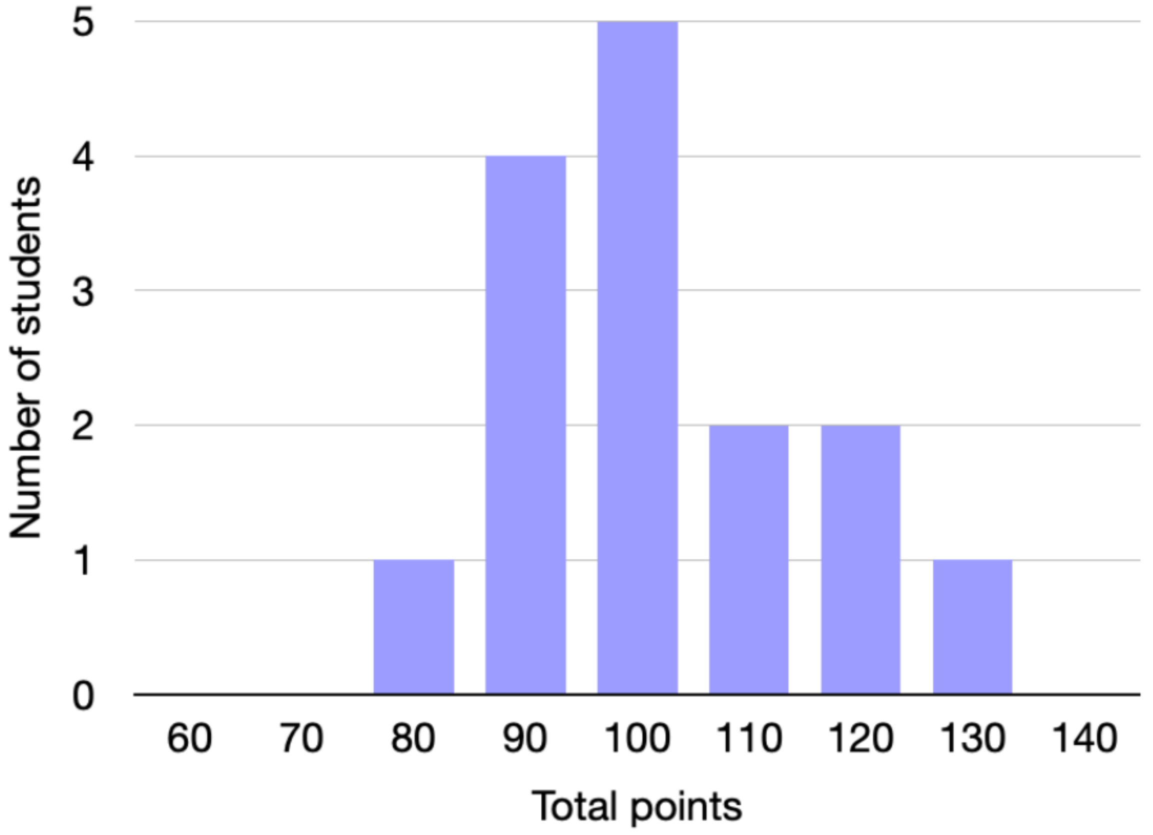
Statistics and Probability

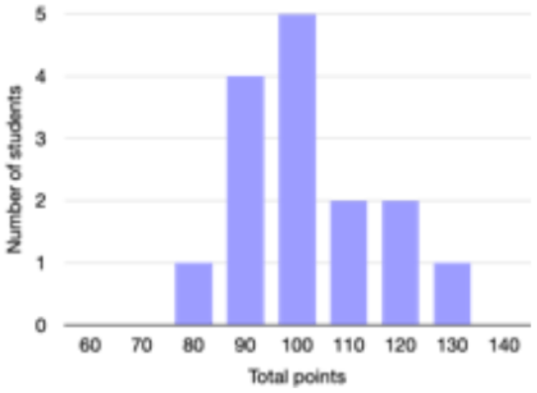
Imagine there is a multiple-choice exam, each question is worth 10 points. There are 15 questions. Let's say these are the grades obtained (the "grade distribution"). Notice that since this is a multiple-choice exam with no partial credit, the grades are "discrete" (e.g., only in multiples of 10 in this case).



What is:

1. The total number of people who took the test?
2. The probability that a randomly selected person got 90 points?
3. The probability that a randomly selected person got 120 points or above?
4. The most probable grade?
5. The median grade?
6. The average (mean) grade (**expectation value**)?
7. The average of the squares of the grades?

Finally, how can we generalize the equations to any function?



1. The total number of people who took the test?
2. The probability that a randomly selected person got 90 points?
3. The probability that a randomly selected person got 120 points or above?
4. The most probable grade?
5. The median grade?
6. The average (mean) grade (**expectation value**)?
7. The average of the squares of the grades?

$N(j)$: number who got j

1. 15 $\quad \sum_{j=0}^{\infty} N(j) = 15$ (N_{Tot})

2. $P(90) = \frac{N(90)}{\sum_{j=0}^{\infty} N(j)} = \frac{N(90)}{N_{tot}} = \frac{4}{15} = 0.267$ $\qquad P(j) = \frac{N(j)}{N_{Tot}}$

3. $P(120) + P(130) = \frac{N(120) + N(130)}{N_{Tot}} = \frac{3}{15} = 0.20$

4. 100 $\qquad$ 5. 100

6. $\frac{1(80) + 4(90) + 5(100) + 2(110) + 2(120) + 130}{15} = \boxed{102}$

$N(j) \times j$

Finally, how can we generalize the equations to any function?

discrete | continuous

$$\langle j \rangle = \frac{\sum_{j=0}^{\infty} j N(j)}{N_{tot}} = \sum_{j=0}^{\infty} j P(j) \qquad \langle j \rangle = \int_0^{\infty} j P(j) dj$$

$$\langle j^2 \rangle = \frac{1 \times 80^2 + 4 \times 90^2 \ldots}{15} = \frac{\sum_{j=0}^{\infty} j^2 N(j)}{N_{tot}} = \sum_{j=0}^{\infty} j^2 P(j)$$

$$= 10580$$

$$\langle j \rangle^2 \neq \langle j^2 \rangle$$

$$102^2 \neq 10580$$

$$10404$$

$$\langle f(j) \rangle = \sum_{j=0}^{\infty} f(j) P(j)$$

discrete

$$\langle f(x) \rangle = \int_0^{\infty} f(x) P(x) dx$$

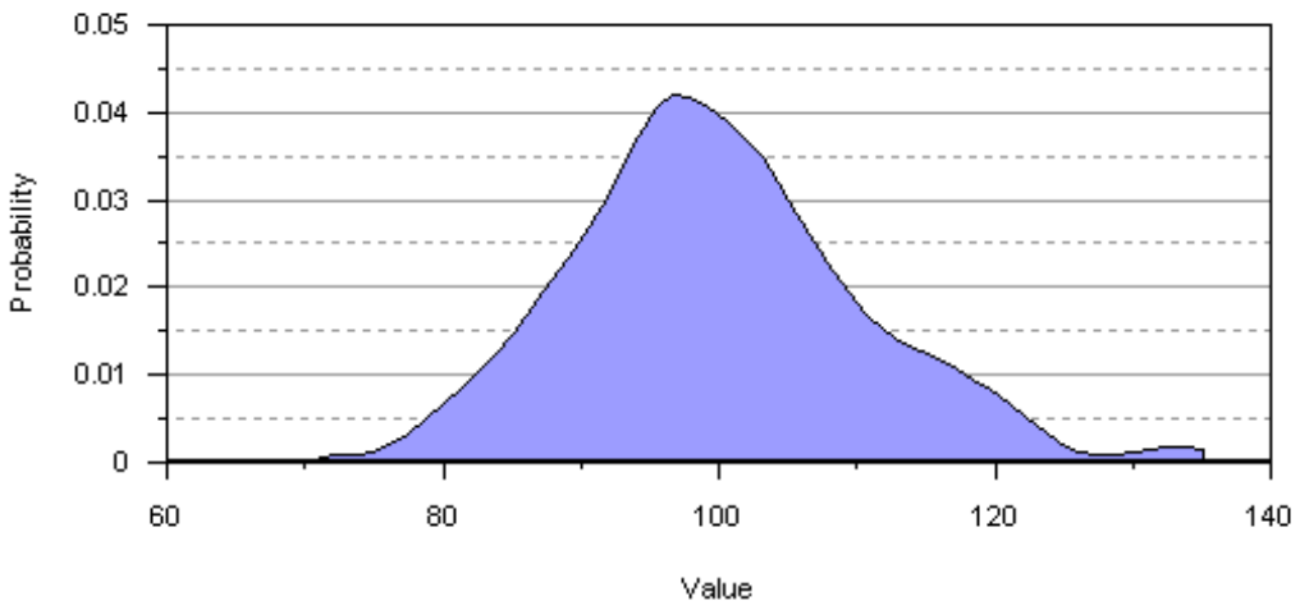
continuous

discrete

$$\sum P(j) = 1$$

$$\int_{-\infty}^{\infty} P(x) dx = 1$$

Continuous variables



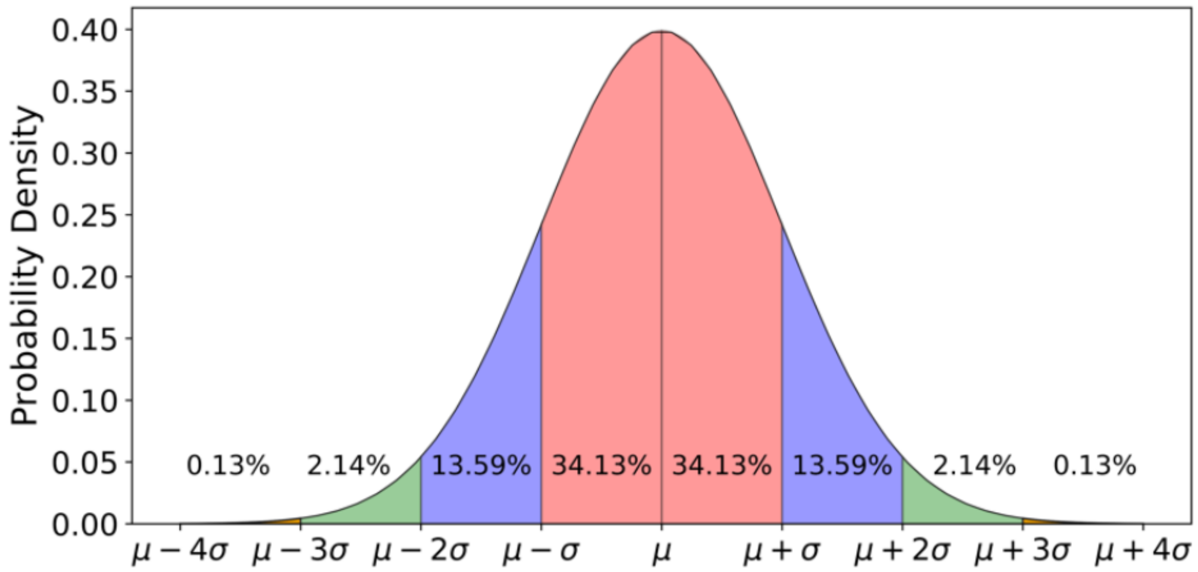
What if the exam had many problems, instead of multiple choice? i.e., the grades are not discrete, and it is possible to score 101.51 points on the exam? What is the probability of finding someone who scored 101.5? 101.51? 101.512? Obviously, the smaller the interval, the less likely it is to find someone to fit that criteria.

Also, what if thousands, or even millions of people had taken the exam?

In this case, it makes much more sense to use a so-called continuous probability distribution to describe the probabilities. You can describe the probability as a function of the variable, called the **probability density function**. This applies to any case where the sample size is very large and the variable (e.g., the grade) is “continuous,” as opposed to “discrete.”

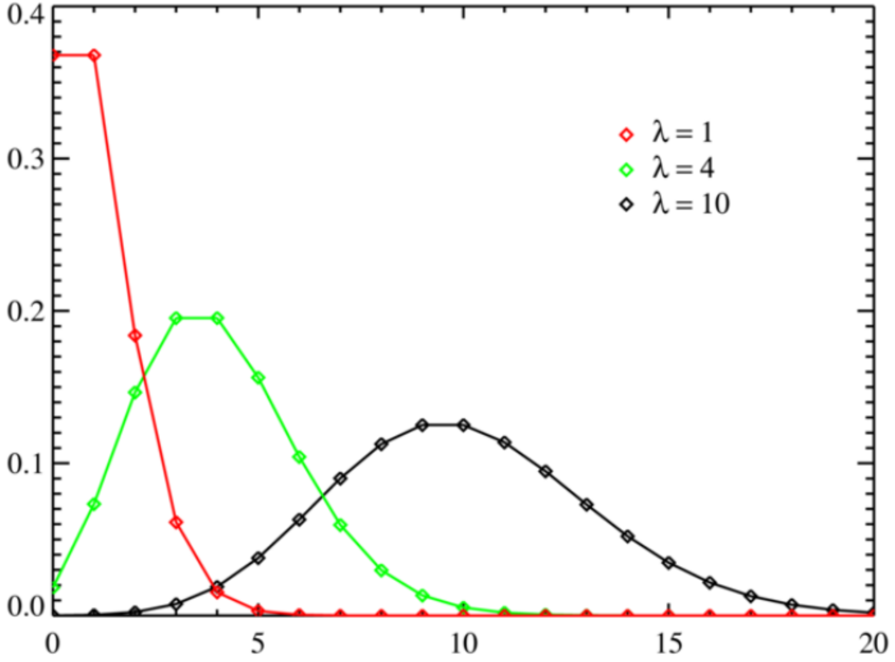
Probability Density Functions: Some common examples

Normal Distribution



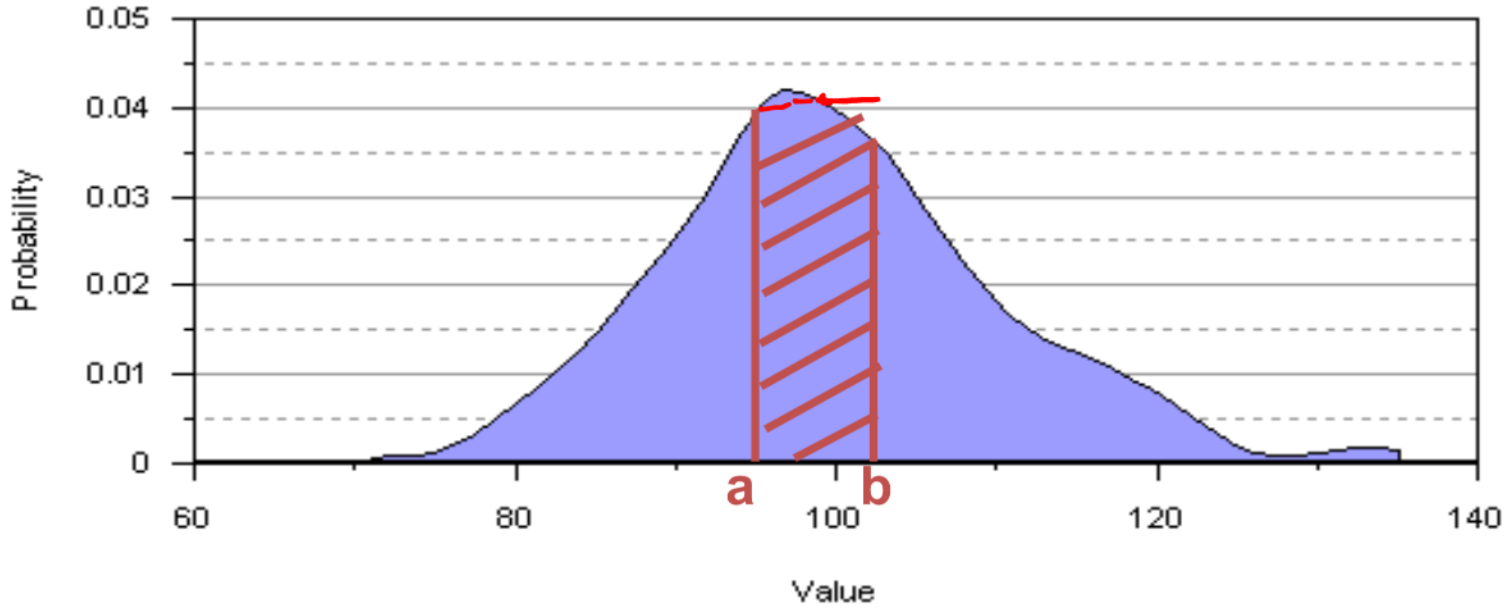
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Poisson distribution



$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

Probability Density Functions



100 – 101

~~100 x 1~~

0.04 (1) = 0.04

The probability that an individual point lies between x and x+dx is: $\rho(x)dx$

ρ(x) is called the **probability density**.
The probability that x lies between **a** and **b** is: $P_{a-b} = \int_a^b \rho(x)dx$

Probability Density Functions

$$\int_{-\infty}^{\infty} \rho(x)dx = 1$$

$$\langle x \rangle = \int_{-\infty}^{\infty} x\rho(x)dx$$

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x)\rho(x)dx$$