مشخصه سازی بر اساس آنتروپی رکوردها در یک مدل تصادفی

علی خسروی طناک، معصومه فشندی

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده: فرض کنید X_1، X_2، ...، X_N یک نمونه تصادفی از جامعه‌ای با تابع توزیع پیوسته F باشد که N یک متغیر تصادفی گسسته و مثبت است. اگر R_i ها رکوردهای حاصل از نمونه فوق باشند، آنگاه گویند مدل تصادفی رکوردها داریم. در این مقاله ضمن معرفی مدل فوق ابتدا مروری بر مشخصه سازی توزیع جامعه در مدل تصادفی رکوردها خواهیم داشت. سپس نشان خواهیم داد با استفاده از خواص دنباله توابع کامل بر حسب آنتروپی رکوردها در یک مدل تصادفی هندسی، می‌توان توزیع جامعه را به طور یکتا مشخص کرد.

واژه‌های کلیدی: آنتروپی، دنباله توابع کامل، مدل تصادفی، مشخصه‌سازی

۱ مقدمه

فرض کنید $\{X_i, i \geq 1\}$ یک دنباله از متغیرهای تصادفی مستقل و همتوزیع (iid) با تابع توزیع F و تابع چگالی f باشد. مشاهده X_j را یک رکورد بالا گویند هرگاه مقدار آن از مشاهدات قبل از خود بزرگتر باشد. بنابراین X_j یک رکورد بالا است هرگاه برای هر $j > i$، $X_j > X_i$. به طور مشابه رکوردهای پایین تعریف می‌شوند. قرار دهید $L(0) = 1$ و $L(j) = \min\{k : X_k > X_{L(j-1)}\}$ برای $j \geq 1$، و $R_n = X_{L(n)}$ برای $n \geq 0$. در این صورت R_0 را رکورد بدیهی و R_n را n ـ امین رکورد بالا و $L(n)$ را زمان رخداد رکورد n ـ ام گویند. در این مقاله برای اختصار به رکورد بالا، رکورد گوئیم. برای جزئیات بیشتر در خصوص تئوری و کاربرد آماره‌های رکوردی به آرنولد و همکاران (۱۹۹۸) مراجعه شود.

فرض کنید X یک متغیر تصادفی با تابع توزیع پیوسته $F(x)$ و تابع چگالی $f(x)$ باشد. آنتروپی متغیر تصادفی X به صورت زیر تعریف می‌شود

$$H(X) = -\int_{-\infty}^{+\infty} f(x) log f(x) dx. \qquad (1)$$

$H(X)$ اولین بار توسط شانون (۱۹۴۸) تعریف شده است و تاکنون تحقیقات زیادی

دربارهٔ خواص تئوری و کاربردی آن صورت گرفته است. آنتروپی متغیر تصادفی X، اندازه ای از اطلاع است که عدم حتمیت آن را نشان می‌دهد. با اطلاع از آنتروپی یک متغیر تصادفی نمی‌توان توزیع آن را مشخص کرد، زیرا دو توزیع متفاوت ممکن است آنتروپی یکسانی داشته باشند. براتپور و همکاران (۲۰۰۷) نشان داده‌اند آنتروپی رکوردها در مدل کلاسیک، توزیع F را به طور یکتا مشخص می‌کند. در این مقاله نشان می‌دهیم آنتروپی رکوردها در مدل تصادفی نیز توزیع F را به طور یکتا مشخص می‌کند. مدل تصادفی رکوردها در بخش ۲ معرفی خواهد شد. از لم‌های زیر برای اثبات برخی قضایا استفاده می‌کنیم.

لم ۱ اگر η یک تابع پیوسته در بازه $[0,1]$ باشد به طوری که برای هر $n \geq 0$، $\int_0^1 x^n\eta(x)dx = 0$، آن‌گاه برای تقریباً هر $x \in [0,1]$ داریم: $\eta(x) = 0$.

در مقالات تعمیمی از لم ۱ وجود دارد که به قضیه کلاسیک مونتز ـ زاس[۱] معروف است.

لم ۲ اگر η یک تابع پیوسته در بازه $[0,1]$ باشد به طوری که برای هر n_j، $j \geq 1$، که $0 < n_1 < n_2 < \ldots$ و $\sum_{j=1}^{\infty} n_j^{-1} = +\infty$، داشته باشیم $\int_0^1 x^{n_j}\eta(x)dx = 0$، آن‌گاه برای تقریباً هر $x \in [0,1]$ داریم: $\eta(x) = 0$.

در بخش ۲، مدل تصادفی رکوردها معرفی و توابع چگالی احتمال آنها ارائه می‌شود. در بخش ۳، نشان داده شده است که زیر دنباله‌های مناسب از امید ریاضی رکوردها در مدل تصادفی، توزیع F را مشخص می‌کند. در بخش ۴، نشان می‌دهیم زیر دنباله‌هایی از آنتروپی رکوردها در یک مدل تصادفی نیز توزیع F را به طور یکتا مشخص می‌کند.

۲ مدل تصادفی رکوردها

در یک دنباله از متغیرهای تصادفی فرض کنید تنها X_1، X_2، ...، X_N مشاهده شوند که N یک متغیر تصادفی صحیح مقدار و مستقل از X_1 ، X_2 ،... است. رکوردهای حاصل از این دنباله را در نظر بگیرید. این مدل به مدل تصادفی رکوردها معروف است. اگر N متغیر تصادفی هندسی با تابع احتمال $P(N = k) = p\ q^{k-1}$، برای $k \geq 1$ باشد، که در آن $0 < p < 1$ و $q = 1 - p$، آن‌گاه یک مدل تصادفی هندسی

[۱] Münt-Szász

رکوردها (GRR)[۲] داریم. اگر N با احتمال ۱ بینهایت شود، مدل کلاسیک رکوردها حاصل می‌شود. برای آشنایی بیشتر به آرنولد و همکاران (۱۹۹۸) مراجعه کنید.

حال یک مدل GRR را در نظر بگیرید. فرض کنید M تعداد رکوردهای غیربدیهی مشاهده‌شده و تابع توزیع F پیوسته با تابع چگالی f باشد. دراین‌صورت تابع درستنمایی $(R_0, \ldots, R_n, \{M \geq n\})$ برابرست با

$$f(r_0, r_1, \ldots, r_n; M \geq n) = f(r_n) \prod_{i=0}^{n-1} \frac{qf(r_i)}{1 - qF(r_i)} \,,\, r_0 < r_1 < \ldots < r_n. \quad (۲)$$

برای جزئیات بیشتر آرنولد و همکاران (۱۹۹۸، صفحه ۲۲۹) را ببینید.
از رابطه (۲) داریم

$$f(r_0, r_1, \ldots, r_{n-1}; M \geq n) = (1 - F(r_{n-1})) \prod_{i=0}^{n-1} \frac{qf(r_i)}{1 - qF(r_i)} \,. \quad (۳)$$

چون

$$f(r_0, \ldots, r_n; M = n) = f(r_0, \ldots, r_n; M \geq n) - f(r_0, \ldots, r_n; M \geq n+1),$$

بنابراین با استفاده از روابط (۲) و (۳) داریم

$$f(r_0, r_1, \ldots, r_n; M = n) = \frac{p}{q} \prod_{i=0}^{n} \frac{qf(r_i)}{1 - qF(r_i)} \,,\, r_0 < r_1 < \ldots < r_n. \quad (۴)$$

با استفاده از روابط (۲) و (۴) می توانیم توابع درستنمایی $(R_n, \{M \geq n\})$ و $(R_n, \{M = n\})$ را به دست آوریم که به ترتیب عبارتند از

$$f(r, M \geq n) = f(r) \frac{[-log(1 - qF(r))]^n}{n!} \,, \quad -\infty < r < +\infty \,, \quad (۵)$$

و

$$f(r, M = n) = \frac{pf(r)}{1 - qF(r)} \frac{[-log(1 - qF(r))]^n}{n!} \,, \quad -\infty < r < +\infty \,. \quad (۶)$$

روابط (۳) تا (۶) در ناگاراجا و بارلوی (۲۰۰۳) نیز آمده است.

[۲] Geometric Random Record

۳ مشخصه سازی بر اساس امید ریاضی

کرمانی و بگ (۱۹۸۴) نشان داده‌اند در مدل کلاسیک، دنباله $\{E(R_n), n \geq m\}$، توزیع F را در بین توزیع‌های پیوسته‌ای که گشتاور p ـ ام آن‌ها برای یک $p > ۱$، متناهی است، مشخص می‌کند در صورتی که لین و هوانگ (۱۹۸۷) ثابت کرده‌اند برای این مدل زیر دنباله $\{E(R_{n_j}), j \geq ۱\}$، که $\sum_{j=۱}^{\infty} n_j^{-۱} = +\infty$، ممکن است توزیع F را مشخص نکند. ناگاراجا و بارلوی (۲۰۰۳) نشان داده‌اند که در یک مدل GRR، اگر $\mathrm{E}(|X|)$ متناهی باشد آنگاه $\mathrm{E}(R_n|M \geq n)$ و $\mathrm{E}(R_n|M = n)$ برای هر $n \geq ۱$ وجود دارند، سپس ثابت کرده‌اند زیر دنباله‌های مناسب از هر یک از آن‌ها توزیع F را به طور یکتا مشخص می‌کند.

قضیه **۹** (ناگاراجا و بارلوی، ۲۰۰۳) در مدل GRR، اگر $E(|X|) < \infty$، آن‌گاه هر یک از دنباله‌های $\mathrm{E}(R_{n_j}|M \geq n_j)$ و $\mathrm{E}(R_{n_j}|M = n_j)$، که در آن $\sum_{j=۱}^{\infty} n_j^{-۱} = +\infty$، توزیع F را در خانواده توزیع‌های پیوسته به طور یکتا مشخص می کند.

گوپتا (۱۹۸۴) با استفاده از نتایج کرمانی و بگ (۱۹۸۴) که در بالا به آن اشاره شد، نشان داد در مدل کلاسیک، دنباله $\{E(R_{n+۱}) - E(R_n), n \geq m\}$، توزیع F را در خانواده توزیع‌های پیوسته مکانی مشخص می‌کند. اما ناگاراجا و بارلوی (۲۰۰۳) نشان داده‌اند در مدل GRR، زیردنباله‌ای ازگشتاورهای تفاضل رکوردها برای مشخصه‌سازی F کافی است. نتایج آن‌ها در قضیه بعدی آمده است.

قضیه ۱۰ (ناگاراجا و بارلوی، ۲۰۰۳) در مدل GRR، اگر $E(|X|) < \infty$، آن‌گاه هر یک از زیردنباله‌های $\mathrm{E}(R_{n_j} - R_{n_{j-۱}}|M \geq n_j)$ و $\mathrm{E}(R_{n_j} - R_{n_{j-۱}}|M = n_j)$، که در آن $\sum_{j=۱}^{\infty} n_j^{-۱} = +\infty$، توزیع F را در خانواده توزیع‌های مکانی پیوسته مشخص می کند.

۴ مشخصه سازی بر اساس آنتروپی

در این بخش با بیان قضیه‌هایی نشان می‌دهیم آنتروپی رکوردها در یک مدل تصادفی هندسی، توزیع جامعه را مشخص می‌کند.

قضیه **۱۱** در مدل GRR، دنباله $H(R_{n_j}|M \geq n_j)$، که در آن $\sum_{j=۱}^{\infty} n_j^{-۱} = +\infty$، توزیع F را در خانواده توزیع‌های مکانی پیوسته به طور یکتا مشخص می کند.

اثبات : با توجه به رابطه (۱)، آنتروپی n ـ امین رکورد از یک مدل GRR با تابع توزیع F و تابع چگالی f برابرست با

$$H(R_n^X|M \geq n) = -\int_{S_X} f_{R_n}(r|M \geq n) log(f_{R_n}(r|M \geq n))dr.$$

بنابر رابطه (۵) و با تغییر متغیر $u(-\log p) = [-\log(1 - qF(r))]$، داریم $qF(r) = 1 - p^u$ و $qf(r)dr = -p^u \log p \, du$، بنابراین

$$H(R_n^X|M \geq n) = \frac{K}{q}\int_0^1 u^n p^u log\left(K(-log p)u^n f\left(F^{-1}\left(\frac{1-p^u}{q}\right)\right)\right) du, \quad (7)$$

که در آن $K = \frac{(-\log p)^n}{n! P(M \geq n)}$. حال فرض می‌کنیم X و Y دو متغیر تصادفی با توابع توزیع به ترتیب F و G باشند و برای هر $n \geq 0$ داشته باشیم

$$H(R_n^X|M \geq n) = H(R_n^Y|M \geq n).$$

پس با توجه به رابطه (۷) و پس از ساده کردن طرفین تساوی داریم

$$\int_0^1 u^n p^u log\left(f\left(F^{-1}\left(\frac{1-p^u}{q}\right)\right)\right) du = \int_0^1 u^n p^u log\left(g\left(G^{-1}\left(\frac{1-p^u}{q}\right)\right)\right) du.$$

در نتیجه

$$\int_0^1 p^u log\left(\frac{f\left(F^{-1}\left(\frac{1-p^u}{q}\right)\right)}{g\left(G^{-1}\left(\frac{1-p^u}{q}\right)\right)}\right) u^n \, du = 0. \quad (8)$$

اگر رابطه (۸) برای $n = n_j$، $j \geq 1$، به طوری که $\sum_{j=1}^{\infty} n_j^{-1} = +\infty$، برقرارباشد آن‌گاه بنابر لم ۲، تقریباً برای هر $u \in (0, 1)$،

$$f\left(F^{-1}\left(\frac{1-p^u}{q}\right)\right) = g\left(G^{-1}\left(\frac{1-p^u}{q}\right)\right). \quad (9)$$

یا تقریباً برای هر $v \in (0, 1)$، $f(F^{-1}(v)) = g(G^{-1}(v))$. بنابراین $(F^{-1})'(v) = (G^{-1})'(v)$ پس $F^{-1}(v) = G^{-1}(v) + c$، که c یک عدد ثابت است.

قضیه **۱۲** در مدل GRR، دنباله $H(R_{n_j}|M = n_j)$، که در آن $\sum_{j=1}^{\infty} n_j^{-1} = +\infty$، توزیع F را در خانواده توزیع های پیوسته مکانی به طور یکتا مشخص می کند.

اثبات : بنا به تعریف آنتروپی

$$H(R_n^X|M=n) = -\int_{S_X} f_{R_n}(r|M=n) log(f_{R_n}(r|M=n))dr.$$

با توجه به رابطه (۶) داریم

$$H(R_n^X|M=n) = -\frac{C}{q}\int_0^1 u^n log\left(Cu^n f\left(F^{-1}\left(\frac{1-p^u}{q}\right)\right)\right)du,$$

که در آن $C = \frac{p(-logp)^n}{n!P(M=n)}$. اکنون فرض می‌کنیم برای هر $n \geq 0$،

$$H(R_n^X|M=n) = H(R_n^Y|M=n).$$

ادامه اثبات مشابه قضیه ۳ است.

بحث و جمع‌بندی

می‌دانیم به طور کلی نتایج مشخصه سازی توزیع‌های آماری می‌تواند در مسائل آزمون فرض مفید واقع شود که تحقیقات در این زمینه در سال‌های اخیر رونق پیدا کرده است (آزمون نیکویی برازش بر اساس آنتروپی). قبلاً براتپور و همکاران (۲۰۰۷) نشان داده‌اند در مدل کلاسیک، برابری آنتروپی رکوردهای حاصل از دو نمونه مستقل، به ازای هر $n \geq 1$، معادل با برابری توزیع دو جامعه مربوط است. در این مقاله نتایج آن‌ها را به مدل تصادفی رکوردها تعمیم داده‌ایم، به علاوه نشان داده‌ایم برابری آنتروپی رکوردها برای یک زیردنباله مناسب از آن‌ها برای رسیدن به این هدف کافی است.

مراجع

Arnold, B. C., Balakrishnan, N., Nagaraja, H. N. (1998). *Records*. New York: John Wiley and Sons.

Baratpour, S., Ahmadi, J., Arghami, N. R. (2007). Some Characterization Based on Entropy of Order Statistics and Record Values. *Comm. Statist.- Theory and Methods*, **36**, 47-57.

Gupta, R. C. (1984). Relationships Between Order Statistics and Record Values and Some Characterization Results. *J. Appl. Prob.*, **21**, 425-430.

Kirmani, S. N. U. A., Beg, M. I. (1984). On Characterization of Distributions by Expected Records. *Sankhya A*, **46**, 463-465.

Lin, G. D., Huang, J. S. (1987). A Note on the Sequence of Expectations of Maxima and of Record Values. *Sankhya A*, **49**, 272-273.

Nagaraja, H. N., Barlevy, G. (2003). Characterization Using Record Moments in a Random Record Model and Applications. *J. Appl. Prob.*, **40**, 826-833.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, **27**, 379-432.