3.3. Інформаційний підхід до систематизації показників узгодженості експертних оцінок

Аналіз основних змістовних характеристик проекту (функціональних, фінансових, тимчасових і ін.), як правило, виконується із залученням групової експертизи, що допомагає приймати рішення в ситуаціях невизначеності. При цьому виникає проблема визначення узгодженості експертних висловлень (оцінок), що є показником достовірності інформації, одержаної від експертів.

Існує твердження про те, що узгодженість експертних оцінок визначається унімодальністю і симетричністю функції щільності розподілу імовірностей $p(x)$ [8]. Однак таке твердження, мабуть, є наближеним і несе в собі невизначеність, тому що наявність унімодальності не говорить про те, що вона точно відповідає визначеному закону розподілу або класу розподілів.

Зручною мірою невизначеності законів розподілу імовірностей є ентропія $H(X)$, яка обумовлена відомим функціоналом [30]:

$$H(X) = -\int_{-\infty}^{+\infty} p(x) \ln p(x) \, \mathrm{d}x. \qquad (3.10)$$

Відомо, що між $H(X)$ і $p(x)$ існує залежність, яка сформульована в положеннях інформаційної теорії вимірів [30] і яка характеризується тим, що величина ентропії, а отже, кількість інформації (інформативність), визначається видом закону розподілу.

Звідси виникає задача визначення величин $H(X)$ для ряду симетричних розподілів імовірностей, що дозволить оцінювати ентропійну, або інформаційну, міру узгодженості експертних оцінок.

Попередньо помітимо, що основні положення інформаційної теорії вимірів можуть бути поширені і на процедуру проведення експертизи.

Даний висновок ґрунтується на тому факті, що задача експертного оцінювання може трактуватись як задача виміру яких-небудь характеристик проектів з використанням шкали, наприклад, інтервального типу, а сам експерт може розглядатися як "вимірювач", на виході якого формуються ймовірнісні оцінки [8].

Рішення поставленої задачі почнемо з узагальнення виразу (3.10) на інтервальну шкалу експертного оцінювання (виміру), для чого розіб'ємо інтервал можливих значень X, обумовлених безперервним розподілом щільності імовірностей на рівні непересічні відрізки Δx, і розглянемо множину дискретних станів x_1, x_2,..., x_m з імовірностями $P_i = p(x_i)\,\Delta x$, $(i = 1, 2,\ldots, m)$.

Тоді

$$H(X) = -\sum_{i=1}^{m} p(x_i)\,\Delta x \log p(x_i)\,\Delta x =$$

$$= -\sum_{i=1}^{m} p(x_i)\Delta x \log p(x_i) - \sum_{i=1}^{m} p(x_i)\Delta x \log \Delta x.$$

У межі при $\Delta x \to 0$ з урахуванням співвідношення $\int\limits_{-\infty}^{+\infty} p(x)dx = 1$ одержимо

$$H(X) = -\int\limits_{-\infty}^{+\infty} p(x) \ln p(x) dx - \log \Delta\, x. \tag{3.11}$$

Перший доданок у цій сумі називається приведеною ентропією, цілком визначає інформативність експертних повідомлень, а величина log Δx залежить тільки від обраного інтервалу Δx, що визначає точність квантування станів, і при Δx = const вона постійна.

Визначимо, наприклад, величину $H(X)$ при заданій дисперсії $\sigma^2 = \int_{-\infty}^{+\infty} x^2 p(x)dx = \text{const}$, тобто для нормально розподілених значень експертних вимірів:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

Маємо

$$H(X) = -\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} \log\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\right] dx - \log\Delta x =$$

$$= \log\left(\sigma\sqrt{2\pi}\right)\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx + \frac{\log e}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} \frac{x^2}{2\sigma^2} dx - \log\Delta x = \tag{3.12}$$

$$= \log\left(\frac{\sigma}{\Delta x}\sqrt{2\pi e}\right).$$

Аналогічно для будь-якого вираженого аналітично закону розподілу результатів експертних вимірів величина $H(X)$ може бути визначена однозначно.

Наприклад, при розподілі Лапласа $H(X) = \log\left(\sigma\sqrt{2e}\right)$, для рівномірного прямокутного закону розподілу $H(X) = \log\left(\sigma 2\sqrt{3}\right)$, для трикутного закону Сімпсона $H(X) = \log\left(\sigma\sqrt{6e}\right)$ і т.д.

Для практичного використання величин $H(X)$ були введені [30]:

– інтервал невизначеності d результату виміру, що визначається через ентропію розглянутих вище законів розподілу відповідно таким способом:

$$H(X) = \log\left(\sqrt{2\pi e}\,\sigma\right) \quad \text{і} \quad d = \sqrt{2\pi e}\,\sigma \approx 4{,}133\sigma;$$

$$H(X) = \log\left(\sqrt{2e}\,\sigma\right) \quad \text{і} \quad d = \sqrt{2e}\,\sigma \approx 3{,}86\sigma;$$

$$H(X) = \log\left(2\sqrt{3}\,\sigma\right) \quad \text{і} \quad d = 2\sqrt{3}\,\sigma \approx 3{,}46\sigma; \tag{3.13}$$

$H(X) = \log\left(\sqrt{6e}\,\sigma\right)$ і $d = \sqrt{6e}\,\sigma \approx 4{,}04\,\sigma$;

– ентропійне значення погрішності виміру $\Delta_e = d/2$;
– ентропійний коефіцієнт $k = \Delta_e/\sigma$ даного закону розподілу. Так, для нормального розподілу, як було показано вище, $\Delta_e = \sigma\sqrt{2\pi e}/2 = \sqrt{2\pi e/2}\,\sigma = 2{,}066$ і $k = 2{,}066$; для рівномірного розподілу $\Delta_e = \sqrt{3}\sigma \approx 1{,}736$ та $k = 1{,}73$; для трикутного розподілу Сімпсона $k = \sqrt{6e}/2 \approx 2{,}02$; для розподілу Лапласа $k = 1{,}93$ і т.д.

К. Шеннон показав, що максимально можливе значення ентропійного коефіцієнта $k = 2{,}066$ має нормальний розподіл [30] і для найбільшого числа розподілів, що часто зустрічаються на практиці, коливається в досить широких межах.

Виходячи з відзначеного факту, відзначимо що для прийняття рішень про ступінь погодженості експертних оцінок з використанням ентропійного коефіцієнта k необхідно вибрати такі класи або сімейства симетричних законів розподілу імовірностей, що знаходяться у взаємозв'язку з нормальним розподілом і при зміні параметрів, які їх визначають, можуть прагнути до нього.

Зупинимо наш вибір на класі експоненційних розподілів і сімействі законів розподілу Стьюдента.

Клас експоненційних розподілів являє собою широкий клас симетричних розподілів, що може бути описаний наступною аналітичною моделлю вигляду [20]:

$$p(x) = \frac{\alpha}{2\lambda\sigma\Gamma(1/\alpha)}\exp\left(-\left|\frac{x - X_ц}{\lambda\sigma}\right|^{\alpha}\right) \qquad (3.14)$$

де $\lambda = \sqrt{\dfrac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$; σ – середньоквадратичне відхилення; $X_ц$ – координата центру розподілу; Γ – гама-функція; α – деяка характерна для даного розподілу постійна, його показник ступеня.

Для ілюстрації впливу показника ступеня α на форму розподілу, який описується, покладемо $X_ц = 0$, а добуток $\lambda\sigma = 1$. Тоді

$$p(x) = \frac{\alpha}{2\Gamma(1/\alpha)}\exp\left(-|x|\right)^{\alpha} = A(\alpha)\exp\left(-|x|^{\alpha}\right), \qquad (3.15)$$

де $A(\alpha)$ – множник розподілу, що нормує, залежний від його показника ступеня α.

При $\alpha < 1$ аналітична модель (3.14) і (3.15) описує розподіл з дуже пологими спадами, близькими за своїми властивостями до розподілу Коші. При $\alpha = 1$ вона відповідає розподілові Лапласа, при $\alpha = 2$ – нормальному розподілові Гаусса, при $\alpha > 2$ – вона описує розподіли, за своїми властивостями близькі до трапецієподібних, і, нарешті, при $\alpha \to \infty$ вона відповідає рівномірному розподілу.

Узагальнена модель (3.14) показує, що всі перераховані вище розподіли взаємозалежні і є представниками єдиного великого класу експериментальних розподілів, а параметр α характеризує їх форму і властивості. Ентропійний коефіцієнт експонентних розподілів також є однозначною функцією α:

$$k = \frac{1}{\alpha} e^{1/\alpha} \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \tag{3.16}$$

При цьому α можуть бути рівними не тільки 1, 2 або ∞, що відповідає розподілам Лапласа, Гаусса і рівномірному, але і приймати будь-які дробові і цілі значення, тобто модель (3.14) класу експонентних розподілів винятково зручна для опису розподілів значень вимірів [3].

Сімейство законів розподілу Стьюдента описує щільність імовірності значень середнього арифметичного, обчисленого по вибірці з n випадкових відліків з нормально розподіленої генеральної сукупності.

Відзначимо, що мова йде не про один якийсь "закон розподілу Стьюдента", а про ціле сімейство законів, тому що вид цього розподілу залежить від числа n відліків, за якими розраховується середнє значення.

У центрованому і нормованому вигляді сімейство розподілів Стьюдента описується виразом

$$p(x) = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\upsilon\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)\left(1+\frac{x^2}{\upsilon}\right)^{\frac{\upsilon+1}{\upsilon}}} \tag{3.17}$$

де Γ – гама-функція, а $\upsilon = n + 1$ – число ступенів свободи, що залежить від числа n усереднених відліків.

Ентропійний коефіцієнт k для розподілів Стьюдента з числом ступенів свободи $\upsilon > 4$ визначається в такий спосіб [20]:

$$k = \frac{\sqrt{\pi(\upsilon - 2)}\Gamma\left(\frac{\upsilon}{2}\right)}{2\Gamma\left(\frac{\upsilon+1}{2}\right)} e^{(\upsilon+1)\beta(\upsilon)} \tag{3.18}$$

де $\beta(\upsilon) = \sum_{-\infty}^{+\infty} \frac{(-1)^m}{\upsilon + m}$, тобто $\beta(1) = \ln 2, \quad \beta(2) = 1 - \ln 2$,

$$\beta(\upsilon) = \left[1 - \frac{1}{2} + \frac{1}{3} \Lambda - \frac{(-1)^{\upsilon-1}}{\upsilon - 1} - \ln 2\right](-1)^{\upsilon}$$

Розподілу Стьюдента з числом ступенів свободи від $\upsilon = 4$ до $\upsilon = \infty$ збігаються з розподілами класу експонентних з показниками ступеня від $\alpha = 0$ до $\alpha = 2$, але різко відрізняються від них за значеннями ентропійного коефіцієнта, що у розподілів Стьюдента значно більше, ніж у експонентних розподілів. У міру збільшення числа ступенів свободи ці розходження монотонно зменшуються і при $\upsilon \to \infty$ розподіл Стьюдента прагне до нормального розподілу з $k = 2{,}066$.

Викладені особливості класу експонентних розподілів і сімейства законів розподілу Стьюдента представлені в табл. 3.2, у якій зведені і систематизовані значення ентропійних коефіцієнтів k у порядку їх зростання. З метою додання монотонного характеру зростання ентропійного коефіцієнта в таблицю введений арксинусоїдальний закон розподілу зі значенням $k = 1{,}11$ і закон Симпсона (трикутний) зі значенням $k = 2{,}02$, що також часто використовуються для опису результатів вимірів. Наведені в таблиці значення ентропійного коефіцієнта можна умовно розбити на градації й одержувати оцінки погодженості експертних повідомлень (вимірів) якісного плану.

Таблиця 3.2

Закон розподілу	***k***	**Закон розподілу**	***k***
1. Експонентний (α = 1/4) $p(x)=\frac{1}{48}e^{-\sqrt[4]{\lvert x\rvert}}$	0,085	8. Лапласа (α = 1) $p(x)=\frac{1}{2}\mathrm{e}^{-\lvert x\rvert}$	1,92
2. Експонентний (α = 1/3) $p(x)=\frac{1}{12}\mathrm{e}^{-\sqrt[3]{\lvert x\rvert}}$	0,424	9. t-розподіл $n=6,\ \upsilon=5$	1,97
3. Арксинусоїдальний $p(x)=\begin{cases}0, & x<-a\\ \frac{1}{\pi\sqrt{a^2-x^2}}, & -a<x<a\\ 0, & x>a\end{cases}$	1,11	10. t-розподіл $n=7,\ \upsilon=6$	2,0
4. Експонентний (α = 1/4) $p(x)=\frac{1}{48}\mathrm{e}^{-\sqrt[4]{\lvert x\rvert}}$	1,35	11. t-розподіл $n=8,\ \upsilon=7$	2,013
5. Рівномірний (прямокутний) $p(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & x<a,\ x>b\end{cases}$	1,73	12. Сімпсона (трикутний) $p(x)=\begin{cases}0, & x<a\\ \frac{4(x-a)}{(b-a)^2}, & a<x<\frac{b+a}{2}\\ \frac{4(b-x)}{(b-a)^2}, & \frac{a+b}{2}<x<b \quad x>b\end{cases}$	2,02
6. Експонентний (α = 7) $p(x)=\frac{7}{2\Gamma(1/7)}\mathrm{e}^{-\lvert x\rvert^7}$	1,87	13. t-розподіл $n=11,\ \upsilon=10$	2,047
7. Стьюдента (t-розподіл) $p(x)=\frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\upsilon\pi}\Gamma\left(\frac{\upsilon}{2}\right)\left(1+\frac{x^2}{\upsilon}\right)^{\frac{\upsilon+1}{\upsilon}}}$ $\upsilon=4,\quad n=5$	1,90	14. Гаусса (нормальний) (α = 2) $p(x)=\frac{1}{\sqrt{\pi}}\mathrm{e}^{-\frac{x^2}{2}}$	2,066

Представимо таку розбивку в наступному вигляді:

$$\begin{cases} 0 < k_1 < 1{,}90 & \text{– низька узгодженість} \\ & \text{(експоненційні розподіли)} \\ & \text{з } 0 < \alpha < 1{,}0;\ \alpha = 7;\ \alpha = \infty \text{ (арксинусоїдальний} \\ & \text{закон; } t\text{ - розподіл з } \upsilon = 4 \text{ і } n = 5); \\ 1{,}90 \le k_2 \le 2{,}0 & \text{– висока узгодженість} \\ & \text{(розподіл Лапласа з } \alpha = 1; \\ & t\text{ - розподіл з } \upsilon = 5 \text{ і } n = 6;\ \upsilon = 6 \text{ і } n = 5); \\ 2{,}0 < k_3 \le 2{,}066 & \text{– практично повна узгодженість} \\ & (t\text{ - розподіл з } \upsilon = 7 \text{ і } n = 8; \text{ закон Сімпсона;} \\ & t\text{ - розподіл з } \upsilon = 10 \text{ і } n = 11; \text{ нормальний} \\ & \text{закон).} \end{cases} \tag{3.19}$$

Таким чином, розглянутий підхід дозволяє з позицій інформаційної теорії вимірів приймати рішення про ступінь узгодженості групових експертних оцінок.