

Discrete and continuous distributions, Part 2

Kris Gunsalus

9/23/2019

Contents

Reading assignment	1
Review	1
Continuous PDFs	1
Uniform distribution	1
Normal distribution	2
Standard normal	2
Examples	5
Uniform distribution	5
Normal distribution	5

Reading assignment

- **Aho - Chapter 3.1-3.2: Probability Density Functions (Part 1)**
- **Aho - Chapter 4.3: Statistics**

Review

- Bernoulli distribution
- Binomial distribution
- **Tranchina - Elements of Calculus**

Continuous PDFs

Uniform distribution

The PDF is defined, for $-\infty < a \leq x \leq b < \infty$, as:

$$f(x) = \frac{1}{b-a}$$

The CDF is:

$$F(x) = \int_a^x \frac{1}{b-a} dt = 1$$

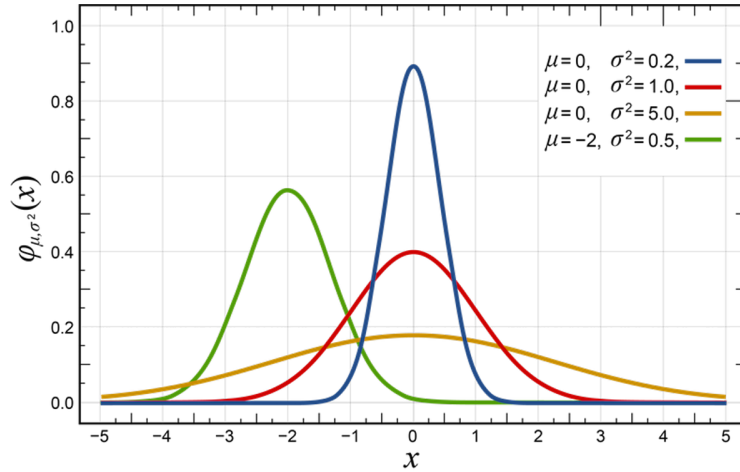


Figure 1: Normal distributions with different mean and variance

In geometric terms, the width of a uniform distribution is $(b - a)$ and the height is $1/(b - a)$, so the area under the curve always equals one.

We write $X \sim UNIF(a, b)$. The mean and variance are: $E(X) = \frac{b-a}{2}$ and $V(X) = \frac{(b-a)^2}{12}$.

Normal distribution

The normal distribution is the most common distribution in statistics and has many applications in biology. It is important because it represents many processes in which the most likely outcome is the average. Large sums of (small) random variables are often normally distributed.

PDF:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

CDF:

$$F(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

where

1. Outcomes x are continuous and independent.
2. $x \in \mathbb{R}$
3. $\mu \in \mathbb{R}$
4. $\sigma > 0$

We write $X \sim N(\mu, \sigma^2)$, where $E(X) = \mu$ and $V(X) = \sigma^2$.

The Normal distribution has a characteristic bell shape, the precise dimensions of which are governed by μ and σ .

Standard normal

The standard normal distribution, or *Z-distribution*, is a normal or Gaussian distribution with $\mu = 0$ and $\sigma = 1$: $Z \sim N(0, 1)$

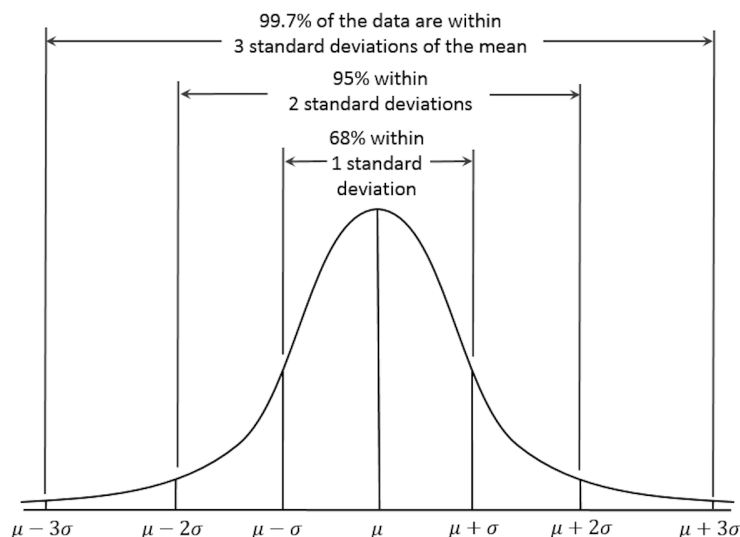


Figure 2: Empirical rule for the normal distribution

The normal distribution is standardized by subtracting the mean and dividing over the SD:

$$Z = \frac{X - \mu}{\sigma}$$

Any outcome x_i from a normal distribution can be turned into a z -score in the same manner.

As with any other distribution, we can ask questions like, “What is the chance of observing a value of x or less? x or more? between x and y ?”

We can also use the **inverse CDF**, or the **quantile** function, to find the value for some percentile in the population (e.g. the femur length that 80% of the population is under, which could be a factor in setting the seat pitch in airplanes).

Empirical rule for the standard normal

There is a general rule of thumb about how much of a distribution falls within some number of standard deviations from the mean: a little over 2/3 of the data fall within 1SD, ~95% fall within 2SD, and ~99% fall within 3SD.

If you were to sample any Normal distribution, it might look something like the histogram here.

What controls the shape of the distribution?

Bell-shaped curve: This is governed by the general formula $y = e^{-x^2}$. The height of this curve at $x = 0$ (i.e. the y-intercept) is $y = 1$.

The total area under the curve e^{-x^2} is:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \approx 1.77$$

Width of the curve: This is governed by the value of the exponent. The Normal distribution is expressed in terms of the mean, μ , and the variance, σ^2 , of the data. Substituting the exponent with the term $-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2$ gives a curve that is zero-centered around the mean μ and has a standard deviation of σ .

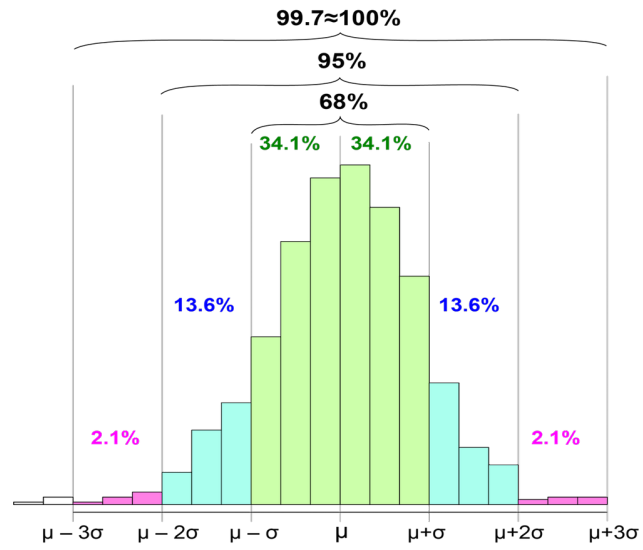


Figure 3: Sampling from the normal distribution

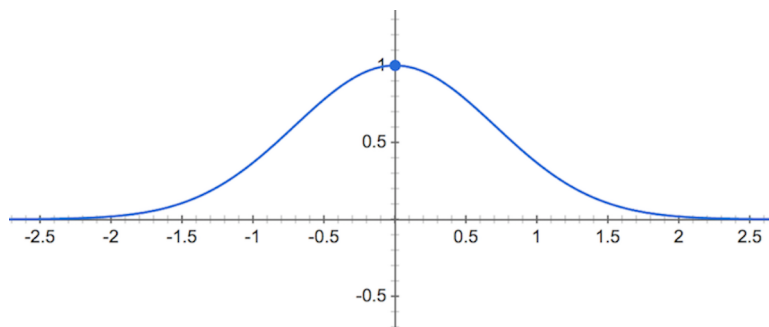


Figure 4: Exponential bell curve

Area under the curve: In order to make the total probability equal to one, we use a scaling factor (let's call this a constant C) that is a multiplier of the exponential formula. This constant turns out to be $C = \frac{1}{\sigma\sqrt{2\pi}}$.

Deriving these equations is not trivial. We end up with the magical number π in the equation because it turns out that using a polar coordinate system, instead of Cartesian coordinates, is more natural to describe this shape. If you want to get an idea of how the Normal PDF is derived, check out this video on YouTube: ****Quick derivation of the Normal PDF (4'50")**** (<https://www.youtube.com/watch?v=ebewBjZmZTw>).

Examples

Uniform distribution

Example: If Starbucks on the corner of Mercer and Greene sells between 2500 and 5000 cups of coffee per day, what's the probability that on any particular day it sold between 3000 and 3500 cups of coffee?

```
punif(c(3500),min=2500, max=5000) - punif(c(3000),min=2500, max=5000)
```

```
## [1] 0.2
```

Normal distribution

The average weight of an adult female greyhound is 63 pounds, with a standard deviation of 8 pounds. What proportion of female greyhounds weigh less than or equal to 55 pounds?

```
pnorm(55, mean = 63, sd = 8)
```

```
## [1] 0.1586553
```

What proportion weigh between 60 and 65 pounds?

$$P(60 \leq X \leq 65) = \int_{60}^{65} f(x)dx = \left[F(x)\right]_{60}^{65} = \int_{-\infty}^{65} f(x)dx - \int_{-\infty}^{60} f(x)dx = F(65) - F(60)$$

```
# note that you cannot use the summation method for a continuous distribution;  
# you have to subtract CDF up to each point to get the probability for the interval  
pnorm(65, mean = 63, sd = 8) - pnorm(60, mean = 63, sd = 8)
```

```
## [1] 0.2448761
```

It can safely be said that 75% weigh no more than what amount? This is a question that calls for the quantile function (inverse CDF):

```
qnorm(0.75, mean = 63, sd = 8)
```

```
## [1] 68.39592
```