§ 1.3 Mixture, Direct Union, Tree-Shaped Union

a) Probability Variables that Take Generic Values

In the previous section, real probability variables and probability vectors are considered. By the same thinking, a probability variable that takes general values can also be considered. For example, depending on whether the rolled dice turns up an even or odd number, a probability variable that takes values in "even" and "odd" can be considered. As stated previously, the probability space of the trial of rolling dice is given by

$$\Omega = \{1,2,3,4,5,6\}\,, \qquad P\{i\} = \frac{1}{6}\,.$$

The probability variable X considered is a function on Ω and

$$X(1) = X(3) = X(5) = \text{odd}\,, \qquad X(2) = X(4) = X(6) = \text{even}\,.$$

The sample space Ω^X in this case is

$$\Omega^X = \{\text{odd}, \text{even}\}\,,$$

and the probability law P^X is given by

$$\begin{aligned} P^X\{\text{odd}\} &= P\big\{\omega | X(\omega) = \text{odd}\big\} = P\{1,3,5\} = 1/2 \\ P^X\{\text{even}\} &= P\big\{\omega | X(\omega) = \text{even}\big\} = P\{2,4,6\} = 1/2\,. \end{aligned}$$

Suppose a card is drawn from a box containing cards labeled $1, 2, \cdots 10$. The probability space (Ω, P) is given by

$$\Omega = \{1,2,\cdots,10\}\,, \qquad P\{i\} = 1/10\,.$$

Now, the cards labelled 1,2 are red, 3,4,5 are blue, and the 5 remaining cards are white. Let Y be the color of the trial of card drawing, and the probability of Y be given by

$$\begin{aligned} &Y(1) = Y(2) = \text{red}\,, \qquad Y(3) = (4) = Y(5) = \text{green}\,, \\ &Y(6) = Y(7) = \cdots = Y(10) = \text{white}\,. \end{aligned}$$

Its sample space Ω^Y and probability law P^Y are, respectively, given by

$$\begin{aligned} &\Omega^Y = \big\{\text{red}, \text{green}, \text{white}\big\}\,, \\ &P^Y\{\text{red}\} = \frac{2}{10} = \frac{1}{5}\,, \quad P^Y\{\text{blue}\} = \frac{3}{10}\,, \quad P^Y\{\text{white}\} = \frac{5}{10} = \frac{1}{2}\,. \end{aligned}$$

For a probability variable that takes generic value, the sample space, probability law, and the probability space are defined in the same way as in the previous section. However, it is not possible to define the average value.

b) Mixture of Trials

Denote a trial by $\boldsymbol{T}$, probability space by (Ω, P), and probability variable on (Ω, P) by $X(\omega)$. If one looks at and observe only the value of $X(\omega)$ as a result of $\boldsymbol{T}$, then one obtains a new trial. Denoted by $\boldsymbol{T_X}$, it is called the **mixture** resulting from $X(\omega)$. If one looks at the odd-or-even number only turning up in rolling a dice, or at the colour only of the card drawn from a box, and so on, then one has many examples of mixture as one desires. The sample space of $\boldsymbol{T}_X$ is of course the same as the sample space Ω^X of X. Since the fact that the point of the result $B\big(\subset \Omega^X\big)$ of $\boldsymbol{T}_X$ appears is the same as, returning to the original $\boldsymbol{T}$, the point of $X^{-1}(B)$ appears, its probability is $P\big(X^{-1}(B)\big)$ or $P^X(B)$. Consequently the probability law of $\boldsymbol{T_X}$ coincides with the probability law P^X of X. To think of $\boldsymbol{T}^X$ is about treating all equally the points $X^{-1}(x)\big(x \in \Omega^X\big)$ in Ω and calling it by the name of 'x', and that is the root of the terminology of mixture.

The function $\varphi(x)$ on $\big(\Omega^X, P^X\big)$ is a probability variable with respect to $\boldsymbol{T}^X$, Its probability law—denoted by $(P^X)^\varphi$—is given by

$$(P^X)^\varphi(\Lambda) = P^X\big(\varphi(x) \in \Lambda\big) = P^X\big(\varphi^{-1}(\Lambda)\big)\,, \qquad \Lambda \subset \Omega^X\,.$$

When $\varphi(x)$ is real-valued (or vector-valued), its average value—denoted by $E^X\varphi$—is given by

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

For the same function with respect to $\boldsymbol{T}$, it is given by $Y(\omega) = \varphi\big(X(\omega)\big)$, which is a function on Ω. This is because when the result ω of $\boldsymbol{T}$ appears, the result $\varphi\big(X(\omega)\big)$ of $\boldsymbol{T}^X$ appears as well. In this respect, its probability law is of course given by

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

When φ is real-valued, its average value is of course given by EY. Though respectively, $(P^X)^\varphi$ and P^Y, $E^X\varphi$ and EY should coincide, they are in fact guaranteed by the following theorem.

Theorem 1.9 Let $Y(\omega) = \varphi\big(X(\omega)\big)$,

(i) $P\{Y \in \Lambda\} - P^X\{\varphi \in \Lambda\}\,, \qquad \Lambda \subset \Omega^X\,,$

(ii) When φ is real-valued (or vector-valued),

$$EY = E^X\varphi\,.$$

Proof (i) $P\{Y \in \Lambda\} = P\big(Y^{-1}(\Lambda)\big) = P\big((\varphi \circ X)^{-1}(\Lambda)\big) = P\big(X^{-1}(\varphi^{-1}(\Lambda))\big) = P^X\big(\varphi^{-1}(\Lambda)\big) = P^X\{\varphi \in \Lambda\}$.

(ii) In the same manner as Theorem 1.3's (v),

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

is proven and the right hand side is $E^X\varphi$. ❚

Since variance, covariance, standard deviation and so on are defined by the average value, all of these have the same property as (ii) in the above theorem. In other words, if

$$Y(\omega) = \varphi\big(X(\omega)\big) \,, \qquad Z(\omega) = \psi\big(X(\omega)\big)$$

then

$$V(Y) = V^X(\varphi) \,, \qquad V(Y, Z) = V^X(\varphi, \psi)$$
$$\sigma(Y) = \sigma^X(\varphi) \,, \qquad R(Y, Z) = R^X(\varphi, \psi)$$

c) Direct Union of Trials

Let $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ be trials, and their respective probability spaces are denoted by (Ω_1, P_1) and (Ω_2, P_2). Let $\widetilde{\boldsymbol{T}}$ be the big trial obtained from performing the two trials successively. $\widetilde{\boldsymbol{T}}$ is called the **direct union** of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, and it is denoted by $\boldsymbol{T}_1 \times \boldsymbol{T}_2$. Since the result ω of $\widetilde{\boldsymbol{T}}$ is expressed as the combination of the result ω_1 of $\boldsymbol{T}_1$ and the result ω_2 of $\boldsymbol{T}_2$, the sample space $\widetilde{\Omega}$ of $\widetilde{\boldsymbol{T}}$ is

$$\widetilde{\boldsymbol{T}} = \big\{(\omega_1, \omega_2) \big| \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\big\} \,.$$

Next, consider the probability law $\widetilde{P}$ of $\widetilde{\boldsymbol{T}}$. It is to look at and observe $\boldsymbol{T}_i$ $(i = 1, 2)$ as the i-th component of the result $\widetilde{\omega} = (\omega_1, \omega_2)$ of $\widetilde{\boldsymbol{T}}$:

$$\omega_i = \pi_i\big(\widetilde{\omega}\big) \,. \qquad\qquad (\pi_i \text{ is the } i\text{-th image})$$

In other words, $\boldsymbol{T}_i$ is the mixture of $\widetilde{\boldsymbol{T}}$'s $\pi_i\big(\widetilde{\omega}\big)$. Accordingly, of course

$$\widetilde{P}\pi_i^{-1} = P_i \,, \qquad\qquad i = 1, 2$$

should be satisfied. There are many of such $\widetilde{P}$. For example, when tossing the coin twice

$$\begin{aligned}
&\Omega_i = \{0,1\} \qquad \text{(0 is tail, 1 is head)}\,, \\
&P_i\{j\} = 1/2\,, \qquad j = 1,2\,, \\
&\widetilde{\Omega} = \big\{(0,0),(0,1),(1,0),(1,1)\big\}
\end{aligned}$$

and the probability measure $\widetilde{P}$ on $\widetilde{\Omega}$ that satisfies the above conditions is given by

$$\begin{aligned}
\widetilde{P}\big\{(0,0)\big\} &= \alpha\,, \qquad \widetilde{P}\big\{(0,1)\big\} = 1/2-\alpha\,, \\
\widetilde{P}\big\{(1,0)\big\} &= 1/2-\alpha\,, \qquad \widetilde{P}\big\{(1,1)\big\} = \alpha\,,
\end{aligned}$$

$$\alpha \text{ is any number in } \big[0,1/2\big].$$

and there are infinitely many of these.

Now returning to the generic case, to determine $\widetilde{P}$, one must choose amongst $\widetilde{P}$ one that satisfies $\widetilde{P}\pi_i^{-1} = P_i$. The method is

Multiplicative Law $\qquad \widetilde{P}\big\{(\omega_1,\omega_2)\big\} = P_1\{\omega_1\}P_2\{\omega_2\}$

Hence, for any $\widetilde{A} \subset \widetilde{\Omega}$, it is defined by

$$\widetilde{P}\big(\widetilde{A}\big) = \sum_{\widetilde{\omega}\in\widetilde{\Lambda}} \widetilde{P}\{\omega\} = \sum_{(\omega_1,\omega_2)\in\Lambda} P_1\{\omega_1\}P_2\{\omega_2\}\,.$$

Hence,

$$\widetilde{P}\big(A_1 \times A_2\big) = P_1(A_1)P_2(A_2)\,,$$

and in particular

$$\begin{aligned}
&\widetilde{P}\big(\widetilde{\Omega}\big) = \widetilde{P}\big(\Omega_1 \times \Omega_2\big) = P_1\big(\Omega_1\big)\,P_2\big(\Omega_2\big) = 1\,, \\
&\widetilde{P}\big(\pi_1^{-1}(A_1)\big) = \widetilde{P}\big(A_1 \times \Omega_2\big) = P_1(A_1)\,P_2(\Omega_2) = P_1(A_1)\,, \\
&\widetilde{P}\big(\pi_2^{-1}(A_2)\big) = P_2(A_2)\,,
\end{aligned}$$

$\widetilde{P}$ is the probability measure on Ω that satisfies the $\widetilde{P}\pi_i^{-1} = P_i(i = 1,2)$. In the above example of coin tossing, its $\widetilde{P}$ corresponds to $\alpha = 1/4$.

Here a question of why the multiplicative law is used may arise. It should be recognized as an **axiom** concerning direct union, and cannot be proven. In this sense, it is the same as the probability P in § 1.1 that requires (P.1), (P.2), and (P.3).

In general, $\widetilde{P}$ as defined by the multiplicative law is called the **direct product** of P_1 and P_2, and it is denoted as $P_1 \times P_2$. In this notation, the probability space $(\widetilde{\Omega},\widetilde{P})$ of the direct union $\widetilde{\boldsymbol{T}} = \boldsymbol{T}_1 \times \boldsymbol{T}_2$ is given by

$$\widetilde{\Omega} = \Omega_1 \times \Omega_2 \,, \qquad \widetilde{P} = P_1 \times P_2 \,.$$

Moreover, $(\widetilde{\Omega}, \widetilde{P})$ is the **direct product** of (Ω_1, P_1) and (Ω_2, P_2), and it is denoted by $(\Omega_1, P_1) \times (\Omega_2, P_2)$.

The multiplicative law is an axiom that picks up $P_1 \times P_2$ from amongst $\widetilde{P}$, which satisfies $\widetilde{P}\pi_i^{-1} = P_i (i = 1, 2)$. This axiom can be derived from the **principle of maximum entropy** (This is also an axiom). It is not because of this that the axiom no longer needed, compared to the multiplicative law, the principle of maximum entropy is just more natural to some people. In general, let P be the probability measure on

$$\Omega = \{a_1, a_2, \cdots, a_m\} \,.$$

Then the entropy of P is

$$\epsilon(P) = \sum_{i=1}^{m} P\{a_i\} \log \frac{1}{P\{a_i\}} \,,$$

and it is restricted that

$$0 \log \frac{1}{0} = \lim_{x \longrightarrow 0} x \log \frac{1}{x} = 0 \,.$$

Now P_1 and P_2 are respectively the probability measures on

$$\Omega_1 = \{a_1, a_2, \cdots, a_\ell\} \,, \qquad \Omega_2 = \{b_1, b_2, \cdots, b_m\} \,.$$

The probability measure $\widetilde{P}$ on $\Omega_1 \times \Omega_2$ satisfies

$$\widetilde{P}\pi_1^{-1} = P_1 \,, \qquad \widetilde{P}\pi_2^{-1} = P_2 \,.$$

Let

$$p_i = P_1\{a_i\} \,, \qquad q_i = P_2\{b_i\} \,, \qquad r_{ij} = \widetilde{P}\{(a_i, b_j)\} \,.$$

From the above conditions,

$$\begin{aligned} \textstyle\sum_{j=1}^{m} r_{ij} &= p_i \,, \qquad i = 1, 2, \cdots, \ell \,, \\ \textstyle\sum_{i=1}^{\ell} r_{ij} &= q_j \,, \qquad j = 1, 2, \cdots, m \,. \end{aligned}$$

The entropy of $\widetilde{P}$ is

$$\epsilon(\widetilde{P}) = \sum_{i,j} r_{ij} \log \frac{1}{r_{ij}} \,.$$

Accordingly, to derive the multiplicative law from the principle of maximum entropy, it is to maximize

$\epsilon(\widetilde{P}))$ with respect to the attached conditions (p_i, q_j are assumed given) of r_{ij}

$$r_{ij} = p_i\, q_j\,, \qquad i = 1, 2, \cdots, \ell;\ j = 1, 2, \cdots, m\,,$$

and it can be proven using the method of Lagrange multiplier. The sketch is discussed as follows.

Let

$$F(r_{11}, r_{12}, \cdots, r_{\ell m}) = \sum_{i,j} r_{ij} \log \frac{1}{r_{ij}} - \sum \alpha_j \left(\sum_j r_{ij} - p_i \right) - \sum \beta_i \left(\sum_i r_{ij} - q_j \right)$$

$$\frac{\partial F}{\partial r_{ij}} = \log \frac{1}{r_{ij}} - 1 - \alpha_i - \beta_j = 0\,.$$

That is,

$$r_{ij} = e^{-1-\alpha_i-\beta_j}\,.$$

This and the attached conditions $p_i = \sum_j r_{ij}$, $q_j = \sum_i r_{ij}$ obtain

$$p_i = e^{-1} e^{-\alpha_i} \sum_j e^{-\beta_j}\,, \qquad q_j = e^{-1} e^{-\beta_j} \sum_i e^{-\beta_i}\,.$$

Hence

$$\frac{r_{ij}}{p_i\, q_j} = \frac{1}{e^{-1} \sum_i e^{-\alpha_i} \sum_j e^{-\beta_j}}\,.$$

The right hand side is unrelated to i, j, and denote it by γ. Then

$$1 = \sum_{i,j} r_{ij} = \gamma \sum_i p_i \sum_j q_j = \gamma\,.$$

Consequently $r_{ij} = p_i\, q_j$ and the expected result is obtained.

It is possible to carry the same observation over to the direct unions of n trials; the probability space of direct union is the direct product of component trials. Clearly

Theorem 1.10 If $\widetilde{P} = P_1 \times P_2 \times \cdots \times P_n$, i.e.,

$$\widetilde{P}\{(\omega_1, \omega_2, \cdots \omega_n)\} = P_1\{\omega_1\} P_2\{\omega_2\} \cdots P_n\{\omega_n\}$$

Then

$$\widetilde{P}(A_1 \times A_2 \times \cdots \times A_n) = P_1(A_1) P_2(A_2) \cdots P_n(A_n)\,.$$

Tree-Like Union of Trials

The combination of two operations of union and mixture can lead to a new trial. This is illustrated with examples.

Example 1.1 Two persons A, B and toss a coin alternatively (beginning with A) 10 times. The person who has a toss of head first wins the game. If none of the 10 times has a head, then it is a draw. The proceeding of this game is considered a trial $\boldsymbol{T}$. A head is represented by 1 and a tail is represented by 0. Its sample space Ω is made up of the following 11 points:

$$\begin{aligned}
a_1 &= 1 && \text{(A wins)}\,, \\
a_2 &= 0\,1 && \text{(B wins)}\,, \\
a_3 &= 0\,0\,1 && \text{(A wins)}\,, \\
&\cdots\cdots && \\
a_{10} &= 0\,0\,0\,0\,0\,0\,0\,0\,0\,1 && \text{(B wins)}\,, \\
a_{11} &= 0\,0\,0\,0\,0\,0\,0\,0\,0\,0 && \text{(draw)}\,.
\end{aligned}$$

The probability law P of $\boldsymbol{T}$ is given by

$$P\{a_k\} = 2^{-k} \quad (k = 1, 2, \cdots, 10)\,, \qquad P\{a_{11}\} = 2^{-10}\,,$$

for which the readers will find it evident. To explain the reason, a combination of the two procedures of direct union and mixture is necessary.

When the result of $\boldsymbol{T}$ is a_1, even if it is a win for A and the first toss is exhaustive, assume the toss goes on up to 10 times. The same applies to the cases of $a_2, a_3, \cdots a_9$. Let $\widetilde{\boldsymbol{T}}$ be such resulting trial of 'toss up to 10 times.' Because $\widetilde{\boldsymbol{T}}$ is the direct union of first toss, second toss, $\cdots$, and the tenth toss, its probability space $\big(\widetilde{\Omega}, \widetilde{P}\big)$ is

$$\begin{aligned}
\widetilde{\Omega} &= \{0,1\}^1 0 = \big\{(i_i, i_2, \cdots, i_1 0)\big| i_1, i_2, \cdots, i_{10} = 0, 1\big\}\,, \\
\widetilde{P} &= \big\{(i_i, i_2, \cdots, i_{10})\big\} = 2^{-10}\,.
\end{aligned}$$

Next, define a mapping $X\big(\widetilde{\omega}\big)$ from $\widetilde{\Omega}$ onto Ω

$$X\big(i_1, i_2, \cdots, i_1 0\big) = \begin{cases}
a_1 & \text{(when } i_1 = 1)\,, \\
a_2 & \text{(when } i_1 = 0, i_2 = 1)\,, \\
\vdots & \\
a_{10} & \text{(when } i_1 = i_2 = \cdots = i_9 = 0, i_{10} = 1)\,, \\
a_{11} & \text{(when } i_1 = i_2 = \cdots = i_{10} = 0)\,.
\end{cases}$$

$X(\widetilde{\omega})$ is the Ω-valued probability variable on $\big(\widetilde{\Omega}, \widetilde{P}\big)$. Clearly, the trial $\boldsymbol{T}$ considered earlier is a mixture of

$\widetilde{\boldsymbol{T}}$ through $X(\widetilde{\omega})$. Accordingly, the probability law P of $\boldsymbol{T}$ equals the probability law P_X of X. With respect to $k = 1, 2, \cdots, 10$,

$$\begin{aligned} P\{a_k\} &= \widetilde{P}_X\{a_k\} = \widetilde{P}\{X(\widetilde{\omega}) = a_k\} \\ &= \widetilde{P}\{(i_1, i_2, \cdots i_{10}) | i_1 = i_2 = \cdots = i_{k-1} = 0, i_k = 1\} \\ &= 10^{10-k} \frac{1}{2^{10}} = 2^{-k}. \end{aligned}$$

Similarly,

$$P\{a_{10}\} = 2^{-10}.$$

The probability of A winning is

$$\sum_{k=1}^{5} P\{a_{2k-1}\} = \sum_{k=1}^{5} 2^{-(2k-1)} = \frac{2}{3}(1 - 2^{-10}),$$

and in the same fashion, the probability of B winning is

$$\sum_{k=1}^{5} P\{a_{2k}\} = \sum_{k=1}^{5} 2^{-2k} = \frac{1}{3}(1 - 2^{-10}).$$

The probability of a draw is of course 2^{-10}.

Or if one is to look only at the winning and losing and consider the trial $\boldsymbol{T}'$, which is a mixtures of $\boldsymbol{T}$, then its probability space (Ω', P') is given by

$$\begin{aligned} &\Omega' = \{\text{A wins, B wins, draw}\}, \\ &P'\{\text{A wins}\} = \frac{2}{3}(1 - 2^{-10}), \qquad P'\{\text{B wins}\} = \frac{1}{3}(1 - 2^{-10}), \qquad P'\{\text{draw}\} = 2^{-10}. \end{aligned}$$

Example 1.2 Let there be n cards with labels $1, 2, \cdots n$ each are in the box B_0. First a card is drawn, and consider the trial $\boldsymbol{T}$ of drawing one more card from the remaining $n-1$ cards. The probability space (Ω, P) of $\boldsymbol{T}$ is given by

$$\begin{aligned} \Omega &= \{(i,j) | i, j = 1, 2, \cdots, n (i \neq j)\} \\ P\{(i,j)\} &= \frac{1}{n(n-1)}, \end{aligned}$$

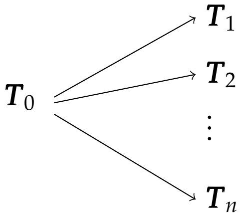
which is evident to the readers, though its reason can be explained by the combination of the two procedures of union and mixture. To begin with, denote the first draw $\boldsymbol{T}_0$ from the given box B_0. In the second draw, when the result i of $\boldsymbol{T}$ appears, draw from the box with $n-1$ cards

$$1, 2, \cdots, i-1, i+1, \cdots, n\,.$$

Denote such box by B_i, and the subsequent draw is denoted by $\boldsymbol{T}_i$. The combination of these two draws $\boldsymbol{T}$ is related to $n+1$ trials $\boldsymbol{T}_0, \boldsymbol{T}_1, \cdots \boldsymbol{T}_n$, and to obtain $\boldsymbol{T}$ from these trials, the following tree diagram is needed.

$$\boldsymbol{T}_0 \begin{cases} \text{card 1 is drawn} \longrightarrow \boldsymbol{T}_1 \\ \text{card 2 is drawn} \longrightarrow \boldsymbol{T}_2 \\ \text{..............................} \\ \text{card } n \text{ is drawn} \longrightarrow \boldsymbol{T}_n \end{cases}$$

or simply written as



In this sense, it is called the **tree union**.

Let $\widetilde{\boldsymbol{T}}$ be the direct union of $\boldsymbol{T}_0, \boldsymbol{T}_1, \cdots \boldsymbol{T}_n$. It is different from $\boldsymbol{T}$. Regardless of the result of $\boldsymbol{T}_0$, $\boldsymbol{T}_1, \boldsymbol{T}_2, \cdots \boldsymbol{T}_n$ are performed. In other words, from each of the $n+1$ boxes B_0, B_1, $\cdots$, B_n, a card is drawn and the vector $(i, j_1, j_2, \cdots, j_n)$ obtained from these drawn cards is considered to be the the result of $\widetilde{\boldsymbol{T}}$. Now, to relate the first trial $\boldsymbol{T}$ of interest to $\widetilde{\boldsymbol{T}}$, one can consider the following point. "Regardless of the result of B_0, from $B_1, B_2, \cdots, B_n$, respectively, $j_1, j_2, \cdots, j_n$ are drawn, and disregard the cards other than j_i, and let (i, j_i) be the result of $\boldsymbol{T}$." In this way, $\boldsymbol{T}$ is the mixture of $\widetilde{\boldsymbol{T}}$.

The probability space (Ω_0, P_0) of $\boldsymbol{T}_0$ is given by

$$\Omega_0 = \{1, 2, \cdots, n\}, \qquad P_0\{j\} = 1/n\,.$$

The probability space (Ω_i, P_i) of $\boldsymbol{T}_i (i = 1, 2, \cdots, n)$ is given by

$$\Omega_i = \{1, 2, \cdots, i-1, i+1, \cdots, n\}, \qquad P_i\{j\} = 1/(n-1)\,.$$

The trial $\boldsymbol{T}$ of drawing one card each is the direct union $\boldsymbol{T}_0 \times \boldsymbol{T}_1 \times \cdots \times \boldsymbol{T}_n$ of $\boldsymbol{T}_0$, $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, $\cdots$, $\boldsymbol{T}_n$. Its probability space $\big(\widetilde{\Omega}, \widetilde{P}\big)$ is given by

$$\widetilde{\Omega} = \Omega_0 \times \Omega_1 \times \cdots \Omega_n\,, \qquad \widetilde{P} = P_0 \times P_1 \times \cdots P_n\,.$$

Let $\pi_i : \widetilde{\Omega} \longrightarrow \Omega_i$ be the i-th image. Then $\pi_i(\widetilde{\omega})$ represents the result of $\boldsymbol{T}_i$ as a function on $\big(\widetilde{\Omega}, \widetilde{P}\big)$. Accordingly, the result of $\boldsymbol{T}$ is a function on $\big(\widetilde{\Omega}, \widetilde{P}\big)$ and is expressed as

$$X(\widetilde{\omega}) = \big(\pi_0(\widetilde{\omega}), \pi_{\pi_0(\widetilde{\omega})}\big)\,.$$

And $\boldsymbol{T}$ coincides with $\widetilde{\boldsymbol{T}}_X$, which is $\widetilde{\boldsymbol{T}}$ mixed by $X(\widetilde{\omega})$.

With the above consideration as background, determine the probability space (Ω, P) of $\boldsymbol{T}$. It is clear that

$$\Omega = \left\{ (i,j) \middle| i,j = 1,2,\cdots,n \; (i \neq j) \right\}.$$

With respect to $(i,j) \in \Omega$,

$$\begin{aligned}
P\{(i,j)\} &= \widetilde{P}^X\{(i,j)\} = \widetilde{P}\{X(\widetilde{\omega}) = (i,j)\} \\
&= \widetilde{P}\{\pi_0(\widetilde{\omega}) = i\,,\ \pi_{\pi_0(\widetilde{\omega})} = j\} \\
&= \widetilde{P}\{\pi_0(\widetilde{\omega}) = i\,,\ \pi_i(\widetilde{\omega}) = j\} \\
&= \widetilde{P}\{\pi_0(\widetilde{\omega}) = i\,,\ \pi_i(\widetilde{\omega}) = j\,,\ \pi_k(\widetilde{\omega}) \in \Omega_k \; (k \neq i,j)\} \\
&= P_0\{i\}\, P_i\{j\} \prod_{k \neq i,j} P(\Omega_k) = P_0\{i\}\, P_i\{j\} \\
&= \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{n(n-1)}\,.
\end{aligned}$$

Thus (Ω, P) is as expected from the previous discussion.

For the general trial, it is possible to define **tree union** with the same consideration. Let the probability space of $\boldsymbol{T}_0$ be

$$\left(\Omega_0 \equiv \{a_1, a_2, \cdots, a_n\}, P_0\right).$$

With respect to each $a \in \Omega_0$, define the trial $\boldsymbol{T}_a$, and the probability space is (Ω_a, P_a).

$$\boldsymbol{T}_0 \begin{cases} \text{when } a_1 \text{ appears} \longrightarrow \boldsymbol{T}_{a_1} \\ \text{when } a_2 \text{ appears} \longrightarrow \boldsymbol{T}_{a_2} \\ \text{............................} \\ \text{when } a_n \text{ appears} \longrightarrow \boldsymbol{T}_{a_n} \end{cases} \qquad \text{or simply}$$

$$\begin{tikzcd}
 & \boldsymbol{T}_{a_1} \\
\boldsymbol{T}_0 \arrow[ru] \arrow[r] \arrow[rdd] & \boldsymbol{T}_{a_2} \\
 & \vdots \\
 & \boldsymbol{T}_{a_n}
\end{tikzcd}$$

Let the tree union be $\boldsymbol{T}$. As explained in the previous example, it is obtainable from the mixture of the direct union $\widetilde{\boldsymbol{T}} = \boldsymbol{T}_0 \times \boldsymbol{T}_1 \times \cdots \times \boldsymbol{T}_n$. The sample space of $\boldsymbol{T}$ is of course given by

$$\Omega = \left\{ (a,b) \middle| a \in \Omega_0, b \in \Omega_a \right\},$$

while its probability law is determined by the following theorem:

Theorem 1.11 The multiplicative law of tree union

$$P\{(a,b)\} = P_0\{a\}\, P_a\{b\}\,.$$

Proof The probability space $(\widetilde{\Omega}, \widetilde{P})$ of $\widetilde{\boldsymbol{T}} = \boldsymbol{T}_0 \times \boldsymbol{T}_{a_1} \times \boldsymbol{T}_{a_2} \times \cdots \times \boldsymbol{T}_{a_n}$ is

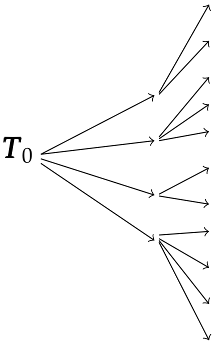
$$\widetilde{\Omega} = \Omega_0 \times \Omega_{a_1} \times \Omega_{a_2} \times \cdots \times \Omega_{a_n}\,, \qquad \widetilde{P} = P_0 \times P_{a_1} \times P_{a_2} \times \cdots \times P_{a_n}\,.$$

Since $\boldsymbol{T}$ is obtained from $\widetilde{\boldsymbol{T}}$ by

$$X(\widetilde{\omega}) = \bigl(\pi_0(\widetilde{\omega}), \pi_{\pi_0(\widetilde{\omega})}\bigr)\,,$$

same as 'drawing from the box', the equality of the theorem is obtained. ▌

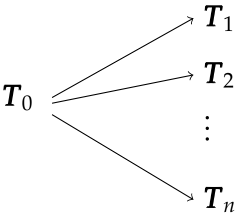
The more complicated tree union can also be defined (refer to the following diagram).



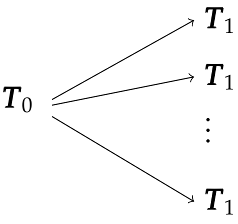
In this instance, the multiplicative law takes the following form:

$$P\bigl\{(a,b,c)\bigr\} = P_0\bigl\{(a,b,c)\bigr\} = P_0\{a\}\,P_a\{b\}\,P_b\{c\}\,.$$

If $\boldsymbol{T}_1 = \boldsymbol{T}_2 = \cdots = \boldsymbol{T}_n$, then the tree union



coincides with the direct union $\boldsymbol{T}_0 \times \boldsymbol{T}_1$. Conversely, if the direct union $\boldsymbol{T}_0 \times \boldsymbol{T}_1$ is a tree union:



Accordingly, it is sufficient to consider only the tree union, as it encompasses direct union as a special case.

Exercise 1.3 (i) Let the number of the card first drawn in Example 1.2 be X, and the second drawn card be Y. Then, find the following

$$P^X\,,\ P^Y\,,\ EX\,,\ EY\,,\ V(X)\,,\ V(Y)\,,\ \sigma(X)\,,\ \sigma(Y)\,,\ P^{(X,Y)}\,,\ V(X,Y)\,,\ R(X,Y)\,.$$

(ii) A total of 18 cards in the box marked two each as $1, 2, 3, 4, 5, 6, 7, 8, 9$. Four cards are drawn and placed from left to right according to the order of appearance to make a 4-digit number. Find the probability that the number is larger than 5283.

[Hint] The probability of drawing out i from the box in question is $P_0\{i\}$. The probability of drawing out j from the rest is $P_i\{j\}$. Furthermore, the probability of drawing out k from the remaining is $P_{ij}\{k\}$, and the probability of drawing l from the rest is $P_{ijk}\{l\}$. Note that the probability of interest is

$$P_0\{6,7,8,9\} + P_0\{5\}\,P_5\{3,4,\cdots,9\} + P_0\{5\}\,P_5\{2\}\,P_{52}\{9\} + P_0\{5\}\,P_5\{2\}\,P_{52}\{8\}\,P_{528}\{4,5,\cdots,9\}\,.$$

Also,

$$P_i\{A\} = \sum_{j\in A} P_i\{j\}\,; \qquad \text{the same applies to } P_{ij}, P_{ijk}\,.$$

(iii) Show that amongst the probability measures on $\Omega = \{a_0, a_1, \cdots, a_n\}$, the one with the largest entropy is $P\{a_i\} \equiv 1/n$.

Christopher Ting
Translation Version 1.1
Date: 23 May, 2017