Statistical basis for pharmacometrics: random variables and their distribution functions, expected values, and correlation coefficient

Kyungmee Choi*+
Division of Mathematics, College of Science and Technology, Hongik University at Sejong, Jochiwon, Sejong 30016, South Korea
*Correspondence: K. Choi; Tel: +82-44-860-2237, Fax: +82-44-863-2648, E-mail: kmchoi@hongik.ac.kr
+This work was supported by the 2016 Hongik University Academic Research Support Fund.

For pharmacometricians, probability theory is the very first obstacle towards the statistics since it is solely founded on mathematics. The purpose of this tutorial is to provide a simple version of introduction to a univariate random variable, its mean, variance, and the correlation coefficient of two random variables using as simple mathematics as possible. The definitions and theorems in this tutorial appear in most of the statistics books in common. Most examples are small and free of subjects like coins, dice, and binary signals so that the readers can intuitively understand them.

Introduction

When the population is too big to study all, it is necessary to randomly sample a proper size of data in order to figure out its distribution including center and dispersion (Fig. 1). Let us define two important parameters \( \mu \) and \( \sigma^2 \) as the true mean and variance of the population which are unknown unless the whole population is studied. When \( \mu \) and \( \sigma^2 \) are estimated from the sample, the probability theory provides their estimators with the theoretical background like how close the estimators are to the true values.

Suppose \( X_1, X_2, ..., X_n \) are a random sample of size \( n \). Since each of them is supposed to be independent and identically represent the population while each observation \( X_i, i = 1, 2, ..., n \) varies. The histogram of a sample is supposed to predict the distribution curve of the population. As an analogy of the population mean and variance, we can define the sample mean \( \bar{X} \) and variance \( S^2 \) as follows:

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \\
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2
\]

For the notation of estimator, we often use hats such as \( \hat{\mu} = \bar{X}, \hat{\sigma}^2 = S^2 \). Naturally we expect that \( \bar{X} \) converges to \( \mu \) and \( S^2 \) converges to \( \sigma^2 \) as \( n \) is large enough. Then, the probability theory provides the mathematical basis to those asymptotic properties.

Whenever the readers want more details, they can refer to the following excellent literatures on mathematical statistics: Hogg and Craig 4th ed. (pp 1-149),[1] Hogg and Tanis 3rd ed. (pp 1-60),[2] Kim et al. 4th ed. (pp 1-93),[3] Mood 3rd ed. (pp 1-84),[4] Peebles 4th (pp 1-202),[5] Rice 3rd ed. (pp 1-176),[6] Rosner 6th ed. (pp 1-121),[7] and Ross 8th ed.[8]

In this tutorial, we start from defining the probability. The sections are composed of probability, expectations, the mean
and the variance, the moment generating function, and finally the correlation coefficient. Still, the readers are encouraged to review derivative and integral of polynomial function and exponential function if they want to follow the examples of continuous cases. Otherwise, they can skip those examples. The contents are proper for a lecture of two or three hours. The readers who want more subject related examples can look into the biostatistics book by Rosner.[7] This tutorial focuses on mathematics and logic.

**Probability**

Let us first define some words in view of experiments as follows [2]:

- **Experimental unit:** an object such as person, thing, or event about which we collect data.
- **Population:** a set of units that we want to study.
- **Sample:** a subset of a population.
- **Trial:** one observation in an experiment.
- **Outcome:** the result of a trial.
- **Sample space:** a set of all possible outcomes.
- **Event:** a subset of the sample space $S$.
- **Population:** a set of units that we want to study.
- **Experimental unit:** an object such as person, thing, or event about which we collect data.

**Definition 1**

For a given event $A$, the probability of the event $A$ is defined by

$$ P(A) = \frac{n(A)}{n(S)} $$

where $n(A)$ is the size or number of outcomes of the event $A$.

**Axiom 1** Three axioms of probability

- $P(A) \geq 0$
- $P(S) = 1$
- $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

The associated properties are as follows:

- $P(A') = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

**Definition 2** For $A, B \subseteq S$, the conditional probability $P(A | B)$ is said to be the probability of the event $A$ given that the event $B$ has occurred and is defined by

$$ P(A | B) = \frac{P(A \cap B)}{P(B)} $$

if $P(B) > 0$.

**Example 2 Probability and conditional probability**

Suppose we roll a die and observe the upface. Then $S = \{1, 2, 3, 4, 5, 6\}$. Let $A = \{1\}$ and $B = \{1, 3, 5\}$, a set of odd numbers, and $C = \{2, 3\}$. Then, $P(A) = \frac{1}{6}$, $P(B) = \frac{1}{3}$, $P(C) = \frac{1}{3}$, $P(B \cap C) = 1/6$.

$$ P(B \cup C) = P(B) + P(C) - P(B \cap C) = 2/3 $$

If we have an information that the outcome is odd, then $P(\{1\} | odd)$ is one of the three. Thus

$$ P(\{1\} | odd) = \frac{P(\{1\} \cap odd)}{P(odd)} = \frac{n(\{1\})}{n(\{1, 3, 5\})} = \frac{1}{3} $$

Note that we divide $n(A \cap B)$ by $n(B)$ because the sample space $S$ becomes $B$. Here,[2]

$$ P(A | B) = \frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B) / n(S)}{n(B) / n(S)} = \frac{P(A \cap B)}{P(B)} $$

Therefore,

$$ P(A | B) = \frac{P(A \cap B)}{P(B)} $$

From the definition of the conditional probability, the multiplication rule is obtained.

**Theorem 1 Multiplication rule**

$$ P(A \cap B) = P(A)P(B | A) = P(B)P(A | B) $$

**Example 3 Sensitivity and specificity [7]**

Conditional probability is very useful in screening tests. Suppose we have a table of a new test result for a disease. There are two widely used conditional probabilities to measure test abilities.

- Sensitivity = True Positive (TP) rate = $P(\text{+} | \text{D})$ = $\frac{TP}{D}$ = $\frac{11}{14}$
- Specificity = True Negative (TN) rate = $P(\text{-} | \text{No D})$ = $\frac{TN}{\text{No D}}$ = $\frac{84}{86}$

There are two types of error probabilities False Positive (FP) rate and False Negative (FN) rate as follows:

- FP rate = $P(\text{+} | \text{No D})$ = $\frac{FP}{\text{No D}}$ = $\frac{2}{86}$
- FN rate = $P(\text{-} | \text{D})$ = $\frac{FN}{D}$ = $\frac{3}{14}$

| disease/test               | + | - | total |
|----------------------------|---|---|-------|
| D                          | 11 (TP) | 3 (FN) | 14 |
| no D                       | 2 (FP) | 84 (TN) | 86 |
| total                      | 13 | 87 | 100 |

Let us also think about the conditional probability of having
the disease given the + test result.

\[ P(D|+) = \frac{P(TP)}{P(+) = \frac{11}{13} } \]

Similarly, the conditional probability of having the disease given the - test result is as follows:

\[ P(D|-) = \frac{P(FN)}{P(-) = \frac{3}{87} } \]

The last two probabilities are related to the famous Bayes’ formula which we want to skip its derivation. Let us still take a look at an interesting small example in a signal transmission problem through a channel.

**Example 4 Bayes’ formula.** [4]

Let us transmit a binary signal through a channel. There are four cases: (1) send 0 and receive 0, (2) send 0 and receive 1, (3) send 1 and receive 0, (4) send 1 and receive 1. Let us define events \( S_0 \) as send 0, \( S_1 \) as send 1, \( R_0 \) as receive 0, \( R_1 \) as receive 1. Suppose that \( P(S_0) = 0.3, P(S_1) = 0.7, \) and \( P(R_0|S_0) = P(R_1|S_1) = 0.99, P(R_1|S_0) = P(R_0|S_1) = 0.01. \) Then what is the probability that the received signal 0 is the true signal? Then we need to calculate

\[
P(S_0|R_0) = \frac{P(S_0\cap R_0)}{P(R_0)} = \frac{P(S_0)P(R_0|S_0)}{P(S_0)P(R_0|S_0) + P(S_1)P(R_0|S_1)} = \frac{(0.3)(0.99)}{(0.3)(0.99) + (0.7)(0.01)} = 0.9769737
\]

Note that there are only two cases of receiving 0 for the denominator: (1) send 0 and receive 0 given that 0 is sent (2) send 1 and receive 0 given that 1 is sent.

If events \( A \) and \( B \) are independent each other, then the conditional probability of \( A \) given \( B \) would not depend on whether \( B \) has occurred or not. In other words, the occurrence of \( B \) may not change the probability of the occurrence of \( A \). The concept of independence is very important in statistics because the assumption of independence makes all calculation much simpler and easier.

**Definition 3 Events \( A \) and \( B \) are independent if \( P(A|B) = P(A) \) or \( P(B|A) = P(B) \).**

From the definition of conditional probability, the following important theorem holds right away.

**Theorem 2 Events \( A \) and \( B \) are independent if and only if \( P(A \cap B) = P(A)P(B) \).**

Proof. \( P(A|B) = P(A \cap B)/P(B) = P(A) \), so \( P(A \cap B) = P(A)P(B). \)

**Example 5 independent events [4]**

Suppose we have a signal transmission system composed of two parts connected in parallel: upper part (UP) and lower part (LP). The UP has one router and the LP has two routers serially connected: 1 in the UP and 2 and 3 in the LP (Fig. 2). Let \( R \), be the \( i \)th router failure, and assume that each failure is independent with the failure probabilities \( P(R_i) = 0.005, P(R_j) = P(R_k) = 0.008 \). What is the failure probability of transmitting a signal from \( a \) to \( b \)?

![Figure 2. The three routers in a communication system.](image)

\[
P(\text{transmission failure of a signal}) = P(UP \text{ fails} \cap LP \text{ fails}) = P(UP \text{ fails})P(LP \text{ fails})
\]

by independence of UP and LP

\[
= P(R_1)P(R_2 \cup R_3) = P(R_1)(P(R_2) + P(R_3) - P(R_2 \cap R_3)) = P(R_1)(P(R_2) + P(R_3) - P(R_2)P(R_3))
\]

by independence of the three routers’ failure

\[
= (0.005)(0.008 + 0.008 - (0.008)^2)
\]

\[
= 0.00007968
\]

**A random variable and its probability density function**

Let us now define a random variable (rv) \( X \) as a function from the sample space \( S \) to a set of real numbers \( R \) with its corresponding probability density function (pdf).

**Definition 4 X(s) : S→R for s ∈ S. The probability density function (pdf) for a discrete rv \( X \) is defined by**

\[
f(x) = P(X = x)
\]

**Example 6 a rv and its pdf**

Let \( X \) be the number of heads in two coins. Then

\[
S = \{HH, HT, TH, TT\}
\]

\[
X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0
\]

\[
\begin{array}{c|cccc}
 x & 0 & 1 & 2 & \text{total} \\
\hline
 P(X = x) & 1/4 & 1/2 & 1/4 & 1
\end{array}
\]

The pdf is uniquely defined for a rv. Note that the difference between a variable and a random variable depends on whether or not it has its corresponding pdf. For example, in the function
for a simple line \( y = a + bx \), both \( x \) and \( y \) are variables since their values change, but not random variables since they don’t have corresponding pdf’s. The rv is called discrete if its values are countable or continuous if its values are not countable.

**Definition 5** For \( f(x) = P(X = x) \) to be the pdf of a discrete rv, it should satisfy the following two conditions:

\[
  f(x) \geq 0, \quad \sum_x f(x) = 1
\]

Calculate the probability of event \( A \) as follows:

\[
P(X \in A) = \sum_{x \in A} f(x)
\]

For a discrete rv, the probability mass function is more widely used than pdf. In this tutorial, we will stick to pdf just for convenience.

**Example 7** the pdf of a discrete rv

Let \( f(x) = \frac{x}{2} \), for \( x = 1, 2, 3 \) and 0 otherwise. Then \( P(X = 3) = \frac{1}{2} \), \( P(\frac{1}{2} < X < \frac{3}{2}) = f(1) + f(2) = \frac{1}{2} + \frac{1}{2} = \frac{1}{2} \).

**Definition 6** For \( f(x) \) to be a pdf of a continuous rv it should satisfy the following two conditions:

\[
f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1
\]

Calculate the probability of set \( A \) as follows:

\[
P(X \in A) = \int_A f(x)dx
\]

Note that the area under the curve \( f(x) \) is 1 and \( P(A) \) is the area under the curve \( f(x) \) for \( x \in A \). Also \( P(X = x) = 0 \) since the area at a point is zero, so that

\[
P(X < x) = P(X \leq x) = \int_{-\infty}^{x} f(v)dv
\]

**Example 8** the pdf of a continuous rv

For \( f(x) = cx \), \( 0 < x < 1 \) to be a pdf, what is \( c \)?

\[
\int_{0}^{1} cx \, dx = \left[ \frac{cx^2}{2} \right]_{0}^{1} = \frac{c}{2} = 1,
\]

therefore \( c = 2 \).

\[
P(1/3 < X < 1/2) = \int_{1/3}^{1/2} 2xdx = \left[ x^2 \right]_{1/3}^{1/2} = \frac{1}{4} - \frac{1}{9} = \frac{5}{36}
\]

**Definition 7** The cumulative distribution function (cdf) of a rv \( X \) is defined by

\[
F(x) = P(X \leq x)
\]

The distribution function is often used instead of cdf.

**Example 9** the cdf of a discrete rv

Suppose a rv \( X \) has a pdf \( P(X = 0) = 0.3, P(X = 1) = 0.2, P(X = 2) = 0.5 \).

Then its cdf is as follows (Fig. 3):

\[
F(x) = \begin{cases} 
0, & x < 0 \\
0.3, & 0 \leq x < 1 \\
0.5, & 1 \leq x < 2 \\
1, & 2 \leq x 
\end{cases}
\]

**Example 10** the cdf of a continuous rv

Suppose a rv \( X \) has a pdf \( f(x) = 1, 0 < x < 1 \). Then its cdf is as follows (Fig. 3):

\[
F(x) = \begin{cases} 
0, & x < 0 \\
x, & 0 \leq x < 1 \\
1, & 1 \leq x 
\end{cases}
\]

Figure 3. (A) the pdf of Example 9 (B) cdf of Example 9 (C) the pdf of Example 10 (D) the cdf of Example 10.

From the two examples, we can easily derive the following properties of the cdf \( F(x) \):

- \( F(-\infty) = 0, F(\infty) = 1, 0 \leq F(x) \leq 1 \)
- If \( x_1 < x_2 \) then \( F(x_1) \leq F(x_2) \). That is, \( F(x) \) is increasing, but not strictly increasing. We say the cdf is non-decreasing.
- \( P(X = x) = F(x+) - F(x-) \), where \( F(x+) \) is the right limit and \( F(x-) \) is the left limit of \( F(x) \). There could be a set of \( X \) values corresponding to a value of \( F(x) \). We say the distribution function is right-continuous.
- \( P(X > x) = 1 - F(x) \)
Proof. Let us consider the discrete case only.

Definition 8 The joint pdf of the two discrete random variables X and Y is defined by

\[ f(x, y) = P(X = x, Y = y), \]

which is positive and should satisfy

\[ \sum_x \sum_y f(x, y) = 1 \]

For the event A, \( P(A) \) can be evaluated by

\[ P((X, Y) \in A) = \sum_{(x,y) \in A} f(x, y) \]

For the discrete random variables X and Y with the given joint pdf \( f(x, y) \), the marginal pdf’s \( f_x(x) \) and \( f_y(y) \) are defined by

\[ f_x(x) = \sum_y f(x, y), \quad f_y(y) = \sum_x f(x, y) \]

If a vector \((x, y)\) is given, then a value of the joint pdf \( f(x, y) \) is determined since it is a bivariate function. In other words, one joint pdf \( f(x, y) \) corresponds to a vector \((X, Y)\) where both X and Y are random variables.

For continuous random variables we replace \( \sum \) with \( \int \).

Definition 9 For the function \( f(x, y) \) to be the joint pdf of two continuous random variables X and Y, it should satisfy the following two conditions:

\[ f(x, y) \geq 0, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \]

For calculation of the probability of the set \( A \), use

\[ P((X, Y) \in A) = \int_A f(x, y) dx dy \]

Its marginal pdfs are defined by

\[ f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx \]

More details about the joint distribution functions \( F(x, y) = P(X \leq x, Y \leq y) \) are found in Peebles.[5]

Theorem 3 Two random variables X and Y are independent if and only if

\[ f(x, y) = f_x(x)f_y(y) \]

Example 11 independence of discrete random variables

Suppose we flip a coin twice and define the random variables X and Y as follows:

- X = the number of heads in the first flip
- Y = the number of heads in the two flips

Then

\[ S = \{HH, HT, TH, TT\} \]

\[ X(TT) = X(TH) = 0, X(HT) = X(HH) = 1 \]

\[ Y(TT) = 0, Y(HT) = Y(TH) = 1, Y(HH) = 2 \]

Therefore the joint pdf is given as the following table.

| X/Y | 0    | 1    | 2    | \( f_x(x) \) |
|-----|------|------|------|---------------|
| TT  | (1/4)| 0    | 1/2  |               |
| TH  | 0    | (1/4)| 1/2  |               |
| HT  | 0    | 1/4  | 1/4  |               |
| HH  | 0    | 1/4  | 1/4  | 1             |

Let us calculate some probabilities.

\[ P(X \geq 1, Y \geq 1) = f(1, 1) + f(1, 2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \]

\[ P(X < Y) = f(0, 1) + f(0, 2) + f(1, 2) = \frac{1}{2} \]

To check the independence of X and Y, we calculate \( P(X = 0, Y = 0) \) and \( P(X = 0)P(Y = 0) \) and compare them.

\[ P(X = 0, Y = 0) = \frac{1}{4} \neq P(X = 0)P(Y = 0) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} \]

Thus, X and Y are not independent.

Example 12 independence of continuous rv’s

The joint pdf \( f(x, y) \) of two continuous random variables X and Y is given by

\[ f(x, y) = \frac{1}{2} e^{-x-\frac{y}{2}} \quad x > 0, y > 0 \]

Then the marginal pdfs are obtained as follows:

\[ f_x(x) = \frac{1}{2} \int_0^{\infty} e^{-x} e^{-\frac{y}{2}} dy = \frac{1}{2} e^{-x} \left[ -2e^{-\frac{y}{2}} \right]_0^{\infty} = e^{-x}, \quad x > 0 \]

\[ f_y(y) = \frac{1}{2} \int_0^{\infty} e^{-x} e^{-\frac{y}{2}} dx = \frac{1}{2} e^{-\frac{y}{2}} \int_0^{\infty} e^{-x} dx = \frac{1}{2} e^{-\frac{y}{2}}, \quad y > 0 \]

Since \( f(x, y) = f_x(x)f_y(y) \), X and Y are independent.

The mean and the variance

Let us now consider how we can estimate the center of the population distribution. The top priority estimator is the population mean \( \mu \) mentioned at the beginning. The \( \mu \) is the simple
average of the whole observations for a rv $X$ in the population, which is said to be the expected value of a rv $X$. Since we often do not know the whole observations for $X$ and we assume its pdf, we need to define the expected value of a rv $X$ based on its pdf.

**Definition 10** The expected value of a rv $X$ is defined by

$$ \mu = E[X] = \begin{cases} \sum_x x f(x) \text{ for a discrete r.v.} \\ \int_{-\infty}^{\infty} x f(x)dx \text{ for a continuous r.v.} \end{cases} $$

If $X$ is a rv, then a function $g(X)$ is also a rv. Thus the following theorem holds immediately.

**Theorem 4** For a rv $X$ with its pdf $f(x)$, the expected value of $g(X)$ is obtained by

$$ E[g(X)] = \begin{cases} \sum_x g(x)f(x) \text{ for a discrete r.v.} \\ \int_{-\infty}^{\infty} g(x)f(x)dx \text{ for a continuous r.v.} \end{cases} $$

Let us look at a very simple example.

**Example 13 $\mu$ of a rv**

Let a rv $X$ have a pdf as the following table.

| $X=x$ | -1 | 0 | 1 |
|-------|----|---|---|
| $p(x)$ | 0.2 | 0.6 | 0.2 |

Then

$$ \mu = E[X] = (-1)(0.2) + (0)(0.6) + (1)(0.2) = 0. $$

Now let us define another rv $Y$ which is a function of $X$. Let

$$ Y = X^2 $$

Then

$$ E[Y] = E[X^2] = \sum_x x^2 f(x) = (-1)^2(0.2) + (0)^2(0.6) + (1)^2(0.2) = 0.4 $$

Due to the linearity of summation and integration, the following properties can be derived.

**Theorem 5** Let a rv $X$ have the pdf $f(x)$.  
1. $E[a] = \sum_x a(x) = a$  
2. $E[a + bX] = a + bE[X]$ (Linearity of the expectation)

Proof of 2.

$$ E[a + bX] = \sum_x (a + bx)f(x) $$

$$ = a \sum_x f(x) + b \sum_x xf(x) $$

$$ = a + bE[X] $$

As a measure of the data dispersion from the center $\mu$, the variance $\sigma^2$ is used.

**Definition 11** The variance of a rv $X$ with its pdf $f(x)$ is defined by

$$ \sigma^2 = \text{Var}(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx $$

One easy way of calculating $\sigma^2$ is

$$ \sigma^2 = E[(X - \mu)^2] = E[X^2] - 2\mu E[X] + \mu^2 $$

The standard deviation $\sigma$ is defined by $\sqrt{\sigma^2}$. Some useful properties of $\text{Var}(X)$ are as follows:

- $\text{Var}(a) = 0$ since $E[X] = a$ and $E[X^2] = a^2$.
- $\text{Var}(aX + b) = E[(aX + b)^2] - (a\mu + b)^2 = a^2 \text{Var}(X)$  
- For a standardized rv $Z = \frac{X - \mu}{\sigma}$, $E[Z] = 0$, $\text{Var}(Z) = 1$

**Example 14 the mean and the variance of a rv**

Let a rv $X$ have a distribution as the following table.

| $x$  | -1 | 0 | 1 |
|------|----|---|---|
| $f(x)$ | 0.2 | 0.6 | 0.2 |

Then, $\mu = E[X] = 0$, $E[X^2] = 0.4$, $\sigma^2 = E[X^2] - (E[X])^2 = 0.4$

**Example 15 the expected values of a transformed rv**

Let $E[X] = \mu$, $\text{Var}(X) = \sigma^2$, $Z = \frac{X - \mu}{\sigma}$, and $T = 5Z + 1$. Then, $E[T] = 1$, $\text{Var}(T) = 5^2$.

Let us now define the expectation of two random variables.

**Definition 12** Let the rv’s $X$ and $Y$ have a joint pdf $f(x, y)$. Then, for the function $g(X, Y)$,

$$ E[r(X, Y)] = \int \int r(x, y)f(x, y)dx dy $$

As before the linearity of the expectation holds because of the linearity of summation and integral.

**Theorem 6** Let the rv’s $X$ and $Y$ have the joint pdf $f(x, y)$. Then

1. $E[ag(X) + bh(Y)] = aE[g(X)] + bE[h(Y)]$ (Linearity of the expectation)
2. If $X$ and $Y$ are independent, then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

Proof of 1.

$$ E[ag(X) + bh(Y)] = \sum_x \sum_y (ag(x) + bh(y))f(x, y) $$

$$ = a \sum_x \sum_y g(x)f(x, y) + b \sum_x \sum_y h(y)f(x, y) $$

$$ = aE[g(X)] + bE[h(Y)] $$

The expectation of the linear combination of $g(X)$ and $h(Y)$ is the linear combination of $E[g(X)]$ and $E[h(Y)]$.

Proof of 2.
The expectation of product of $g(X)$ and $h(Y)$ is the product of each expectation $E[g(X)]$ and $E[h(Y)]$. We will take a look at the examples at the last section.

**Moment generating function**

In order to measure the location center $\mu$ and the dispersion of the distribution $\sigma^2$, we need $E[X]$ and $E[X^2]$, which we call the first moment and the second moment of the distribution.

**Definition 13** The two types of moments of a rv $X$ are defined by

$$m_n = E[X^n],$$

where $m_1 = \mu$, the population mean and

$$\mu = E[(X-\mu)^n],$$

where the variance is $\sigma^2 = \mu_2/\sigma^2$, and the kurtosis, thickness of the distribution tails, is $\mu_4/\sigma^4 - 3$.

We define a moment generating function from which we can generate moments of the distribution.

**Definition 14** The moment generating function (MGF) of a rv $X$ is given by

$$M_X(t) = E[e^{tX}]$$

Then

$$M'_X(t) = E[(e^{tX})'] = E[Xe^{tX}], \quad M'_X(0) = E[X]$$

Similarly, the $n$th derivative of MGF at 0 generates the $n$th moment of $X$

$$M''_X(0) = E[X^2]$$

Since a MGF corresponds to a rv as in pdf, the MGF uniquely defines the distribution of a rv.

**Example 16** the MGF of a discrete rv

Let us consider the previous rv $X$ that has a pdf as the following table.

| $X = x$ | -1 | 0 | 1 |
|---------|----|---|---|
| $f(x)$  | 0.2| 0.6| 0.2|

Then its MGF is

$$M_X(t) = 0.2e^t + 0.6 + 0.2e^t$$

and

$$M'_X(t) = -0.2e^t + 0.2e^t, \quad M''_X(0) = E[X] = 0$$

$$M'_X(t) = 0.2e^{-t} + 0.2e^t, \quad M''_X(0) = E[X^2] = 0.4, \quad \text{Var}(X) = 0.4$$

**Example 17** the MGF of a continuous rv

Suppose a rv $X$ have pdf $f(x) = e^{-x}, x > 0$. Then its MGF is for $t < 1$,

$$M_X(t) = \int_0^\infty e^{-tx}e^{tx}dx$$

$$= \int_0^\infty e^{(1-t)x}dx$$

$$= \left[ \frac{e^{(1-t)x}}{1-t} \right]_0^\infty$$

$$= \lim_{x \to \infty} \frac{e^{(1-t)x}}{1-t} + \frac{1}{1-t}$$

$$= \frac{1}{1-t}$$

since $\lim_{x \to \infty} e^{(1-t)x} = 0$ for $t < 1$. After differentiating $M_X(t)$, we get

$$M'_X(t) = \frac{1}{(1-t)^2}, \quad M''_X(t) = \frac{2}{(1-t)^3}$$

$$M'_X(0) = E[X] = 1, \quad M''_X(0) = E[X^2] = 2$$

Therefore,

$$\mu = E[X] = 1, \quad \sigma^2 = \text{Var}(X) = E[X^2] - (E[X])^2 = 1$$

For the calculation, the readers should be familiar with the following integral and derivative: $\int e^{ax}dx = (1/a)e^{ax} + C$, $(fg)' = (f'g) + (fg')$.

**Correlation coefficient**

Two random variables $X$ and $Y$ can be correlated. For example, $Y$ increases as $X$ increases, or vise versa. As a measure of their relationship, covariance and correlation coefficient are often used.

**Definition 15** The covariance of rv’s $X$ and $Y$ are defined by

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

If both $X$ and $Y$ increase, then the covariance is positive. If $Y$ decreases as $X$ increases, then the covariance is negative. If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$. Thus, $\text{Cov}(X, Y) = 0$.

**Theorem 7** Let $X$ and $Y$ be two random variables. Then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

The readers are recommended to prove this theorem. If $X$ and $Y$ are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Note that the amount of covariance depends on the measure-
The correlation coefficient of the two random variables \( X \) and \( Y \) is defined by

\[
\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad -1 < \rho < 1
\]

Then \( \rho \) does not depend on the measurement units and \( \rho = 0 \) if \( X \) and \( Y \) are independent.

**Example 18** The correlation coefficient between two discrete random variables

Let us go back to Example 11 of flipping two coins.

\( X = \) the number of heads in the first flip  
\( Y = \) the number of heads in the two flips

| \( X/Y \) | 0  | 1  | 2  | \( f_X(x) \) |
|-----------|----|----|----|-------------|
| 0         | 1/4| 1/4| 0  | 1/2         |
| 1         | 0  | 1/4| 1/4| 1/2         |

\( f_Y(y) \mid \frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 1 \)

Then,

\[
E[X] = (0)(\frac{1}{4}) + (1)(\frac{1}{2}) + (2)(\frac{1}{4}) = \frac{3}{2}, \quad \text{Var}(X) = \frac{1}{4}
\]

\[
E[Y] = (0)(\frac{1}{4}) + (1)(\frac{1}{2}) + (2)(\frac{1}{4}) = 1, \quad E[Y^2] = \frac{3}{2}, \quad \text{Var}(Y) = \frac{1}{2}
\]

\[
\text{Cov}(X, Y) = (0)(1)(\frac{1}{4}) + (1)(1)(\frac{1}{4}) = \frac{3}{4}
\]

Therefore

\[
\rho(X, Y) = \frac{\frac{3}{4}}{\sqrt{\frac{3}{8}} \sqrt{\frac{3}{8}}} = \frac{1.414}{2} = 0.707
\]

The following example would be a challenging example to the readers who are familiar with double integrals.

**Example 19** The correlation coefficient between two continuous random variables

Let \( f(x, y) = 2, 0 \leq x + y \leq 1 \). Then,

\[
P(0 \leq X \leq 1/2, 0 \leq Y \leq 1/2) = 1/2
\]

\[
f_X(x) = \int_0^{1-x} 2dy = 2(1-x), 0 < x < 1
\]

\[
E[X] = \int_0^1 2x(1-x)dx = 1/3, \quad E[X^2] = \int_0^1 2x^2(1-x)dx = 1/6
\]

Thus

\[
\text{Var}(X) = 1/18
\]

Similarly,

\[
f_Y(y) = 2(1-y), 0 < y < 1, E[Y] = 1/3, E[Y^2] = 1/6, \text{Var}(Y) = 1/18
\]

Also,

\[
E[XY] = \int_0^1 \int_0^{1-y} 2xydxdy = 1/12
\]

\[
\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = -1/36
\]

Therefore,

\[
\rho = \frac{-1/36}{\sqrt{1/18} \sqrt{1/18}} = -1/2
\]

**Discussion**

Since the contents of this tutorial is more like a lecture note and limited, readers are strongly encouraged to read references for further contents. Hopefully, pharmacometricians who read this tutorial get more familiar with statistical notations and probability theories.

**Conflict of interest**

The author has no conflict of interest.

**References**

1. Hogg RV, Craig AT. Introduction to Mathematical Statistics, 4th ed. Macmillan, Publishing Company, 1978.
2. Hogg RV, Tanis EA. Probability and Statistical Inference, 3rd ed. Macmillan, Publishing Company, 1983.
3. Kim WC et al. Introduction to Statistics, 4th ed. Youngchi, Seoul, 2005.
4. Mood AM, Graybill FA, Boes DC. Introduction to the Theory of Statistics, McGraw-Hill, New York, 1974.
5. Peebles PZ JR. Probability, Random Variables, and Random Signal Principles, 4th ed, McGraw-Hill Korea, 2001.
6. Rice JA. Mathematical Statistics and Data Analysis, 3rd ed. Brooks/Cole Cengage Learning, 2007.
7. Rosner B. Fundamentals of Biostatistics, 6th ed. 2006.
8. Ross S. A First Course in Probability and Statistical Inference, 8th ed. Pearson Education, New Jersey, 2010.