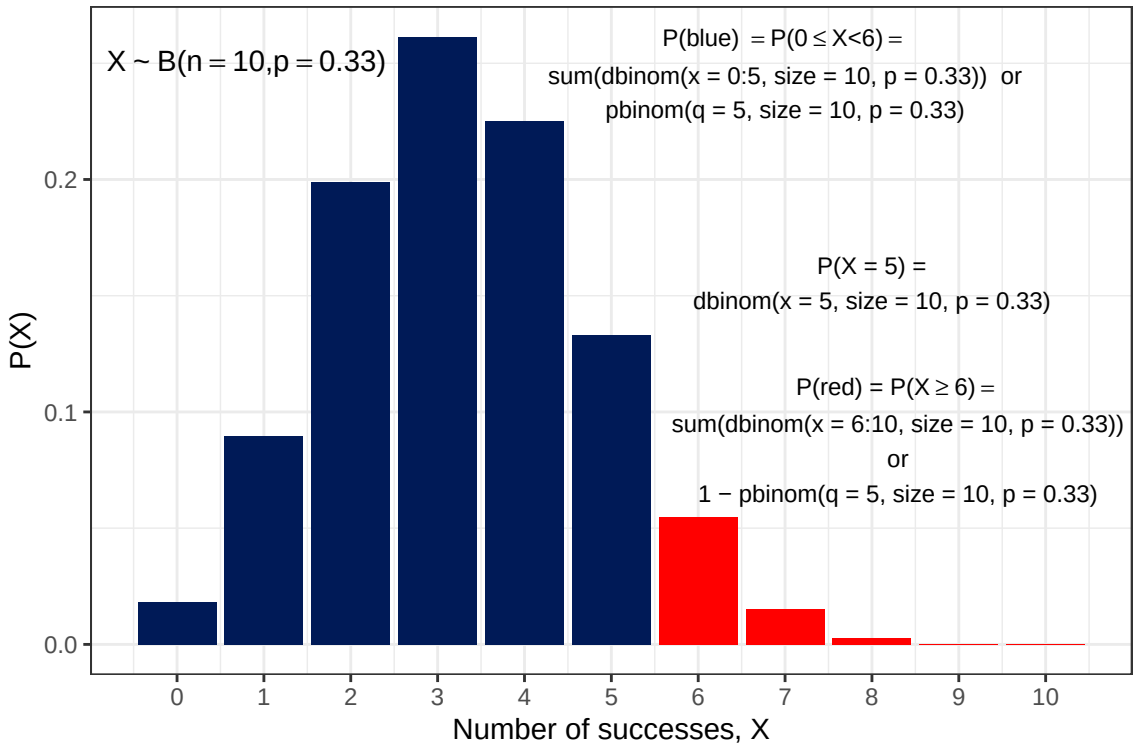
Discrete and Continuous Distributions

Discrete Distributions

In discrete distributions like the binomial or Poisson distributions, we can calculate the probabilities of obtaining specific integer values. For example, consider the probability of going fishing and catching a brown trout in ten casts if the probability of catching a brown trout is 33% per cast. For each cast we either catch (success) or don't catch (failure) a fish - a Bernouilli trial.



To find **specific values** for a discrete probability function, we use `dbinom`. For example, the probability of getting a 5 (catching 5 fish in 10 casts) in this example would be:

`dbinom(x = 5, size = 10, prob = 0.33) = 0.133`.

To get a **cumulative probability** like $P(0 \leq X < 6)$ – the blue bars in the above figure – we could either use `dbinom` to sum across the probabilities of each value, 0 through 5, or use `pbinom` to find the cumulative probability:

`sum(dbinom(x = 0:5, size = 10, prob = 0.33)) = 0.927`

`pbinom(q = 5, size = 10, prob = 0.33) = 0.927`.

The probability of catching 5 or fewer fish in 10 casts is very high, 92.7%, if the probability of catching a fish is 33% on each cast.

If we want to simulate 10 casts (size = 10) and see the number of fish we would catch, we could do it using `rbinom`. `rbinom` generates random numbers from the Binomial distribution (similarly, `rnorm` would generate random numbers from a normal distribution) with the specific number of trials and the probability of success.

```
rbinom(n = 1, size = 10, prob = 0.33).
```

If you run this line of code a few times, the number of fish you 'catch' will vary between 0 and 10. If you run this line a whole bunch of times, say 1000, on average you should catch: $E(X) = p \cdot n = 0.33 \cdot 10 = 3.3$ fish.

The following line of code takes the average of 1000 simulations of casting 10 tens and should produce a result of approximately 3.3 fish.

```
mean(rbinom(n = 1000, size = 10, prob = 0.33))
```

Continuous Distributions

Like discrete distributions, we use continuous distributions like the Uniform and Normal distributions to estimate the probability of getting values or a range of values. However, because the distribution is continuous we cannot calculate the probability of any specific number. Take for example, a value of -2 on the x-axis (see below figure):

```
dnorm(x = -2, mean = 0, sd = 1) = 0.054.
```

The line of code using `dnorm` determines the height of the curve for a x-value of -2, which is 0.054 (see the red line of Fig. 2). Again, this is the density or height of the curve at that point, not its probability. We could *estimate* the probability of getting a -2 by finding the area under the curve between -1.99999 and -2:

```
dnorm(x = -1.99999, mean = 0, sd = 1) - dnorm(x = -2, mean = 0, sd = 1)
```

This probability is very small, $1.08 \cdot 10^{-6}$, and would get smaller and smaller if we tried to estimate it with more precision. The point is that *on continuous distributions we find probabilities for area under the curve, not for individual numbers.*

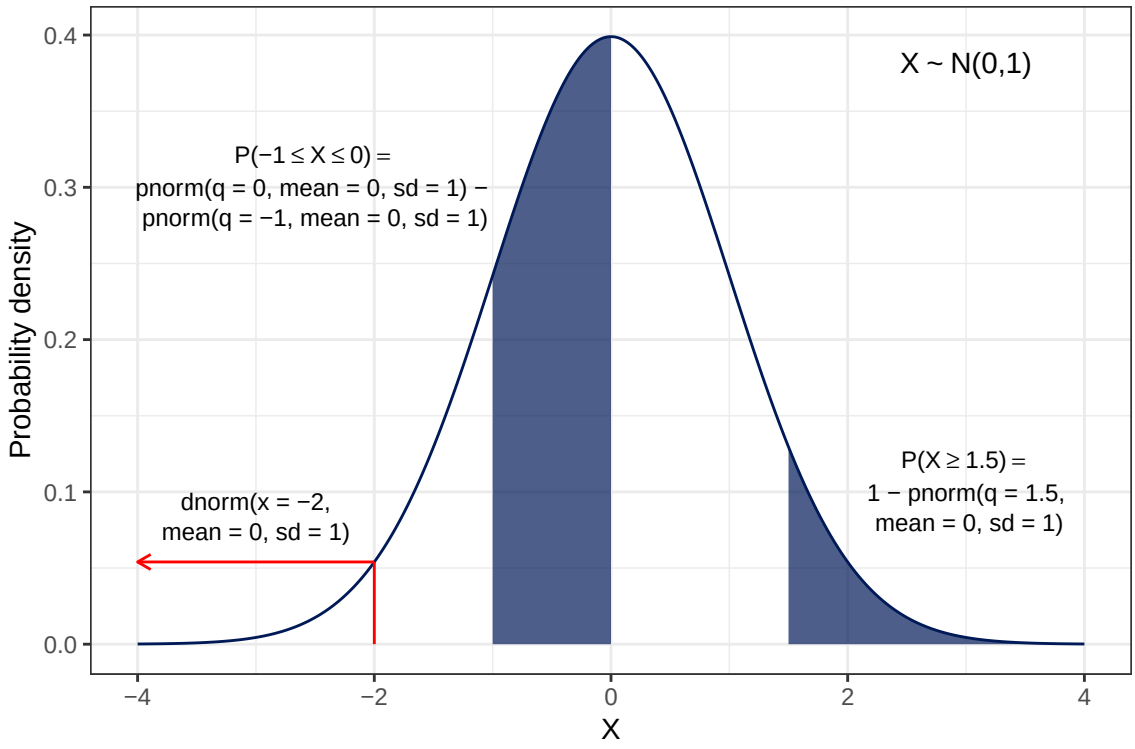
But that is OK because we usually want to calculate bigger areas under the Normal curve (see below figure). For example, we will ask questions like 'what is the probability of getting a value bigger than 1.5?'. This is how we evaluate whether the probability of getting a number (e.g., 1.5) is bigger or smaller than expected by chance given a null hypothesis. In this example, the probability $P(X \geq 1.5)$ is 0.067:

```
1 - pnorm(q = 1.5, mean = 0, sd = 1).
```

Incidentally, `qnorm` is the inverse of `pnorm`. Whereas `pnorm` finds the cumulative probability of a quantile (or any x-value you are interested in), `qnorm` finds the quantile of any cumulative probability that you give it. For example, below `pnorm` finds the cumulative probability for the quantile -1.96, and `qnorm` finds the quantile for the cumulative probability of 0.025.

```
pnorm(q = -1.96, mean = 0, sd = 1) ~ 0.025

qnorm(p = 0.025, mean = 0, sd = 1) ~ -1.96
```


X ~ N(0,1)
P(−1 ≤ X ≤ 0) =
pnorm(q = 0, mean = 0, sd = 1) −
pnorm(q = −1, mean = 0, sd = 1)
P(X ≥ 1.5) =
1 − pnorm(q = 1.5,
mean = 0, sd = 1)
dnorm(x = −2,
mean = 0, sd = 1)
Probability density
0.4
0.3
0.2
0.1
0.0
−4
−2
0
2
4
X