Linear Transformation of Normal Random Variables and Its Statistical Application

LIN Guozhang (061801907)

The normal distribution plays an important role in the fields of probability and statistics. This report is to discuss about the linear transformation of normal random variables, with the following properties to be proved.

Properties of the linear transformation of normal random variables

1. If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$, where a, $b \in \mathbb{R}$.
2. If X_1, ..., X_n are independent, and $X_i \sim N(\mu_i, \sigma_i^2)$, where $i = 1$, 2, ... n,, then

$$Y := \sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2), \text{ for } a_i \in \mathbb{R}.$$

The usefulness of the two properties will be discussed after their proofs.

Proof of property 1

To prove property 1, two results on page 64 of the lecture notes are used here. They are

1. $M_{aX+b}(t) = \exp(tb) M_X(at)$.
2. For $X \sim N(\mu, \sigma^2)$, $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

The proofs of the results can be found in Lin Guozhang's report, *Properties Related to MGF*.
According to the two results, for $X \sim N(\mu, \sigma^2)$,

$$\begin{aligned} M_{aX+b}(t) &= \exp(tb) \exp(\mu a t + \frac{1}{2}\sigma^2 a^2 t^2) \\ &= \exp[(a\mu + b)t + \frac{1}{2}(a\sigma)^2 t^2] \end{aligned}$$

By the uniqueness theorem of mgf, $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Q.E.D.

Sketch of proof of property 2

By property 1, for two independent random variables $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$,

$$a_1 X_1 \sim N(a_1\mu_1, a_1^2\sigma_1^2)$$

$$a_2 X_2 \sim N(a_2\mu_2, a_2^2\sigma_2^2)$$

As proved in Example 7.57 in the textbook, for two independent random variables $Y_\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$ and $Y_\beta \sim N(\mu_\beta, \sigma_\beta^2)$,

$$Y_\alpha + Y_\beta \sim N(\mu_\alpha + \mu_\beta, \sigma_\alpha^2 + \sigma_\beta^2)$$

Let $Y_\alpha = a_1 X_1$ and $Y_\beta = a_2 X_2$. Then $\mu_\alpha = a_1\mu_1$, $\mu_\beta = a_2\mu_2$, $\sigma_\alpha^2 = a_1^2\sigma_1^2$, $\sigma_\beta^2 = a_2^2\sigma_2^2$

Thus,

$$a_1X_1 + a_2X_2 \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$$

The above is the proof of property 2 for $n = 2$. The general case can be proved by induction.

The two properties are related to the concepts of the sample mean and the standardized normal distribution, which are important in probability and statistics.

Standardized normal distribution
A random variable $X \sim N(\mu, \sigma^2)$ is standardized under the linear transformation

$$Z := \frac{X - \mu}{\sigma}.$$

By property 1, $a = \frac{1}{\sigma}$ and $b = \frac{-\mu}{\sigma}$,

$$Z \sim N(\frac{1}{\sigma}\mu + \frac{-\mu}{\sigma}, \frac{1}{\sigma^2}^2 \sigma^2)$$

That is,

$$Z \sim N(0, 1)$$

One says that Z obeys the standard normal distribution.

Sample mean
For i.i.d. X_1, ..., $X_n \sim N(\mu, \sigma^2)$, where $i = 1$, 2, ... n,, define the sample mean $\overline{X}$ as follows.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

By property 2, let $a_i = \frac{1}{n}$, for $i = 1$, 2, ... n. Then

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n}\frac{1}{n}\mu, \sum_{i=1}^{n}\frac{1}{n^2}\sigma^2)$$

That is,

$$\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$$

As the above shows, the sample mean is also a normal random variable with the same mean μ.
The sample mean can be standardized to obtain

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Application
This example is adapted from Example 4 on page 81 of *Biostatistical Analysis, Fifth Edition* by Jerrold H. Zar.

Horses were given an antibiotic for 2 weeks.

Let random variable X_i denote the weight change of the i-th horse. Assume $X_i \sim N(\mu, 13.4621)$, and that the weight changes for different horses are independent. One is interested in the mean of weight change to decide what effect the antibiotic has on the horses' weight.

Form hypotheses about the mean and check.

Null hypothesis (the hypothesis assumed to be true in the check) :

$$H_0 : \mu = 0.$$

Alternative hypothesis (all the other possibilities) :

$$H_a : \mu \neq 0.$$

In order to check if the null hypothesis is acceptable at the confidence level of 0.05 (to be explained below), measurements of X_i are obtained from 17 horses. They are as follows.

2.0, 1.1, 4.4, -3.1, -1.3, 3.9, 3.2, -1.6, 3.5, 1.2, 2.5, 2.3, 1.9, 1.8, 2.9, -0.3, -2.4 kg.

From these 17 data, the sample mean $\overline{X} = 1.29$ kg.
Assume the null hypothesis is true for the check, that is, $\mu = 0$. Then,

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{1.29 - 0}{\sqrt{13.6421}/\sqrt{17}} = 1.45$$

1.45 is obtained from the calculation with the data. Such a value is called the realization of Z. Note that Z is a random variable that obeys the standard normal distribution, not a number.

By referring to the Z-table (URL in the section of Reference), which provides probabilistic information on Z, one has

$$\mathbb{P}(Z \geq 1.45) = 0.5 - 0.4265 = 0.0735$$

Since the pdf of Z is symmetric about the y-axis,

$$\mathbb{P}(Z \leq -1.45) = 0.0735$$

Define the p-value as the probability of obtaining a value with the same or higher deviation from the mean than the realization.
Then, in this example,

$$\text{p-value} = \mathbb{P}(Z \geq 1.45) + \mathbb{P}(Z \leq -1.45) = 0.1470$$

The criterion to reject the null hypothesis is

$$\text{p-value} \leq \text{confidence level}$$

The intuition is that, assuming the null hypothesis is true, the probability of obtaining the realization or some more extreme value is so small that it is smaller than or equal to the confidence level, a very small number. However, the realization is what one has already obtained. If the null hypothesis is true, then an almost impossible event has occurred. One concludes that that event did not occur, but the problem lies in the null hypothesis and rejects it.

To finish this example, compare the p-value with the confidence level 0.05.

$$0.1470 > 0.05$$

Thus, one cannot reject $H_0 : \mu = 0$, and it is accepted.
In conclusion, the check does not rule out the possibility that the antibiotic has no effect on horses' weight change.

Remark:

1. In a more mathematical term, the check is called hypothesis testing, which is very important in statistics.
2. The test used in this example is two-sided because $H_a : \mu \neq 0$ means that under the alternative hypothesis, μ can be greater or smaller than 0. I will discuss the use of the one-sided test in my later reports, so stay tuned.

References

- *Probability, an introduction* from Grimmett and Welsh
- Lecture notes for SML: Probability by Richard, S.
- *Properties Related to MGF* by Lin Guozhang.
- *Biostatistical Analysis, Fifth Edition* by Jerrold H. Zar
- Z-table: `https://www.statisticshowto.com/tables/z-table/`