

## استقلال و توزیع $F$

جی. چن<sup>۱</sup> . ا. آداتیا

مترجم: مریم شریف دوست<sup>۱</sup>

### 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

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

حال فرض کنید که  $X_1, X_2, \dots, X_n$  متغیرهای تصادفی مستقل و هم‌توزیع با  $N(\mu, \sigma^2)$  باشند. می‌دانیم که  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  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 سوالات زیر برای دانشجویان پیش می‌آید:

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

<sup>۱</sup> مریم شریف دوست، دانشگاه آزاد اسلامی واحد خمینی شهر

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

۳- در صورت عدم استقلال  $U$  و  $W$  آیا  $T \sim t_{(r)}$ ؟  
در بخش دوم نشان خواهیم داد که هر سه سؤال فوق جواب منفی دارند و بطور کل حالتی که صورت و مخرج کسر در متغیر تصادفی  $T$  مستقل نیستند مورد بررسی قرار خواهند گرفت.

### ۲ چند مثال

مثال ۱:

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

تابع چگالی احتمال حاشیه‌ای  $Y_2$  برابر است با

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

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

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

$$U = X \quad V = \frac{Y}{\left(\frac{X^2}{1}\right)^{1/2}} = \frac{Y}{|X|}$$

خواهیم داشت

$$g(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \times \\ \exp\left\{-\frac{1}{2\pi(1-\rho^2)}(u^2 - 2\rho uv + u^2v^2)\right\} |u| \quad (4-2)$$

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

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

برای مثال فرض کنید  $n = 2$ ، خواهیم داشت

$X_1, X_2 \stackrel{i.i.d}{\sim} N(0, 1)$   $S^2 = \frac{1}{2}(X_1 - X_2)^2$  متغیرهای تصادفی  $Y_1$  و  $Y_2$  را به صورت زیر تعریف می‌کنیم

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

تابع چگالی توأم  $Y_1$  و  $Y_2$  عبارتست از

$$g(y_1, y_2) = \frac{1}{2\pi\sqrt{2}y_2} \left\{ \exp[-(y_1^2 + y_2 - y_1\sqrt{2}y_2)] \right. \\ \left. + \exp[-(y_1^2 + y_2 - y_1\sqrt{2}y_2)] \right\} \quad (1-2) \\ -\infty < y_1 < \infty, \quad 0 < y_2 < \infty$$

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  $S^2$  و  $X_i - \bar{X}$  نیز از یکدیگر مستقل نیستند.

مثال ۲:

فرض کنید  $X_1, X_2, \dots, X_n$  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

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

$Y_1$  و  $Y_2$  را به صورت زیر تعریف می‌کنیم

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

تابع چگالی احتمال توأم  $Y_1$  و  $Y_2$  عبارتست از

$$g(y_1, y_2) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[y_1^2 + y_1^2\left(\frac{1-\sqrt{2}}{y_2}\right)^2\right]\right\} \\ \times \frac{\sqrt{2}|y_1|}{y_2^2} \\ + \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[y_1^2 + y_1^2\left(\frac{1+\sqrt{2}}{y_2}\right)^2\right]\right\} \\ \times \frac{\sqrt{2}|y_1|}{y_2^2} \quad (2-2)$$

در جاییکه

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

از  $X_1, X_2, \dots, X_{10}$  از یک  $ARMA(1, 1)$  با مدل زیر  
شبیه سازی نموده

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

که در آن  $\varepsilon_t \stackrel{i.i.d}{\sim} N(0, 1)$  می باشند. برای جفت ثابت  $\theta, \phi$   
آماره  $t$  معمول به صورت

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

برای هر ۱۰۰۰۰ نمونه محاسبه و چندک  $P$  درصد توزیع  
نمونه ای  $T_T$  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 می شود، توزیع نمونه ای  $T_T$  با  $t_{(10-1)}$   
یکسان نیست و استنباط بر اساس  $t_{(10-1)}$  ما را به نتیجه گیری  
غلط رهنمون می سازد.

### ۳ نتیجه

اگر  $U \sim N(0, 1)$  و  $W \sim \chi^2_{(r)}$ ، شرط لازم و کافی برای آنکه  
 $T_1 = \frac{U}{(W/r)^{1/2}} \sim t_{(r)}$  این است که  $U$  و  $W$  از یکدیگر مستقل  
باشند.  
اگر  $X_1, X_2, \dots, X_{10}$  مستقل و هم توزیع با  $N(\mu, \sigma^2)$  باشند  
پس:

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

برای جفت های متعدد از  $\theta, \phi$ ، ۱۰۰۰۰ نمونه تنها آماره ای است که دارای توزیع  $t_{(n-1)}$  است.

جدول ۱: درصد توزیع نمونه ای  $T_T$ :

| $n$ | $P$    |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|
|     | 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 |

چون توزیع  $t$  اغلب برای برآورد فاصله اطمینان و مسایل آزمون  
فرض در نمونه های کوچک مورد استفاده قرار می گیرد، دو مثال  
بعد را بر این اساس و با توجه به عدم استقلال بیان می کنیم.

مثال ۴:

برای  $n$  معلوم، ۱۰۰۰۰ نمونه تصادفی از  $X_1, X_2, \dots, X_n$  با  
توزیع  $N(0, 1)$  را شبیه سازی می کنیم. پس اگر

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

آنگاه

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

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

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

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

به خوبی دیده می شود که  $T_T$  دارای توزیع  $t_{(n-1)}$  نمی باشد، اما  
اثر عدم استقلال بین  $U$  و  $W$  با افزایش نمونه کاهش می یابد.  
چون توزیع  $t$  برای نمونه های کوچک بیشتر مورد استفاده قرار  
می گیرد، در صورت عدم استقلال بین  $U$  و  $W$  به هیچ عنوان  
نمی توان توزیع  $T_T$  را با  $t_{(n-1)}$  تقریب زد.

مثال ۵:

برای جفت های متعدد از  $\theta, \phi$ ، ۱۰۰۰۰ نمونه

جدول ۲: درصد توزیع نمونه‌ای  $T_2$ :  
ردیف Exact شامل چندکهای  $P$  درصدی توزیع  $t_{(1)}$  می‌باشد.

|        |          | $P$    |        |        |         |         |         |
|--------|----------|--------|--------|--------|---------|---------|---------|
| Exact  |          | ۰/۸۰۰  | ۰/۸۵   | ۰/۹۰۰  | ۰/۹۵    | ۰/۹۷۵   | ۰/۹۹۵   |
| $\phi$ | $\theta$ | ۰/۸۸۳۴ | ۱/۰۹۹۷ | ۱/۳۸۳۰ | ۱/۸۳۳۱  | ۲/۲۶۲۲  | ۳/۲۴۹۶  |
| ۰/۰    | ۰/۰      | ۰/۸۸۱۱ | ۱/۰۰۰۱ | ۱/۳۷۴۴ | ۱/۸۶۷۶  | ۲/۲۹۱۸  | ۳/۲۷۰۱  |
| ۰/۳    | ۰/۰      | ۱/۱۹۹۰ | ۱/۵۳۱۴ | ۱/۹۲۷۵ | ۲/۵۶۳۷  | ۳/۱۲۷۳  | ۴/۶۷۶۵  |
| ۰/۶    | ۰/۰      | ۱/۸۹۷۳ | ۲/۳۵۹۲ | ۳/۰۱۶۷ | ۴/۱۲۵۸  | ۵/۲۳۱۵  | ۷/۲۹۵۰  |
| ۰/۹    | ۰/۱      | ۴/۴۵۵۳ | ۵/۶۱۰۹ | ۷/۳۷۴۷ | ۱۰/۱۴۶۳ | ۱۳/۰۰۰۸ | ۱۹/۰۵۸۳ |
| ۰/۱۰   | ۰/۳      | ۱/۱۰۵۸ | ۱/۳۸۰۴ | ۱/۷۴۲۲ | ۲/۲۹۸۱  | ۲/۸۱۹۱  | ۴/۰۹۳۸  |
| ۰/۱۰   | ۰/۶      | ۱/۲۶۶۳ | ۱/۵۸۹۳ | ۲/۰۲۸۰ | ۲/۷۵۷۱  | ۳/۴۳۷۱  | ۴/۷۹۵۱  |
| ۰/۱۰   | 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 cliffs, NJ: prentice. Hall.

اصل این مقاله با عنوان "Independence and  $t$  Distribution"

نوشته A. Adatia و Gemai Chen است که در

The American Statistician, May 1997, Vol. 51, NO. 2

چاپ شده است.