M1 Maths

P6-5 Continuous Random Variables

- probability density functions
- expected value, quantiles and spread
- cumulative distribution functions
- mean and variance of $aX + b$

Summary Learn Solve Revise Answers

Summary

With a continuous random variable, the probability of any particular value is zero, so we use the probability density function – the probability that the variable will lie within a very small range divided by the width of that range. The probability of lying in a larger range can be obtained by integration.

The expected value of a continuous random variable, $E(X) = \int xP(x)dx$ over all possible values x of X.

If the 60th percentile is n, then n can be obtained from the equation $\int_a^n P(x)dx = 0.6$ where a is the lowest possible value for X. Similarly for other quantiles.

The variance of X is given by $\text{Var}(X) = \text{E}(X^2) - \mu^2$, where $\mu = \text{E}(X)$.

A cumulative distribution function for a continuous random variable, X, is the relation between n and $\int_a^n P(x)ds$, where a is the lowest possible value for x. It can be used to find the probability that X lies within a given interval without having to integrate.

If X is a random variable with expected value (mean) μ and variance σ^2, then $aX + b$ is a random variable with expected value $a\mu + b$ and variance $a^2\sigma^2$.

Learn

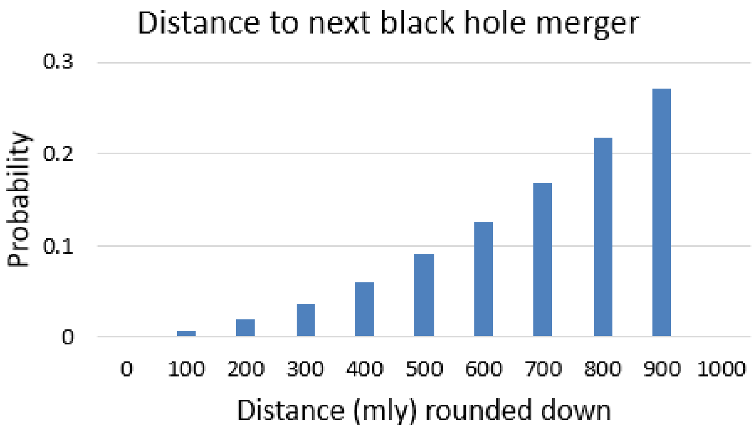
Continuous Random Variables and Probability Density Functions

The LIGO (Laser Interferometer Gravitational-wave Observatory) detectors can pick up gravity waves from mergers of black holes. Black hole mergers happen quite frequently throughout the universe.

Image from Wikipedia – Binary Black Hole

The scientist in charge is interested in black hole mergers within 1000 million light years (mly) of Earth. She wants to know the probability distribution for how far from Earth the next merger will take place.

Let the random variable S be the distance from Earth to the next merger rounded down to the next 100 mly. The probability distribution for S looks like this:

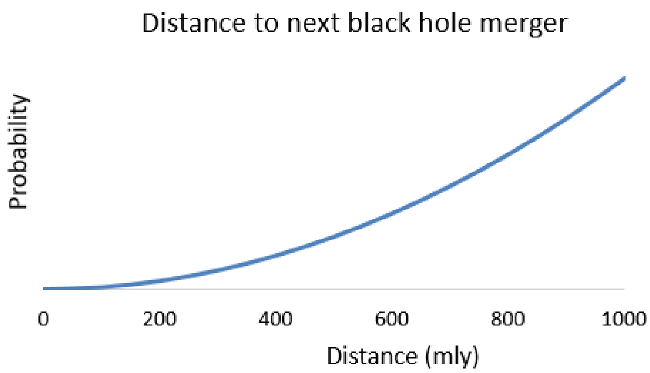


Can you see why the next merger is more likely to be at 900 mly than at 200 mly? It's because the shell of space from 900 mly away to 1000 mly away has a larger volume than the shell from 200 mly away to 300 mly away.

Because all values, s, of the variable S are a whole number of hundreds of millions of light years, S is a discrete random variable with allowable values 100 mly, 200 mly etc.

As a discrete random variable, each allowable value has a definite probability. For example, from the graph, $P(S = 600) = 0.127$.

But if S was the actual observed distance rather than being rounded down, it would be a continuous random variable with any value between 0 and 1 billion, e.g. 273 591 347.89. The probability distribution would then look like this:



Being a continuous random variable, S has a continuous probability distribution. The problem with this is that now $P(S = 600) = 0$. In fact the probability of being any specific distance s from Earth is zero. Why?

When S was a discrete variable, it had to take on one of 10 values, so each had quite a reasonable probability. The sum of the probabilities has to be 1 and so the average of the probabilities is 0.1.

If S is continuous, however, it can take on any of an infinite number of values. The sum of all the infinite number of probabilities must still be 1, so each probability can be no greater than zero.

We can get around this problem by considering **probability density** rather than probability. The probability density at a particular value, s, of S is the probability that S will lie within a one unit interval centred on s. If we use millions of light years as our unit of distance, then the probability density at $S = 432$ could be thought of as the probability that S lies between 431.5 and 432.5.

In actual fact, because the probability density is slightly higher at 432.5 than at 431.5, we define the probability density at 432 as the probability that S will lie in an infinitely small interval centred on 432, divided by the width of that interval. This will generally come to roughly the same probability density that we would obtain using a 1-unit interval.

The line graph above for the continuous random variable S gives the probability density for each value of S. We can see that the probability density increases as the value of S increases. However, we haven't got a scale on the vertical axis.

To get the scale on the vertical axis, we have to use a bit of calculus.

The area under the curve can be considered as the sum of infinitely thin vertical slices of width ds and height $P(s)$, where $P(s)$ is the probability density at that s. Each slice has an area equal to $P(s)ds$. This will be the probability that S lies within that interval.

The sum of the probabilities for all the intervals must be 1. So $\int_0^{1000} P(s)ds = 1$.

Now we don't know the probability for any value, s, of S, but we do know that the probability is proportional to s^2. This is because the volume of any thin shell of thickness ds at distance s from Earth will be proportional to s^2. So we can say $P = ks^2$, where k is the constant of proportionality.

Subbing this for $P(s)$ in our integral, we get $\int_0^{1000} ks^2 ds = 1$.

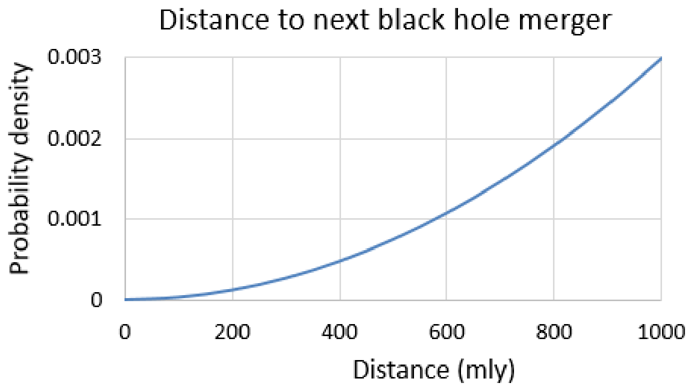
Integrating, we get $\left[\frac{ks^3}{3}\right]_0^{1000} = 1$

$$\frac{1000000000k}{3} - \frac{0k}{3} = 1$$

$$k = 0.000\,000\,003$$

So, when $s = 1000$, $P(s) = 0.000\,000\,003 \times 1000^2 = 0.003$

So now we can add a vertical scale to the graph to get:



This probability function is called a **probability density function**. We generally use that term when dealing with continuous random variables.

The probability density function is $P(s) = 0.000\,000\,003\, s^2$.

Using a Probability Density Function

Suppose we want to use the probability density function above to find the probability that the next merger will occur at a distance between 700 mly and 750 mly.

We can get an approximate answer like this:

Looking at the graph, we can see that in the range 700 to 750 mly, the probability density averages about 0.0016. That's 0.0016 per million light years. The interval from 700 to 750 mly is 50 wide. So the total probability of lying in that interval is about 50 × 0.0016 = 0.08.

To get an exact answer, we have to integrate:

The probability density function is $P(s) = 0.000\,000\,003\, s^2$.

The probability that S will be between 700 and 750 is

$$\int_{700}^{750} 0.000\,000\,003 s^2 ds$$

$$= 0.0789$$

[**Note:** As this module is not about learning to integrate functions, it is quite appropriate where possible to evaluate the integrals using the numerical integration function on your graphics calculator.]

Practice

Use the calculus method above to answer these questions.

Q1 Using the information above, find the probability that the next black hole merger will occur between 300 and 330 mly from Earth.

Q2 Find the probability that the next black hole merger will occur less than 420 mly away.

Q3 If LIGO detects an average of 15 mergers a year within 1000 mly, how long would it be on average between mergers that occur within 200 mly of Earth?

Q4 After an upgrade, the LIGO detector can pick up all mergers within 2500 mly and the project is extended to take in all mergers within that distance. Find the probability that the next merger within that range occurs more than 2000 mly away.

Q5 Suroshito's Sushi Suppliers grow seaweed in a large tank. After growing for two weeks, the probability density function for the mass m of a plant for $0 < m < 20$ is $P(m) = 0.000075m^3 - 0.003m^2 + 0.03m$, where m is in grams. The ideal mass is between 10 and 15 g. What percentage are in the ideal range?

Q6 A tennis serving machine can be set to fire balls in a set direction, but, to add a bit of unpredictability, the balls can go up to 5° from that direction. The probability of being x° from the set direction is proportional to $\frac{1}{x+1}$ for $0 < x < 5$. Find the probability density function and use it to find the probability that the ball will go more than 3° from the set direction.

Q7 A sample of ^{40}K (a radioactive isotope of potassium) decays with an average of 9 disintegrations per hour. But the disintegrations occur at random. The probability density function for t, the time in hours from the start of observation to the first decay is $9e^{-9t}$. Find the probability that no disintegrations will occur in the first 6 minutes.

Expected Value

Let's go back to the merging black holes. We might be interested in the expected distance to the next merger. This is another way of saying the mean distance to mergers.

We know that for a discrete distribution, the expected value is given by $E(X) = \Sigma xP(x)$ over all possible values x of X.

For a continuous distribution, it is given by $E(X) = \int xP(x)dx$ over all possible values x of X.

For the black holes, $P(s) = 0.000\,000\,003\, s^2$.

So $E(S) = \int s\,P(s)\,ds = \int_0^{1000} 0.000\,000\,003 s^3 ds = 750.$

So the average or expected distance to black hole mergers within 1000 mly is 750 mly.

Practice

Q8 For Suroshito's seaweed, the probability density function for the mass m of a plant for $0 < m < 20$ is $P(m) = 0.000075m^3 – 0.003m^2 + 0.03m$, where m is in grams. Find the expected mass for the plants.

Q9 After being upgraded, the LIGO detector can pick up all mergers within 2500 mly. Find the expected distance to detected mergers now.

Q10 Find the expected value of x for the tennis serving machine in Q6.

Q11 Find the expected wait time for the first disintegration of the ^{40}K sample in Q7.

Quantiles

Suppose we wanted to find the median distance to black hole mergers occurring between 0 and 1000 mly away.

To find this, we have to find the distance at which the probability of the next merger being less than that is 50%.

i.e. the value of x for which $\int_0^{x} 0.000\,000\,003 s^2 ds = 0.5$.

For this, we have to integrate algebraically. We get:

$[0.000000001s^3]_0^x = 0.5$

$0.000000001x^3 = 0.5$

$x^3 = 500\,000\,000$

$x = 794$ mly

To find, say, the 20th percentile, we just use 0.2 instead of 0.5 in the equation.

Practice

Q12 If the LIGO detector can detect black hole mergers up to 2500 mly away, what is

(a) the median distance for detections?

(b) the 90th percentile distance?

(c) the first quartile distance?

(d) the interquartile range?

Spread

To calculate the variance of a probability distribution, we can use the computational formula $\text{Var}(X) = \text{E}(X^2) - \mu^2$, where $\text{E}(X^2)$ is the expected value of X^2 and μ is the expected value of X or $\text{E}(X)$. The derivation of this formula is given in the box below if you need it.

Let's look again at the black holes. We saw that the mean distance to mergers closer than 1000 mly was 750 mly. So $\mu = 750$.

$\text{E}(S^2) = \int s^2\, P(s)\, ds = \int_0^{1000} 0.000\,000\,003 s^4 ds = 600\,000$

So $\text{Var}(X) = 600\,000 - 750^2 = 37\,500$

The standard deviation is the square root of the variance $= \sqrt{37\,500} = 194$ mly.

Derivation of the Computation Formula for Variance

$$\begin{aligned}\text{Var}(X) &= \int (x - \mu)^2\, p(x)\, dx \\ &= \int (x^2 - 2\mu x + \mu^2)\, p(x)\, dx \\ &= \int x^2\, p(x)\, dx - \int 2\mu x\, p(x)\, dx + \int \mu^2\, p(x)\, dx \\ &= \text{E}(X^2) - 2\mu \int x\, p(x)\, dx + \mu^2 \int p(x)\, dx\end{aligned}$$

As $\int x\, p(x)\, dx = \mu$ and $\int p(x)\, dx = 1$,

$$\begin{aligned}\text{Var}(X) &= \text{E}(X^2) - 2\mu^2 + \mu^2 \\ &= \text{E}(X^2) - \mu^2\end{aligned}$$

Practice

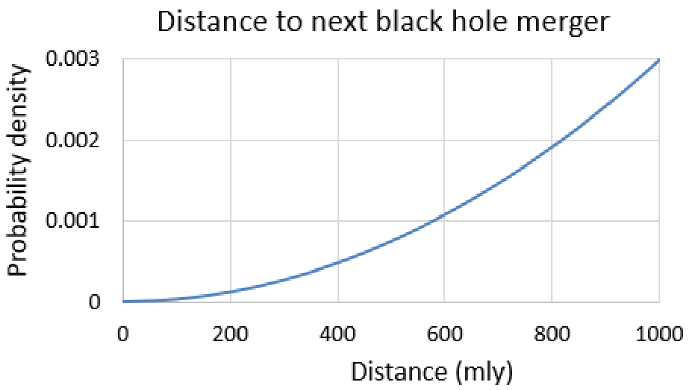
Q13 If LIGO can detect mergers out to 4000 mly, find the variance and standard deviation of the distance probability distribution.

Q14 For Suroshito's seaweed, the probability density function for the mass m of a plant for $0 < m < 20$ is $P = 0.000075m^3 - 0.003m^2 + 0.03m$, where m is in grams. Find the variance and standard deviation for the masses of the plants.

Q15 Find the variance and standard deviation of the value of x for the tennis serving machine in Q6.

Cumulative Distribution Function

Let's have another look at the probability density function for the black hole mergers. We worked out that $P = 0.000\ 000\ 003\ s^2$. The graph of this function is shown below.



The same information could also be expressed as a cumulative probability distribution function or cumulative distribution function. The cumulative distribution function gives, for each distance, x, the probability that the next merger will occur at distance less than x. It is $\int_0^x 0.000\ 000\ 003\ s^2 ds$. This is the area under the curve between $s = 0$ and $s = x$.

$$\int_0^x 0.000\ 000\ 003\ s^2 ds \quad = \left[\frac{0.000\ 000\ 003\ s^3}{3}\right]_0^x \quad = 0.000\ 000\ 001\ x^3$$

Using this function, if we want to know the probability that the next merger is less than 250 mly away, we just sub 250 for x to get $0.000\ 000\ 001 \times 250^3 = 0.0156$.

The probability that it will be more than 250 mly away will obviously be $1 - 0.0156$, which is 0.9844.

To find the probability that the next merger occurs between 500 mly and 700 mly away, we subtract the probability that it is less than 500 mly away from the probability that it is less than 700 mly away:

$0.000\ 000\ 001 \times 700^3 - 0.000\ 000\ 001 \times 500^3 = 0.218$

This is an alternative way of finding the probabilities. It has the advantage that we don't have to integrate for each calculation: the integration is done in getting the formula.

Practice

Q16 Using the cumulative distribution function for black hole mergers above, find the probabilities that the next merger is

(a) less than 450 mly away

(b) more than 450 mly away

(c) more than 900 mly away

(d) between 700 mly and 720 mly away

Q17 If LIGO can now detect any black hole merger up to 5000 mly away, find

(a) the probability density function for the next merger

(b) the cumulative distribution function for the next merger

(c) the probability that the next merger is less than 1000 mly away

(d) the probability that the next merger is more than 4000 mly away

(e) the probability that the next merger is between 3000 mly and 4000 mly away.

Q18 As we saw earlier, in Suroshito's seaweed tanks, after growing for two weeks, the probability density function for the mass m of a plant for $0 < m < 20$ is $P = 0.000075m^3 - 0.003m^2 + 0.03m$, where m is in grams. Find

(a) the cumulative probability function

(b) the probability that a plant will have a mass between 10 g and 12 g

(c) the percentage of plants less than 6 g

(d) the percentage of plants between 16 and 20 g.

Mean and Variance of $aX + b$

If X is a random variable with expected value (mean) μ and variance σ^2, then $aX + b$ is a random variable with expected value $a\mu + b$ *and variance* $a^2\sigma^2$.

Practice

Q19 For a continuous random variable, X, $E(X) = 23$ and $Var(X) = 4$. Find the expected value and the variance of $2X - 5$.

Q20 For a random variable, S, the mean is 14.2 and the standard deviation is 1.5. Find the mean and standard deviation of $S/2 + 12$.

Q21 In 1296 an attempt was made to eradicate rats from the town of Hamelin. Residents who brought dead rats to the town hall were paid 3 pfennigs per rat minus a 1 pfennig processing fee. [Based on past experience, this was considered preferable to employing a piper.] The number of rats people brought in averaged 4.5 with a standard deviation of 1.2. What were the mean and standard deviation of the pay-outs.

Solve

Q51 The cumulative distribution function for x, the distance walked by a toad before it falls asleep, is $P = 0.008\ (15x^2 - 2x^3)$ for $0 < x < 5$. Find the probability density function. Plot both functions on the same axes.

Q52 A machine produces rectangular off-cuts of sheet metal. Their vertical lengths have a rectangular continuous distribution from 0 to 10 cm. Likewise for their horizontal lengths. Find:
(a) the probability density function for their perimeters, x
(b) their mean perimeter
(c) the percentage with perimeters >30 cm^2

These off-cuts can be sold for 30c plus 2c per centimetre of perimeter.
(d) What would be the expected value of 1000 off-cuts?

Revise

Revision Set 1

Q71 Zamboobooland is a round island 40 km in diameter. The king lives in a round compound 1 km in diameter right in the middle of the island. The inhabitants of nearby Mankie Island don't like the Zambbobbolanders and periodically lob bombs onto their island. The bombs are equally likely to land anywhere on the island. The king is interested in the probability distribution for how far from the centre of his compound they land. Obviously the probability of landing a distance x from the centre would be proportional to x.
(a) Find the probability density function.
(b) Use this function to find the probability that the next bomb will land in his compound.
(c) Find the expected distance from the centre of the island.
(d) Find the median distance from the centre of the island and the inter-quartile range.
(e) Find the standard deviation of the distance from the centre.
(f) Find the cumulative distribution function.
(g) Use this to find the probability that the next bomb will land within 2 km of the coast.
(h) As the richer people live nearer to the coast, the average cost of repairing the damage from a bomb strike in dollars is equal to 200 times the distance from the centre plus 3000. Find the mean repair cost and the variance.

Q72 People arrive at a supermarket at random, but on average 120 people arrive per hour. The probability distribution for the time, t, in minutes till the next arrival is thus $2e^{-2t}$.
(a) Find the probability that the next customer will arrive between 10 seconds and 15 seconds from now.

(b) Find the cumulative distribution function.
(c) Use the cumulative distribution function to find the probability that the next customer will arrive within the next 20 seconds.

Answers

Q1 0.00894
Q2 0.0741
Q3 8.33 years
Q4 0.488
Q5 26.2%
Q6 $P(x) = 1/(1.79176(x+1))$, 22.6%
Q7 0.407
Q8 8 g
Q9 1875 mly
Q10 1.79°
Q11 0.111 hours (6 h 40 min)
Q12 (a) 1984 mly (b) 2414 mly (c) 1575 mly (d) 696 mly
Q13 Variance = 600 000 mly^2, sd = 775 mly
Q14 Variance = 16 g^2, sd = 4 g
Q15 Variance = 1.98, sd = 1.41°
Q16 (a) 0.091 (b) 0.909 (c) 0.271 (d) 0.0302
Q17 (a) $P = 2.4 \times 10^{-11} s^2$ (b) $8 \times 10^{-12} s^3$ (c) 0.008 (d) 0.488 (e) 0.296
Q18 (a) $0.00001875m^4 - 0.001m^3 + 0.015m^2$ (b) 0.133 (c) 34.8% (d) 2.7%
Q19 $E(2X - 5) = 41$, $Var(2X - 5) = 8$
Q20 Mean = 19.1, sd = 0.75
Q21 Mean = 12.5, sd = 3.6

Q51 $P(x) = 0.008(30x - 6x^2)$; see calculator for plots.
Q52 (a) $P(x) = \begin{cases} x/400 & \text{if } 0 < x < 20 \\ 1/10 - x/400 & \text{if } 20 < x < 40 \end{cases}$ or $P(x) = 1/20 - \frac{|x-20|}{400}$
(b) 20 cm (c) 12.5% (d) $700

Q71 (a) $P(x) = 0.005x$ (b) 0.000 625 (c) 13.3 km
(d) median = 14.1 km, IQR = 7.25 km (e) 4.71 km
(f) cdf = $0.0025x^2$ (g) 0.19 (h) mean = $5667, variance = $888 872
Q72 (a) 0.110 (b) cdf = $1 - e^{-2t}$ (c) 0.487