X1-089 (2013/01/28)

سنجش ِ احتمال

محمد خرمی

mamwad@mailaps.org

تخمین ِ احتمال و عدم- ِ-قطعیت ِ آن، و نیز اثر ِ آزمایش و پیش-فرض بر این تخمینها بررسی میشود.

0 مقدمه

یک تعریف ِ غیر- ِ-رسمی یِ احتمال این است: یک رویداد را n بار می‌آزماییم. اگر m بار رخ داد، احتمال ِ رخ-دادن ِ آن رویداد (m/n) است. دقیقتر-شده یِ این تعریف آن است که این نتیجه وقت ی درست است که تعداد ِ آزمایشها زیاد باشد. این سئالها پیش می‌آیند.

- «تعداد ِآزمایشها زیاد باشد» یعنی چه؟ یک N هست که اگر تعداد ِ آزمایشها بیشتر از N باشد تعداد ِ آزمایشها زیاد (کافی) است؟

  با فرض ِ این که چنین N ی هست،

- اگر تعداد ِ آزمایشها از N بیشتر باشد احتمال ِ رویداد دقیقَن معلوم است؟

- اگر تعداد ِ آزمایشها از N کمتر باشد در باره یِ احتمال ِ رویداد هیچ نمیدانیم؟

- اگر از پیش چیز ی در باره یِ احتمال ِ آن رویداد بدانیم، آن داده چه طُر وارد ِ سنجش میشود؟

لابد جواب ِ سه سئال ِ اول منفی است. یک سکه (که از سالم-بودن َش مطمئن نیستیم) اگر ده بار شیر بیاید هم ممکن است بار ِ یازدهُم خط بیاید. آن ده بار که شیر آمده نشان نمیدهد احتمال ِ شیر-آمدن َش یک است. اگر فقط دُ بار آزمایش کنیم و هر دُ بار نتیجه شیر باشد، حدس میزنیم احتمال ِ شیر-آمدن بیشتر باشد، ولی نَ آن قدر که از ده بار آزمایش نتیجه میگیریم.

در باره یِ سئال ِ آخر، اگر بدانیم سکه سالم است (یعنی بدانیم احتمال ِ شیر-یا-خط-آمدن َش 0.5 است)، آن گاه نتیجه یِ هیچ آزمایش ی این داده (پیشفرض) را عوض نمیکند: اگر یک سکه یِ سالم صد بار پشت ِ سر ِ هم شیر بیاید هم، احتمال ِ شیر-آمدن َش 0.5 است نَ 1. اما اگر به سالم-بودن ِ سکه شَک داشته باشیم، هر چه تعداد ِ موارد ِ آزمایش بیشتر شود (به شرط ی که نتیجه یِ آزمایش با پیشفرض متفاوت باشد) شَک ِمان تقویت میشود.

دنبال ِ مدل ی برا یِ سنجش ِ احتمال یم که ویژه‌گیها یِ بالا را داشته باشد: با افزایش ِ تعداد ِ آزمایشها مقدار ی که برا یِ احتمال میدهد قطعیتر باشد، و با افزایش ِ تعداد ِ آزمایشها اثر ِ پیشفرض کمتر شود، مگر از اول احتمال را با قطعیت بدانیم.

1 قضیه یِ بیز

احتمال ِ رخ-دادن ِ $\mathbb{A}$ را با $P(\mathbb{A})$ نشان میدهیم. احتمال ِ رخ-دادن ِ $\mathbb{A}$ و $\mathbb{B}$ را با $P(\mathbb{A}\cap\mathbb{B})$، احتمال ِ رخ-دادن ِ $\mathbb{A}$ یا $\mathbb{B}$ را با $P(\mathbb{A}\cup\mathbb{B})$، و احتمال ِ رخ-دادن ِ $\mathbb{B}$ به شرط ِ $\mathbb{A}$ را با $P(\mathbb{B}|\mathbb{A})$ نشان میدهیم. داریم

$$P(\mathbb{A}\cup\mathbb{B}) = P(\mathbb{A}) + P(\mathbb{B}) - P(\mathbb{B}\cap\mathbb{A}). \tag{1}$$

$$P(\mathbb{B}|\mathbb{A}) = \frac{P(\mathbb{B}\cap\mathbb{A})}{P(\mathbb{A})}. \tag{2}$$

مجموعه‌ها یِ $\mathbb{E}_i$ با این ویژه‌گیها را در نظر بگیرید.

$$\mathbb{E}_i \cap \mathbb{E}_j = \varnothing, \qquad i \neq j. \tag{3}$$

$$\bigcup_i \mathbb{E}_i = \mathbb{M}, \tag{4}$$

که $\mathbb{M}$ مجموعه یِ مرجع است (که احتمال َش یک است). داریم

$$\begin{aligned} P(\mathbb{A}) &= P(\mathbb{A}\cap\mathbb{M}),\\ &= P[\mathbb{A}\cap(\bigcup_i \mathbb{E}_i)],\\ &= P[\bigcup_i(\mathbb{A}\cap\mathbb{E}_i)],\\ &= \sum_i P(\mathbb{A}\cap\mathbb{E}_i), \end{aligned} \tag{5}$$

که در برابری یِ آخر (1) به کار رفته، و این که به خاطر ِ (3)،

$$(\mathbb{A}\cap\mathbb{E}_i)\cap(\mathbb{A}\cap\mathbb{E}_j)=\varnothing,\qquad i\ne j. \tag{6}$$

داریم

$$\begin{aligned} P(\mathbb{E}_i|\mathbb{A}) &= \frac{P(\mathbb{E}_i\cap\mathbb{A})}{P(\mathbb{A})},\\ &= \frac{P(\mathbb{E}_i\cap\mathbb{A})}{\sum_j P(\mathbb{E}_j\cap\mathbb{A})}, \end{aligned} \tag{7}$$

که نتیجه میدهد

$$P(\mathbb{E}_i|\mathbb{A}) = \frac{P(\mathbb{E}_i)\,P(\mathbb{A}|\mathbb{E}_i)}{\sum_j P(\mathbb{E}_j)\,P(\mathbb{A}|\mathbb{E}_j)}. \tag{8}$$

به این رابطه قضیه یِ بیز [1] میگویند. اینها را میشود در مثلَن فصل ِ 4 از [2] یافت.

وقت ی مجموعه یِ مرجع پیوسته است و بر آن انتگرالگیری تعریف شده، چگالی یِ احتمال را تابع ی تعریف میکنیم که انتگرال َش احتمال است. چگالی یِ احتمال را با D نشان میدهیم. در این صورت،

$$P(\mathbb{S}) =: \int_{\mathbb{S}} (\mathrm{d}V)(\xi)\; D(\xi), \tag{9}$$

که $\mathrm{d}V$ اندازه یِ انتگرالگیری است. متغیر ِ تصادفی یِ X نگاشت ی از مجموعه یِ مرجع است. چگالی یِ احتمال برا یِ X را با D_X نشان میدهیم:

$$\begin{aligned} \int_{\mathbb{U}} (\mathrm{d}V_X)(x)\; D_X(x) &:= P[X^{-1}(\mathbb{U})],\\ &= \int_{X^{-1}(\mathbb{U})} (\mathrm{d}V)(\xi)\; D(\xi). \end{aligned} \tag{10}$$

$\mathrm{d}V_X$ اندازه یِ انتگرالگیری در مجموعه یِ مقدارها یِ X، یعنی در $X(\mathbb{M})$، است. به این ترتیب، اگر Y یک متغیر ِ تصادفی باشد چنان که (X,Y) یک-به-یک باشد و

$$(\mathrm{d}V)(\xi)=(\mathrm{d}V_X)[X(\xi)]\,(\mathrm{d}V_Y)[X(\xi),Y(\xi)], \tag{11}$$

آنگاه،

$$D_X(x)=\int_{X^{-1}(x)}\mathrm{d}V_Y[x,Y(\xi)]\;D(\xi). \tag{12}$$

به هم ین ترتیب،

$$\begin{aligned}\int_{\mathbb{U}}(\mathrm{d}V_X)(x)\;D_{X\cap\mathbb{A}}(x)&:=P\{[X^{-1}(\mathbb{U})]\cap\mathbb{A}\},\\ &=\int_{[X^{-1}(\mathbb{U})]\cap\mathbb{A}}(\mathrm{d}V)(\xi)\;D(\xi),\end{aligned} \tag{13}$$

که با تعریف ِ $\Theta_{\mathbb{A}}$ با

$$\Theta_{\mathbb{A}}(\xi):=\begin{cases}1, & \xi\in\mathbb{A}\\ 0, & \xi\notin\mathbb{A}\end{cases}, \tag{14}$$

نتیجه میدهد

$$D_{X\cap\mathbb{A}}(x)=\int_{[X^{-1}(x)]\cap\mathbb{A}}\mathrm{d}V_Y[x,Y(\xi)]\;\Theta(\xi)\,D(\xi). \tag{15}$$

سرانجام،

$$\begin{aligned}P(\mathbb{A})&=\int_{\mathbb{A}}(\mathrm{d}V)(\xi)\;D(\xi),\\ &=\int_{\mathbb{M}}(\mathrm{d}V)(\xi)\;\Theta(\xi)\,D(\xi),\\ &=\int_{\mathbb{M}}(\mathrm{d}V_X)[X(\xi)]\,(\mathrm{d}V_Y)[X(\xi),Y(\xi)]\;\Theta(\xi)\,D(\xi),\\ &=\int(\mathrm{d}V_X)(x)\;D_{X\cap\mathbb{A}}(x),\end{aligned} \tag{16}$$

$$D_{X|\mathbb{A}}=\frac{D_{X\cap\mathbb{A}}}{P(\mathbb{A})}, \tag{17}$$

$$P(\mathbb{A}|X)=\frac{D_{X\cap\mathbb{A}}}{D_X}. \tag{18}$$

به این ترتیب،

$$D_{X|\mathbb{A}}(x) = \frac{D_X(x)\,[P(\mathbb{A}|X)](x)}{\displaystyle\int (\mathrm{d}V_X)(x')\,D_X(x')\,[P(\mathbb{A}|X)](x')}. \tag{19}$$

2 احتمال ِ یک رویداد، بر اساس ِ آزمایش

رویداد ی را n بار می‌آزماییم. m بار رخ میدهد. خاسته احتمال ِ رخ-دادن ِ آن رویداد است. این (احتمال) را با متغیر ِ تصادفی یِ X نشان میدهیم. این که طی ِ n آزمایش m بار رویداد ِ مُرد- ِ-نظر رخ بدهد را با $\mathbb{A}_m$ نشان میدهیم. رابطه یِ (19) میشود

$$D_{X|\mathbb{A}_{m|n}}(x) = \frac{D_X(x)\,[P(\mathbb{A}_{m|n}|X)](x)}{\displaystyle\int (\mathrm{d}V_X)(x')\,D_X(x')\,[P(\mathbb{A}_{m|n}|X)](x')}. \tag{20}$$

$[P(\mathbb{A}_{m|n}|X)](x)$ احتمال ِ این است که با n بار آزمایش m بار رویداد ِ مُرد- ِ-نظر رخ دهد، به شرط ِ که احتمال ِ رخ-دادن رویداد ِ مُرد- ِ-نظر با یک آزمایش x باشد. اگر آزمایشها نابسته باشند، این احتمال از تُزیع ِ دُجمله‌ای به دست می‌آید:

$$[P(\mathbb{A}_{m|n}|X)](x) = \binom{n}{m}\,x^m\,(1-x)^{n-m}, \tag{21}$$

(مثلَن فصل ِ 3 از [2]). به این ترتیب،

$$(\mathrm{d}V_X)(x)\,D_{X|\mathbb{A}_{m|n}}(x) = \frac{(\mathrm{d}V_X)(x)\,D_X(x)\,x^m\,(1-x)^{n-m}}{\displaystyle\int \mathrm{d}x'\,D_X(x')\,x'^m\,(1-x')^{n-m}}, \tag{22}$$

که با تعریف ِ D_X و $\mathsf{D}_{X|\mathbb{A}_{m|n}}$ با

$$\begin{aligned}\mathrm{d}x\,\mathsf{D}_X(x) &:= (\mathrm{d}V_X)(x)\,D_X(x),\\ \mathrm{d}x\,\mathsf{D}_{X|\mathbb{A}_{m|n}}(x) &:= (\mathrm{d}V_X)(x)\,D_{X|\mathbb{A}_{m|n}}(x),\end{aligned} \tag{23}$$

میشود

$$\mathsf{D}_{X|\mathbb{A}_{m|n}}(x) = \frac{\mathsf{D}_X(x)\,x^m\,(1-x)^{n-m}}{\displaystyle\int_0^1 \mathrm{d}x'\,\mathsf{D}_X(x')\,x'^m\,(1-x')^{n-m}}, \tag{24}$$

$\mathsf{D}_X(x)$ چگالی یِ احتمال ِ این است که احتمال ِ رخ-دادن ِ رویداد ِ مُرد- ِ-نظر طی ِ یک آزمایش x باشد، به این معنی که $\mathrm{d}x\,\mathsf{D}_X(x)$ احتمال ِ این است که احتمال ِ رخ-دادن ِ رویداد ِ مُرد- ِ-نظر طی ِ یک آزمایش بین ِ x بین ِ $(x+\mathrm{d}x)$ باشد. $\mathsf{D}_{X|\mathbb{A}_{m|n}}(x)$ چگالی یِ احتمال ِ این است که

احتمالِ رخ-دادنِ رویدادِ مُرد- ِ-نظر طیِ یک آزمایش x باشد، به شرطِ این که قبلَن طیِ n بار آزمایش، m بار آن رویداد رخ داده باشد.

با چگالی-یِ-احتمالِ $\mathsf{D}_{X|\mathbb{A}_{m|n}}$ میشود میانگینِ $f(X)$ را حساب کرد:

$$\langle f(X)\rangle_{\mathbb{A}_{m|n}} = \int_0^1 \mathrm{d}x\, \mathsf{D}_{X|\mathbb{A}_{m|n}}(x)\, f(x). \tag{25}$$

از جمله،

$$\langle X\rangle_{\mathbb{A}_{m|n}} = \int_0^1 \mathrm{d}x\, \mathsf{D}_{X|\mathbb{A}_{m|n}}(x)\, x, \tag{26}$$

$$[\mathrm{Var}(X)]_{\mathbb{A}_{m|n}} = \int_0^1 \mathrm{d}x\, \mathsf{D}_{X|\mathbb{A}_{m|n}}(x)\, \left(x - \langle X\rangle_{\mathbb{A}_{m|n}}\right)^2. \tag{27}$$

اما برایِ به-دست-آوردنِ $\mathsf{D}_{X|\mathbb{A}_{m|n}}$ باید D_X را بدانیم، یعنی باید چگالی یِ احتمال پیش از انجامِ آزمایش را بدانیم.

3 سنجشِ احتمالِ یک رویداد، بدونِ پیش-فرض

اگر هیچ پیش-فرض ی نداشته باشیم (یعنی مثلَن در مُردِ پرتابِ سکه ندانیم احتمالِ شیر-آمدن 0.5 یا نزدیکِ آن است)، یک انتخاب برا یِ D_X این است که آن را در دامنه یِ تعریفَ ش ثابت بگیریم:

$$\mathsf{D}_X^{\mathrm{uni}}(x) = 1, \qquad 0 \le x \le 1. \tag{28}$$

به این ترتیب، با استفاده از

$$\int_0^1 \mathrm{d}x\, x^m\, (1-x)^{n-m} = \frac{m!\,(n-m)!}{(n+1)!}, \tag{29}$$

نتیجه میشود

$$\mathsf{D}^{\mathrm{uni}}_{X|\mathbb{A}_{m|n}}(x) = \frac{(n+1)!}{m!\,(n-m)!}\, x^m\, (1-x)^{n-m}. \tag{30}$$

بیشینه یِ این چگالی در x_0 رخ میدهد، که

$$x_0 = \frac{m}{n}. \tag{31}$$

همچنین،

$$\langle X^k\rangle^{\rm uni}_{\mathbb{A}_{m|n}} = \frac{(n+1)!}{m!\,(n-m)!}\,x^m\,(1-x)^{n-m}\,\frac{(m+k)!\,(n-m)!}{(n+k+1)!}, = \frac{(m+k)!\,(n+1)!}{m!\,(n+k+1)!}, \tag{32}$$

که از جمله نتیجه میدهد

$$\langle X\rangle^{\rm uni}_{\mathbb{A}_{m|n}} = \frac{m+1}{n+2}, \tag{33}$$

$$[{\rm Var}(X)]^{\rm uni}_{\mathbb{A}_{m|n}} = \frac{(m+1)\,(n-m+1)}{(n+2)^2\,(n+3)}. \tag{34}$$

دیده میشود مقدارِ چشمداشتی یِ احتمال، لزومَن با x_0 (جا یی که چگالی بیشینه میشود) برابر نیست. البته با افزایشِ تعدادِ آزمایشها جا یی که چگالی بیشینه میشود و مقدارِ چشمداشتی یِ احتمال به هم نزدیک میشوند:

$$\langle X\rangle^{\rm uni}_{\mathbb{A}_{m|n}} - x_0 = \frac{n-2\,m}{n\,(n+2)}. \tag{35}$$

با افزایشِ تعدادِ آزمایشها، وردایی هم کم میشود. به این ترتیب، با افزایشِ تعدادِ آزمایشها اطلاعاتِ مان در باره یِ احتمالِ رخ-دادنِ رویداد بیشتر میشود، و مقدار ی که برا یِ احتمالِ رخ-دادنِ رویداد تخمین میزنیم به هم ان مقدار ی که میگویند، یعنی به (m/n)، میگراید.

4 سنجشِ احتمالِ یک رویداد، با پیش-فرض

وجودِ پیش-فرض، یعنی پیش از آزمایش چیز ی در باره یِ X میدانیم. مثلَن بعید است احتمالِ شیر-آمدنِ یک سکه خیل ی با 0.5 فرق کند. این اطلاعات را میشود در D_X وارد کرد. یک راه این است که این اطلاعات را هم-ارز با تعداد ی آزمایش بگیریم. یعنی پیش-فرض را هم-ارز با این بگیریم که ν آزمایش انجام شده، و μ بار رویدادِ مُرد- -نظر رخ داده. این یعنی D_X را $\mathsf{D}_X^{\mathbb{A}_{\mu|\nu}}$ بگیریم، که

$$\mathsf{D}_X^{\mathbb{A}_{\mu|\nu}}(x) = \frac{(\nu+1)!}{\mu!\,(\nu-\mu)!}\,x^\mu\,(1-x)^{\nu-\mu}. \tag{36}$$

اگر بخاهیم مقدارِ چشمداشتی یِ X با این چگالی ξ باشد،

$$\xi = \frac{\mu+1}{\nu+2}. \tag{37}$$

این یک معادله برا یِ دُ متغیرِ μ و ν است. معادله یِ دیگر لابد از این جا می‌آید که چه قدر به این که احتمال x است مطمئن یم. از رابطه یِ وردایی با تعدادِ آزمایشها دیده میشود هر چه تعدادِ آزمایشها بیشتر باشد وردایی کمتر است، یعنی اطمینانِ مان بیشتر است. پس به عنوانِ یک معیار از مقدارِ اطمینان میشود خُدِ ν را داد.

میشود به صحیح-بودنِ μ و ν مقید نبود. البته μ و ν باید چنان باشند که $\mathsf{D}_X^{\mathbb{A}_{\mu|\nu}}$ انتگرال-پذبر باشد. این یعنی باید μ و $(\nu-\mu)$ بزرگتر از (-1) باشند. داریم

$$\mu=(\nu+2)\,\xi-1, \tag{38}$$

$$\nu-\mu=\nu+1-(\nu+2)\,\xi. \tag{39}$$

به این ترتیب، اگر ξ صفر یا یک نباشد، شرطِ انتگرال-پذیری یِ میشود

$$\nu>-2. \tag{40}$$

در حالتها یی که x_0 صفر یا یک است، در واقع احتمال را میدانیم: اگر یک متغیرِ تصادفی فقط بینِ صفر و یک مقدار بگیرد، و مقدارِ چشمداشتی یَش صفر یا یک باشد، آنگاه چگالی یِ احتمال برا یِ آن متغیر دلتا یِ دیرَک [3] است. البته حالتها یِ $\xi=0$ و $\xi=1$ را میشود با $\nu\to\infty$ و به ترتیب $(\mu/\nu)\to 0$ و $(\mu/\nu)\to 1$ هم به دست آورد.

سرانجام، دیده میشود

$$\mathsf{D}_X^{\mathrm{uni}}=\mathsf{D}_X^{\mathbb{A}_{0|0}}. \tag{41}$$

این یعنی حالتِ بدونِ پیش-فرض، هم ارز با آن است که تعدادِ مئثرِ آزمایشها یِ متناظر با پیش-فرض صفر است، چنان که انتظار میرفت.

با گذاشتنِ $\mathsf{D}_X^{\mathbb{A}_{\mu|\nu}}$ به جا یِ D_X در (24)، نتیجه میشود

$$\mathsf{D}_{X|\mathbb{A}_{m|n}}^{\mathbb{A}_{\mu|\nu}}(x)=\frac{(n+\nu+1)!}{(m+\mu)!\,(n+\nu-m-\mu)!}\,x^{m+\mu}\,(1-x)^{n+\nu-m-\mu}. \tag{42}$$

به این ترتیب، اثرِ پیش-فرض فقط آن است که n و m به، به ترتیب، $(n+\nu)$ و $(m+\mu)$ تبدیل

میشوند. از جمله،

$$\langle X\rangle^{\mathbb{A}_{\mu|\nu}}_{\mathbb{A}_{m|n}} = \frac{m+\mu+1}{n+\nu+2}, \tag{43}$$

$$[\mathrm{Var}(X)]^{\mathbb{A}_{\mu|\nu}}_{\mathbb{A}_{m|n}} = \frac{(m+\mu+1)\,(n+\nu-m-\mu+1)}{(n+\nu+2)^2\,(n+\nu+3)}. \tag{44}$$

رُشن است که با این مدل، هر چه ν بزرگتر باشد پیش-فرض مهمتر میشود، و هر چه n بزرگتر باشد آزمایشها مهمتر میشوند.

5 پانوشتها

[1] Bayes

[2] Morris H. deGroot & Mark J. Schervish; “probability and statistics” 4th edition (Addison-Wesley, 2002)

[3] Dirac