Continuous Distributions

SOC 512 & CSSS 505
written by Laina Mercer & Jessica Godwin

Aaron Osgood-Zimmerman

Department of Statistics
University of Washington

February 25, 2021

Outline

A continuous random variable could have a number of different probability distributions. Today we will focus on the following continuous distributions.

- Uniform
- Univariate Normal
- Chi-Square
- Student's t-distribution
- Exponential

Uniform Distribution

Discrete

A uniform probability distribution assigns equal probability to every possible value for the random variable. A uniform distribution may be discrete or continuous.

A *discrete uniform* random variable takes on a finite number of values.

Examples:

- Let X be a random integer between 1 and 10. Then X takes on the values $\{1, 2, 3, ..., 10\}$, each with probability $1/10$.
- If Y is the roll of a die, Y is a discrete uniform random variable with range $\{1, 2, 3, 4, 5, 6\}$, each of which has probability $1/6$.
- The Powerball lottery winner results from drawing five independent random numbers between 1 and 59. Let Z be the first of these numbers, so Z takes values $\{1, 2, 3, ..., 59\}$, each with probability $1/59$. The next four numbers drawn have the same probability distribution as Z.

Uniform Distribution

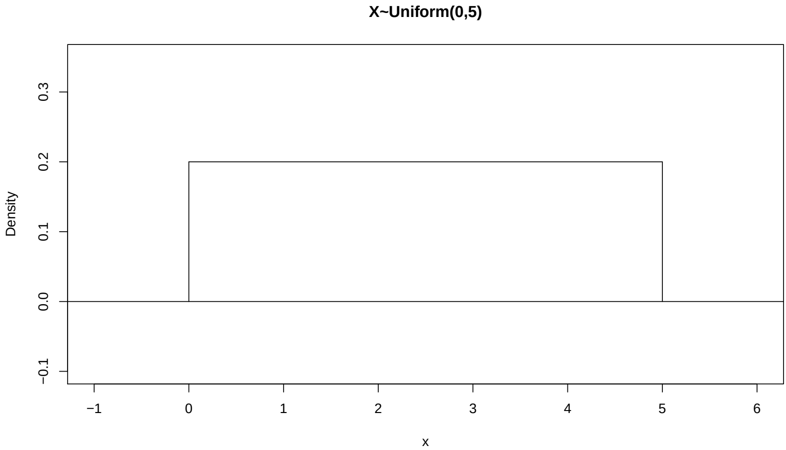
Continuous

A *continuous uniform* random variable has as its range a continuous interval on the real number line. Any point in the interval is equally likely. Since there are an infinite number of points, we can't write down the probability of a single point, because they would sum to more than one. We can only express the probability of the variable falling within a certain interval.

For continuous random variables, we don't have a probability distribution function, since we can only express probabilities of intervals rather than probabilities of points. Instead we use a probability density function (pdf). By integrating the probability density function we can compute the probability of X falling in a given interval.

Uniform Distribution

Continuous



Recall our definition of the expectation:

$$E[X] = \sum_{i=1}^{n} P(X = x_i) \cdot x_i = \sum_{i=1}^{n} p_i \cdot x_i$$

We can extend this definition to continuous distributions. For a continuous distribution, $P(X = x_i)$ is always zero, but we can compute the probability that X falls within a certain interval by integrating the pdf over that interval. If we divide up the real line into very small intervals, we can estimate $E[X]$ with

$$E[X] \cong \sum_{i=1}^{n} P(x_i < X < x_{i+1}) \cdot x_i$$

By letting the number of rectangles approach infinity while their width approaches zero (and taking the limit), we obtain:

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

Example, continuous uniform distribution on [0,5]. $f(x) = 1/5$ for $x \in [0, 5]$.

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{5} x \cdot 1/5dx = 1/5 \int_{0}^{5} xdx \\ &= 1/5[x^2/2]_0^5 = 1/10[5^2 - 0^2] = 25/10 = 2.5 \end{aligned}$$

We can extend the formula for the variance to continuous random variables in the same way. Recall the formula for the variance for a discrete distribution:

$$Var[X] = \sum_{i=1}^{n} (x_i - E[X])^2 \cdot P(X = x_i) = \sum_{i=1}^{n} (x_i - E[X])^2 \cdot p_i$$

Following the arguments we used for the expectation, we obtain the following formula for the variance:

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

Variance
Continuous Random Variables

Example, continuous uniform distribution on [0,5]. $f(x) = 1/5$ for $x \in [0,5]$. Previously we found that $E[X] = 2.5$.

$$\begin{aligned}
Var[X] &= \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx = \int_0^5 (x - 2.5)^2 \cdot 1/5dx \\
&= 1/5 \int_0^5 (x - 2.5)^2 dx = 1/5 \left[\frac{1}{3}(x - 2.5)^3\right]_0^5 \\
&= 1/15 \left[(5 - 2.5)^3 - (0 - 2.5)^3\right] = 31.25/15 = 2.0833
\end{aligned}$$

Continuous Uniform

Expectation & Variance

In general, for $X \sim Uniform(a, b)$ has

$$E[X] = \frac{a+b}{2}$$

$$V[X] = \frac{(b-a)^2}{12}$$

Example, continuous uniform distribution on [0,5].

$$E[X] = \frac{5-0}{2} = 2.5 \; \& \; V[X] = \frac{(5-0)^2}{12} = \frac{25}{12} = 2.083$$

Normal Distribution

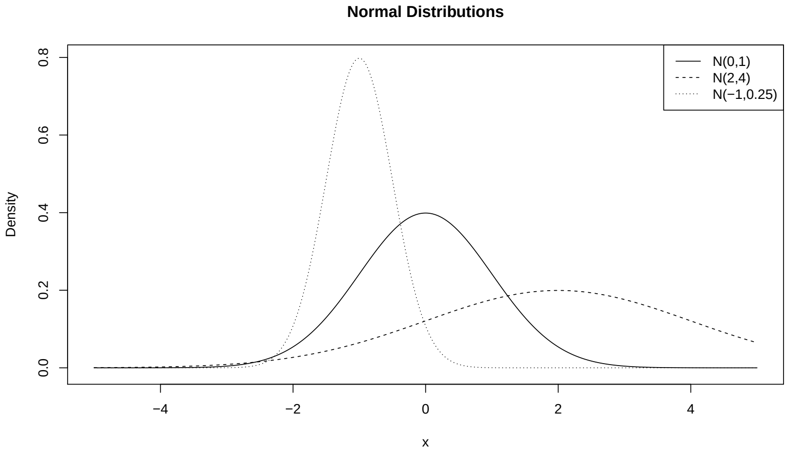
The *Gaussian* or *normal* distribution is the most commonly used distribution in statistics. It looks much like a bell curve and is often used to represent large populations. (The underlying idea in many statical concepts is that the higher n gets, the more normal the distribution of the average gets).

Probability Distribution:

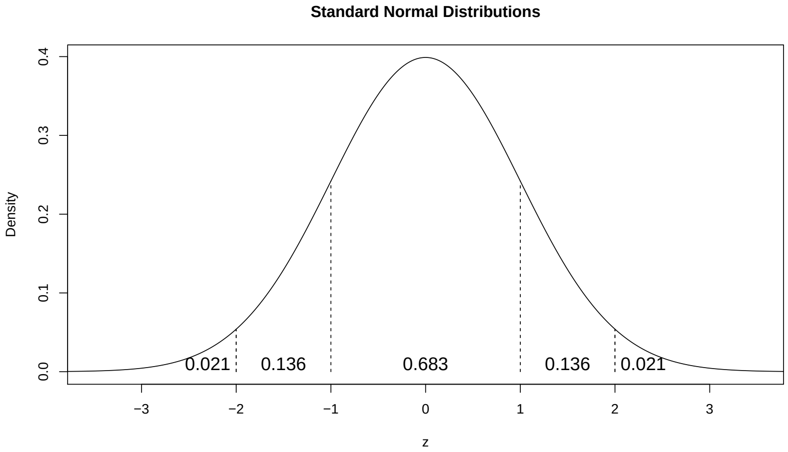
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \text{ where } -\infty < x < \infty$$

Each normal distribution has two parameters, the mean (μ) and variance (σ^2). The μ is the mean or central tendency of the distribution: σ^2 is the variance or measure of spread. The σ by itself is the *standard deviation*, the square root of the variance.

Normal Distributions



Standard Normal Distribution



Standard Normal Distribution

In order to find $P(a < X < b)$ we need to integrate the probability distribution function of a Normal distribution. This is a difficult integral to compute (by hand). A common approach is to 'standardize' a distributions.

If $X \sim N(\mu, \sigma^2)$ and we would like to 'standardize' the distribution to look like a $N(0, 1)$ we need to define a random variable

$$Z = \frac{X - \mu}{\sigma}.$$

We have subtracted the mean and divided by the standard deviation.

Standard Normal Distribution

If $X \sim N(\mu, \sigma^2)$ then $E[X] = \mu$ and $Var[X] = \sigma^2$. So, for $Z = \frac{X-\mu}{\sigma}$ we have:

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{E[X]-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

$$Var[Z] = Var\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma^2}Var[X-\mu] = \frac{1}{\sigma^2}Var[X] = \frac{\sigma^2}{\sigma^2} = 1$$

Standard Normal Distribution

Example

If $X \sim N(3, 4)$ what is $P(5 < X < 7)$? Let's define $Z = \frac{X-\mu}{\sigma}$ as we know .

$$
\begin{aligned}
P(5 < X < 7) &= P\left(5 - \mu < X - \mu < 7 - \mu\right) \\
&= P\left(\frac{5-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{7-\mu}{\sigma}\right) \\
&= P\left(\frac{5-3}{2} < Z < \frac{7-3}{2}\right) \\
&= P\left(1 < Z < 2\right) \\
&= 0.136 \text{ (from slide 13)}
\end{aligned}
$$

Chi-Square Distribution

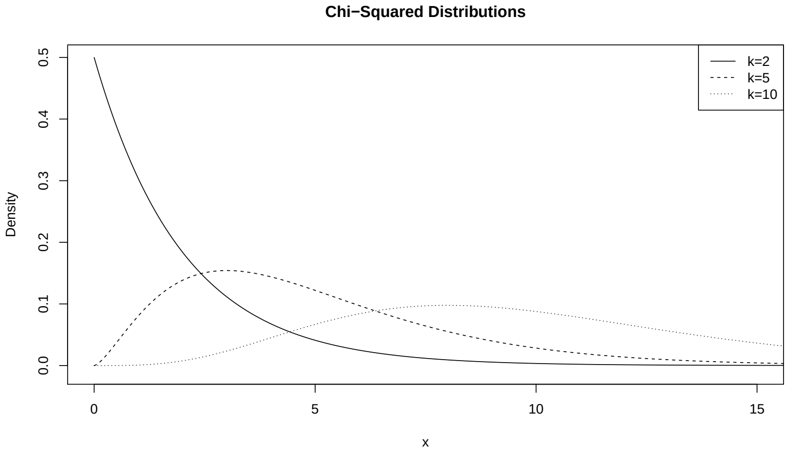
The chi-square (χ^2) is defined as the sum of k squared normal distributions.

Probability distribution:

$$f(x) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2} \text{ where } 0 \leq x < \infty$$

$$E[X] = k \;\&\; Var[X] = 2k$$

Chi-Square Distribution



Chi-Square Distribution

One of the more common uses of the χ^2 distribution is its goodness-of-fit test. When looking at categorical data, it measures the difference between what we would expect to see and what we saw. The results of the test tell us whether our observed values were extreme.

For example, let's say we asked 100 people their favorite soft drink. We received the following responses:

Coke	Cherry Coke	Sprite	Dr. Pepper
27	30	28	15

If all the soft drinks were equally likely, what would we expect to see?

Coke	Cherry Coke	Sprite	Dr. Pepper
25	25	25	25

Chi-Square Distribution

The χ^2 statistic measures how different these (the observed vs. expected) are:

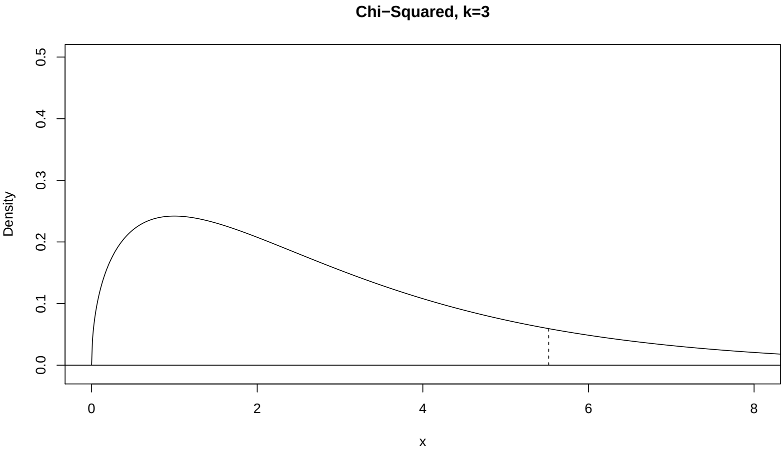
$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(observed - expected)^2}{expected}$$

For our data, this is

$$\frac{(27-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(28-25)^2}{25} + \frac{(15-25)^2}{25} = 5.52$$

The 5.52 is compared to the chi-squared curve of $k-1$ to see if it is an extreme value. If yes, then we think that not all soft drinks are equally preferred. Here $k-1$ is the *degrees of freedom*.

Chi-Squared Statistic
Example



Student's t-distribution

The Student's t-distribution arises when estimating the mean of a normally distributed population when the population variance is unknown.

The probability distribution:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$E[X] = 0$$

$$Var[X] = \frac{\nu}{\nu-2} \text{ for } \nu > 2, \infty \text{ for } 1 < \nu \leq 2, \text{ otherwise undefined}$$

Student's t-distribution

The t-distribution with ν degrees of freedom can be defined as the random variable T defined by

$$T = \frac{Z}{\sqrt{V/\nu}}$$

Where Z is normally distributed with mean 0 and variance 1 ($Z \sim N(0,1)$) and V has a χ^2 distribution with ν degrees of freedom ($V \sim \chi^2_\nu$).

Student's t-distribution

In practice, if $x_1, ..., x_n$ are observed from some normal distribution with mean μ, then the sample mean and variance would be:

$$\bar{x} = \frac{x_1 + ... + x_n}{n}$$

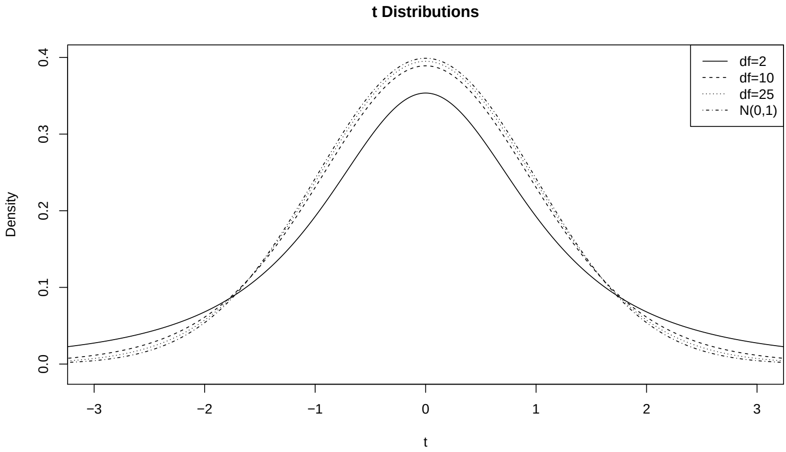
$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

then

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

has a t-distribution with $n-1$ degrees of freedom. Applications of this distribution will be discussed next week!

Student's t-distribution



Exponential Distribution

The *Exponential* distribution is continuous distribution that is somewhat similar to the geometric distribution. Often we think of it as a way to model the time until a 'failure'.

Examples:

- Population Decline
- Radioactive Decay
- How long a patient will live after surgery

This distribution is sometimes also called a type of *survival* function.

Exponential Distribution

The probability distribution

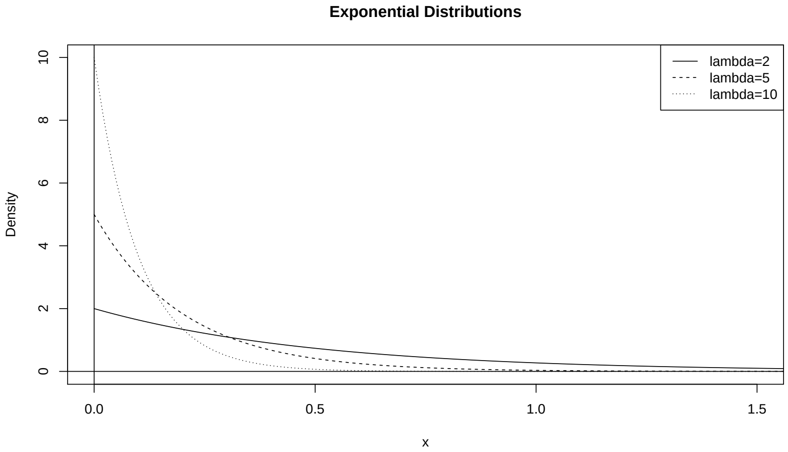
$$f(x) = \lambda e^{-\lambda x}$$

resulting in

$$E[X] = 1/\lambda$$

$$Var[X] = 1/\lambda^2$$

Exponential Distributions



The End

Questions?