X1-142 (2019/08/12)

نایقینی در سنجش ِ عملی یِ احتمال

محمد خرمی

mamwad@mailaps.org

یک راه ِ سنجش ِ احتمال ِ یک پدیده این است که آن پدیده را بارها را میئازمایند و تعداد ِ رخ-دادها یِ پدیده تقسیم بر تعداد ِ آزمایشها را برابر با احتمال ِ آن پدیده میگیرند. مقدار ی که به این طریق به دست میئاید نایقینی دارد. رابطه یِ این نایقینی با تعداد ِ بارها یِ آزمایش بررسی میشود.

0 درامد

یک پدیده با احتمال ِ p رخ میدهد. اگر این پدیده N بار آزموده شود، احتمال ِ این که $\mathfrak{n}$ (تعداد ِ بارها یی که این پدیده رخ میدهد) n باشد از تُزیع ِ دُ-جملِئی به دست میئاید:

$$\mathsf{Pro}(\mathfrak{n} = n) = \binom{N}{n}\, p^n\, (1-p)^{N-n}, \tag{1}$$

که Pro احتمال است. مقدار ِ چشمداشتی و وردایی را با، به ترتیب، E و Var نشان میدهم.

$$\mathsf{E}(\mathfrak{n}) = N\,p. \tag{2}$$

$$\mathsf{Var}(\mathfrak{n}) = N\,p\,(1-p). \tag{3}$$

پس با N بار آزمایش و n بار رخ-دادن، $\mathsf{es}(p)$ با

$$\mathsf{es}(p) = \frac{n}{N} \tag{4}$$

یک تخمین برا یِ p است. و هر چه N بزرگتر شود این تخمین دقیقتر میشود. سئال این است که این تخمین چه قدر به واقعیت (p) نزدیک است. کمی-تر، با $\mathfrak{p}$ که چنین تعریف میشود

$$\mathfrak{p} = \frac{\mathfrak{n}}{N}, \tag{5}$$

حد ِ پایین برا یِ N چنان که

$$\mathsf{Pro}(|\mathfrak{p} - p| > \alpha) < \beta \tag{6}$$

چیست؟

1 چگالی یِ احتمال برا یِ تعداد ِ زیاد یِ آزمون

بر اساس ِ قضیه یِ حد ِ مرکزی، در حد ِ N ها یِ بزرگ، تُزیع ِ احتمال برا یِ $\mathfrak{p}$ گاؤسی میشود، مثلَن [1]. چگالی-یِ-احتمال ِ متناظر با یک تُزیع ِ گاؤسی با میانگین و انحراف-ِ-معیار ِ تُزیع مشخص میشود. چگالی یِ احتمال را با pro، میانگین ِ $\mathfrak{p}$ را با μ، و انحراف-ِ-معیار ِ $\mathfrak{p}$ را با σ نشان میدهم:

$$\mu = \mathsf{E}(\mathfrak{p}). \tag{7}$$

$$\sigma = \sqrt{\mathsf{Var}(\mathfrak{p})}. \tag{8}$$

$$\mathsf{pro}(\mathfrak{p} = x) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\,\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]. \tag{9}$$

از (2) و (3) و (5) دیده میشود

$$\mu = p. \tag{10}$$

$$\sigma = \sqrt{\frac{p\,(1-p)}{N}}. \tag{11}$$

2 دقت ِ تخمین ِ احتمال

از (9) و (10) دیده میشود برا یِ N ها یِ بزرگ،

$$\begin{aligned}\mathsf{Pro}(|\mathfrak{p} - p| > \alpha) &= 1 - \int_{-\alpha}^{\alpha} \frac{\mathrm{d}\,y}{\sqrt{2\,\pi\,\sigma^2}}\,\exp\left(-\frac{y^2}{2\,\sigma^2}\right), \\ &= 1 - \mathrm{erf}\left(\frac{\alpha}{\sqrt{2\,\sigma^2}}\right),\end{aligned} \tag{12}$$

که erf تابع ِ خطا ست. با استفاده از (11)،

$$\mathsf{Pro}(|\mathfrak{p} - p| > \alpha) = 1 - \mathrm{erf}\left(\frac{\alpha\,\sqrt{N}}{\sqrt{2\,p\,(1-p)}}\right). \tag{13}$$

به این ترتیب شرط ِ (6) میشود

$$N > M, \tag{14}$$

که

$$M = \frac{2\,p\,(1-p)}{\alpha^2}\,\mathrm{erf}^{-1}(1-\beta). \tag{15}$$

M کمینه یِ تعداد ِ آزمایشها یِ لازم برا یِ این است که شرط ِ (6) براورده شود.

نایقینی با یک احتمال کمتر از α است. شرط ِ (6)، یا هم-ارز با آن (14) و (15) ، این است که این احتمال از $(1-\beta)$ بیشتر باشد. دیده میشود M با α^{-2} متناسب است. بستِگی یِ M به β ملایمتر است:

$$\mathrm{erf}^{-1}(1-\beta) = \begin{cases} 3.46, & \beta = 10^{-6} \\ 1.16, & \beta = 10^{-1} \end{cases}. \tag{16}$$

برا یِ β ها یِ کوچک،

$$\operatorname{erf}^{-1}(1-\beta) = \sqrt{-\ln(\sqrt{\pi}\,\beta) - \ln\sqrt{-\ln(\sqrt{\pi}\,\beta)} + \cdots}, \quad \beta \ll 1. \tag{17}$$

نتیجه این که برا یِ یک احتمال ِ معقول (نزدیک به یک، ولی نَ بسیار نزدیک به یک)،

$$M = \frac{\gamma}{\alpha^2}, \tag{18}$$

که γ از مرتبه یِ یک است.

3 پانوشتها

[1] William Feller; “an introduction to probability theory and its applications, volume I” 3rd edition (John Wiley & Sons, 1968) chapter X