آشنایی با جک نایف

حسینعلی نیرومند*

چکیده

در این مقاله روش جک نایف را به کمک چند مثال معرفی می‌کنیم و نشان می‌دهیم چگونه گاهی می‌توان اریبی برآورد کننده‌ها را کاهش داد.

مقدمه

یکی از اهداف اساسی جک نایف کاهش اریبی برآوردگرها و همچنین برآورد دقت (واریانس) یک برآوردگر و سایر ملاکهای خطاست.

روش جک نایف ابتدا توسط کوئنویل در سال ۱۹۴۹ برای حذف اریبی در همبستگی پیاپی مرتبهٔ اول در سریهای زمانی پیشنهاد شد. وی در سال ۱۹۵۶ نظریه‌ای کلی را توسعه داد که این نظریه، استخوانبندی روش جک نایف است. برای آشنایی با روش نخست مثالی را که در آن جک نایف ضرورتی ندارد، معرفی می‌کنیم.

مثال ۱. فرض کنید $X_1, \ldots, X_n$ متغیرهای تصادفی همتوزیع از جامعه F و $\theta(F)$ میانگین F است. اگر $\hat{\theta} = \bar{X}$ میانگین نمونه‌ای برآوردگر θ و $\frac{\sigma^2}{n}$ (σ^2 واریانس F است) دقت برآوردگر باشد، برآوردگر این دقت خود برابر است با

$$\left(\frac{\hat{\sigma}}{n}\right) = \frac{1}{n(n-1)} \sum (X_i - \bar{X})^2$$

در این حالت محاسبه $(\frac{\sigma^2}{n})$ آسان است امّا محاسبه واریانس و برآورد واریانس سایر مشخصه برآورد شدهٔ F چندان روشن نیست. مثلاً اگر $\theta(F)$ میانه F باشد و برآورد آن را میانه نمونه اختیار کنیم برآورد دقت $\hat{\theta}$ را چگونه برآورد کنیم؟ در این حالت معلوم است که کار بسیار مشکل است. قاعدهٔ زیر روشی برای پاسخ به این مشکل است.

فرض کنید $\hat{\theta}_i = \hat{\theta}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ برآوردگری برای θ بر اساس $X_1, \ldots, X_n$ باشد که X_i از آن حذف شده است. تعریف می‌کنیم $\hat{\theta}(\cdot) = \frac{1}{n}\sum \hat{\theta}(i)$، آنگاه دقت $\hat{\theta}$ را می‌توان به صورت زیر برآورد کرد:

$$\hat{\mathrm{Var}}(\hat{\theta}) = \frac{n-1}{n} \sum (\hat{\theta}(i) - \hat{\theta}(\cdot))^2$$

روش ساده بالا حالتی از روش جک نایف است. برآورد واریانس $\bar{X}$ به روش پیش وقتی $\hat{\theta}(i)$ میانگین نمونه‌ای حاصل از کل نمونه بجز مشاهدهٔ X_i است برابر است با

$$\hat{\theta}(i) = \hat{\theta}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

$$= \frac{1}{n-1} \sum_{j \neq i} X_j = \frac{n\bar{X} - X_i}{n-1}$$

که در نتیجه $\hat{\theta}(\cdot)$ عبارت است از

$$\hat{\theta}(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}(i) = \bar{X} = \hat{\theta}$$

و دقت برآورد شده به روش حاضر عبارت خواهد بود از

*) گروه آمار، دانشکدهٔ علوم ریاضی، دانشگاه فردوسی مشهد

$$\hat{\mathrm{Var}}(\hat\theta(\cdot))=\frac{n-1}{n}\sum(\hat\theta(i)-\hat\theta)^2$$
$$=\frac{n-1}{n}\sum_{i=1}^{n}\frac{(X_i-\bar X)^2}{(n-1)^2}$$
$$=\frac{1}{n(n-1)}\sum(X_i-\bar X)^2$$

می‌بینیم که همان ضابطه‌های کلاسیک به دست می‌آیند.

حال اگر طبق معمول اریبی $\hat\theta$ برابر $E(\hat\theta)-\theta=b_\theta(\hat\Theta)$ باشد. می‌خواهیم این اریبی را برآورد کنیم. $\hat\Theta$ برآوردگری برای θ است. برآورد اریبی را به صورت زیر تعریف می‌کنیم:

$$\hat b(\hat\theta)=(n-1)(\hat\theta(\cdot)-\hat\theta)$$

به کمک این اریبی برآورد شده می‌توان برآورد جک نایف θ را تعریف کرد. چنانچه این کمیت را از برآورد اصلی که $\hat\theta$ است کم کنیم خواهیم داشت:

$$\tilde\theta=\hat\theta-\hat b(\hat\theta)$$
$$=n\hat\theta-(n-1)\hat\theta(\cdot)$$

$\tilde\theta$ همان برآورد جک نایف θ است. اریبی آن کمتر از $\hat\theta$ است. (چرا؟) برای آنکه علت درستی ادعای اخیر را دریابیم قدری با مبنای منطقی روش جک نایف آشنا می‌شویم.

فرض کنید برای امید ریاضی برآوردکنندهٔ $\hat\theta$ بتوانیم بنویسیم:

$$E_F(\hat\theta)=\theta+\frac{a_1(F)}{n}+\frac{a_2(F)}{n^2}+\cdots=\theta+\frac{a_1(F)}{n}+O(\frac{1}{n^2})$$

که $a_i(F)$ به n بستگی ندارند. برای چنین برآوردگرهایی چنانچه بسط فوق را روی برآورد جک نایف خورده و اجزای سازندهٔ آن اعمال کنیم، خواهیم داشت:

$$E[\hat\theta(i)]=\theta+\frac{a_1(F)}{n-1}+O(\frac{1}{n^2})$$
$$E(\hat\theta(\cdot))=\theta+\frac{a_1(F)}{n-1}+O(\frac{1}{n^2})$$

و از آنجا $E(\tilde\theta)=nE_F(\hat\theta)-(n-1)E(\hat\theta(\cdot))=\theta+O(\frac{1}{n^2})$

می‌بینیم اریبی مرتبهٔ $\frac{1}{n}$ حذف شده است. یعنی ادعای کاهش اریبی اثبات شده است.

مثال ۲. فرض کنید $\hat\theta=\bar X$ و $\hat\theta(i)=\frac{1}{n-1}\sum_{j\neq i}X)j$. در نتیجه همان طور که دیدیم $\hat\theta(\cdot)$ همان $\bar X$ خواهد شد. می‌دانیم این برآوردگر نااریبی است. این نکته توسط برآورد آن نیز تایید می‌شود؛

برآوردگر اریبی $=(n-1)(\hat\theta(\cdot)-\hat\theta)=0$

برآورد کننده حاصل از جک نایف واضح است که همان $\bar X$ خواهد شد.

مثال ۳. فرض کنید θ واریانس F باشد و $\hat\theta=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$، در این حالت اریبی برابر است با:

$$E(\hat\theta)-\theta=\frac{n-1}{n}\theta-\theta=-\frac{\theta}{n}$$

و $\hat\Theta(i)=\frac{1}{n-1}\sum_{j\neq i}(X_j-\bar X_{(i)})^2$ که درآن $\bar X_{(i)}$ میانگین همه مشاهدات بجز X_iام است و به علاوه $\bar X_{(i)}=\frac{n\bar X-X_i}{n-1}$ در نتیجه

$$\hat\theta_{(i)}=\frac{1}{n-1}\sum_{j\neq i}X_j^2-[\bar X_{(i)}]^2$$
$$=\frac{1}{n-1}\sum_{j\neq i}X_j^2-\frac{1}{(n-1)^2}[n^2\bar X^2-2nX_i\bar X+X_i^2]$$

به سادگی با توجه به محاسبات بالا دیده می‌شود که

$$\hat\theta(\cdot)=\frac{n^2-2n}{(n-1)^2}\left[\frac{1}{n}\sum X_i^2-\bar X^2\right]=\frac{n^2-2n}{(n-1)^2}\hat\theta$$

بنابراین برآورد اریبی برابر است با:

$$(n-1)(\hat\theta(\cdot)-\hat\theta)=\frac{1}{n(n-1)}\sum(X_i-\bar X)^2$$

که از آنجا برآورد جک نایف شده برابر خواهد بود با:

$$\tilde\theta=n\hat\theta-(n-1)\hat\theta(\cdot)=\frac{1}{n-1}\sum(X_i-\bar X)^2$$

یعنی همان برآوردگر نااریب معمول حاصل شده است.

مراجع

[1] Quenouille, M. (1949), Approximate tests of Correlation in time series, JRSS, B. 11. 68-84.

[2] Quenouille, M. (1956), Notes on Bias in Estimation, Biometrika, 43, 353-360.

[3] Schucany, W. R. Gray, H. L. & Own, D. B. (1971), On Bias Reduction in Estimation, JASA, GG, 524- 533.