§ 1.2 Real Random Variable, Random Vector

Let (Ω, P) be the probability space of a trial $\boldsymbol{T}$. The real-valued function $X(\omega)$ defined on Ω is called the **real random variable** on (Ω, P). It represents the quantity of randomness in the value taken by $X(\omega)$ when the resulting sample point ω of $\boldsymbol{T}$ has appeared.

As discussed in the previous section, the probability space (Ω, P) of the trial of rolling a dice is given by

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \qquad P(A) = \frac{\sharp A}{6}.$$

Now, when the wager on rolling a dice is such that the payoff received is twice the points, the payoff X is a real random variable defined by

$$X(\omega) = 2\omega .$$

The probability space (Ω, P) of the trial of rolling the dice twice consecutively is given by

$$\Omega = \big\{(i,j) \,\big|\, i = 1,2,3,4,5,6\big\} = \{1,2,3,4,5,6\}^2, \qquad P(A) = \frac{\sharp A}{36}.$$

In this trial, let the number of points appearing in the first roll be X_1, and that of the second roll be X_2, and their sum be X. These are real probability variables defined by, respectively,

$$X_1(i,j) = i, \qquad X_2(1,j) = j, \qquad X(i,j) = i + j.$$

Moving back to the earlier generic case, let $X = X(\omega)$ be the real probability variable. The set of all possible values taken by $X(\omega)$, that is, $X(\Omega)$ is called e the **sample space** of X, and it is denoted by Ω^X. For any $B \subset \Omega^X$, the probability of the event "the value of X is in B" is $P\{\omega | X(\omega)\}$, that is, equal to $P\big(X^{-1}(B)\big)$. When viewed as a function of B and written as $P^X(B)$, it is easy to understand why, with P^X being a set function on Ω^X, $P^X(B)$, satisfies the conditions for probability measure, and

$$X^{-1}(B_1 + B_2) = X^{-1}(B_1) + X^{-1}(B_2), \qquad X^{-1}\big(\Omega^X\big) = \Omega$$

The probability measure P^X is called the **probability law** of X. In particular

$$P^X\{b\} = P\big(X^{-1}(b)\big), \qquad b \in \Omega^X.$$

Since P^X is the probability measure on Ω^X, for any $B \subset \Omega^X$,

$$P^X(B) = \sum_{b \in B} P^X\{b\},$$

and P^X is completely determined by having given $P^X\{b\}$ $(b \in \Omega^X)$.

From the above example of rolling the dice twice, let us find the probability space of real probability variables X_1, X_2, X that appear as a result. As can be seen from Figure 1.1,

$$\Omega^{X_1} = \Omega^{X_2} = \{1,2,3,4,5,6\}\,, \qquad \Omega^X = \{1,2,3,4,5,6\}\,,$$
$$P^{X_1}\{k\} = P\big(X_i^{-1}(k)\big) = \frac{\sharp X_i^{-1}(k)}{36} = \frac{6}{36} = \frac{1}{6}\,, \quad i = 1,2,\,,$$
$$P^X\{k\} = P\big(X^{-1}(k)\big) = \frac{\sharp X^{-1}(k)}{36} = \begin{cases} k - 1/36 & (2 \le k \le 7)\,, \\ (13-k)/36 & (8 \le k \le 12)\,. \end{cases}$$

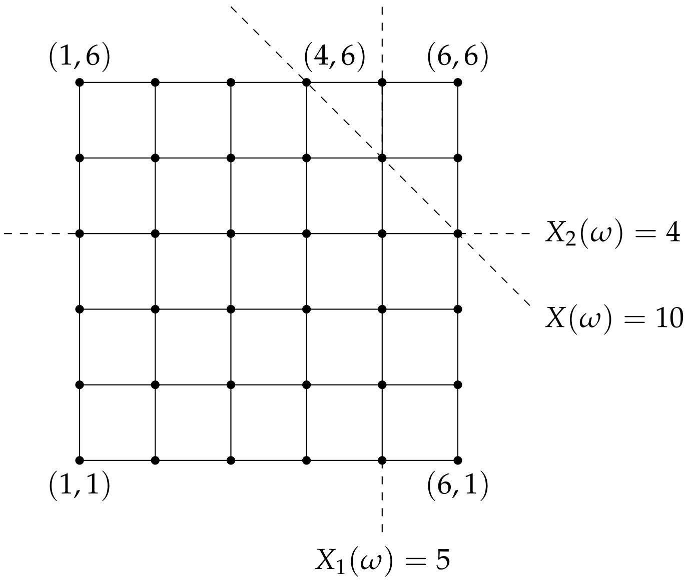

Figure 1.1

Returning to the general probability space (Ω, P), when the real probability variables $X_1(\omega)$, $X_2(\omega)$ are juxtaposed, a function that takes 2-dimensional values is obtained:

$$X(\omega) = \big(X_1(\omega), X_2(\omega)\big)\,.$$

It is termed 2-dimensional **probability vector variables** (or simply **probability vector**). Even in this instance, the sample space Ω^X of X, the probability law P^X are defined as

$$\Omega^X = X(\Omega)\,,$$
$$P^X(B) = P\big\{\omega | X(\omega) \in B\big\} = P\big(X^{-1}(B)\big)\,, \qquad B \subset \Omega^X\,.$$

Ω^X is a subspace of the 2-dimensional space $\boldsymbol{R}^2$, and P^X is the probability measure on Ω^X. Since

$$\Omega^X \subset \Omega^{X_1} \times \Omega^{X_2}$$

it is necessary that the sets on both sides are not equal. $X_1(\omega)$, $X_2(\omega)$ are the **component variables** of $X(\omega)$, and $X(\omega)$ is the **composite variable** of $X_1(\omega)$, $X_2(\omega)$. The probability law P^X is the **composite probability law** of X_1 and X_2 (or the **combined distribution**). For a generic natural number n, the ***n*-dimensional probability vector** is defined in the same manner.

To a point $(x_1, x_2, \cdots, x_n)$ in n-dimensional space $\boldsymbol{R}^n$, the function (can also be said as mapping) corresponding to its k-th component x_k is called the ***k*-image**, and is denoted by π_k. In other words

$$\pi_k(x_1, x_2, \cdots, x_n) = x_k \,.$$

If $X(\omega)$ is the composite variable of $X_1(\omega), X_2(\omega), \cdots X_k(\omega)$, then

$$X_k(\omega) = \pi_k\big(X(\omega)\big) \qquad \text{and hence} \quad X_k = \pi_k \circ X \,.$$

The small circle $\circ$ denotes the composition of functions.

Let $X(\omega)$ be the probability variable, Ω^X, P^X be the sample space and probability law. For a real-valued function defined on Ω^X, $\varphi : \Omega^X \longrightarrow \boldsymbol{R}^1$, if we let

$$Y(\omega) = \varphi\big(X(\omega)\big) \,, \qquad \text{i.e.} \quad Y = \varphi \circ X \,,$$

then a new real probability variable $Y(\omega)$ is obtained. Its sample space Ω^Y, probability law P^Y are of course given by

$$\begin{aligned} \Omega^Y &= (\varphi \circ X)(\Omega) \\ P^Y(C) &= P\big((\varphi \circ X)^{-1}(C)\big) \end{aligned}$$

The real probability variable $Y(\omega)$ obtained in this way is called the **function** of $X(\omega)$. More generically, when $X(\omega)$ is a n-dimensional probability vector, φ is a mapping from $\Omega^X(\subset \boldsymbol{R}^n$ onto $\boldsymbol{R}^m$, $Y(\omega) = \varphi\big(X(\omega)\big)$ becomes the m-dimensional probability vector. Even in this case, $Y(\omega)$ is the **function** of $X(\omega)$. In particular, if φ is taken as the k-image π_k, $\pi_k\big(X(\omega)\big)$ is equal to the k-th component of $X(\omega)$, and hence $X_k(\omega)$ is a function of $X(\omega)$.

Theorem 1.2 When $Y(\omega) = \varphi\big(X(\omega)\big)$

$$\Omega^Y = \varphi\big(\Omega^X\big) \,, \qquad P^X(C) = P^X\big(\varphi^{-1}(C)\big) \,.$$

Proof Since $Y(\omega) = \varphi(X(\omega)) = (\varphi \circ X)(\omega)$,

$$\Omega^Y = (\varphi \circ X)(\Omega) = \varphi(X(\Omega)) = \varphi(\Omega^X)$$
$$P^Y(C) = P\Big((\varphi \circ X)^{-1}(C)\Big) = P\Big(X^{-1}(\varphi^{-1}(C))\Big) = P^X(\varphi^{-1}(C))\,.$$

In the above proof, $(\varphi \circ X)^{-1}(C) = X^{-1}(\varphi^{-1}(C))$ is used. Its proof is

$$\omega \in (\varphi \circ X)^{-1}(C) \Longleftrightarrow (\varphi \circ X)(\omega) \in C \Longleftrightarrow \varphi(X(\omega)) \in C$$
$$\Longleftrightarrow X(\omega) \in \varphi^{-1}(C) \Longleftrightarrow \omega \in X^{-1}(\varphi^{-1}(C))\,. \blacksquare$$

With respect to the real probability variable $X(\omega)$,

$$EX = \sum_{\omega \in \Omega} X(\omega) P\{\omega\}$$

is the **average value** (or **expected value**). Suppose $A \in \Omega$ is any set. Then

$$E(X, A) = \sum_{\omega \in A} X(\omega) P\{\omega\}\,.$$

When $X(\omega)$ is the probability vector, the **average value vector** EX is defined component-wise. In other words, if $X_k(\omega)$ is the k-th component of $X(\omega)$, then

$$EX = \big(EX_1, EX_2, \cdots, EX_n\big) \in \boldsymbol{R}^n\,.$$

Theorem 1.3 If X and Y are probability vectors, then

(i) (**Additive property of average value**) $E(aX + bY) = aEX + b\,EY$ (a and b are constants),

(ii) $E\Big(X, \sum_{i=1}^n A_i\Big) = \sum_{i=1}^n E(X, A_i)$,

(iii) If $X(\omega) \equiv a$ (a is a constant vector) in on A, then $E(X, A) \equiv a\,P(A)$, and in particular $E(a) = a$ (the a on the left is a probability vector that is constantly equal to a),

(iv) $EX = \sum_{x \in \Omega^X} x P^X\{x\}$,

(v) If $Y(\omega) = \varphi(X(\omega))$, then $EY = \sum_{x \in \Omega^X} \varphi(x) P^X\{x\}$.

If X and Y are real probability variables,

(vi) $X(\omega) \geqq Y(\omega) \Longrightarrow EX \geqq EY,\ E(X, A) \geqq E(Y, A)$,

(vii) $X(\omega) \geqq 0\ \ A \subset B \Longrightarrow E(X, A) \leqq E(X, B)$.

Proof Except (iv) and (v), the rest are easy and left them to the reader. Prove only (iv) and (v).

(iv) Let $A_x = X^{-1}(x)$, and

$$\Omega = \sum_{x\in\Omega^X} A_x\,,$$

hence

$$E(X) = E(X,\Omega) = \sum_{x\in\Omega^X} E(X, A_x)\,,$$

As $X(\omega) = x$ on A_x,

$$E(X, A_x) = xP(A_x) = xP^X\{x\}\,.$$

From these two formulas, the formula (iv) is obtained.

(v) In the same way as above,

$$E\big(\varphi(X)\big) = \sum_{x\in\Omega^X} E\big(\varphi(X), A_x\big) = \sum_{x\in\Omega^X} \varphi(x)P(A_x) = \sum_{x\in\Omega^X} \varphi(x)P^X\{x\}\,,$$

and the formula (v) is obtained. ∎

When A is any arbitrary subset of Ω, the **indicator function** 1_A is defined by

$$1_A(\omega) = \begin{cases} 1, & \omega \in A\,, \\ 0, & \omega \in A^c\,. \end{cases}$$

It is a probability variable that takes value of 1 if event A occurs, and 0 if it does not happen. Clearly,

$$1_\Omega(\omega) \equiv 1\,.$$

Under Theorem 1.3's (iii), it is mentioned that a is a probability variable that is constantly equal to a, more precisely it should be written as $a\,1_\Omega(\omega)$. However, such abuse of language is quite common across the field of mathematics, and henceforth as long as confusion does not occur, $a\,1_\Omega(\omega)$ is simply written as a. Clearly,

$$E1_A = P(A)\,, \qquad\qquad E(X, A) = E(X1_A)\,.$$

Theorem 1.4 (i) $1_A^c = 1 - 1_A$,

(ii) If $A = \bigcap_{i=1}^n A_i$, then

$$1_A = \prod_{i=1}^n 1_{A_i} = \min\left(1_{A_1}, 1_{A_2}, \cdots, 1_{A_n}\right).$$

(iii) If $A = \bigcup_{i=1}^n A_i$, then

$$1_A = 1 - \prod_{i=1}^n \left(1 - 1_{A_i}\right) = \max\left(1_{A_1}, 1_{A_2}, \cdots, 1_{A_n}\right).$$

Proof (i) and (ii) are simple. To show the first part of (iii), using (i) and (ii), it suffices that

$$1 - 1_A = 1_{A^c} = \prod_{i=1}^{n} 1_{A^c} \qquad \left(\text{Note that } \left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c\right)$$
$$= \prod_{i=1}^{n} \left(1 - 1_{A_i}\right)$$

Moreover, it is easy to understand that 1_A is equal to $\max\left(1_{A_1}, 1_{A_2}, \cdots, 1_{A_n}\right)$. ∎

The **variance** $V(X)$ of probability variable X is defined by

$$V(X) = E\big((X - EX)^2\big).$$

Under Theorem 1.3's (v), let $\varphi(x) = (x - EX)^2$,

$$V(X) = \sum_{x \in \Omega^X} \big(x - E(X)\big)^2 P^X\{x\}.$$

The **covariance** of two probability variables X and Y is defined by

$$V(X, Y) = E\big((X - EX)(Y - EY)\big).$$

Under Theorem 1.3's (v), use the composite variable (X, Y) instead of X, let $\varphi(x, y) = (x - EX)(y - EY)$, and

$$V(X, Y) = \sum_{(x,y) \in \Omega^{(X,Y)}} (x - EX)(y - EY) P^{(X,y)}\{(x, y)\}.$$

Clearly, $V(X, Y) = V(Y, X)$, $V(X, a) = 0$ (a is constant), $V(X, X) = V(X) \geqq 0$.

Theorem 1.5 (i) $V(aX + b) = a^2 V(X)$, $\;V(aX + b, cY + d) = acV(X, Y)$,

(ii) $V(aX + bY) = a^2 V(X) + 2abV(X, Y) + b^2 V(Y)$

(iii) $V(X) = EX^2 - (EX)^2$, $V(X, Y) = E(XY) - EX\,EY$

(iv) $\left|V(X, Y)\right| \leqq \sqrt{V(X)V(Y)}$

Proof To show (i), (ii), and (iii), take the average value on both sides of the following equations:

$$\big((aX + b) - E(aX + b)\big)^2 = a^2\big(X - E(X)\big)^2,$$
$$\big((aX + b) - E(aX + b)\big)\big(cY + d - E(cY + d)\big) = ac(X - EX)(Y - EY),$$

$$
\begin{aligned}
&\big((aX+bY)-E(aX+bY)\big)^2 = a^2(X-EX)^2 + 2ab(X-EX)(Y-EY) + b^2(Y-EY)^2\,,\\
&(X-EX)^2 = X^2 - 2(EX)X + (EX)^2\,,\\
&(X-EX)(Y-EY) = XY - (EX)Y - (EY)X + EX\,EY\,.
\end{aligned}
$$

To show (iv), for any real number t,

$$
\begin{aligned}
0 \leqq V(tX+Y) &= E(tX+Y) - \big(E(tX+Y)\big)^2\\
&= E\Big(t^2(X-EX)^2 + 2t(X-EX)(Y-EY) + (Y-EY)^2\Big)\\
&= t^2V(X) + 2tV(X,Y) + V(Y)\,.
\end{aligned}
$$

Accordingly, in the last quadratic expression with respect to t, regardless of the value of t, it is non-negative. Taking the determinant

$$
V(X,Y)^2 - V(X)V(Y) \leqq 0\,, \qquad \text{namely} \qquad V(X,Y) \leqq \sqrt{V(X)V(Y)}\,. \ \blacksquare
$$

With regard to the conditions of real probability variables $X(\omega), Y(\omega)$, and $Z(\omega)$, for example,

$$
X(\omega) + Y(\omega) \geqq Z(\omega)
$$

is a condition concerning ω also, hence may be considered as an event. Its probability is

$$
P\big\{\omega \big| X(\omega) + Y(\omega) \geqq Z(\omega)\big\}
$$

and written simply as

$$
P\big\{X(\omega) + Y(\omega) \geqq Z(\omega)\big\} \qquad \text{or} \qquad P\big\{X + Y \geqq Z\big\}\,.
$$

When the probability of an event $\alpha(\omega)$ is 1, it is said to be "**almost surely** $\alpha(\omega)$ is valid," and the notation is

$$
\alpha(\omega) \quad \text{a.s.} \qquad \text{(a.s. = almost surely)}
$$

It is also called "$\alpha(\omega)$ is valid **with probability 1**."

The **standard deviation** $\sigma(X)$ of X, and the **correlation coefficient** $R(X,Y)$ of X and Y, are respectively defined as

$$\sigma(X) = \sqrt{V(X)}\,, \qquad R(X,Y) = \frac{V(X,Y)}{\sigma(X)\sigma(Y)} \quad \text{when } \sigma(X), \sigma(Y) > 0\,.$$

$$X(\omega) = X'(\omega) \text{ a.s.} \Longrightarrow EX = EX' \qquad V(X) = V(X')\,, \quad \sigma(X) = \sigma(X')$$

$$X(\omega) = X'(\omega) \text{ a.s.} \quad Y(\omega) = Y'(\omega) \text{ a.s.} \Longrightarrow V(X,Y) = V(X',Y')\,, \quad R(X,Y) = R(X',Y')\,.$$

To show that these are true, it is sufficient to prove the statements with respect to E. Let $A = \{\omega | X(\omega) \neq X(\omega')\}$, $P(A) = 0$, then for any $\omega \in A$, $P\{\omega\} = 0$. As a result, $E(X,A) = E(X',A) = 0$, for $\omega \in A^c$, $X(\omega) = X'(\omega)$. Consequently $E(X,A^c) = E(A',A^c)$, and $EX = EX'$.

Theorem 1.6 $X(\omega) \geqq 0\,, EX = 0 \Longrightarrow X(\omega) = 0$ a.s.

Proof Let $A = \{\omega | X(\omega) > 0\}$. On $\omega \in A^c$, $X(\omega) = 0$. Hence

$$EX = E(X,A) = \sum_{\omega \in A} X(\omega) P\{\omega\}\,.$$

If $P(A) > 0$, then there exists $\omega_0 \in A$ such that $P\{\omega_0\} > 0$. Of course $X(\omega_0) > 0$. And for all ω, since $X(\omega) \geqq 0$, $EX \geqq X(\omega_0)P\{\omega_0\} > 0$. Hence if $EX = 0$, then $P(A) = 0$, that is, $X(\omega) = 0$ a.s. ▮

Theorem 1.7 (i) $\sigma(X) = 0 \Longleftrightarrow X(\omega) = EX$ a.s.,

(ii) $-1 \leqq R(X,Y) \leqq 1$,

(iii) If $R(X,Y) = \pm 1$, then

$$Y(\omega) - EY = \pm \frac{\sigma(Y)}{\sigma(X)} \Big(X(\omega) - Y(\omega \Big) \text{ a.s.}$$

Proof (i) Since $\sigma(X)^2 = E(X - EX)^2$, it is clear from Theorem 1.6.

(ii) It is clear from Theorem 1.5's (iv).

(iii) Let $k = \sigma(Y)/\sigma(X)$. From Theorem 1.5's (ii) and the definitions of $\sigma(X)$ and $R(X,Y)$,

$$\begin{aligned} \sigma(Y - kX)^2 &= \sigma(Y)^2 - 2k\sigma(X)\sigma(Y)R(X,Y) + k^2\sigma(X)^2 \\ &= \sigma(Y)^2 - 2\sigma(Y)^2 R(X,Y) + \sigma(Y)^2 \\ &= 2\sigma(Y)^2\big(1 - R(X,Y)\big)\,, \end{aligned}$$

and it follows that

$$R(X,Y) = 1 \Longleftrightarrow \sigma(Y - kX) \Longleftrightarrow Y(\omega) - kX(\omega) = EY - kEX \text{ a.s.}$$

$$Y(\omega) \Longleftrightarrow k\big(X(\omega) - EX\big) \text{ a.s.}$$

The case of $R(X,Y) = -1$ can be discussed in the same way. ▮

Theorem 1.8 (**Čebysěv's Inequality**) For $a > 0$,

$$P\{|X(\omega) - EX| \geqq a\,\sigma(X)\} \leqq 1/a^2\,,$$

that is

$$P\{EX - a\,\sigma(X) < X(\omega) < EX + a\,\sigma(X)\} \geqq 1 - 1/a^2\,.$$

Proof When $\sigma(X) = 0$, $X(\omega) = EX$ a.s. from (i) in the previous theorem. Accordingly, the above inequality obviously holds. Let $\sigma(X) > 0$ and

$$A = \{\omega \big| |X(\omega) - EX| \geqq a\,\sigma(X)\}\,.$$

Then

$$\sigma(X)^2 = E(X - EX)^2 \geqq E\big((X - EX)^2, A\big) \geqq a^2\,\sigma(X)^2 P(A)\,.$$

Hence $P(A) \leqq 1/a^2$. ▮

Exercise 1.2 Other than (ii), X and Y are real probability variables.

(i) Prove that $\big|EX\big| \leqq E|X|\,, \qquad |EX| \leqq \sqrt{EX^2}$.

[Hint] The former equation is $-|X| \leqq X \leqq |X|$. The latter equation is $EX^2 - (EX)^2 = V(X)$.

(ii) Use the second equation of (i) and for a probability vector $X(\omega)$, show that the following inequality holds.

$$||EX|| \leqq E||X||\,.$$

(iii) Use $E1_A = P(A)$ and Theorem 1.4, give a different proof of Exercise 1.1 (inclusion-exclusion formula)'s (i). Then, use $P(A^c) = 1 - P(A)$, derive (ii) of Theorem 1.4. In this instance, apply

$$\sum_{k=0}^{n} (-1)^k \begin{pmatrix} n \\ k \end{pmatrix} = (1-1)^n = 0\,.$$

(iv) When $\sigma(X) > 0, a \neq 0$, prove that $R\big(X, aX + b\big) = \operatorname{sgn} a$. Here, $\operatorname{sgn} a$ is a function that takes value ± 1 according to $\gtreqless a$.

(v) Use Čebysěv's inequality and prove the following with $\sigma(X), \sigma(Y) > 0$:

$$P\left\{\left|\frac{Y(\omega) - EY}{\sigma(Y)} - \frac{X(\omega) - EX}{\sigma(X)}\right| > a\sqrt{2\big(1 - R(X,Y)\big)}\right\} \leqq \frac{1}{a^2} \qquad (a > 0)$$

(vi) Suppose X_1, X_2 are the points obtained from rolling a dice twice. Express

$$R\big(aX_1 + a_2X_2\,,\, b_1X_1 + b_2X_2\big)$$

in terms of a_1, a_2, b_1, b_2.

(vii) Find the value of t such that $f(t) = E(X-t)^2$ is the minimum; also find the minimum value of $f(t)$.

(viii) Find the values of t, s such that $g(t,s) = E\big(Y - tX - s\big)^2$ is minimum; also find the minimum value of $g(t,s)$. Here, $\sigma(X) > 0$ is assumed.

Christopher Ting
Translation Version 1.0
Date: 10 May, 2017