X1-099 (2014/04/25)

سنجش، تخمین، نایقینی

محمد خرمی

mamwad@mailaps.org

تخمین ِ مقدار ِ یک کمیت، بر اساس ِ چند بار سنجش ِ آن، بررسی میشود. بر اساس ِ هم ین سنجشها، نایقینی یِ کمیت هم تخمین زده میشود.

0 درامد

قرار است با سنجش مقدار ِ کمیت ِ X معلوم شود. n بار سنجش انجام میشود. مقدارها یِ x^1 تا x^n به دست میآید. مقدار ِ X کدام است؟ نایقینی یِ مقدار ی که بر اساس ِ این سنجشها به دست میآید کدام است؟ یک جواب ِ معمول به سئال ِ اول، میانگین ِ مقدارها یِ x^1 تا x^n است. میانگین ِ $\boldsymbol{x}$ با وزن ِ $\boldsymbol{\alpha}$ را با $\mu(\boldsymbol{\alpha};\boldsymbol{x})$ نشان میدهم:

$$\mu(\boldsymbol{\alpha};\boldsymbol{x}) = \frac{\boldsymbol{\alpha}\,\boldsymbol{x}}{\boldsymbol{\alpha}\,\boldsymbol{s}}, \tag{1}$$

که

$$s^i = 1, \tag{2}$$

$$\boldsymbol{\gamma}\,\boldsymbol{v} = \gamma_i\,v^i, \tag{3}$$

و α_i ها نامنفی یند و همه یِشان صفر نیستند. البته رُشن است که $\mu(\boldsymbol{\alpha};\boldsymbol{x})$ با مقدار ِ چشمداشتی یِ کمیت ِ X فرق دارد. مقدار ِ چشمداشتی یِ کمیت ِ X را با $\mathsf{E}(X)$ نشان میدهم. همبستگی یِ X و Y را هم با $\mathbf{C}(X,Y)$ نشان میدهم:

$$\mathbf{C}(X,Y) = \mathsf{E}\{[X-\mathsf{E}(X)]\,[Y-\mathsf{E}(Y)]\}. \tag{4}$$

برا یِ تخمین ِ Xچه وزن ی باید به کار رود؟

برا یِ سئال ِ دوم، اول باید تعریف ی از نایقینی ارائه کرد. یک کمیت برا یِ بیان ِ نایقینی انحراف-ِ-معیار است. اما از آنجا که مقدار ِ واقعی یِ کمیت معلوم نیست، انحراف ِ معیار را باید تخمین زد. مثلَن ممکن است مقدار ِ واقعی یِ کمیت را با میانگین ِ مقدارها یِ سنجیده تخمین زد. در این صورت به جا یِ انحراف از مقدار ِ واقعی انحراف از میانگین مقدارها یِ سنجیده به کار میرود. نتیجه ای که به این ترتیب به دست میآید چه ربط ی به انحراف از مقدار ِ واقعی دارد؟

1 وزن ِ بهینه برا یِ تخمین، نایقینی یِ کمینه

متناظر با سنجش ِ i، یک متغیر ِ تصادفی یِ X^i میگیرم. مقدار ِ چشمداشتی یِ هر یک از X^i ها با مقدار ِ چشمداشتی یِ (مقدار ِ واقعی یِ) X برابر است. این مقدار را با x نشان میدهم:

$$\mathsf{E}(X) = x. \tag{5}$$

$$\mathsf{E}(X^i) = x\,s^i.$$

$$\mathsf{E}(\boldsymbol{X}) = x\,\boldsymbol{s}. \tag{6}$$

اما در حالت ِ کلی نایقینی یِ X^i ها یکسان نیست. حتا لازم نیست X^i ها ناهمبسته باشند. $\mathbf{C}(X^i,X^j)$ را با $\mathbf{C}^{i\,j}$ نشان میدهم. به این ترتیب،

$$\begin{aligned}\mathbf{C}^{i\,j} &= \mathsf{E}[(X^i - x\,s^i)\,(X^j - x\,s^i)],\\ &= \mathsf{E}(X^i\,X^j) - (x)^2\,s^i\,s^j.\end{aligned} \tag{7}$$

به نظر میرسد مسئله این است. یک تابع (f) از $\boldsymbol{x}$ میخاهم، که انحراف َش از x کمینه باشد. یعنی میخاهم $|f(\boldsymbol{x})-x|$ کمینه باشد. اما این مسئله یِ خوب ی نیست. x معلوم نیست. به جا یِ آن این مسئله

را بگیرم که f چنان باشد که $\mathsf{E}[|f(\boldsymbol{X})-x|]$ کمینه باشد. این که $\mathsf{E}[|f(\boldsymbol{X})-x|]$ کمینه باشد شاید خُشتعریف باشد، ولی کارکردن با قدر- ِ-مطلق سخت است. کمینه-بودن ِ $|f(\boldsymbol{x})-x|$ با کمینه-بودن ِ $[f(\boldsymbol{x})-x]^2$ هم‌ارز است، ولی کمینه-بودن ِ $\mathsf{E}[|f(\boldsymbol{X})-x|]$ با کمینه-بودن ِ $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ هم‌ارز نیست. و البته به جا یِ قدر- ِ-مطلق یا توان ِ دوم ممکن است هر تابع ِ دیگر ی هم به کار برد که در صفر برابر ِ صفر است و در بقیه یِ مقدارها مثبت است. توان ِ دوم از همه یِ اینها ساده‌تر است.

پس مسئله را این طُر تعریف میکنم. f چه باشد تا $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ کمینه شود؟ اگر برا یِ f یک بسط ِ توانی بگیرم، مجهول ضریبها یِ این بسط خاهد بود:

$$f(\boldsymbol{\xi})=\beta+\alpha_i\,\xi^i+\gamma_{i\,j}\,\xi^i\,\xi^j+\cdots\,. \tag{8}$$

در این صورت $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ شامل ِ گشتاورها یِ $\boldsymbol{X}$ از مرتبه یِ بزرگتر از 2 خاهد بود، مگر f دست- ِ-بالا خطی باشد. این نشان میدهد اگر تنها-چیز ی که از تُزیع ِ $\boldsymbol{X}$ معلوم است داده‌ها یی در باره یِ گشتاورها یِ اول و دوم ِ آن باشد، رابطه‌ها یِ (6) و (7)، آنگاه شرط ِ کمینه-شدن ِ $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ ضریبها یِ بسط ِ f را به طُر ِ یکتا مشخص نخاهد کرد.

پس مسئله را خاصتر میکنم: f که دست- ِ-بالا خطی است چه باشد تا $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ کمینه شود؟ این یعنی

$$f(\boldsymbol{\xi})=\beta+\boldsymbol{\alpha}\,\boldsymbol{\xi}, \tag{9}$$

و هدف یافتن ِ $\boldsymbol{\alpha}$ و β ست، چنان که $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ کمینه شود. البته x معلوم نیست. پس مقدار ی که برا یِ $\boldsymbol{\alpha}$ و β به دست میآید نباید شامل ِ x باشد. از (9) و (7) دیده میشود

$$\begin{aligned}\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}&=(\beta-x)^2+2\,(\beta-x)\,\alpha_i\,s^i\,x+\alpha_i\,\alpha_j\,[\mathbf{C}^{i\,j}+(x)^2\,s^i\,s^j],\\&=(\beta-x)^2+2\,(\beta-x)\,x\,\boldsymbol{\alpha}\,\boldsymbol{s}+\boldsymbol{\alpha}\,\mathbf{C}\,\boldsymbol{\alpha}+(x)^2\,(\boldsymbol{\alpha}\,\boldsymbol{s})^2.\end{aligned} \tag{10}$$

یک شرط ِ لازم برا یِ کمینه-شدن ِ $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ این است که مشتق ِ $\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}$ نسبت به β صفر باشد. این یعنی

$$\beta-x+x\,\boldsymbol{\alpha}\,\boldsymbol{s}=0, \tag{11}$$

یا

$$\beta=(1-\boldsymbol{\alpha}\,\boldsymbol{s})\,x. \tag{12}$$

قرار است $\boldsymbol{\alpha}$ تابع ِ x نباشد. تنها-حالت ی که چنین است و β هم تابع ِ x نیست این است که پرانتز در طرف ِ راست ِ صفر شود. در این صورت،

$$\boldsymbol{\alpha}\, \boldsymbol{s} = 1. \tag{13}$$

$$\beta = 0. \tag{14}$$

با گذاشتن ِ اینها در (10) نتیجه میشود

$$\mathsf{E}\{[f(\boldsymbol{X}) - x]^2\} = \boldsymbol{\alpha}\, \mathbf{C}\, \boldsymbol{\alpha}. \tag{15}$$

این را باید با قید ِ (13) کمینه کرد. روش ِ ضریبها یِ لَگرانژ [1] را به کار میبرم:

$$g(\boldsymbol{\alpha}, \lambda) := \boldsymbol{\alpha}\, \mathbf{C}\, \boldsymbol{\alpha} + 2\, \lambda\, (1 - \boldsymbol{\alpha}\, \boldsymbol{s}), \tag{16}$$

و باید g را نسبت به $\boldsymbol{\alpha}$ و λ فرینه کرد. (ضریب ِ 2 کنار ِ λ اهمیت ی ندارد و فقط رابطه‌ها یِ بعدی یِ شامل ِ λ را مختصرتر میکند.) مشتق ِ g نسبت به $\boldsymbol{\alpha}$ را صفر میگذارم:

$$\mathbf{C}\, \boldsymbol{\alpha} - \lambda\, \boldsymbol{s} = 0, \tag{17}$$

که تقارن ِ $\mathbf{C}$ به کار رفته. نتیجه میشود

$$\boldsymbol{\alpha} = \lambda\, \mathbf{C}^{-1}\, \boldsymbol{s}. \tag{18}$$

ثابت ِ λ هم از قید ِ (13)، که هم ان شرط ِ فرینه-شدن ِ g نسبت به λ ست، به دست میآید:

$$\lambda = \frac{1}{\boldsymbol{s}\, \mathbf{C}^{-1}\, \boldsymbol{s}}. \tag{19}$$

به این ترتیب،

$$\boldsymbol{\alpha} = \frac{\mathbf{C}^{-1}\, \boldsymbol{s}}{\boldsymbol{s}\, \mathbf{C}^{-1}\, \boldsymbol{s}}. \tag{20}$$

$\mathbf{C}$ مثبت ِ شبهِ‌معین است. مثبت ِ معین است، اگر هیچ یک از $\boldsymbol{X}^i$ ها یک تابع ِ دست- ِ-بالا-خطی از بقیه نباشد. اگر $\mathbf{C}$ مثبت ِ معین باشد $\mathbf{C}^{-1}$ وجود دارد. اگر $\mathbf{C}$ مثبت ِ معین نباشد، یک بردار ِ (حقیقی یِ ناصفر ِ) $\boldsymbol{u}$ هست که

$$\mathbf{C}\, \boldsymbol{u} = 0. \tag{21}$$

از (7) دیده میشود به ازا یِ هر $\boldsymbol{u}$ که (21) را براوَرَد،

$$\mathsf{E}[(\boldsymbol{u}\, \boldsymbol{X} - x\, \boldsymbol{u}\, \boldsymbol{s})^2] = 0, \tag{22}$$

که نتیجه میدهد

$$\boldsymbol{u}\,\boldsymbol{X} - x\,\boldsymbol{u}\,\boldsymbol{s} = 0. \tag{23}$$

فرض میکنم این نتیجه مستقل از مقدار ِ (نامعلوم ِ) x برقرار باشد. در این صورت،

$$\boldsymbol{u}\,\boldsymbol{X} = 0. \tag{24}$$

$$\boldsymbol{u}\,\boldsymbol{s} = 0. \tag{25}$$

بر عکس، اگر $\boldsymbol{u}$ رابطه‌ها یِ (24) و (25) را براوردَ، از (7) نتیجه میشود (22) برقرار است، که این هم (23) و در نتیجه (21) را میدهد. پس (21) برقرار است، اگر و تنها اگر (24) و (25) برقرار باشند.

(25) تضمین میکند معادله یِ (17) برا یِ $\boldsymbol{\alpha}$ جواب دارد، هر چند این جواب یکتا نیست، و به شکل ِ (20) هم نیست. در واقع متناظر با هر $\boldsymbol{u}$ که (21) را براوردَ،

$$\boldsymbol{u}\,(\mathbf{C}\,\boldsymbol{\alpha} - \lambda\,\boldsymbol{s}) \equiv 0, \tag{26}$$

که نشان میدهد تعداد ِ معادله‌ها یِ مستقل برا یِ α کمتر از n است (به اندازه یِ بُعد ِ هسته یِ $\mathbf{C}$).
ضمنَن دیده میشود اگر $\boldsymbol{\alpha}$ یک جواب ِ (17) باشد، $\boldsymbol{\alpha}'$ هم یک جواب ِ (17) است اگر و تنها اگر

$$\mathbf{C}\,(\boldsymbol{\alpha}' - \boldsymbol{\alpha}) = 0, \tag{27}$$

که نتیجه میدهد

$$(\boldsymbol{\alpha}' - \boldsymbol{\alpha})\,\boldsymbol{X} = 0. \tag{28}$$

این نشان میدهد هر چند $\boldsymbol{\alpha}$ یکتا به دست نمیآید، $f(\boldsymbol{X})$ یکتا به دست میآید. در واقع حالت ِ کلی یِ (20) میشود

$$\boldsymbol{\alpha} = \frac{\mathbf{B}\,\boldsymbol{s}}{\boldsymbol{s}\,\mathbf{B}\,\boldsymbol{s}}, \tag{29}$$

که $\mathbf{B}$ یک وارون ِ راست ِ $\mathbf{C}$ است:

$$\mathrm{dom}(\mathbf{B}) = \mathrm{img}(\mathbf{C}). \tag{30}$$

$$\mathbf{C}\,\mathbf{B} = \mathbf{1}_{\mathrm{img}(\mathbf{C})}. \tag{31}$$

اگر $\mathbf{C}$ وارونپذیر نباشد، $\mathbf{B}$ یکتا نیست.

حالت ی را در نظر میگیرم که هسته یِ $\mathbf{C}$ یک فضا یِ $(n-\tilde{n})$ بُعدی ست، یعنی فضا یِ $\boldsymbol{u}$ ها یی که (21)، یا همارز با آن (24) و (25)، را برمیآورَند $(n-\tilde{n})$ بُعدی ست. پس $(n-\tilde{n})$ تا از X^i ها ترکیبها یی خطی از $\tilde{n}$ تا X^i یِ دیگر ند. بدون ِ کاستن از کلیت، دسته یِ اخیر را X^1 تا $X^{\tilde{n}}$ میگیرم. به این ترتیب برا یِ هسته یِ $\mathbf{C}$ یک پایه یِ $\{(i,\boldsymbol{e}^i)\mid 1\le i\le (n-\tilde{n})\}$ هست که

$$(\boldsymbol{e}^i)_j=\delta^i_{j-\tilde{n}}, \qquad j>\tilde{n}. \tag{32}$$

e^i را به جا یِ u در (26) میگذارم. نتیجه میشود معادله یِ شماره یِ $(\tilde{n}+i)$ یک ترکیب ِ خطی از $\tilde{n}$ معادله یِ اول است. پس میشود معادلهها یِ شماره یِ $(\tilde{n}+1)$ تا n را حذف کرد. همچنین، از این که با افزودن ِ یک عضو ِ هسته یِ $\mathbf{C}$ به یک جواب برا یِ $\boldsymbol{\alpha}$ یک جواب ِ دیگر برا یِ $\boldsymbol{\alpha}$ به دست میآید استفاده میکنم و جواب ی برا یِ $\boldsymbol{\alpha}$ میگیرم که

$$\alpha_i=0, \qquad i>\tilde{n}. \tag{33}$$

به این ترتیب (17) میشود

$$\tilde{\mathbf{C}}\,\tilde{\boldsymbol{\alpha}}-\lambda\,\tilde{\boldsymbol{s}}=0, \tag{34}$$

که شاخصها یِ $\tilde{Y}$ از 1 تا $\tilde{n}$ مقدار میگیرند، چنان که $\tilde{Y}$ به ازا یِ شاخصها یِ نابزرگتر از $\tilde{n}$ هم ان Y با شاخصها یِ متناظر است. رُشن است که $\tilde{\mathbf{C}}$ متقارن و مثبت ِ معین (و در نتیجه وارونپذیر) است. به این ترتیب،

$$\tilde{\boldsymbol{\alpha}}=\frac{\tilde{\mathbf{C}}^{-1}\,\tilde{\boldsymbol{s}}}{\tilde{\boldsymbol{s}}\,\tilde{\mathbf{C}}^{-1}\,\tilde{\boldsymbol{s}}}, \tag{35}$$

پس در حالت ِ کلی، دستورالعمل برا یِ تخمین ِ x چنین است. سنجشها یِ کاملَن-وابسته ($x^{\tilde{n}+1}$ تا x^n) را حذف میکنم. اینها سنجشها یی اند که نتیجه یِ هر کدام ِشان یک ترکیب ِ خطی یِ ثابت از سنجشها یِ دیگر (x^1 تا $x^{\tilde{n}}$) است. همبستهگی یِ سنجشها یِ باقیمانده $\tilde{\mathbf{C}}$ ست. از اینجا به بعد، برا یِ ساده-کردن ِ نمادگذاری $\tilde{Y}$ را به Y تبدیل میکنم، البته منظور این است که سنجشها یِ کاملَن-وابسته حذف شده اند. به این ترتیب،

$$f(\boldsymbol{\xi})=\mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{\xi}), \tag{36}$$

و $\mathsf{es}(x)$ (تخمین برا یِ x) میشود

$$\begin{aligned}\mathsf{es}(x) &= \mathsf{E}[f(\boldsymbol{X})],\\ &= \mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{x}).\end{aligned}\tag{37}$$

$\mathsf{un}(x)$ (نایقینی یِ x) هم چنین میشود.

$$\mathsf{un}(x) = \sqrt{\mathsf{E}\{[f(\boldsymbol{X})-x]^2\}},\tag{38}$$

که به این مینجامد.

$$\begin{aligned}[\mathsf{un}(x)]^2 &= \mathsf{E}\{[\mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{X})-x]^2\},\\ &= \mathsf{E}\{[\mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{X})]^2\} - 2\,x\,\mathsf{E}[\mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{X})] + x^2.\end{aligned}\tag{39}$$

از (6) و (7) نتیجه میشود

$$\mathsf{E}[\mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{X})] = x.\tag{40}$$

$$\mathsf{E}\{[\mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{X})]^2\} = \frac{1}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}} + (x)^2.\tag{41}$$

به این ترتیب،

$$[\mathsf{un}(x)]^2 = \frac{1}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}.\tag{42}$$

یک حالت ِ ساده این است که سنجشها مستقل از هم باشند. این یعنی همبسته‌گی قطری است:

$$\mathbf{C}^{i\,j} = (\sigma^i)^2\,\delta^{i\,j},\tag{43}$$

که σ^i نایقینی یِ سنجش ِ i (و البته مثبت) است. در این صورت $\mathbf{C}^{-1}$ هم قطری است، و،

$$\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s} = \sum_i (\sigma^i)^{-2}.\tag{44}$$

$$\alpha_i = \Big[\sum_j (\sigma^j)^{-2}\Big]^{-1}\,(\sigma_i)^{-2}.\tag{45}$$

$$\mathsf{es}(x) = \Big[\sum_j (\sigma^j)^{-2}\Big]^{-1}\,\Big[\sum_i (\sigma^i)^{-2}\,x^i\Big].\tag{46}$$

$$\mathsf{un}(x) = \Big[\sum_j (\sigma^j)^{-2}\Big]^{-1/2}.\tag{47}$$

یعنی تخمین ِ x یک میانگین ِ وزندار از x^i ها ست، چنان که وزن ِ x^i متناسب با عکس ِ مجذور ِ نایقینی یِ سنجش ِ i است. عکس ِ مجذور ِ نایقینی هم مجموع ِ عکس ِ مجذور ِ نایقینی یِ سنجشها ست.

از این ساده‌تر وقت ی ست که همبسته‌گی نَ تنها قطری ست، بل که عنصرها یِ قطری یِ آن هم یکسان َند:

$$\mathbf{C}^{i\,j} = (\sigma)^2\,\delta^{i\,j}, \tag{48}$$

که σ مثبت است. در این صورت،

$$\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s} = n\,(\sigma)^{-2}. \tag{49}$$

$$\alpha_i = \frac{1}{n}. \tag{50}$$

$$\mathsf{es}(x) = \frac{1}{n}\,\sum_i x^i. \tag{51}$$

$$\mathsf{un}(x) = \frac{\sigma}{\sqrt{n}}. \tag{52}$$

یعنی تخمین ِ x یک میانگین ِ بدون ِ وزن (با وزن ِ یکسان) از x^i ها ست. نایقینی هم برابر ِ نایقینی یِ هر یک از سنجشها تقسیم بر جذر ِ تعداد ِ سنجشها ست.

2 تخمین ِ نایقینی با سنجشها

نایقینی یِ تخمین از (42) به دست می‌آید. اما برا یِ محاسبه یِ طرف ِ راست ِ (42) باید $\mathbf{C}$ معلوم باشد. میشود تخمین ی برا یِ نایقینی داد که اگر $\mathbf{C}$ یک ثابت ِ ضربی یِ نامعین داشته باشد هم قابل ِ محاسبه است. از (7) نتیجه میشود

$$\begin{aligned}\mathsf{E}(\boldsymbol{X}\,\mathbf{C}^{-1}\,\boldsymbol{X}) &= (\mathbf{C}^{-1})_{i\,j}\,\mathbf{C}^{j\,i} + (x)^2\,(\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}),\\ &= n + (x)^2\,(\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}).\end{aligned} \tag{53}$$

از این و (41) نتیجه میشود

$$\mathsf{E}\left[\frac{\boldsymbol{X}\,\mathbf{C}^{-1}\,\boldsymbol{X}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}} - \left(\frac{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{X}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}\right)^2\right] = \frac{n-1}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}. \tag{54}$$

به این ترتیب،

$$[\mathsf{un}(x)]^2 = \frac{1}{n-1}\,\mathsf{E}\bigg[\frac{\boldsymbol{X}\,\mathbf{C}^{-1}\,\boldsymbol{X}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}} - \bigg(\frac{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{X}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}\bigg)^2\bigg]. \tag{55}$$

از اینجا یک تخمین برا یِ $\mathsf{un}(x)$ به دست می‌آید:

$$\mathsf{es}\{[\mathsf{un}(x)]^2\} = \frac{1}{n-1}\,\bigg[\frac{\boldsymbol{x}\,\mathbf{C}^{-1}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}} - \bigg(\frac{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}\bigg)^2\bigg]. \tag{56}$$

وردایی یِ $\boldsymbol{x}$ با وزنِ $\mathbf{M}$ را با $\mathsf{var}(\mathbf{M};\boldsymbol{x})$ نشان می‌دهم:

$$\begin{aligned}\mathsf{var}(\mathbf{M};\boldsymbol{x}) &= \frac{\{\boldsymbol{x}-[\mu(\boldsymbol{s}\,\mathbf{M};\boldsymbol{x})]\,\boldsymbol{s}\}\,\mathbf{M}\,\{\boldsymbol{x}-[\mu(\boldsymbol{s}\,\mathbf{M};\boldsymbol{x})]\,\boldsymbol{s}\}}{\boldsymbol{s}\,\mathbf{M}\,\boldsymbol{s}},\\ &= \frac{\boldsymbol{x}\,\mathbf{M}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{M}\,\boldsymbol{s}} - [\mu(\boldsymbol{s}\,\mathbf{M};\boldsymbol{x})]^2,\\ &= \frac{\boldsymbol{x}\,\mathbf{M}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{M}\,\boldsymbol{s}} - \bigg(\frac{\boldsymbol{s}\,\mathbf{M}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{M}\,\boldsymbol{s}}\bigg)^2,\end{aligned} \tag{57}$$

که $\mathbf{M}$ یک ماتریسِ متقارنِ مثبتِ معین است. به این ترتیب،

$$\mathsf{es}\{[\mathsf{un}(x)]^2\} = \frac{1}{n-1}\,\mathsf{var}(\mathbf{C}^{-1};\boldsymbol{x}). \tag{58}$$

این یک تخمین برا یِ $[\mathsf{un}(x)]^2$، و در نتیجه برا یِ $[\mathsf{un}(x)]$، است، خُدِ $[\mathsf{un}(x)]^2$ نیست. اما برا یِ محاسبه اَش $\mathbf{C}$ تا حدِ یک ثابتِ ضربی یِ نامعین، و داده‌ها یِ حاصل از سنجش (یعنی x^i ها) کافی یَند. چون $[\mathsf{un}(x)]^2$ به خُدِ $\mathbf{C}$ مربوط است، رابطه یِ (42)، تخمینِ $[\mathsf{un}(x)]^2$ را می‌شود برا یِ محاسبه یِ تخمین ی برا یِ $\mathbf{C}$ به کار بُرد:

$$\mathsf{es}\bigg(\frac{1}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}\bigg) = \frac{1}{n-1}\,\bigg[\frac{\boldsymbol{x}\,\mathbf{C}^{-1}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}} - \bigg(\frac{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}\bigg)^2\bigg]. \tag{59}$$

پس اگر

$$\mathbf{C} = (\sigma)^2\,\underset{\circ}{\mathbf{C}}, \tag{60}$$

که σ یک عددِ مثبتِ نامعلوم است و $\underset{\circ}{\mathbf{C}}$ معلوم است،

$$\mathsf{es}[(\sigma)^2] = \frac{1}{n-1}\,\bigg[\boldsymbol{x}\,\underset{\circ}{\mathbf{C}}^{-1}\,\boldsymbol{x} - \frac{(\boldsymbol{s}\,\underset{\circ}{\mathbf{C}}^{-1}\,\boldsymbol{x})^2}{\boldsymbol{s}\,\underset{\circ}{\mathbf{C}}^{-1}\,\boldsymbol{s}}\bigg]. \tag{61}$$

به این ترتیب،

$$\begin{aligned}\mathsf{es}\{[\mathsf{un}(x)]^2\} &= \frac{1}{n-1}\,\bigg[\frac{\boldsymbol{x}\,\underset{\circ}{\mathbf{C}}^{-1}\,\boldsymbol{x}}{\boldsymbol{s}\,\underset{\circ}{\mathbf{C}}^{-1}\,\boldsymbol{s}} - \bigg(\frac{\boldsymbol{s}\,\underset{\circ}{\mathbf{C}}^{-1}\,\boldsymbol{x}}{\boldsymbol{s}\,\underset{\circ}{\mathbf{C}}^{-1}\,\boldsymbol{s}}\bigg)^2\bigg],\\ &= \frac{1}{n-1}\,\mathsf{var}(\underset{\circ}{\mathbf{C}}^{-1};\boldsymbol{x}).\end{aligned} \tag{62}$$

این رابطه خیل ی شبیه ِ (56) یا (58) است، در واقع هم ان است با این تفاوت که $\mathop{\mathbf{C}}\limits_{\circ}$ جا یِ $\mathbf{C}$ را گرفته. این را مستقیمَن هم میشد از (56) یا (58) نتیجه گرفت. نتیجه این است که اگر همبسته‌گی یِ سنجشها تا حد ِ یک ضریب ِ نامعلوم معین باشد، نایقینی یِ سنجش را میشود تخمین زد.

برا یِ حالت ِ ساده یِ همبستگی یِ قطری، (43)، رابطه یِ (60) میشود

$$\sigma^i = \sigma\, \mathop{\sigma}\limits_{\circ}{}^i, \qquad (63)$$

که $\mathop{\sigma}\limits_{\circ}{}^i$ ها (مثبت و) معلوم َند، و σ مثبت ولی نامعلوم است. (61) میشود

$$\begin{aligned}\mathsf{es}[(\sigma)^2] &= \frac{1}{n-1}\Big\{\Big[\sum_i (\mathop{\sigma}\limits_{\circ}{}^i)^{-2}\,(x^i)^2\Big] - \Big[\sum_j (\mathop{\sigma}\limits_{\circ}{}^j)^{-2}\Big]^{-1}\Big[\sum_i (\mathop{\sigma}\limits_{\circ}{}^i)^{-2}\,x^i\Big]^2\Big\},\\ &= \frac{1}{n-1}\sum_i (\mathop{\sigma}\limits_{\circ}{}^i)^{-2}\Big\{x^i - \Big[\sum_j (\mathop{\sigma}\limits_{\circ}{}^j)^{-2}\Big]^{-1}\Big[\sum_k (\mathop{\sigma}\limits_{\circ}{}^k)^{-2}\,x^k\Big]\Big\}^2, \qquad (64)\end{aligned}$$

که با تعریف ِ

$$\mu_0(\boldsymbol{x}) := \Big[\sum_j (\mathop{\sigma}\limits_{\circ}{}^j)^{-2}\Big]^{-1}\Big[\sum_k (\mathop{\sigma}\limits_{\circ}{}^k)^{-2}\,x^k\Big], \qquad (65)$$

میشود

$$\mathsf{es}[(\sigma)^2] = \frac{1}{n-1}\sum_i (\mathop{\sigma}\limits_{\circ}{}^i)^{-2}\,[x^i - \mu_0(\boldsymbol{x})]^2. \qquad (66)$$

در نتیجه،

$$\mathsf{es}\{[\mathsf{un}(x)]^2\} = \frac{1}{n-1}\Big[\sum_j (\mathop{\sigma}\limits_{\circ}{}^j)^{-2}\Big]^{-1}\sum_i (\mathop{\sigma}\limits_{\circ}{}^i)^{-2}\,[x^i - \mu_0(\boldsymbol{x})]^2. \qquad (67)$$

این هم با تعریف ِ

$$\mathsf{var}_0(\boldsymbol{x}) = \Big[\sum_j (\mathop{\sigma}\limits_{\circ}{}^j)^{-2}\Big]^{-1}\sum_i (\mathop{\sigma}\limits_{\circ}{}^i)^{-2}\,[x^i - \mu_0(\boldsymbol{x})]^2, \qquad (68)$$

میشود

$$\mathsf{es}\{[\mathsf{un}(x)]^2\} = \frac{1}{n-1}\,\mathsf{var}_0(\boldsymbol{x}). \qquad (69)$$

نتیجه این است که برا یِ محاسبه یِ چه میانگین و چه وردایی، وزن ی متناسب با عکس ِ مجذور ِ نایقینی به کار برود.

در حالت ِ ساده‌تر ی که همبسته‌گی نَ تنها قطری ست، بل که عنصرها یِ قطری یِ آن هم یکسان َند،

(48)، رابطه یِ (60) با σ یِ مثبت ولی نامعلوم، (61) میشود

$$\mathsf{es}[(\sigma)^2] = \frac{1}{n-1}\left\{\Big[\sum_i (x^i)^2\Big] - \frac{1}{n}\Big(\sum_i x^i\Big)^2\right\},$$
$$= \frac{1}{n-1}\sum_i\left(x^i - \frac{1}{n}\sum_k x^k\right)^2, \tag{70}$$

که با تعریفِ

$$\mu(\boldsymbol{x}) := \frac{1}{n}\sum_k x^k, \tag{71}$$

میشود

$$\mathsf{es}[(\sigma)^2] = \frac{1}{n-1}\sum_i [x^i - \mu(\boldsymbol{x})]^2. \tag{72}$$

در نتیجه،

$$\mathsf{es}\{[\mathsf{un}(x)]^2\} = \frac{1}{n-1}\,\frac{1}{n}\sum_i [x^i - \mu(\boldsymbol{x})]^2. \tag{73}$$

این هم با تعریفِ

$$\mathsf{var}(\boldsymbol{x}) = \frac{1}{n}\sum_i [x^i - \mu(\boldsymbol{x})]^2, \tag{74}$$

میشود

$$\mathsf{es}\{[\mathsf{un}(x)]^2\} = \frac{1}{n-1}\,\mathsf{var}(\boldsymbol{x}). \tag{75}$$

3 تخمین با استفاده از قضیه یِ قضیه یِ بِیز

چنان که در [2] آمده، قضیه یِ بِیز [3] میگوید تُزیع ی احتمال برا یِ متغیرِ تصادفی یِ X به شرطِ رخ-دادنِ $\mathbb{A}$ میشود

$$D_{X|\mathbb{A}}(x) = \frac{D_X(x)\,[\mathsf{Pro}(\mathbb{A}|X)](x)}{\displaystyle\int (\mathrm{d}V_X)(x')\,D_X(x')\,[\mathsf{Pro}(\mathbb{A}|X)](x')}. \tag{76}$$

$\mathrm{d}V_X$ اندازه یِ انتگرالگیری، D_X چگالی یِ احتمال برا یِ X (بدونِ شرط)، $D_{X|\mathbb{A}}$ چگالی یِ احتمال برا یِ X به شرطِ این که $\mathbb{A}$ رخ داده باشد، و $[\mathsf{Pro}(\mathbb{A}|X)](x)$ احتمالِ رخ-دادنِ $\mathbb{A}$ است به شرطِ این که X برابرِ x باشد. به جا یِ رویدادِ $\mathbb{A}$ متغیرِ تصادفی یِ Y را میگیرم. (76) میشود

$$D_{X|Y}(x,y) = \frac{D_X(x)\,D_{Y|X}(y,x)}{\displaystyle\int (\mathrm{d}V_X)(x')\,D_X(x')\,D_{Y|X}(y,x')}. \tag{77}$$

$D_{Y|X}(y,x)$ چگالی ی احتمال ِ این است که Y برابر ِ y باشد، به شرط ِ این که X برابر ِ x باشد.

مقدار ِ چشمداشتی ی تابع ِ F از متغیر ِ تصادفی ی X، به شرط ِ این که Y برابر ِ y باشد را با $\mathsf{E_c}[Y=y;F(X)]$ نشان میدهم. دیده میشود

$$\mathsf{E_c}[Y=y;F(X)]=\frac{\displaystyle\int(\mathrm{d}V_X)(x)\,D_X(x)\,F(x)\,D_{Y|X}(y,x)}{\displaystyle\int(\mathrm{d}V_X)(x')\,D_X(x')\,D_{Y|X}(y,x')}. \tag{78}$$

برا ی محاسبه ی این مقدار- ِ-چشمداشتی ی شرطی، هم $(\mathrm{d}V_X\,D_X)$ (اندازه ی انتگرالگیری ضرب در چگالی- ِ-احتمال ِ غیر- ِ-شرطی)، و هم $D_{Y|X}$ لازم است. اگر پیشفرض-نداشتن را چنین معنی کنم که $(\mathrm{d}V_X\,D_X)$ مستقل از متغیر َش باشد، همچنان دانستن ِ $D_{Y|X}$ لازم است، هر چند خُد ِ این که $(\mathrm{d}V_X\,D_X)$ مستقل از متغیر است خُشتعریف نیست، مگر رُشن باشد منظور ناوردادماندن تحت ِ کدام انتقال است. مسئله سادهتر میشود اگر $D_{Y|X}(y,x)$ تیز باشد. در این حالت میشود انتگرال ی که انتگرالده اَش $D_{Y|X}(y,x)$ ضرب در یک تابع ِ نرم ِ G از x است را چنین حساب کرد که به جا ی $G(x)$ در انتگرالده اولین جمله ی ناصفر ِ بسط ِ تیلِر [4] ِ G حُل ِ $\xi(y)$ را گذاشت، که $\xi(y)$ مقدار ی است که با نشاندن ِ آن به جا ی x در $D_{Y|X}(y,x)$، مقدار ِ $D_{Y|X}(y,x)$ بیشینه میشود. این در واقع یک روش ِ تندترین شیب است. پس $D_{Y|X}(y,x)$ را با یک تابع ِ گاؤسی نسبت به y، و ضریبها ی آن در انتگرالده را با اولین جمله ی ناصفر ِ بسط ِ تیلِر [4] شان جایگزین میکنم:

$$\mathsf{E_c}[Y=y;F(X)]=\frac{\displaystyle\int\mathrm{d}x\,(\mathrm{T}F)[\xi(y);x]\,(\mathrm{G}D)_{Y|X}(y,x)}{\displaystyle\int\mathrm{d}x'\,(\mathrm{G}D)_{Y|X}(y,x')}, \tag{79}$$

که $(\mathrm{T}F)(a,\centerdot)$ اولین جمله ی ناصفر ِ بسط ِ تیلِر [4] ِ F حُل ِ a، و $(\mathrm{G}D)$ تقریب ِ گاؤسی ی D است.

حالا مسئله را خاص میکنم. Y هم ان $\boldsymbol{X}$ در بخشها ی پیش است: X^i نتیجه ی سنجش ِ i برا ی متغیر ِ X است. گشتاورها ی اول و دوم ِ $\boldsymbol{X}$ هم از (6) و (7) به دست میآیند. به این ترتیب،

$$(\mathrm{G}D)_{\boldsymbol{X}|X}(\boldsymbol{x},x)=N\,\exp\left[-\frac{1}{2}\,(\boldsymbol{x}-x\,\boldsymbol{s})\,\mathbf{C}^{-1}\,(\boldsymbol{x}-x\,\boldsymbol{s})\right], \tag{80}$$

که N یک ضریب ِ بهنجارش است (که در محاسبه ی مقدارها ی چشمداشتی ی مشروط حذف میشود). با استفاده از

$$(\boldsymbol{x}-x\,\boldsymbol{s})\,\mathbf{C}^{-1}\,(\boldsymbol{x}-x\,\boldsymbol{s})=(\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s})\,x^2-2\,(\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{x})\,x+(\boldsymbol{x}\,\mathbf{C}^{-1}\,\boldsymbol{x}), \tag{81}$$

از (80) دیده میشود

$$(\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s})\,[\xi(\boldsymbol{x})] = \boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{x}. \tag{82}$$

به این ترتیب،

$$\begin{aligned}\mathsf{E}_{\mathsf{c}}(\boldsymbol{X}=\boldsymbol{x};X) &= \frac{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{x}}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}},\\ &= \mu(\boldsymbol{s}\,\mathbf{C}^{-1};\boldsymbol{x}).\end{aligned} \tag{83}$$

از مقایسه یِ این با (37) دیده میشود

$$\mathsf{E}_{\mathsf{c}}(\boldsymbol{X}=\boldsymbol{x};X) = \mathsf{es}(x). \tag{84}$$

مشابه با مقدار ِ چشمداشتی یِ مشروط، خُدبسته‌گی یِ مشروط ِ Z به شرط ِ این که Y برابر ِ y باشد را با $[\Delta_{\mathsf{c}}(Y=y;Z)]^2$ نشان میدهم و آن را چنین تعریف میکنم.

$$\Delta_{\mathsf{c}}(Y=y;Z) = \sqrt{\mathsf{E}_{\mathsf{c}}\{Y=y;[Z-\mathsf{E}_{\mathsf{c}}(Y=y;Z)]^2\}}. \tag{85}$$

$(\mathrm{G}D)_{\boldsymbol{X}|X}(\boldsymbol{x},x)$ نسبت به x گاؤسی ست.. پس در تقریب ِ تندترین شیب، $\Delta_{\mathsf{c}}(\boldsymbol{X}=\boldsymbol{x};X)$ از ضریب ِ جمله یِ مجذوری-نسبت-به- x در نما یِ $(\mathrm{G}D)_{\boldsymbol{X}|X}(\boldsymbol{x},x)$ به دست میآید: با

$$(\mathrm{G}D)_{\boldsymbol{X}|X}(\boldsymbol{x},x) = \frac{1}{\sqrt{2\,\pi\,\Delta^2}}\,\exp\left[-\frac{(x-\mathsf{E})^2}{2\,\Delta^2}\right], \tag{86}$$

E مقدار ی چشمداشتی یِ مشروط ِ X و Δ^2 خُدبسته‌گی یِ مشروط ِ X است. (82) را در (81) میگذارم:

$$(\boldsymbol{x}-x\,\boldsymbol{s})\,\mathbf{C}^{-1}\,(\boldsymbol{x}-x\,\boldsymbol{s}) = (\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s})\,[x-\xi(\boldsymbol{x})]^2 - (\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s})\,[\xi(\boldsymbol{x})]^2 + (\boldsymbol{x}\,\mathbf{C}^{-1}\,\boldsymbol{x}). \tag{87}$$

با گذاشتن ِ این در (80)، از (86) نتیجه میشود

$$[\Delta_{\mathsf{c}}(\boldsymbol{X}=\boldsymbol{x};X)]^2 = \frac{1}{\boldsymbol{s}\,\mathbf{C}^{-1}\,\boldsymbol{s}}. \tag{88}$$

از مقایسه یِ این با (42) دیده میشود

$$\Delta_{\mathsf{c}}(\boldsymbol{X}=\boldsymbol{x};X) = \mathsf{un}(x). \tag{89}$$

4 پانوشتها

[1] Lagrange

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

[3] Bayes

[4] Taylor