

INFO 1301

Prof. Michael Paul

Prof. William Aspray

The Normal Distribution

10, 12 October 2016

[have your textbook open in a separate tab on your computer and turned to page 132]

Background

- Scaling triangles
- To simplify we will work only with right triangles, but scaling works for all triangles.
- Unit circle
- trigonometric table – how does the value of x and y vary as the angle with vertex $(0,0)$ increases
- Calculating lengths of sides on triangles

The normal distribution 1

- The most common curve in all of statistics and in all of the applications of statistics to science and engineering
- Unimodal, symmetric, bell curve
- Few distributions are perfectly normal, but many are almost normal and many applications gain from treating the distribution curve as normal
- Known as the normal distribution (or the Gaussian distribution)
- First mathematical analysis of the normal distribution by Carl Frederic Gauss (1809) – German child prodigy

The normal distribution 2

- The normal distribution is defined by the mean (mu, written as μ) and the standard deviation (sigma, written as σ)
- Written as $N(\mu, \sigma)$
- What does changing the mean do?
- What does changing the standard deviation do?
- Why don't we have to worry about the height?

The normal distribution 3

- Any point x on the standard deviation can be identified by the number of standard deviations it is away from the mean
- This is called the Z-score of x , or $Z(x)$
- It is given by the formula $Z(x) = [x - \mu]/\sigma$
- If $Z(x)$ is negative, x is to the left of the mean.
- If $Z(x) = 0$, x is the mean.
- If $Z(x)$ is positive, x is to the right of the mean.
- The larger the absolute value of $Z(x)$, the more unusual the observation is.

The normal distribution 4

- Let's do an example that appears in the book
- The normal distribution of the SAT is given by $N(1500,21)$, while the normal distribution of the ACT is given by $N(300,5)$.
- Ann gets an 1800 on the SAT, while Tom gets a 24 on the ACT. Which student did better on the college entrance exam?
- Hint: compare the Z scores for Tom and Ann.

The normal distribution 5

- $N(0,1)$ is called the standard normal distribution.
- It plays a similar role in studying the statistics of normal distributions that the circle of radius 1 centered at the origin $(0,0)$ does in trigonometry.
- Draw $N(0,1)$. Note where its mean is located. Note that few points are less than -3 or greater than 3 in this distribution.
- An important fact about $N(0,1)$, which we won't prove is that the area under the curve = 1.

The normal distribution 6

- Suppose we cut the x-axis into a lot of small pieces, each $1/100$ long, and tried to figure out the area of the curve that is to the left of a given point along the x-axis called x_a .
- As x_a moves to the right, the area increases.
- What happens when $x_a = 0$?
- Important fact: In a normal distribution, the median and the mean are identical. Both are at $x_a = 0$.

The normal distribution 7

- Since the mean and median are identical on a normal distribution, on this drawing of $N(0,1)$, the point x_a where $Z=0$ is not only the mean; it is also the median and thus x_a is at the 50th percentile.
- [Remember that percentiles go generally with medians, not means; but in a normal distribution, where the mean = the median, we can talk about percentiles.
- So, let's go back to Ann's score on the SAT. What percentile is she in among all the test-takers of the SAT?

The normal distribution 8

- Ann's percentile would be hard to calculate, though we could eyeball the graph and see approximately where she is.
- However, just like there are trigonometry tables, there are normal distribution tables. They measure the area under the curve to that point = percentile (given the Z value).
- Turn to table 3.8 in your textbook (p.132) [This is a fragment of the full table, which is given in Appendix B.1 (p. 427).]
- How do we read this table?
- What happens to the table entries as the input numbers increase? Is this what we should expect?

The normal distribution 9

- Example: Harken back to the early 2000s. Michael Paul is a high school junior getting ready to take his SATs while listening to 50 Cent and the Black Eyed Peas. What is the probability that his SAT scores on the test will be greater than 1800? (Assume the mean is 1500 and standard deviation is 300 on the three-part 2400 point exam.)
- Do you think his scores would have been higher if he had been listening to Beethoven and Schubert instead?

The normal distribution 10

- Identify the percentage of events that fall within 1, 2, and 3 standard deviations of the mean in a normal probability distribution.
- Use these approximations for everyday rough estimates