Lecture 34: Calculus and Statistics

In this lecture, we look at an application of calculus to statistics. We have already defined the probability density function \( f \) called PDF and its anti-derivative, the cumulative distribution function CDF.

Probability density

Recall that a probability density function is a function \( f \) satisfying \( \int f(x) \, dx = 1 \) and which has the property that it is \( \geq 0 \) everywhere. We say \( f \) is a probability density function on an interval \([a, b]\) if \( \int_a^b f(x) \, dx = 1 \) and \( f(x) \geq 0 \) there. In such a case, we assume that \( f \) is zero outside the interval.

Recall also that we called the antiderivative of \( f \) the cumulative distribution function \( F(x) \) (CDF).

Expectation

The expectation of probability density function \( f \) is

\[
m = \int_{-\infty}^{\infty} x f(x) \, dx .
\]

In the case, when the probability density function is zero outside some interval, we have

The expectation of probability density function \( f \) defined on some interval \([a, b]\) is

\[
m = \int_{a}^{b} x f(x) \, dx .
\]

As the name tells, the expectation tells what is the average value we expect to get.
Variance and Standard deviation

The variance of probability density function $f$ is

$$\int_{-\infty}^{\infty} x^2 f(x) \, dx - m^2,$$

where $m$ is the expectation.

Again, if the probability density function is defined on some interval $[a, b]$ then

$$\int_{a}^{b} x^2 f(x) \, dx - m^2,$$

where $m$ is the expectation of $f$.

The square root of the variance is called the standard deviation.

The standard deviation tells us what deviation we expect from the mean.

Examples

In the lecture, we will compute this in some examples. Here is some sample.

1. The expectation of the geometric distribution $f(x) = e^{-x}$

$$\int x e^{-x} \, dx = 1.$$ 

The variance of the geometric distribution $f(x) = e^{-x}$ is 1 and the standard deviation 1 too.

Remember that we can compute also with Tic-Tac-Toe:

$$\int x^2 e^{-x} \, dx$$

<table>
<thead>
<tr>
<th>$x^2$</th>
<th>$e^{-x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2x$</td>
<td>$-e^{-x}$</td>
</tr>
<tr>
<td>2</td>
<td>$e^{-x}$</td>
</tr>
<tr>
<td>0</td>
<td>$e^{-x}$</td>
</tr>
</tbody>
</table>

2. The expectation of the standard Normal distribution $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$\int_0^{\infty} x (2\pi)^{-1/2} e^{-x^2/2} \, dx = 0.$$ 

3. The variance of the standard Normal distribution $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$\int_0^{\infty} x^2 (2\pi)^{-1/2} e^{-x^2/2} \, dx = 0.$$ 

We can do that by partial integration too. Its a bit more tricky.

The next example is for trig substitution:

4. The distribution on $[-1, 1]$ with function $(1/\pi)(1 - x^2)^{-1/2}$ is called the arcsin distribution. What is the cumulative distribution function? What is the mean $m$? What is the standard deviation $\sigma$? We will compute this in class. The answers are $m = 0, \sigma = 1/\sqrt{2}$. 
The function \( f(x) = \cos(x)/2 \) on \([-\pi/2, \pi/2]\) is a probability density function. Its mean is 0. Find its variance
\[
\int_{-\pi/2}^{\pi/2} x^2 \cos(x) \, dx.
\]

The uniform distribution on \([a, b]\) is a distribution, where any real number between \(a\) and \(b\) is equally likely to occur. The probability density function is \( f(x) = 1/(b - a) \) for \( a \leq x \leq b \) and 0 elsewhere. Verify that \( f(x) \) is a valid probability density function.

Verify that the function which is 0 for \( x < 0 \) and equal to
\[
f(x) = \frac{1}{\log(2)} \frac{e^{-x}}{1 + e^{-x}}
\]
for \( x \geq 0 \) is a probability density function.

A particular Cauchy distribution has the probability density
\[
f(x) = \frac{1}{\pi} \frac{1}{(x - 1)^2 + 1}.
\]
Verify that \( f(x) \) is a valid probability density function.

Find the cumulative distribution function (CDF) \( F(x) \) of \( f \) in the previous problem.