矩阵求逆的一种新方法

目 录

这是 80 年前的一篇旧文, 也是一篇短文. 原文于 1941 年发表在 The Journal of the American Statistical Association, Dec., 1941, Vol. 36, No. 216, pp. 530-534. 它讨论了计算逆矩阵的一种 "新" 方法, 内容涉及行列式, 伴随矩阵, 特征方程和特征根, 韦达定理, 牛顿恒等式, 凯莱—哈密尔顿定理等内容.

编者重新录入该文全文, 并在文后对若干概念进行了简单说明, 对其中的逻辑链条进行了梳理.

— 编者 (5070319@qq.com)

2021.06.03 ~ 2021.07.16

1 原文

A New Method For Obtaining The Inverse Matrix

By M. D. Bingham

THE INVERSE MATRIX is playing an increasingly important role modern statistical analyses. For example in fitting regression equations by the method of least squares the inverse matrix not only facilitates computation of the regression coefficients but also is required for obtaining the variances and covariances of these coefficients. Again the inverse matrix is essential in discriminant analysis[1] and in testing certain statistical hypotheses concerning sets of means.[2] It is, therefore, important to develop efficient methods for obtaining from a given matrix the numerical values of its inverse.

Aside from certain refinements introduced by the author, the method here presented was developed jointly by Professor Harold Hotelling of Columbia University and Mr. M. A. Girshick of the United States Department of Agriculture. This method has been successfully used in the United States Department of Agriculture, Bureau of Home Economics' Body Measurement Study.[3] We desire to calculate from a given non-singular, square matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

its inverse, say A^{-1}, where by definition[4]

$$A = \begin{bmatrix} A_{11}/a & \cdots & A_{n1}/a \\ \vdots & & \vdots \\ A_{1n}/a & \cdots & A_{nn}/a \end{bmatrix}, \tag{2}$$

A_{ij} being the cofactor of a_{ij} in the determinant a of A.

Consider the characteristic polynomial in λ,

$$|A - \lambda I| = \lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \cdots (-1)^n p_n = 0, \tag{3}$$

where I is the identity matrix and $p_r(r = 1, 2, \cdots, n)$ is the sum of the r-rowed principal minors of the matrix A. By a well known theorem in algebra[5], it can be shown that when the matrix A is substituted for λ, equation (3) still holds. That is,

[1] R. A Fisher, *Statistical Methods for Research Workers*, 7th ed

[2] Harold Hotelling, "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, Vol. 2 (1931), pp. 360 to 378.

[3] U.S. Department of Agriculture Misc. Publications #365 and #366.

[4] M. Bôcher, *Introduction to Higher Algebra*, pp. 74-75.

[5] Ibid., pp. 282,296-297.

$$A^n - p_1 A^{n-1} + p_2 A^{n-2} - \cdots (-1)^n p_n I = 0, \tag{4}$$

where the symbol 0 represents an n-rowed square matrix whose elements are all zero. Multiplying equation (4) by A^{-1} we get

$$A^{n-1} - p_1 A^{n-2} + p_2 A^{n-3} - \cdots (-1)^n p_n A^{-1} = 0, \tag{5}$$

or

$$p_n A^{-1} = (-1)^{n-1}(A^{n-1} - p_1 A^{n-2} + p_2 A^{n-3} - \cdots (-1)^{n-1} p_{n-1} I) = 0, \tag{6}$$

Since A is non singular and p_n equals the value of the determinant of A, equation (6) can be solved for A^{-1} in terms of the powers of A. Thus, if the numerical values for the p_r's are available the inverse matrix A^{-1} of A can be calculated from equation (6) by raising A to the appropriate powers, multiplying each A^{n-r-1} by $p_r/p_n (r = 0, 1, 2, \cdots, n-1; p_0 = 1)$ and adding algebraically the matrices thus obtained. In actual practice it has been found advisable first to calculate the adjoint $\overline{A}$ of A rather than A^{-1}. The adjoint $\overline{A}$ is by definition equal to A^{-1} times the value of the determinant of A. Since p_n equals the value of the determinant of A, $\overline{A} = p_n A^{-1}$. Hence, equation (6) becomes

$$\overline{A} = (-1)^{n-1}(A^{n-1} - p_1 A^{n-2} + p_2 A^{n-3} - \cdots (-1)^{n-1} p_{n-1} I) = 0. \tag{7}$$

In practice the values of p_r are as a rule not known and direct calculation of them is a laborious process. Fortunately, the process of raising the matrix to powers of itself yields the numerical values p_r. This can be seen from the following considerations. The quantities p_r are symmetric functions of the roots of equation (3). Moreover, if s_r stands for the sum of the rth powers of the n roots of equation (3) then by Newton's symmetric equations [6]

$$\begin{aligned} s_1 - p_1 &\equiv 0 \\ s_2 - p_1 s_1 + 2p_2 &\equiv 0 \\ s_3 - p_1 s_2 + p_2 s_1 - 3p_3 &\equiv 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ s_k - p_1 s_{k-1} + \cdots + (-1)^{k-1} p_{k-1} s_1 + (-1)^k k p_k &\equiv 0. \end{aligned} \tag{8}$$

Now it can be easily shown that s_1, (i.e. the sum of the roots of equation (3)) is equal to the sum of the principal diagonal elements of A, s_2 (i.e. the sum of the squares of the roots of equation (3)) is equal to the sum of the principal diagonal elements of A^2, and so on. Thus, knowing A, A^2, A^3, etc. it is a simple matter to obtain s_1, s_2, s_3, etc. Using the values of the s's thus obtained together with equations (8), the values of p's are easily obtainable.

It can readily be seen that the great bulk of the work is in the determination of the adjoint matrix, $\overline{A}$. Once it is determined, the division of each of its elements by the determinant p_n yields the desired inverse matrix A^{-1}. The matrix relation $A\overline{A} = p_n I$ yields the determinant p_n as elements in the principal

[6] Ibid., p. 243.

diagonal of the matrix $p_n I$, and also affords a check on the computations. So also can the matrix relation $AA^{-1} = I$ be used as a checking device. The details of the foregoing method can best be explained by working through an actual example involving four variates.

In the study of 5760 eight year old boys measured in fifteen states the intercorrelations for the variates stature, chest girth, hip girth and arm length were determined. These are given in the following matrix of correlations. Since the correlation matrix is symmetrical ($r_{ij} = r_{ji}$) only the necessary $n(n+1)/2$ terms are given in the matrix

$$A = \begin{array}{c} \\ \begin{bmatrix} 1.000000 & .615429 & .674646 & .852162 \\ & 1.000000 & .815032 & .608666 \\ & & 1.000000 & .627702 \\ & & & 1.000000 \end{bmatrix} \end{array} \begin{array}{c} \text{Check} \\ 3.142237 \\ 3.039127 \\ 3.117380 \\ 3.088530 \end{array}$$

Since n, the number of variates, is equal to four, equation (7) becomes

$$\overline{A} = -A^3 + p_1 A^2 - p_2 A + p_3 I.$$

Given A, it remains to compute A^2, A^3, p_1, p_2 and p_3.

To insure accuracy, as each row of the succeeding matrix is evolved, a check column is added, each element of which is equal to the sum of the elements appearing and understood on the particular row to which it refers.

To square A, multiply it by itself. That is, the element which lies in the ith row and jth column of A^2 is obtained by multiplying each element in the ith row of A by the corresponding element in the jth column of A, and summing the results.[7]

$$A^2 = \begin{array}{c} \\ \begin{bmatrix} 2.560080 & 2.299398 & 2.385790 & 2.502391 \\ & 2.413504 & 2.427322 & 2.253374 \\ & & 2.513434 & 2.326394 \\ & & & 2.490664 \end{bmatrix} \end{array} \begin{array}{c} \text{Check} \\ 9.747660 \\ 9.393598 \\ 9.652940 \\ 9.572824 \end{array}$$

To obtain A^3 multiply A by A^2. That is, to obtain the element which lies in the ith row and jth column of A^3, multiply each element of the ith row of A by the corresponding element of the jth column of A^2, and sum the results. The matrices are symmetrical, therefore, rows and columns are interchangeable.

$$A^3 = \begin{array}{c} \\ \begin{bmatrix} 7.717202 & 7.342561 & 7.557777 & 7.581124 \\ & 7.178517 & 7.360132 & 7.205486 \\ & & 7.561625 & 7.414589 \\ & & & 7.454941 \end{bmatrix} \end{array} \begin{array}{c} \text{Check} \\ 30.198667 \\ 29.086698 \\ 29.894124 \\ 29.656143 \end{array}$$

We now solve for p_1, p_2 and p_3. To do this we utilize equations (8). Substituting $s_1 = 4.000000$, $s_2 = 9.977682$, $s_3 = 29.912285$ in these equations; we obtain $p_1 = 4.000000, p_2 = 3.011159, p_3 = .682065$.

[7] Ibid., p. 63.

Knowing A, A^2, A^3, p_1, p_2, p_3; we need only multiply A^3 by -1, A^2 by p_1, A by p_2 and I by p_3. This is readily accomplished since the multiplication of a matrix by a scalar is a matrix each of whose elements is the scalar times the corresponding elements of the matrix. [8]

$$\begin{array}{rll} & & \text{Check} \\ -A^3 = & \begin{bmatrix} -7.717202 & -7.342561 & -7.557777 & -7.581124 \\ & -7.178517 & -7.360132 & -7.205486 \\ & & -7.561625 & -7.414589 \\ & & & -7.454941 \end{bmatrix} & \begin{matrix} -30.198667 \\ -29.086698 \\ -29.894124 \\ -29.656143 \end{matrix} \\ p_1A^2 = & \begin{bmatrix} 10.240320 & 9.197592 & 9.543160 & 10.009564 \\ & 9.654016 & 9.709288 & 9.013496 \\ & & 10.053736 & 9.305576 \\ & & & 9.962656 \end{bmatrix} & \begin{matrix} 38.990636 \\ 37.574392 \\ 38.611760 \\ 38.291292 \end{matrix} \\ -p_2A = & \begin{bmatrix} -3.011159 & -1.853155 & -2.031466 & -2.565995 \\ & -3.011159 & -2.454191 & -1.832790 \\ & & -3.011159 & -1.890111 \\ & & & -3.011159 \end{bmatrix} & \begin{matrix} -9.461775 \\ -9.151295 \\ -9.386927 \\ -9.300055 \end{matrix} \\ p_3I = & \begin{bmatrix} .682065 & & & \\ & .682065 & & \\ & & .682065 & \\ & & & .682065 \end{bmatrix} & \begin{matrix} .682065 \\ .682065 \\ .682065 \\ .682065 \end{matrix} \end{array}$$

To arrive at A it remains simply to add the corresponding elements of the matrices $-A^3, p_1A^2, -p_2A$ and p_3I.

$$\overline{A} = \begin{bmatrix} .194024 & .001876 & -.046083 & -.137555 \\ & .146405 & -.105035 & -.024780 \\ & & .163017 & .000876 \\ & & & .178621 \end{bmatrix} \begin{matrix} .012263 \\ .018460 \\ .012776 \\ .017165 \end{matrix}$$

Finally, to obtain the desired inverse matrix A^{-1}, we divide $\overline{A}$ by p_4. To obtain p^4 we make use of the matrix relation $A\overline{A} = Ip_n$. Multiplying A by $\overline{A}$ we get [9]

$$A\overline{A} = \begin{bmatrix} .046870 & .000000 & .000001 & .000000 \\ & .046870 & .000001 & -.000001 \\ & & .046870 & .000001 \\ & & & .046870 \end{bmatrix}.$$

Dividing A by p_4 (i.e., 0.046870), we obtain the desired inverse matrix

[8] Ibid., p. 62.

[9] It is of interest to note that $p_4 = \frac{1}{4}(-s_4 + p_1 s3 - p_2 s_2 + p_3 s_1)$ which in this case equals .046870. For practical reasons, however, it is perhaps better to obtain p_n from $A\overline{A}$, since to obtain p_n by Newton's formulae (8) it is necessary to compute the principal diagonal elements of A^n.

$$
A^{-1} = \begin{array}{c}
\begin{bmatrix}
4.139262 & .040026 & -.983209 & -2.934820 \\
 & -3.123640 & 2.240986 & -.538696 \\
 & & 3.478066 & .018690 \\
 & & & 3.810988
\end{bmatrix}
\end{array}
\begin{array}{l}
\text{Check} \\
.261617 \\
.393983 \\
.272562 \\
.366162
\end{array}.
$$

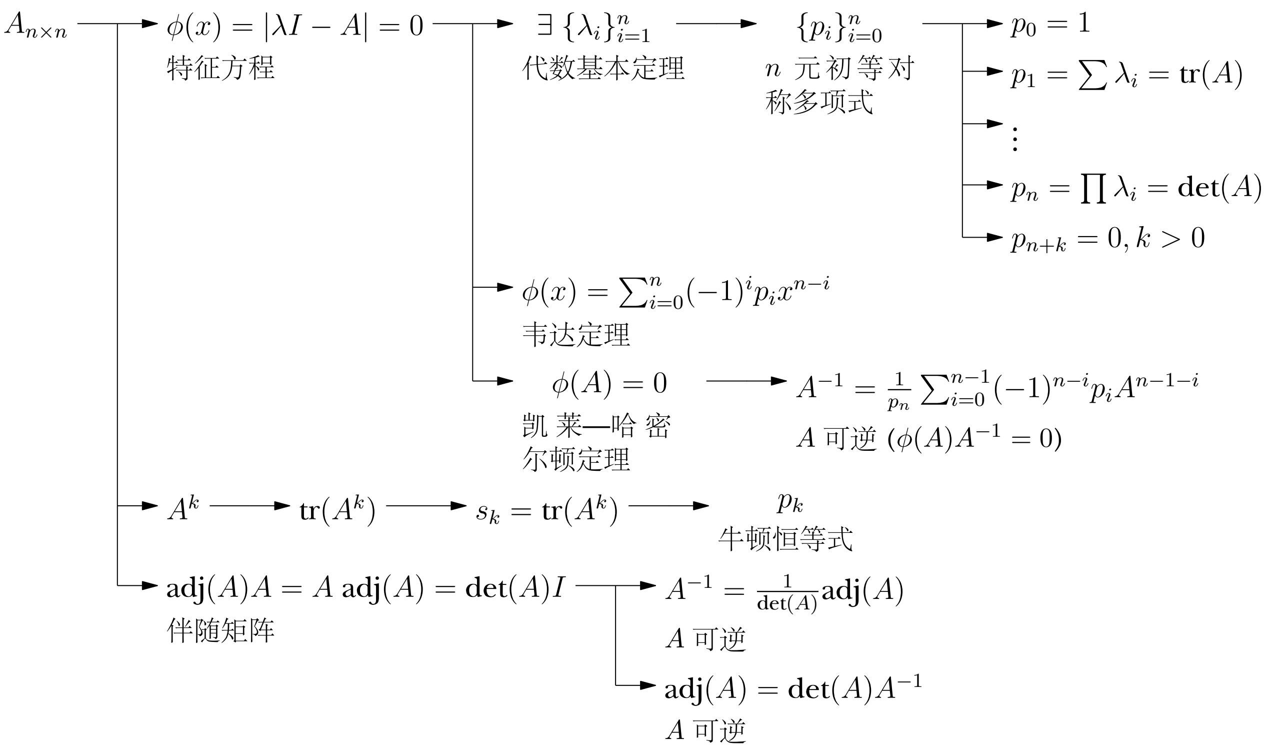
As a final check on the calculations, we multiply A by A^{-1}.

$$
AA^{-1} = \begin{bmatrix}
.999993 & -.000002 & .000014 & -.000007 \\
 & .999999 & .000031 & -.000018 \\
 & & 1.000005 & -.000014 \\
 & & & .999978
\end{bmatrix}.
$$

It can readily be seen that the principal diagonal elements are very close to unity, and the other terms close to zero.

2 注解

原文的主要逻辑推理链条如下图所示:



下面对图中的各个知识点做简单说明.

- 特征多项式/特征方程: 根据 Wikipedia, 特征多项式一般定义为 $\det(\lambda I - A)$ 或 $\det(A - \lambda I)$, 这两者的差别在于各项的符号差一个 $(-1)^n$, 而前者的优点在于从它得到的总是一个首一多项式, 后者仅当 n 为偶数时才为首一多项式. 由此可见, 原文第 (3) 式就显得稍欠严谨.

- n 元对称多项式:

  - 初等对称多项式 $p_i = \sum_{1\leqslant j_1<j_2<\cdots<j_k\leqslant n} X_{j_1}\cdots X_{j_k} (i \geqslant 0)$: 若以 $\{\lambda_i\}_{i=1}^n$ 为 n 个变元, 则有 $p_1 = \sum_{i=1}^n \lambda_i$, $p_n = \prod_{i=1}^n \lambda_i$. 另外, 有定义 $p_0 = 1$; $\forall k \geq 1, p_{n+k} = 0$.
  - 对称多项式基本定理: 任意的 n 元对称多项式都可以唯一地表示成初等对称多项式的多项式. 故幂和对称多项式 $\{s_k = \sum_{i=1}^n \lambda_i^k\}_{k=1}^\infty$ 可以唯一表示成 $\{p_i\}$ 的多项式.

- 牛顿恒等式: 即初等对称多项式和幂和对称多项式的关系 $kp_k = \sum_{i=1}^k (-1)^{i-1} p_{k-i} s_i$ 对 $n \geqslant 1, n \geqslant k \geqslant 1$ 成立. 该式展开为 $kp_k = p_{k-1}s_1 - p_{k-2}s_2 + \cdots + (-1)^{k-1}p_1 s_{k-1} + (-1)^k s_k$, 即:

$$
\begin{aligned}
k=1: \quad p_1 &= & p_0 s_1 \\
k=2: \quad 2p_2 &= & p_1 s_1 - p_0 s_2 \\
k=3: \quad 3p_3 &= & p_2 s_1 - p_1 s_2 + p_0 s_3 \\
k=4: \quad 4p_4 &= & p_3 s_1 - p_2 s_2 + p_1 s_3 - p_0 s_4 \\
\cdots\cdots
\end{aligned}
$$

- 韦达定理: 描述了多项式的根与系数的关系. 特别地, 对 n 阶方阵 A 的特征多项式, 有 $p_1 = \mathrm{tr}(A)$ 和 $p_n = \det(A)$. 另外, 若 λ_i 为 A 的特征根, 则 λ_i^k 为 A^k 的特征根, 故又有 $s_k = \mathrm{tr}(A^k)$.
- 行列式和拉普拉斯展开定理: 行列式的定义本身是用归纳法递归定义的, 它对应的定理 (拉普拉斯展开定理) 也只能用归纳法证明. 根据 Matrix Analysis, 2nd, 0.8.9, 定理的一般形式为:

> Let $A \in M_n(F)$, let $k \in \{1, \cdots, n\}$ be given, and let $\beta \in \{1, \cdots, n\}$ be any given index set of cardinality k. Then
>
> $$
> \begin{aligned}
> \det A &= \sum_\alpha (-1)^{p(\alpha,\beta)} \det A[\alpha, \beta] \det A[\alpha^c, \beta^c] \\
> &= \sum_\alpha (-1)^{p(\alpha,\beta)} \det A[\beta, \alpha] \det A[\beta^c, \alpha^c]
> \end{aligned}
> $$
>
> in which the sums are over all index sets $\alpha \subseteq \{1, \cdots, n\}$ of cardinality k, and $p(\alpha, \beta) = \sum_{i\in\alpha} i + \sum_{j\in\beta} j$.

当 $k = 1$ 时, 就是通常的按一行或一列展开的情况. 关于拉普拉斯展开的更友好的图解可以参考这里. 行列式的性质有很多, 其中一条是关于拉普拉斯展开的, 根据张禾瑞《高等代数》第四版定理 3.4.3, 行列式的某一行 (列) 的元素与另外一行 (列) 的对应元素的代数余子式的乘积的和等于零. 从这个性质可以推导出伴随矩阵的性质.

- 伴随矩阵 (adjugate): 方阵 A 的伴随矩阵是其余矩阵 (comatrix) 的转置矩阵.
  - 子式 (minor) 是 A 的子矩阵 (submatrix) 的行列式; A 的子矩阵是从 A 中挑选一些行和列后得到的.
  - 余子式也是一种子式, 它对应的子矩阵是通过删除 A 的一些行和一些列后得到的. 一般指的都是删除一行和一列以后得到的余子式.
  - 代数余子式 (cofactor): 余子式前面加上相应的正负符号后, 称为代数余子式.
  - 余矩阵 (comatrix): 由余子式组成的矩阵. 每个余子式所在的位置, 就是为了得到它对应的余矩阵, 从 A 中删除的那一行和那一列所交叉的位置.
  - 由于历史的原因, 伴随矩阵又常被称为 adjoint 或 classical adjoint, 这造成一些混淆. 关于这方面的讨论, 可以参考这里. 现在的通行规则似乎是, 用 adjugate 表示余矩阵的转置; 用 adjoint 表示线性变换的伴随变换, 它反映到变换所对应的矩阵, 就是矩阵的转置 ($\mathbb{R}$ 上) 或共轭转置 ($\mathbb{C}$ 上).
  - 伴随矩阵的性质 $\mathrm{adj}(A)A = A\mathrm{adj}(A) = \det(A)\ I$ 表明 $\mathrm{adj}(A)$ 具有类似于 A^{-1} 的特性, 事实上正是如此. 另外, 从行列式 $|\mathrm{adj}(A)A| = \det(A)^n$ 的角度看, $|\mathrm{adj}(A)| = \det(A)^{n-1}$,

这可以直观理解为 adj(A) 对体积的拉伸效果是 A 的 $n-1$ 次方, 因为其每个元素所对应的子矩阵为 $(n-1)\times(n-1)$.

- 克莱姆法则: 由伴随矩阵的性质马上可以得到克莱姆法则, $x = A^{-1}b = \frac{1}{\det(A)}\text{adj}(A)b$, 其中向量 adj$(A)b$ 的第 i 行就是 $\det(A \underset{i}{\leftarrow} b)$. 这里采用 *Matrix Analysis* 中的记号 $A \underset{i}{\leftarrow} b$ 表示把 A 中的第 i 列用 b 替换. 该书 0.8.3 节用行列式乘积的性质也简洁地推导出了克莱姆法则: $A(I \underset{i}{\leftarrow} x) = A \underset{i}{\leftarrow} b \implies \det(A)\det(I \underset{i}{\leftarrow} x) = \det(A \underset{i}{\leftarrow} b) \implies$ 克莱姆法则.

- 特征多项式的系数和主子式的关系: 原文在式 (3) 后面指出了特征多项式的系数和主子式的关系. 这个关系在原文的算法中并没有用到, 所以它没有出现在上面的逻辑推理链条图中. 关于这一论断, 以及确定特征多项式系数的算法, 可以参考下面的链接:
  - The coefficients of the characteristic polynomial in terms of the eigenvalues and the elements of an $n \times n$ matrix
  - On the coefficients of the characteristic polynomial of a matrix

- Cayley-Hamilton 定理: 该定理是原文算法的重要依据, 而且它还有其它很多方面的应用. 不过编者对此了解也不多, 具体内容可参考张贤达《矩阵分析与应用》第二版第 7.3 节 "Cayley-Hamilton 定理及其应用" 等.

- Faddeev-LeVerrier 算法: 本文所谓的 "新方法" 其实已经不新, 早在 1840 年法国数学家勒维耶首次发表了计算方阵的特征多项式系数的递归算法, 后来又有多人重新发现和改进. 目前一般以前苏联数学家Dmitrii Faddeev和勒维耶来命名. 关于该算法的更多信息, 可以参考下面的链接:
  - Wikipedia: Faddeev-LeVerrier 算法
  - MATLAB: Faddeev-LeVerrier Algorithm
  - math.stackexchange: What is the fastest way to find the characteristic polynomial of a matrix?