# سه اثبات استقرایی ساده در آمار و احتمال

دوریان فِلدمَن          مارتین فوکس

ترجمه : مریم شریف دوست[1]

## 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 $X_1, \ldots, X_n$ یک نمونه از $N(\mu, 1)$ است، ارایه شده است.

## ۱ مقدمه

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

**الف —** تقریب توزیع دوجمله ای 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 $X_1, \ldots, X_n$ یک نمونه از $N(\mu, 1)$ است و $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$.

روش استفاده شده در قسمت (ب) می تواند 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 به‌دست آوردن تقریب توزیع دوجمله‌ای به پوآسون به‌کار گرفته شده است ([۱]،صفحه ۱۵۳).

[1] مریم شریف دوست ، گروه آمار دانشگاه شیراز

## 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 می‌دهیم (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 دوجمله‌ای مربوط را با $b(k|n, p)$ نمایش می‌دهیم.

فرض کنید $X_1$، $X_2$، ... متغیرهای تصادفی نشانگر با مقادیر ممکن ۰ و ۱ ، به‌صورت زیر باشند :

$$X_i = \begin{cases} ۱ & \text{اگر } i \text{ امین عضو نمونه، یک موفقیت باشد} \\ ۰ & \text{در غیر این‌صورت} \end{cases}$$

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

$$\begin{aligned} &P(S_{n+1} = k) \\ &\quad = P(S_n = k)P(X_{n+1} = 0|S_n = k) \qquad (۱) \\ &\quad + P(S_n = k-1)P(X_{n+1} = 1|S_n = k-1) \end{aligned}$$

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

**قضیه ۱ —** اگر $\lim_{N\to\infty} \frac{M_N}{N} = p$ آنگاه $\lim_{N\to\infty} h(k|n, N, M_N) = b(k|n, p)$.

**اثبات —** ابتدا به راحتی می بینیم که اگر $N \to \infty$، آنگاه $h(k|1, N, M_N) \to b(k|1, p)$. 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 آماره‌ی ترتیبی

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

**قضیه ۲ —** چگالی $Y_K$ به صورت زیر است :

$$\begin{aligned} &g_k(y) \qquad (۲) \\ &= \frac{n!}{(k-1)!(n-k)!} f(y)[F(y)]^{k-1}[1-F(y)]^k . \end{aligned}$$

**اثبات —** ابتدا (۲) را برای $k = 1$ به‌وسیله روش معمولی به‌دست 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 :

$$\begin{aligned} G_l(y) - G_{l+1}(y) &= P(Y_l \le y < Y_{l+1}) \qquad (۳) \\ &= \binom{n}{i} [F(y)]^i[1-F(y)]^{n-i} . \end{aligned}$$

چون $Y_l < y < Y_{l+1}$ ، پیشامدِ آن‌است که فقط $l$ تا از $X_i$ ها ، کمتر یا مساوی $y$ باشند؛ از (۳) مشتق می گیریم.

$$\begin{aligned} &g_l(y) - g_{l+1}(y) \qquad (۴) \\ &= \frac{n!}{(l-1)!(n-l)!} f(y)[F(y)]^{l-1}[1-F(y)]^{n-l} \\ &- \frac{n!}{l!(n-l-1)!} f(y)[F(y)]^l[1-F(y)]^{n-l-1} \end{aligned}$$

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

این روش می تواند برای مشتق گیری از چگالی‌های توأم آماره‌های ترتیبی نیز بسط داده شود.

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

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

$T_n$ را به‌صورت $T_n = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ 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 $Z_1, \ldots, Z_n$ مستقل و هم توزیع با $N(0, 1)$ باشند، بنابراین $\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$.

**قضیه ۳ —** فرض کنیم $X_1, \ldots, X_n$ یک نمونه از $N(\mu, 1)$ باشند، در این‌صورت $T_n \sim \chi^2_{(n-1)}$.

**اثبات —** ابتدا برای $n = 1$ ثابت می کنیم :

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

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

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

$T_m$ و $\bar{X}_m$ و $X_{m+1}$ از هم مستقل هستند، پس دو جمله طرف راست (۵) 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:

$$T_m \sim \chi^2_{m-1}$$

پس

$$T_{m+1} \sim \chi^2_m .$$

## 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 Statistis
نوشته Dorian Feldmand و Martin Fox است که در
The American Statistician, Vol. 34, No.1 (1980)
چاپ شده است.