MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
ÚSTAV TEORETICKÉ FYZIKY A ASTROFYZIKY

Praktická astrofyzika — základy

Príklady

Marek Chrastina

Brno, 14. novembra 2013

Masaryk University
Faculty of Science
Department of Theoretical Physics and Astrophysics
Kotlářská 2, 611 37 Brno

Practical astrophysics — elements

Examples

Mgr. Marek Chrastina, Ph.D.

The work has been partially supported by grants and research projects 7AMB13PL019, 7AMB12AT003 and LH12175 of the Ministry of Education, Youth and Sports of the Czech Republic. It was typeset in system TEX/LATEX.

Obsah

Kapitola 1

Deskriptívna štatistika

Majme súbor n meraní nejakej fyzikálnej veličiny. Hodnotu i-teho merania si označme x_i a jeho neistotu δx_i. Súbor môžeme skrátene zapísať ako $\mathbf{Z} = \{x_i, \delta x_i\}_{i=1,\dots,n}$ a nazývame ho štatistický súbor. Váhu merania w_i stanovíme ako $w_i = (\delta x_i)^{-2}$. Suma váh S_w je potom daná vzťahom $S_w = \sum\limits_{i=1}^{n} w_i$. Ak štatistický súbor $\mathbf{Z}$ zoradíme vzostupne podľa hodnôt x_i, bude v ňom platiť nerovnosť $x_1 \leq x_2 \leq \ldots \leq x_n$ a dostaneme usporiadaný súbor, ktorý si označíme $\mathbf{Z}^*$. Celkové rozpätie R je rozdiel najväčšej a najmenšej hodnoty súboru.

1.1 Stredné hodnoty a neistoty

1.1.1 Aritmetický priemer

Aritmetický priemer $\overline{x}$ a váhovaný aritmetický priemer $\overline{x}_w$ sú dané vzťahmi 1.1. Ich neistoty $\overline{\delta x}$ a $\overline{\delta x}_w$ sú dané vzťahmi 1.2.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{x}_w = \frac{1}{S_w}\sum_{i=1}^{n} x_i w_i \tag{1.1}$$

$$\overline{\delta x} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \delta x_i^2} \qquad \overline{\delta x}_w = \sqrt{\frac{1}{S_w}\sum_{i=1}^{n} \delta x_i^2 w_i} \tag{1.2}$$

1.1.2 Orezaný aritmetický priemer

Orezaný aritmetický priemer $\overline{x}_{T,p}$ získame tak, že z usporiadaného štatistického súboru $\mathbf{Z}^*$ odstránime $\mathrm{round}(p/2)$ najvyšších a rovnaký počet najnižších hodnôt a zo zvyšku spočítame aritmetický priemer. Parameter p sa udáva v percentách. Funkcia round je definovaná vzťahom 1.3. Funkcia $\mathrm{int}(x)$ vracia celočíselnú časť reálneho čísla x a funkcia $\mathrm{sgn}(x)$ je definovaná vzťahom 1.4.

$$\mathrm{round}(x) = \begin{cases} \mathrm{sqn}(x)\mathrm{int}(|x|), & |x| - \mathrm{int}(|x|) < 0.5 \\ \mathrm{sqn}(x)\mathrm{int}(|x|) + 1, & |x| - \mathrm{int}(|x|) \geq 0.5 \end{cases} \tag{1.3}$$

$$\mathrm{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} \tag{1.4}$$

1.1.3 Kvadratický priemer

Kvadratický priemer $\overline{x}_K$ a váhovaný kvadratický priemer $\overline{x}_{K,w}$ sú dané vzťahmi 1.5.

$$\overline{x}_K = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \qquad \overline{x}_{K,w} = \sqrt{\frac{1}{S_w}\sum_{i=1}^{n} x_i^2 w_i} \tag{1.5}$$

1.1.4 Geometrický priemer

Geometrický priemer $\overline{x}_G$ a váhovaný geometrický priemer $\overline{x}_{G,w}$ sú dané vzťahmi 1.6.

$$\overline{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad \overline{x}_{G,w} = \sqrt[S_w]{\prod_{i=1}^{n} x_i^{w_i}} \tag{1.6}$$

1.1.5 Harmonický priemer

Harmonický priemer $\overline{x}_H$ a váhovaný harmonický priemer $\overline{x}_{H,w}$ sú dané vzťahmi 1.7.

$$\overline{x}_H = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{-1}\right)^{-1} \qquad \overline{x}_{H,w} = \left(\frac{1}{S_w}\sum_{i=1}^{n} x_i^{-1} w_i\right)^{-1} \tag{1.7}$$

1.1.6 Medián

Medián $\widetilde{x}$ je hodnota, ktorá delí usporiadaný štatistický súbor $\mathbf{Z}^*$ na dve rovnako početné časti. Platí, že najmenej 50% nameraných hodnôt je menších alebo rovných a najmenej 50% nameraných hodnôt je väčších alebo rovných mediánu. Medián získame zo vzťahu 1.8, kde $k \in \mathbb{N}$.

$$\widetilde{x} = \begin{cases} x_{k+1}, & n = 2k+1 \\ \dfrac{x_{k+1} - x_k}{2}, & n = 2k \end{cases} \tag{1.8}$$

Váhovaný medián $\widetilde{x}_w$ usporiadaného štatistického súboru $\mathbf{Z}^*$ je daný vzťahom 1.9, kde každému bodu x_k prináleží hodnota W_k podľa vzťahu 1.10. Medián nemusí byť určený jednoznačne.

$$\widetilde{x}_w = \frac{(W_{j+1} - 0{,}5)x_j + (0{,}5 - W_j)x_{j+1}}{W_{j+1} - W_j}, \quad W_j < 0{,}5 < W_{j+1} \tag{1.9}$$

$$W_k = \begin{cases} \dfrac{1}{S_w}\dfrac{w_1}{2}, & k = 1 \\ \dfrac{1}{S_w}\left(\displaystyle\sum_{i=1}^{k-1} w_i + \frac{1}{2}w_k\right), & k > 1 \end{cases} \tag{1.10}$$

1.1.7 Modus

Modus $\widehat{x}$ je najčastejšie sa vyskytujúca hodnota v súbore meraní. Takýchto hodnôt môže byť viacero a preto modus nemusí byť určený jednoznačne.

1.1.8 Odchýlky

Rozptyl alebo variancia či disperzia $D(x) = s^2$ a štandardná odchýlka s sú dané vzťahmi 1.11, kde a je centrum rozpylu čiže stredná hodnota. Sú to vychýlené odhady neistôt. Nevychýlený odhad neistoty predstavuje σ_{odh} daná vzťahmi 1.12, kde $\overline{w} = S_w/n$ je stredná váha. Štandardná odchýlka v odhade skutočnej hodnoty μ aritmetickým priemerom $\overline{x}$ z n meraní je daná vzťahom $\sigma(\overline{x}) = \dfrac{\sigma_{odh}}{\sqrt{n}}$. Stredná veľkosť odchýlky (mad) a váhovaná stredná veľkosť odchýlky (wmad) sú dané vzťahmi 1.13.

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 \qquad s_w^2 = \frac{1}{S_w}\sum_{i=1}^{n}(x_i - a)^2 w_i \tag{1.11}$$

$$\sigma_{odh} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - a)^2}{n-1}} = \sqrt{\frac{n}{n-1}}s \qquad \sigma_{odh,w} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - a)^2 w_i}{\overline{w}(n-1)}} \tag{1.12}$$

$$\mathrm{mad} = \frac{1}{n}\sum_{i=1}^{n}|x_i - a| \qquad \mathrm{wmad} = \frac{1}{S_w}\sum_{i=1}^{n}|x_i - a|w_i \tag{1.13}$$

1.2 Pokročilá štatistika

1.2.1 Všeobecný centrálny moment. Šikmosť a špicatosť

Všeobecný centrálny moment k-teho rádu okolo bodu a je daný vzťahmi 1.14. Takto definovaný moment je vychýlený. Bezrozmerná veličina šikmosť $a_3 = \dfrac{m_3}{s^3}$ charakterizuje mieru asymetrie rozdeľovacej funkcie a pre symetrické rozdelenia je nulová. Kladná šikmosť naznačuje, že odľahlé body sa vyskytujú skôr vpravo od centra než vľavo (rozdelenie má tzv. pravý chvost) a väčšina hodnôt sa nachádza blízko vľavo od priemeru. U zápornej šikmosti je to napoak. Bezrozmerná veličina špicatosť $a_4 = \dfrac{m_4}{s^4} - 3$ charakterizuje mieru koncentrácie bodov okolo strednej hodnoty. Ak je a_4 blízke 0, potom hovoríme o súboroch s normálnou špicatosťou, ak je $a_4 < 0$, tak hovoríme o súboroch plochých a pre $a_4 > 0$ o súboroch špicatých. Špicaté súbory sú charakteristické tým, že väčšina hodnôt leží v blízkosti strednej hodnoty a hlavný vplyv na rozptyl majú málo pravdepodobné odľahlé body.

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^k \qquad m_{k,w} = \frac{1}{S_w}\sum_{i=1}^{n}(x_i - a)^k w_i \tag{1.14}$$

1.2.2 Kumulatívna distribučná funkcia

Kumulatívna distribučná funkcia príp. váhovaná kumulatívna funkcia $\Phi(x)$ je reprezentovaná lomenou čiarou s uzlovými bodmi $\{x_i, p_i\}$, kde p_i je dané vzťahmi 1.15.

$$p_i = \begin{cases} \dfrac{1}{2n}, & i = 1 \\ \dfrac{2i-1}{2n}, & i > 1 \end{cases} \qquad p_i = \begin{cases} \dfrac{w_1}{2S_w}, & i = 1 \\ \dfrac{2\sum_{j=1}^{j-1} w_j + w_i}{2S_w}, & i > 1 \end{cases} \tag{1.15}$$

1.2.3 Kvantily

Kvantil určený číslom $p \in (0, 1)$ je také číslo Q z intervalu $\langle x_1, x_n \rangle$, pre ktoré platí, že np hodnôt súboru je menších než Q a $(1 - p)n$ hodnôt je väčších ako Q. Vyjadruje odhad pravdepodobnosti, že náhodne vybrané číslo x bude menšie ako Q. Ak je p vyjadrené v percentách, tak sa mu hovorí percentil. Prvý kvartil Q_1 je určený $p = 0{,}25$ a tretí kvartil Q_3 je určený $p = 0{,}75$. Medzikvartilné rozpätie IQR je dané rozdielom tretieho a prvého kvartilu a slúži ako robustný odhad rozptylu. V prípade súboru s normálnym rozdelením môžeme kvantil určiť zo vzťahu 1.16.

$$Q = \Phi^{-1}(p) \tag{1.16}$$

1.2.4 Normálne rozdelenie

Rozdeľovacia alebo distribučná funkcia $f(x)$ udáva hustotu pravdepodobnosti. Distribučná funkcia normálneho rozdelenia s parametrami μ a σ je daná Gaussovou funkciou normovanou na 1 (vzťah 1.17). Kumulatívna distribučná funkcia normálneho rozdelenia je daná vzťahom 1.18, ktorý môžeme prepísať do tvaru 1.19, kde erf je tzv. chybová funkcia definovaná vzťahom 1.20. Početnosť výskytu hodnôt v intervale $\langle \mu - b/2, \mu + b/2 \rangle$ je daná vzťahom 1.21. Graf normálnej pravdepodobnosti je konštruovaný tak, aby sa na ňom súbory s normálnym rozdelením zobrazily ako priamky. Na x-ovú os sa vynášajú hodnoty meraní x_i a na pravdepodobnostnú (kvantilovú) y-ovú os hodnoty inverznej kumulatívnej distribučnej funkcie normálneho rozdelenia v bodoch p_i podľa vzťahu 1.22 prečíslované na p_i. Priamku pre príslušné normálne rozdelenie získame zo vzťahov 1.23.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{1.17}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] \mathrm{d}t \tag{1.18}$$

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right] \tag{1.19}$$

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \mathrm{d}t \tag{1.20}$$

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{n}{2}\left[\operatorname{erf}\left(\frac{x-(\mu-b/2)}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{x-(\mu+b/2)}{\sqrt{2}\sigma}\right)\right] \tag{1.21}$$

$$\Phi^{-1}(p) = \sqrt{2}\,\operatorname{erf}^{-1}(2p-1) \tag{1.22}$$

$$y = kx + b, \qquad k = \frac{\Phi^{-1}(0{,}75) - \Phi^{-1}(0{,}25)}{Q_3 - Q_1}, \qquad b = \Phi^{-1}(0{,}25) - kQ_1 \tag{1.23}$$

1.2.5 Histogram

Doporučený počet stĺpcov n_h v histograme pre súbor n meraní udáva vzťah 1.24 známy ako Sturgesovo pravidlo (Sturges 1926).

$$n_h = 1 + 3{,}322 \log n \tag{1.24}$$

Kapitola 2

Príklady

2.1 Príklad č. 1

Výsledkom meraní atmosférickej extinkcie z pozorovaní komét na observatóriu Skalnaté Pleso sú hodnoty extinkčných koeficientov vo vlnovej dĺžke 416 nm. Hodnoty sú uvedené v tab. 2.1 v jednotkách mag/vzdušnú hmotu (Mikulášek 2012).

Tabuľka 2.1: Namerané hodnoty extinkčných koeficientov

0,82±0,07	0,39±0,03	0,54±0,05	0,57±0,03	0,42±0,04
0,39±0,07	0,69±0,05	0,81±0,05	0,33±0,05	0,41±0,04
0,11±0,07	0,23±0,04	0,39±0,04	0,43±0,04	0,97±0,03
0,26±0,05	0,47±0,04	0,41±0,05	0,52±0,04	0,45±0,03

Nástrojmi deskriptívnej štatistiky charakterizujte tento súbor a špeciálne uveďte:

i. počet meraní a ich charakter (spojitý, diskrétny), minimálnu a maximálnu hodnotu extinkcie a celkové rozpätie

ii. stanovte váhy jednotlivých meraní a diskutujte, či je v tomto prípade vhodné tieto váhy použiť

iii. aritmetický priemer, váhovaný aritmetický priemer a ich neistoty

iv. geometrický priemer a váhovaný geometrický priemer, harmonický priemer a váhovaný harmonický priemer, kvadratický priemer a váhovaný kvadratický priemer

v. medián a váhovaný medián

vi. orezaný aritmetický priemer pre 10% a 20%

vii. rozptyl s^2, štandardnú odchýlku s, odhad neistoty σ_{odh} a strednú veľkosť odchýlky. Veličiny spočítajte s centrom v aritmetickom priemere a v mediáne a v oboch variantách (bez váh a s váhami)

viii. porovnajte odhady μ a σ pre normálne rozdelenie získané rôznymi metódami

ix. šikmosť a špicatosť s centrom v aritmetickom priemere v oboch variantách (bez váh a s váhami) a porovnajte ich s normálnym rozdelením. Aký je to typ súboru?

x. graf kumulatívnej distribučnej funkcie pre obe varianty (bez váh a s váhami) a pomocou neho stanovte hodnoty prvého a tretieho kvartilu a medzikvartilného rozpätia pre obe varianty (bez váh a s váhami). Do grafu vykreslite aj kumulatívnu distribučnú funkciu normálneho rozdelenia.

xi. stanovte optimálny počet stĺpcov v histograme a zostrojte ho. Doporučujem stĺpce histogramu centrovať na násobky 0,2. Do histogramu vykreslite aj krivku, ktorá by zodpovedala prípadu, že súbor má normálne rozdelenie a diskutujte.

xii. odhadnite modus rozdelenia

xiii. diskutujte tvar rozdeľovacej funkcie s vedomím, že konštantná zložka extinkčného koeficientu vo vlnovej dĺžke 416 nm spôsobená Rayleigho rozptylom na náhodných zhlukoch molekúl vzduchu je 0,262 mag/vzdušnú hmotu

xiv. zostrojte graf normálnej pravdepodobnosti

2.2 Príklad č. 2

Majme štatistický súbor $\mathbf{Z} = \{x_i, \delta x_i\}_{i=1,\dots,n}$. Ukážte, že:

i. funkcionál $S_w(a) = \sum_{i=1}^{n} (x_i - a)^2 w_i$ nadobúda minimum práve pre váhovaný aritmetický priemer $\overline{x}_w$

ii. funkcionál $\text{mad} = \frac{1}{n}\sum_{i=1}^{n} |x_i - a|$ nadobúda minimum práve pre neváhovaný medián $\widetilde{x}$

iii. prvý kvartil súboru s normálnym rozdelním sa nachádza vo vzdialenosti $0{,}6745\,\sigma$ od centra rozptylu μ

iv. σ_{odh} je nevychýleným odhadom neistoty aritmetického priemeru $\overline{x}$

v. pre centrálny moment k-teho rádu $\mathbb{E}((X-\mu)^k) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}(x-\mu)^k \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ súboru s normálnym rozdelením platí vzťah 2.1 a na jeho základe vyčíslite šikmosť a špicatosť

$$\mathbb{E}((X-\mu)^k) = \begin{cases} \sigma^k \dfrac{(2n)!}{n!2^n}, & k = 2n \\ 0, & k = 2n+1 \end{cases} \tag{2.1}$$