Continuous Random Variables

Owen G. Ward

2021-05-24

Introduction

We now want to consider similar problem to the previous setup, but now the variables can take on all values in some range.

For example, we want a distribution to describe height or weight, or the length of time, etc. Any quantities which can take or **continuous** values.

Instead of a pmf for discrete variables, we now have a **probability density function** (pdf), $f_X(x)$, describing the relationship between a random variable and the values in can take.

In the discrete case, we could give the exact probability of a rv taking a specific value, but we cannot do that for continuous random variables.

To determine $P(3 \leq X \leq 6)$ we integrate the pdf over this range

$$P(3 \leq X \leq 6) = \int_3^6 f_X(x)dx.$$

This shows that $P(X = 4)$, say is $\int_4^4 f_X(x)dx = 0$.

We have the same properties that we had before:

- $f_X(x) \geq 0$
- $\int_X f_X(x)dx = 1$, where we integrate over all possible values of X.

Example

Suppose we have a random variable X with pdf given by

$$f_X(x) = 2e^{-2x}, \ x \geq 0.$$

X could be used to model the lifetime of a bulb or a machine component, in days.

What is $P(X < 1)$, the probability fails in less than a day?

What is $P(X > 2)$, the probability the bulb works for longer than 2 days?

Cdf

Similar to discrete variables, the cdf $F(x) = P(X \leq x)$ is given by

$$F(x) = \int_{-\infty}^{x} f_X(t)dt,$$

where we use t as the dummy variable for integration.

Instead of the lower limit being $-\infty$ it will be the smallest non zero value of the pdf.

We can actually get the pdf by taking the derivative of the cdf

$$\frac{d}{dx} F_X(x) = f_X(x).$$

Example

Compute the cdf of

$$f_X(x) = 2e^{-2x},\ x \geq 0.$$

Continuous CDF

Unlike the cdf of a discrete random variable, the cdf of a continuous random variable is a continuous non decreasing function, with the same limit properties.

Expectation and Variance

The expectation and variance of continuous random variables is similar to discrete, with integrals replacing sums.

If X has pdf $f_X(x)$ then

$$\mathbb{E}(X) = \int_X x f_X(x)dx.$$

We still have

$$\mathbb{E}(X^2) = \int_X x^2 f_X(x)dx$$

which can use to compute $Var(X)$,

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Example

For $f_X(x) = cx^2$ on $[0, 1]$, 0 otherwise, find the value of c to make this a valid pdf.

Compute $\mathbb{E}(X)$

Compute $\mathbb{E}(X^2)$

For $f_X(x) = 2e^{-2x},\ x \geq 0$, compute $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$.

$\mathbb{E}(X)$

$\mathbb{E}(X^2)$

Common Distributions

Uniform Distribution

Simple distribution, where every value is equally likely between some values $[a, b]$.

pdf given by

$$f_X(x) = \frac{1}{b-a},\ a \leq x \leq b$$

Commonly consider $a = 0, b = 1$.

For $X \sim Unif(a, b)$ then $\mathbb{E}(X) = \frac{a+b}{2}$ and $Var(X) = \frac{1}{12}(b-a)^2$.

Example

Suppose the arrival of the next subway is uniformly distributed on $[0, 10]$ minutes, so you will never be waiting more than 10 minutes.

- What is the expected length of time you will be waiting?

- What is the probability you will be waiting more than 8 minutes?

Exponential Distribution

Exponential distribution can be used for random variables which take on positive values, such as the length of time you wait in line.

Has pdf given by

$$f_X(x) = \lambda e^{-\lambda x},\ x \geq 0.$$

Say X follows an exponential distribution with parameter λ.

As we sort of saw before, $\mathbb{E}(X) = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.

The cdf is given by $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$.

Example

Suppose instead that you could be waiting longer than 10 minutes for the next subway, but that the mean waiting time is still 5 minutes. If we use an exponential distribution instead, what is the probability you will be waiting longer than 8 minutes?

Memoryless Property of Exponential Distribution

Suppose we have $X \sim Exp(\lambda = 1/24)$, some component in a server which on average is replaced every 24 weeks. What is the probability it lasts more than 30 weeks?

Given that this part has lasted 100 weeks, what is the probability it will last for 30 more weeks?

To get this we have

$$P(X > 130|X > 100) = \frac{P(X > 130, X > 100)}{P(X > 100)} = \frac{P(X > 130)}{P(X > 100)}.$$

When we compute this, we see

$$P(X > 130|X > 100) = P(X > 30),$$

which is the memoryless property of the exponential distribution.

Normal Distribution

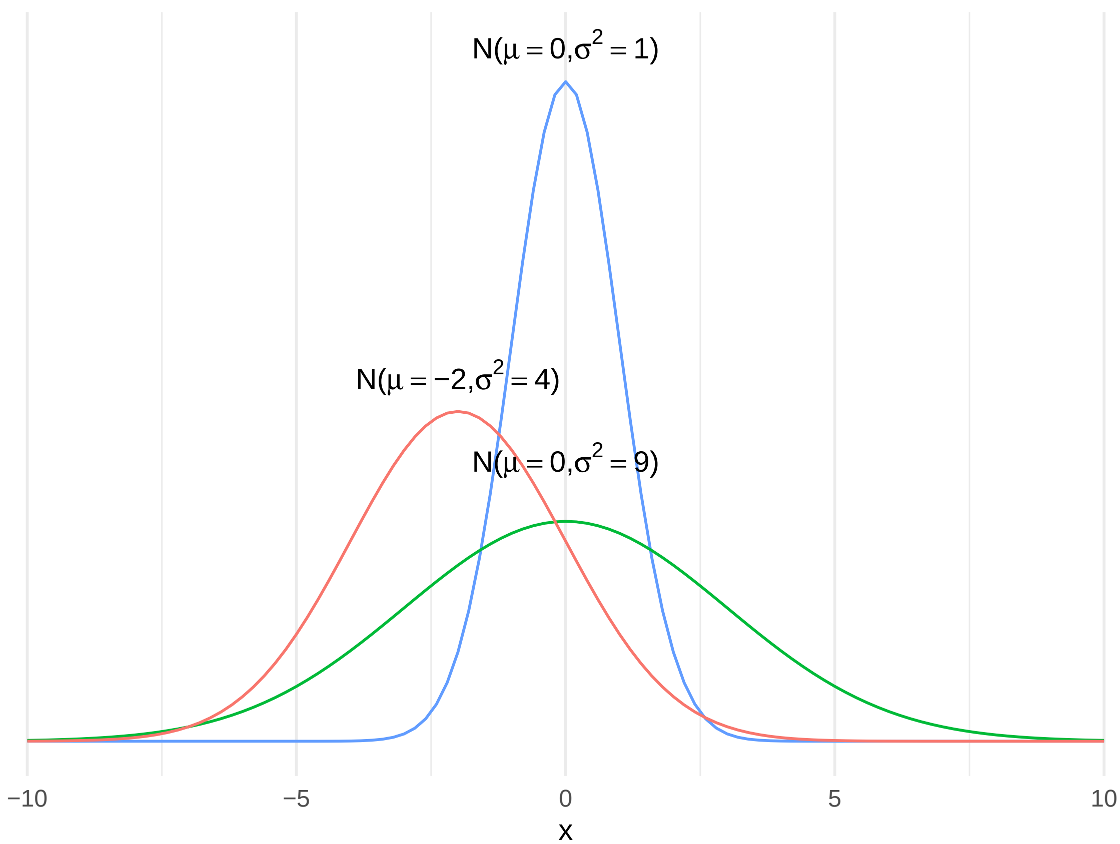
This is the most well known and important distribution!

Can be used for real data taking on positive and negative values, such as in natural sciences.

Say X follows a normal distribution with parameters μ, σ^2 with pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

Some Normal pdfs



Expectation and variance

The pdf's above are symmetric about the value of μ, and this actually means that for $X \sim \mathcal{N}(\mu, \sigma^2)$ that

$$\mathbb{E}(X) = \mu.$$

By doing some slightly complicated integrals, we have

$$Var(X) = \sigma^2.$$

Getting a mean zero variance one variable

For any random variable X with $\mathbb{E}(X) = \mu$ and $Var(X) = \sigma^2$ then we can make this mean zero.

Similarly, we can make it have variance 1.

So we can do both these things to make any random variable mean 0 and variance 1.

If we do this to a normal random variable, it is still normally distributed!

Transforming to a standard Normal

If we have $X \sim \mathcal{N}(\mu, \sigma^2)$ then for

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1),$$

which is a **Z-score**, which we can compute easily.

For $Z \sim \mathcal{N}(0, 1)$ then $P(Z \leq z) = \Phi(z)$, where Φ is a function that you can look up.

Area of a standard Normal

Example

Suppose $Z \sim \mathcal{N}(0, 1)$. What is

- $P(Z < -1)$.
- $P(Z > 1.5)$.
- $P(Z < -1)$ or $P(Z > 2)$

Standard Deviations from the Mean

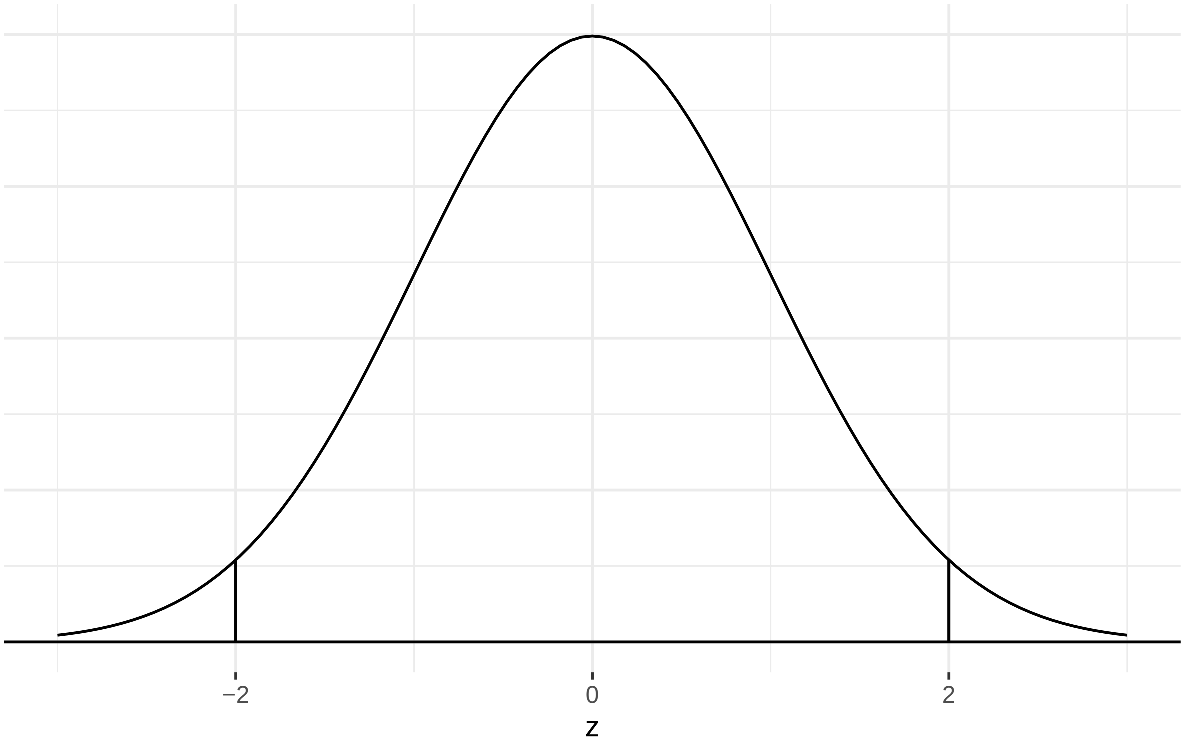
To get the area between 2 and -2 say we can compute

$$\Phi(2) - \Phi(-2) = 0.9544.$$

So over 95% of values within 2 standard deviations of the mean. To get exactly 95% use $z = 1.96$, which we will see later.

Approximately 68% within 1 standard deviation, 99.7% within 3 standard deviations.

Symmetry of the Normal CDF



Transforming Probabilities

An extremely useful property of probability statements is we can take linear transformations and still get the same probability, because we are still getting the probability of the same statement.

Example

Suppose SAT scores are normally distributed with mean $\mu = 1150$ and $\sigma = 200$. What is the probability a SAT score is less than 1200?

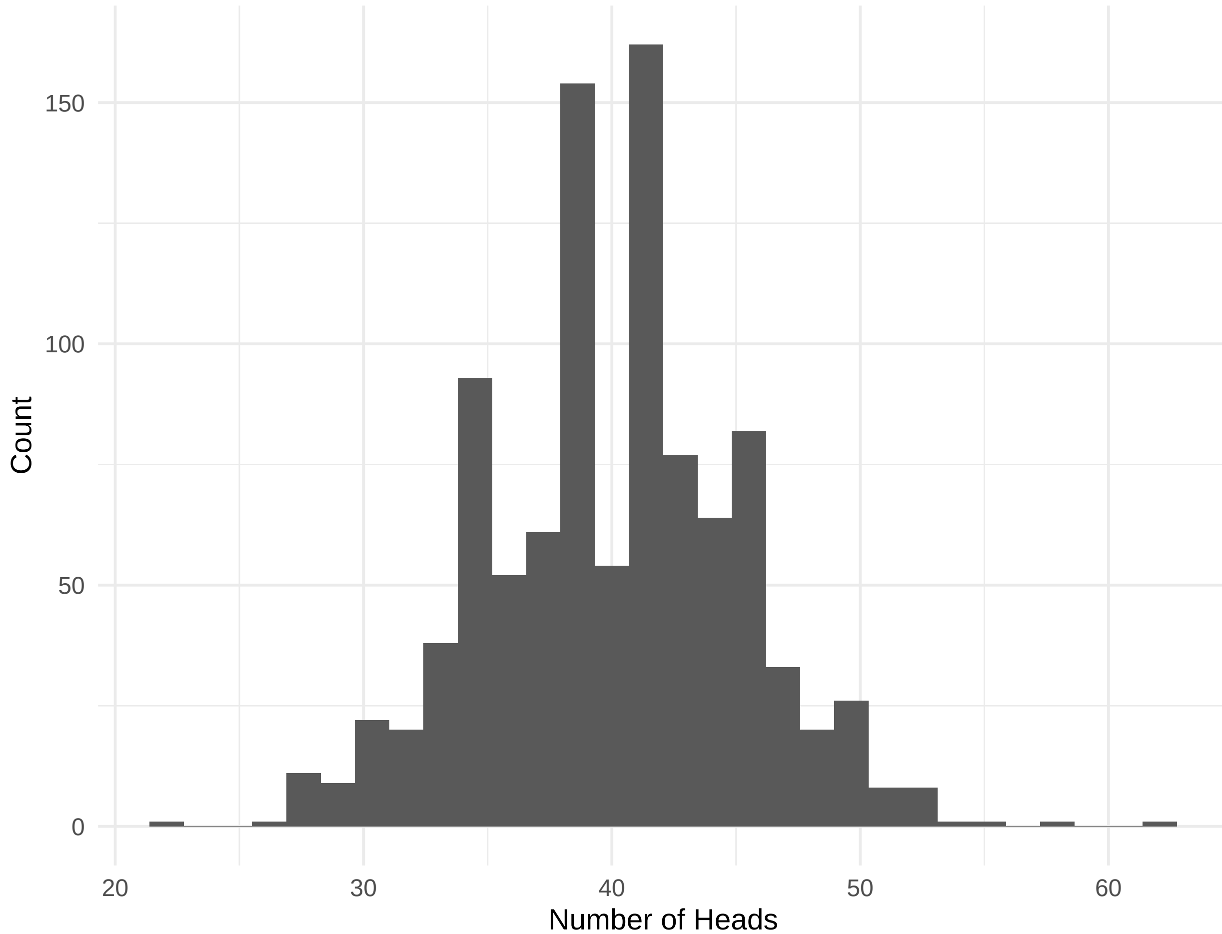
What is the probability a SAT score is greater than 1500?

Example

Cholesterol levels for women aged 20 to 34 follow an approximately normal distribution with mean 185 milligrams per deciliter (mg/dl). Women with cholesterol levels above 220 mg/dl are con- sidered to have high cholesterol and about 18.5% of women fall into this category. What is the standard deviation of the distribution of cholesterol levels for women aged 20 to 34?

Approximating using a Normal Distribution

It turns out that in many cases the normal distribution is a good approximation to other distributions, even non continuous ones. Suppose we have a Binomial distribution with large n. If we do a histogram of 1000 draws from that distribution, we see that…



this looks quite like a normal distribution!

In truth $X \sim Binom(n = 100, p = 0.4)$ but we will approximate it with a normal distribution Y with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. We want to compute $P(X \leq 30)$.

Instead of computing $P(X \leq 30)$ directly for the Binomial we can compute $P(Y \leq 30)$ for $Y \sim \mathcal{N}(40, \sigma^2 = 24)$. This is much easier to compute.

$P(X \leq 30) = 0.0248$

While $P(Y \leq 30)$

Central Limit Theorem

The previous example is true more generally. For almost all distributions, if you take a large sample $X_1, X_2, \ldots, X_n$ from that distribution, the mean of that sample is well approximated by a normal distribution.

The mean of that normal distribution will be $\mathbb{E}(X_1)$ and the variance will be $\frac{1}{n}Var(X_1)$.

So you compute things about the average of data from any distribution, once you know the mean and the variance of the distribution.

Example

Suppose the number of times a college student checks social media in a day has mean 15 and standard deviation 4. If you ask 100 students how many times they checked social media yesterday, what is the probability the average number will be greater than 10?

If $X_1, X_2, \ldots, X_{100}$ is the number of times each student checked, we know that $\bar{X} = \frac{1}{100}(X_1 + X_2 + \ldots + X_{100})$. By the CLT this is approximately normally distributed with mean 15 and variance $16/n$.

$$\bar{X} \approx \mathcal{N}\left(15, \sigma^2 = \frac{16}{n}\right),$$

and we can compute $P(\bar{X} > 10)$ similar to previous examples, by transforming to a standard normal.

Recall that if $Y \sim \mathcal{N}(\mu, \sigma^2)$ then $\frac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

So

$$P(\bar{X} > 10) = P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} > \frac{\sqrt{n}(10-15)}{\sigma}\right) = P\left(Z > \frac{10(-5)}{4}\right) = P\left(Z > -12.5\right)$$

Notice, we don't know anything about the distribution here except its mean and variance.

Recap

Joint Continuous Random Variables

We saw previously that we can have joint distributions of two discrete random variables. We can do the same thing with continuous random variables. This means we have a joint pdf

$$f_{X,Y}(x,y),$$

which has the same properties of a standard pdf.

To get the marginal distribution of one variable, we integrate out the other.

$$f_Y(y) = \int_X f_{X,Y}(x,y)dx.$$

Similarly, we can get expectations of joint distributions

$$\mathbb{E}(g(X,Y)) = \int_X \int_Y g(x,y) f_{X,Y}(x,y)dxdy$$

We say X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Example

Suppose X and Y have joint pdf

$$f_{X,Y}(x,y) = \frac{6}{5}(x+y^2),\ 0 \le x \le 1, 0 \le y \le 1.$$

Compute the marginals and determine if they are independent.

Compute $\mathbb{E}(X)$.

Other Transformations

If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ then

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

If $X_1 \sim Poisson(\lambda_1)$ and $X_2 \sim Poisson(\lambda_2)$ then

$$X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2).$$

If $X_1 \sim Binom(n,p)$ and $X_2 \sim Binom(m,p)$ then

$$X_1 + X_2 \sim Binom(n+m,p).$$

The sum of Uniforms and Exponentials is not Uniform/Exponential. But if they are independent we can get joint distributions easily.

Example

Suppose the lifetime of a bulb has an exponential distribution with mean 100 days.
You go to the store and buy 20 bulbs.What is the chance that at least one bulb will work for more than 200 days?
Assume the lifetime of different bulbs are mutually independent.

Want $P(X_i > 200)$ for at least one i in $1, \ldots, 20$.