مجله علوم آماری، پاییز و زمستان ۱۳۹۰
جلد ۵، شماره ۲، ص ۱۷۹–۱۸۸

طرح‌های E_{tc}-بهینه بلوکی برای مقایسه تیمارها با یک کنترل و مشاهدات همبسته

محبوبه دوستی ایرانی، سعید پولادساز
دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان

تاریخ دریافت: ۱۳۹۰/۵/۲ تاریخ آخرین بازنگری: ۱۳۹۰/۱۲/۲۲

چکیده : در این مقاله روشی برای دستیابی به طرح بلوکی ناقص E_{tc}-بهینه برای مقایسهٔ تیمارهای آزمایش با تیمار کنترل تحت این فرض که مشاهدات درون بلوک‌ها همبسته‌اند، ارائه می‌شود. سپس الگوریتمی برای ساختن طرح بهینه بیان می‌شود که برای هر ساختار همبستگی با درایه‌های غیر قطری نامثبت قابل استفاده است.

واژه‌های کلیدی : طرح بهینه، خطاهای همبسته، طرح بلوکی ناقص، فرایند اتورگرسیو.

۱ مقدمه

در بعضی از آزمایش‌های علمی تیماری وجود دارد که نسبت به تیمارهای آزمایش دارای نقش مهم‌تری است و تیمار استاندارد یا کنترل نامیده می‌شود. در این حالت تمام مقایسه‌های زوجی بین تیمارها به یک اندازه دارای اهمیت نیستند، بلکه تنها مقایسهٔ تیمارهای آزمایش با تیمار کنترل مورد توجه قرار می‌گیرد. به عنوان مثال، در مطالعات داروسازی، داروهای جدید تیمارهای آزمایش هستند و یک دارونما به

آدرس الکترونیک مسئول مقاله: محبوبه دوستی ایرانی، m.doosti66@gmil.com
کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۲K۰۵ و ۶۲K۱۰

عنوان تیمار کنترل در نظر گرفته می‌شود. برای انجام چنین آزمایش‌هایی، از طرح‌هایی موسوم به طرح‌های آزمایش-کنترل استفاده می‌شود. طرح بلوکی یکی از متداولترین طرح‌ها برای انجام این‌گونه آزمایش‌ها است. فرض کنید قرار است $v \geq ۲$ تیمار آزمایش با یک تیمار کنترل مقایسه شوند. تیمارها با برچسب‌های ۰، ۱، ...، v نشان داده می‌شوند که در آن ۰ نشان دهندهٔ تیمار کنترل و ۱، ...، v نشان دهندهٔ تیمارهای آزمایش هستند. تیمارها در b بلوک با اندازه برابر k مرتب می‌شوند، که در آن $k \leq v$، یعنی طرح‌ها بلوکی ناقص هستند. کلاس همه طرح‌های توصیف شده با $\mathcal{D}(v+۱, b, k)$ نشان داده می‌شود.

تا کنون مطالعات گسترده‌ای در ارتباط با مسألهٔ یافتن طرح بهینه در این کلاس انجام شده است. از جملهٔ آنها می‌توان به ماجامدار و نوتز (۱۹۸۳)، هدایت و ماجامدار (۱۹۸۴)، چنگ و همکاران (۱۹۸۸) و استافکن (۱۹۸۷ و ۱۹۸۸) اشاره کرد. در اکثر مطالعات انجام شده در مورد بهینگی طرح‌های آزمایش-کنترل بلوکی، مشاهدات ناهمبسته در نظر گرفته شده‌اند. در حالی که آزمایش‌های بسیاری وجود دارند که در آنها مشاهدات درون یک بلوک همبسته هستند (وو و همکاران، ۲۰۰۵). کاتلر (۱۹۹۳) به کمک برآوردهای کمترین توان‌های دوم تعمیم یافته نتایجی در زمینهٔ A_{tc}-بهینگی طرح‌های بلوکی با اندازه بلوک برابر تحت یک فرآیند اتورگرسیو مرتبه اول AR(1) به دست آورد. کانرت و همکاران (۲۰۱۰) به ارائهٔ روشی برای یافتن طرح‌های بلوکی A_{tc}-بهینه با اندازه بلوک برابر تحت بعضی از ساختارهای همبستگی پرداختند و با به کارگیری آن، طرح‌های A_{tc}-بهینه را تحت فرآیند اتورگرسیو مرتبه اول برای طرح‌های بلوکی با اندازه‌های ۳ یا ۴ تعیین کردند.

در این مقاله با در نظر گرفتن یک تیمار کنترل و v تیمار آزمایشی در بخش مدل آماری طرح و ماتریس اطلاع با مرور مقاله کانرت و همکاران (۲۰۱۰) معرفی می‌شوند. در بخش ۳ روشی برای یافتن طرح آزمایش-کنترل E_{tc}-بهینه در کلاس طرح‌های بلوکی ناقص با اندازه بلوک برابر ارایه می‌شود. در بخش ۴ الگوریتمی بر اساس نتایج به دست آمده از بخش ۳، به منظور دستیابی به طرح بهینه ارائه می‌شود و در نهایت با استفاده از الگوریتم مطرح شده، طرح‌های آزمایش-کنترل E_{tc}-بهینه تحت فرآیند AR(1) به دست می‌آیند.

۲ مدل آماری

فرض کنید آزمایشی با $v+1$ تیمار در b بلوک هر یک با اندازه k باید انجام گیرد. یک مدل آماری مناسب برای این طرح به صورت

$$y = T_d\tau + B\beta + e$$

است، که در آن $y=[y_{11},\ldots,y_{1k},y_{21},\ldots,y_{bk}]'$، y_{ij} j–امین مشاهده در i–امین بلوک، T_d ماتریس طرح تیمارها، $B=I_b\otimes \mathbf{1}_k$ ماتریس بلوک‌های طرح و e بردار خطاهای تصادفی با ماتریس کوواریانس Σ است. ماتریس کوواریانس به صورت $\Sigma=\sigma^2(I_b\otimes V)$ در نظر گرفته می‌شود، که در آن V یک ماتریس معلوم پررتبه است، یعنی همبستگی درون بلوک‌ها برای همه بلوک‌ها یکسان و خطاها در بلوک‌های متفاوت ناهمبسته هستند. به منظور برآورد پارامترهای مدل از روش کمترین توان‌های دوم تعمیم یافته استفاده می‌شود. با تعریف ماتریس

$$W = V^{-1} - V^{-1}1_k(1_k'V^{-1}1_k)^{-1}1_k'V^{-1}$$

ماتریس اطلاع طرح به صورت $C_d=T_d'[I_b\otimes W]T_d$ به دست می‌آید، که می‌توان آن را به صورت

$$C_d = \begin{pmatrix} c_{00} & c_{n0}' \\ c_{n0} & M_d \end{pmatrix} \qquad (1)$$

افراز کرد، که در آن c_{00} مقداری ثابت، c_{n0} برداری v بعدی و M_d ماتریسی است که از حذف سطر و ستون مربوط به تیمار کنترل به دست می‌آید. در واقع ماتریس M_d، ماتریس اطلاع مقابله‌های تیماری $(\tau_0-\tau_1,\ldots,\tau_0-\tau_v)'$ است (بچهوفر و تمهانه، ۱۹۸۱). از طرفی هر طرح بلوکی را می‌توان به صورت مجموعه‌ای از b دنباله تیماری s در نظر گرفت. در این صورت ماتریس اطلاع طرح را می‌توان با استفاده از جمع ماتریس‌های اطلاع مربوط به هر بلوک به دست آورد. به عبارت دقیق‌تر ماتریس اطلاع را می‌توان به صورت

$$C_d = \sum_{u=1}^{b} C(s_u) \qquad (2)$$

بازنویسی کرد، که در آن C_{s_u} ماتریس اطلاع متناظر با u-امین بلوک است و با استفاده از دنباله تیماری s_u ساخته شده است.

لازم به ذکر است که درایه‌های غیر قطری ساختارهای همبستگی در نظر گرفته شده در این مقاله نامثبت هستند.

۳ طرح E_{tc}-بهینه

طرح d را E_{tc}-بهینه گویند اگر تابع $\phi(M_d) = \max_{\{i=1,\ldots,v\}} \mu_i^{-1}$ را کمینه کند، که در آن μ_iها مقادیر ویژه ماتریس اطلاع M_d هستند. در حالتی که خطاها همبسته باشند به دست آوردن μ_iها و مقایسه آن‌ها به انجام محاسبات طولانی نیاز دارد. اما به کمک لم‌هایی که در ادامه خواهند آمد می‌توان کران پایینی برای ϕ به دست آورد که منجر به ساده‌تر شدن محاسبات می‌شود.

فرض کنید $d \in \mathcal{D}(v+1, b, k)$ یک طرح دلخواه باشد. Ω را مجموعه همه $v!$ جایگشت تیمارهای آزمایشی ۱، ...، v در نظر بگیرید. همچنین فرض کنید $\omega \in \Omega$ و ωd طرح نتیجه شده از d به وسیله جایگشت ω از تیمارها در d باشد. حال $\overline{M}_d$ را به صورت

$$\overline{M}_d = \sum_{\omega \in \Omega} M_{\omega d}/v! = \sum_{\pi \in \Pi} \pi' M_d \pi / v! \qquad (۳)$$

تعریف کنید، که در آن Π مجموعه همه ماتریس‌های جایگشت $v \times v$ است.

لم ۱ : اگر $d \in \mathcal{D}(v+1, b, k)$، آن‌گاه $\overline{M}_d$ دارای یک مقدار ویژه برابر

$$m_1(d) = \frac{1}{v} 1_v' M_d 1_v = \frac{1}{v} c_{d,00} \qquad (۴)$$

و $v-1$ مقدار ویژه برابر

$$m_2(d) = \frac{1}{v-1}[tr(M_d) - m_1(d)] = \frac{1}{v-1}[tr(C_d) - (v+1)m_1(d)] \qquad (۵)$$

است.

برهان : با توجه به ماتریس (۱) و چون $C_d\mathbf{1} = \mathbf{0}$، مجموع درایه‌های iـامین سطر یا iـامین ستون ماتریس M_d برابر $-c_{\circ i}$ است. بنابراین با استفاده از (۳) ماتریس $\overline{M}_d$ به صورت

$$\begin{aligned}\overline{M}_d &= \frac{1}{v}\Big[\sum_{i=1}^{v} c_{ii} - \frac{\sum\sum_{1\le i\neq j\le v} c_{ij}}{v-1}\Big]I_v + \frac{1}{v(v-1)}\Big(\sum\sum_{1\le i\neq j\le v} c_{ij}\Big)J_v \\ &= \frac{1}{v}\Big[\frac{v\,tr(M_d) - vm_1(d)}{v-1}\Big]I_v + \Big[\frac{vm_1(d) - tr(M_d)}{v(v-1)}\Big]J_v \\ &= \frac{1}{v-1}[tr(M_d) - m_1(d)]I_v + \frac{1}{v(v-1)}[vm_1(d) - tr(M_d)]J_v\end{aligned}$$

به دست می‌آید. حال با توجه به این‌که $aI_v + bJ_v$ دارای یک مقدار ویژه برابر $a+bv$ و تعداد $v-1$ مقدار ویژه برابر a است اثبات کامل می‌شود.

ماجامدار و نوتز (۱۹۸۳) نشان دادند اگر ϕ یک تابع محدب باشد، آن‌گاه $\phi(\overline{M}_d) \le \phi(M_d)$. همچنین کانرت و مارتین (۲۰۱۰) نشان دادند اگر طرح $d \in \mathcal{D}(v+1, b, k)$ در تیمارهای آزمایش غیردودویی باشد و همهٔ درایه‌های خارج از قطر اصلی ماتریس W نامثبت باشند، آن‌گاه طرح d^* در این کلاس وجود دارد به طوری که d^* در تیمارهای آزمایش دودویی است و $\phi(\overline{M}_{d^*}) \le \phi(\overline{M}_d)$. بنابراین می‌توان ادعا کرد که مقدار $\phi_{E_{tc}}$ حداقل برابر

$$e_d = \max\{\frac{1}{m_1(d)}, \frac{1}{m_2(d)}\} \qquad (6)$$

است و طرحی بهینه است که مقدار e_d را کمینه کند، به طوری که d در تیمارهای آزمایش دودویی است. در واقع باید یک کران پایین برای e_d به دست آورد. بنابر لم ۱ برای حالتی که W متقارن مرکزی است، مقادیر m_1 و m_2 تنها به محل قرار گرفتن تیمار کنترل بستگی دارند.

لم ۲ : اگر تمام درایه‌های غیر قطری ماتریس W نامثبت باشند، آن‌گاه به ازای هر دنباله s که در تیمارهای آزمایش دودویی باشد، $m_1(s) \le m_2(s)$ است.

برهان : طبق لم ۱ برای آن‌که $m_1(s) \le m_2(s)$ باید

$$(v-1)c_{\circ\circ} < vtr(C_s) - (v+1)c_{\circ\circ}$$

یا به طور معادل $2c_{00} < tr(\mathbf{C}_s)$ باشد. این نامساوی در صورتی برقرار است که

$$2\sum_{m=1}^{v-1}\sum_{n=m+1}^{v} c_{mn} < 0 \qquad (7)$$

طبق تعریف ماتریس اطلاع، اگر دو تیمار m و n در دنباله s به ترتیب در موقعیت‌های i و j قرار گرفته باشند، آنگاه $c_{mn} = w_{ij}$ و در غیر این صورت برابر صفر است. چون درایه‌های غیر قطری ماتریس W نامثبت هستند، نامساوی (۷) همیشه برقرار است.

طبق لم ۲ کران e_d را می‌توان به صورت $e_d = \frac{1}{m_1(d)}$ در نظر گرفت. از طرفی برای هر طرح دلخواه $d \in \mathcal{D}(v+1, b, k)$ و هر دنباله تیمار دلخواه، $m_1 \geq 0$ است. بنابراین برای یافتن طرح E_{tc}–بهینه تنها کافی است کلاس تیماری s^* را که به ازای آن مقدار m_1 ماکسیمم می‌شود، یافت. با استفاده از این موضوع در بخش بعد به بررسی ساختار طرح E_{tc}–بهینه تحت فرآیند اتورگرسیو مرتبه اول با پارامتر $\rho > 0$ پرداخته می‌شود.

۴ الگوریتمی برای یافتن طرح بهینه

حال به کمک نتایج بخش ۳ می‌توان با الگوریتم زیر طرح E_{tc}–بهینه را یافت.

• گام اول: به ازای مقدار مشخص k همه کلاس‌های دنباله‌های تیماری ممکن ساخته شوند به طوری که در تیمارهای آزمایش دودویی باشند. (به ازای هر k، تعداد $\sum_{i=1}^{k}\binom{k}{i}$ دنباله تیماری وجود دارد که بعضی از آن‌ها با یکدیگر معادل هستند، یعنی مقادیر m_1 آن‌ها با هم برابر است، دنباله‌های معادل در یک کلاس قرار می‌گیرند).

• گام دوم: مقدار $m_1(s)$ برای هر کلاس محاسبه شود.

• گام سوم: بیشترین مقدار $m_1(s)$ و کلاس دنباله‌های تیماری متناظر با آن با نماد s^* مشخص شود.

به این ترتیب چنانچه بلوک‌های طرح d^* با استفاده از دنباله‌های تیماری متعلق به کلاس s^* ساخته شوند، آن‌گاه طرح d^*، E_{tc}–بهینه است.

به منظور ساختن کلاس دنباله‌های تیماری در گام اول روش‌های متعددی وجود دارند که می‌توانند مورد استفاده قرار گیرند. در این مقاله از شبیه‌سازی دنباله‌ها به کمک نمونه‌گیری پارتو استفاده می‌شود. اساس نمونه‌گیری پارتو به این صورت است که از بین N واحد جامعه، n واحد را به عنوان نمونه انتخاب می‌کند. شانس انتخاب هر واحد جامعه برابر p_i است که باید در شرط $\sum_{i=1}^{N} p_i = n$ صدق کند. مقدار p_i در واقع تعیین کننده احتمال شمول واحد iام است. به عبارت دیگر چنانچه به طور مثال ۱۰۰ مرتبه شبیه‌سازی شود انتظار می‌رود که تعداد $100p_i$ از نمونه‌های انتخاب شده، شامل واحد iام باشند. دلیل استفاده از این روش آن است که آنتروپی یا بی‌نظمی لازم در نمونه‌های انتخاب شده را دارد، بنابراین می‌توان مطمئن بود که با تعداد دفعات شبیه‌سازی به اندازه کافی زیاد، کلاس‌های تیماری متفاوت بسیاری تولید شوند.

بر اساس تعداد تکرار تیمار کنترل در هر کلاس می‌توان کلاس‌های تیماری را به k گروه تقسیم کرد. حال برای ساختن کلاس‌های تیماری، تعداد واحد جامعه برابر با k اختیار کرده و تعداد تکرار تیمار کنترل برابر با حجم نمونه در نظر گرفته می‌شود. بدین ترتیب چنانچه به ازای مقادیر مختلف ممکن از تکرار تیمار کنترل، عمل نمونه‌گیری انجام شود آن‌گاه کلاس‌های متفاوت تیماری به دست می‌آیند. برای انتخاب مقادیر p_i ها باید این نکته را مد نظر قرار داد که اگرچه روش توضیح داده شده مربوط به نمونه‌گیری با احتمال‌های نابرابر است، ولی در این‌جا تیمارها نسبت به یکدیگر ترجیح نداشته و هریک شانس یکسان برای قرار گرفتن در بلوک را دارا هستند. به همین علت، برای ایجاد هر گروه با توجه به تعداد تکرار تیمار کنترل در آن گروه که با r_0 نشان داده می‌شود، می‌توان p_iها را مقادیر نزدیک به r_0/k انتخاب کرد، به طوری که $\sum_{i=1}^{k} p_i = r_0$ باشد.

این الگوریتم در تمامی محیط‌های برنامه‌نویسی قابل اجرا و به ازای همه مقادیر $v \geq k$ و $v \geq 2$ قابل استفاده است. همچنین تنها محدودیتی که برای ساختار همبستگی باید در نظر گرفته شود آن است که درایه‌های غیر قطری آن باید نامثبت باشند.

جدول ۱: دنباله‌های تیماری ممکن برای $k = ۴$

دنباله	نوع دنباله s	$m_۱(s)$
۱	[۰, ۱, ۰, ۲]	۰/۴۶۸۷۵
۲	[۰, ۱, ۲, ۰]	۰/۳۳۳۳۳
۳	[۰, ۱, ۲, ۳]	۰/۲۰۸۳۳
۴	[۱, ۰, ۲, ۱]	۰/۳۰۲۰۸
۵	[۱, ۲, ۳, ۴]	۰

جدول ۲: طرح بهینه برای AR(1)

k	طرح بهینه	k	طرح بهینه
۵	[۱, ۰, ۲, ۰, ۳]	۸	[۰, ۱, ۰, ۲, ۰, ۳, ۰, ۴]
۶	[۰, ۱, ۰, ۲, ۰, ۳]	۹	[۱, ۰, ۲, ۰, ۳, ۰, ۴, ۰, ۵]
۷	[۱, ۰, ۲, ۰, ۳, ۰, ۴]	۱۰	[۰, ۱, ۰, ۲, ۰, ۳, ۰, ۴, ۰, ۵]

به عنوان مثال چنانچه در محیط برنامه نویسی R، $k = ۴$، $v = ۴$ و ساختار همبستگی فرآیند اتورگرسیو مرتبه اول با پارامتر $\rho = ۰/۵$ در نظر گرفته شود آن‌گاه در زمانی کمتر از چند صدم ثانیه خروجی نشان داده شده در جدول ۱ محاسبه می‌شود. واضح است که در این حالت طرح E_{tc}-بهینه با استفاده از کلاس تیماری ۱ ساخته می‌شود.

جدول ۲ طرح‌های بهینه تحت فرایند اتورگرسیو مرتبه اول با پارامتر $۰ \leq \rho \leq ۱$، $v = k$ و $۴ \leq k \leq ۱۰$ را به ازای برخی از مقادیر k نشان می‌دهد. همان‌طور که ملاحظه می‌شود طرح بهینه بر حسب اینکه k زوج یا فرد باشد، به ترتیب با استفاده از دنباله‌های $[CTCT\ldots CT]$ و $[TCTC\ldots CT]$ ساخته می‌شود. این موضوع با استفاده از ویژگی‌های فرایند اتورگرسیو مرتبه اول نیز قابل اثبات است.

۵ بحث و نتیجه‌گیری

اگرچه طرح‌های آزمایش -کنترل بلوکی بهینه با مشاهدات همبسته در سال‌های اخیر مورد توجه برخی آماردانان قرار گرفته است، اما معمولاً طرح‌ها از نظر معیار A_{tc} بهینگی مورد مطالعه قرار گرفته‌اند. از طرفی به دلیل محاسبات سنگین یافتن طرح بهینه، استفاده از این تئوری‌ها در عمل وقتی اندازه بلوک بزرگ‌تر از ۴ باشد، بسیار مشکل است. در حالی که الگوریتم ارایه شده در این مقاله تنها با پیمودن سه گام قادر است برای هر مقدار k طرح بهینه را در کوتاه‌ترین زمان ممکن تعیین کند. همچنین روشی که برای تولید کلاس‌های تیماری در این الگوریتم ارایه شده است را می‌توان برای یافتن طرح بهینه تحت هر معیار بهینگی به کار برد.

تقدیر و تشکر

مؤلفین مقاله از داوران و هیئت تحریریهٔ محترم مجله علوم آماری به خاطر پیشنهادات ارزنده‌ای که موجب بهبود مقاله گردید قدردانی و تشکر می‌کنند.

مراجع

Behhofer, R. E. and Tamhane, A. C. (1981), Incomplete Block Designs for Comparing Treatments with a Control: General Theory, *Technometrics*, **23**, 45-57.

Cheng, C. S, Majumdar, D. and Stufken J. (1988), Optimal Step-Type Designs for Comparing Test Treatments with a Control, *American Statistical Association*, **402**, 477-482.

Cutler, D. R. (1993), Effcient Block Designs for Comparing Test Treatments to a Control When the Errors Are Correlated, *Statistical Planning and Inference*, **36**, 107-125.

Grafstrom A. (2010), Entropy of Unequal Probability Sampling Designs, *Statistical Methodology*, **7**, 84-97.

Hedayat, A. S. and Majumdar, D. (1984), A-Optimal Incomplete Block Designs For Control-Test Treatment Comparisons, *Technometrics*, **26**, 363-370.

Kunert, J., Martin R. J. and Eccleston J. (2010), Optimal Block Designs Comparing Treatments with a Control When the Errors Are Correlated, *Statistical Planning and Inference*, **140**, 2719-2738.

Majumdar D. and Notz, W. I. (1983), Optimal Incomplete Block Designs for Comparing Treatments with a Control, *Annals of Statistics*, **11**, 258-266.

Stufken, J. (1988), On Bounds for the Efficiency of Block Designs for Comparing Test Treatments with a Control, *Statistical Planning and Inference*, **19**, 361-372.

Stufken, J. (1987), A-Optimal Block Designs for Comparing Test Treatments with a Control, *Annals of Statistics*, **15**, 1629-1638.

Woo, Y., Krueger, W., Kaur, A. and Churchill, G. (2005), Experimental Designs for Three-Color and Four-color Gene Expression Microarrays, *Bioinformatics*, **21**, i459-i467.