فصل توزیع های نمونه ای

* تعریف: n متغیر تصادفی $X_1, \dots, X_n$ که از هم مستقل و با یکدیگر هم توزیع (دارای توزیع یکسان $f_X(x)$) باشند را یک نمونه ی تصادفی n تایی از جامعه ای با توزیع $f_X(x)$ می نامند و به صورت زیر نمایش می دهند:

$$X_1, X_2, \dots, X_n \overset{i.i.d.}{\sim} f_X(x)$$

که در آن، i.i.d. مخفف کلمات independent identically distribution به معنی مستقل و هم توزیع است.

* تعریف: هر ویژگی یک جامعه را یک پارامتر می نامند.

* تعریف: ویژگی متناظر را در نمونه، آماره می نامند. (آماره تابعی از نمونه ی تصادفی است که به پارامتر مجهول وابسته نیست. مثل $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$)

* توزیع هر آماره را توزیع نمونه ای می نامند.

در ادامه ابتدا آماره های مهم در انجام تحلیل های آماری را معرفی کرده و سپس توزیع نمونه ای هر یک را بیان می کنیم.

* جامعه ای با میانگین μ و واریانس σ^2 را در نظر بگیرید و فرض کنید $X_1, \dots, X_n$ نمونه ای تصادفی از این جامعه است. دو آماره ی مهم در این نمونه عبارتند از میانگین نمونه و واریانس نمونه که به شرح زیر تعریف می شوند:

میانگین نمونه : $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

واریانس نمونه : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$

برای یافتن توزیع نمونه ای $\bar{X}$ و S^2 دو حالت کلی وجود دارد:

(الف) توزیع جامعه ای که $X_1, \dots, X_n$ از آن آمده است نرمال باشد. ← در این صورت هر ترکیب خطی از آن نیز نرمال است.

(ب) " " " " " نرمال نباشد ← در این صورت از قضیه زیر استفاده می شود:

* قضیه حد مرکزی: اگر $X_1, X_2, \dots$ دنباله ای از متغیرهای تصادفی هم توزیع با میانگین μ و واریانس $\sigma^2 < \infty$ باشند به طوری که هر تعداد متناهی از آن ها مستقل از هم باشند. در این صورت اگر $n \to +\infty$ آنگاه $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

بنابراین در مورد توزیع نمونه‌ای میانگین نمونه داریم:

* توزیع نمونه‌ای $\bar{X}$ (وقتی واریانس جامعه معلوم است) $\downarrow \sigma^2$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + \dots + X_n)$$

الف) $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \frac{1}{n}(X_1 + \dots + X_n) \sim N(\mu, \frac{\sigma^2}{n})$

زیرا: $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$

$$Var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}$$

$$\Rightarrow \boxed{Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)}$$

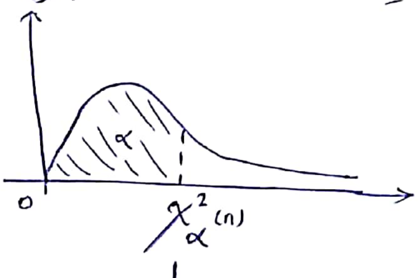
ب) $X_1, \dots, X_n \overset{iid}{\sim} f_X(x)$ غیرنرمال $\&$ $n \to +\infty \xRightarrow{\text{طبق قضیه حد مرکزی}} \boxed{Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)}$

* توزیع نمونه‌ای S^2:

$$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}}$$

☺ توزیع کای دو یک توزیع شناخته شده دیگر است که مقادیر احتمالات تجمعی آن در جداول کای دو وجود دارد و این توزیع دارای منحنی به صورت مقابل است:

(منحنی نامتقارن)



↓

عددی در توزیع کای دو با n درجه آزادی که احتمال کمتر از آن α است.

* توزیع نمونه‌ای میانگین نمونه (وقتی واریانس جامعه مجهول است):

$$x_1, \dots, x_n \overset{iid}{\sim} N(\mu, \underset{\text{مجهول}}{\sigma^2})$$

ابتدا σ^2 را به وسیله‌ی S^2 برآورد (تخمین) می‌کنیم. بنابراین داریم:

$$\boxed{\frac{\bar X-\mu}{S/\sqrt{n}} \sim t_{(n-1)}}$$

نکته: توزیع t یک توزیع شناخته شده‌ی دیگر است که مقادیر احتمالات تجمعی آن در جدول t وجود دارد

و منحنی توزیع t به صورت زیر است:

(منحنی متقارن)

$t_{\alpha}^{(n)}$

جایی در توزیع t با n درجه آزادی که احتمال قبل از آن α است.

* توزیع نمونه‌ای تفاضل میانگین‌های نمونه:

دو جامعه داریم:
- جامعه اول دارای میانگین μ_1 و واریانس σ_1^2 ← نمونه‌ای n_1 تایی از این جامعه می‌گیریم ← میانگین نمونه $\bar X_1$
- جامعه دوم " " μ_2 " σ_2^2 ← " n_2 " " ← " $\bar X_2$

هدف تعیین توزیع $\bar X_1-\bar X_2$ است. (باز دو حالت را در نظر می‌گیریم. اول اینکه دو جامعه نرمال باشند و دوم اینکه نرمال نباشند ولی حجم نمونه‌ها بزرگ‌تر از 30 باشد که از قضیه حد مرکزی تقریباً حساب می‌کنیم.)

توزیع $\bar X_1-\bar X_2$:

- σ_1^2 و σ_2^2 معلوم:
$$\bar X_1-\bar X_2 \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right) \Rightarrow \boxed{\frac{(\bar X_1-\bar X_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)}$$

- $\sigma_1^2=\sigma_2^2=\sigma^2$ مجهول:
$$\boxed{\frac{(\bar X_1-\bar X_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}} \quad ; \quad S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

توزیع نمونه‌ای نسبت واریانس های نمونه :

دو جامعه با شرایط بالا داریم.

هدف: یافتن توزیع $\frac{S_1^2}{S_2^2}$ است.

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{(n_1-1,\,n_2-1)}$$

دیگه آخرشه قول می‌دم

نکته: توزیع F (فیشر) نیز یک توزیع شناخته شده‌ی دیگر است که احتمالات تجمعی آن در جداول F وجود دارد و نمودار توزیع F به صورت زیر است:

(منحنی نامتقارن)

$F_{\alpha(m,n)}$

↓

مقداری از توزیع F با m و n درجه آزادی که احتمال کمتر از آن α است.

خلاصه‌ی این فصل :

در فصل های بعدی هر وقت سروکارمان با یک آماره بود، یاد توزیع نمونه‌ای مرتبط با آن بیفتید. یعنی:

$\bar{X}$ — واریانس جامعه معلوم $\longrightarrow Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

$\bar{X}$ — " مجهول $\longrightarrow T=\frac{\bar{x}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$

$S^2\longrightarrow \chi^2=\frac{(n-1)S^2}{\sigma^2}\sim \chi^2_{(n-1)}$

$\bar{X}_1-\bar{X}_2$ — معلوم $\sigma_1^2,\sigma_2^2\longrightarrow Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$

$\bar{X}_1-\bar{X}_2$ — مجهول $\sigma_1^2=\sigma_2^2=\sigma^2\longrightarrow T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(n_1+n_2-2)}$

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

$\frac{S_1^2}{S_2^2}\longrightarrow F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{(n_1-1,\,n_2-1)}$