# Applications of Calculus to Probability

**Basic terminology** Given numerical data (for example, test scores) there are two common statistics that "summarize" the data:

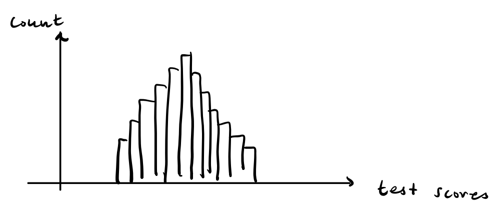
**mean** $\mu = \dfrac{x_1 + \cdots + x_n}{n}$

**standard deviation**

$$\sigma = \sqrt{\underbrace{\frac{(x_1-\mu)^2 + \cdots + (x_n-\mu)^2}{n}}_{\text{average square "deviation" from the mean}}}$$

(square and square root are used for technical reasons)

And a **histogram** can be used to visualize data:



# Normal Distribution

The function $e^{-x^2}$ arises in probability and statistics for the following reason:

if a data set is normally distributed (meaning the data values fall under a bell-shaped normal curve), with mean $\mu$ and standard deviation $\sigma$, the probability that a randomly chosen value lies between a and b is given by

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

$u = \frac{x-\mu}{\sigma}$

$du = \frac{1}{\sigma}\, dx$

$$= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}}\, du$$

This integral cannot be computed by hand, $e^{-u^2/2}$ doesn't have an antiderivative.

<u>Exercise</u> Find the first 6 terms of the Maclaurin series of $e^{-x^2/2}$ and write the series in summation notation.

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5$$

$$e^{-x^2/2} = 1 - \frac{1}{2}x^2 + \frac{1}{2!}\frac{1}{2^2}x^4 - \frac{1}{3!}\frac{1}{2^3}x^6 + \frac{1}{4!}\frac{1}{2^4}x^8 - \frac{1}{5!}\frac{1}{2^5}x^{10} + \ldots$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n} x^{2n}$$

<u>Exercise</u> Suppose a set of test scores has mean $\mu = 100$ and standard deviation $\sigma = 50$.

Use the work in the previous exercise and term-by-term integration to approximate the probability that a randomly chosen test score is between 100 and 200.

$$\int_{\frac{100-100}{50}}^{\frac{200-100}{50}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

$$= \int_0^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^2 \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n} x^{2n}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n} \int_0^2 x^{2n}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n(2n+1)} \, x^{2n+1} \Big|_0^2$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n(2n+1)} \, 2^{2n+1}$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n\,2^{n+1}}{n!\,(2n+1)}$$

$$\approx \frac{1}{\sqrt{2\pi}} \left( 2 - \frac{4}{3} + \frac{8}{2!\cdot 5} - \frac{16}{3!\,7} + \frac{32}{4!\cdot 9} - \frac{64}{5!\cdot 11} \right)$$

$$\approx 0.4729$$

**Exercise** What is probability of score between 0 and 200?

$$\int_{\frac{0-100}{50}}^{\frac{200-100}{50}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

$$= 2\int_0^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \quad \text{(symmetry)}$$

$$\approx 2(0.4729)$$

$$\approx 95\%$$

**Rule of Thumb** 95% of data is within 2 standard deviations of the mean for normally distributed data

**Problem 1.** Suppose a set of standardized test scores is normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 10$. Set up an integral that represents the probability that a test score will be between 90 and 110. Use a Maclaurin series and term-by-term integration to write the value of the integral as an infinite series. Find the sum of the first 6 terms of the series as an approximation for the value of the integral.

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}} \int_0^1 e^{-x^2/2}\,dx$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^1 \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n} x^{2n}\,dx$$

$$= \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n} \int_0^1 x^{2n}\,dx$$

$$= \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n} \frac{1}{2n+1} x^{2n+1}\Big|_0^1$$

$$= \frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n(2n+1)}$$

$$\approx \frac{2}{\sqrt{2\pi}} \left(1 - \frac{1}{6} + \frac{1}{40} - \frac{1}{336} + \frac{1}{3456} - \frac{1}{42240}\right)$$

$$\approx 0.6827$$

68% of data within 1 std dev.

**Problem 2.** You might recall the following fact, that we mentioned in passing earlier in the semester: if the alternating series $\sum_{n=0}^{\infty}(-1)^n a_n$ converges to $L$ then $|s_n - L| \le a_{n+1}$. In words, this says the error in using the $n$th partial sum to approximate the sum of an alternating series series is at most the absolute value of the $n+1$st term of the series. Use this fact to give an estimate for your approximation of the actual probabilities in Problem 1.

This fact says that the value of

$$\frac{2}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,2^n(2n+1)}$$

and

$$\frac{2}{\sqrt{2\pi}} \Big(\underbrace{1}_{n=0} - \underbrace{\frac{1}{6}}_{n=1} + \underbrace{\frac{1}{40}}_{n=2} - \underbrace{\frac{1}{336}}_{n=3} + \underbrace{\frac{1}{3456}}_{n=4} - \underbrace{\frac{1}{42240}}_{n=5}\Big)$$

differ by at most the $n=6$ (next) term:

$$\frac{2}{\sqrt{2\pi}} \frac{1}{6!\,2^6(13)} \approx 1.3319 \times 10^{-6}$$