بسم الله الرحمن الرحیم   جلسه   ۱۴۰۰/۸/۱۱

* Bias
* Data modeling

عامل سوگیری   Bias factor

در رویکرد تدوین مدل نظری ← اصول موضوعه / ← ساختارهای پدیده‌شناختی } مدل شکل می‌گیرد

$\rightarrow f = ma \leftarrow$

$\uparrow$

$f = \checkmark$

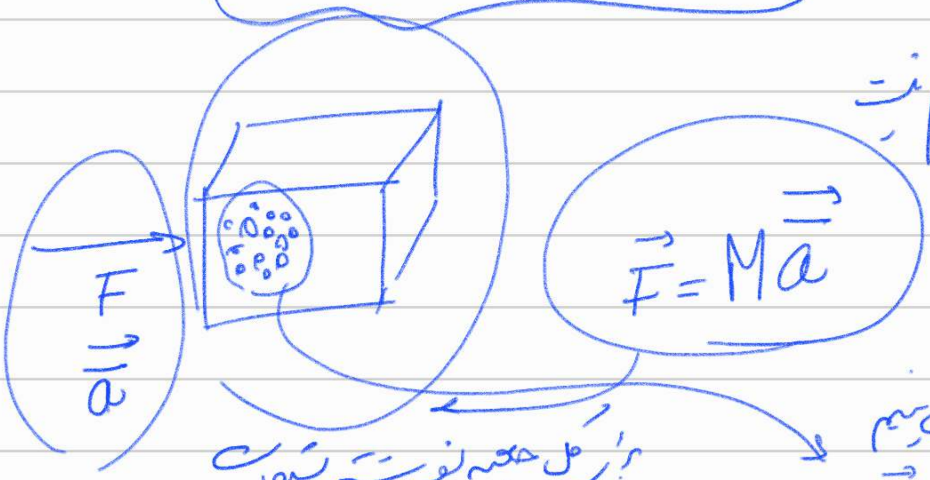
$a = \frac{dv}{dt} \checkmark$ $\rightarrow$ $m = \checkmark$

اندازه‌گیری انجام دهیم

$f = \checkmark$

$a = \checkmark$

حجم ثابت



از مدل جعبه نوشته شود

تکرار را در آزمایش می‌دهیم

$\vec{f} = \checkmark$

$\vec{a} = \checkmark$

$m = \checkmark$ — $\underline{M}$

مشاهده آزمایش   مدل ثابت

$$m \propto M$$

$$m = BM \quad — \quad M = B'm$$

Bias factor

$\underbrace{m = \checkmark}$ $\longrightarrow$ $\underbrace{M}$

$\rightarrow \boxed{B = ?} \leftarrow$

مدیریت

انگیزه ←

تباین فیزیکی در کوه Δ

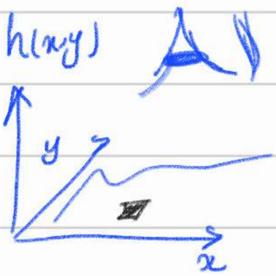
$$\delta_\Delta \equiv \frac{n_\Delta - \langle n_\Delta \rangle}{\langle n_\Delta \rangle}$$

مثال: فرض کنید در یک آزمایش ← AFM از:

$h(x,y)$

یک سطح به دست می‌آید

$$\delta_\Delta \stackrel{\Delta}{=} B\,\delta$$



$h(x,y) \rightarrow \boxed{(1+2)D}$

$$\delta_{peak} = B\,\delta_h$$

$$\delta_h \equiv \frac{h(x,y) - \langle h \rangle}{\langle h \rangle}$$

$$\delta_\Delta = \frac{n_{peak} - \langle n_{peak} \rangle}{\langle n_{peak} \rangle}$$

حالت کلی

$$\boxed{\delta_{peak}(\vec{r},\vartheta) = B(\vec{r},\vartheta)\,\delta_h(\vec{r})}$$

حالت خاص

$$\boxed{B = cts}$$

اگر B می‌دانیم به صورت غیرمستقیم به مطالعه کمیت مورد توجه می‌پردازیم

$$\langle \delta_\Delta(\vec{r}_i)\,\delta_\Delta(\vec{r}_j) \rangle = \left\langle \left(\frac{n_\Delta(\vec{r}_i) - \langle n_\Delta \rangle}{\langle n_\Delta \rangle} \right) \left(\frac{n_\Delta(\vec{r}_j) - \langle n_\Delta \rangle}{\langle n_\Delta \rangle} \right) \right\rangle$$

فرض $R = |\vec{r}_i - \vec{r}_j|$

$$= \frac{\langle n_\Delta(\vec{r}_i)\, n_\Delta(\vec{r}_j) \rangle}{\langle n_\Delta \rangle \langle n_\Delta \rangle} - 1$$

$$= \Psi_\Delta(R) \quad \text{Un-Weighted TPCF of } \Delta$$

خود میدان $\langle \delta \delta \rangle = C_{\delta\delta}(R)$

فرض ات

if $\delta_\Delta = B\,\delta \rightarrow \langle \delta_\Delta \delta_\Delta \rangle = B^2 \langle \delta\delta \rangle$

$$\psi_\Delta(R) = B^2\, \underbrace{C_{\delta\delta}(R)}_{\text{Weighted TPCF}}$$

Un-Weighted TPCF

Weighted TPCF

$B = ?$

Example. sharp clip

برش تیز

$$n_{sc} = \Theta(h-\alpha)$$

$$\boxed{\psi_{sc}(R) \overset{?}{=} B^2 \langle \underset{r_i}{h}\, \underset{r_j}{h} \rangle \overset{?}{=} B^2 C_{hh}(R)} \;\longrightarrow\; B^2 = ?$$

$$\psi_{sc}(R) = \frac{\langle \Theta(h-\alpha)\, \Theta(h'-\alpha) \rangle}{\langle \Theta(h-\alpha) \rangle \langle \Theta(h-\alpha) \rangle} - 1$$

$R = |r_i - r_j|$

Un-Weighted TPCF

$$= \frac{\int dh(r_i)\, \Theta(h(r_i)-\alpha) \int dh(r_j)\, \Theta(h(r_j)-\alpha)\, \overbrace{p_{joint}(h(r_i)\, h(r_j))}^{C_{hh}(R)}}{\underbrace{\int_{-\infty \to \alpha}^{+\infty} dh(r_i)\, \Theta(h(r_i)-\alpha)\, p(h(r_i))}_{\Delta} \underbrace{\int dh(r_j)\, \Theta(h(r_j)-\alpha)\, p(h(r_j))}_{\Delta}} - 1$$

$P_{joint}(h(r_i)\, h(r_j)) = ?$ Bivariate Gaussian function

فرض می‌کنیم

$$= \frac{1}{2\pi\left[\mathrm{Det}(C)\right]^{1/2}}\, e^{-\frac{h^T \cdot C^{-1} \cdot h}{2}}$$

Weighted TPCF

$$C = \begin{bmatrix} \langle h(r_i)\, h(r_i)\rangle & \langle h(r_i)\, h(r_j)\rangle \\ \langle h(r_j)\, h(r_i)\rangle & \langle h(r_j)\, h(r_j)\rangle \end{bmatrix} = \begin{bmatrix} \sigma_h^2 & C_{hn}(R) \\ C_{hh}(R) & \sigma_h^2 \end{bmatrix}$$

$$\psi_{sc}(R) = \frac{\text{صورت کسر}}{\text{مخرج کسر}} - 1$$

$$\gamma \equiv \text{صورت کسر} = \int_{\alpha}^{+\infty} dh(r_i) \int_{\alpha}^{+\infty} dh(r_j)\, \frac{1}{2\pi\sqrt{\mathrm{Det}(C)}}\, e^{\frac{(h(r_i), h(r_j)) \cdot C^{-1} \cdot \begin{pmatrix} h(r_i) \\ h(r_j) \end{pmatrix}}{2}}$$

$$= \frac{1}{2\pi}\left[\sigma^4 - C_{hn}^2(R)\right]^{-1/2} e^{-\frac{1}{2}\frac{\sigma(h^2(r_i) + h^2(r_j)) - 2C_{hn}(R)\, h(r_i)\, h(r_j)}{2[\sigma^4 - C_{hn}^2(R)]}}$$

$$\Delta \equiv \text{مخرج کسر} = \int_{\alpha}^{+\infty} dh(r_i)\, \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{h^2(r_i)}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{\alpha^2}{2\sigma^2}}$$

$p(h(r))$ زیر

$\boxed{\alpha >> 1}$

Δ هم که را بسط دادم

$$\boxed{\lim_{\alpha \to \infty}} \psi_{sc}(R) = \frac{\gamma}{\Delta^2} - 1 = \frac{\gamma - \Delta^2}{\Delta^2}$$

$$= e^{\left(\frac{\alpha^2\, C_{hn}(R)}{\sigma^2}\right)} - 1$$

$$\boxed{\lim_{\substack{\alpha\to\infty \\ \left\{ \begin{array}{l} R\to\infty \\ C_{hh}(R)\to 0 \end{array}\right.}} \psi_{sc}(R)} = \alpha^2 \frac{C_{hh}(R)}{\sigma^2} = \frac{\alpha^2}{\sigma^2} C_{hh}(R)$$

$$= B^2\, C_{hh}(R)$$

تابع همبستگی بدون وزن ← خوشگی نواحی (حبابک) ای که در مقادیر زیاد در فواصل زیاد

α

$$\psi_{sc}(R) = \frac{\alpha^2}{\sigma^2} C_{hh}(R)$$

$\boxed{\alpha \gg 1}$ $\boxed{R\to\infty}$

$$B^2 = \frac{\alpha^2}{\sigma^2} = \frac{\psi_{sc}(R)}{C_{hh}(R)}$$

B^2

$$\langle \delta_{sc}\delta_{sc}\rangle = \psi_{sc} = B^2 C(R)$$

↑ Bias factor

فقط آنهایی که مقدارشان از α بالاتر است →

$$\boxed{\delta_{sc} \triangleq B\, h(\bar{r})}$$

ثابت

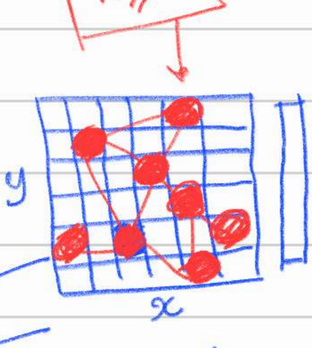
بدون واریانس

$$= \frac{\alpha}{\sigma} h(r) = \alpha\, \tilde{h}(\vec{r})$$

$\boxed{h>\alpha}$

در این وزن‌دهی فقط نواحی‌ای که در مقدار $h>\alpha$ نشان می‌دهد.

$$\boxed{\delta_{sc}(h>\alpha\gg 1) = \alpha\, \tilde{h}(\vec{r})}$$



Density plot

$\boxed{\begin{array}{l}\text{Un-Weighted TPCF} \\ \text{Weighted TPCF of field}\end{array}}$

داستان کلی و خلاصه‌اش

بود

$$\boxed{\delta_{se} \triangleq \alpha\, \bar{h}(\bar{r})}$$

$\alpha \gg 1$

پژوهش نیز ۱۳۹۹

مجید سهرابی و خودم

Data modeling

هدف ما یافتن مقادیر بهینه پارامترهای آزاد مدل است

← بیشترین تطابق بین پیش‌بینی‌های مدل‌ها و نظریه‌ها با آنچه که در آزمایش بدست آمده است.

Outline

1. Measurements ← آزمایشگر انجام می‌دهد.
2. Errors Estimation and propagation ← قبلاً دیدیم
3. Model Selection ← براساس آنکه چه پارامترهایی لازم است تعیین بشود

$\{\Theta\}$: Model's free Parameters

هیچ سازوکار خوش‌تعریفی از خود مدل برای تعیین دقیق مقدار آنها بدست نمی‌آید.

4. Datamodeling
   - Frequentist ← سنتی
   - Bayesian ← مدرن → تعداد داده‌ها زیاد شد / الگوریتم‌های محاسباتی و سخت‌افزار

1702 (Thomas Bayes)

1812 (Laplace)

1990 (smith)

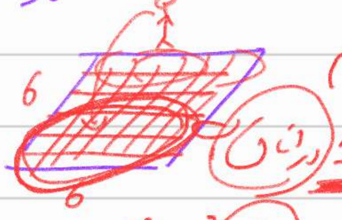
Now

? 5 Determination of Confidence Intervals for free Parameter

بازه تطابق

ریشه در استفاده از منطق و رویکرد Bayesian دارد.

پارامترهای مدل کمیت های مجهولی هستند که ما آنها به صورت احتمالی مقدار بهینه آنها را پیدا می‌کنیم.



مکان و اندازه جسم

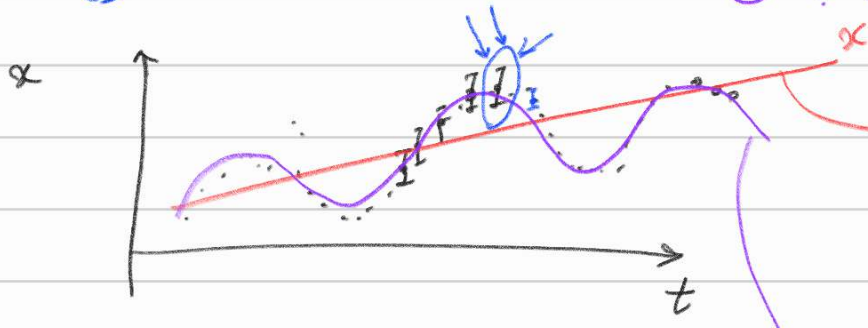
گم شده با احتمال 1 در زمین

با احتمال $\frac{1}{36}$ در یکی از سوراخ‌ها است

مقدار بهینه پارامتر مدل با احتمال ؟ به ازای ؟ است

? 6 Goodness of fit

میزان خوبی برازش



$x(t) = at$

$\{\theta\} = a \;—\; a = 1$

$x(t) = g(t)$

$= at \sin(bt)$

رشد را نشان دهد — تناوب

$\{\theta\} = \{a, b\}$

a_{best} ✓ $\quad b_{best}$ ✓

$M = 2$

$x_2 = a_{best}\, t \sin(b_{best}\, t) \leftarrow$ Goodness of fit

$M = 1$

$x_1 = a'_{best}\, t \leftarrow$ Goodness of fit

واضح است بدیهی است که هر چقدر تابع مدل پیچیده‌تر باشد $\leftarrow \{\theta_i\},\ i = 1, \dots, M$

تعداد پارامترهای آزاد مدل

خوبی برازش بیشتر می‌شود → تا حد نوشته ← مدلی معقولیت بیشتری دارد که در عین سادگی تطابق خوبی داشته باشد

→ AIC — Akaike Information Criterion

BIC — Bayesian Information Criteria

? 7 Fisher forecast , Fisher information Method

بدون انجام آزمایش ← خوبی مدل و جایگاه آن در قیاس با سایر مدل‌ها و ناحیه تمرکز تطابق پارامترها صحبت می‌کنیم، محاسبات انجام می‌دهد.

از اطلاعات مناسب استفاده می‌کنیم.

در جلسه آینده مفهوم Bayesian و frequentist و ارائه کتب.