Probability Distributions

Random Variables

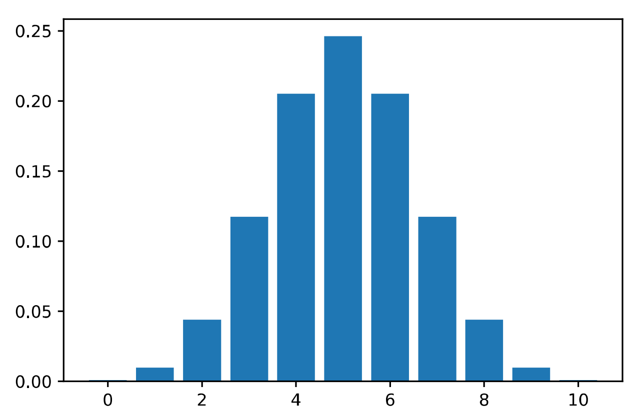
Random variables are functions with numerical outcomes that occur with some level of uncertainty. For example, rolling a 6-sided die could be considered a random variable with possible outcomes {1,2,3,4,5,6}.

Discrete and Continuous Random Variables

Discrete random variables have countable values, such as the outcome of a 6-sided die roll.
Continuous random variables have an uncountable amount of possible values and are typically measurements, such as the height of a randomly chosen person or the temperature on a randomly chosen day.

Probability Mass Functions

A probability mass function (PMF) defines the probability that a discrete random variable is equal to an exact value. In the provided graph, the height of each bar represents the probability of observing a particular number of heads (the numbers on the x-axis) in 10 fair coin flips.



codecademy

Probability Mass Functions in Python

The binom.pmf() method from the scipy.stats module can be used to calculate the probability of observing a specific value in a random experiment.
For example, the provided code calculates the probability of observing exactly 4 heads from 10 fair coin flips.

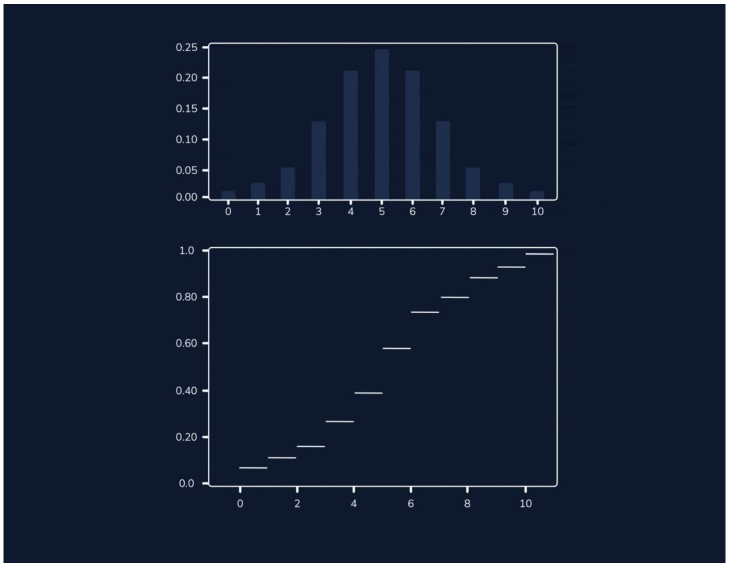
```
import scipy.stats as stats

print(stats.binom.pmf(4, 10, 0.5))

# Output:
# 0.20507812500000022
```

Cumulative Distribution Function

A cumulative distribution function (CDF) for a random variable is defined as the probability that the random variable is less than or equal to a specific value.
In the provided GIF, we can see that as x increases, the height of the CDF is equal to the total height of equal or smaller values from the PMF.



Calculating Probability Using the CDF

The binom.cdf() method from the scipy.stats module can be used to calculate the probability of observing a specific value or less using the cumulative density function.
The given code calculates the probability of observing 4 or fewer heads from 10 fair coin flips.

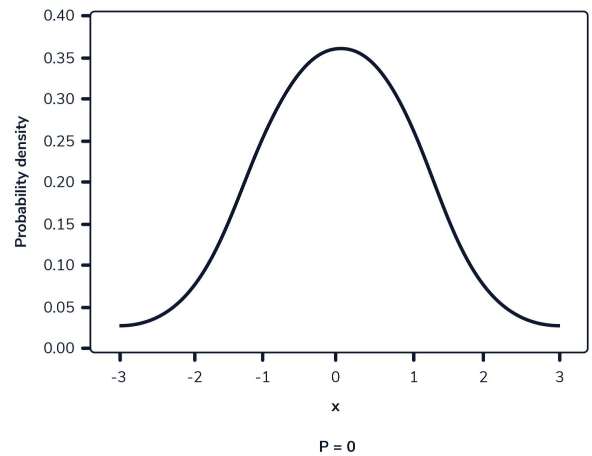
```
import scipy.stats as stats

print(stats.binom.cdf(4, 10, 0.5))

# Output:
# 0.3769531250000001
```

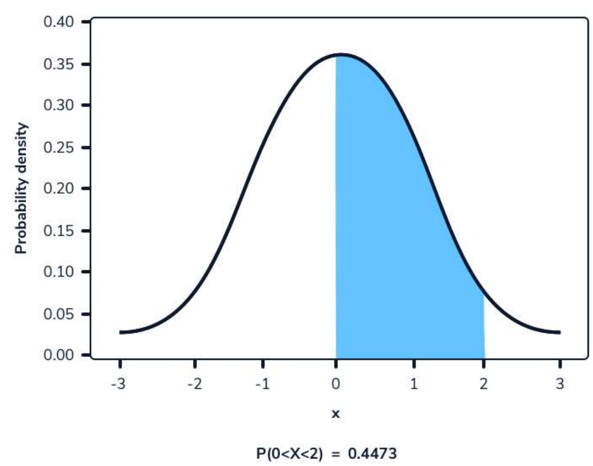
Probability Density Functions

For a continuous random variable, the probability density function (PDF) is defined such that the area underneath the PDF curve in a given range is equal to the probability of the random variable equalling a value in that range. The provided gif shows how we can visualize the area under the curve between two values.



Probability Density Function at a Single Point

The probability that a continuous random variable equals any exact value is zero. This is because the area underneath the PDF for a single point is zero.
In the provided gif, as the endpoints on the x-axis get closer together, the area under the curve decreases. When we try to take the area of a single point, we get 0.



Parameters of Probability Distributions

Probability distributions have parameters that control the exact shape of the distribution.
For example, the binomial probability distribution describes a random variable that represents the number of sucesses in a number of trials (n) with some fixed probability of success in each trial (p). The parameters of the binomial distribution are therefore *n* and *p*. For example, the number of heads observed in 10 flips of a fair coin follows a binomial distribution with n=10 and p=0.5.

The Poisson Distribution

The Poisson distribution is a probability distribution that represents the number of times an event occurs in a fixed time and/or space interval and is defined by parameter λ (lambda).
Examples of events that can be described by the Poisson distribution include the number of bikes crossing an intersection in a specific hour and the number of meteors seen in a minute of a meteor shower.

Expected Value

The *expected value* of a probability distribution is the weighted (by probability) average of all possible outcomes. For different random variables, we can generally derive a formula for the expected value based on the parameters.
For example, the expected value of the binomial distribution is n*p.
The expected value of the Poisson distribution is the parameter λ (lambda).
Mathematically:

$$X \sim Binomial(n, p),\ E(X) = n \times p$$

$$Y \sim Poisson(\lambda),\ E(Y) = \lambda$$

Variance of a Probability Distribution

The *variance* of a probability distribution measures the spread of possible values. Similarly to expected value, we can generally write an equation for the variance of a particular distribution as a function of the parameters. For example:

$$X \sim Binomial(n, p),\ Var(X) = n \times$$

$$Y \sim Poisson(\lambda),\ Var(Y) = \lambda$$

Sum of Expected Values

For two random variables, *X* and *Y*, the expected value of the sum of *X* and *Y* is equal to the sum of the expected values.
Mathematically:

$$E(X + Y) = E(X) + E(Y)$$

Adding a Constant to an Expected Value

If we add a constant *c* to a random variable *X*, the expected value of *X + c* is equal to the original expected value of *X* plus *c*.
Mathematically:

$$E(X+c) = E(X) + c$$

Multiplying an Expectation by a Constant

If we multiply a random variable *X* by a constant *c*, the expected value of *c*X* equals the original expected value of *X* times *c*.
Mathematically:

$$E(c \times X) = c \times E(X)$$

Adding a Constant to Variance

If we add a constant *c* to a random variable *X*, the variance of the random variable will not change.
Mathematically:

$$Var(X+c) = Var(X)$$

Multiplying Variance by a Constant

If we multiply a random variable *X* by a constant *c*, the variance of *c**X* equals the original expected value of *X* times *c* squared.
Mathematically:

$$Var(c \times X) = c^2 \times Var(X)$$

Print Share