Academic Achievements of Professor P.L. Hsu

CHEN Jiading and ZHENG Zhongguo

This year we celebrate the 100'th birthday of Professor Pao-Lu Hsu, a world-class statistician and the founder of probability and statistics in China. He was a member of the Chinese Academy of Science, and a Professor of Rank One at Peking University. With reverence we recall his contributions to probability and statistics in China.

Hsu was born in Beijing on September 1, 1910. But his forefathers were natives of Hangzhou, Zhejiang Province. He was from a prominent intellectual family. In his childhood, he received solid training in both traditional Chinese and modern western cultures [?]. He graduated from Tsinghua University in 1933, majoring in mathematics. After his graduation, he worked at Peking University as a teacher. In the meantime, he published a joint paper with Tsai-han Kiang (Jiang Zehan) on the numbers of nondegenerate critical points, which showed his solid mathematical foundation and research capability. In 1936, he went to University College, London and spend 4 years for studying mathematical statistics. During this period, by his strong mathematical skill combining with advanced statistical ideas, he wrote a series remarkable papers. He earned his Ph.D. in 1938 and Sc.D. in 1940. From London, he returned to China accepting a professorship in the Department of Mathematics, Peking University. In 1945, he went to the United States, visiting the University of California at Berkeley, Columbia University and the University of North Carolina at Chapel Hill. In 1947, he returned to Beijing and thereafter he was engaged in teaching mathematics at Peking University for more than 20 years. In December 18, 1970, he died in his home in the Peking University campus.

He set the highest standard for his research work. "The value of a paper is not determined by its publication. Instead, it is validated when it is cited frequently by others later on". "I don't like to get famous because my papers are published on the journals with good reputation, I prefer that a journal builds its reputation because of my papers."[?] Owing to his high standard, his published papers have profound impact in the statistics area.

Professor Hsu's main research areas were mathematical statistics and probability theory. Besides, his works on matrix theory and integral transformation were excellent. He was the first Chinese who was internationally recognized in the area of probability and statistics. In 1979, in connection with his 70th birthday, the editor of *the Annals of Statistics* invited three specialists to jointly write a paper on Hsu's life and work [?].

They also wrote papers separately introducing Hsu's works in details. More specifically, American statistician E. L. Lehmann (who was a member of the National Academy of Science) wrote a paper introducing Hsu's work on statistical inference [?]; Professor T. W. Anderson wrote a paper about Hsu's work in multivariate area [?]. The famed probabilist, Professor K. L. Chung, wrote a paper about Hsu's contributions in probability theory [?]. In 1981, the book entitled "Hsu Pao-Lu Collected Papers" (in Chinese, [?]) was published by the Science Press of China, with a preface by Professor Tsai-han Kiang and Professor Hsio-Fu Tuan (both were members of the Chinese Academy of Science), who highly praised Professor Hsu for his contributions to the development of the theory of probability and statistics in China. The book of "Pao-Lu Hsu Collected Papers"[?], edited by K.L. Chung and published by Springer-Verlag in 1983, includes almost all Hsu's papers. Some of Hsu's papers were written in Chinese. All the papers in Chinese were translated into English in the book. The book entitled "Leading Personalities in Statistical Science from the Seventeenth Century to the Present"[?], edited by N. J. Johnson and S. Koty and published in the United States in 1997, includes 114 people who had great influence in the development of statistical science since early 17th century. Among them I. Newton, C. F. Gauss, P. S. Laplace, R. A. Fisher, J. Neyman, A. N. Kolmogorov are on the list. Professor Hsu is the only Chinese statistician included in the book.

The important accomplishments of Hsu's works can be divided into the following ten areas:

(1) The first paper of Hsu in the statistics area [?], published in 1938, was concerned with Behrens-Fisher problem. Let $X_1, X_2, ..., X_n$, $Y_1, Y_2, ..., Y_m$ be samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively, where σ_1 and σ_2 are unknown. The problem is to test the null hypothesis $H_0 : \mu_1 = \mu_2$. (When $\sigma_1 = \sigma_2$, it is solved by t-test. The difficult point is that σ_1 and σ_2 are not necessary equal). Hsu considered the class of statistics

$$U = (\bar{X} - \bar{Y})^2/(A_1 S_1^2 + A_2 S_2^2),$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i,$$

$$S_1^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2, \quad S_2^2 = \sum_{i=1}^{m}(Y_i - \bar{Y})^2,$$

and A_1 and A_2 are two constants. When $A_1 = A_2 = (m+n)/[(n+m-2)nm]$, U is just the Student's t-statistic u_1, and when $A_1 = 1/(n(n-1))$ and $A_2 = 1/(m(m-1))$, U is the Behrens-Fisher statistic u_2. Hsu found the series expansion for the density of U and utilized this formula to study the power function of the rejection area of $\{U > C\}$. It shows that the power function relies only on the parameters $\theta = \sigma_1^2/\sigma_2^2$

and $\lambda = (\mu_1 - \mu_2)^2/(\frac{1}{n}\sigma_1^2 + \frac{1}{m}\sigma_2^2)$. It is an exact result (not an asymptotic result), which H. Scheffé [?] described as " a model of mathematical rigor". Hsu's main conclusion, obtained by a combination of his analytical study with some numerical work, was that, for $\lambda = 0$ and varying θ, neither u_1 nor u_2 control the rejection probability (except when $m = n$). But the statistic u_2 is less sensitive to variation of θ. Up to now, the most common used method to solve this test problem is to utilize the rejection area $\{U > C\}$. Owing to Hsu's work, in the Behrens-Fisher problem, the method utilizing the rejection area $\{U > C\}$ is called "Hsu's Method".

(2) Another paper [?] of Hsu was concerned with the optimal estimate of σ^2 in the Gauss-Markov model. Let

$$y = A\beta + \varepsilon \tag{1}$$

be a linear model, where $y = (y_1, ..., y_n)^T$ (hereinafter M^T denotes the transpose of matrix M), $\beta = (\beta_1, ..., \beta_p)^T$, A is a known $n \times p$ matrix (rank p), $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^T$ with independent components satisfying

$$E\varepsilon_i = 0, \quad E\varepsilon_i^2 = \sigma^2, \quad E\varepsilon_i^4 = \alpha_i \sigma^4 \qquad (i = 1, ..., n),$$

where $\sigma > 0$ is parameter and $\alpha_i > 0$, $i = 1, \ldots, n$, are constants independent of σ. Let $Q = Q(y_1, y_2, ..., y_n)$ be the quadratic of $y_1, y_2, ..., y_n$. Hsu considered the class of quadratics Q which satisfy the following condition (i) unbiased, i.e. $EQ = \sigma^2$ for all β and σ^2, (ii) the variance of Q is independent of β. A quadratic Q in the class is said to be an optimal quadratic estimate of σ^2, if the variance of Q reaches the minimum within the class of quadratics.

Let $Q = y^T \Lambda y$, where Λ is a $n \times n$ symmetric matrix. Hsu obtained a necessary and sufficient condition for Q to be an optimal quadratic estimate of σ^2. A necessary and sufficient condition is that the matrix Λ has the following form

$$\Lambda = M D_\tau M,$$

where $M = I - A(A^T A)^{-1} A^T$ and A is the matrix in the model (1), I is the unit matrix,

$$D_\tau = \begin{pmatrix} \tau_1 & 0 & \cdots & 0 \\ 0 & \tau_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tau_n \end{pmatrix},$$

where $\tau_1, \tau_2, ..., \tau_n$ are the values at which the quadratic function

$$F = \sum_{i,j} \mu_{ij} \tau_i \tau_j$$

takes the minimum value under the constraint condition $\sum_{i=1}^{n} m_{ii}\tau_i = 1$, where

$$\mu_{ij} = \sum_{k=1}^{n} (\alpha_k - 3) m_{ki}^2 m_{kj}^2 + 2m_{ij}^2$$

$(i = 1, ..., n; j = 1, .., n)$, m_{ij} are elements of matrix $M = (m_{ij})_{n\times n}$, $\alpha_i = \sigma^{-4} E\varepsilon_i^4 \quad (i = 1, ..., n)$.

Let

$$S_0^2 = \frac{1}{n-p} \parallel y - A\hat{\beta} \parallel^2$$

where $\hat{\beta}$ is the least square estimate of β. Hsu obtained the result that a necessary and sufficient condition for S_0^2 to be an optimal quadratic estimate of σ^2 is

$$(n-p) \sum_{i=1}^{n} (\alpha_i - 3) m_{ii} m_{ik}^2 = m_{kk} \sum_{i=1}^{n} (\alpha_i - 3) m_{ii}^2.$$

This part of Hsu's work is regarded as the origin of the large literature on the optimal quadratic estimates of variance and variance components.

(3) For the small sample size inference, Hsu was concerned with the problem of testing univariate and multivariate linear hypotheses, particularly with power properties of the tests of these hypotheses. Linear hypotheses mean the hypotheses about linear relations between the parameters of the model. In [?], Hsu obtained the power function of Hotelling's T^2-test and pointed out under nonnull hypothesis the distribution of T^2 is the distribution of the ratio of a noncentral χ^2 to an independent central χ^2 variable. He also showed that the test is in a certain sense locally most powerful. Hsu formulated the general multivariate linear hypothesis in its canonical form. He obtained the nonnull distribution of the likelihood ratio test statistic when the covariance matrix is known. For the case of unknown covariance matrix, let θ_i be the nonzero roots of the associated determinantal equation. Hsu considered test statistics $W = \prod(1-\theta_i)$ and $V = \sum \theta_i/(1-\theta_i)$ and pointed out that their asymptotic power are the same when the sample size tends to infinity.(see [?])

In the small sample size inference, probably the most important result of Hsu's work is the finding and proving first optimum property for the likelihood ratio test of the linear hypothesis (see [?]). Without loss of generality, we present the result in the canonical form.

Let the joint density of $Y_1, Y_2, ..., Y_m, Z_1, ..., Z_n$ be

$$p(y_1, ..., y_m, z_1, ..., z_n) = (\sqrt{2\pi}\sigma)^{-(m+n)} \exp\{-\frac{1}{2\sigma^2}[\sum_{i=1}^{m} (y_i - \eta_i)^2 + \sum_{i=1}^{n} z_i^2]\},$$

where $\eta_1, ..., \eta_m$ are m arbitrary real numbers, σ is an arbitrary positive number, and all numbers are unknown. The problem is to test the hypothesis

$$H_0 : \eta_1 = \eta_2 = ... = \eta_{n_1} = 0,$$

where n_1 is a positive integer less or equal to m. Let

$$F = \frac{\sum_{i=1}^{n_1} y_i^2}{\sum_{i=1}^{n_1} y_i^2 + \sum_{i=1}^{n} z_i^2},$$

$$W_0 = \{(y_1, ..., y_{n_1}, z_1, ..., z_n) : F \geq F_\alpha\},$$

where F_α is the $1-\alpha$ quantile of $\tilde{F}$ under the null hypothesis, i.e. F_α is determined by the equation $P(\tilde{F} \geq F_\alpha | H_0) = \alpha$, where $\tilde{F} = \sum_{i=1}^{n_1} Y_i^2 / (\sum_{i=1}^{n_1} Y_i^2 + \sum_{i=1}^{n} Z_i^2)$.

When W_0 is used as the rejection area in the test problem, it is shown that the power function of the test has the form of $\beta_0(\lambda)$, where

$$\lambda = \frac{1}{2\sigma^2} \sum_{i=1}^{n_1} \eta_i^2.$$

Hsu obtained the following result: Suppose that the rejection area W is the set of points $(y_1, ..., y_m, z_1, ..., z_n)$ satisfying the conditions: the level of W is α, and the power function of W relies on the parameters only through the parameter λ, i.e. the power function has the form of $\beta(\lambda)$, then $\beta(\lambda) \leq \beta_0(\lambda)$ for all $\lambda > 0$.

In other words, for the problem of testing the null hypothesis H_0, the rejection area W_0 has the maximum power function within the class of rejection areas with level α whose power function relies on the parameter λ only. This is the first result about the optimal property for F-test, and the associated theorem was called "P. L. Hsu Theorem" (see the book "Analysis and Design of Experiments" by H.B. Mann (1949).) This work initiated two lines of developments. On the one hand, Hsu's work was applied to the multivariate problem, Hotelling's T^2 and the multiple correlation coefficient, by Simaika (1941). On the other hand, Hsu's paper offered a new method for obtaining all the similar tests and was formulated by means of the concept of completeness by Lehmann and Scheffé (see [?]).

(4) From 1938 to 1945, Hsu published several papers in the forefront of the development of the theory of multivariate analysis. He obtained several exact or asymptotic distributions of important statistics in the theory of multivariate analysis.

A crucial element of multivariate theory is the distribution of the sample covariance matrix S. Let $X_1, X_2, ..., X_N$ be a sample from p dimensional normal population $N(0, \Sigma)$, then

$$A \triangleq (N-1)S = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X})^T \tag{2}$$

has the so-called Wishart distribution $W(\Sigma, N-1)$. Hsu [?] derived the density function of the Wishart distribution based on algebra and analysis by mathematical induction. Up to now, it is said that Hsu's method of proof is the most elegant. (see [?])

In [?], Hsu obtained the joint distribution of roots of certain determinantal equation which is a basic result in the multivariate analysis. Let A and B be independent and from the Wishart distribution $W(\Sigma, m)$ and $W(\Sigma, n)$ respectively, where $m \geq p, n \geq p$ and p is the size of the matrix Σ. Let $\theta_1 \geq \theta_2 \geq ... \geq \theta_p$ be the roots of the determinantal equation

$$|A - \theta(A+B)| = 0.$$

After complicated calculations, Hsu proved that the joint density of $\theta_1, ..., \theta_p$ equals to a constant times

$$\prod_{i=1}^{p} \theta_i^{\frac{1}{2}(m-p-1)} \prod_{i=1}^{p} (1-\theta_i)^{\frac{1}{2}(n+p-1)}) \prod_{i=1}^{p} \prod_{j=i+1}^{p} (\theta_i - \theta_j).$$

Now suppose that the matrices A and Σ in (2) are partitioned into

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where A_{11} and Σ_{11} are matrices with size $p_1 \times p_1$, and A_{22} and Σ_{22} with size $p_2 \times p_2$. Sample and population canonical correlation are defined as the roots of the following equations respectively,

$$\begin{vmatrix} -\lambda A_{11} & A_{12} \\ A_{21} & -\lambda A_{22} \end{vmatrix} = 0, \quad \begin{vmatrix} -\lambda \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda \Sigma_{22} \end{vmatrix} = 0.$$

Hsu obtained the asymptotic distribution for the normalized sample canonical correlation ([**?**], page 142-149).

Let A and B be independent random matrices, where A has a noncentral Wishart distribution and B has a central Wishart distribution. Hsu considered the joint distribution of the roots of the following determinantal equation

$$|A - \phi B| = 0.$$

Under certain conditions, he obtained the asymptotic distribution of the roots ([**?**], page 129-140).

We should point out that Hsu's result about random matrix (for example, matrices A and B mentioned above) is one of the first results of random matrix appeared in modern research history. In the preface of the book " Random Matrices", the author Madan Lal Mehta took [**?**] of Hsu as one of the two first important papers in the random matrix theory.

To test the independence of identically distributed random variables $X_1, X_2, ..., X_N$, Hsu considered the statistic $T = Q/S$, where

$$Q = \sum_{i,j=1}^{N} a_{ij}(X_i - \bar{X})(X_j - \bar{X}), \quad S = \sum_{i=1}^{N} (X_i - \bar{X})^2.$$

Let a_{ij} rely on the sample size N. Under certain conditions, Hsu obtained the asymptotic expansion for distribution function of T as N tends to infinity ([**?**], page 224-228).

Besides, Hsu investigated the asymptotic behavior of function $f(\bar{u}_1, \ldots, \bar{u}_k)$ when the sample size tends to infinity, where $\bar{u}_1, \ldots, \bar{u}_k$ are independent vector sample means, and f is a smooth multivariate function. By applying the central limit theorem to the means, and by the Taylor expansion of $f(\cdot)$, Hsu obtained the result that the limit distribution is normal or the distribution of weighted sum of squares of normal random variables. We should point out that Hsu's method (by using central limit theorem and Taylor expansion) is a general method which is well known today in the large sample research area and is called the Δ-method.

(5) Hsu's important results in the probability theory and their applications were derived by his ability of proficient manipulation of characteristic functions. He was a true virtuoso in the method of characteristic function (see [**?**]). Hsu's result in [**?**] is a crucial improvement of Berry's result. Suppose that $\xi_1, \xi_2, ..., \xi_n$ are *i.i.d.* random variables with mean zero and variance one. Let

$$\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i, \quad \eta = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2$$

and denote by $\Phi(x)$ the standard normal distribution. An important problem both in theory and in application is the convergence rate of the normalized $\bar{\xi}$ and η towards the standard normal distribution. Let

$$F_n(x) \triangleq P(\sqrt{n}\bar{\xi} \leq x).$$

Cramér's showed that

$$F_n(x) = \Phi(x) + \psi(x) + R(x),$$

where $\psi(x)$ and $R(x)$ rely on the distribution of ξ_1 and $\lim_{n\to\infty} R(x) = 0$. Berry obtained the formula that for all x,

$$|F_n(x) - \Phi(x)| \leq A\beta_3 n^{-\frac{1}{2}},$$

where $\beta_3 = E|\xi_1|^3$ and A is an absolute constant which relies neither on n nor on the distribution of ξ_1.

In [**?**], Hsu extended Berry's method to give a simpler proof of Cramér's result on the asymptotic expansion of $F_n(x)$. Furthermore, instead of considering the sample mean $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$, Hsu investigated the asymptotic property of sample variance $\eta = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2$. Let

$$G_n(x) = P(\frac{\sqrt{n}(\eta - 1)}{\sqrt{\alpha_4} - 1} \leq x),$$

where $\alpha_4 = E\xi_1^4$. Under the condition that $\alpha_6 = E\xi_1^6 < \infty$ and $\alpha_4 - 1 - \alpha_3^2 \neq 0$ ($\alpha_3 = E\xi_1^3$), Hsu proved that for all x

$$|G_n(x) - \Phi(x)| \leq \frac{A}{\sqrt{n}}\Big(\frac{\alpha_6}{\alpha_4 - 1 - \alpha_3^2}\Big)^{\frac{3}{2}}, \tag{3}$$

where A is an absolute constant. Hsu also obtained the asymptotic expansion of $G_n(x)$ under the condition $E\xi_1^{2k} < \infty$ for some $k > 3$, and the reminder of the expansion has an upper bound.

By the way, influenced of Hsu's paper [?], Chen Xiru *et al* extended Hsu's result (3) to estimate of σ^2 in the linear model (1), and obtained a series of important results (see [?]).

(6) Hsu's paper [?] (with Robbins as the co-author) deals with the complete convergence of series of *i.i.d.* random variables, which is Hsu's another important contribution to probability theory. Let $\{\xi_n, n \geq 1\}$ be *i.i.d.* series with common mean μ and finite variance. Hsu proved that for arbitrary $\varepsilon > 0$, the following

$$\sum_{n=1}^{\infty} P(|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \mu| \geq \varepsilon) < \infty \tag{4}$$

holds. This result strengthens the classical Strong Law of Large Numbers. In [?], when (4) holds, the random series $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ is called complete convergence to the common mean μ of ξ_i. Hsu and Robbins further conjectured in [?] that the condition of finiteness of the variance of ξ is also a necessary condition for (4) to hold. Two years later, the famed mathematician P. Erdös proved the conjecture.

(7) Around 1940, a challenging problem was to find a solution of the most general form of the Central Limit Theorem, which drew the attention of many famed mathematicians, such as Levy, Feller, Kolmogorov and Gnedenko. Hsu was a competitor and the competition showed that he was also on the peak. Paper [?] was Professor Hsu's manuscript which Hsu mailed to Professor K.L. Chung in 1947. In this paper Hsu independently obtained the necessary and sufficient condition under which the row sums of a triangular array of infinitesimal random variables, independent in each row, converges in distribution to a given infinitely divisible distribution. Despite the fact that Gnedenko obtained the same result in 1944, Hsu's method is direct and has its own trait. When Professor K.L. Chung translated the book "Limit Theorems of Sums of Independent Random Variables" by B. V. Gnedenko and A. N. Kolmogorov in 1968, he decided to include Hsu's paper in the book as Appendix III.

(8) Hsu was an expert in manipulating characteristic functions. He used characteristic functions as a tool to obtain distribution of certain random variables, to calculate the power function in a test problem, to determine the limit distribution of series of random variables. He also obtained some important properties of characteristic functions.

Let $F(x)$ be the distribution function of random variable X. The characteristic function of $F(x)$ is defined as

$$f(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x).$$

In [?], Hsu obtained the necessary and sufficient condition, in terms of the property of its characteristic function on certain interval $(-\delta, +\delta)$ where $\delta > 0$ is a small number, for the finiteness of the βth absolute moment of the corresponding distribution, i.e.

$$M_\beta(F) = \int_{-\infty}^{+\infty} |x|^\beta dF(x) < \infty.$$

Paper [?] deals with the problem of identifying a characteristic function, i.e. to find a condition under which the values of a characteristic function on the interval $(-\infty, \infty)$ are determined by the values of the characteristic function on a small interval $(-\delta, \delta)$ with some $\delta > 0$. It is related to the moment problem, i.e., to find a condition under which the distribution function is determined by its moments. Gnedenko found a counterexample that two different characteristic functions coincide in a small interval $(-\delta, \delta)$, where δ is a positive number. Hsu classified characteristic functions into two groups. A characteristic function belongs to the class $(\bar{U})$, if it can be equal to another characteristic function in a neighborhood of zero without being equal to it identically. Hsu gave three subclasses of the class $(\bar{U})$. The first subclass is the simplest. However it includes all the known counterexamples. The second subclass consists of those characteristic functions of some stable distributions. The third subclass is composed of characteristic functions whose corresponding distribution function $F(x)$ has density $p(x)$ of the following form:

$$p(x) = O[\exp(-\frac{|x|}{\psi(|x|)})], \qquad |x| \to \infty;$$

where $\psi(x)$ has one of the following forms:

$$(\ln x)^\lambda, \quad (\ln x)(\ln \ln x)^\lambda, \quad \ldots \quad \lambda > 1.$$

(9) Hsu was an expert in applying matrices as a tool to solve mathematical problems. He obtained several theorems in matrix theory. In [?], Hsu studied the transform from square matrix A to square matrix B, where all the elements of the matrices are complex numbers,

$$A \to B = PA(\bar{P})^{-1},$$

where matrix P is nonsingular square matrix, and $\bar{P}$ is the conjugate matrix of P. If A and B can be transformed to each other, then A and B are called *similar*. Hsu obtained the canonical form under the sort of transformations for all matrices and a necessary and sufficient condition for two matrices to be *similar*.

In [?], Hsu studied a transform from a matrix pair (A_1, A_2) to another pair (B_1, B_2), where A_i and B_i are matrices of the same size, with complex number as their elements. If (A_1, A_2) can be transformed into (B_1, B_2) through the following equations

$$A_1 \to B_1 = PA_1Q, \quad A_2 \to B_2 = PA_2\bar{Q},$$

where P and Q are nonsingular square matrices, then the two pairs are called equivalent. Hsu found the canonical form for the equivalent pairs and a necessary and sufficient condition under which two pairs are equivalent.

In the long article [?] (38 pages in the original Chinese version and 54 pages in the English version), Hsu investigated profoundly the property of joint transformation of Hermitian matrix and a symmetric (or skew symmetric) matrix. Let A_1 be a Hermitian matrix (i.e. $A_1^T = \bar{A}_1$), and A_2 be a symmetric matrix (or skew symmetric). The transformation for the pair (A_1, A_2) is

$$A_1 \to B_1 = PA_1(\bar{P})^T, \quad A_2 \to B_2 = PA_2P^T,$$

where P is some nonsingular matrix, and the transpose of P is denoted by P^T. If the two pairs can be transformed to each other through these transforms, then the two pairs are said to be *congruent*. By a careful derivation and complex calculation, Hsu obtained the following conclusions:

(i) Let A_1 be a Hermitian matrix, A_2 be a symmetric matrix. The canonical forms of the pairs were found (there exist 7 different forms totally). A necessary and sufficient condition for two pairs to be *congruent* was obtained.

(ii) Let A_1 be a Hermitian matrix, A_2 be a skew symmetric matrix. The canonical forms of the pairs were found (there exist 8 different forms totally). A necessary and sufficient condition for two pairs to be *congruent* was obtained.

These results in Hsu's papers are important contributions to the matrix theory. The arguments in the proof exhibited Hsu's skills at manipulating matrices and his attention to details

(10) In [?], Hsu investigated the differentiability of probability transition function of a purely discontinuous homogeneous Markov process on the Euclidian space. Let X be the n dimensional Euclidian space, and $\mathcal{F}$ be the Borel σ-field in the space X and $p(t, x, E)$ be the probability transition function of a purely discontinuous homogeneous Markov process on X, i.e., for $x \in X, t > 0$ and $E \in \mathcal{F}$, $p(t, x, E)$ is the conditional probability of the event when the process is at state in the set E at time $s + t$ under the condition that the process is at x at time s. In [?], Hsu proved the differentiability of probability transition function $p(t, x, E)$ with respect to the variable t. Hsu also derived several integral equations for the differential of $p(t, x, E)$ which was a generalization of the results of Austin for the probability transition function on the discrete space. Hsu's method was more elementary than Austin's and Hsu's results were sharper.

In addition to the 10 aspects of his results, Professor P. L. Hsu led several seminars at Peking University in 1957-1966. Under his guidance, the participants obtained valuable results especially in the area of experiment of design and order statistics. Some results were published in the journals by pseudonym "BanCheng". Paper [**?**] deals with partial balanced incomplete block design (PBIB design). For certain design parameters, BanCheng obtained the condition of existence for the PBIB design with m-associate classes and constructed the design. In [**?**], the limit distribution of order statistics is investigated. Let $X_1, ..., X_n$ be *i.i.d.* random variables with common distribution function $F(x)$ and their order statistics be denoted by $\xi_1^{(n)} \leq ... \leq \xi_n^{(n)}$. BanCheng proved that under certain conditions, the series of normalized statistics $\xi_{k_n}^{(n)}$ $(k_n \to \infty, k_n/n \to \lambda \in [0, 1))$ has one of the following distributions as the limit distribution:

$$\phi_1(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

$$\phi_2(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha \ln x + \beta} e^{-\frac{t^2}{2}} dt & x > 0, \alpha > 0 \end{cases}$$

$$\phi_3(x) = \begin{cases} 1 & x \geq 0 \\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha \ln |x| + \beta} e^{-\frac{t^2}{2}} dt & x < 0, \alpha > 0 \end{cases}$$

The conditions for the common distribution function $F(x)$ to be in the domain of attraction of $\phi_i(x), i = 1, 2, 3,$ were also obtained respectively in [**?**].

Above we summarized Prof. Hsu's major scientific achievements. The book of "Pao-Lo Hsu Collected Papers" showed his accomplishment, and also reflected his high standard in research quality and scientific spirit. This kind of spirit is especially important for scientists today.

References

[1] Editorial board for “Pao-Lu Hsu Memorial Collection”, “Pao-Lu Hsu Memorial Collection”, Peking University Press, (2010), Beijing.

[2] School of Math. Sci., Peking University, In Memory of Professor P. L. Hsu on the Occasion of His 90th Birthday. Adv. Math., 29: 369-374.

[3] Anderson, T. W., Chung, K. L. and Lehmann, E. L., *Pao-Lu Hsu 1909-1970*, Ann. Statist, 1979, 7: 467-470.

[4] Lehmann, E. L., *Hsu’s Work on Inference*. Ann. Statist. 1979, 7: 471-472.

[5] Anderson, T. W., *Hsu’s Work in Multivariate Analysis*. Ann. Statist, 1979, 7: 474-478.

[6] Chung, K. L., *Hsu’s Work in Probability*. Ann. Statist, 1979, 7: 479-483.

[7] Jiang Zehan, Duan Xuefu, Preface of the book “Hsu Pao-Lu Collected Papers”, “Hsu Pao-Lu Collected Papers”, The Science Press of China, (1981), Beijing.

[8] John, N. L. and Kotz, S., Leading Personalities in Statistical Science from the Seventeenth Century to the Present. Wiley and Sons.1997.

[9] M. L. Mehta, Random Matrices, Elsevier Academic Press, 2004.

[10] Chen Xiru et al, Theory of Parameter Estimation in Linear Model, The Science Press of China, 1985, Beijing.

[11] Contribution to the theory of “ Student’s” t-test as applied to the problem of two-samples. Statist. Res. Mem, 2 (1938), 1-24.

[12] On the best unbiased quadratic estimate of the variance. Statist. Res. Mem. 2 (1938), 91-104.

[13] Notes on Hotelling’s generalized T. Ann. Math. Statist. 9 (1938), 231-243.

[14] On generalized analysis of variance. Biometrika. 31 (1940), 221-237.

[15] Analysis of variance from the power function standpoint. Biometrika. 32 (1941), 62-69.

[16] A new proof of the joint product moment distribution. Proc. Cambridge Philos. Soc. 35 (1939), 336-338.

[17] On the distribution of roots of certain determinantal equations. Ann. Eugenics. 9 (1939), 250-258.

[18] The approximate distribution of the mean and variance of a sample of independent variables. Ann. Math. Statist 16 (1945), 1-29.

[19] Complete convergence and the law of large numbers. (with H. Robbins) Proc. Nat. Acad. Sci. U.S.A. 33 (1947), 25-31.

[20] A general weak limit theorem for independent distributions. (Appendix III in Limit Theorems of Sums of Independent Random Variables by B. V. Gnedenko and A. N. Kolmogorov, translated by K. L. Chung. revised edition, Addison-Wesley 1968).

[21] Absolute moments and characteristic function, J. Chinese Math. Soc. (New series) 1 (1951), 257-280.

[22] On characteristic functions which coincide in a neighborhood of zero, Acta. Math. Sinica 4 (1954), 21-32.

[23] On a kind of transformations of matrices, Acta Math. Sinica 5 (1955), 333-346.

[24] On a kind of transformations of matrix pairs, Acta Sci. Natur. Univ. Pekinensis, 1 (1955), 1-15.

[25] Simultaneous transformation of a Hermitian matrix and a symmetric or skew-symmetric matrix, Acta Sci. Natur. Univ. Pekinensis, 3 (1957), 167-209.

[26] Differentiability of probability transition function of a purely discontinuous stationary Markov process on the Euclidian space, Acta Sci. Natur. Univ. Pekinensis, 4 (1958), 257-270.

[27] BanCheng, Partial balanced incomplete block design, (BanCheng is a pseudonym) Adv. Math., 7 (1964)，240-281.

[28] BanCheng, The limit distribution of order statistics, (BanCheng is a pseudonym), Acta Math. Sinica, 14 (1964)，694-714.

[29] Theory of Sampling, Peking University Press, (1982), Beijing.

[30] Hsu Pao-Lu Collected Papers, The Science Press of China, (1981), Beijing.

[31] Pao-Lu Hsu Collected Papers, Springer-Verlag (1983), New York.

[32] H. Scheffe, Practical solutions of of the Behrens-Fisher problem. J. Amer. Statist. Assoc. 65 (1970), 1501-1508.

Note: For the contents of P. L. Hsu's publications, see [**?**] or [**?**].