

معرفی و خواص سانسور نوع دوم پیشرو تعمیم یافته

آمنه سادات میرنیام^{1*}؛ زهرا شناوری²؛ عبدالرسول برهانی حقیقی³

دریافت: 1392/10/09

پذیرش: 1394/03/03

چکیده

این مقاله ابتدا به مفهوم زمان شکست پرداخته و سپس سانسور و انواع آن، به خصوص طرح سانسور نوع دوم پیشرو تعمیم یافته را به همراه خواص آن در قالب قضایایی معرفی نموده است. همچنین مثالی برای روشن تر شدن این نوع طرح بیان گردیده است.

واژگان کلیدی: آنالیز بقا، زمان شکست، سانسور، آزمون عمر.

1. دانشجوی دکتری، آمار دانشگاه شیراز (*نویسنده مسئول) bahareh104@gmail.com

3. دانشجوی دکتری، آمار دانشگاه شیراز zshnavari@yahoo.com

3. استادیار، بخش آمار دانشگاه شیراز borhani@susc.ac.ir

مقدمه

در آنالیز بقاء کلاسیک، به مطالعه گروهی از افراد پرداخته می‌شود که برای تک تک آنها یک رویداد لحظه‌ای، به نام شکست، تعریف می‌شود. مدت زمانی که طول می‌کشد تا این شکست رخ دهد، زمان شکست نامیده می‌شود. روی هر فرد شکست حداکثر یک بار می‌تواند رخ دهد. داده‌های بقاء در حقیقت زمان‌های شکست را از یک لحظه ورود تعریف شده تا وقوع یک رویداد نهایی اندازه‌گیری می‌کنند. تعیین دقیق زمان شکست مستلزم آن است که:

الف- معنی و مفهوم شکست کاملاً واضح باشد.

ب- یک مبدأ زمانی برای هر یک از افراد مورد مطالعه تعریف شده باشد.

ج- یک مقیاس برای اندازه‌گیری زمان تعیین شده باشد. (حیدری، 1377)

دانستن زمان شکست در آنالیز بقا برای ما حائز اهمیت است. گاهی این زمان‌ها را نمی‌توان مشاهده کرد چون به دلایلی سانسور می‌شوند. محققان بسیاری در مورد داده‌های سانسور شده تحقیق و بررسی کرده‌اند. از جمله سینها¹ (1986)، دیویس و فلدستین² (1979)، سن³ (1979)، هالپرین و همکاران⁴ (1989)، ویورس و بالاکریشنان⁵ (1994) باندیوپا-دهیای و چاتوپادهیای⁶ (1995)، یین و تسه⁷ (1996) آگاروالا و بالاکریشنان⁸ (1998) همکاران⁹ (2001)، گیلباد¹⁰ (2001). در ادامه مفهوم سانسور و انواع آن را بررسی کرده و خواص آنها را نیز در قالب قضایایی بیان خواهیم کرد.

سانسور¹¹

همان‌گونه که قبلاً اشاره شد در آنالیز بقاء کلاسیک گروهی از افراد را از یک لحظه ورود

تعریف شده تا وقوع یک رویداد نهایی (شکست) مورد مطالعه قرار می‌دهند.

هدف از مطالعه داده‌های بقاء مشاهده و بررسی زمان‌های شکست برای افراد تحت مطالعه می‌باشد غالباً امکان‌پذیر نیست که صبر کنیم تا اینکه این رویداد برای همه افراد اتفاق بیفتد. برای بعضی، فقط می‌دانیم که این رویداد هنوز در یک زمان خاص رخ نداده است، یعنی زمان شکست هنوز مشاهده نشده است. در این حالت مشاهده زمان شکست سانسور شده است. به عبارت دیگر زمانی را که در آن هنوز شکست برای افراد مشاهده نشده، زمان سانسور آن فرد می‌نامند. مثلاً در آزمون عمر در قابلیت اعتماد صنعتی، بعضی از قطعات (مولفه‌ها) ممکن است از کار نیفتند، یعنی شکست برای آنها مشاهده نشود.

اساساً داده‌های زمانی را سانسور شده گوئیم هرگاه مولفه‌هایی در نمونه وجود داشته باشند که تنها یک کران پائین (یا بالا) روی طول عمر آنها در دسترس باشد.

دانستن علت سانسور در مسائل، برای تحلیل داده‌های بقاء دارای اهمیت است. یادآور می‌شویم که همانند شکست، سانسور نیز یک رویداد لحظه‌ای است و زمان مشاهده برای افراد سانسور شده (زمان سانسور) بایستی ثبت گردد.

انواع داده‌های سانسور شده طول عمر

برای جمع‌آوری داده‌های طول عمر دو روش وجود دارد:

1- روش اول به این صورت است که قبل از شروع آزمایش زمان t_0 را مشخص کنیم و فقط طول عمر مولفه‌هایی را که تا زمان t_0 شکست خورده‌اند یادداشت کنیم این روش را زمان-سانسور می‌نامند، زیرا زمان قطعی آزمایش از ابتدا مشخص شده است.¹²

1. Sinha
2. Davis & Feldstain
3. Sen
4. Halperin et al.
5. Viveros & Balakrishnan
6. Bandyopadhyay & Chatopadhyay
7. Yuen & Tse
8. Aggarwala & Balakrishnan
9. Balakrishnan et al.
10. Guilbaud
11. Censoring

(1979)، سن (1979)، هالپرین و همکاران (1989)، ویورس و بالاکریشنان (1994) باندیوپادهای و چاتوپادهای (1995)، ین و تسه (1996) آگاروالو و بالاکریشنان (1998) بالاکریشنان و همکاران (2001) و گیلبلد (2001) در نظر گرفته شده است.

این طرح به صورت زیر است:

1- یک گروه n تایی از واحدهای آزمایشی در زمان صفر در یک آزمایش **آزمون عمر** قرار داده می شوند.

2- به طور ناگهانی پس از مشاهده اولین شکست (مرگ)، یک تعداد از پیش تعیین شده $R_1 \geq 0$ از $n - 1$ واحد باقی مانده به تصادف انتخاب و از آزمایش خارج می شوند. بنابراین تنها $n - 1 - R_1$ واحد قابل مشاهده در آزمایش باقی می ماند.

3- به طور ناگهانی پس از مشاهده دومین شکست (مرگ)، یک تعداد از پیش تعیین شده $R_2 \geq 0$ از $n - 1 - R_1 - 1$ واحد باقی مانده به تصادف از آزمایش خارج می شوند. بنابراین تنها $n - 2 - (R_1 + R_2)$ واحد قابل مشاهده در آزمایش باقی می ماند و به همین ترتیب ...

بنابراین واحدهای خارج شده، در لحظه شکست (مرگ) واحدهای دیگر، از راست سانسور می شوند، یعنی طول عمر آنها کران بالا ندارد. مقادیر R_1 و R_2 و ... و R_m ، یعنی همان واحدهای خارج شده، طرح سانسور را تعیین می کنند.

سانسور نوع دوم پیشرو تعمیم یافته⁹

بالاکریشنان و سند هو¹⁰ (1995)، علاوه بر مدل سانسور پیشرو، مدل سانسور از سمت چپ¹¹ داده ها را در نظر گرفتند که سانسور نوع دوم پیشرو تعمیم یافته نامیده می شود. آنها فرض کردند که:

1- در زمان $\mathbf{x}_0 = \mathbf{0}$ واحد در آزمون قرار داده می شوند.

2- روش دوم به این صورت است که قبل از شروع آزمایش یک عدد صحیح کوچکتر از n (n) تعداد کل مولفه ها) را معین کنیم و آن را r بنامیم و آزمایش را تا زمانی که طول عمر r تا مولفه اول به پایان برسد ادامه دهیم و دقیقاً در زمان $T(r)$ یعنی طول عمر r امین مولفه ای که شکست خورده، آزمایش را قطع می کنیم. این روش را شکست سانسور می نامیم زیرا تعداد شکست های ثبت شده از ابتدا مشخص شده است¹.

همچنین اگر همه مولفه ها آزمایش را با هم شروع کنند، داده ها را سانسور منفرد² و اگر مولفه ها در زمان های مختلفی از بازه زمانی شروع به فعالیت کنند، و آزمایش در زمان t_0 قطع شود داده ها را سانسور چندگانه³ می نامند.

سانسور نوع اول⁴

به داده های منفرد زمان- سانسور⁵، داده های سانسور شده نوع اول می گویند.

سانسور نوع دوم⁶

به داده های منفرد شکست سانسور⁷ داده های سانسور شده نوع دوم می گویند.

سانسور نوع دوم پیشرو⁸

سانسور نوع دوم پیشرو، طرح مفید و عمومی تری است که در آن به آزمایشگر اجازه می دهد تا تعداد مشخصی از مولفه ها را در هر یک از زمان های متعدد شکست از مطالعه خارج کند (یعنی سانسور کند). حذف کردن های متوالی از این نوع به دلایل مختلفی طراحی می شوند. صرفه جویی در زمان و هزینه می تواند نتیجه چنین طرح نمونه گیری باشد. این نوع نمونه های سانسور شده توسط دیویس و فلدستین

1. Failure Censoring
2. Single Censoring
3. Multiple Censoring
4. Type I Censoring
5. Single Time Censoring
6. Type-II Censoring (Right Censoring)
7. Single Failure – Censoring
8. Progressive Type-II Censoring

9. General Progressive Type-II Censoring

10. Balakrishnan & Sandhu,

11. Left Censoring

تعمیم یافته و زمان های شکستی که دارای توزیع نمایی هستند ارائه می گردد.

خواص آماره های ترتیبی سانسور شده پیشرو تعمیم یافته

قضیه 1: تحت شرایط بالا داریم:

(الف) توزیع X_{r+i} مستقل از $R_{r+i}, R_{r+i+1}, \dots, R_m$ است. $(i = 1, \dots, m-r)$

(ب) $X_{r+1}, X_{r+2}, \dots, X_{r+i}$ یک نمونه سانسور شده نوع دوم پیشرو تعمیم یافته، به اندازه i از n واحد موجود در آزمون با طرح سانسور r و $(R_{r+1}, \dots, R_{r+i-1})$ و $n - (R_{r+1}, \dots, R_{r+i-1}) - (r+i)$ اثبات: در منبع (2).

نکته 1: این قضیه در اثبات برخی روابط بازگشتی برای گشتاورهای آماره های ترتیبی سانسور شده نوع دوم پیشرو مفید خواهد بود.

قضیه 2: آماره های ترتیبی سانسور شده نوع دوم پیشرو تعمیم یافته از یک توزیع پیوسته دلخواه، یک زنجیره مارکف تشکیل می دهند.

اثبات: هدف ما این است که نشان دهیم:

$$f_{X_{r+i-1}|X_{r+i}, X_{r+i-1}, \dots, X_{r+1}}(x_{r+i-1}|x_{r+i}, x_{r+i-1}, \dots, x_{r+1}) = f_{X_{r+i-1}|X_{r+i}}(x_{r+i-1}|x_{r+i}) \quad (1)$$

سمت چپ تساوی (1)، می شود:

$$\frac{(n-r-R_{r+1}-\dots-R_{r+i}-i)}{\left[\frac{1-F(x_{r+i-1})}{1-F(x_{r+i})}\right]^{n-r-R_{r+1}-\dots-R_{r+i-1}-1} \frac{f(x_{r+i})}{[1-F(x_{r+i})]}}$$

$$= f_{X_{r+i-1}|X_{r+i}}(x_{r+i-1}|x_{r+i})$$

بنابراین آماره های ترتیبی سانسور شده نوع دوم پیشرو تعمیم یافته، یک زنجیره مارکف را تشکیل می دهند.

قضیه 3: X_{r+i+1}, \dots, X_m ($i \geq 1$)، به شرط $X_{r+i} = X_{r+j}$ توأمأ بصورت یک نمونه سانسور شده از راست، نوع دوم پیشرو به اندازه $m-r-i$ از $n - R_{r+1} - \dots - R_{r+i} - r - i$ از چگالی $f(\cdot)$ ، که از سمت چپ در X_{r+j} بریده شده¹ توزیع می شوند یعنی با چگالی

2- اگر X_i زمان i امین شکست مرتب شده باشد، X_1, \dots, X_r مشاهده نمی شوند.

3- در زمان X_i ($i = r+1, \dots, m-1$)، واحد به طور تصادفی از آزمون خارج می شوند. بنابراین پیش از $(i+1)$ امین شکست، $n_i = n - i - \sum_{j=r+1}^i R_j$ واحد در آزمون وجود دارند.

4- در زمان m امین شکست، X_m آزمایش خاتمه می یابد که R_m واحد باقیمانده از آزمون خارج می شوند.

توضیح بیشتر و مثال:

فرض کنیم $X_0 = 0$ و $r = 3$ یعنی زمان های شکست X_1, X_2, X_3 مشاهده نمی شوند. در زمان X_4 که $X_1 \leq X_2 \leq X_3 \leq X_4$ واحد به طور تصادفی از آزمون خارج می شوند بنابراین پیش از مشاهده پنجمین شکست، تعداد واحدهای موجود در آزمون عبارتند از:

$$n_4 = n - 3 - 1 - \sum_{j=4}^4 R_j = n - 4 - R_4$$

اگر $m = 7$ باشد با مشاهده زمان X_6 واحد R_6 به طور تصادفی از آزمون خارج می شوند و تعداد باقیمانده پیش از مشاهده شکست هفتم عبارت است از:

$$n_6 = n - 6 - (R_4 + R_5 + R_6).$$

با مشاهده X_7 واحد R_7 باقیمانده از آزمون خارج شده و آزمایش خاتمه می یابد.

شرایط:

R_i ها، m ، r اعداد صحیح از پیش تعیین شده ای هستند که باید در شرایط زیر صدق کنند:

$$0 \leq r < m \leq n - 1$$

$$2- \quad R_m = n_{m-1} - 1 \quad \text{و} \quad i = r + 1, \dots, m - 1$$

$$0 \leq R_i < n_{i-1} - 1 \quad \text{و} \quad n_r = n - r$$

برای راحتی نوشتاری فرض می کنیم

$$\vec{R} = (R_{r+1}, \dots, R_m) \quad \text{و} \quad \vec{x} = (x_{r+1}, \dots, x_m)$$

فصل بعدی یک طرح سانسور نوع دوم پیشرو

این نوع سانسور در قالب قضایایی همراه یک مثال مورد مطالعه قرار گرفته است.

به عنوان مثال نشان داده شده که این داده‌ها از هر توزیعی که باشند دارای خاصیت مارکف هستند و این خاصیت ما را در یافتن چگالی توأم شرطی بخشی از داده‌های سانسور شده یاری می‌دهد.

$f(x)/(1 - F(x_{r+i}))$ و با طرح سانسور پیشرو R_{r+i+1}, \dots, R_m
اثبات: در منبع (2).

نتیجه گیری

در این مقاله ابتدا داده‌های سانسور شده طول عمر معرفی گردیده ولی تاکید اصلی بر روی سانسور نوع دوم پیشرو تعمیم یافته می‌باشد. همچنین خواص

منابع

حیدری، فریرز. (1377). تابع راست‌نمایی برای مدل‌های سانسور معمولی در آنالیز بقاء، گلچین ریاضی، جلد پنجم، شماره 1.

- Aggarwala, R. and Balakrishnan, N. (1998). Some properties of progressive censored order statistics from arbitrary and uniform distributions with applications to inference and simulation. *Journal of statistical planning and inference*. 70, 35-49.
- Balakrishnan, N. and Sandhu, R.A. (1995). A simple simulational algorithm for generating progressive type II censored samples. *The American Statistician*, Vol. 49, No.2. pp.229-230.
- Balakrishnan, N., Cramer, E., kamps, U. and sckenk, N. (2001). Progressive type II censored order statistics from exponential distributions. *Statistics* 35, 537-556.
- Bandyopadhyay, U. and Chattopadhyay, G. (1995). Progressive censoring under inverse sampling for nonparametric two-sample problems. *Sequential Anal.* 14, 1-28.
- Davis, H.T. and Feldstein, M.L. (1979). The generalized pareto law as a model for progressively censored survival data. *Biometrika* 66, 299-306.
- Guilbaud, O. (2001). Exact non-parametric confidence intervals for quantiles with progressive type II censoring. *Scand. J. statis.* 28, 699-713.
- Halperin, M., Hamdy, M.I. and Thall, P.F. (1989). Distribution - free confidence interval for a parameter of Wilcoxon - Mann - Whitney type for ordered categories and progressive censoring. *Biometrics* 45, 509-521.
- Sen, P.K. (1979). Weak convergence of some quantile processes arising in progressively censored tests. *Ann. Statist.* 7, 414-431.
- Sinha, S.K. (1986). *Reliability and life testing*. Wiley Eatern Ltd, India :Dehli.
- Viveros, R. and Balakrishnan, N. (1994). Interval estimation of life characteristics from progressively censored samples. *Technometrics* 36, 84-91.
- Yuen, H.K. and Tse, S.K. (1996). Parameter estimation for Weibull distributed lifetimes under progressive censoring with random removals. *J. Statistics. Comput. Simulation* 55, 57-71.