

معدل وزنی تصادفی و برخی از خواص آن

هژیر حومئی^۱، ریاب سلیم پوراردی^۲، منیره حامل دربندی^۳

چکیده:

مفهوم معدل وزنی تصادفی برای دو متغیر تصادفی تعریف شده و سپس آن را به n متغیر تصادفی تعمیم داده ایم. نحوه محاسبه توزیع و مشخصه سازی برای این مدل پیچیده توجه زیادی را به خود جلب کرده است. در این مقاله با تکنیک قوی و جدیدی به نام تبدیل اشتبلیس توزیع این مدل ها را بدست آورده و مشخصه سازی می کنیم. **واژه‌های کلیدی:** معدل وزنی تصادفی، تبدیل اشتبلیس، توزیع آرک سینوس، توزیع کشی.

۱ مقدمه

به دنبال کاروان اشو (۱۹۸۷)، خانواده ای جدید از توزیع ها با نام معدل وزنی تصادفی توسط جانسن و کاتز (۱۹۹۰) معرفی شد $U, Z = UX_1 + (1 - U)X_2$ دارای توزیع یکنواخت و مستقل از X_1, X_2 است (آن ها ضمن اشاره به کاربرد فراوان این خانواده از توزیع ها، در مقابل کاروان اشو، نتایج وان اشورا به روشی ساده مورد مطالعه قرار دادند. جانسن و کاتز با شروع یک بحث ناتمام در پایان مقاله خود این خانواده را به شکل $\sum W_i X_i$ تعمیم دادند، (X_i) ها متغیرهای تصادفی و W_i ها متغیرهای تصادفی دارای توزیع دیریکله (اما در نوشته های آن ها درباره تصدیق و تعمیم نتایج وان اشو مطلبی ارائه نگردیده است. سلطانی و حومئی (۲۰۰۹) کار آن ها را دنبال کردند و برقراری نتایج وان اشو و همینطور تعمیم آن را بررسی نمودند.

تا کنون مقالات بسیاری درباره متغیرهای تصادفی آمیخته در مجلات معتبر منتشر شده است. جامعه شناسی، بیولوژی و به طور کلی توزیع هایی که در علوم مختلف استفاده می شود به شکل مدل آمیخته تعریف شده اند و معمولاً در احتمال کل، قضیه بیز و احتمال پیشین نیز کاربرد دارند. بررسی توزیع های این نوع مدل از اواخر قرن نوزدهم آغاز شد سپس با پیشرفت نظریه توزیع ها، انواع تعمیم یافته این توزیع ها مورد توجه آماردانان قرار گرفت. در چند دهه اخیر مفهوم تازه ای بنام معدل وزنی تصادفی مورد توجه زیادی قرار گرفته است. در این مقاله درباره این خانواده جدید از توزیع ها بحث می کنیم. البته یکی از اهداف این مقاله، گسترش و شناساندن مفهوم معدل وزنی تصادفی و ارائه برخی از خواص آن می باشد.

^۱گروه آمار دانشکده ریاضی، دانشگاه تبریز
^۲گروه آمار دانشکده علوم، دانشگاه صنعتی خواجه نصیرالدین طوسی
^۳گروه آمار دانشکده ریاضی، دانشگاه تبریز

الف) Z^* دارای توزیع یکنواخت روی $[1, -1]$ می باشد اگر و تنها اگر X_1 و X_2 دارای توزیع آرک سینوس^۵ باشند.

ب) Z^* همان توزیع X_1 و X_2 را دارد اگر و تنها اگر X_1 و X_2 تباهیده یا دارای توزیع کشی^۶ باشند. با استفاده از این تعریف، جانسن و کاتز ثابت کردند که توزیع Z^* با توزیع معدل وزنی تصادفی زیر هم توزیع است:

$$Z = WX_1 + (1 - W)X_2$$

که در آن W و X_1 و X_2 متغیرهای تصادفی مستقل و W دارای توزیع یکنواخت $[0, 1]$ می باشد. بنابراین اولین تعریف رسمی معدل وزنی تصادفی به وسیله آن ها ارائه شد.

۲.۲ معدل وزنی تصادفی

به طور مشابه با استفاده از نکته ای که ذکر شد سلطانی و حومئی، کار جانسن و کاتز را دنبال کردند و معدل وزنی تصادفی را روی n متغیر تصادفی تعریف کردند.

تعریف ۲ فرض کنید $X_1, X_2, \dots, X_n$ متغیرهای تصادفی مستقل و پیوسته باشند و آماره های ترتیبی $U_{(1)}, \dots, U_{(n)}$ از توزیع یکنواخت روی $[0, 1]$ باشند و $U_0 = 0$ و $W_i = U_{(i)} - U_{(i-1)}, i = 1, \dots, n-1$

واضح است به دلیل پیچیدگی مدل ممکن است محاسبه توزیع با مشکل مواجه شود، اما می توان در مشخص سازی توزیع ها، که نقش مهمی در شناسایی غیرمستقیم و الگوسازی آن دارند، از تبدیل اشتیلیس استفاده نمود. در این مبحث، تبدیل اشتیلیس نقش بسیار اساسی در محاسبه توزیع معدل وزنی تصادفی دارد.

در این مقاله ابتدا در بخش دوم معدل وزنی تصادفی را تعریف کرده و به بیان برخی از تعاریف اولیه در این زمینه می پردازیم. در بخش سوم توزیع شرطی حاصل از این مدل آمیخته را معرفی می کنیم سپس نحوه محاسبه توزیع را که نقش ویژه در شاخه های مختلف آمار دارند را در بخش چهارم مورد بررسی قرار می دهیم و بالاخره، در بخش پنجم کاربردهایی از این مدل را بیان می کنیم.

۲ مفاهیم و تعاریف اولیه

۱.۲ توزیع یکنواخت بین دو متغیر تصادفی

تعریف ۱ متغیر تصادفی Z^* دارای توزیع یکنواخت بین دو متغیر تصادفی مستقل X_1 و X_2 است اگر:

$$P(Z^* \leq z | x_1, x_2) = \begin{cases} 0 & z \leq \min(x_1, x_2) \\ \frac{z-x_1}{x_2-x_1} & x_1 \leq z < x_2 \\ \frac{z-x_2}{x_1-x_2} & x_2 \leq z < x_1 \\ 1 & z \geq \max(x_1, x_2) \end{cases}$$

وان اشو^۴ (۱۹۸۷) مفهوم متغیر تصادفی Z^* را مطرح نموده و توزیع Z^* را در حالت های خاص مطالعه نمود و نتایج زیر را بدست آورد.

^۴ Van Assche
^۵ Arcsin Distribution
^۶ Cauchy Distribution

و $W_n = 1 - \sum_{i=1}^{n-1} W_i$ باشد.

متغیر تصادفی Z_n را در نظر می گیریم

$$Z_n = W_1 X_1 + W_2 X_2 + \dots + W_{n-1} X_{n-1} + W_n X_n$$

آنگاه Z_n را معدل وزنی تصادفی می گویند.

تعمیم معدل وزنی تصادفی اخیراً توسط سلطانی و حومئی ارائه گردیده است. کار آن ها علاوه بر آنکه کاروان اشو را در بر می گیرد درستی تعمیم نتایج به دست آمده توسط وان اشورا نیز نشان می دهد. نتایج حاصل از کار آن ها به شرح زیر می باشد:

الف) Z_n و $(X_1, \dots, X_n)$ توزیع یکسان دارند اگر و تنها اگر $X_1, X_2, \dots, X_n$ تباهیده یا دارای توزیع کشی باشند.
ب) نتیجه آرک سینوس وان اشو تنها برای $n = 2$ درست می باشد.

که در آن

$$C(x_{n-j}; x_1, \dots, x_n) = \prod_{k=1}^{n-j-1} (x_k - x_{n-j}) \prod_{k=n-j+1}^n (x_k - x_{n-j})$$

به ازای $j = 0, \dots, n-1$. با استفاده از تابع نشانگر $U(x)$ که برای $x < 0$ برابر صفر و برای $x \geq 0$ برابر یک می باشد، توزیع شرطی Z_n به ازای $z \in [\min\{x_1, \dots, x_n\}, \max\{x_1, \dots, x_n\}]$ برابر با (۱)، به ازای $z < \min\{x_1, \dots, x_n\}$ برابر صفر و به ازای $z > \max\{x_1, \dots, x_n\}$ برابر یک می باشد.

$$\sum_{i=0}^{n-1} \frac{(z - x_{n-i})^{n-1} u(z - x_{n-i})}{C(x_{n-i}; x_1, \dots, x_n)} \quad (1)$$

بنابراین توزیع شرطی Z_n برای مقادیر داده شده $X_1 = x_1, \dots, X_n = x_n$ در (۱)، به دست آمده است. توزیع $K(z)$ در حالت خاص $n = 3$ ، توزیع مثلثی را نتیجه می دهد بنابراین این توزیع را مدل تعمیم یافته توزیع مثلثی می نامیم.

۴ توزیع

محاسبه توزیع معدل وزنی تصادفی با روش های کلاسیک شناخته شده بسیار پیچیده و دشوار است. در این بخش در حالت خاص $n = 2$ ، مثال های متعددی را بررسی می نمائیم. برای محاسبه توزیع Z_n می توان از روش های زیر استفاده کرد.

الف) روش تابع توزیع: در این روش می توان با استفاده از امید ریاضی مکرر توزیع Z_n را محاسبه نمود.

$$F_{Z_n}(z) = E_W(F_{Z_n}(z | w)) \quad (2)$$

۳ توزیع شرطی

توزیع شرطی یک بحث کلیدی در به دست آوردن توزیع معدل وزنی تصادفی به شمار می آید، به همین دلیل در این بخش توزیع شرطی معدل وزنی را نشان می دهیم. توزیع شرطی Z_n به ازای $x_1, \dots, x_n$ در z ، برای $x_1 > x_2 > \dots > x_n$ و $x_{n-i} < z \leq x_{n-i-1}, i = 0, 1, \dots, n-2$ به صورت زیر محاسبه شده است.

$$\begin{cases} K(z | x_1, x_2, \dots, x_n) = \sum_{j=0}^i \frac{(z - x_{n-j})^{n-1}}{C(x_{n-i}; x_1, \dots, x_n)} \\ x_{n-i} < z \leq x_{n-i-1}, i = 0, 1, \dots, n-2. \end{cases}$$

مثال ۱ فرض کنید که W, X_2, X_1 متغیرهای

تصادفی مستقل و دارای توزیع یکنواخت روی فاصله $[0, 1]$ باشند. واضح است معدل وزنی آن $Z = WX_1 + (1 - W)X_2$ بوده و توزیع شرطی Z

بصورت:

$$\begin{aligned} F_Z(z | w) &= \int P(Z \leq z | X_1 = x_1) f(x_1) dx_1 \\ &= \int P(X_2 \leq \frac{z - wx_1}{1 - w}) f(x_1) dx_1 \end{aligned}$$

بنابراین با استفاده از (۲)، به آسانی نتیجه می شود که

$$f_Z(z) = \begin{cases} -2(1-z)\ln(1-z) - 2z\ln z & 0 < z < 1 \\ 0 & \text{o.w} \end{cases}$$

مثال ۲ اگر X_2, X_1 دارای توزیع نرمال استاندارد

باشند، آنگاه توزیع شرطی Z به ازای $W = w$ دارای توزیع نرمال با میانگین صفر و واریانس $W^2 + (1 - W)^2$ می باشد، در نتیجه

$$P[Z \leq z | W] = \Phi\left(\frac{z}{\{w^2 + (1 - w)^2\}^{1/2}}\right)$$

که در آن

$$\Phi(y) = (\sqrt{2\pi})^{-1} \int_{-\infty}^y e^{-\frac{u^2}{2}} du$$

و

$$P[Z \leq z] = E_W(P[Z \leq z | W])$$

این نتیجه برای هر توزیع W درست می باشد. اگر W دارای توزیع یکنواخت $[0, 1]$ باشد، آنگاه

$$P[Z \leq z] = \int_0^1 \Phi\left\{\frac{z}{\{w^2 + (1 - w)^2\}^{1/2}}\right\} dw$$

Stieltjes Transform^۷

(ب) روش تغییر متغیر: با استفاده از این روش نیز می توان توزیع معدل وزنی تصادفی را به دست آورد. برای $n = 2$ به مثال زیر توجه کنید.

مثال ۳ اگر W, X_2, X_1 متغیرهای تصادفی مستقل باشند و X_1 دارای توزیع یکنواخت $[-1, 1]$ و X_2 دارای توزیع آرک سینوس روی $[-1, 1]$ و W دارای توزیع $Beta(2, 1)$ باشد، آنگاه با تعریف تغییر متغیرهای زیر

$$X_1 = Z_1, X_2 = Z_2, W = \frac{Z - Z_1}{Z_1 - Z_2}$$

داریم

$$f(z_1, z_2, z) = \frac{y - z_2}{\pi(z_1 - z_2)^2 \sqrt{1 - z_2^2}}$$

بنابراین چگالی معدل وزنی به صورت زیر می باشد.

$$f(z) = \frac{2\sqrt{1 - z^2}}{\pi}.$$

(ج) روش تبدیل اشتیلیس^۷: مثال های گفته شده در حالت خاص $n = 2$ بررسی گردیده است، در صورتی که $n \geq 3$ باشد مدل Z_n پیچیده می شود و محاسبه توزیع با استفاده از روش های شناخته شده امکان پذیر نبوده و یا محاسبات پیچیده ای دارد. اکنون به روش دیگری که توسط سلطانی و حومئی ارائه گردیده است اشاره می کنیم. قضیه زیر با استفاده از تکنیک تبدیل اشتیلیس، توزیع Z_n را مشخص می کند و به دلیل طولانی بودن محاسبات، آن را بدون اثبات بیان می کنیم.

قضیه ۱ تحت فرض اینکه Z_n معدل وزنی تصادفی و $X_1, X_2, \dots, X_n$ متغیرهای تصادفی مستقل و پیوسته هستند، داریم

$$\frac{(-1)^{n-1}}{(n-1)!} \frac{d^{n-1}}{dz^{n-1}} S(F_Z, z) = \prod_{i=1}^n S(F_{X_i}, z),$$

که در آن $z \in C \cap_{i=1}^n (\text{supp} F_{X_i})^c$ و $S(F, z)$ تبدیل اشتیلیس تابع $f(x)$ می باشد که به صورت زیر تعریف می شود.

$$S(f, z) = \int_R \frac{f(x)}{z - x} dx, \quad z \in C$$

و C مجموعه اعداد مختلط می باشد.

روش بیان شده در قضیه ۱ در مقایسه با روش های دیگر برتری خاصی را نشان می دهد، زیرا:

الف) محاسبه را ساده تر می کند.

ب) در صورت تعیین توزیع Z_n می توان توزیع $-X_i$ ها را نیز به دست آورد.

ابتدا مثالی برای محاسبه توزیع Z_n با استفاده از قضیه ۱ مطرح می کنیم.

مثال ۴ اگر X_1, X_2, X_3 متغیرهای تصادفی $i.i.d$ دارای توزیع آرک سینوس باشند، آنگاه با استفاده از قضیه ۱ داریم

$$\begin{aligned} \frac{1}{3} \frac{d^3}{dz^3} S(F_Z, z) &= \prod_{i=1}^3 S(F_{X_i}, z) \\ &= \frac{1}{(z^2 - 1)^{\frac{3}{2}}} \end{aligned}$$

بنابراین $S(F_Z, z) = \sqrt{z^2 - 1}$ ، که این تبدیل اشتیلیس توزیع نیم دایره A روی $[-1, 1]$ می باشد. حال مثال هایی را دنبال می کنیم که در آن توزیع Z_n مشخص شده است و با استفاده از قضیه ۱ توزیع X_i را محاسبه می کنیم.

مثال ۵ اگر X_1, X_2 متغیرهای تصادفی $i.i.d$ روی $[-1, 1]$ باشند با فرض اینکه Z دارای توزیع یکنواخت روی $[-1, 1]$ باشد، با استفاده از قضیه ۱ داریم

$$\begin{aligned} (S(F_{X_1}, z))^2 &= -S'(F_Z, z) \\ &= \frac{1}{z^2 - 1} \end{aligned}$$

بنابراین $S(F_{X_1}, z) = \frac{1}{\sqrt{z^2 - 1}}$ ، که این تبدیل اشتیلیس توزیع آرک سینوس روی $[-1, 1]$ می باشد.

مثال ۶ اگر X_1, X_2, X_3 متغیرهای تصادفی $i.i.d$ باشند با فرض اینکه Z معدل وزنی X_i ها و دارای توزیع کشی باشد، با استفاده از قضیه ۱ داریم

$$(S(F_{X_1}, z))^3 = \frac{1}{3} S''(F_Z, z)$$

آن به آسانی نتیجه می شود که

$$S(F_X, z) = \frac{1}{z - a + ib}, \quad \text{Im}(z) > 0, \quad b \neq 0$$

و این تبدیل اشتیلیس توزیع کشی می باشد.

۵ نتایج

که در آن X مستقل از W می باشد. هنگامی که X ها دارای توزیع یکنواخت $[0, 1]$ باشند به خانواده ای جالب از توزیع های متقارن منجر می شود و در ادامه این روش می تواند به توزیع متغیرهایی از این نوع،

$$Y = \sum_{j=0}^{\infty} (-1)^j \prod_{i=0}^j X_i$$

بسط داده شود که کاربرد فراوانی در فیزیک، در حاصلضرب ماتریس های تصادفی 2×2 دارد.

بردارهای تصادفی $X = (X_1, X_2, \dots, X_n)$ و $W = (W_1, W_2, \dots, W_n)$ را در نظر می گیریم که در آن $X_1, X_2, \dots, X_n$ متغیرهای تصادفی مستقل و پیوسته و $-W_i$ ها دارای توزیع دیریکله هستند، معدل وزنی تصادفی زیر را تعریف می کنیم

$$Z_n = W_1 X_1 + W_2 X_2 + \dots + W_{n-1} X_{n-1} + W_n X_n;$$

مراجع

- [1] Homei, H. (2010), Weighted averages with random proportion, 10-Th Iranian Statistical Conference University of Tabriz.
- [2] Johnson, N.L. and Kotz, S. (1990), Randomly weighted averages: Some aspects and extensions, American Statistician, 44, 245-249.
- [3] Rudin, W. (1987), Real and Complex Analysis, McGraw and Hill.
- [4] Rudin, W. (1991), Functional Analysis, McGraw and Hill.
- [5] Soltani, A.R. and Homei, H. (2009), Weighted averages with random proportions that are jointly uniformly distributed over the unit simplex, Statistics and Probability Letters, 9, 1215-1218.
- [6] Van Assche, W. (1987), A random variable uniformly distributed between two independent random variables, Sankhya, Ser. A, 49, 207-211.
- [7] Zayed, A.I. (1996), Handbook of Function and Generalized Function Transformations, CRC Press, London.