



# Lecture 2

Probability refresher (or introduction)

WILD6900 (Spring 2020)

# Readings:

| Hobbs & Hooten 29-70

**Warning:** The material presented in this lecture is tedious. But the concepts in this lecture are critical to everything that will follow in this course. So push through and try your best to understand these topics. You do not need to be an expert in probability at the end of this lecture - we will reinforce these concepts over and over again throughout the semester - but getting the gist now will help you grasp other topics as we move forward

# Stochasticity and uncertainty in ecological models

In each level of our models, we differentiate between:

- a *deterministic* model  $g()$ , and
- a *stochastic* model  $[a|b, c]$

The deterministic portion of the model contains no uncertainty <sup>1</sup>

*Stochastic* processes are different:

- given an input, the model will not always return the same answer
- the output of stochastic processes are *uncertain*
- Even though stochastic processes are inherently uncertain, they are not unpredictable.

# Stochasticity and uncertainty in ecological models

In Bayesian models, all unobserved quantities are treated as a **random variables**, that is they can take on different values due to chance (i.e., stochastic)

Each random variable in our model is governed by a **probabilty distribution** <sup>2</sup>

Our goal is to *use our data to learn about those distributions*

# Probability

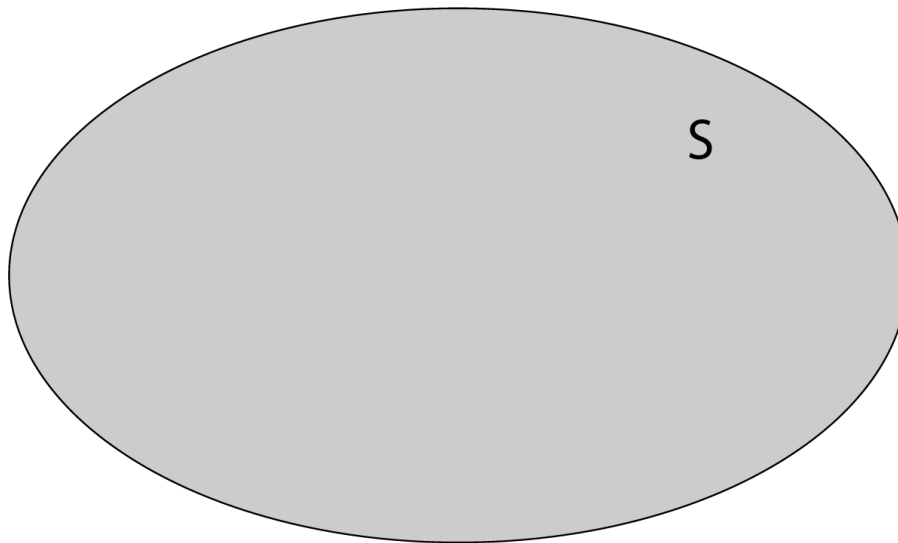
Uncertain events are not necessarily unpredictable

*Probability* allows us to summarize how likely each possible value of a random variable is to occur

# Sample space

For any given random variable, the *sample space*  $s$  includes all of the possible values the variable can take

For example, for an single-species occupancy model,  $S$  would be present or absent. For a model of species abundance,  $S$  would be  $0, 1, 2, 3, \dots, \infty$ .



# Sample space

## Example

Imagine an occupancy model in which we want to know if species  $x$  is present at a given location

We will denote the occupancy status  $z_x$  and the sample space includes just two possible values:

$$S_{z_x} = \{0, 1\}$$

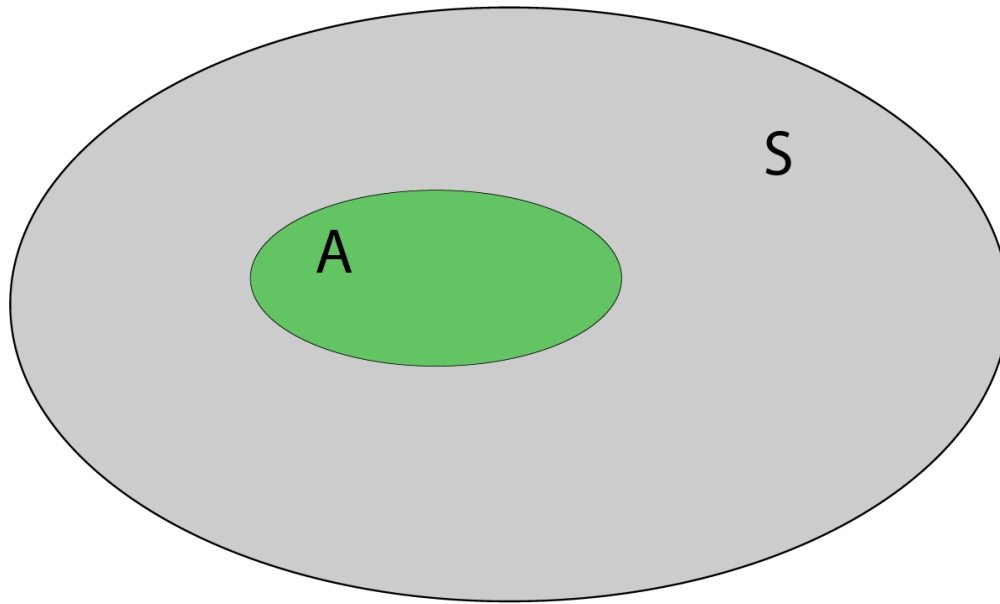


# Probability of single events

# Probability of single events

The probability that  $A$  will occur is the area of  $A$  divided by the area of  $S$ <sup>3</sup>

$$Pr(A) = \frac{\text{area of } A}{\text{area of } S}$$



# Probability of single events

## Example

In our example, let's say that the probability of occupancy for species  $x$  is:

$$Pr(z_x = 1) = 0.4$$

This means that the probability that the site is not occupied is:

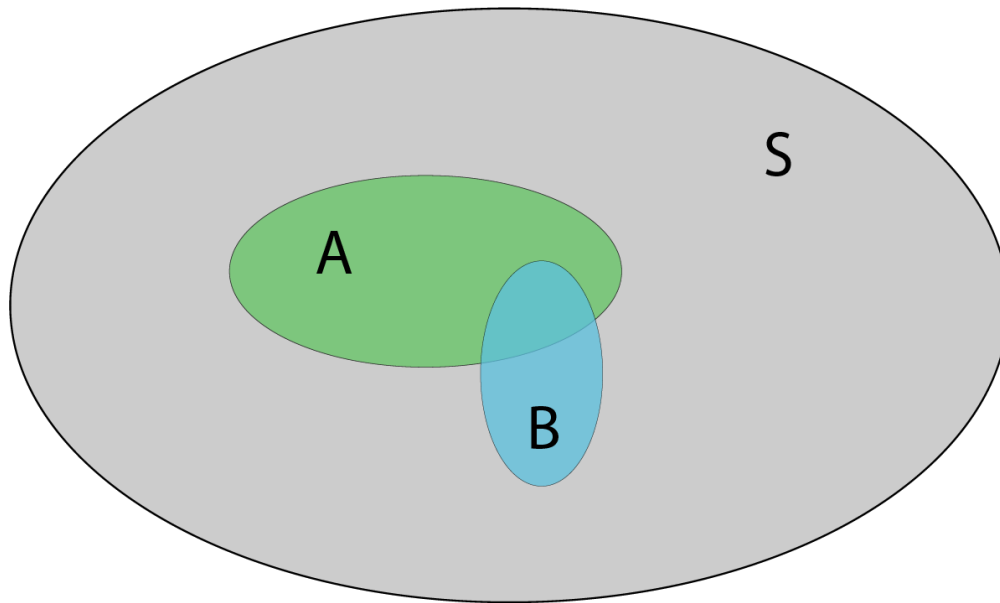
$$Pr(z_x = 0) = 1 - 0.4 = 0.6$$

# Probability of multiple events

# Probability of multiple events

Often, we are not interested in the probability of a single event happening but instead of more than one events

The *joint* probability refers to the probability that two or more events occur and is usually denoted  $Pr(A, B)$  <sup>4</sup>



# Probability of multiple events

## Example

To extend our simple example, let's imagine we are interested in the occupancy status of two species -  $x$  and  $y$ . Our sample space is now:

$$S_{z_x, z_y} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

The question we want to know now is: what is the probability that a site is occupied by both species:

$$Pr(z_x = 1, z_y = 1) = Pr(z_x, z_y)$$

The answer to that question depends on the relationship between  $Pr(z_x)$  and  $Pr(z_y)$

# Conditional probability

# Conditional probability

In some cases, knowing the status of one random variable tells us something about the status of another random variable

Let's say we know that species  $x$  is present, that is  $z_x = 1$

Knowing that  $z_x = 1$  does two things:

- 1) It shrinks the possible range of sample space (if  $z_x = 1$  occurred, the remainder of our sample space (in this case  $z_x = 0$ ) did not occur)
- 2) It effectively shrinks the area  $z_y$  - we know that the area of  $z_y$  outside of  $z_x$  didn't occur

You can see this very clearly in this [awesome visualization](#)



# Conditional probability

$Pr(z_y|z_x)$  is the area shared by the two events divided by the area of  $z_x$  (not  $S$ !) <sup>5</sup>

$$Pr(z_y|z_x) = \frac{\text{area shared by } z_x \text{ and } z_y}{\text{area of } z_x} = \frac{Pr(z_x \cap z_y)}{Pr(z_x)}$$

likewise,

$$Pr(z_x|z_y) = \frac{Pr(z_x \cap z_y)}{Pr(z_y)}$$

# Conditional probability

For conditional events, the joint probability is:

$$Pr(z_y, z_x) = Pr(z_y|z_x)Pr(z_x) = Pr(z_x|z_y)Pr(z_y)$$

# Probability of independent events

In some cases, the probability of one event occurring is *independent* of whether or not the other event occurs <sup>6</sup>

In our example, the occupancy of the two species may be totally unrelated

- if they occur together, it happens by complete chance <sup>7</sup>

In this case, knowing that  $z_x = 1$  gives us no new information about the probability of  $z_y = 1$

Mathematically, this means that:

$$Pr(z_y|z_x) = Pr(z_y)$$

and

$$Pr(z_x|z_y) = Pr(z_x)$$

Thus,

$$Pr(z_x, z_y) = Pr(z_x)Pr(z_y)$$

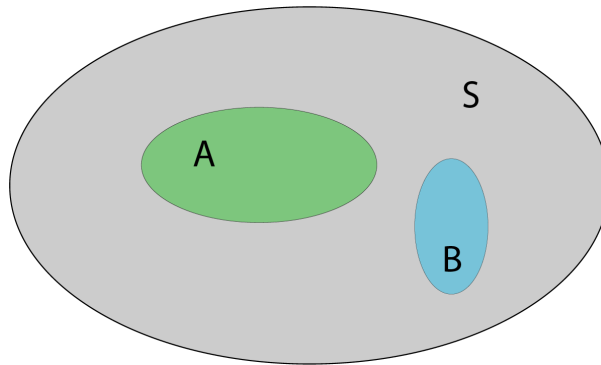
# Disjoint events

A special case of conditional probability occurs when events are *disjoint*

In our example, maybe species  $x$  and species  $y$  **never** occur together <sup>8</sup>

In this case, knowing that  $z_x = 1$  means that  $z_y = 0$ . In other words,

$$Pr(z_y | z_x) = Pr(z_x | z_y) = 0$$



# Probability of one event or the other

In some cases, we might want to know the probability that one event *or* the other occurs

For example, what is the probability that species  $x$  or species  $y$  is present but not both?

This is the area in  $z_x$  and  $z_y$  not including the area of overlap:

$$Pr(z_x \cup z_y) = Pr(z_x) + Pr(z_y) - Pr(z_x, z_y)$$

# Probability of one event or the other

When  $z_x$  and  $z_y$  are independent,

$$Pr(z_x \cup z_y) = Pr(z_x) + Pr(z_y) - Pr(z_x)Pr(z_y)$$

If they are conditional,

$$Pr(z_x \cup z_y) = Pr(z_x) + Pr(z_y) - Pr(z_x|z_y)Pr(z_y) = Pr(z_x) + Pr(z_y) - Pr(z_y|z_x)Pr(z_x)$$

If they are disjoint,

$$Pr(z_x \cup z_y) = Pr(z_x) + Pr(z_y)$$

# Marginal probability

A critical concept in Bayesian models is **marginal probability**, that is the probability of one event happening regardless of the state of other events

Imagine that our occupancy model includes the effect of 3 different habitats on the occupancy probability of species  $x$ , so:

$$Pr(z_x | H_i) = \frac{Pr(z_x \cap H_i)}{Pr(H_i)}$$

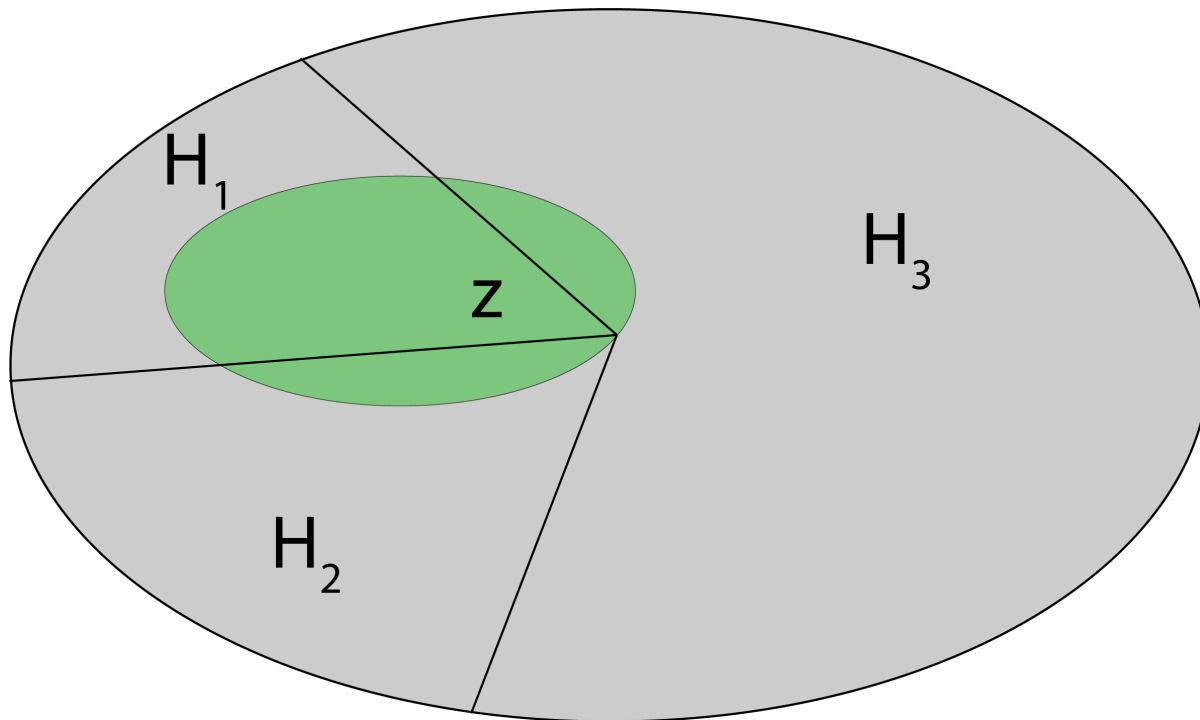
What is the overall probability that species  $x$  occurs regardless of habitat type? That is,  $Pr(z_x)$ ?

In this case, we *marginalize* over the different habitat types by summing the conditional probabilities weighted by probability of each  $H_i$ :

$$Pr(z_x) = \sum_{i=1}^3 Pr(z_x | H_i) Pr(H_i)$$

Think of this as a weighted average - the probability that  $z_x = 1$  in each habitat type weighted by the probability that each habitat type occurs.

# Marginal probability





# Marginal probability

	H1	H2	H3	Total
Occupied	60	10	10	80
Unoccupied	20	70	250	340
Total	80	80	260	420

# Factoring joint probabilities

# Factoring joint probabilities

Many of the models you will work with as an ecologist will contain multiple random variables

$$\left[ z, \theta_p, \theta_o, \sigma_p^2, \sigma_s^2, \sigma_o^2, u_i | y_i \right] \propto [y_i | d(\Theta_o, u_i), \sigma_o^2] [u_i | z, \sigma_s^2] [z | g(\theta_p, x), \sigma_p^2] [\theta_p] [\theta_o] [\sigma_p^2] [\sigma_s^2] [\sigma_o^2]$$

The rules of probability allow us to express the complex *joint* probabilities as a series of more simple *conditional* probabilities

- These concepts may feel a little abstract now but they will be very important when we learn about implementing Bayesian models

Determining the dependences between parameters in the models is aided by **Bayesian network models**

# Bayesian networks

Bayesian networks graphically display the dependence among random variables

- Random variables are *nodes*
- Arrows point from *parents* to *children*



# Bayesian networks

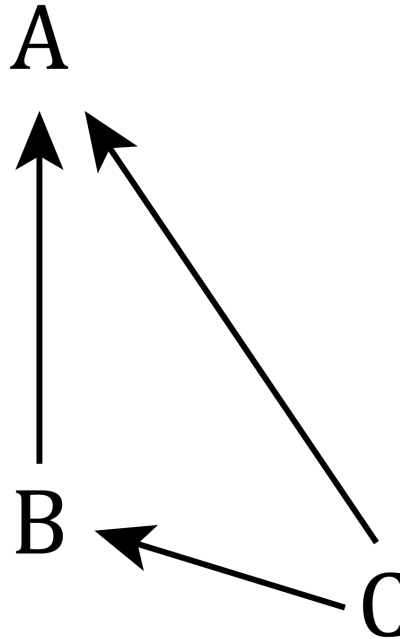
Bayesian networks graphically display the dependence among random variables <sup>9</sup>

- Children nodes are on the LHS of conditioning symbols
- Parent nodes are on the RHS of conditioning symbols
- Nodes without a parent are expressed unconditionally



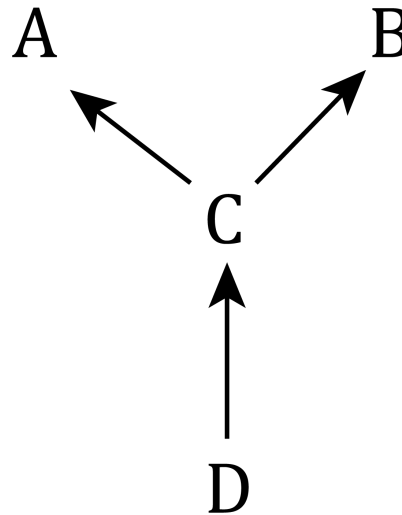
$$Pr(A, B) = Pr(A|B)Pr(B)$$

# Bayesian networks



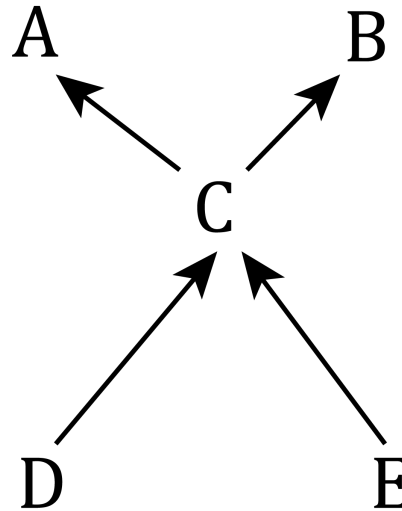
$$Pr(A, B, C) = Pr(A|B, C)Pr(B|C)Pr(C)$$

# Bayesian networks



$$Pr(A, B, C, D) = Pr(A|C)Pr(B|C)Pr(C|D)Pr(D)$$

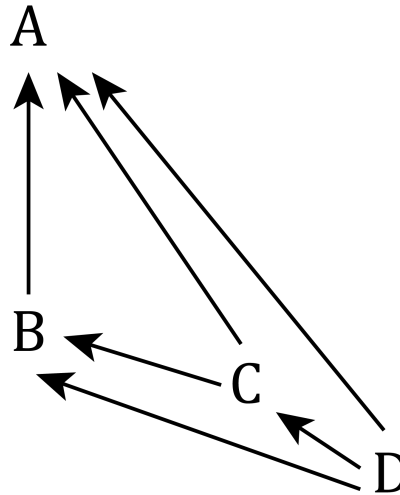
# Bayesian networks



$$Pr(A, B, C, D, E) = Pr(A|C)Pr(B|C)Pr(C|D, E)Pr(D)Pr(E)$$

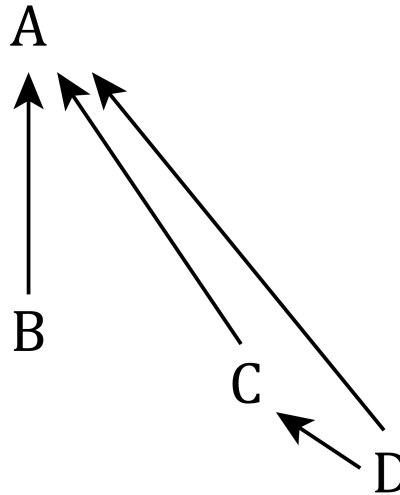


# Bayesian networks



$$Pr(A, B, C, D) = Pr(A|B, C, D)Pr(B|C, D)Pr(C|D)Pr(D)$$

# Bayesian networks



$$Pr(A, B, C, D) = Pr(A|B, C, D)Pr(C|D)Pr(D)Pr(B)$$

# Properties of probability distributions

Because all unobserved quantities are treated as random variables governed by probability distributions, using and understanding Bayesian methods requires understanding probability distributions.

As ecologists, there are a number of very common probability distributions that we encounter and use regularly <sup>10</sup>:

- normal
- Poisson
- binomial
- gamma

# Discrete vs. continuous distributions

**Continuous** random variables can take on an infinite number of values on a specific interval <sup>11</sup>

- Normal ( $-\infty$  to  $\infty$ )
- Gamma (0 to  $\infty$ )
- Beta (0 to 1)
- Uniform (? to ?)

**Discrete** random variables are those that take on distinct values, usually integers <sup>12</sup>

- Poisson (integers  $\geq 0$ )
- Bernoulli (0 or 1)
- Binomial
- Multinomial

# Probability functions

Very often we want to know the probability that a random variable will take a specific value  $z$

Answering this question requires the use of probability functions, which we will denote  $[z]^{13}$

Probability functions differ between *continuous* and *discrete* distributions so we will discuss these separately

# Probability mass functions

For *discrete* random variables, the probability that the variable will take a specific value  $z$  is defined by the *probability mass function* (pmf)

All pmf's share two properties:

$$0 \leq [z] \leq 1$$

$$\sum_{z \in S} [z] = 1$$

where  $S$  is the set of all  $z$  for which  $[z] > 0$  (the range of possible values of  $z$ ).

# Probability mass functions

As an example, let's assume a random variable that follows a Poisson distribution

- Poisson random variables can take any integer value  $> 0$  ( $0, 1, 2, \dots$ )
- e.g., the number of individuals at a site or the number of seeds produced by a flower

The shape of the Poisson distribution is determined by 1 parameter called  $\lambda$

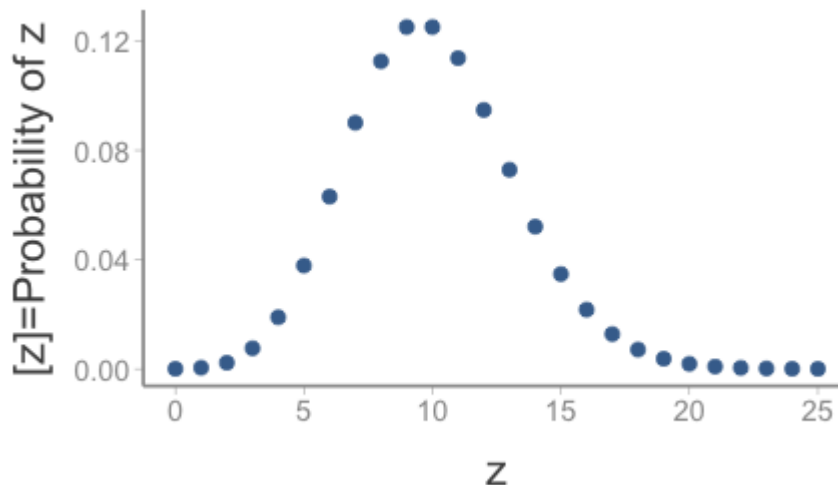
- $\lambda$  is the expected value (the most likely value) of a random variable generated from the Poisson distribution
- larger  $\lambda$  means larger values of the variable

# Probability mass functions

If  $\lambda = 10$ , what is the probability that  $z$  will equal 10? Or 8? Or 15?

In R, probability mass is estimating using the `dpois()` function (or the equivalent for other discrete distributions)

- takes two arguments: the value we are interested in estimating the probability of ( $z$ )<sup>14</sup> and the expected value of our distribution ( $\lambda$ )
- `dpois(x = seq(0,25), lambda = 10)`



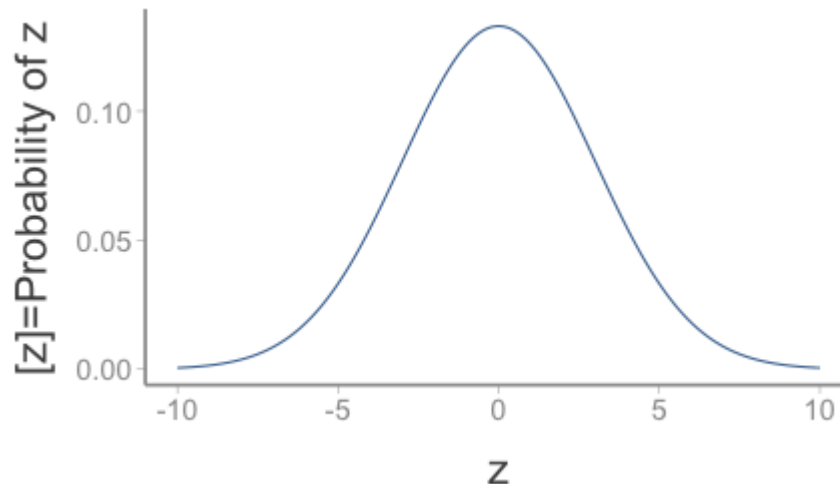


# Probability density functions

Probability mass functions provide the probability that a discrete random variable takes on a specific value  $z$

For continuous variables, estimating probabilities is a little trickier because  $Pr(z) = 0$  for any specific value  $z$

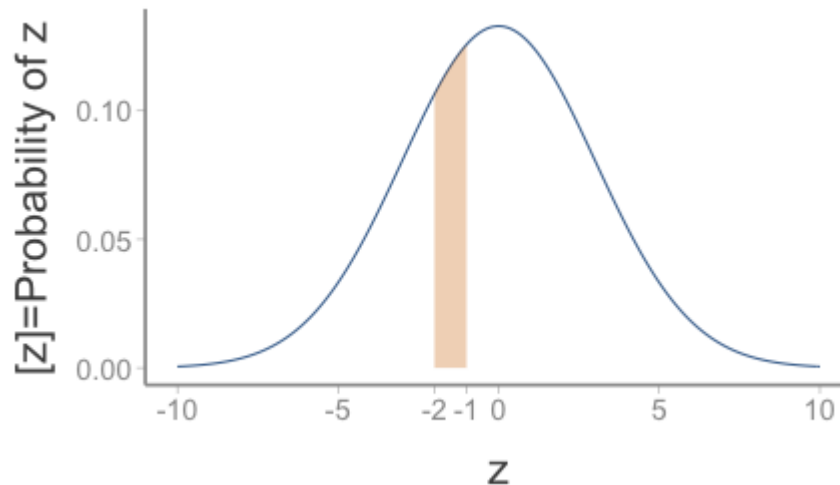
Why? Let's look at the probability distribution for a normal random variable with mean  $= 0$  and standard deviation  $= 3$ :



# Probability density functions

The *probability density* is the area under the curve for an interval between  $a$  and  $b$ , which we'll call  $\Delta_z = (a - b)$ .

For example, the shaded area below shows the probability density  $Pr(-2 \leq z \leq -1)$ :

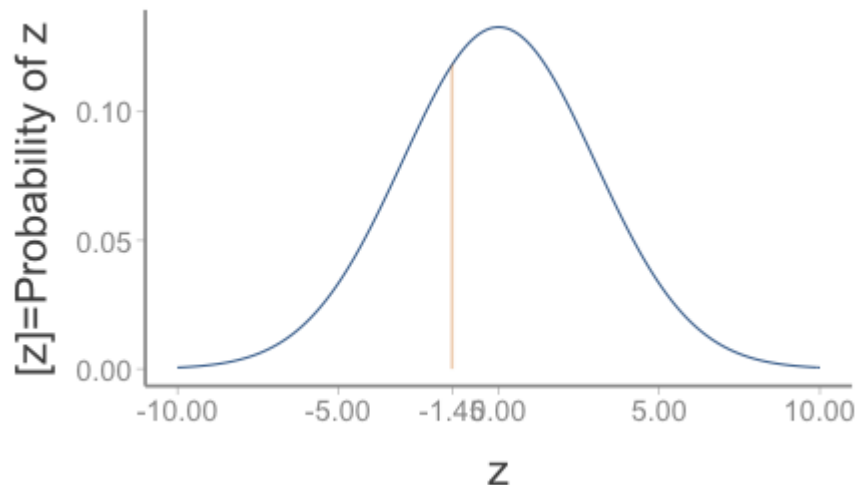


# Probability density functions

This area can be approximated by multiplying the width times the (average) height of the rectangle:

$$Pr(a \leq z \leq b) \approx \Delta_z [(a + b)/2]$$

By making the range  $\Delta_z = a - b$  smaller and smaller, we get closer to  $Pr(z)$ :



# Probability density functions

At  $z$ ,  $\Delta_z = 0$ , thus  $[z] = 0$

However, we can use calculus to estimate the height of the line ( $[z]$ ) as  $\Delta_z$  approaches 0

So for continuous random variables, the *probability density* of  $z$  as the area under the curve between  $a \leq z \leq b$  as  $\Delta_z$  approaches zero

Estimating probability density in R is the same as for discrete variables:

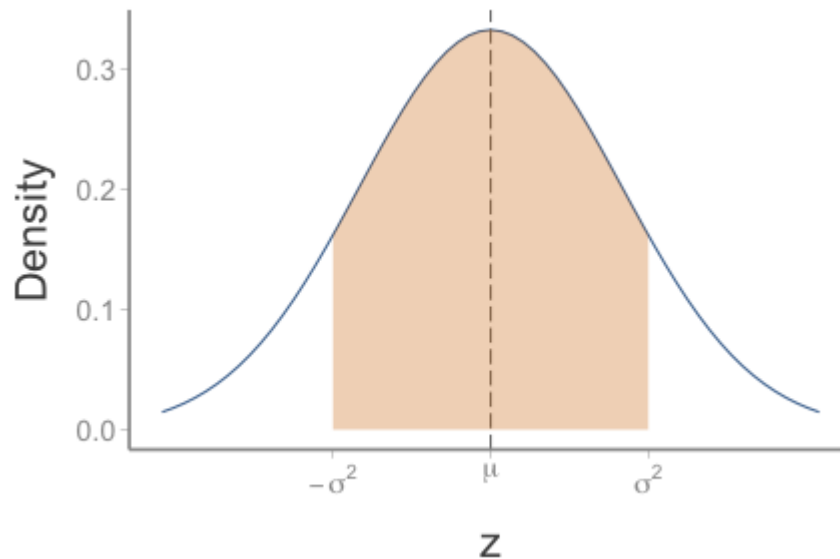
`dnorm()` <sup>15</sup>

# Moments

# Moments

Every probability distribution we will use in the course can be described by its *moments*

- 1<sup>st</sup> moment is the expected value (i.e., mean)
- 2<sup>nd</sup> moment is the variance



# Expected value (i.e., the mean)

The first moment of a distribution describes its central tendency (denoted  $\mu$ ) or *expected value* (denoted  $E(z)$ )

This is the most probable value of  $z$

Think of this as a weighted average - the mean of all possible values of  $z$  weighted by the probability mass or density of each value  $[z]$

For discrete variables, the first moment can be calculated as

$$\mu = E(z) = \sum_{z \in S} z[z]$$

For continuous variables, we need to use an integral instead of a sum:

$$\mu = E(z) = \int_{-\infty}^{\infty} z[z] dz$$

# Variance

The second moment of a distribution describes the *variance* - that is, the spread of the distribution around its mean

On average how far is a random value drawn from the distribution from the mean of the distribution

For discrete variables, variance can be estimated as the weighted average of the squared difference (squared to prevent negative values) between each value  $z$  and the mean  $\mu$  of the distribution:

$$\sigma^2 = E((z - \mu)^2) = \sum_{z \in S} (z - \mu)^2 [z]$$

and for continuous variables:

$$\sigma^2 = E((z - \mu)^2) = \int_{-\infty}^{\infty} (z - \mu)^2 [z] dz$$



# Exercise: Estimating moments using *Monte Carlo integration* <sup>16</sup>

One way to estimate moments is by simulating a large number of values from a probability distribution and then using these samples to calculate the first and second moments <sup>17</sup>

This approach is very easy to do in **R** using the **r** class of functions (e.g., **rnorm()**, **rpois()**, etc.)

- These functions generate specified number of random draws (**r** for random) from a given probability distribution.

# Exercise: Estimating moments using *Monte Carlo integration*

Let's estimate the first and second moments of a gamma distribution <sup>18</sup>

The shape of the gamma distribution is governed by two parameters,  $\alpha$  (referred to as the shape) and  $\beta$  (referred to as the rate or sometimes the scale) <sup>19</sup>

In **R**, we can generate and visualize a large number (e.g., 10000) random draws from the gamma distribution using the following code:

```
n <- 10000 # Sample size  
samp <- rgamma(n, shape = 0.5, rate = 2)
```

# Exercise: Estimating moments using *Monte Carlo integration*

Now let's use these sample to estimate the first moment (the mean) and the second moment (the variance) of the distribution

We estimate the first moment by taking the arithmetic mean of our samples ( $\frac{1}{n} \sum_{i=1}^n z_i$ ) and the variance as ( $\frac{1}{n} \sum_{i=1}^n (z_i - \mu)^2$ ):

```
mu <- sum(samp)/n # mean of the sample  
sigma2 <- sum((samp - mu)^2)/n # variance of the sample
```

# Exercise: Estimating moments using *Monte Carlo integration*

How close are these values to the true moments? For the gamma distribution:

$$\mu = \frac{\alpha}{\beta}$$

$$\sigma^2 = \frac{\alpha}{\beta}$$

For our samples: <sup>1</sup>

```
mu # Estimated mean
```

```
## [1] 0.2536
```

```
0.5/2 # True mean
```

```
## [1] 0.25
```

# Exercise: Estimating moments using *Monte Carlo integration*

How close are these values to the true moments? For the gamma distribution:

$$\mu = \frac{\alpha}{\beta}$$

$$\sigma^2 = \frac{\alpha}{\beta^1}$$

For our distribution:

```
sigma2 # Estimated variance
```

```
## [1] 0.1339
```

```
0.5/2^2
```

```
## [1] 0.125
```

# Exercise: Estimating moments using *Monte Carlo integration*

Try this on your own - simulate data from a Poisson distribution and see if the moments you estimate from the sample are close to the true moments

**Hint** - the Poisson distribution has a single parameter  $\lambda$ , which is both the mean and the variance of the distribution

Change both  $\lambda$  and  $n$ . *Does varying these values change how well your sample estimates the moments?*<sup>20</sup>

# Moment matching

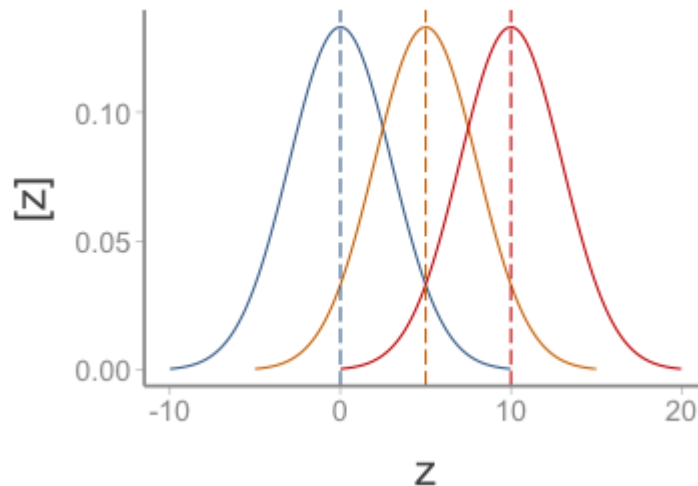
What if you know the mean and variance of a distribution and need the parameters?

Rather than using simulation, each distribution has a set of formulas for converting between parameters and moments (called *moment matching*)

Moment matching is very important because often we have the mean and variance of distributions but need to convert those summaries into the parameters of the underlying distribution <sup>21,22</sup>

# Moment matching

For the normal distribution, it is relatively easy to understand moments because the parameters of the distribution (mean and standard deviation) *are* the first and second moments





# Moment matching

The normal distribution has an interesting property - you can change the first moment without changing the second moment

This is not true of all probability distributions

For example, the beta distribution is a continuous distribution with values between 0 and 1<sup>23,24</sup>. Its first and second moments are:

$$\mu = \frac{\alpha}{\alpha + \beta}$$

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

# Moment matching

What if you know the mean of a beta distribution is 0.3 and the variance is 0.025? What are  $\alpha$  and  $\beta$ ?

$$\alpha = \left( \frac{1 - \mu}{\sigma^2} - \frac{1}{\mu} \right) \mu^2$$

$$\beta = \alpha \left( \frac{1}{\mu} - 1 \right)$$

For our model, that means <sup>25</sup>:

```
(alpha <- ( (1 - 0.3)/0.025 - (1/0.3) ) * 0.3^2)
```

```
## [1] 2.22
```

```
(beta <- alpha * ( (1/0.3) - 1 ))
```

```
## [1] 5.18
```