

سه اثبات استقرایی ساده در آمار و احتمال

دوریان فلدمن مارتین فوکس

ترجمه: مریم شریف دوست^۱

چکیده

در این مقاله، اثبات‌های استقرایی مقدماتی، برای تقریب توزیع دوجمله‌ای به فوق هندسی نمایش داده شده، همچنین چگالی یک آماره‌ی ترتیبی و توزیع $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ وقتی که X_1, \dots, X_n یک نمونه از $N(\mu, 1)$ است، ارائه شده است.

۱ مقدمه

سه اثبات استقرایی خیلی ساده که محاسبات برخی نتایج آمار و احتمال را آسان می‌سازد، نمایش می‌دهیم. این نتایج عبارتند از:

الف - تقریب توزیع دوجمله‌ای به فوق هندسی

ب - چگالی k امین آماره‌ی ترتیبی برای یک نمونه از توزیع پیوسته

ج - توزیع $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ وقتی که X_1, \dots, X_n یک

^۱مریم شریف دوست، گروه آمار دانشگاه شیراز

نمونه از $N(\mu, 1)$ است و $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$.

روش استفاده شده در قسمت (ب) می‌تواند برای بدست آوردن چگالی توأم هر k آماره‌ی ترتیبی نیز به‌کار گرفته شود. استفاده از روش استقرا، تا آنجا که اطلاع داریم تنها در مورد مسایل مشابه، برای بدست آوردن تقریب توزیع دوجمله‌ای به پواسون به‌کار گرفته شده است ([۱]، صفحه ۱۵۳).

۲ تقریب توزیع دوجمله ای به فوق هندسی

فرض کنید یک جامعه شامل N عضو است که M عضو آن موفقیت هستند. برای یک نمونه به حجم n از این جامعه، تابع احتمال فوق هندسی مربوط را با $h(k|n, N, M)$ نمایش می دهیم (در صورتی که k تعداد موفقیت در نمونه باشد).

برای n تکرار مستقل از یک آزمایش که احتمال موفقیت در هر تکرار، p باشد، تابع احتمال دوجمله ای مربوط را با $b(k|n, p)$ نمایش می دهیم.

فرض کنید X_1, X_2, \dots متغیرهای تصادفی نشانگر با مقادیر ممکن ۰ و ۱، به صورت زیر باشند:

$$X_i = \begin{cases} 1 & \text{اگر } i \text{ امین عضو نمونه، یک موفقیت باشد} \\ 0 & \text{در غیر این صورت} \end{cases}$$

بنابراین، با فرض $S_n = X_1 + X_2 + \dots + X_n$ داریم:

$$\begin{aligned} P(S_{n+1} = k) &= P(S_n = k)P(X_{n+1} = 0 | S_n = k) \\ &+ P(S_n = k - 1)P(X_{n+1} = 1 | S_n = k - 1) \end{aligned} \quad (1)$$

در هر دو مدل فوق هندسی و دوجمله ای، S_n تعداد موفقیتها در n تکرار است. در توزیع فوق هندسی، $P(X_{n+1} = 1 | S_n = k) = \frac{M-k}{N-k}$ و $P(X_{n+1} = 0) = p$ مستقل هستند و X_1, X_2, \dots, X_{n+1} بنابراین (۱) یک رابطه بازگشتی برای $h(k|n+1, N, M)$ و $b(k|n+1, p)$ خواهد بود.

قضیه ۱ - اگر $\lim_{N \rightarrow \infty} \frac{M}{N} = p$ آنگاه $\lim_{N \rightarrow \infty} h(k|n, N, M) = b(k|n, p)$

اثبات - ابتدا به راحتی می بینیم که اگر $N \rightarrow \infty$ ، آنگاه $h(k|1, N, M) \rightarrow b(k|1, p)$ حال فرض می کنیم که قضیه

برای $n = m$ درست است و از (۱) استفاده کرده و درستی آن را برای $n = m + 1$ اثبات می کنیم. این اثبات می تواند در مورد تقریب توزیع چند جمله ای به فوق هندسی چند متغیره نیز بسط داده شود.

۳ چگالی آماره ی ترتیبی

فرض کنید X_1, \dots, X_n یک نمونه از توزیع به طور مطلق پیوسته F با چگالی f باشد و همچنین فرض کنید که $Y_1 < Y_2 < \dots < Y_n$ آماره های ترتیبی مربوط به X_i ها باشند.

قضیه ۲ - چگالی Y_k به صورت زیر است:

$$\begin{aligned} g_k(y) &= \frac{n!}{(k-1)!(n-k)!} f(y) [F(y)]^{k-1} [1-F(y)]^k \end{aligned} \quad (2)$$

اثبات - ابتدا (۲) را برای $k = 1$ به وسیله روش معمولی به دست می آوریم. حال فرض کنید که G_k تابع توزیع Y_k باشد و همچنین فرض کنید که (۲) برای $k = l$ درست باشد، بنابراین:

$$\begin{aligned} G_l(y) - G_{l+1}(y) &= P(Y_l \leq y < Y_{l+1}) \\ &= \binom{n}{l} [F(y)]^l [1-F(y)]^{n-l} \end{aligned} \quad (3)$$

چون $Y_l < y < Y_{l+1}$ ، پیشامد آن است که فقط l تا از X_i ها، کمترین مساوی y باشند؛ از (۳) مشتق می گیریم.

$$\begin{aligned} g_l(y) - g_{l+1}(y) &= \frac{n!}{(l-1)!(n-l)!} f(y) [F(y)]^{l-1} [1-F(y)]^{n-l} \\ &- \frac{n!}{l!(n-l-1)!} f(y) [F(y)]^l [1-F(y)]^{n-l-1} \end{aligned} \quad (4)$$

اما جمله ی تفریق شده از طرف راست (۴)، $g_{l+1}(y)$ با توجه به فرض استقرا است.

این روش می تواند برای مشتق گیری از چگالی های توأم آماره های ترتیبی نیز بسط داده شود.

اگر برای مثال، چگالی توأم Y_k و Y_l ($l < k$) را بخواهیم، با چگالی توأم Y_{n-1} و Y_n شروع می کنیم. سپس فرض می کنیم که نتایج برای $l \geq z$ و $k \geq m$ ، با حداقل یک نامساوی اکید درست باشد و برای $z = l$ و $k = m$ ثابت می کنیم.

اثبات - ابتدا برای $n = 1$ ثابت می کنیم:

$$\begin{aligned} T_2 &= (X_1 - \bar{X}_2)^2 + (X_2 - \bar{X}_2)^2 \\ &= \left(\frac{X_1 - X_2}{\sqrt{2}} \right)^2 \sim \chi_{(1)}^2 \end{aligned}$$

حال فرض می کنیم که قضیه برای $n = m$ درست است، پس

$$T_{m+1} = T_m + \left[\left(\frac{m}{m+1} \right)^{\frac{1}{2}} (X_{m+1} - \bar{X}_m) \right]^2 \quad (5)$$

T_m و \bar{X}_m و X_{m+1} از هم مستقل هستند، پس دو جمله طرف راست (5) مستقل هستند، علاوه بر آن

$$\left(\frac{m}{m+1} \right)^{\frac{1}{2}} (X_{m+1} - \bar{X}_m) \sim N(0, 1)$$

و تحت فرض استقرا:

$$T_m \sim \chi_{(m-1)}^2$$

پس

$$T_{m+1} \sim \chi_m^2$$

۴ توزیع $\sum_{i=1}^n (X_i - \bar{X}_n)^2$

T_n را به صورت $T_n = \sum_{i=1}^n (X_i - \bar{X}_n)^2$ تعریف می کنیم. فرض می کنیم که استقلال \bar{X}_n و T_n را برای نمونه های توزیع نرمال می دانیم. (اگر مفهوم نرمال چندبعدی مطرح باشد، این استقلال به راحتی با مشاهده $Cov(\bar{X}_n, X_i - \bar{X}_n) = 0$ ثابت می شود). همچنین فرض می کنیم که اگر Z_1, \dots, Z_n مستقل و هم توزیع با $N(0, 1)$ باشند، بنابراین $\sum_{i=1}^n Z_i^2 \sim \chi_{(n)}^2$.

قضیه ۳ - فرض کنیم X_1, \dots, X_n یک نمونه از

$$N(\mu, 1)$$

مراجع

- [1] Feller, William, *An Introduction to Probability Theory and Its Applications (Vol. I, 3rd ed.)*, New York: John Wiley & Sons, (1968).

اصل این مقاله با عنوان

Three Simple Inductive Proofs in Probability and Statistics

نوشته Dorian Feldmand و Martin Fox است که در

The American Statistician, Vol. 34, No.1 (1980)

چاپ شده است.