Center for Statistics and the Social Sciences Math Camp 2021

Lecture 8: Continuous Distributions

Peter Gao & Jessica Kunke

Department of Statistics
University of Washington

September 16, 2021

Outline

$X \in (a, b)$

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
- 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.

  $$X \in \{1, 2, 3, ..., 10\} \text{ and } P(X = x_i) = \frac{1}{10} \ \ , i = 1, \dots, 10$$

- Let Y be the outcome of a die roll.

  $$Y \in \{1, 2, 3, 4, 5, 6\} \text{ and } P(Y = y_i) = \frac{1}{6} \ \ i = 1, \dots, 6$$

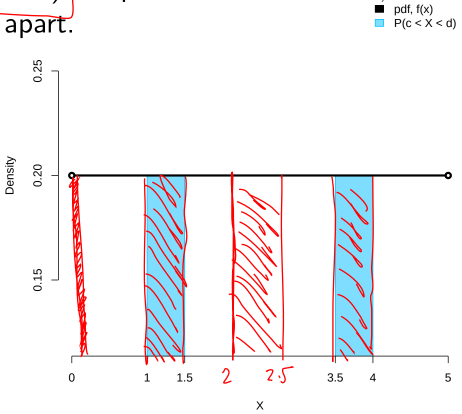
Uniform Distribution

Continuous

$P(X = c) = 0$

A **continuous uniform** random variable takes on values on any continuous interval (a, b) where $a, b \in \mathbb{R}$.

- **RECALL:** $P(X = x) = 0$ for any x when X is continuous
- However, $P(c < X < d) = \frac{d-c}{b-a}$ for all numbers $c, d \in (a, b)$
- $P(c < X < d)$ is equal for all numbers c, d that are the same distance apart.



$X \sim Unif[0, 5]$

Expectations

Continuous Random Variables

Recall our definition of the **expectation** for **discrete** random variables:

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

We can extend this definition to **continuous** distributions. If we divide up the real line into very small intervals, we can estimate $E[X]$ with

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

By letting $n \to \infty$ or the distance between each x_i and x_{i+1} get smaller, i.e. $x_{i+1} - x_i \to 0$, we obtain:

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

like $P(X=x)$ for discrete (pmf)

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

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

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:

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

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

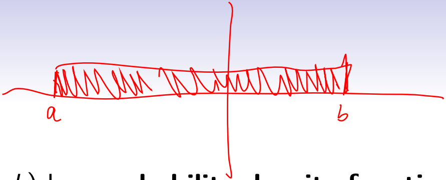
$$Var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx$$

Variance

Continuous Random Variables

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

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



In general, for $X \sim Uniform(a, b)$ has **probability density function (pdf)**

$$\boxed{f(x) = \frac{1}{b-a}}$$

and

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

Example, $X \sim Uniform(0, 5)$.

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

Normal Distribution

The **Gaussian** or **normal** distribution is the most commonly used distribution in statistics.

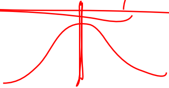
- looks much like a bell curve

- often used to represent ~~large populations~~ population means
  - The larger n gets, the distribution of $\bar{X}$ looks more and more like a normal $\Rightarrow$ confidence intervals.

Probability density function (pdf):

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

Parameters



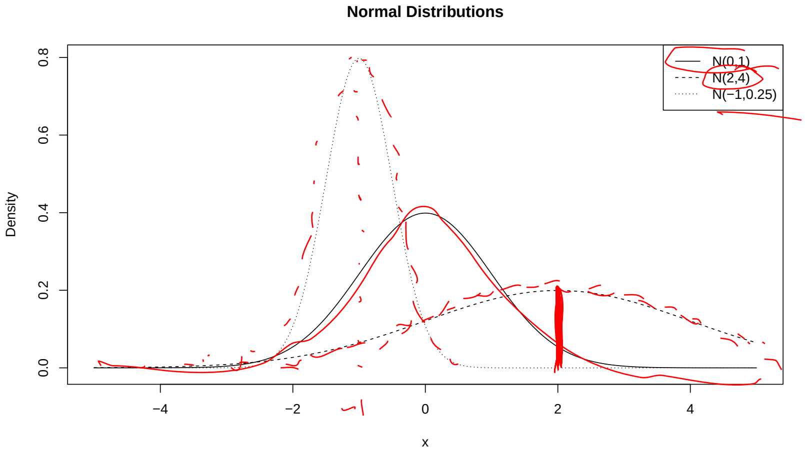
μ

σ^2 determines "spread"

- mean, (μ),
- variance, σ^2,
- **standard deviation**, σ, square root of the variance

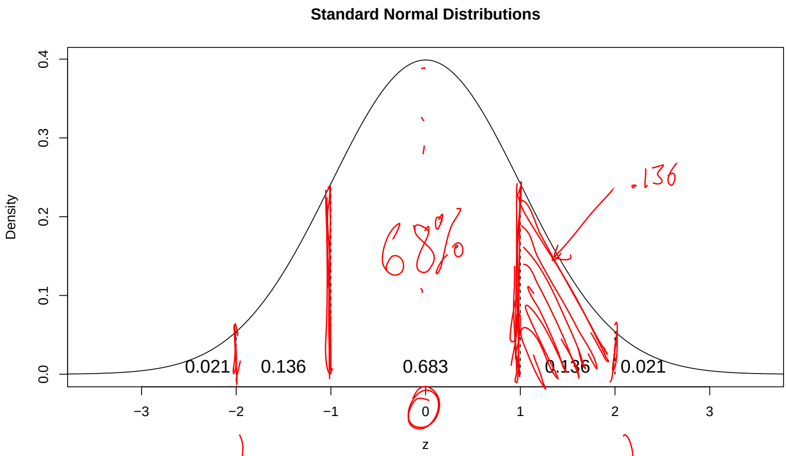
Normal Distributions

$N(\mu, \sigma^2)$



Standard Normal Distribution

N(0, 1)



95% within 2 SD

mean household income of a sample

In order to find $\boxed{P(a < X < b)}$ or the area under $f(x)$ between $x = a$ and $x = b$ 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 $\boxed{X \sim N(\mu, \sigma^2)}$, to **standardize** the distribution to look like a $Z \sim N(0, 1)$ we need to define a random variable

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

$Z \sim N(0,1)$

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

Standard Normal Distribution

$Y = aX + b \Rightarrow E[Y] = aE[X] + b \ Var(Y) = a^2 Var(X)$

If $X \sim N(\mu, \sigma^2)$, then $E[X] = \mu$ and $Var[X] = \sigma^2$.

For $Z = \frac{X-\mu}{\sigma}$,

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

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

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

$$
\begin{aligned}
\boxed{P(5 < X < 7)} &= P(5-\mu < X-\mu < 7-\mu) \\
&= P\left(\frac{5-\mu}{\sigma} < \boxed{\frac{X-\mu}{\sigma}} < \frac{7-\mu}{\sigma}\right) \\
&= P\left(\frac{5-3}{2} < Z < \frac{7-3}{2}\right) \\
&= P(1 < Z < 2) \\
&= 0.136 \text{ [From slide 13, or } \texttt{pnorm(2) - pnorm(1)} \text{ in R.]}
\end{aligned}
$$

$Z \sim N(0,1)$

Chi-Square Distribution

$X \sim N(\mu, \sigma^2)$

$\boxed{Y_1 = (X_1 - \mu)^2}$ $\quad Y_2 = (X_2 - \mu)^2$

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

Probability density function (pdf):

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

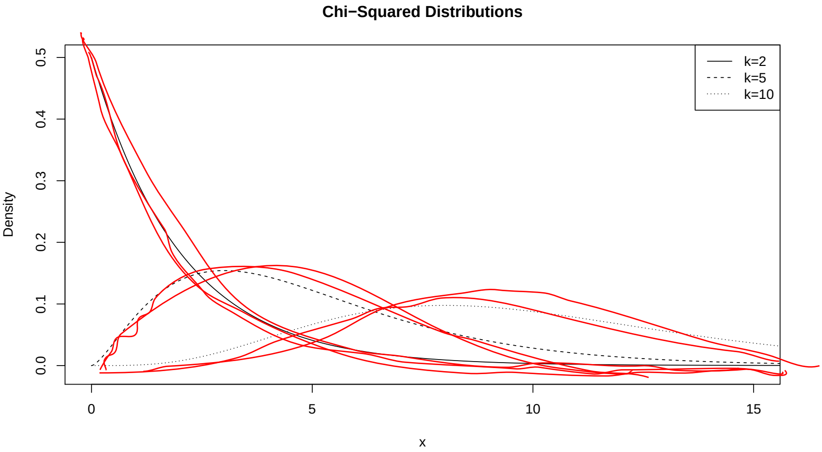
density for sum of standard normals squared.

$$E[X] = k \quad Var[X] = 2k$$

Gamma function

k degrees of freedom

Chi-Square Distribution



Chi-Square Distribution

One of the more common uses of the χ^2 distribution is its **goodness-of-fit test**.

For 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

data

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

Coke	Cherry Coke	Sprite	Dr. Pepper
25	25	25	25

expected

Chi-Square Distribution

The χ^2 statistic measures how different the observed and expected values are:

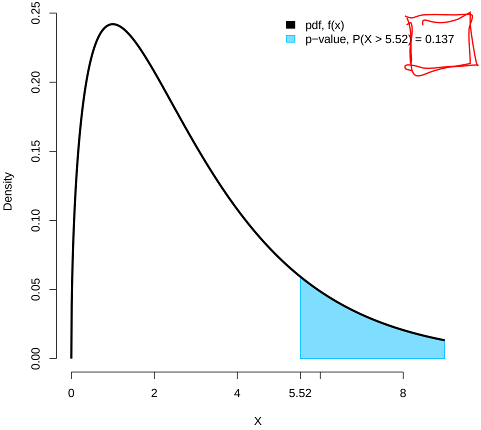
$$\text{test statistic} = \chi^2_{k-1} = \sum_{i=1}^{k} \frac{(observed - expected)^2}{expected}$$

Expected is in terms of the **null hypothesis**. In our example, we're assuming all sodas are equally liked. Our data will either provide enough evidence for us to refute the idea that sodas are equally liked or we will not have enough evidence to reject the hypothesis that sodas are equally liked. For our data, this is

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

Now we compare our test statistic, 5.52, to a χ^2_{k-1} to see if it is an extreme value, i.e. is $P(\chi^2_{k-1} > 5.52)$ small?

- If yes, reject the null hypothesis to conclude not all sodas are equally preferred. If no, cannot reject the null.



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

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

resulting in

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

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

Exponential Distributions