CS 3430: S19: SciComp with Py
Lecture 16

Definite Integral Approximation
Part 2

Vladimir Kulyukin
Department of Computer Science
Utah State University

Review

Riemann Sum

Suppose there is a partition with n subintervals. Let's pick a point x_i in each subinterval so that x_1 is in subinterval 1, x_2 is in subinterval 2, etc.

Let Δx be the width of each subinterval. Then $f(x_1)\Delta x$ is the area of the rectangle above the first subinterval, $f(x_2)\Delta x$ is the area of the rectangle above the second subinterval, etc.

The Riemann sum is defined as

$$f(x_1)\Delta x + f(x_2)\Delta x + ... + f(x_n)\Delta x = [f(x_1) + f(x_2) + ... + f(x_n)]\Delta x.$$

Midpoint Rule

The midpoint rule is the Riemann sum approximation where we take the middle point from each subinterval in a given partition.

Trapezoidal Rule

$$\int_a^b f(x)dx \approx$$
$$[f(a_0) + 2f(a_1) + ... + 2f(a_{n-1}) + f(a_n)]\frac{\Delta x}{2}.$$

Simpson's Rule

Let M and T be the estimates from the midpoint and trapezoidal rules. Simpson's rule esitmation is

$$S = \frac{2M+T}{3}.$$

Error of Approximation Theorem

Let n be the number of subintervals used in an approximation of the definite integral $\int_a^b f(x)dx$.

1) The error for the midpoint rule is at most $\frac{A(b-a)^3}{24n^2}$, where A is a number such that $|f''(x)| \leq A$ for all $x \in [a, b]$.

2) The error for the trapezoidal rule is at most $\frac{A(b-a)^3}{12n^2}$, where A is a number such that $|f''(x)| \leq A$ for all $x \in [a, b]$.

3) The error for Simpson's rule is at most $\frac{A(b-a)^5}{2880n^4}$, where A is a number such that $|f''''(x)| \leq A$ for all $x \in [a, b]$.

Error Analysis

Motivation

We can use the Error of Approximation Theorem to estimate the error of our intergral approximation.

The main caveat is, of course, that we have to be able to differentiate the function whose definite integral we want to approximate.

Even with this caveat, the estimation of error is valuable for us, because it allows us to compare various rules and heuristics of definite integral approximation.

Problem

Bound the error of using the trapezoidal rule with n = 20 to approximate

$$\int_0^1 e^{x^2}\, dx.$$

Solution

Part 2 is of the Error of Approximation Theorem tells us that we need to compute A such that $|f''(x)| \leq A$ for all $x \in [0,1]$.

Let's compute $f''(x)$. We get $f''(x) = e^{x^2}(4x^2+2)$.

How large $|f''(x)|$ can be? Since $e^{x^2}(4x^2+2)$ is increasing on $[0,1]$ (and we can use the computational tools we've developed in this class to assess it!), the greatest value of $f''(x)$ occurs at $x = 1$.

The greatest value is $(4 \cdot 1^2 + 2)e^{1^2} = 6e$. So the error of approximation using the trapezoidal rule is at most

$$\frac{A(b-a)^3}{12n^2} = \frac{6e(1-0)^3}{12(20)^2} \approx 0.003398.$$

Solution

```
from deriv import deriv
from tof import tof

fex = make_e_expr(make_pwr('x', 2.0))
ddfex = deriv(deriv(fex))
ddfexf = tof(ddfex)
A = ddfexf(1.0)
n = 20
err = 0.0001
trp_err = A*((1 - 0)**3)/(12*(n**2))
assert abs(trp_err - 0.003398) <= err
```

Definite Integral Approximation and Gaussian Curves

Variance and Mean

Let X be a random variable. Let x_1, x_2, ... x_n be its possible values.

The variance of a set of n equally likely values is defined as

$$Var(X) = \frac{1}{n} \sum_1^n (x_i - \mu)^2,$$

where

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i.$$

The standard deviation σ is $\sqrt{Var(X)}$.

Gaussian Functions

A Gaussian function (named after Carl Friedrich Gauss) is any function of the form

$$f(x) = a \cdot e^{-\frac{(x-b)^2}{2c^2}},$$

where a, b, and c are arbitrary real constants.

Gaussian Probability Density Functions

Gaussian functions are used to represent the probability density function (PDF) of a normally distributed variable with the expected value μ and the standard deviation σ

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

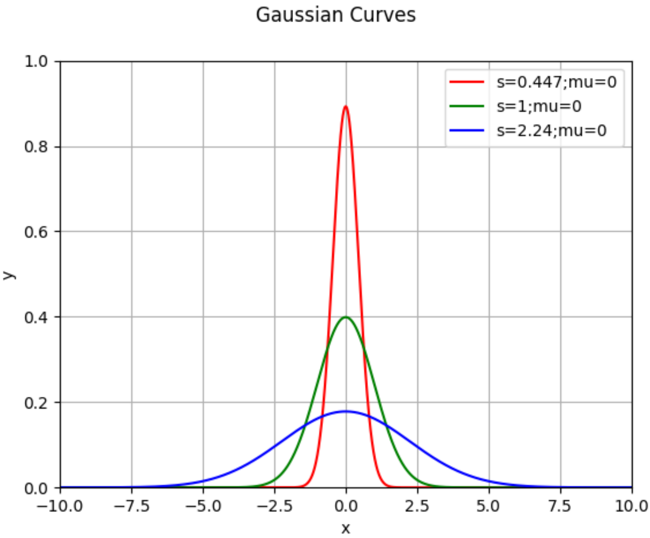
Gaussian PDF in Python

Let's implement this Gaussian PDF in Python

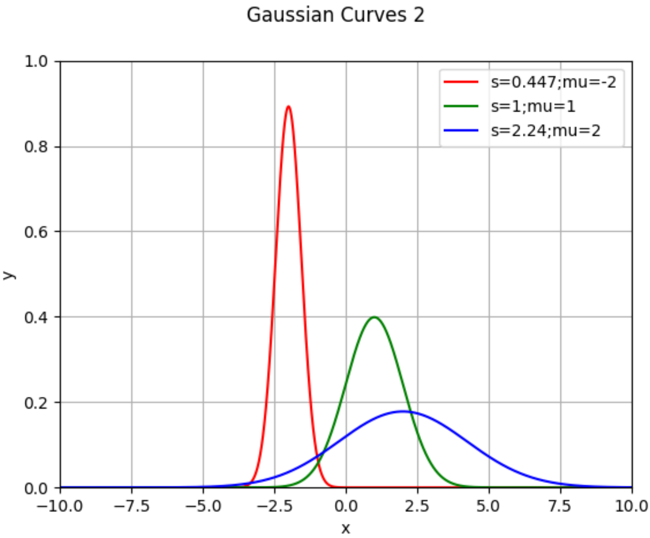
$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

```
def gaussian_pdf(x, sigma=1, mu=0):
    a = 1.0/(sigma*math.sqrt(2*math.pi))
    b = math.e**(-0.5*(((x - mu)/sigma)**2))
    return a*b
```

Plotting Gaussian Curves



Plotting Gaussian Curves



Distribution of IQs

Psychologists use various standard tests to measure intelligence, some less controversial than others.

The coefficient commonly used to describe the results of such tests is an intelligence quotient (IQ).

An IQ is a positive number that, theoretically speaking, indicates how a person's mental age compares with the person's physical age.

The Bell Curve IQ Model

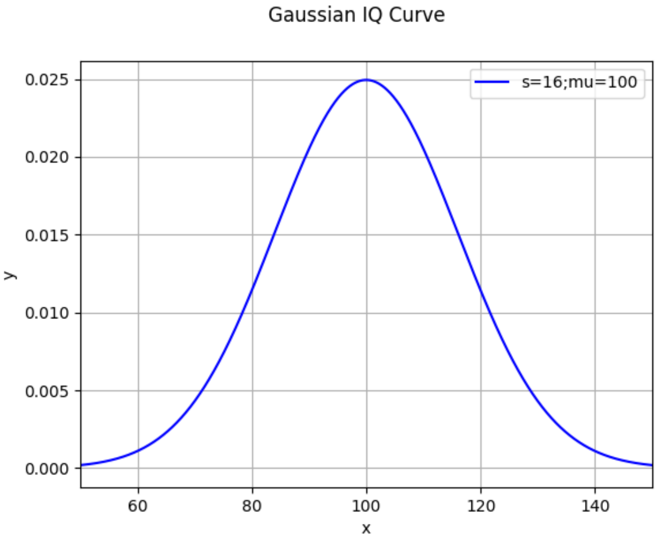
In the Bell Curve IQ model, the median IQ is arbitrarily set at 100, so half the population has an IQ less than 100 and half greater than 100.

IQs are assumed to distribute according to a bell-shaped curve called a **normal curve**.

The proportion of all people having IQs between A and B is given by

$$\int_A^B \frac{1}{16\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-100}{16}\right)^2}.$$

Gaussian IQ Curve (aka Bell Curve)



Problem

Estimate the proportion of the population having IQs between 120 and 126.

Solution

We need to approximate

$$\int_{A=120}^{B=126} \frac{1}{16\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-100}{16}\right)^2}.$$

Let's approximate this definite integral by the simpson rule with $n = 6$ so that $\Delta x = (126 - 120)/6 = 1$.

Solution

```
def gaussian_pdf(x, sigma=1, mu=0):
    a = 1.0/(sigma*math.sqrt(2*math.pi))
    b = math.e**(-0.5*(((x - mu)/sigma)**2))
    return a*b

def approximate_iq():
    iqc = lambda x: gaussian_pdf(x, sigma=16.0, mu=100.0)
    print(simpson_rule_aux(iqc, 120, 126, 6))
```

Here is the output in the Py shell.

```
>>> approximate_iq()
0.0535684928417
```

So, $\approx$ 5.36% of the population have IQs between 120 and 126.

References

1. L. Goldstein, D. Lay, D. Schneider, N. Asmar. *Calculus and its Applications*, Chapters 9. Pearson.
2. `www.python.org`.