

## برآورد آنتروپی گذشته در داده‌های طول-اریب

فرزانه اولیاءزاده<sup>۱</sup>، انیس ایرانمنش<sup>۱</sup> و وحید فکور<sup>۲</sup>

<sup>۱</sup>گروه ریاضی و آمار، واحد مشهد، دانشگاه آزاد اسلامی

<sup>۲</sup>گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۹۹/۰۷/۲۲ تاریخ پذیرش و انتشار: ۱۴۰۰/۰۶/۳۰

**چکیده:** در این مقاله دو برآوردگر ناپارامتری برای آنتروپی گذشته بر اساس داده‌های طول-اریب معرفی و سازگاری قوی این برآوردگرها اثبات شده است. به منظور مقایسه عملکرد آن‌ها مطالعه شبیه‌سازی انجام شده و بر اساس نتایج به دست آمده، نشان داده می‌شود که هر کدام در نواحی مختلفی از تکیه‌گاه توزیع احتمال متغیر تصادفی طول-اریب، بهتر از دیگری عمل می‌کند. واژه‌های کلیدی: آنتروپی گذشته، برآوردگر هسته‌ای تابع چگالی، داده‌های طول-اریب، سازگاری قوی.

## ۱ مقدمه

در بسیاری از موقعیت‌های واقعی ممکن است که عدم قطعیت در مورد یک پدیده تصادفی به گذشته مربوط باشد. به عنوان مثال سیستمی را در نظر بگیرید که وضعیت آن فقط در زمان معین بازرسی از پیش تعیین شده، مشاهده می‌شود. اگر در زمان  $t$ ، سیستم برای اولین بار مورد بازرسی قرار گیرد و مشخص شود که خراب است، پس عدم قطعیت به گذشته متکی است، یعنی که سیستم در لحظه  $t$  از کار افتاده است. بنابراین می‌توان مفهوم عدم قطعیت را برای زمان گذشته نیز تعریف کرد. براساس این ایده **دی کرنسنزو و لانگوباردی (۲۰۰۲)** آنتروپی گذشته<sup>۱</sup> را روی بازه  $(\circ, t)$  تعریف کردند و آن را

آدرس الکترونیکی نویسنده مسئول مقاله: انیس ایرانمنش، iranmanesh0030@mshdiau.ac.ir  
کد موضوع‌بندی ریاضی (۲۰۱۰): 94A17 62B10 .

<sup>۱</sup>Past entropy

$\bar{H}(t)$  نامیدند. همچنین آن‌ها ارتباط  $\bar{H}(t)$  را با آنتروپی مانده<sup>۱</sup> به‌دست آوردند. از  $\bar{H}(t)$  در علوم پزشکی که در آن زمان دقیق شکست (مثلاً زمان فوت انسان) مهم است، استفاده می‌شود. **گوپتا و ناندا** (۲۰۰۲) تعمیم مرتبه  $\beta$  ام آنتروپی گذشته را به‌دست آوردند. **ناندا و پال** (۲۰۰۶) بعضی از ویژگی‌ها و کاربردهای آنتروپی گذشته را بررسی کردند. **گوپتا** (۲۰۰۹) ثابت کرد تحت شرایط معینی  $\bar{H}(t)$  می‌تواند به‌طور یکتا یک توزیع را مشخص کند. **ساجلاس و پاپینو** (۲۰۱۴) کاربردهایی از آنتروپی گذشته را در تحلیل بقا ارائه کردند. **مایا** (۲۰۱۳) برآوردگر هسته‌ای برای آنتروپی گذشته را در داده‌های وابسته به‌دست آورد. **مایا و همکاران** (۲۰۱۴) برآورد ناپارامتری برای آنتروپی گذشته تعمیم یافته، در داده‌های سانسور شده را به‌دست آوردند. **توانگر و اسدی** (۲۰۱۰) و **دی کرنسزو و همکاران** (۲۰۲۰) کاربرد آنتروپی گذشته را ارائه کردند. آنتروپی مانده و همچنین آنتروپی به مفهوم کلی آن در **حوتی و احمدی** (۱۳۹۵) و **پورکاظمی و همکاران** (۱۳۹۸) مطرح شده است. تاکنون مطالعاتی که در زمینه برآورد آنتروپی گذشته انجام شده برای داده‌های کامل و سانسور شده بوده است. در این مقاله برای اولین بار برآورد  $\bar{H}(t)$  برای داده‌های طول-اریب مطرح شده و دو برآوردگر برای آنتروپی گذشته بر اساس این داده‌ها، به همراه خاصیت سازگاری آن‌ها ارائه می‌شود.

در بخش ۲، مفاهیم آنتروپی گذشته و طول-اریبی به اختصار ارائه خواهد شد. در بخش ۳، ابتدا برای  $\bar{H}(t)$  دو برآوردگر براساس داده‌های طول-اریب ارائه، و نحوه ساختن آن‌ها شرح داده خواهد شد. در بخش ۴، سازگاری قوی دو برآوردگر  $\bar{H}(t)$  مورد بررسی قرار گرفته و در این رابطه دو قضیه اثبات خواهد شد. به‌منظور ارزیابی عملکرد برآوردگرهای معرفی شده، مطالعات شبیه‌سازی در بخش ۵ انجام شده است. برای مقایسه بهتر اریبی‌ها و  $MSE$  های دو برآوردگر، نمودارهایی نیز رسم شده است. در خاتمه بحث و نتیجه‌گیری در بخش ۶ ارائه می‌گردد.

## ۲ مفهوم طول-اریبی و آنتروپی گذشته

داده‌های طول-اریب نخستین بار توسط **ویکسل** (۱۹۲۶) در آناتومی مطرح شد. وی هنگام مشاهده گلبول‌ها در میکروسکوپ متوجه شد که فقط گلبول‌هایی قابل مشاهده هستند که اندازه بزرگی آن‌ها، از حد معینی بیشتر باشد و گلبول‌های کوچکتر قابل دیدن در میکروسکوپ نمی‌باشند. بعدها این موضوع توسط **کاکس** (۱۹۶۹) در زمینه برآورد توزیع طول الیاف پارچه به صورت آماری مطالعه شد. کاکس متوجه شد که الیاف با طول بلندتر، با احتمال بیشتری وارد نمونه می‌شوند و این موضوع نوعی اریبی را به نتایج

<sup>1</sup>Residual entropy

تحلیل می‌کرد. نتایج بنیادی و اساسی در نظریه داده‌های طول-اریب حاصل تلاش‌های بسیار کاکس است. اطلاعات بیشتر در زمینه داده‌های اریب، به خصوص طول-اریب توسط کین (۲۰۱۷) ارائه شده است.

فرض کنید متغیر تصادفی  $X$  با تکیه‌گاه نامنفی، با تابع توزیع مطلقاً پیوسته  $F(\cdot)$  و تابع چگالی  $f(\cdot)$  باشد. متغیر تصادفی  $Y$  دارای توزیع طول-اریب  $G(\cdot)$  است هرگاه توزیع  $Y$  برای هر  $y \geq 0$  به صورت  $G(y) = \frac{1}{\mu} \int_0^y x f(x) dx$  باشد، که در آن  $\mu = \int_0^\infty x f(x) dx$ . تابع چگالی  $Y$  به صورت

$$g(y) = \frac{y f(y)}{\mu}, \quad y \geq 0, \quad (۱)$$

است. توزیع  $G(\cdot)$  را نسخه طول-اریب توزیع  $F(\cdot)$  می‌نامند. واضح است که

$$F(y) = \mu \int_0^y x^{-1} g(x) dx, \quad f(y) = \frac{\mu g(y)}{y}, \quad y \geq 0. \quad (۲)$$

اگر  $X$  نشان‌دهنده طول عمر یک مؤلفه (یا سیستم) با تابع چگالی  $f(\cdot)$  و تابع توزیع  $F(\cdot)$  باشد، متغیر تصادفی  $X(t) \stackrel{d}{=} (t - X | X \leq t)$ ، مدت زمان سپری شده از پایان عمر آن مؤلفه تا زمان  $t$  است به شرط آن که مؤلفه حداکثر تا زمان  $t$  عمر کند. به این متغیر تصادفی طول عمر غیر فعال گویند و تابع توزیع احتمال و تابع چگالی آن برای هر  $t$  ثابت به ترتیب به صورت

$$F_t(x) = 1 - \frac{F(t-x)}{F(t)}, \quad f_t(x) = \frac{f(t-x)}{F(t)}, \quad x \in (0, t).$$

هستند. به عنوان مثال فرض کنید در زمان  $t$  یک آزمایش برای بررسی ابتلا به یک نوع بیماری، روی فرد خاصی انجام شود و نتیجه آزمایش مثبت باشد. اگر  $X$  را زمان ابتلا به بیماری در نظر بگیریم، آنگاه مطمئناً  $X < t$  است. حال این سوال مطرح می‌شود که چه مدت زمانی است که بیمار، به بیماری مبتلا شده است. به عنوان مثال دیگر فرض کنید که یک سیستم در زمان  $t = 0$  توسط مهندسی تحت آزمایش قرار بگیرد. معمولاً در آزمون طول عمر، سیستم‌های تحت آزمون به طور پیوسته و مداوم مورد بررسی قرار نمی‌گیرند. حال اگر فرض شود در زمان  $t$  که مهندس سیستم‌ها را بررسی می‌کند، بعضی از آن‌ها از کار افتاده باشند، آنگاه مهندس می‌خواهد بداند که چه مدت زمانی است که دستگاه از کار افتاده است. چنانچه آنتروپی شانون برای متغیر تصادفی  $X(t)$  تعریف شود، آن را آنتروپی گذشته متغیر تصادفی  $X$  می‌نامند.

آنتروپی گذشته برای متغیر تصادفی  $X$  در زمان  $t$  با  $\bar{H}(t)$  نشان داده و به صورت

$$\bar{H}(t) = - \int_0^t \log f_t(x) dF_t(x) = 1 - \frac{1}{F(t)} \int_0^t f(x) \log \frac{f(x)}{F(t)} dx, \quad (۳)$$

تعریف می‌شود. اگر  $f(\cdot)$  نامعلوم باشد،  $\bar{H}(t)$  نیز نامعلوم است و باید با استفاده از داده‌ها برآورد شود. بدین منظور فرض می‌شود نمونه تصادفی از جامعه طول-اریب در دسترس است و در بخش ۳ بر اساس داده‌های طول-اریب دو برآوردگر برای  $\bar{H}(t)$  ارائه و سازگاری آن‌ها اثبات می‌شود.

### ۳ معرفی برآوردگرها

در این بخش دو برآوردگر برای آنتروپی مانده معرفی می‌کنیم. فرض کنید  $Y_1, \dots, Y_n$  نمونه‌ای تصادفی از توزیع  $G(\cdot)$  با تابع توزیع تجربی  $G_n(\cdot)$  باشد که برای  $t > 0$  به صورت  $G_n(t) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq t)$  تعریف می‌شود، که در آن  $I(A)$  تابع نشانگر پیشامد  $A$  است. براساس این نمونه تصادفی، **کاکس** (۱۹۶۹) برآوردگر تابع توزیع  $F(\cdot)$  را، که در رابطه (۲) تعریف شده است، به صورت

$$F_n(t) = \mu_n \int_0^t y^{-1} dG_n(y), \quad (۴)$$

ارائه داد، که در آن

$$\mu_n^{-1} = \int_0^\infty y^{-1} dG_n(y). \quad (۵)$$

سازگاری قوی برآوردگر  $F_n(\cdot)$  به  $F(\cdot)$  در **هوروات** (۱۹۸۵) اثبات شده است.

لم ۱. (هوروات، ۱۹۸۵) اگر  $0 < \mu < \infty$ ، آنگاه  $\lim_{n \rightarrow \infty} \sup_{0 < t < \infty} |F_n(t) - F(t)| \stackrel{a.s.}{=} 0$ .

اولین برآوردگری که برای  $\bar{H}(t)$  ارائه می‌شود برآوردگر جایگزین  $\bar{H}_n(t)$  است، که با جایگذاری

---

<sup>1</sup>Plug-in

$f_n(\cdot)$  و  $F_n(\cdot)$  به جای  $f(\cdot)$  و  $F(\cdot)$  در (۳) به صورت

$$\bar{H}_n(t) = - \int_0^t \frac{f_n(x)}{F_n(t)} \log \frac{f_n(x)}{F_n(t)} dx = 1 - \frac{1}{F_n(t)} \int_0^t f_n(x) \log \frac{f_n(x)}{F_n(t)} dx.$$

حاصل می‌شود، که در آن  $F_n(\cdot)$  در رابطه (۴) تعریف و  $f_n(\cdot)$  برآوردگر تابع چگالی با داده‌های طول-اریب است که توسط **جونز (۱۹۹۱)** به صورت

$$f_n(x) = \frac{\mu_n}{nh_n} \sum_{i=1}^n \frac{1}{Y_i} K\left(\frac{x - Y_i}{h_n}\right). \quad (۶)$$

معرفی شده است. در اینجا  $K(\cdot)$  را تابع هسته گویند که در شرط  $\int_{-\infty}^{\infty} K(x) dx = 1$  صدق می‌کند،  $h_n$  دنباله پهنای نوار است که وقتی  $n \rightarrow \infty$  به صفر میل می‌کند و  $\mu_n$  در رابطه (۵) تعریف شده است. دومین برآوردگر که با ایده گرفتن از **بلزونس و همکاران (۲۰۰۱)** ارائه می‌شود و آن را با  $\hat{H}_n(t)$  نشان می‌دهیم، با استفاده از رابطه (۳) به دست می‌آید. به جای  $F_t(\cdot)$  در رابطه (۳)، برآوردگر تجربی کاکس از رابطه (۴) را، فقط با در نظر گرفتن مشاهدات  $Y_i$  کمتر از  $t$  به دست آورده، در آن قرار می‌دهیم و برآوردگر

$$\hat{I}(t) = \int_0^t f(x) dF_n(x) = \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} \log f(Y_i) I(Y_i \leq t),$$

به دست می‌آید. چون  $f(\cdot)$  نامعلوم است برآوردگر آن  $f_n(\cdot)$  به جای آن قرار داده می‌شود که همان برآوردگر جونز (۶) است و بر اساس مشاهدات  $Y_i$  کمتر از  $t$  ساخته می‌شود

$$\hat{H}_n(t) = \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} \log f_n(Y_i) I(Y_i \leq t). \quad (۷)$$

**نکته ۱:** در **هال و مورتون (۱۹۹۳)** برای داده‌های کامل و غیر طول-اریب ثابت شده است که  $\hat{H}_n(t)$  یک برآوردگر اریب برای  $\bar{H}(t)$  است، پس بدیهی است که برای داده‌ها طول-اریب نیز این برآوردگر اریب است. با توجه به اینکه  $f_n(\cdot)$  به  $Y_i$  ها وابستگی دارد، به منظور کاهش اریبی برآوردگر  $\hat{H}_n(t)$ ، در شبیه‌سازی به

طور عملی از برآوردگر

$$\tilde{H}_n(t) = \frac{1}{h_n \sum_{j \neq i} Y_j^{-1}(t)} \sum_{\substack{j=1 \\ j \neq i}} Y_j^{-1}(t) \log f_{n,i}(Y_i(t)). \quad (۸)$$

استفاده می‌شود که تعدیل شده برآوردگر (۷) است و در آن  $Y_i(t) = Y_i \cdot I(Y_i \leq t)$  و  $f_{n,i}(\cdot)$  همان برآوردگر تابع چگالی در (۶) است که با  $n - 1$  مشاهده و بدون مشاهده  $i$ ام تشکیل شده است.

## ۴ سازگاری

در این بخش سازگاری قوی هر دو برآوردگر تحت فرض‌هایی مورد مطالعه قرار می‌گیرد.

**A۱** فرض کنید  $K$  یک تابع هسته متقارن، با تغییرات کراندار و حدود انتگرال‌گیری  $(-1, 1)$  باشد و داشته باشیم

$$\int_{-1}^1 K(x) dx = 1, \quad \int_{-1}^1 x K(x) dx = 0, \quad \int_{-1}^1 x^2 K(x) dx < \infty.$$

**A۲** وقتی  $t \rightarrow 0$  آنگاه  $f(t) = O(t)$ .

**A۳** برای  $r > 2$ ,  $\nu(r) = \int_0^\infty u^{-r} G^{1/r}(u) du < \infty$ .

**A۴**  $\inf_{0 < x < \tau} f(x) > 0$ .

**H** برای یک  $r > 2$  و  $0 < \lambda < \frac{1}{r} - \frac{1}{\lambda}$  وقتی  $n \rightarrow \infty$  داشته باشیم  $\frac{\log n}{n^{\frac{1}{r} + \lambda} h_n} = o(1)$ .

**نکته ۲:** فرض **A۱** فرض متعارفی برای برآورد تابع چگالی به روش هسته است. فرض‌های **A۲**، **A۳** و **H** در اثبات سازگاری برآورد  $f_n(\cdot)$  به کار رفته است که این فرض‌ها از **عجمی و همکاران (۲۰۱۳)** است. در قضیه ۲ برای اثبات سازگاری برآوردگر دوم به فرض **A۴** نیاز است. از فرض **H** می‌توان برای انتخاب پهنای نوار نیز استفاده کرد. همان‌طور که در نکته ۱ بیان شده  $h_n$  می‌تواند از مرتبه  $n^{-\beta}$  باشد که  $0 < \beta < \frac{1}{r} + \lambda$  است.

**قضیه ۱.** فرض کنید  $\tau = \sup \{x, F(x) < 1\} < \infty$  برای هر  $\varepsilon > 0$ ، با برقراری فرض‌های

$$\lim_{n \rightarrow \infty} \sup_{\varepsilon \leq t < \tau} |\bar{H}_n(t) - \bar{H}(t)| \stackrel{a.s.}{=} 0 \quad \text{داریم} \quad A1 - A3.$$

**برهان:** ابتدا داریم

$$\bar{H}_n(t) - \bar{H}(t) = (\log F_n(t) - \log F(t)) - \left( \frac{\mathbb{H}_n(t)}{F_n(t)} - \frac{\mathbb{H}(t)}{F(t)} \right) = J_{1n}(t) - J_{2n}(t)$$

که در آن

$$\begin{aligned} \mathbb{H}_n(t) &= \int_0^t f_n(x) \log f_n(x) dx, & \mathbb{H}(t) &= \int_0^t f(x) \log f(x) dx, \\ J_{1n}(t) &= (\log F_n(t) - \log F(t)), & J_{2n}(t) &= \left( \frac{\mathbb{H}_n(t)}{F_n(t)} - \frac{\mathbb{H}(t)}{F(t)} \right). \end{aligned}$$

با استفاده از بسط تیلور، رابطه  $\log F_n(t) - \log F(t) = \frac{F_n(t) - F(t)}{F_n^*(t)}$  را داریم، که در آن

$$\min\{F_n(t), F(t)\} < F_n^*(t) < \max\{F_n(t), F(t)\}, \quad \varepsilon < t \leq \tau.$$

بنابراین وقتی  $n \rightarrow \infty$  نتیجه می‌شود

$$\sup_{\varepsilon \leq t < \tau} |\log F_n(t) - \log F(t)| \leq F^{-1}(\varepsilon) \sup_{\varepsilon \leq t < \tau} |F_n(t) - F(t)| \stackrel{a.s.}{=} o(1),$$

برای بررسی جمله  $J_{2n}$ ، برای هر  $t \geq 0$  داریم

$$\begin{aligned} \frac{\mathbb{H}_n(t)}{F_n(t)} - \frac{\mathbb{H}(t)}{F(t)} &= \frac{\mathbb{H}_n(t) - \mathbb{H}(t)}{F(t)} - \frac{\mathbb{H}_n(t)(F_n(t) - F(t))}{F^2(t)} \\ &+ \frac{\mathbb{H}_n(t)(F_n(t) - F(t))^2}{F_n(t)F^2(t)}. \end{aligned} \quad (9)$$

ابتدا نشان می‌دهیم برای جمله اول در سمت راست رابطه (۹) داریم

$$\sup_{\varepsilon \leq t < \tau} \left| \frac{\mathbb{H}_n(t) - \mathbb{H}(t)}{F(t)} \right| \stackrel{a.s.}{=} o(1), \quad n \rightarrow \infty. \quad (10)$$

۳۷۰ ..... برآورد آنتروپی گذشته در داده‌های طول-اریب

بنابراین طبق نامساوی مثلث می‌توان نوشت

$$\begin{aligned}
 |\mathbb{H}_n(t) - \mathbb{H}(t)| &\leq \left| \int_0^t [f_n(x) \log f_n(x) - f_n(x) \log f(x)] dx \right| \\
 &+ \left| \int_0^t [f(x) \log f(x) - f_n(x) \log f(x)] dx \right| \\
 &\leq \int_0^t f(x) \left| \frac{f_n(x)}{f(x)} \log \frac{f_n(x)}{f(x)} \right| dx \\
 &+ \|f_n - f\| \int_0^t |\log f(x)| dx \\
 &= I_n(t) + I_{\nabla n}(t).
 \end{aligned}$$

که در آن  $I_n(t) = \|f_n - f\| \int_0^t |\log f(x)| dx$ ،  $I_{\nabla n}(t) = \int_0^t f(x) \left| \frac{f_n(x)}{f(x)} \log \frac{f_n(x)}{f(x)} \right| dx$  و  $\|\cdot\|$  به صورت  $\sup\{|f(x)| : x > 0\}$  تعریف می‌شود. ابتدا  $I_n(t)$  را در نظر می‌گیریم. چون برای  $z \geq 0$  داریم،  $|z \log z| \leq |z - 1|(1 + z)$  می‌توان نوشت

$$\left| \frac{f_n(x)}{f(x)} \log \frac{f_n(x)}{f(x)} \right| \leq \left| \frac{f_n(x)}{f(x)} - 1 \right| \left( 1 + \frac{f_n(x)}{f(x)} \right)$$

پس نتیجه می‌شود

$$\begin{aligned}
 I_n(t) &\leq \int_0^t f(x) \left| \frac{f_n(x)}{f(x)} - 1 \right| \left( 1 + \frac{f_n(x)}{f(x)} \right) dx \\
 &\leq \|f_n - f\| \int_0^t \left| 1 + \frac{f_n(x) + f(x) - f(x)}{f(x)} \right| dx \\
 &\leq \|f_n - f\| \left( 2t + \int_0^t \left| \frac{f_n(x) - f(x)}{f(x)} \right| dx \right) \\
 &\leq \tau \|f_n - f\| \left( 2 + \frac{\|f_n - f\|}{\inf_{0 \leq t \leq \tau} f(t)} \right) \xrightarrow{a.s.} 0.
 \end{aligned}$$

برای جمله  $I_{\nabla n}(t)$  با استفاده از فرض‌های  $A1 - A3$  و  $H$  و قضیه ۲ در [عجمی و همکاران \(۲۰۱۳\)](#) داریم  $I_{\nabla n}(t) \leq C \|f_n - f\| \xrightarrow{a.s.} 0$  بنابراین

$$\sup_{0 \leq t \leq \tau} |\mathbb{H}_n(t) - \mathbb{H}(t)| \stackrel{a.s.}{=} o(1) \quad n \rightarrow \infty. \quad (11)$$



با استفاده از همگرایی یکنواخت  $\mathbb{H}_n(\cdot)$  و لم ۱، برای قسمت‌های دوم و سوم رابطه (۹) به‌ترتیب داریم

$$\sup_{\varepsilon \leq t < \tau} \left| \frac{\mathbb{H}_n(t)(F_n(t) - F(t))}{F^\gamma(t)} \right| \leq \frac{1}{F^\gamma(\varepsilon)} \sup_{\varepsilon \leq t < \tau} |\mathbb{H}_n(t)| \|F_n - F\| \stackrel{a.s.}{=} o(1), \quad (12)$$

$$\sup_{\varepsilon \leq t < \tau} \left| \frac{\mathbb{H}_n(t)(F_n(t) - F(t))^\gamma}{F_n(t)F^\gamma(t)} \right| \leq \frac{1}{F^\gamma(\varepsilon)} \sup_{\varepsilon \leq t < \tau} |\mathbb{H}_n(t)| \|F_n - F\|^\gamma \stackrel{a.s.}{=} o(1), \quad (13)$$

بنابراین از رابطه‌های (۹)، (۱۰)، (۱۲) و (۱۳) می‌توان نتیجه گرفت

$$\sup_{\varepsilon \leq t < \tau} \left| \frac{\mathbb{H}_n(t)}{F_n(t)} - \frac{\mathbb{H}(t)}{F(t)} \right| \stackrel{a.s.}{=} o(1) \quad n \rightarrow \infty. \quad (14)$$

و با استفاده از رابطه‌های (۱۱)، (۱۳) و (۱۴) نتیجه  $\sup_{\varepsilon \leq t < \tau} |\bar{H}_n(t) - \bar{H}(t)| \xrightarrow{a.s.} 0$  به دست می‌آید، بنابراین اثبات کامل می‌شود.

**قضیه ۲.** تحت فرض‌های  $A1 - A4$  و  $H$  برآوردگر  $\hat{H}_n(\cdot)$  سازگار قوی یکنواخت است. یعنی

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq \tau} |\hat{H}_n(t) - \bar{H}(t)| \stackrel{a.s.}{=} 0, \quad (15)$$

**برهان:** از مشتق رابطه (۴) نسبت به  $t$  داریم  $dF_n(t) = \frac{\mu_n}{t} dG_n(t)$ . از طرفی طبق رابطه (۷) می‌توان نوشت

$$\begin{aligned} \hat{H}_n(t) &= \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} \log f_n(Y_i) I(Y_i \leq t) \\ &= \mu_n \int_0^t \frac{1}{y} \log f_n(y) dG_n(y) = \int_0^t \log f_n(y) dF_n(y). \end{aligned}$$

بنابراین

$$\begin{aligned} \hat{H}_n(t) - \bar{H}(t) &= \int_0^t (\log f_n(x) - \log f(x)) dF_n(x) \\ &\quad + \int_0^t \log f(x) dF_n(x) - \int_0^t \log f(x) dF(x) \\ &= S_{1n}(t) + S_{2n}(t) \end{aligned} \quad (16)$$

با استفاده از بسط تیلور تابع  $\log f$  و فرض‌های  $A_1 - A_3$  و  $H$  و کاربرد قضیه ۲ در **عجمی و همکاران** (۲۰۱۳) رابطه  $\lim_{n \rightarrow \infty} \sup_{0 \leq x \leq \tau} |\log f_n(x) - \log f(x)| \xrightarrow{a.s.} 0$  را داریم. چون دنباله  $F_n(\cdot)$  دنباله‌ای همگرای یکنواخت به  $F(\cdot)$  است، رابطه

$\sup_{0 < t < \tau} |S_n(t)| \leq \sup_{0 < t < \tau} |\log f_n(t) - \log f(t)| \sup_{0 < t < \tau} |F_n(t)| \rightarrow 0$  به دست می‌آید. با توجه به همگرایی یکنواخت  $F_n(\cdot)$  به  $F(\cdot)$  و این که  $\{F_n(x)\}$  دنباله‌ای صعودی است، با استفاده از قضیه گسترش هلی<sup>۱</sup> (سرفلینگ، ۱۹۸۰) برای هر  $t$  داریم

$$\int_0^{t^\pm} \log f(x) dF_n(x) \rightarrow \int_0^{t^\pm} \log f(x) dF(x)$$

که منظور از  $t^\pm = \lim_{\delta \rightarrow 0} (t \pm \delta t)$  است. با استفاده از لم ۳.۲.۸ در **چو و تیچر** (۱۹۸۸) می‌توان نتیجه گرفت  $\sup_{0 < t < \tau} |S_n(t)| \xrightarrow{a.s.} 0$ . اکنون بنابر (۱۶) رابطه (۱۵) حاصل می‌شود.

## ۵ مطالعه شبیه‌سازی

برای ارزیابی عملکرد برآوردگرهای معرفی شده  $\bar{H}_n(t)$  و  $\tilde{H}_n(t)$  (نسخه تعدیل شده  $\hat{H}_n(t)$ ) در این بخش یک مطالعه شبیه‌سازی انجام شده است. برای  $n = 50, 100, 200$  و برای سه گروه الف-توزیع بتا  $Beta(1, 1)$ ،  $Beta(0.5, 0.5)$ ،  $Beta(2, 2)$ ، ب-توزیع لگ-نرمال  $LN(0, 1)$ ،  $LN(0, 2)$ ،  $LN(0, 0.5)$  و ج-توزیع وایبل  $W(1, 1)$ ،  $W(2, 1)$ ،  $W(0.5, 1)$ . شبیه‌سازی انجام شده است. در توزیع بتا برای  $t$  مقادیر  $0.25$ ،  $0.5$  و  $0.75$  در توزیع لگ-نرمال مقادیر  $0.5$ ،  $1.5$  و  $2$  و در توزیع وایبل مقادیر  $3$ ،  $5$  و  $7$  را در نظر گرفته‌ایم. دلیل انتخاب این توزیع‌ها آن است که توزیع وایبل و لگ-نرمال دو توزیع با مقادیر مثبت، متداول و پرکاربرد در مبحث طول عمر هستند. توزیع بتا نیز به عنوان یک توزیع با دامنه محدود در بیشتر مقالات مورد استفاده قرار می‌گیرد. لازم به ذکر است هر دو پارامتر توزیع‌های بتا که در شبیه‌سازی به کار برده شده برابر در نظر گرفته شده‌اند، در نتیجه این توزیع‌ها همگی متقارن هستند. در حالاتی که توزیع بتا نامتقارن است، نیز شبیه‌سازی انجام شده که نتایج مشابه حالت متقارن است، بنابر این نتایج شبیه‌سازی در حالت نامتقارن در این بخش آورده نشده است. ضمناً باید توجه داشت که مقدار  $t$  نزدیک به صفر برای توزیع‌های مختلف باهم فرق دارد. در واقع انتخاب مقادیر  $t$  نزدیک به صفر یا  $t$  بزرگ برای توزیع‌های مختلف به ماهیت توزیع بستگی دارد. مثلاً در توزیع  $Beta(1, 1)$  که یک توزیع متقارن

<sup>1</sup>Helly's theorem

است، چارک اول، دوم و سوم به ترتیب برابر ۰/۲۵، ۰/۵، ۰/۷۵ است. یا در توزیع  $Beta(2, 2)$  این مقادیر به ترتیب برابر ۰/۳۳، ۰/۵، ۰/۶۷ است. یا در توزیع وایبل که یک توزیع مناسب طول عمر است چارک مرتبه اول برای توزیع  $W(1, 1)$  برابر ۰/۷۸، برای توزیع  $W(2, 1)$  برابر ۰/۴۷ و برای توزیع  $W(0.5, 1)$  برابر ۰/۶ است. در واقع اگر شکست در نزدیکی صفر واقع شود رخداد خوبی برای طول عمر نیست. با استفاده از رابطه (۱) می‌توان توزیع‌های طول-اریب بتا، لگ-نرمال و وایبل را به‌صورت زیر به‌دست آورد.

الف) اگر  $X \sim Beta(\alpha, \beta)$ ، آنگاه توزیع طول -اریب آن  $Beta(\alpha + 1, \beta)$  است.

ب) اگر  $X \sim LN(0, \sigma)$ ، آنگاه توزیع طول -اریب آن عبارت است از

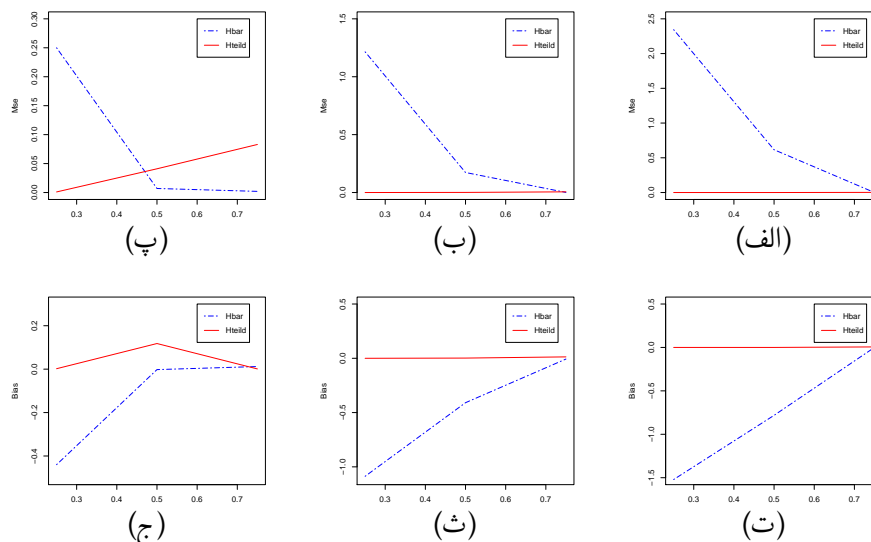
$$g(y) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[ -\frac{1}{2} \left( \frac{\ln y}{\sigma} \right)^2 \right] \exp \left( -\frac{\sigma^2}{2} \right), \quad y > 0.$$

ج) اگر  $X \sim W(\alpha, \beta)$ ، آنگاه توزیع طول -اریب آن به‌صورت

$$g(y) = \frac{\alpha\beta}{\Gamma(1 + \frac{1}{\alpha})} (\beta y)^\alpha \exp [-(\beta y)^\alpha], \quad y > 0, \alpha > 0, \beta > 0.$$

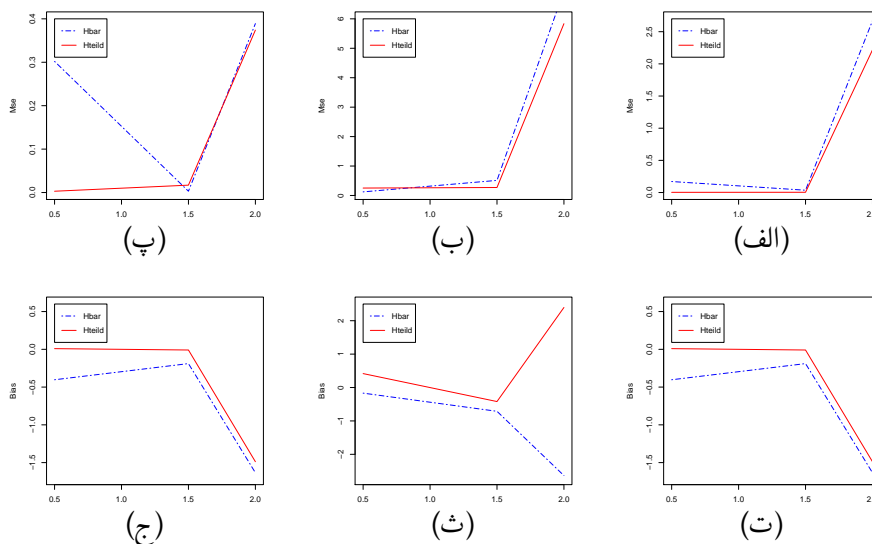
است، که توزیع گامای تعمیم‌یافته  $Y \sim GGa(1 + \frac{1}{\alpha}, \beta, \alpha)$  است (آندرس کورا و ولفسون، ۱۹۹۹).

در هر سه گروه توزیع، داده‌های طول-اریب شبیه‌سازی شده و مقدار واقعی  $\bar{H}(t)$  و دو برآوردگر  $\bar{H}_n(t)$  و  $\tilde{H}_n(t)$  برای آن محاسبه شده‌اند. مقادیر اریبی و  $MSE$  محاسبه شده و نتایج مربوط به توزیع بتا در جداول ۱، تا ۳ ارائه شده است. برای مقایسه بهتر اریبی و  $MSE$  دو برآوردگر، برای هر توزیع شکل رسم شده است. در این شکل‌ها رنگ آبی مربوط به  $\bar{H}_n(t)$  و رنگ قرمز مربوط به  $\tilde{H}_n(t)$  است. در جدول ۱ که مربوط به برآورد آنتروپی گذشته در داده‌های طول-اریب برای توزیع بتا است، نمی‌توان گفت یک برآوردگر بر دیگری ارجحیت دارد، بلکه گاهی  $\bar{H}_n(t)$  دارای اریبی و  $MSE$  کمتر و گاهی  $\tilde{H}_n(t)$  دارای اریبی و  $MSE$  کمتری است. به‌طور مثال به‌ازای ۰/۵، ۰/۲۵،  $t$ ، برآورد  $\tilde{H}_n(t)$  بهتر از  $\bar{H}_n(t)$  عمل کرده است ولی به‌ازای  $t = 0.75$  همه‌جا  $\bar{H}_n(t)$  از  $\tilde{H}_n(t)$  بهتر است. این نتایج در شکل ۱ به وضوح دیده می‌شود. در جدول ۲ که برآورد آنتروپی گذشته در داده‌های طول-اریب مربوط به توزیع لگ-نرمال است  $t = 0.5, 1.5, 2$  اختیار شده است که باز هم نمی‌توان به‌طور قطعی گفت کدام برآوردگر بهتر است. گاهی  $\bar{H}_n(t)$  دارای اریبی و  $MSE$  کمتر و گاهی  $\tilde{H}_n(t)$  دارای اریبی و  $MSE$  کمتری است. در واقع برای



شکل ۱. نمودار اریبی (پایین) و  $MSE$  (بالا) دو برآوردگر در توزیع بتا (الف) و (ت)  $Beta(2, 2)$ ، (ب) و (ث)  $Beta(1, 1)$ ، (پ) و (ج)  $Beta(0.5, 0.5)$ .

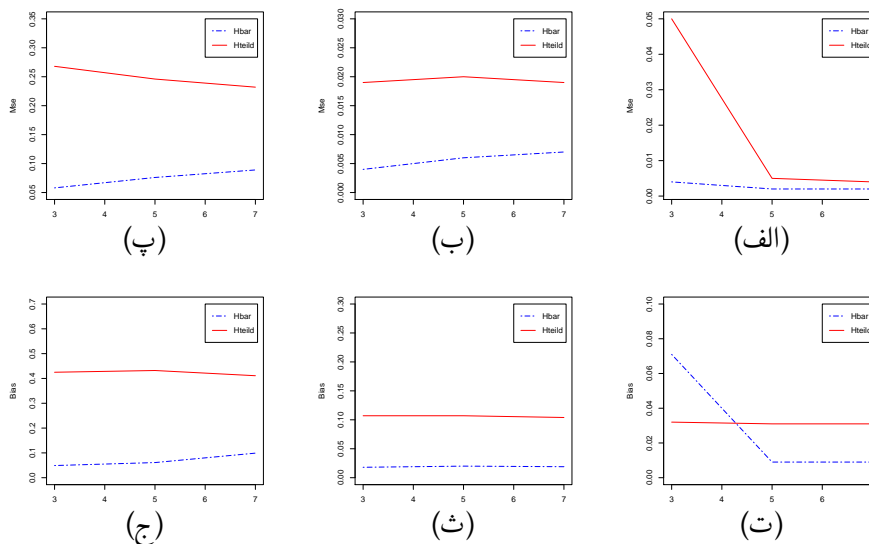
$t = 0.5$  که نزدیک صفر است برآوردگر  $\tilde{H}_n(t)$  اریبی و  $MSE$  کمتری دارد و برای مقادیر  $t = 1.5, 2$  که از صفر دورتر هستند برآوردگر  $\tilde{H}_n(t)$  نتایج بهتری دارد. نتایج مقایسه اریبی و  $MSE$  دو برآوردگر، در توزیع لگ-نرمال را می‌توان در شکل ۲ مشاهده کرد. جدول ۳ برای توزیع وایبل  $t = 3, 5, 7$  انتخاب شده که همه از صفر دور هستند. نتایج نشان می‌دهد که تقریباً همه‌جا برآوردگر  $\tilde{H}_n(t)$  بهتر از برآوردگر  $\tilde{H}_n(t)$  عمل می‌کند. نتایج مقایسه اریبی و  $MSE$  دو برآوردگر، در توزیع وایبل را می‌توان در شکل ۳ مشاهده کرد. طبق این شکل به جز یک مورد در توزیع  $W(2, 1)$ ، در همه جای دیگر برآوردگر  $\tilde{H}_n(t)$  از  $\tilde{H}_n(t)$  بهتر است. با توجه به نتایج شبیه‌سازی ملاحظه می‌شود که برای مقادیر  $t$  نزدیک به صفر برآوردگر  $\tilde{H}_n(t)$  و برای مقادیر دورتر از صفر  $\tilde{H}_n(t)$  بهتر عمل می‌کند. به عبارت دیگر نتایج شبیه‌سازی، مطالب اثبات شده در دو قضیه ۱ و ۲ را تایید می‌کند.



شکل ۲. نمودار اریبی (پایین) و  $MSE$  (بالا) دو برآوردگر در توزیع لگ-نرمال (الف) و (ت)  $LN(0, 1)$ ، (ب) و (ث)  $LN(0, 2)$ ، (پ) و (ج)  $LN(0, 0.5)$ .

## ۶ بحث و نتیجه‌گیری

دو برآوردگر  $\bar{H}_n(t)$  و  $\hat{H}_n(t)$  برای آنتروپی گذشته با داده‌های طول-اریب معرفی شدند، همچنین انگیزه و چگونگی ساختن آن‌ها نیز بیان گردید. سازگاری قوی هر دو برآوردگر تحت فرض‌هایی در دو قضیه اثبات گردید. در قضیه ۱ مقادیر  $t$  در بازه  $\varepsilon < t < \tau$  بود و انتظار داشتیم که  $\bar{H}_n(t)$  برای مقادیر  $t$  دورتر از صفر بهتر عمل کند، که با توجه به نتایج شبیه‌سازی این مطلب تایید گردید. در قضیه ۲ مقادیر  $t$  در بازه  $0 < t < \tau$  در نظر گرفته شد. با توجه به این قضیه انتظار داریم که برآوردگر  $\hat{H}_n(t)$  برای مقادیر  $t$  در نزدیکی صفر بهتر عمل کند، که این مطلب نیز در شبیه‌سازی تایید شد. برای ارزیابی عملکرد دو برآوردگر، مطالعه شبیه‌سازی انجام و با توجه به بحث **هال و مورتون (۱۹۹۳)** برای کاهش اریبی  $\hat{H}_n(t)$  از شکل اصلاح شده آن یعنی  $\tilde{H}_n(t)$ ، در شبیه‌سازی استفاده شد. بدین منظور سه توزیع مختلف بتا، لگ-نرمال و وایبل با پارامترهای مختلف در نظر گرفته و از نسخه طول-اریب آن‌ها نمونه‌های تصادفی تولید شد. مقدار  $\bar{H}(t)$  و دو برآورد آن به ازای مقادیر مختلف  $t$  در هر نمونه محاسبه و مقادیر اریبی و  $MSE$  هر دو برآوردگر محاسبه و نتایج در سه جدول جداگانه ارائه گردید. برای مقایسه بهتر نتایج شبیه‌سازی، شکل‌هایی



شکل ۳. نمودار اریبی (پایین) و  $MSE$  (بالا) دو برآوردگر در توزیع وایبل (الف) و (ت)  $W(2, 1)$ ، (ب) و (ث)  $W(1, 1)$ ، (پ) و (ج)  $W(0.5, 1)$ .

نیز برای اریبی و  $MSE$  رسم شد. خلاصه این تحقیق آن است که در برآورد آنتروپی گذشته در داده‌های طول-اریب، به ازای مقادیر  $t$  نزدیک به صفر بهتر است از  $\hat{H}_n(t)$  و برای مقادیر دور از صفر  $t$ ، بهتر است از  $\bar{H}_n(t)$  استفاده شود.

## تشکر و قدردانی

نویسندگان از پیشنهادهای ارزنده داوران و سردبیر محترم مجله که باعث اصلاحات سازنده در این مقاله شدند، کمال تشکر را دارند.

## مراجع

حوتی، ف.، احمدی، ج.، (۱۳۹۵)، آنتروپی مانده تجمعی پویای چندکی و مشخص‌سازی توزیع‌های یکنواخت، نمایی و پارتو، مجله علوم آماری، ۱۰(۱)، ۶۷-۸۰.

جدول ۱. مقادیر اربیی و  $MSE$  دو برآوردگر برای توزیع بتا

| Beta(۲, ۲) |        | Beta(۰.۵, ۰.۵) |        | Beta(۱, ۱) |        | $n$ | Es.           | $t$  |
|------------|--------|----------------|--------|------------|--------|-----|---------------|------|
| MSE        | Bias   | MSE            | Bias   | MSE        | Bias   |     |               |      |
| ۲,۰۸۱      | -۱,۳۸۵ | ۰,۳۳۶          | -۰,۴۳۹ | ۱,۲۳۷      | -۱,۰۵۸ | ۵۰  | $\hat{H}_n$   |      |
| ۲,۲۸۴      | -۱,۴۹۵ | ۰,۳۰۵          | -۰,۴۷۹ | ۱,۲۶۶      | -۱,۰۹۸ | ۱۰۰ |               |      |
| ۲,۳۴۴      | -۱,۵۲۱ | ۰,۲۵۰          | -۰,۴۴۰ | ۱,۲۱۵      | -۱,۰۸۷ | ۲۰۰ |               |      |
| ۰,۰۰۰      | ۰,۰۰۰  | ۰,۰۰۲          | ۰,۰۰۶  | ۰,۰۰۰      | ۰,۰۰۱  | ۵۰  | $\tilde{H}_n$ | ۰,۲۵ |
| ۰,۰۰۰      | ۰,۰۰۰  | ۰,۰۰۲          | ۰,۰۰۴  | ۰,۰۰۰      | ۰,۰۰۰  | ۱۰۰ |               |      |
| ۰,۰۰۰      | ۰,۰۰۰  | ۰,۰۰۱          | ۰,۰۰۲  | ۰,۰۰۰      | ۰,۰۰۰  | ۲۰۰ |               |      |
| ۰,۶۵۵      | -۰,۸۰۱ | ۰,۰۲۲          | -۰,۰۲۰ | ۰,۲۰۵      | -۰,۴۳۱ | ۵۰  | $\hat{H}_n$   |      |
| ۰,۶۳۰      | -۰,۷۸۹ | ۰,۰۱۲          | -۰,۰۲۳ | ۰,۱۸۷      | -۰,۴۱۹ | ۱۰۰ |               |      |
| ۰,۶۱۶      | -۰,۷۸۲ | ۰,۰۰۷          | -۰,۰۲۱ | ۰,۱۷۳      | -۰,۴۰۹ | ۲۰۰ |               |      |
| ۰,۰۰۰      | ۰,۰۰۰  | ۰,۱۰۷          | ۰,۲۶۰  | ۰,۰۰۲      | ۰,۰۰۶  | ۵۰  | $\tilde{H}_n$ | ۰,۵  |
| ۰,۰۰۰      | ۰,۰۰۰  | ۰,۰۶۶          | ۰,۱۷۱  | ۰,۰۰۱      | ۰,۰۰۳  | ۱۰۰ |               |      |
| ۰,۰۰۰      | ۰,۰۰۰  | ۰,۰۴۱          | ۰,۱۱۸  | ۰,۰۰۱      | ۰,۰۰۲  | ۲۰۰ |               |      |
| ۰,۰۰۲      | ۰,۰۰۱  | ۰,۰۱۱          | ۰,۰۰۴  | ۰,۰۰۴      | -۰,۰۱۴ | ۵۰  | $\hat{H}_n$   |      |
| ۰,۰۰۱      | ۰,۰۰۱  | ۰,۰۰۴          | ۰,۰۰۶  | ۰,۰۰۲      | -۰,۰۱۰ | ۱۰۰ |               |      |
| ۰,۰۰۰      | ۰,۰۰۰  | ۰,۰۰۲          | ۰,۰۱۲  | ۰,۰۰۱      | -۰,۰۰۷ | ۲۰۰ |               |      |
| ۰,۰۰۵      | ۰,۰۱۹  | ۰,۱۹۷          | ۰,۳۷۲  | ۰,۰۱۹      | ۰,۰۶۴  | ۵۰  | $\tilde{H}_n$ | ۰,۷۵ |
| ۰,۰۰۳      | ۰,۰۰۱  | ۰,۱۱۴          | ۰,۲۷۸  | ۰,۰۱۰      | ۰,۲۷۸  | ۱۰۰ |               |      |
| ۰,۰۰۲      | ۰,۰۰۶  | ۰,۰۸۳          | ۰,۲۴۰  | ۰,۰۰۶      | ۰,۰۱۳  | ۲۰۰ |               |      |

پورکاظمی، ع.، علیزاده نوغابی، ه.، جمهوری، س.، (۱۳۹۸)، برآورد آنتروپی با روش های بوت استرپ و جک نایف و کاربرد آن در آزمون نرمال بودن، مجله علوم آماری، ۱۳(۲)، ۳۳۹-۳۶۲.

Ajami, M., Fakoor, V., and Jomhoori, S. (2013), Some Asymptotic Results of Kernel Density Estimator in Length-Biased Sampling, *Journal of Sciences, Islamic Republic of Iran*, **24**(1), 55-62.

Andres Correa, J., and Wolfson, D. B. (1999), Length-Bias some Characterizations and Applications, *Journal of Statistical Computation and Simulation*, **64**(3), 209-219.

Belzunce, F., Guillamón, A., Navarro, J., and Ruiz, J. M. (2001), Kernel

جدول ۲. مقادیر اربیی و  $MSE$  دو برآوردگر برای توزیع لگ-نرمال

| LN(۰, ۲) |        | LN(۰, ۱) |        | LN(۰, ۰.۵) |        | $n$ | Es.           | $t$ |
|----------|--------|----------|--------|------------|--------|-----|---------------|-----|
| MSE      | Bias   | MSE      | Bias   | MSE        | Bias   |     |               |     |
| ۰.۳۷۲    | ۰.۴۸۳  | ۰.۱۳۶    | -۰.۲۸۷ | ۰.۱۳۳      | -۰.۰۸۱ | ۵۰  | $\hat{H}_n$   |     |
| ۰.۲۴۵    | ۰.۳۰۹  | ۰.۱۷۱    | -۰.۲۵۳ | ۰.۲۹۹      | -۰.۲۸۷ | ۱۰۰ |               |     |
| ۰.۱۲۲    | ۰.۱۶۸  | ۰.۱۷۱    | -۰.۴۰۳ | ۰.۳۰۲      | -۰.۳۲۲ | ۲۰۰ |               |     |
| ۰.۱۷۵    | ۰.۲۸۰  | ۰.۰۰۸    | ۰.۰۲   | ۰.۰۰۲      | ۰.۰۰۴  | ۵۰  | $\tilde{H}_n$ | ۰.۵ |
| ۰.۲۸۴    | ۰.۴۵۲  | ۰.۰۰۵    | ۰.۰۰۷  | ۰.۰۰۱      | ۰.۰۰۲  | ۱۰۰ |               |     |
| ۰.۲۵۲    | ۰.۴۱۷  | ۰.۰۰۳    | ۰.۰۰۹  | ۰.۰۰۳      | ۰.۰۰۱  | ۲۰۰ |               |     |
| ۰.۷۰۰    | -۰.۸۱۴ | ۰.۰۴۱    | -۰.۱۹۹ | ۰.۰۰۹      | ۰.۰۰۱  | ۵۰  | $\hat{H}_n$   |     |
| ۰.۵۷۲    | -۰.۷۴۲ | ۰.۰۳۸    | -۰.۱۹۳ | ۰.۰۰۵      | ۰.۰۰۲  | ۱۰۰ |               |     |
| ۰.۵۱۲    | -۰.۷۱۲ | ۰.۰۳۶    | -۰.۱۸۹ | ۰.۰۰۳      | ۰.۰۰۲  | ۲۰۰ |               |     |
| ۰.۴۵۳    | -۰.۴۹۳ | ۰.۰۱۶    | -۰.۰۱۶ | ۰.۰۵۱      | ۰.۱۳   | ۵۰  | $\tilde{H}_n$ | ۱.۵ |
| ۰.۳۵۱    | -۰.۴۳۲ | ۰.۰۰۷    | -۰.۰۱۹ | ۰.۰۲۹      | ۰.۰۶۸  | ۱۰۰ |               |     |
| ۰.۲۷۱    | -۰.۴۲۰ | ۰.۰۰۴    | -۰.۰۰۹ | ۰.۰۱۷      | ۰.۰۳۹  | ۲۰۰ |               |     |
| ۷.۷۲۰    | -۲.۷۶۸ | ۲.۷۰۳    | -۱.۶۴۳ | ۰.۳۹۴      | -۰.۶۲۸ | ۵۰  | $\hat{H}_n$   |     |
| ۷.۰۸۴    | -۲.۶۵۹ | ۲.۶۸۰    | -۱.۶۳۷ | ۰.۳۹۱      | -۰.۶۲۵ | ۱۰۰ |               |     |
| ۶.۹۵۰    | -۲.۶۳۶ | ۲.۶۷۰    | -۱.۶۳۴ | ۰.۳۸۹      | -۰.۶۲۴ | ۲۰۰ |               |     |
| ۶.۵۵۵    | -۰.۴۹۲ | ۲.۲۱۰    | -۱.۴۸۱ | ۰.۳۵۶      | -۰.۵۸۹ | ۵۰  | $\tilde{H}_n$ | ۲   |
| ۶.۰۱۸    | -۲.۴۰۷ | ۲.۲۱۴    | -۱.۴۸۵ | ۰.۳۶۲      | -۰.۵۹۸ | ۱۰۰ |               |     |
| ۵.۸۳۵    | -۲.۳۹۶ | ۲.۲۱۷    | -۱.۴۸۸ | ۰.۳۷۴      | -۰.۶۰۹ | ۲۰۰ |               |     |

Estimation of Residual Entropy, *Communications in Statistics-Theory and Methods*, **30**(7), 1243-1255.

Chow, Y. S. Y., and Teicher, H. (1988), *Probability Theory*. New York, Springer—Verlag.

Cox, D. R. (1969), *Some Sampling Problems in Technology, New Developments in Survey Sampling*, 506-527.

Di Crescenzo, A., and Longobardi, M. (2002), Entropy-Based Measure of Uncertainty in Past Lifetime Distributions, *Journal of Applied Probability*, **39**(2), 434-440.

Di Crescenzo, A., Di Gironimo, P., and Kayal, S. (2020), Analysis of the



جدول ۳. مقادیر اریبی و  $MSE$  دو برآوردگر برای توزیع وایبل

| Weibull(۲, ۱) |       | Weibull(۱, ۱) |        | Weibull(۰.۵, ۱) |         | $n$ | Es.           | $t$ |
|---------------|-------|---------------|--------|-----------------|---------|-----|---------------|-----|
| MSE           | Bias  | MSE           | Bias   | MSE             | Bias    |     |               |     |
| ۰/۰۴۹         | ۰/۰۴۵ | ۰/۰۱۰         | ۰/۰۳۰۰ | ۰/۰۶۱           | ۰/۰۳۶   | ۵۰  | $\hat{H}_n$   |     |
| ۰/۰۴۷         | ۰/۰۶۲ | ۰/۰۰۷         | ۰/۰۲۳  | ۰/۰۶۱           | ۰/۰۴۲   | ۱۰۰ |               |     |
| ۰/۰۴۰         | ۰/۰۷۱ | ۰/۰۰۴         | ۰/۰۱۸  | ۰/۰۵۸           | ۰/۰۴۹   | ۲۰۰ |               |     |
| ۰/۰۱۷         | ۰/۰۶۸ | ۰/۰۶۱         | ۰/۰۱۹۶ | ۰/۰۱۵۴          | -۰/۰۱۸  | ۵۰  | $\tilde{H}_n$ | ۳   |
| ۰/۰۰۹         | ۰/۰۴۹ | ۰/۰۳۸         | ۰/۰۱۵۳ | ۰/۰۴۳۵          | ۰/۰۶۱۳  | ۱۰۰ |               |     |
| ۰/۰۰۵         | ۰/۰۳۲ | ۰/۰۱۹         | ۰/۰۱۰۷ | ۰/۰۲۶۸          | ۰/۰۴۲۵  | ۲۰۰ |               |     |
| ۰/۰۰۵         | ۰/۰۱۵ | ۰/۰۱۶         | ۰/۰۲۸  | ۰/۰۸۶           | ۰/۰۳۱   | ۵۰  | $\hat{H}_n$   |     |
| ۰/۰۰۳         | ۰/۰۱۲ | ۰/۰۱۰         | ۰/۰۲۳  | ۰/۰۸۶           | ۰/۰۴۴   | ۱۰۰ |               |     |
| ۰/۰۰۲         | ۰/۰۰۹ | ۰/۰۰۶         | ۰/۰۲۰  | ۰/۰۷۶           | ۰/۰۶۱   | ۲۰۰ |               |     |
| ۰/۰۱۷         | ۰/۰۶۸ | ۰/۰۶۱         | ۰/۰۱۹۳ | ۰/۰۱۵۶          | -۰/۰۷۵  | ۵۰  | $\tilde{H}_n$ | ۵   |
| ۰/۰۰۹         | ۰/۰۴۸ | ۰/۰۳۷         | ۰/۰۱۵۱ | ۰/۰۳۹۵          | ۰/۰۵۸۰  | ۱۰۰ |               |     |
| ۰/۰۰۵         | ۰/۰۳۱ | ۰/۰۲۰         | ۰/۰۱۰۷ | ۰/۰۲۴۶          | ۰/۰۴۳۲  | ۲۰۰ |               |     |
| ۰/۰۰۵         | ۰/۰۱۶ | ۰/۰۱۷         | ۰/۰۲۷  | ۰/۰۱۰۵          | ۰/۰۲۶   | ۵۰  | $\hat{H}_n$   |     |
| ۰/۰۰۳         | ۰/۰۱۸ | ۰/۰۱۱         | ۰/۰۲۳  | ۰/۰۱۰۳          | ۰/۰۶۰   | ۱۰۰ |               |     |
| ۰/۰۰۲         | ۰/۰۰۹ | ۰/۰۰۷         | ۰/۰۱۹  | ۰/۰۸۹           | ۰/۰۹۹   | ۲۰۰ |               |     |
| ۰/۰۱۷         | ۰/۰۶۸ | ۰/۰۵۹         | ۰/۰۱۸۳ | ۰/۰۱۷۵          | -۰/۰۱۳۸ | ۵۰  | $\tilde{H}_n$ | ۷   |
| ۰/۰۰۹         | ۰/۰۴۸ | ۰/۰۳۶         | ۰/۰۱۴۵ | ۰/۰۳۷۵          | ۰/۰۵۵۷  | ۱۰۰ |               |     |
| ۰/۰۰۴         | ۰/۰۳۱ | ۰/۰۱۹         | ۰/۰۱۰۴ | ۰/۰۲۳۲          | ۰/۰۴۱۱  | ۲۰۰ |               |     |

Past Lifetime in a Replacement Model through Stochastic Comparisons and Differential Entropy, *Mathematics*, **8**(8), 1203.

Gupta, R. C. (2009), Some Characterization Results Based on Residual Entropy Function, *Journal of Statistical Theory and Applications*, **8**(1), 45-59.

Gupta, R. D., and Nanda, A. K. (2002),  $\alpha$ - and  $\beta$ - Entropies and Relative Entropies of Distributions, *Journal of Statistical Theory and Applications*, **1**(3), 177-190.

Hall, P., and Morton, S. (1993), On the Estimation of Entropy, *Annals of the Institute of Statistical Mathematics*, **45**(1), 69-88.

- Horváth, L. (1985), Estimation from a Length-Biased Distribution, *Statistics and Decisions*, **3**, 91-113.
- Jones, M. C. (1991), Kernel Density Estimation for Length Biased Data, *Biometrika*, **78**(3), 511-519.
- Maya, R. (2013), Kernel Estimation of the Past Entropy Function with Dependent Data, *Journal of the Kerala Statistical Association*, **24**, 12-36.
- Maya, R., Abdul-Sathar, E. I., and Rajesh, G. (2014), Non-Parametric Estimation of the Generalized Past Entropy Function with Censored Dependent Data, *Statistics and Probability Letters*, **90**, 129-135.
- Nanda, A., and Paul, P. (2006), Some Properties of Past Entropy and their Applications, *Metrika*, **64**(1), 47-61.
- Qin, J. (2017), *Biased Sampling, Over-Identified Parameter Problems and Beyond*, Singapore Springer.
- Sachlas, A., and Papaioannou, T. (2014), Residual and Past Entropy in Actuarial Science and Survival Models, *Methodology and Computing in Applied Probability*, **16**(1), 79-99.
- Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons.
- Tavangar, M., and Asadi, M. (2010), A Study on the Mean Past Lifetime of the Components of  $(nk + 1)$ -out-of- $n$  System at the System Level, *Metrika*, **72**(1), 59-73.
- Wicksell, S. D. (1926), The Corpuscle Problem Second Memoir Case of Ellipsoidal Corpuscles, *Biometrika*, **72**, 151-172.

## Estimation of Past Entropy in Length-Biased data

Farzaneh Oliazadeh<sup>1</sup>, Anis Iranmanesh<sup>1</sup>, Vahid Fakoor<sup>2</sup>

<sup>1</sup>Department of Mathematics and Statistics, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

<sup>2</sup>Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran.

**Abstract** In this article, we propose two non-parametric estimators for the past entropy based on length-biased data and the strong consistency of the proposed estimators is proved. In addition, some simulations are conducted to evaluate the performance of the proposed estimators. Based on the results, we show that they have better performance in a different region of the probability distribution for length biased random variables.

**Keywords:** Length-biased data; Kernel density estimation; Past entropy; Simulation, Strong consistency

**Mathematics Subject Classification (2010):** 94A17, 62B10.