## Probability density function

In our earlier article "Expectation value and Standard deviation" we have seen how the expectation value can be calculated in terms of the possible values and their probabilities. In this article, we will see how this can be done in the case of "continuous" variables.

By closely following Wikipedia article on "Probability density function" let us introduce what it is. Suppose the life time of a certain species of bacteria is around 4 to 6 hours. i.e. it dies around 4 to 6 hours after its birth. Then, what is the probability that a bacterium lives exactly 5 hours? It is zero percent, since it is very unlikely that a bacteria dies exactly $5.000000000000000 \cdots$ hours after its birth, even though it can die approximately 5 hours after its birth.

Therefore, a better and meaningful question to ask is: What is the probability that a bacterium lives longer than 5 hours but shorter than 5.01 hours? Let's say that the answer is 0.02 . Then, we can easily guess that the probability that a bacterium lives between 5 hours and 5.001 hours is around one-tenth of 0.02 , which is 0.002 , as the new interval is one-tenth of the previous one. Similarly, the chance that the bacterium lives between 5 hours and 5.00001 hours is around 0.00002 and so on.

We see here in these examples that the probability a bacteria dies during certain interval divided by that interval is constant, as long as it dies approximately 5 hours after its birth. It is about 2 hour $^{-1}$ (or $2 /$ hour). For example, $0.02 /(0.01$ hour $)=0.002 /(0.001$ hour $)=0.00002 /(0.00001$ hour $)=2 /$ hour . This quantity 2 /hour is called the "probability density" for the bacteria to die 5 hours after its birth.

Therefore, if somebody asks "What is the probability that the bacteria dies 5 hours after its birth?," a meaningful answer would be $2 \operatorname{hour}^{-1}(d t)$ where $d t$ is the infinitesimal interval of time, in which the bacterium dies. For example, the probability that a bacterium dies between 5 hours and 5.00000002 hours after its birth is given by 2 hour $^{-1} \times 0.00000002$ hour which is 0.00000004 .

Let's write this out more mathematically. If we call the life time of bacterium by $t$ and the probability density function for $t$ by $f(t)$, then we have $f(5$ hour $)=2$ hour $^{-1}$. Furthermore, the probability that the bacterium dies between $t$ and $t+d t$ is given by $f(t) d t$. Noticing that the total probability is

1, we also have the following constraint for the probability density function.

$$
\begin{equation*}
\int_{0}^{\infty} f(t) d t=1 \tag{1}
\end{equation*}
$$

In other words, the probability that the life time of a bacterium is between 0 and infinity (i.e. any number) is 1.

Returning to our earlier discussion in the beginning of this article, notice that the life time of a bacterium, measured in hours, is a continuous variable. It can have any values such as $5.00013492 \cdots$ hours or $4.216487 \cdots$ hours. On the other hand, things such as SAT score are discrete variables. You can never get $1537.6109 \cdots$ on SATs. It's either 1530 or 1540 . It is easy to see that it is not really natural to use probability density function for discrete variables such as SAT scores.

Now, let's express how the expectation value can be calculated in terms of the probability density function. Remembering that expectation value can be calculated by summing over the possible values multiplied by their probabilities, we conclude:

$$
\begin{equation*}
\langle x\rangle=\int_{-\infty}^{\infty} x f(x) d x \tag{2}
\end{equation*}
$$

In our later articles on quantum mechanics, we will also see that the probability density function is useful. For example, in quantum mechanics, the probability density function for position $x$ is given by $|\psi(x)|^{2}$ where $\psi(x)$ is what is called a "wave function." Therefore, the expectation value for the position $x$ is naturally given as follows:

$$
\begin{equation*}
\langle x\rangle=\int d x x|\psi(x)|^{2} \tag{3}
\end{equation*}
$$

where

$$
\begin{equation*}
\int d x|\psi(x)|^{2}=1 \tag{4}
\end{equation*}
$$

which amounts to the fact that the total probability is 1.
Problem 1. Express the variance using the integral and the probability density function $f(x)$.

Problem 2. Can the probability density function ever be negative? How about zero?

## Summary

- For continuous variables, it makes more sense to use probability density functions than probability.
- If the probability density function is given by $f(t)$, the probability that the concerned value will be between $t$ and $t+d t$ is given by $f(t) d t$.
- The probability density function satisfies

$$
\int f(t) d t=1
$$

as the total probability is always 1 .

