طرحهای بهینه در مسائل رگرسیون چندجمله‌ای

حسن شاهقلیان*

چکیده

تعیین روابط بین متغیرهایی که اندازه‌گیری آنها با عدم حتمیت صورت می‌گیرد، در رشته‌های مختلف علمی از اهمیّت زیادی برخوردار است. برای تعیین این روابط به طور تجربی، لازم است که آزمایشهایی با طرحهای معیّنی انجام شوند. مشخص کردن مقادیر متغیر مستقل به طوری که نتایج از نوعی بهینگی برخوردار باشند مسألهٔ طرحهای بهینه را پیش آورده است. در این مقاله، با تعریف ملاکهای بهینگی، نحوهٔ تعیین طرحهای بهینه از نوع Dـ بهینه را تشریح می‌کنیم.

مقدمه

تجربه علمی نشان داده است که در برخورد با پدیده‌های طبیعت، مطالعهٔ رفتارهای این گونه پدیده‌ها نیاز به ساختن قالب یا قالبهایی دارد که موارد مشابه را بتواند در خود جای داده و بیان کنندهٔ خصایص مشترک آنها باشد. یکی از راههای ساختن این قالبها، استفاده از فرمولهای ریاضی و مدلهای آماری است. از رایجترین این مدلها، مدل رگرسیونی است که از آن برای مطالعهٔ میزان تأثیر متغیرهای ورودی $X = (x_1, \ldots, x_p)$ (که قابل مشاهده‌اند) بر روی یک متغیر پاسخ Y استفاده می‌شود. فرض می‌کنیم رابطهٔ X و Y (مدل) به صورت

$$Y = f'(X)\theta + \varepsilon$$

باشد که در آن $f = (f_1, \ldots, f_k)$ برداری از توابع معلوم، $\theta = (\theta_1, \ldots, \theta_k)$ برداری از پارامترهای نامعلوم و ε خطای تصادفی (خطای انحراف از مدل) با توزیع $N(0, \sigma^2)$ است.

حال اگر n مشاهده از Y داشته باشیم، می‌توان مشاهدات را با $Y_1, \ldots, Y_n$ نمایش داد که هر Y_i به صورت $Y_i = f'(X_i)\theta_i + \varepsilon_i$ خواهد بود. به صورت ماتریسی مدل را می‌توان به شکل زیر نمایش داد:

$$Y = W\theta + \varepsilon$$

که در آن Y بردار $n \times 1$، W ماتریس $n \times k$ از مشاهدات است که i ـامین سطر آن بردار $f'(X_i)$ بوده و X_iها مقادیری در مجموعهٔ Ω، زیرمجموعه‌ای از اعداد حقیقی‌اند. با فرض اینکه $W'W$ وارون‌پذیر باشد، با استفاده از روش کمترین توانهای دوم برآورد θ عبارت است از

$$\hat{\theta} = (W'W)^{-1}W'Y$$

با واریانس

$$V(\hat{\theta}) = \sigma^2(W'W)^{-1}$$

۱ طرح‌ریزی

مسألهٔ طرح‌ریزی، انتخاب مقادیر X_i، $i = 1, 2, \ldots, n$، از فضای Ω به

*) حسن شاهقلیان، عضو هیأت علمی دانشگاه شهرکرد

بهترین صورت است. واضح است که با تغییر X_iها متغیر پاسخ نیز عوض خواهد شد و برآورد θ یعنی $\hat{\theta}$ نیز تحت تأثیر قرار خواهد گرفت، به عنوان مثال اگر مسأله رگرسیون، مطالعهٔ برخی از خصوصیات کودکان و به ویژه اگر متغیر ورودی وزن کودک و متغیر پاسخ قد کودک باشد، انتخاب کودک خود یکی از مشکلات اولیه کار است، از این رو نیاز به معیارهایی داریم که انتخاب X_iها را برای دستیابی به «بهترین طرح» میسر سازد. ذیلاً برخی از این معیارها را معرفی می‌کنیم یکی از راههایی که می‌توان از طریق آن معیاری برای بهینگی پیدا کرد، توجه به ماتریس واریانس ـ کواریانس $\hat{\theta}$ است که برحسب ماتریس اطلاع به صورت $\frac{\sigma^2}{n}M^{-1}(h)$ است (رجوع کنید به [۲]) که در آن σ^2 واریانس خطای تصادفی و n مجموع تعداد تکرارها، مقادیری ثابت‌اند.

چنانچه هر X_i به تعداد n_i بار $(i = 1, 2, \dots, k)$ تکرار شود، ماتریس اطلاع به صورت

$$M = nW'W = \sum_{i=1}^{K} n_i f(X_i)f'(X_i), \sum_{i=1}^{K} n_i = n$$

خواهد بود، زیرا این وضعیت معادل است با اینکه متغیر تصادفی‌ای با توزیع h داشته باشیم که مقادیر X_1، ...، X_k را با احتمالهای $\frac{n_1}{n}$، ...، $\frac{n_k}{n}$ اختیار کند. در این صورت ماتریس M به صورت

$$M(h) = E(f(X)f'(X))$$

خواهد بود. لذا یک طرح را می‌توان یک اندازهٔ شمارشی روی فضای Ω در نظر گرفت که به نقاط (متناهی) $X_1, \dots, X_k$ اندازه‌های مثبت $\frac{n_1}{n}, \dots, \frac{n_k}{n}$ را منتسب می‌کند.

در طرح‌ریزی بهینه، هدف به نحوی ماکسیمم کردن ماتریس M (یا مینیمم کردن M^{-1}) روی کلیهٔ hهایی است که دارای تکیه‌گاه متناهی هستند.

۲ تعریف برخی از معیارهای بهینگی

(الف) معیار Dـبهینه

چون ماتریس واریانس ـ کواریانس $\hat{\theta}$ برحسب ماتریس اطلاع برابر $\frac{\sigma^2}{n}M^{-1}(h)$ است، در این معیار قصد آن است که دترمینان این ماتریس مینیمم شود، به عبارت دیگر طرح h^* برای $\theta_1, \theta_2, \dots, \theta_k$ را Dـ بهینه گویند هرگاه

$$\det M^{-1}(h^*) = \min_h \det M^{-1}(h).$$

(ب) معیار G بهینه

اگر واریانس $\hat{\theta}$ را برای مقداری خاص از $\underset{\sim}{X}$ در نظر بگیریم، این ماتریس به صورت $\frac{\sigma^2}{n}f'(X)M^{-1}(h)f(X)$ در خواهد آمد که برای راحتی آن را به صورت $\frac{\sigma^2}{n}d(X,g)$ نشان می‌دهیم، چنانچه مقدار $d(X,g)$ را برای بدترین مشاهدهٔ X بتوانیم مینیمم کنیم، معیار دیگری به نام G بهینه به دست می‌آید. لذا h^* را یک طرح Gـ بهینه برای $\theta_1, \theta_2, \dots, \theta_k$ گویند هرگاه

$$\min_h \max_{X\in\Omega} d(X,h) = \max_{X\in\Omega} d(X,h^*).$$

(پ) معیار کلی ϕـ بهینه

این معیار کلی‌ترین معیار بهینگی است که در آن به جای ماتریس $M^{-1}(h)$، ماتریس $J(h) = (KM^{-}(h)K')^{-1}$ در نظر گرفته می‌شود. لازم به ذکر است که از این معیار می‌توان در دو حالت استفاده کرد، حالتی که تمام پارامترهای بردار θ موردنظر است و حالتی که فقط بعضی از آنها مورد نظر هستند. به همین منظور به جای θ بردار $K\theta$ در نظر گرفته شده که K ماتریس معلوم $s \times k$ با مرتبهٔ k است. نظر به اینکه ممکن است ماتریس اطلاع همیشه وارون‌پذیر نباشد، در این معیار به جای $M^{-1}(h)$ از وارون تعمیم‌یافتهٔ ماتریس $M(h)$ استفاده کرده آن را با $M^{-}(h)$ نشان داده‌ایم. بدیهی است چنانچه ماتریس $M(h)$ وارون‌پذیر باشد در فرمول $J(h)$ همان $M^{-1}(h)$ قرار می‌گیرد.

معیارهای دیگری از قبیل Eـ بهینه و طرحهای بلوکی و Aـ بهینه و Cـ بهینه و ... وجود دارند که ما در اینجا از ذکر آنها خودداری می‌کنیم.

مثال: رگرسیون درجه دوم روی فضای $\Omega = [-1, 1]^q$.

در مدل خطی $y = f'(X)\theta + \varepsilon$ فرض کنیم Ω شامل تمام نقاط به شکل

$$\Omega = \{X = (x_1, \dots, x_q); -1 \leq x_i \leq 1, i = 1, 2, \dots, q\}$$

باشد که اصطلاحاً آن را مکعب می‌گوییم. بردار $f(X)$ را به صورت زیر در نظر می‌گیریم

$$f'(X) = (1, x_1^2, \dots, x_q^2, x_1, \dots, x_q, x_1x_2, \dots, x_{q-1}x_q)$$

و نیز

$$\theta' = \left(\theta_0, \theta_1, \dots, \theta_q, \theta_{q+1}, \dots, \theta_{2q}, \theta_{2q+1}, \dots, \theta_{\frac{(q+1)(q+2)}{2}}\right).$$

یک طرح Dـ بهینهٔ h عبارت است از اندازهٔ احتمال h که به هر کدام از گوشه‌های این مکعب اندازه‌ای را نسبت دهد. این طرح برای حالت $q = 1$ و $q = 2$ ساده‌است، لذا حالت $q \geq 2$ را بررسی می‌کنیم.

h را به این صورت معرفی می‌کنیم که تنها به نقاط گوشه و نقاط میانی هر ضلع و نقاط مرکز هر وجه به ترتیب اندازه‌های α، β، و γ را نسبت داده و به بقیهٔ نقاط فضا اندازهٔ صفر نسبت دهد. به منظور ایجاد تقارن در فضای q بعدی (مکعب) گوشه به نقاطی می‌گوییم که مختصات آنها تنها اعداد ۱ و ۱− (بدون صفر)، نقطهٔ میانی ضلع را نقطه‌ای می‌گیریم که مختصات آن اعداد ۱ و ۰ و ۱− (دقیقاً دارای یک صفر)، و منظور از نقطهٔ مرکز وجه نقطه‌ای است که مختصات آن اعداد ۱ و ۰ و ۰ و ۱− (دقیقاً دارای دو صفر) باشد.

حالت	:	$q = 2$
نقاط گوشه	:	A_1, A_2, A_3, A_4
نقاط میانی	:	B_1, B_2, B_3, B_4
نقطهٔ مرکز	:	C

حالت	:	$q = 3$
نقاط گوشه	:	A_1
نقاط میانی	:	B_1, B_2, B_3
نقطهٔ مرکز	:	C_1, C_2, C_3

در این صورت به کمک روشهای شمارش می‌توان نشان داد که این فضا دارای 2^q گوشه، $q2^{q-1}$ نقطهٔ میانی و $q(q-2)2^{q-3}$ نقطهٔ مرکز است. (توجه کنید که در شکل بالا برای حالت $q = 3$ نصف فضای Ω نمایش داده شده است، در حقیقت حالتی نمایش داده شده که $(x_1, x_2, x_3) \in \Omega : 0 \leq x_i \leq 1$. همچنین فرمولهای تعداد گوشه و میانی و مرکز برای $q > 2$ صدق می‌کند). از آنجا که محاسبات برای $q \geq 6$ بسیار طولانی و خسته کننده است تنها برای $q = 2, 3, 4, 5$ طرح بهینه را تعیین خواهیم کرد.

ماتریس اطلاع حاوی درایه‌های زیر است:

$$m_{ij}(h) = \sum_{\{-1,0,1\}} f_i(X) f_j(X) h(X).$$

در این ماتریس چهار نوع مؤلفه وجود دارد:

$$m_{11}(h) = 1$$

مؤلفه‌های دوم تا qـام سطر اول همگی برابرند با:

$$\begin{aligned} u &= \sum_{\{-1,0,1\}} x_1^2 h(X) \\ &= 2^{q-3}(8\alpha + 4(q-1)\beta + (q-1)(q-2)\gamma) \end{aligned}$$

مؤلفه‌های سوم تا qـام سطر دوم همگی برابرند با:

$$\begin{aligned} v &= \sum_{\{-1,0,1\}} x_1^2 x_2^2 h(x_1, x_2) \\ &= 2^{q-3}(8\alpha + 4(q-2)\beta + (q-2)(q-3)\gamma) \end{aligned}$$

حال اگر این مقادیر را در ماتریس $M(h)$ قرار داده و آن را وارون کنیم به صورت:

$$M^{-1}(h) = \begin{bmatrix} H^{-1} & & \\ & U^{-1} I_q & \\ & & V^{-1} I_{\frac{q(q-1)}{2}} \end{bmatrix}$$

به دست خواهد آمد که در آن

$$H^{-1} = \begin{bmatrix} a & b & & \dots & b \\ b & c & d & \dots & d \\ \vdots & d & \ddots & & \\ & & & & \\ b & d & & \ddots & c \end{bmatrix}$$

$$a = \frac{(q-1)v + u}{(q-1)v + u - qu^2}$$

$$b = \frac{-u}{(q-1)v + u - qu^2}$$

$$c = \frac{(q-1)v + u - (q-1)u^2}{(q-v)((q-1)v + u - qu^2)}$$

$$d = \frac{u^2 - v}{(u-v)((q-1)v + u - qu^2)}.$$

بدین ترتیب دترمینان ماتریس $M(h)$ برابر خواهد بود با

$$\det M(h) = u^q v^{\frac{q(q-1)}{2}} (q-v)^{q-1}(u + (q-1)v - qu^2)$$

اینک باید α و β و γ را به گونه‌ای محاسبه کنیم که اندازهٔ h بتواند Dـ بهینه باشد که بعد از محاسبات بسیار طولانی خواهیم داشت:

$$\begin{aligned} \alpha = (2^{q+2}(q+2)^3(q+1))^{-1}\{(&4q^6 + 12q^5 - 25q^4 - 107q^3 \\ &+ 85q^2 + 479q + 128 - (2q^2 - q - 19)q(q-1) \\ &\times (q+3)\sqrt{4q^2 + 12q + 17}\} \end{aligned}$$

$$\begin{aligned} \beta = (2^{q+2}(q+2)^3(q+1))^{-1}\{-(&4q^5 + 16q^4 - 11q^3 - 143q^2 \\ -149q + 139) + (&q+3)(q-1)(2q^2 + q - 15)\sqrt{4q^2 + 12q + 17}\} \end{aligned}$$

$$\begin{aligned} \gamma = (2^{q+1}(q+2)^3(q+1))^{-1}\{(&4q^4 + 24q^3 + 43q^2 - 24q \\ &- 119 - (q+3)(2q^2 + 3q - 11)\sqrt{4q^2 + 12q + 17}\} \end{aligned}$$

و بدین ترتیب طرح D- بهینهٔ h به دست می‌آید. به عنوان مثال برای چند مقدار خاص q؛ α، β، و γ به شرح زیرند:

q	α	β	γ
۲	۰٫۱۴۵۸	۰٫۰۸۰۱۵	۰٫۰۹۶۲
۳	۰٫۰۷۱۹۷۵	۰٫۰۱۸۹۵	۰٫۰۳۲۸
۴	۰٫۰۳۷۰۵	۰٫۰۰۳۸۳۷۵	۰٫۰۱۱۸۵
۵	۰٫۰۱۹۲۸	۰٫۰۰۰۳۱۲۵	۰٫۰۰۴۴۷۵

مراجع

[1] Studden, W. J. (1989) "Note on some p-omital designs for polynomial regression"The Annl. of Stat. Vol. 17, No. 2, pp 618-623.

[2] Gaffke, N. (1985) "Singular information matrices, directional derivatives, and subgradients in optimal design theory"In linear stat. inf., Proc. Internat. Conf. Poznan (Poland) (T. Calinski and W. Klenechki, eds) pp 61-77.

[3] Calil, Z. and Kiefer. J. (1980) "D-optimal weighing designs"Annl. Stat. Vol. 8 No. 6, pp 1239-1306.

[4] Gaffke, N. (1987) "Further Characterizations of design optimality and admissibility for partial paramter estimation in linear regression"The Annl. of Stat. Vol. 15. No, 3, pp 942-957.

[5] John, R. C. St. and Draper, N. R. (1975) "Optimality for regression designs"Vol. 17, No. 1, pp 15-23.

[6] Fedorov, V., V. (1972) "Theory of optimal experiments", Academic Press, New York.