ENV 710

continuous distributions

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

roadmap

- where are we?
- review of continuous distributions
- pod work

descriptive statistics

↓

discrete probability/distributions

↓

continuous probability/distributions

PDF, CDF
uniform distribution
normal distribution

⇣

inference

sampling, central limit theorem, confidence intervals, t-distribution

SUMMARY

uniform distribution

1. $$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b. \\ 0 & \text{for } x < a \text{ or } x > b. \end{cases}$$

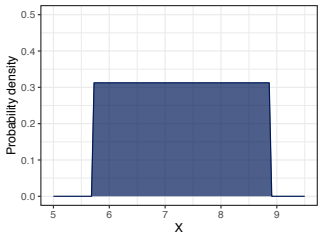
   uniform probability density function for continuous data

2. $$E(X) = (b+a)/2$$
   $$\sigma^2(X) = \frac{(b-a)^2}{12}$$

   mean and variance of uniform distribution

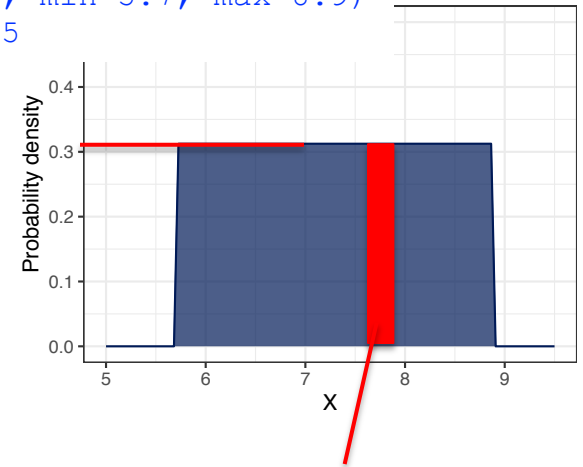
3. `dunif(x, min, max)`
   `punif(q, min, max)`

   `dunif()` calculates density and `punif()` calculates cumulative probability

4. 

   uniform distribution with a minimum of 5.7 and maximum of 8.9

```
dunif(x=7, min=5.7, max=8.9)
[1] 0.3125
```




```
punif(q=7.9, min=5.7, max=8.9)-
punif(q=7.6, min=5.7, max=8.9)
[1] 0.09375

dunif(x=7, min=5.7, max=8.9) *
      (7.9-7.6)
[1] 0.09375
```

SUMMARY

normal distribution

1

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

normal probability density function for continuous data

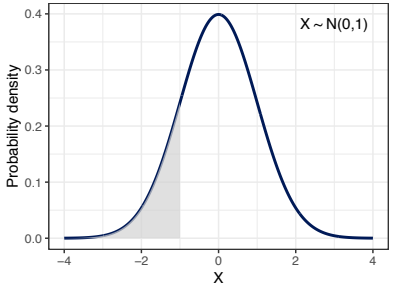
2

```
dnorm(x, mean, sd)
pnorm(q, mean, sd)
```

`dnorm()` calculates probability density

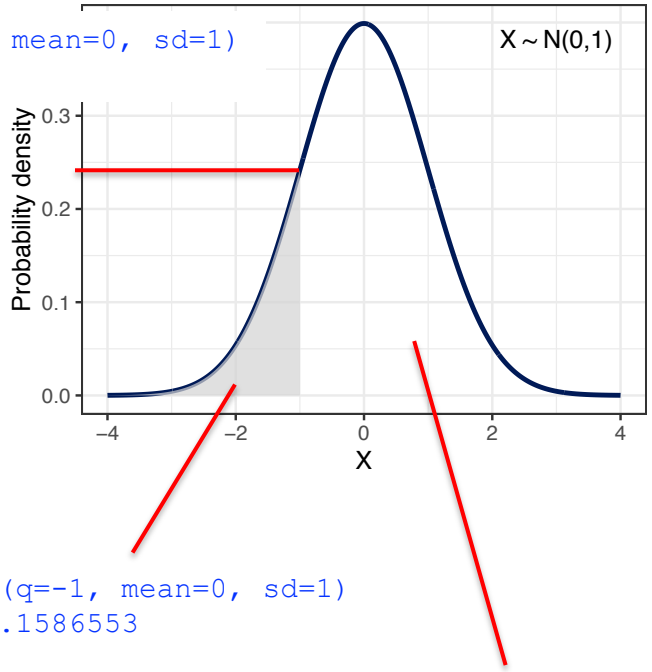
`pnorm()` calculates cumulative probability

3



standard normal distribution (z-distribution) has a mean of 0 and sd of 1

```
dnorm(x=-1, mean=0, sd=1)
[1] 0.242
```



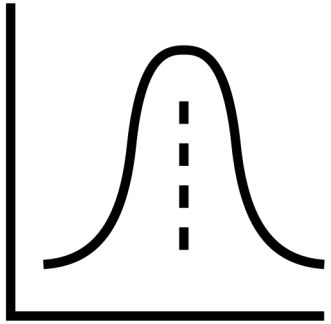

```
pnorm(q=-1, mean=0, sd=1)
[1] 0.1586553
```

```
1-pnorm(q=-1, mean=0, sd=1)
[1] 0.8413447
```

I – summary

Discuss the following terms/subjects (5 min.)

- what are the differences between the discrete and continuous functions/distributions?
- give an example from your work/area of interest of a case when you would use either the uniform or Normal distribution



2 – uniform random variable

Discuss the following and solve the problems together (10 min).

- what is the probability density of finding a plant at any single distance from the woodland border?
- what is the probability of finding plants between 5 and 10 m from the woodland?
- what is the probability of finding plants from 75 to 110 m from the woodland?
- what is the mean, median, mode and variance of the distribution of plants?

a plant species maintains a consistent distance between individuals… and is distributed in a prairie in your study site from 1 to 120 m away from a woodland border.

2 – uniform random variable

Discuss the following and solve the problems together (10 min).

- what is the probability density of finding a plant at any single distance from the woodland border?

1

```
> 1/(120-1)
[1] 0.008403361
```

2

```
> dunif(x = 50, min = 1, max = 120)
[1] 0.008403361
```

3

```
> dunif(x = 6.22453454, min = 1, max = 120)
[1] 0.008403361
```

can insert any value of x in dunif() between 1 and 120 because they all have equal probability

2 – uniform random variable

Discuss the following and solve the problems together (10 min).

- what is the probability of finding plants between 5 and 10 m from the woodland?

```
> 0.008403361 * (10-5)

> dunif(x = 1, min = 1, max = 120) *
(10-5)
[1] 0.04201681

> punif(q=10, min=1, max=120) -
punif(q=5, min=1, max=120)
[1] 0.04201681
```

- what is the probability of finding plants from 75 to 110 m from the woodland?

```
> dunif(x = 1, min = 1, max = 120) * (110-75)
[1] 0.2941176
```

2 – uniform random variable

Discuss the following and solve the problems together (10 min).

- what is the mean, median, and mode of the distribution of plants?

a plant species maintains a consistent distance between individuals… and is distributed in a prairie in your study site from 1 to 120 m away from a woodland border.

1. mean

```
> (120+1)/2
[1] 60.5
```

2. median = mean

3. uniform has no mode because all values have the same probability of occurrence

4. variance

```
> ((120-1)^2)/12
[1] 1180.083
```

3 – random variables

IQ tests are made to have an average score of 100. Psychologists revise the test every few years to maintain 100 as the average. Most people (about 68 percent) have an IQ between 85 and 115 – which corresponds to a σ of roughly 22. A small fraction of people have a very low IQ (below 70) or a very high IQ (above 130).

Several researchers suggest that the concept of intelligence is culturally bound, suggesting IQ tests are biased toward cultures that developed them - white, Western society. Other researchers, suggest that if minority groups score lower on IQ tests, the differences might be real. Discrepancies could be indicative of a weaker educational system, poverty, differences in educational opportunities, etc., which would imply that tests are not, in fact, biased. Others suggest that cultural bias is not the culprit; rather, it could be socioeconomic status.

Conclusion: concept of intelligence, as well as IQ tests, is controversial in scholarly spheres.

3 – random variables

Discuss the following and solve the problems together (5 min).

- how is IQ distributed (i.e., what type of distribution)? how do you know this?
- what is probability density of the distribution for an IQ of 33? what does this density represent?
- what is the highest possible density of the distribution? why?

IQ tests are made to have an average score of 100. Psychologists revise the test every few years to maintain 100 as the average. Most people (about 68 percent) have an IQ between 85 and 115 – which corresponds to a σ of roughly 22. A small fraction of people have a very low IQ (below 70) or a very high IQ (above 130).

3 – random variables

Discuss the following and solve the problems together (5 min).

- how is IQ distributed (i.e., what type of distribution)? how do you know this?

1 distribution is a normal distribution: 68-95-99.7 rule

- what is density of the distribution for an IQ of 33? what does this density represent?

2 density represents the height of the distribution at 33

```
> dnorm(x = 33, mean = 100,  sd = 22)
[1] 0.0001755867
```

- what is the highest possible density of the distribution? why?

3 highest possible density is 0.018 – density at the mean of 100

```
> dnorm(x = 100, mean = 100,  sd = 22)
[1] 0.01813374
```

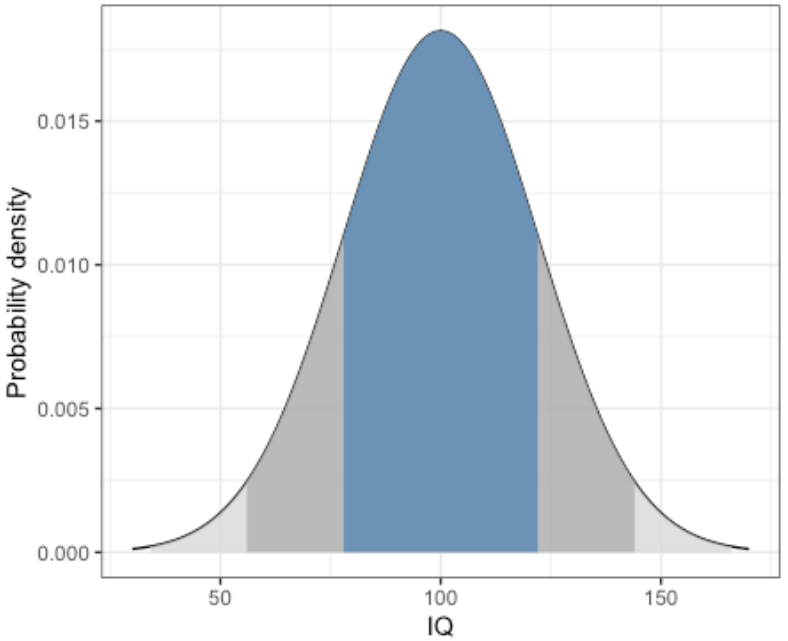
3 – random variables

US ranks 24th globally in IQ

https://www.healthline.com/health/average-iq

Discuss the following and solve the problems together (5 min).

- how is IQ distributed (i.e., what type of distribution)? how do you know this?
- what is density of the distribution for an IQ of 33? what does this density represent?
- what is the highest possible density of the distribution? why?



4 – normal random variable

Discuss the following and solve the problems together (10 min).

- what is the probability of an individual having an IQ between 135 and 140?
- what is the probability of an individual having an IQ less than 70?
- what is the probability of an individual having an IQ of 120 or greater?

IQ tests are made to have an average score of 100. Psychologists revise the test every few years to maintain 100 as the average. Most people (about 68 percent) have an IQ between 85 and 115 – which corresponds to a σ of roughly 22. A small fraction of people have a very low IQ (below 70) or a very high IQ (above 130).

4 – normal random variable

Discuss the following and solve the problems together (10 min).

- what is the probability of an individual having an IQ between 135 and 140?

```
> pnorm(q=140, mean=100, sd=22) - pnorm(q=135, mean=100, sd=22)
[1] 0.02129684
```

4 – normal random variable

Discuss the following and solve the problems together (10 min).

- what is the probability of an individual having an IQ of less than 70?

```
> pnorm(q=69, mean=100, sd=22)
[1] 0.07940414
```

```
> pnorm(70, 100, 22, lower.tail = T)
[1] 0.08634102
```

```
> 1-pnorm(q=70, mean=100, sd=22, lower.tail=F)
[1] 0.08634102
```

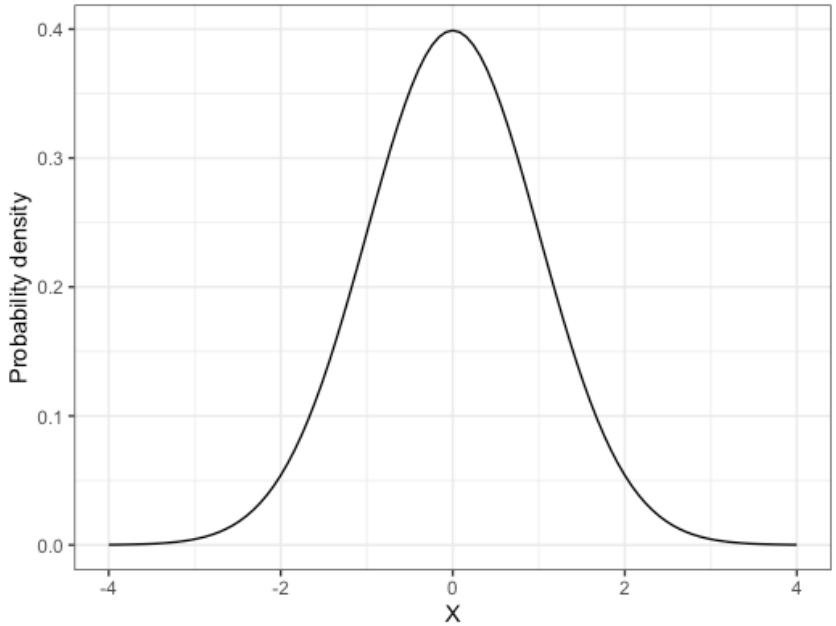
- what is the probability of an individual having an IQ of 120 or greater?

```
> 1 - pnorm(q=120, mean=100, sd=22)
[1] 0.1816511
```

```
> pnorm(q=120, mean=100, sd=22, lower.tail = F)
[1] 0.1816511
```

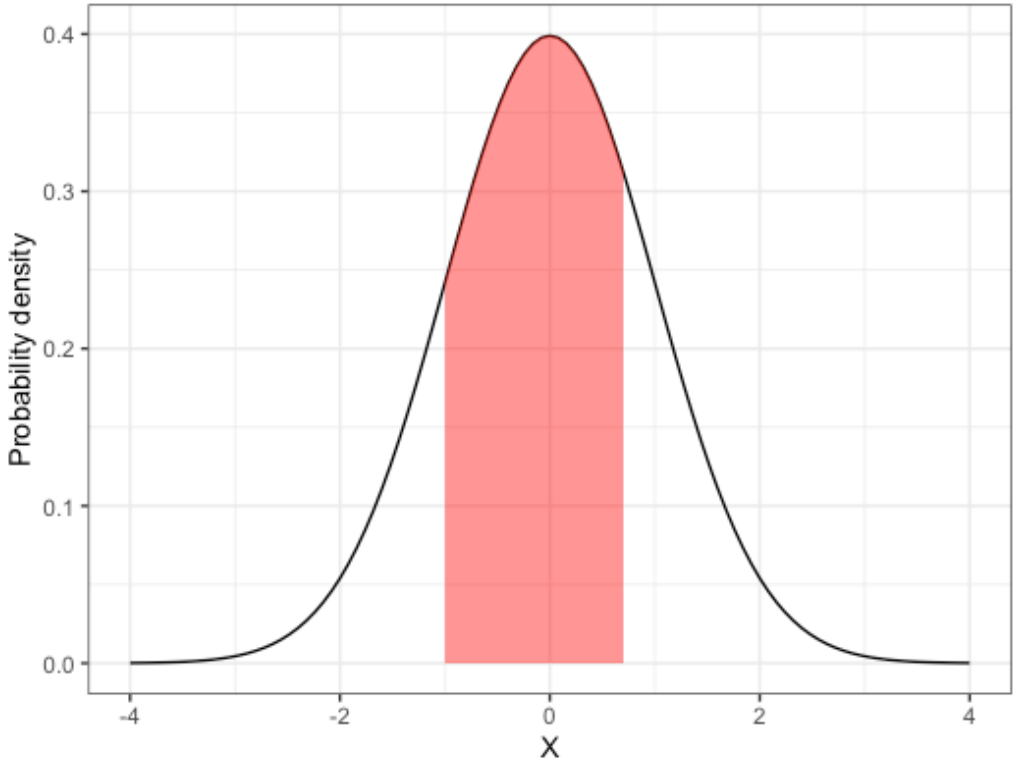
4 – plotting distribution

```
my_norm <- ggplot(data = data.frame(x = c(-4, 4)), aes(x)) +
            stat_function(fun = dnorm, args = list(mean = 0, sd = 1)) +
             ylab("Probability density") + xlab("X") +
            theme_bw()
my_norm
```



4 – plotting distribution

```
my_norm +
stat_function(fun = dnorm, args =
     list(mean = 0, sd = 1),
      xlim = c(-1, 0.7),
       geom = "area",
        fill = "red", alpha = 0.5)
```



Questions?