آزمون نیکویی برازش طول عمر برای توزیع رایلی بر اساس اطلاع کولبک-لیبلر مانده‌ی تجمعی

هدیه افتخاری مودی،محمد خراشادی‌زاده،هادی علیزاده نوقابی

دانشگاه بیرجند

۲۸ اردیبهشت ۱۴۰۱

مقدمه

در این مقاله ابتدا ضمن معرفی برخی از تعمیم‌های اطلاع کولبک–لیبلر و خصوصاً اطلاع کولبک–لیبلر مانده ی تجمعی و خواص آن‌ها به ارائه آزمون نیکویی برازش طول عمر برای توزیع رایلی بر اساس اطلاع کولبک–لیبلر مانده ی تجمعی پرداخته می شود. در ادامه برای یک مجموعه داده واقعی در مورد دفعات شکست بلبرینگ در یک آزمون استقامت، مقادیر بحرانی و توان آزمون‌های پیشنهادی محاسبه و با توان سایر آزمون‌ها مقایسه می‌شود.

مفاهیم و تعاریف اولیه

تعریف (آنتروپی شانون)

فرض کنید متغیر تصادفی X دارای تابع توزیع $F(x)$ و تابع چگالی پیوسته‌ی $f(x)$ باشد، تعمیم آنتروپی شانون به صورت

$$H(f) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx,$$

تعریف می‌شود.

مفاهیم و تعاریف اولیه

تعریف (آنتروپی ماندهی تجمعی)

آنتروپی ماندهی تجمعی CRE [1] توسط رائو و همکاران با جایگزینی تابع بقا $(\bar{F})$ به جای تابع چگالی در آنتروپی شانون معرفی شد که به فرم

$$CRE(F) = -\int_{-\infty}^{+\infty} \bar{F}(x) \log \bar{F}(x) dx,$$

است.

[1] *Cumulative Residual Entropy*

مفاهیم و تعاریف اولیه

تعریف (آنتروپی تجمعی)

دیکرشنزو و لانگوباردی معیار دیگری تحت عنوان آنتروپی تجمعی CE [1] ارائه دادند که با جایگزین کردن تابع توزیع به جای تابع چگالی در آنتروپی شانون حاصل می‌شود و به شکل

$$CE(F) = -\int_{0}^{+\infty} F(x)\log F(x)dx,$$

تعریف می‌شود.

---

[1] Cumulative Entropy

مفاهیم و تعاریف اولیه

تعریف (اطلاع کولبک-لیبلر)

اطلاع کولبک-لیبلر (KL) [۱] اندازه اطلاع اختلاف بین $g(x)$ و $f(x)$ است که به صورت

$$KL(g:f) = \int_{-\infty}^{+\infty} g(x) \log \frac{g(x)}{f(x)} dx,$$

تعریف می‌شود.

---

[۱] *Kullback-Leibler Information*

مفاهیم و تعاریف اولیه

تعریف (اطلاع کولبک-لیبلر ماندهی تجمعی)

براتپور و راد اخیراً تعمیمی از اطلاع کولبک-لیبلر را با استفاده از تابع بقا پیشنهاد دادهاند که اطلاع کولبک-لیبلر ماندهی تجمعی $(CRKL)$ نامیده میشود[۱] و به صورت

$$CRKL(G:F)=\int_{0}^{\infty}\bar{G}(x)\log\frac{\bar{G}(x)}{\bar{F}(x)}dx-(E(Y)-E(X)), \qquad (1)$$

قابل تعریف است.

[۱] *Cumulative Residual Kullback-Leibler Information*

مفاهیم و تعاریف اولیه

تعریف (اطلاع کولبک-لیبلر تجمعی)

همچنین می‌توان مستقیماً تعمیم دیگری برای اطلاع کولبک لیبلر با استفاده از تابع توزیع تجمعی بیان کرد که اطلاع کولبک-لیبلر تجمعی CKL [۱] نامیده می‌شود و به شکل

$$CKL(G:F) = \int_{0}^{\infty} G(x) \log \frac{G(x)}{F(x)} dx + E(Y) - E(X), \qquad (۲)$$

بیان می‌شود.

[۱] *Cumulative Kullback-Leibler Information*

نتایج

قضیه

طبق شرط

$$\int_{-\infty}^{+\infty} g(x)\log f(x)dx = \int_{-\infty}^{+\infty} f(x)\log f(x)dx,$$

می‌توان اطلاع کولبک-لیبلر را به صورت

$$KL(g:f) = H(f) - H(g),$$

نوشت.

نتایج

قضیه

اگر شرط

$$\int_{\circ}^{\infty} \bar{G}(x) \log \bar{F}(x) dx + E(Y) = \int_{\circ}^{\infty} \bar{F}(x) \log \bar{F}(x) dx + E(X),$$

برقرار باشد، آنگاه

$$CRKL(G;F) = CRE(F) - CRE(G). \tag{۳}$$

مثال

فرض کنید $F(x) = ۱ - \exp(\frac{-x^۲}{۲\theta^۲})$ آنگاه با توجه به رابطه‌ی (۳) داریم،

$$\frac{-۱}{۶\theta^۲}E_G(X^۳) + E_G(X) = \frac{-۱}{۶\theta^۲}E_F(X^۳) + E_F(X). \qquad (۴)$$

رابطه‌ی (۴) به ازای هر θ برقرار است اگر $E_G(X^۳) = E_F(X^۳)$ و $E_G(X) = E_F(X)$ باشد. توزیع رایلی با پارامترθ، آنتروپی مانده‌ی تجمعی (CRE) را با شرط $E(X^۳) - ۶\theta^۲E(X) + ۳\theta^۳\sqrt{\frac{\pi}{۲}} = ۰$ در بین توزیع های مطلقاً پیوسته با تکیه گاه $(۰,\infty)$، ماکسیمم می‌کند. بنابراین متغیر تصادفیX، آنتروپی مانده‌ی تجمعی (CRE) را در بین تمام متغیرهای تصادفی مطلقاً پیوسته‌ی نامنفی Y، ماکسیمم می‌کند با این شرط که اگر $E(Y) = v$ و $E(Y^۳) = w$ باشد، داشته باشیم $\theta^۲ = \frac{w}{۳v}$.

تعریف

تعریف

فرض کنید $G(x)$ و $F(x)$ به ترتیب تابع توزیع تجربی $(F_n(x))$ و یک تابع توزیع پارامتری $(F_\theta(x))$ باشند، آنتروپی ماندهٔ تجمعی بر اساس این دو تابع به صورت

$$CRE(F_n) = -\sum_{i=0}^{n} \frac{n-i}{n} \log \frac{n-i}{n} (x_{i+1} - x_i), \qquad (5)$$

به دست می‌آید.

تعریف

تعریف

اطلاع کولبک-لیبلر مانده‌ی تجمعی $(CRKL)$ بین F_n و F_θ به صورت
(۶)

$$CRKL(F_n, F_\theta) = -CRE(F_n) - \sum_{i=0}^{n} \frac{n-i}{n} \int_{x_i}^{x_{i+1}} \log \bar{F}_\theta(x) dx - (\bar{x} - E_{F_\theta}(X)),$$

است.

می‌توان اطلاع کولبک–لیبلر مانده‌ی تجمعی $(CRKL)$ برای توزیع رایلی را با تقسیم بر پارامتر مقیاسی برآورد شده به صورت

$$T_۱ = \frac{۱}{\hat{\theta}_۱} CRKL(F_n, F_{\hat{\theta}_۱})$$
$$= \frac{۱}{\sqrt{\frac{\sum_{i=۱}^{n} x^۲}{۲n}}} \left(\frac{\sum_{i=۱}^{n} {x_i}^۳}{۳\sum_{i=۱}^{n} {x_i}^۲} - \bar{x} + \left(\sqrt{\frac{\sum_{i=۱}^{n} {x_i}^۲}{۲n}} \sqrt{\frac{\pi}{۲}}\right) - CRE(F_n) \right), \quad (۷)$$

نوشت.

حال باید برآوردگری را در نظر بگیریم که $\frac{CRKL(F_n,F_\theta)}{\theta}$ را مینیمم کند. برآوردگر مینیمم کننده برای توزیع رایلی به صورت

$$\hat{\theta} = \left(\frac{\sum_{i=۱}^{n} {x_i}^{۳}}{۲n(\bar{x} + CRE(F_n))} \right)^{\frac{۱}{۲}},$$

حاصل می‌شود و آماره آزمونی که براساس برآوردگر فوق به دست می‌آید برابر

$$T_۲ = \frac{-۲(\overline{x} + CRE(F_n))}{۳\left(\frac{\sum_{i=۱}^{n} x_i^{۳}}{۲n(\overline{x}+CRE(F_n))} \right)^{\frac{۱}{۲}}} + \sqrt{\frac{\pi}{۲}}, \qquad (۸)$$

است.

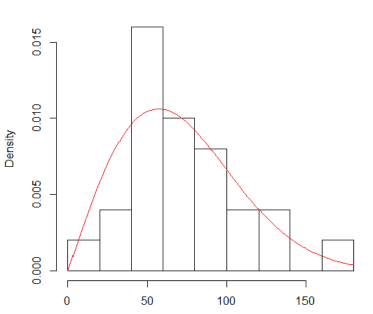
مثال

داده‌های ارائه شده در جدول ۱، را در نظر می‌گیریم. این داده‌ها مربوط به تعداد دفعات شکست ۲۵بلبرینگ در آزمون استقامت است که توسط کارونی ارائه و تحلیل شده است. نویسندگانی مانند براتپور و خدادادی، جهانشاهی و همکاران و صفوی‌نژاد و همکاران نیز این داده‌ها را تجزیه و تحلیل کرده‌اند. به علت استفاده زیاد نویسندگان از این داده‌ها ما نیز این داده‌ها را به کار می‌بریم.
هیستوگرام مجموعه داده‌ها، در شکل ۱ نشان داده شده است.

جدول: تعداد دفعات شکست ۲۵ بلبرینگ در آزمون استقامت

۱۷/۸۸	۲۸/۹۲	۳۳/۰۰	۴۱/۵۲	۴۲/۱۲	۴۵/۶۰	۴۸/۴۸	۵۱/۸۴	۵۱/۹۶	۵۴/۱۲
۵۵/۵۶	۶۷/۸۰	۶۷/۸۰	۶۷/۸۰	۶۸/۶۴	۸۶/۶۴	۶۸/۸۸	۸۴/۱۲	۹۳/۱۲	۹۸/۶۴
۱۰۵/۱۲	۱۰۵/۸۴	۱۲۷/۹۲	۱۲۸/۰۴	۱۷۳/۴۰					

مثال



شکل: هیستوگرام تعداد دفعات شکست بلبرینگ‌ها

مثال

در اینجا از آماره آزمون‌های (۷) و (۸) برای بررسی این که آیا داده‌ها از توزیع رایلی پیروی می‌کنند یا خیر استفاده شده است. برآورد درستنمایی ماکسیمم پارامتر θ به صورت

$$\hat{\theta} = ۵۷/۰۷۶,$$

است. مقادیر بحرانی، مقادیر آماره آزمون‌ها و همچنین p-مقدار تعریف شده به ازای آماره آزمون‌های (۷)، (۸)، آزمون کولموگروف-اسمیرنوف (D)، آزمون کوپر (V)، آزمون کرامر-وان میسز $(W^۲)$، آزمون واتسون $(U^۲)$، و آزمون اندرسون-دارلینگ $(A^۲)$ در جدول ۲ ارائه شده‌اند.

مثال

جدول: مقادیر بحرانی، مقادیر آماره و p-مقدار آزمون‌ها

p-مقدار	آماره آزمون	مقدار بحرانی	
۰/۶۲۰	۰/۱۲۳	۰/۲۰۹	D
۱/۰۰۰	۰/۲۳۵	۰/۳۱۸	V
۰/۷۹۹	۰/۰۲۵	۰/۲۱۷	W^{2}
۰/۶۱۹	۰/۰۵۱	۰/۱۵۸	U^{2}
۰/۸۲۱	۰/۳۴۲	۱/۲۸۹	A^{2}
۰/۹۲۶	۰/۰۱۳۰۹	۰/۰۵۲۱	T_{1}
۰/۹۲۱	۰/۰۱۳۰۸	۰/۰۴۹۲	T_{2}

مراجع

- Baratpour, S., & Khodadadi, F., (2012) . A cumulative residual entropy characterization of the Rayleigh distribution and related goodness-of-fit test. *Journal of Statistical Research of Iran*, 9 , 115-131 .
- Baratpour, S., & Rad, A.H., (2012) . Testing goodness-of fit for exponential distribution based on cumulative residual entropy. *Communications in Statistics Theory and Methods*, 41 , 1387-1396 .
- Caroni, C., (2002) . The correct ball bearing data. *Lifetime Data Analysis*, 8 , 395-399 .