استقلال و توزیع F

جی. چن ا. آداتیا

مترجم: مریم شریف‌دوست[۱]

چکیده

در این مقاله اهمیت استقلال بین متغیرهای تصادفی $N(0,1)$ و $\chi^2_{(r)}$ در ساختن توزیع $t_{(r)}$ بصورت $\frac{N(0,1)}{\sqrt{\chi^2_{(r)}}}$ بیان شده و نشان داده می‌شود در صورت عدم استقلال، تقریب توزیع کسر فوق به $t_{(r)}$ کاری نامناسب خواهد بود.

۱ مقدمه

فرض کنید $U \sim N(0,1)$ و $W \sim \chi^2_{(r)}$ و U و W از یکدیگر مستقل باشند پس خواهیم داشت

$$T = \frac{U}{(W/r)^{\frac{1}{2}}} \sim t_{(r)}$$

حال فرض کنید که $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی مستقل و هم‌توزیع با $N(\mu, \sigma^2)$ باشند. می‌دانیم که $\bar{X} = \frac{X_1+\ldots+X_n}{n}$ و $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ از یکدیگر مستقلند و

$$T^2 = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{(n-1)}$$

هنگام تدریس این مبحث سؤالات زیر برای دانشجویان پیش می‌آید:

۱– آیا X_i و $X_i - \bar{X}$ از S^2 مستقلند؟

۲– آیا $\frac{X_i - \mu}{S} \sim t_{(n-1)}$ ؟

۳– در صورت عدم استقلال U و W آیا $T = \frac{U}{(W/r)^{\frac{1}{2}}} \sim t_{(r)}$؟

در بخش دوم نشان خواهیم داد که هر سه سؤال فوق جواب منفی دارند و بطور کل حالتهایی که صورت و مخرج کسر در متغیر تصادفی T مستقل نیستند مورد بررسی قرار خواهند گرفت.

۲ چند مثال

مثال ۱:

فرض کنید $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی مستقل و هم‌توزیع با $N(\mu, \sigma^2)$، باشند پس برای هر i خواهیم داشت X_i و S^2 مستقل نیستند.

[۱] مریم شریف‌دوست، دانشگاه آزاد اسلامی واحد خمینی‌شهر

[Downloaded from andisheyeamari.irstat.ir on 2025-08-01]

برای مثال فرض کنید $n=2$، خواهیم داشت

$$X_1, X_2 \overset{i.i.d}{\sim} N(0,1) \qquad S^2 = \frac{1}{2}(X_1 - X_2)^2$$

متغیرهای تصادفی Y_1 و Y_2 را به صورت زیر تعریف می‌کنیم

$$Y_1 = X_1 \qquad Y_2 = \frac{1}{2}(X_1 - X_2)^2$$

تابع چگالی توأم Y_1 و Y_2 عبارتست از

$$\begin{aligned} g(y_1, y_2) &= \frac{1}{2\pi\sqrt{2y_2}}\{\exp[-(y_1^2 + y_2 - y_1\sqrt{2y_2})] \\ &+ \exp[-(y_1^2 + y_2 - y_1\sqrt{2y_2})] \qquad (2-1) \\ & -\infty < y_1 < \infty \quad , \quad 0 < y_2 < \infty \end{aligned}$$

چون این تابع چگالی نمی‌تواند به صورت حاصلضرب توابع چگالی Y_1 و Y_2 درآید پس Y_1 و Y_2 مستقل نیستند. به همین ترتیب $X_i - \bar{X}$ و S^2 نیز از یکدیگر مستقل نیستند.

مثال ۲:

فرض کنید $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی مستقل و هم‌توزیع با $N(\mu, \sigma^2)$ باشند، متغیر تصادفی $\frac{X-\mu}{S}$ دارای توزیع $t_{(n-1)}$ نمی‌باشد.

در حالت $n=2$ خواهیم داشت

$$X_1, X_2 \overset{i.i.d}{\sim} N(0,1)$$

Y_1 و Y_2 را به صورت زیر تعریف می‌کنیم

$$Y_1 = X_1 \qquad Y_2 = \frac{X_1}{S} = \frac{\sqrt{2}X_1}{|X_1 - X_2|}$$

تابع چگالی احتمال توأم Y_1 و Y_2 عبارتست از

$$\begin{aligned} g(y_1, y_2) &= \frac{1}{2\pi}\exp\{-\frac{1}{2}[y_1^2 + y_1^2(\frac{1-\sqrt{2}}{y_2})^2]\} \\ &\times \frac{\sqrt{2}|y_1|}{y_2^2} \\ &+ \frac{1}{2\pi}\exp\{-\frac{1}{2}[y_1^2 + y_1^2(\frac{1-\sqrt{2}}{y_2})^2]\} \\ &\times \frac{\sqrt{2}|y_1|}{y_2^2} \qquad (2-2) \end{aligned}$$

در جاییکه

$$(y_1, y_2) \in \{(y_1, y_2) : y_1 < 0, y_2 < 0 \text{ یا } y_1 > 0, y_2 > 0\}$$

تابع چگالی احتمال حاشیه‌ای Y_2 برابر است با

$$g_2(y_2) = \frac{\sqrt{2}}{4\pi}\{\frac{1}{y_2^2 - \sqrt{2}y_2 + 1} + \frac{1}{y_2^2 + \sqrt{2}y_2 + 1}\}$$
$$, -\infty < y_2 < \infty \quad (2-3)$$

واضح است که تابع چگالی احتمال فوق، چگالی $t_{(1)}$ یا حتی چگالی توزیع کوشی نیست که از عدم استقلال X_i و S^2 ناشی می‌شود.

مثال ۳:

فرض کنید $U \sim N(0,1)$ و $W \sim \chi^2_{(r)}$ باشد، پس $T = \frac{U}{(W/r)^{\frac{1}{2}}} \sim t_{(r)}$ اگر و فقط اگر U و W از یکدیگر مستقل باشند.

اثبات شرط کافی از تعریف توزیع t بدست می‌آید. برای اثبات شرط لازم حالت $r=1$ را در نظر می‌گیریم. در حالت کلی اثبات طولانی‌تر خواهدبود.

فرض کنید $(X,Y) \sim N_2(0, \Sigma)$ در صورتی که $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ و $0 \le \rho < 1$ و V, U را بصورت زیر تعریف کنیم

$$U = X \qquad V = \frac{Y}{(\frac{X^2}{1})^{\frac{1}{2}}} = \frac{Y}{|X|}$$

خواهیم داشت

$$\begin{aligned} g(u,v) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \times \\ &\exp\{\frac{-1}{2\pi(1-\rho^2)}(u^2 - 2\rho u|u|v + u^2v^2)\}|u| \ (2-4) \end{aligned}$$

و تابع چگالی احتمال حاشیه‌ای V برای $-\infty < v < \infty$ عبارتست از

$$\begin{aligned} g_2(v) &= \frac{\sqrt{(1-\rho^2)}}{2\pi} \\ &\times \{\frac{1}{v^2 - 2\rho v + 1} + \frac{1}{v^2 + 2\rho v + 1}\} \quad (2-5) \\ &= \frac{1}{\pi}\frac{1}{1+v^2} \text{ اگر و فقط اگر } \rho = 0 \quad (2-6) \end{aligned}$$

[Downloaded from andisheyeamari.irstat.ir on 2025-08-01]

چون توزیع t اغلب برای برآورد فاصله اطمینان و مسایل آزمون فرض در نمونه‌های کوچک مورد استفاده قرار می‌گیرد، دو مثال بعد را بر این اساس و با توجه به عدم استقلال بیان می‌کنیم.

مثال ۴:

برای n معلوم، ۱۰۰۰۰ نمونه تصادفی از $X_1, X_2, \ldots, X_n$ با توزیع $N(0,1)$ را شبیه‌سازی می‌کنیم. پس اگر

$$U = X_1 \quad , \quad W = (n-1)S^2$$

آنگاه

$$U \sim N(0,1) \quad , \quad W \sim \chi^2_{(n-1)}$$

ولی U و W از یکدیگر مستقل نیستند. برای هریک از ۱۰۰۰۰ نمونه تصادفی عبارت

$$T_3 = \frac{X_1 - 0}{[W/(n-1)]^{\frac{1}{2}}}$$

و چندک P درصد توزیع نمونه‌ای T_3 را برای $P = 0/8$ و $0/85$ و $0/9$ و $0/95$ و $0/975$ و $0/995$ محاسبه می‌کنیم و با چندک مربوط در $t_{(n-1)}$ برای $n = 5, 10, 15$ مقایسه می‌کنیم که نتایج در جدول ۱ نمایش داده شده است.

به خوبی دیده می‌شود که T_3 دارای توزیع $t_{(n-1)}$ نمی‌باشد، اما اثر عدم استقلال بین U و W با افزایش نمونه کاهش می‌یابد. چون توزیع t برای نمونه‌های کوچک بیشتر مورد استفاده قرار می‌گیرد، در صورت عدم استقلال بین U و W به هیچ عنوان نمی‌توان توزیع T_3 را با $t_{(n-1)}$ تقریب زد.

مثال ۵:

برای جفت‌های متعدد از θ, ϕ، ۱۰۰۰۰ نمونه از $X_1, X_2, \ldots, X_{10}$ از یک $\mathrm{ARMA}(1,1)$ با مدل زیر شبیه‌سازی نموده

$$X_t - \phi X_{t-1} = \varepsilon_t + \theta\varepsilon_{t-1}$$

که در آن $\varepsilon_t \overset{i.i.d}{\sim} N(0,1)$ می‌باشند. برای جفت ثابت θ, ϕ آماره t معمول به صورت

$$T_2 = \frac{\sqrt{n}(\bar{X} - 0)}{S}$$

برای هر ۱۰۰۰۰ نمونه محاسبه و چندک P درصد توزیع نمونه‌ای T_2 برای $P = 0/8$ و $0/85$ و $0/9$ و $0/95$ و $0/975$ و $0/995$ برآورد شده است، نتایج در جدول شماره ۲ در مقایسه با چندکهای P درصد استنباط بر اساس $t_{(10-1)}$ آورده شده است. همانطور که دیده می‌شود، توزیع نمونه‌ای T_2 با $t_{(10-1)}$ یکسان نیست و استنباط بر اساس $t_{(10-1)}$ ما را به نتیجه‌گیری غلط رهنمون می‌سازد.

۳ نتیجه

اگر $U \sim N(0,1)$ و $W \sim \chi^2_{(r)}$، شرط لازم و کافی برای آنکه $T_1 = \frac{U}{(W/r)^{\frac{1}{2}}} \sim t_{(r)}$ این است که U و W از یکدیگر مستقل باشند.

اگر $X_1, X_2, \ldots, X_{10}$ مستقل و هم‌توزیع با $N(\mu, \sigma^2)$ باشند پس:

$$T_2 = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$$

تنها آماره‌ای است که دارای توزیع $t_{(n-1)}$ است.

جدول ۱: درصد توزیع نمونه‌ای T_3:
برای هر n؛ سطر اول شامل درصدهای $t_{(n-1)}$ و سطر دوم شامل برآورد درصدی T_3 می‌باشد.

	P					
n	۰/۸۰۰	۰/۸۵	۰/۹۰۰	۰/۹۵	۰/۹۷۵	۰/۹۹۵
۵	۰/۹۴۱۰	۱/۱۸۹۶	۱/۵۳۳۲	۲/۱۳۱۸	۲/۷۷۶۴	۴/۶۰۴۱
	۰/۹۶۱۹	۱/۱۵۳۴	۱/۳۹۱۵	۱/۷۱۸۶	۱/۹۹۷۵	۲/۶۴۵۱
۱۰	۰/۸۸۳۴	۱/۰۹۹۷	۱/۳۸۳۰	۱/۸۳۳۱	۲/۲۶۲۲	۳/۲۴۹۸
	۰/۹۱۶۸	۱/۱۱۷۹	۱/۳۵۰۰	۱/۶۸۸۹	۱/۹۸۸۰	۲/۴۴۷۱
۱۵	۰/۸۶۸۱	۱/۰۷۶۳	۱/۳۴۵۰	۱/۷۶۱۳	۲/۱۴۴۸	۲/۹۷۶۸
	۰/۸۹۳۶	۱/۰۷۶۷	۱/۳۲۹۱	۱/۶۷۶۵	۱/۹۴۳۹	۲/۴۶۸۳

[Downloaded from andisheyeamari.irstat.ir on 2025-08-01]

جدول۲: درصد توزیع نمونه‌ای $T_۲$:

ردیف Exact شامل چندکهای P درصدهای توزیع $t_{(۹)}$ می‌باشد.

		P					
		۰/۸۰۰	۰/۸۵	۰/۹۰۰	۰/۹۵	۰/۹۷۵	۰/۹۹۵
Exact		۰/۸۸۳۴	۱/۰۹۹۷	۱/۳۸۳۰	۱/۸۳۳۱	۲/۲۶۲۲	۳/۲۴۹۶
ϕ	θ						
۰/۰	۰۱۰	۰/۸۸۱۱	۱/۰۰۱	۱/۳۷۴۴	۱/۸۶۷۶	۲/۲۹۱۸	۳/۲۷۰۱
۰۳	۰۱۰	۱/۱۹۹۰	۱/۵۳۱۴	۱/۹۲۷۵	۲/۵۶۳۷	۳/۱۲۷۳	۴/۶۷۶۵
۰/۶	۰۱۰	۱/۸۹۷۳	۲/۳۵۹۲	۳/۰۱۶۷	۴/۱۲۵۸	۵/۲۳۱۵	۷/۲۹۵۰
۰/۹	۰۱	۴/۴۵۵۳	۵/۶۱۰۹	۷/۳۷۴۷	۱۰/۱۴۶۳	۱۳/۰۰۸۰	۱۹/۰۵۸۳
۰۱۰	۰/۳	۱/۱۰۵۸	۱/۳۸۰۴	۱/۷۴۲۲	۲/۲۹۸۱	۲/۸۱۹۱	۴/۰۹۳۸
۰۱۰	۰/۶	۱/۲۶۶۳	۱/۵۸۹۳	۲/۰۲۸۰	۲/۷۵۷۱	۳/۴۳۷۱	۴/۷۹۵۱
۰۱۰	۰/۹	۱/۳۶۲۵	۱/۶۷۶۰	۲/۱۲۰۶	۲/۸۴۳۸	۳/۴۷۹۷	۵/۰۱۷۲
۰/۳	−۰/۹	۰/۳۳۲۵	۰/۴۱۳۵	۰/۵۱۸۶	۰/۶۷۸۰	۰/۸۲۸۵	۱/۱۶۸۷
−۰/۹	۰/۳	۰/۴۲۲۳	۰/۵۲۷۶	۰/۶۸۴۷	۰/۹۲۸۳	۱/۲۰۹۳	۱/۷۵۲۹
۰/۸	۰/۴	۳/۶۲۸۵	۴/۵۴۳۷	۵/۸۵۰۴	۷/۷۹۹۳	۹/۸۳۱۱	۱۵/۳۴۷۹
−۰/۴	−۰/۸	۰/۲۵۰۴	۰/۳۰۹۶	۰/۳۸۵۰	۰/۵۰۱۲	۰/۶۱۱۴	۰/۸۴۲۳

مراجع

[1] Browker, A. H., and Lieberman, G. J. (1972), *Engineering Statistics (2nd ed.). Englewood chiffs, NJ: prentice. Hall.*

اصل این مقاله با عنوان "Independence and t Distribution"

نوشته Gemai Chen و A. Adatia است که در

The American Statistician, May 1997, Vol. 51, N0. 2

چاپ شده است.

[Downloaded from andisheyeamari.irstat.ir on 2025-08-01]