **Exponential probability distribution:**
  Useful for modeling the lifetime of electronic components.

- A continuous r.v. X follows the exponential probability distribution with parameter $\lambda > 0$ if its pdf function is given by

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

  Note: From the pdf of the gamma distribution, if we set $\alpha = 1$ and $\beta = \frac{1}{\lambda}$ we get $f(x) = \lambda e^{-\lambda x}$. We see that the exponential distribution is a special case of the gamma distribution.

  - Find cdf of the exponential distribution.
  - Find the mean of the exponential distribution.
  - Find the variance of the exponential distribution.
  - Find the median of the exponential distribution.
  - Find the p_{th} percentile of the exponential distribution.

- Example:
  Let X be an exponential random variable with $\lambda = 0.2$.

  a. Find the mean of X.

  b. Find the median of X.

  c. Find the variance of X.

  d. Find the 80_{th} percentile of this distribution (or find c such that $P(X < c) = 0.80$).

- Memoryless property of the exponential distribution:
  Suppose the lifetime of an electronic component follows the exponential distribution with parameter λ. The memoryless property states that

$$P(X > s + t | X > t) = P(X > s), \quad s > 0, t > 0$$

Example:
Suppose the number of miles a car can run before its battery wears out follows the exponential distribution with mean $\mu = 10000$ miles. If the owner of the car takes a 5000-mile trip what is the probability that he will be able to complete the trip without having to replace the battery of the car?

If the number of miles follow some other distribution with known cumulative distribution function (cdf) give an expression of the probability of completing the trip without having to replace the battery of the car in terms of the cdf.

The distribution of a function of a random variables

Suppose we know the pdf of a random variable X. Many times we want to find the probability density function (pdf) of a function of the random variable X. Suppose $Y = X^n$.

We begin with the cumulative distribution function of Y:

$$F_Y(y) = P(Y \leq y) = P(X^n \leq y) = P(X \leq y^{\frac{1}{n}}).$$

So far we have

$$F_Y(y) = F_X(y^{\frac{1}{n}})$$

To find the pdf of Y we simply differentiate both sides wrt to y:

$$f_Y(y) = \frac{1}{n} y^{\frac{1}{n}-1} \times f_X(y^{\frac{1}{n}}).$$

where, $f_X(\cdot)$ is the pdf of X which is given. Here are some more examples.

Example 1
Suppose X follows the exponential distribution with $\lambda = 1$. If $Y = \sqrt{X}$ find the pdf of Y.

Example 2
Let $X \sim N(0, 1)$. If $Y = e^X$ find the pdf of Y. Note: Y it is said to have a log-normal distribution.

Example 3
Let X be a continuous random variable with pdf $f(x) = 2(1 - x), 0 \leq x \leq 1$. If $Y = 2X - 1$ find the pdf of Y.

Example 4
Let X be a continuous random variable with pdf $f(x) = \frac{3}{2}x^2, -1 \leq x \leq 1$. If $Y = X^2$ find the pdf of Y.

Continuous random variables - Some examples

(Some are from: Sheldon Ross (2002), *A first Course in Probability*, Sixth Edition, Prentice Hall).

Example 1
Suppose X, the lifetime of a certain type of electronic device (in hours), is a continuous random variable with probability density function $f(x) = \frac{10}{x^2}$ for $x > 10$ and $f(x) = 0$ for $x \leq 10$.

a. Find $P(X > 20)$.

b. Find the cumulative distribution function (cdf).

c. Find the 75_{th} percentile of this distribution.

d. What is the probabilty that among 6 such types of devices at least 3 will function for at least 15 hours?

Example 2
Suppose a bus always arrives at a particular stop between 8:00 AM and 8:10 AM. Find the probability that the bus will arrive tomorrow between 8:00 AM and 8:02 AM.

Example 3
A parachutist lands at a random point on a line AB.

a. Find the probability that he is closer to A than to B.

b. Find the probability that his distance to A is more thant 3 times his distance to B.

Example 4
Suppose the length of a phone call in minutes follows the exponential distribution with parameter $\lambda = 0.1$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait

a. more than 10 minutes.

b. between 10 and 20 minutes.

Example 5
Let X be an exponential random variable with $\lambda = 0.2$.

a. Find the mean of X.

b. Find the median of X.

c. Find the variance of X.

d. Find the 80_{th} percentile of this distribution (or find c such that $P(X < c) = 0.80$).

Example 6
The random variable X has probability density function

$$f(x) = \begin{cases} ax + bx^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

If $E(X) = 0.6$ find

a. $P(X < \frac{1}{2})$.

b. $\text{Var}(X)$.

Example 7
For some constant c, the random variable X has probability density function

$$f(x) = \begin{cases} cx^4 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find

a. $E(X)$.

b. $\text{Var}(X)$.

Example 8
To be a winner in the following game, you must be succesful in three succesive rounds. The game depends on the value of X, a uniform random variable on (0,1). If $X > 0.1$, then you are succesful in round 1; if $X > 0.2$, then you are succesful in round 2; if $X > 0.3$, then you are succesful in round 3.

a. Find the probability that you are succesful in round 1.

b. Find the conditional probability that you are succesful in round 2 given that you were succesful in round 1.

c. Find the conditional probability that you are succesful in round 3 given that you were succesful in round 2.

d. Find the probability that you are a winner.

Example 9
There are two types of batteries in a bin. The lifetime of type i battery is an exponential random variable with parameter λ_i, $i = 1, 2$. The probability that a type i battery is chosen from the bin is p_i. If a randomly chosen battery is still operating after t hours of use, what is the probability it willl still be operating after an additional s hours?

Example 10
You bet $1 on a specified number at a roulette table. A roulette wheel has 38 slots, numbered 0, 00, and 1 through 36. Approximate the probability that

a. In 1000 bets you win more than 28 times.

b. In 10000 bets you win more than 270 times.

Continuous random variables - Some examples
Solutions

Example 1
We are given that the pdf of X is $f(x) = \frac{10}{x^2}$ for $x > 10$ and $f(x) = 0$ for $x \leq 10$.

a. $$P(X > 20) = \int_{20}^{\infty} \frac{10}{x^2} dx = -\frac{10}{x}|_{20}^{\infty} = 0 - (-\frac{10}{20}) = \frac{1}{2}.$$

b. The cumulative distribution function (cdf) is defined as $F(x) = P(X \leq x)$.
$$F(x) = P(X \leq x) = \int_{10}^{x} \frac{10}{u^2} du = -\frac{10}{u}|_{10}^{x} \Rightarrow F(x) = 1 - \frac{10}{x}.$$

c. We want to find a value of X (call it p) such that $P(X \leq p) = 0.75$.
$$\int_{10}^{p} \frac{10}{x^2} dx = 0.75 \Rightarrow -\frac{10}{x}|_{10}^{p} = 0.75 \Rightarrow 1 - \frac{10}{p} = 0.75 \Rightarrow p = 40.$$
Therefore the 75_{th} percentile is 40, which means $P(X \leq 40) = 0.75$.

d. We first find the probability that a device willfunction for at least 15 hours:
$$P(X > 15) = \int_{15}^{\infty} \frac{10}{x^2} dx = -\frac{10}{x}|_{15}^{\infty} = 0 - (-\frac{10}{15}) = \frac{2}{3}.$$
Now we have $n = 6$ devices with $p = \frac{2}{3}$. We want to find $P(X \geq 3)$, where X represents the number of devices that function for at least 15 hours among the 6 selected. Therefore $X \sim b(6, \frac{2}{3})$. The answer is:
$$P(X \geq 3) = \sum_{x=3}^{6} \binom{6}{x} (\frac{2}{3})^x (\frac{1}{3})^{6-x}.$$

Example 2
This is a uniform distribution problem with $f(x) = \frac{1}{10}$. We want $P(0 < X < 2) = \int_0^2 \frac{1}{10} dx = \frac{x}{10}|_0^2 = 0.20$.

Example 3
Another uniform distribution problem with $f(x) = \frac{1}{a}$, where a is the length of the line AB with B at the origin.

a. $P(\frac{a}{2} < X < a) = \int_{\frac{a}{2}}^{a} \frac{1}{a} dx = \frac{x}{a}|_{\frac{a}{2}}^{a} = 1 - \frac{1}{2} = \frac{1}{2}$.

b. $P(0 < X < \frac{a}{4}) = \int_0^{\frac{a}{4}} \frac{1}{a} dx = \frac{x}{a}|_0^{\frac{a}{4}} = \frac{1}{4} - 0 = \frac{1}{4}$.

Example 4
We are given that $f(x) = 0.1e^{-0.1x}$, $x \geq 0$.

a. $P(X > 10) = \int_{10}^{\infty} 0.1e^{-0.1x} dx = -e^{-0.1x}|_{10}^{\infty} = e^{-0.1(10)} = e^{-1}$.

b. $P(< 10 < X > 20) = \int_{10}^{20} 0.1e^{-0.1x} dx = -e^{-0.1x}|_{10}^{20} = -e^{-0.1(20)} + e^{-0.1(10)} = e^{-1} - e^{-2}$.

Example 5

Let X be an exponential random variable with $\lambda = 0.2$.

a. $\mu = \frac{1}{\lambda} = \frac{1}{0.2} \Rightarrow \mu = 5$.

b. Let m be the median. We want $P(X \leq m) = 0.50$. Therefore

$$P(X \leq m) = 0.50 \Rightarrow \int_0^m 0.1e^{-0.2x}dx = 0.50 \Rightarrow -e^{-0.2x}|_0^m = 0.50 \Rightarrow m = \frac{ln(0.5)}{-0.2} = 3.47.$$

Or faster way, is to use the cdf function:
$F(m) = 1 - e^{-0.2m} = 0.50 \Rightarrow m = \frac{ln(0.5)}{-0.2} = 3.47$. The median is 3.47, which means $P(X \leq 3.47) = 0.50$.

c. $\mu = \frac{1}{\lambda^2} = \frac{1}{0.2^2} \Rightarrow \mu = 25$.

d. Let p be the 80_{th} percentile. We want $P(X \leq p) = 0.80$. Therefore

$$P(X \leq p) = 0.80 \Rightarrow \int_0^p 0.1e^{-0.2x}dx = 0.80 \Rightarrow -e^{-0.2x}|_0^p = 0.80 \Rightarrow p = \frac{ln(0.2)}{-0.2} = 8.05.$$

Or faster way, is to use the cdf function:
$F(p) = 1 - e^{-0.2p} = 0.80 \Rightarrow p = \frac{ln(0.2)}{-0.2} = 8.05$. The 80_{th} percentile is 8.05, which means $P(X \leq 8.05) = 0.80$.

Example 6

First we must find the constants a, and b. Since $f(x)$ is a pdf we know that $\int_0^1 f(x)dx = 1$. Therefore $\int_0^1 (ax + bx^2)dx = 1 \Rightarrow [a\frac{x^2}{2} + b\frac{x^3}{3}]|_0^1 = 1 \Rightarrow 3a + 2b = 6$ (1).
We also know that $E(X) = \int_0^1 xf(x)dx$ and that $E(X) = 0.6$. Therefore $\int_0^1 x(ax + bx^2)dx = 0.6 \Rightarrow [a\frac{x^3}{3} + b\frac{x^4}{4}]|_0^1 = 0.6 \Rightarrow 4a + 3b = 7.2$ (2). Now solving (1) and (2) for a, and b we get $a = 3.6$ and $b = -2.4$. The pdf is $f(x) = 3.6x - 2.4x^2, 0 < x < 1$.

a. $P(X < \frac{1}{2}) = \int_0^{\frac{1}{2}} (3.6x - 2.4x^2)dx = (3.6\frac{x^2}{2} - 2.4\frac{x^3}{3})|_0^{\frac{1}{2}} = 0.35$.

b. $\sigma^2 = \int_0^1 x^2 f(x)dx - \mu^2 = \int_0^1 x^2(3.6x - 2.4x^2)dx - (0.6)^2 = [3.6\frac{x^4}{4} - 2.4\frac{x^5}{5}]|_0^1 - (0.6)^2 \Rightarrow \sigma^2 = 0.06$.

Example 7

First we must find the constant c. $\int_0^1 cx^4dx = 1 \Rightarrow c\frac{x^5}{5}|_0^1 = 1 \Rightarrow c = 5$. The pdf is $f(x) = 5x^4, 0 < x < 1$.

a. $E(X) = \int_0^1 x5x^4dx = 5\frac{x^6}{6}|_0^1 \Rightarrow E(X) = \frac{5}{6}$.

b. $Var(X) = \int_0^1 x^2 5x^4dx - \mu^2 = 5\frac{x^7}{7}|_0^1 - (\frac{5}{6})^2 \Rightarrow Var(X) = 0.0198$.

Example 8

This is a uniform distribution problem with $f(x) = 1, 0 < x < 1$.

a. $P(X > 0.1) = \int_{0.1}^1 dx = x|_{0.1}^1 = 0.9$.

b. $P(X > 0.2/X > 0.1) = \frac{P(X>0.2)}{P(X>0.1)} = \frac{0.8}{0.9} = \frac{8}{9}$.

c. $P[X > 0.3/(X > 0.1 \cap X > 0.2)] = \frac{P(X>0.3)}{P(X>0.2)} = \frac{0.7}{0.8} = \frac{7}{8}$.

d. The probability that you are a winner is the probability that you are a winner in round 1, and winner in round 2, and winner in round 3. This is equal to the product of the 3 probabilities found in parts a,b,c. $P(win) = \frac{9}{10}\frac{8}{9}\frac{7}{8} = \frac{7}{10}$.

Example 9

We want $P(X > s + t/x > t)$. This is equal to:

$$P(X > s+t|x > t) = \frac{P(X > s+t \cap X > t)}{P(X > t)} =$$

$$\frac{P(X > s+t)}{P(X > t)} = \frac{P(X > s+t)p_1 + P(X > s+t)p_2}{P(X > t)p_1 + P(X > t)p_2} \Rightarrow$$

Using the exponential cdf we find: $P(X > x) = 1 - P(X \leq x) = 1 - [1 - e^{-\lambda x}] = e^{-\lambda x}$

$$P(X > s+t/X > t) = \frac{e^{-\lambda_1(s+t)}p_1 + e^{-\lambda_2(s+t)}p_2}{e^{-\lambda_1 t}p_1 + e^{-\lambda_2 t}p_2}.$$

Example 10

Let X be the number of times you win among the n times you play this game. Then $X \sim b(n, \frac{1}{38})$.

a. We want $P(X > 28)$. We will approximate this probability using the normal distribution. We need μ and σ. These are equal to: $\mu = 1000\frac{1}{38} = 26.32$, and $\sigma^2 = 1000\frac{1}{38}(1 - \frac{1}{38}) = 25.62 \Rightarrow \sigma = 5.06$. Now the desired probability is (we use the continuity correction):
$P(X > 28) = P(Z > \frac{28.5-26.32}{5.06}) = P(Z > 0.43) = 1 - 0.6664 = 0.3336$.

b. We want $P(X > 280)$. Again we will approximate this probability using the normal distribution. We need μ and σ. These are equal to: $\mu = 10000\frac{1}{38} = 263.16$, and $\sigma^2 = 10000\frac{1}{38}(1 - \frac{1}{38}) = 256.23 \Rightarrow \sigma = 16.01$. Now the desired probability is (we use the continuity correction):
$P(X > 280) = P(Z > \frac{280.5-263.16}{16.01}) = P(Z > 1.08) = 1 - 0.8599 = 0.1401$.