Разпределение на Уишарт върху неразложими графи

Евелина Велева

Wishart distribution for non-decomposable graphs: *The aim of this paper is to find the marginal densities of the Wishart distribution, corresponding to non-decomposable graphs. Each non-decomposable graph consists of at least one cycle with 4 or more edges. It is shown how by Monte Carlo method for numerical integration a marginal density corresponding to a cyclic graph may be computed at any point.*

Key words: *Wishart distribution, Model, non-decomposable graph, marginal density, covariance matrix, graphical Gaussian models.*

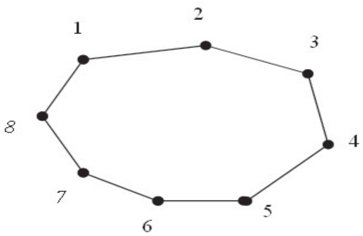
ВЪВЕДЕНИЕ

Разпределението на извадъчната ковариационна матрица за извадка от наблюдения над многомерно нормално разпределение е изведено през 1928 г. от Уишарт ([9]) и е известно под името разпределение на Уишарт. То се явява многомерно обобщение на χ^2 (хи-квадрат) разпределението и присъства в почти всички учебници по многомерен статистически анализ. Разпределението на Уишарт е обект на непрекъснато изучаване и обобщаване от неговото въвеждане през 1928г. до наши дни. Изследването на това разпределение е актуално и към днешна дата, което се вижда с публикуването на все по-нови материали по темата (виж напр. [1], [4], [5]). Една $p \times p$ случайна матрица с разпределение на Уишарт $\mathrm{W}_p(n,\Sigma)$, където $p < n+1$ и Σ е положително определена $p \times p$ матрица, има вероятностна плътност от вида

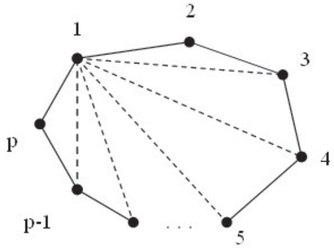
$$f_{p,n,\Sigma}(\mathrm{W}) = \frac{1}{2^{np/2}\Gamma_p\left(n/2\right)(\det\Sigma)^{n/2}}(\det\mathrm{W})^{(n-p-1)/2}e^{-tr(\mathrm{W}\Sigma^{-1})/2} \tag{1}$$

за всяка реална $p \times p$ положително определена матрица W, където $\Gamma_p(\cdot)$ е т. нар. многомерна Гама функция, $\Gamma_p\left(\gamma\right) = \pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma[\gamma+(1-j)/2]$ и $\det(\cdot)$, $tr(\cdot)$ са означения съответно за детерминанта и следа на матрица. Когато наблюдаваните фактори са независими помежду си, както Σ така и нейната обратна матрица Σ^{-1}са диагонални матрици. Ще означаваме елементите на матрицата Σ^{-1} с $\sigma^{i,j}$. Условието $\sigma^{i,j} = 0$ означава, че съответната двойка фактори X_i и X_j са условно независими помежду си, при условие останалите $p-2$ наблюдавани фактора. Наличието на условна независимост между двойки от факторите при условие останалите, лежи в основата на графичните гаусови модели ([3]). В някои научни области, например в генетиката преобладаващата част от двойките фактори са условно независими, при условие останалите. Елементите на извадъчната ковариационна матрица, за които елементите на Σ^{-1} са различни от нула, образуват достатъчно множество от статистики за оценяването на ковариационната матрица Σ ([2]). При моделирането на ковариационни матрици е достатъчно в модела да се включат само тези ковариационни коефициенти, за които $\sigma^{i,j} \neq 0$. Съответната маргинална плътност на разпределението на Уишарт, която ще се използва за оценяване на параметрите в модела се получава като се интегрира плътността (1) по отношение на тези $w_{i,j}$, за които $\sigma^{i,j} = 0$.

В [6], чрез интегриране на плътността (1) са изведени някои от маргиналните плътности на разпределението на Уишарт. Не винаги обаче, границите на интегриране по отношение на променливите, които се изключват от общата съвместна плътност могат да бъдат изразени точно, в явен вид. Нека $\mathbf{W} = (W_{i,j})$ е произволна положително определена случайна матрица от ред p. Нека M е подмножество на множеството $\{W_{i,j}, 1 \le i < j \le p\}$ от извъндиагоналните елементи на $\mathbf{W}$. Маргиналната плътност f_{M^C}, получена след интегриране на плътността на $\mathbf{W}$ по отношение на променливите от M, може да бъде представена чрез граф G_{-M} с множество от върхове $V = \{1, \dots, p\}$. За всеки елемент $W_{i,j}$ от множеството $\{W_{i,j}, 1 \le i < j \le p\} \setminus M$ построяваме между съответните му върхове „i" и „j" ненасочена дъга в графа. Например графът на фигура 1 представя съвместната (маргиналната) плътност на случайните величини от множеството $\{W_{1,2}, W_{2,3}, W_{3,4}, W_{4,5}, W_{5,6}, W_{6,7}, W_{7,8}, W_{1,8}\}$. За граф с $p = 8$ върха, броят на всевъзможните дъги е $C_8^2 = 28$. Множеството M в случая се състои от останалите 20 променливи, по отношение на които е интегрирана плътността на $\mathbf{W}$.



Фиг. 1 Цикличен граф



Фиг. 2 Разложим граф

В [7] е доказано твърдението.

Твърдение 1. *Границите на интегриране на плътността на* $\mathbf{W}$ *по отношение на всичките елементи на* M *могат да бъдат получени в явен вид тогава и само тогава, когато графът* G_{-M} *е разложим.*

Ще приведем някои основни понятия от теория на графите ([3]). Нека $G = (V, E)$ е граф с крайно множество от върхове V и множество от ненасочени дъги E. G се нарича пълен граф ако всяка двойка от върхове на G е съединена с ненасочена дъга в графа. На всяко подмножество $U \subset V$ от върхове съответства подграф G_U на G, състоящ се от върховете в U и дъгите от E, съединяващи върхове от U. Всеки пълен подграф, който е максимален, т.е. не е подграф на никой друг пълен подграф, се нарича клика.

Дефиниция 1. *Графът* G *е разложим тогава и само тогава, когато кликите на* G *могат да бъдат номерирани* C_1, C_2, ..., C_k *така че за всяко* $i = 2, \dots, k$ *ако* $S_i = C_i \bigcap \left(\bigcup_{j=1}^{i-1} C_j \right)$ *то* $S_i \subset C_l$ *за някое* $l < i$.

За разложим граф G_{-M}, при условие че $\sigma^{i,j} = 0$ за всеки елемент $W_{i,j}$ на множеството M, маргиналната плътност f_{M^C} на разпределението на Уишарт има компактен вид ([7]). За произволна матрица A от ред p и произволно подмножество $\alpha \subset \{1, 2, \dots, p\}$ ще означаваме с $\mathrm{A}[\alpha]$ подматрицата на A, образувана от редовете и стълбовете на A с номера от α.

Твърдение 2. *Нека* $\mathbf{W}=(W_{i,j})$ *е случайна матрица с разпределение на Уишарт* $W_p(n,\Sigma)$ *и* M *е подмножество на множеството* $\{W_{i,j}, 1\le i<j\le p\}$ *от извън-диагоналните елементи на* $\mathbf{W}$. *Нека* $\sigma^{i,j}=0$ *за всеки елемент* $W_{i,j}$ *на* M. *Нека графът* G_{-M} е разложим *с* k *клики* C_1, C_2, ..., C_k, *номерирани съгласно Дефиниция 1 и* V_i *е множеството от върховете на* C_i, $i=1,\dots,k$. *Тогава съвместната плътност* f_{M^C} *на елементите от множеството* $M^C=\{W_{i,j}, 1\le i\le j\le p\}\setminus M$ *може да бъде записана във вида*

$$f_{M^C}(W_0)=\frac{\prod_{i=1}^{k} f_{|V_i|,n,\Sigma[V_i]}(W_0[V_i])}{\prod_{i=2}^{k} f_{|U_i|,n,\Sigma[U_i]}(W_0[U_i])}, \tag{2}$$

където $W_0=(w_{i,j})$ *е симетрична* $p\times p$ *матрица, за която* $w_{i,j}=0$ *за* $W_{i,j}\in M$; $f_{p,n,\Sigma}(W)$ *е плътността (1) и* $U_i=(V_1\cup\ldots\cup V_{i-1})\cap V_i$, $i=2,\dots,k$.

Настоящата статия си поставя за цел намирането на маргиналните плътности на разпределението на Уишарт, съответстващи на неразложими графи. Всеки неразложим граф има поне един цикъл (затворен контур, виж фиг. 1), състоящ се от 4 или повече дъги. Показано е как чрез метода Монте Карло може да бъде изчислена в произволна точка маргинална плътност, съответстваща на цикличен граф.

РАЗПРЕДЕЛЕНИЕ НА УИШАРТ ВЪРХУ НЕРАЗЛОЖИМИ ГРАФИ

Нека $\mathbf{W}=(W_{i,j})$ е случайна матрица с разпределение на Уишарт $W_p(n,\Sigma)$. Нека M е подмножество на множеството $\{W_{i,j}, 1\le i<j\le p\}$ от извъндиагоналните елементи на $\mathbf{W}$ такова, че графът G_{-M} е неразложим. Това, съгласно Твърдение 1 означава, че границите на интегриране по отношение на елементите на M не могат да бъдат записани в явен вид. Маргиналната плътност f_{M^C}, получена от плътността (1) след интегриране по отношение на елементите на M следователно също не може да бъде изразена с точна формула. Тя може да бъде пресметната числено за всеки набор от конкретни стойности за случайните величини от множеството $M^C=\{W_{i,j}, 1\le i\le j\le p\}\setminus M$. Този набор от стойности може да бъде записан във вид на матрица W_0, получена от положително определена матрица W след заместване с нули на елементите на W, съответстващи на елементите на множеството M.

Ако от един пълен граф с p върха бъде отстранена дъга, в резултат се получава разложим граф. Следователно всеки неразложим граф се получава чрез отстраняване на дъги в някакъв непълен разложим граф. Затова, с цел намаляване на кратността на интеграла, който трябва да бъде числено пресметнат за определянето на f_{M^C}, интегрирането трябва да започне не директно върху плътността (1), а от някоя маргинална плътност, съответстваща на разложим граф. Такива плътности са разгледани в [6] и [7]. Твърдение 2 дава всички маргинални плътности, съответстващи на разложими графи, при условие че $\sigma^{i,j}=0$ за всяка дъга (i, j), не включена в графа.

Нека M_1 е подмножество на M, такова че графът G_{-M_1} е разложим. Плътността f_{M^C} може да бъде получена чрез числено интегриране на плътността $f_{M_1^C}$ по отношение на променливите от множеството $L = M \setminus M_1$, т.е.

$$f_{M^C}(\mathrm{W}_0) = \int_D f_{M_1^C}(\mathrm{W}_0, \mathrm{x})\, d\mathrm{x} \tag{3}$$

Интегрирането в (3) се извършва върху областта D на дефиниране на плътността $f_{M_1^C}$. Когато кратността на интеграла, който трябва да бъде пресметнат числено е по-голяма или равна на 8, или когато границите на интегриране не могат да се определят в явен вид, в литературата се препоръчва използването на метода Монте Карло. Интегралът (3) се записва във вида

$$f_{M^C}(\mathrm{W}_0) = \int_D \left(\frac{f_{M_1^C}(\mathrm{W}_0, \mathrm{x})}{g(\mathrm{W}_0, \mathrm{x})} \right) g(\mathrm{W}_0, \mathrm{x})\, d\mathrm{x}\,,$$

$g(\mathrm{W}_0, \mathrm{x})$ е вероятностна плътност върху областта D. По този начин $f_{M^C}(\mathrm{W}_0)$ се представя като математическото очакване на величината

$$\frac{f_{M_1^C}(\mathrm{W}_0, \xi)}{g(\mathrm{W}_0, \xi)}, \tag{4}$$

където ξ е случайна величина с плътност $g(\mathrm{W}_0, \mathrm{x})$. Така, изчисляването на $f_{M^C}(\mathrm{W}_0)$ се свежда до намирането на средната аритметична на голям брой реализации на величината (4). Идеалният избор за $g(\mathrm{W}_0, \mathrm{x})$ (който обаче не е реализуем поради непознаването на стойността на търсения интеграл) е ([8])

$$g^*(\mathrm{W}_0, \mathrm{x}) = \frac{f_{M_1^C}(\mathrm{W}_0, \mathrm{x})}{f_{M^C}(\mathrm{W}_0)} = \frac{f_{M^C \cup L}(\mathrm{W}_0, \mathrm{x})}{f_{M^C}(\mathrm{W}_0)}\,. \tag{5}$$

Дясната страна на (5) е равна на условната плътност $f_{L/M^C}(\mathrm{W}_0, \mathrm{x})$ на величините от множеството L при условие величините от множеството M^C. Нека множеството M_2 е такова, че $M \subset M_2 \subset \{W_{i,j}, 1 \le i < j \le p\}$ и графът G_{-M_2} е разложим. Това означава, че графът G_{-M_2} се получава от G_{-M} след отстраняване на дъги, така че да се разкъсат всички затворени контури с дължина по-голяма от 3. Изборът на M_2 следователно не е еднозначен. Нека множеството M_2 освен това е такова, че плътността $f_{M_2^C \cup L}$ се представя също с разложим граф. Тогава вместо идеалният избор f_{L/M^C}, за плътността g може да се избере условната плътност f_{L/M_2^C},

$$g(\mathrm{W}_0, \mathrm{x}) = f_{L/M_2^C}(\mathrm{W}_0, \mathrm{x}) = \frac{f_{M_2^C \cup L}(\mathrm{W}_0, \mathrm{x})}{f_{M_2^C}(\mathrm{W}_0)}\,. \tag{6}$$

Понеже и двете плътности в дясната страна на (6) съответстват на разложими графи, за тяхното определяне ще бъде използвано Твърдение 2.

Ще илюстрираме този подход за пресмятане на съвместната плътност на случайните величини от множеството $M^C = \{W_{1,2}, W_{2,3}, W_{3,4}, \ldots, W_{p-1,p}, W_{1,p}, W_{1,1}, \ldots, W_{p,p}\}$, които са представени с непрекъсната линия посредством граф на фигура 2. За избор на множеството M_1^C, към M^C трябва да бъдат добавени още променливи, така че съответстващия граф G_{-M_1} да бъде разложим. Това може да бъде направено с включването на променливите от множеството $L = \{W_{1,3}, W_{1,4}, W_{1,5}, \ldots, W_{1,p-1}\}$, които са представени с пунктирана линия на фигура 2. Полученият по този начин граф G_{-M_1}, включващ всички дъги (пунктирани и непрекъснати) на фигура 2 е разложим с $p-2$ клики, съответстващи на триъгълниците с върхове съответно $C_1 \to (1,2,3)$, $C_2 \to (1,3,4)$, и т.н., $C_{p-2} \to (1, p-1, p)$. При предположението че $\sigma^{i,j} = 0$ за $W_{i,j} \in M$, плътността $f_{M_1^C}$ има вида (2). За изчисляването на f_{M^C} интегрирането ще се извърши върху $f_{M_1^C}$, по отношение на променливите от множеството L, които са $p-3$ на брой, а не върху общата съвместна плътност (1) по отношение на всички променливи от множеството M, чиито брой е $p(p-3)/2$. За определяне на вероятностната функция g, съгласно (6) трябва да се избере множество M_2, такова че $M \subset M_2$ и графите, съответстващи на маргиналните плътности $f_{M_2^C \cup L}$ и $f_{M_2^C}$ да са разложими. В случая тези условия ще бъдат изпълнени, ако $M_2^C = M^C \setminus \{W_{1,p}\}$.

ЗАКЛЮЧЕНИЕ

Изчисляването на маргиналните плътности на разпределението на Уишарт, съответстващи на неразложими графи, може да бъде извършено единствено чрез числен алгоритъм, базирайки се на получените в [6] и [7] резултати за маргиналните плътности, съответстващи на разложими графи.

ЛИТЕРАТУРА

[1] S. A. Andersson, T. Klein. On Riesz and Wishart distributions associated with decomposable undirected graphs. Journal of Multivariate Analysis, Volume 101, Issue 4, April 2010, 789-810.

[2] T. W. Anderson, An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, N. Y., 3ed., 2003.

[3] S. L. Lauritzen. Graphical Models. Clarendon Press, Oxford, 1996.

[4] G. Letac, H. Massam. Wishart distributions for decomposable graphs. Ann. Statist. 35 (2007), 1278-1323.

[5] H.F. Strydom, N.A.S. Crowther. Maximum likelihood estimation of parameter structures for the Wishart distribution using constraints. J Stat Plan Infer, 143 (4), 2013, 783–794.

[6] E. Veleva. Some marginal densities of the Wishart distribution. Math. and Education in Math., 41 (2012), 273-278.

[7] E. Veleva. Marginal densities of the Wishart distribution. Pliska Stud. Math. Bulgar., 2013, No 22, pp. 225-236.

[8] St. Weinzierl. Introduction to Monte Carlo methods. Eprint arXiv: hep-ph/0006269 (2000).

[9] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. Biometrika, 20 (1928), 32–52.

За контакти:

гл.ас.д-р Евелина Велева, Катедра “Приложна математика и статистика”, Русенски университет “Ангел Кънчев”, e-mail: eveleva@uni-ruse.bg

Докладът е рецензиран.