فصل ۱

مدل کردن داده‌های آماری

در علوم تجربه‌ای و بسیاری از آزمایش‌ها غالباً با برازش یک مدل پیچیده تئوری با نتایج بدست آمده از آزمایش روبرو می‌شویم. در این رهیافت بایستی پارامترهای تئوری را تخمین زده و محدوده اعتبار آنها را تعیین کنیم. گاهی مواقع نیز به دنبال مدلی که بهترین تطابق را با مشاهدات و اندازه‌گیری‌های مستقیم دارد، می‌باشیم. ممکن است که مدل شکل بسیار ساده‌ای مانند یک چند جمله‌ای یا یک شکل گوسی داشته باشد. در چنین مواقعی به سادگی با روش‌های استاندارد می‌توان مدل را کمّی کرد. از طرف دیگر ممکن است نیاز باشد که هرکدام از پارامترها، مدل مجزایی را ارضا کنند. در این وضعیت مدل کردن با چالش‌هایی روبرو می‌شود که روش‌های متداول راهگشا نمی‌باشد. در مدل کردن داده‌ها اساساً با مجموعه‌ای از درونیابی‌های مقید شده روبرو هستیم که به کمک آنها مجموعه‌ای از نقاط گسسته را به توابع مناسب پیوسته مرتبط می کنیم. برای اهداف ذکر شده بایستی یک تابع مناسب[۱] انتخاب یا طراحی کنیم. تطابق مدل و داده‌ها به کمک پارامترهای تابع مناسب تعیین می‌شود و در واقع بهترین پارامترها، تابع مناسب را کمینه می‌کنند که به آنها بهترین پارامترهای برازش می‌گوییم. به سبب وجود عدم قطعیت در داده‌های تجربی، هرگز انتظار نداریم که یک مشاهده با مدل تطابق کامل داشته باشد، حتی اگر مدل کاملاً صحیح باشد. برای تعیین اینکه مدل به کار رفته برای توصیف داده‌ها مناسب است یا نه، از محک ارزش

[۱] Meritfunction

http://facultymembers.sbu.ac.ir/movahed/

برازش [۲] استفاده می‌کنیم که در ادامه آن را معرفی می‌کنیم. در بیشتر مواقع پس از بدست آوردن مقدار مناسب برای پارامترها بایستی محدوده اعتبار و خطای آنها را محاسبه کنیم. این کار باعث می‌شود که اثر خطاهای آماری در نظر گرفته شود. به طور خلاصه برای مدل کردن داده‌ها با مراحل زیر روبرو هستیم:

۱) تعیین شکل مدل و پارامترهای آن

۲) بهرگیری از روش درست‌نمایی (MLE) [۳]

۳) تخمین خطای پارامترها و همبستگی بین آنها با استفاده از درست‌نمایی نسبی بی اثر شده (MRL) [۴] و کانتورهای تراز تطابق [۵]

۴) تعیین میزان درستی برازش مدل با نتایج تجربی

اجازه دهید قبل از اینکه به روش عام تعیین مدل و تحلیل درست‌نمایی بپردازیم، روش معمول کمترین مربعات [۶] را معرفی کنیم و محدودیت‌های آنرا برشماریم و سپس به روش کلی آن بپردازیم.

۱.۱ کمترین مربعات به عنوان یک تخمین‌گر بیشینه احتمال

فرض کنید مشاهدات ما شامل N نقطه به صورت

$$(x_i, y_i) \quad , i = ۱, ۲, ..., N \tag{۱.۱}$$

باشد که x_i پارامتر و y_i کمیت منتسب به x_i می‌باشد. اگر مدل در نظر گرفته شده دارای M پارامتر باشد آنگاه:

$$a_j \qquad , \qquad j = ۱, ۲, ..M \tag{۱.۲}$$

[۲] Goodness of fit

[۳] Maximum Likelihood Estimation

[۴] Marginalized relative likelihood

[۵] Level og confidence

[۶] Least square

http://facultymembers.sbu.ac.ir/movahed/

و مدل یک رابطه تابعی بین اندازه‌گیری‌های مستقل انجام شده و متغیرهای وابسته پیش بینی می‌کند:

$$\begin{aligned} y(x) &= y(x; a_1, a_2, ..., a_M) \\ &= a_1 + a_2 x + ... + a_M x^{M-1} \end{aligned} \tag{۱.۳}$$

شکل کلی‌تر معادله (۱.۳) می‌تواند به صورت زیر باشد:

$$y(x) = \sum_{k=1}^{M} a_k X_k(x) \tag{۱.۴}$$

در معادله (۱.۴) تابع $X_k(x)$، هر تابع دلخواهی از x است که تابع پایه نامیده می‌شود. آنچه که در رهیافت کمترین مربعات انجام می‌گیرد این است که تابع مناسب[۷] را تشکیل می‌دهیم و در فضای پارامترها جستجو کرده دسته پارامترهایی از مدل را انتخاب می‌کنیم که این تابع را کمینه کند شکل این تابع به صورت زیر است:

$$\chi^2(\{a\}) = \sum_{i=1}^{N} \frac{\left[y_i - \sum_{k=1}^{M} a_k X(x_i)\right]^2}{\sigma_i^2} \tag{۱.۵}$$

در این معادله σ_i در واقع واریانس iامین اندازه‌گیری می‌باشد. اگر خطای اندازه‌گیری‌ها معلوم نباشد آنها را یکسان و برابر با واحد در نظر می‌گیریم. کمینه کردن تابع χ^2 به روشهای متعددی انجام می‌شود [? ،?]. اجازه دهید به بررسی دو سئوال اساسی در مورد معادله (۱.۵) بپردازیم:

۱ – چرا بهترین پارامترهای مدل با کمینه کردن این تابع بدست می آیند؟

۲ – اساساً، اصل کلی که معادله (۱.۵) بر آن استوار است چیست؟

برای پاسخ به این سوالات به تخمین‌گر بیشینه احتمال نیاز داریم. به دنبال پاسخ به این سئوال خواهیم بود که هر دسته از پارامترهای تعیین شده {a} با چه احتمالی درست خواهد بود؟ منظور از درست بودن یعنی به چه میزان مدل با آزمایش همخوانی دارد؟ و به بیانی دیگر هر دسته از پارامترها {a}، با چه احتمالی نتایج بدست آمده از آزمایش را بدست می‌دهند؟ اکنون آماده‌ایم تا ارتباط معنایی با معادله (۱.۵) برقرار کنیم. فرض کنید هر اندازه‌گیری y_i مستقل از دیگر اندازه‌گیری‌ها بوده و حول مقداری که توسط مدل

[۷] mericfunction

http://facultymembers.sbu.ac.ir/novahed/

داده می‌شود یعنی $y(x_i)$ دارای تابع توزیع گوسی[۸] با انحراف معیار σ_i باشد بنابراین برای هر اندازه‌گیری (x_i, y_i) داریم:

$$P_i \equiv P(x_i, y_i; a_1, .., a_M) \tag{۱.۶}$$

پس احتمال کل برابر است با:

$$\begin{aligned} L(a|(x,y)) = P((x,y)|a_1, ..., a_M) &= \prod_{i=1}^{N} P_i \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[\frac{(y_i - y(x_i))^2}{2\sigma_i^2}\right] \end{aligned} \tag{۱.۷}$$

واضح است تلاش برای بیشینه کردن احتمال $P((\mathbf{x},\mathbf{y})|a_1, ..., a_M)$ به کمینه کردن معادله (۱.۵) برای دستیابی به مدل بهینه می‌انجامد. برازش کمترین مربعات در حقیقت یک تخمین‌گر بیشینه احتمال است، اگر اندازه‌گیری‌ها مستقل از یکدیگر باشند و حول مقدار متوسط به صورت یک تابع گوسی توزیع شده باشند. در بیشتر موارد که تعداد اندازه‌گیری‌ها زیاد نباشد و در ضمن مستقل از یکدیگر نباشند، فرض گوسی بودن تابع توزیع با چالش روبرو شده و نیازمند به استفاده مستقیم از تخمین‌گر بیشینه احتمال هستیم.

در رهیافت کمترین مربعات آنچه که در عمل انجام می‌شود این است که در فضای پارامترهای مدل حرکت کرده و دسته پارامترهای مدل مطابق با کمینه مطلق χ^2 پیدا می‌کنیم. هیچ‌گاه به دلیل محدودیت های موجود در اندازه‌گیری نمی‌توان مدلی دقیقاً منطبق با اندازه‌گیری و بالعکس یافت. بنابراین تعیین کیفیت مدل بدست آمده از اهمیت ویژه‌ای برخوردار است. محک کیفیت برازش در واقع یافتن میزان احتمال اینکه دسته پارامترهای موجود در فضای پارامتری، χ^2هایی بزرگتر و مساوی χ^2_{min} بدست می‌دهد. یعنی :

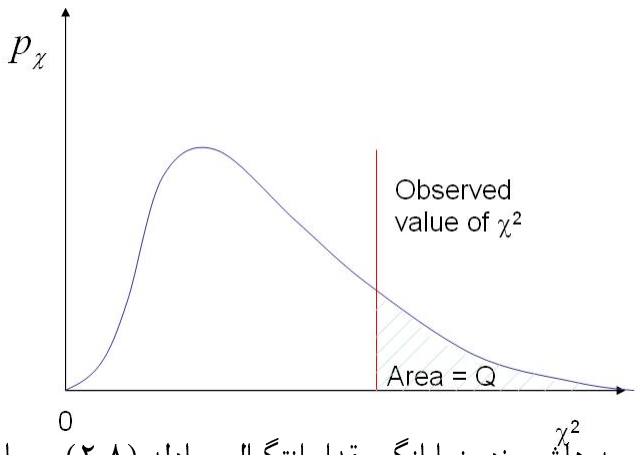
$$P_\chi(\chi^2; \nu) = \int_{\chi^2}^{\infty} p_\mu(\mu, \nu) d\mu = Q(\frac{\nu - 2}{2}, \frac{\chi^2}{2}) \tag{۱.۸}$$

در معادله (۱.۸)، $p_\mu(\mu, \nu)$ برابر است با [?] :

$$p_\mu(\mu, \nu) = \frac{\mu^{\frac{\nu-2}{2}} e^{-\frac{\mu}{2}}}{2^{\nu/2}\Gamma(\nu/2)} \tag{۱.۹}$$

[۸] فرض گوسی بودن در حقیقت از قضیه حد مرکزی آمده است ولی همیشه انتخاب تابع توزیع گوسی صحیح نمی‌باشد

http://facultymembers.sbu.ac.ir/movahed/



شکل ۱—۱: ناحیه هاشور زده نمایانگر مقدار انتگرال معادله (۲.۸) می‌باشد

و $\nu = N - M$ نمایانگر تعداد درجه آزادی مدل می‌باشد و $Q(\frac{\nu-۲}{۲}, \frac{\chi^۲}{۲})$ تابع گامای ناکامل می‌باشد. اگر $Q < ۰.۱$ یا $Q > ۰.۹$ بدست آید بدین معنی است که مدل ارائه شده خوب نمی‌باشد. شکل (B—۱) به صورت طرحوار، تابع احتمال P_χ را نشان می‌دهد. برای $\nu = ۱$، شکل تابع توزیع $\chi^۲$ را اثبات می‌کنیم برای حالت کلی می‌توان به مرجع [?] مراجعه کرد. در حقیقت تابع $\chi^۲$ به صورت زیر ساخته می‌شود یعنی :

$$\chi^۲ = x_۱^۲ + x_۲^۲ + ... + x_\nu^۲ \qquad (۱.۱۰)$$

و همچنین تابع توزیع xها گوسی می‌باشد. اکنون با این فرض می‌خواهیم شکل تابع توزیع $\chi^۲$ با یک درجه آزادی تعیین کنیم. پس

$$\begin{aligned} P(\chi^۲; ۱) = P(x^۲ < \chi^۲) &= P(-\sqrt{\chi^۲} < x < \sqrt{\chi^۲}) \\ &= \frac{۱}{\sqrt{۲\pi}} \int_{-\sqrt{\chi^۲}}^{+\sqrt{\chi^۲}} e^{\frac{۱}{۲}x^۲} dx = \frac{۲}{\sqrt{۲\pi}} \int_{۰}^{+\sqrt{\chi^۲}} e^{\frac{۱}{۲}x^۲} dx \qquad (۱.۱۱) \end{aligned}$$

با تغییر متغیر $x^۲ = \mu$ داریم:

$$P(\chi^۲; ۱) = \frac{۱}{\sqrt{۲\pi}} \int_{۰}^{\chi^۲} \mu^{-\frac{۱}{۲}} e^{-\frac{۱}{۲}\mu} d\mu \qquad (۱.۱۲)$$

انتگرالده معادله (۱.۱۲) دقیقاً برابر با معادله (۱.۹) به ازاء $\nu = ۱$ می‌باشد.

اکنون در مرحله‌ای هستیم که می‌توان به طور دقیق به برآورد دسته پارامترهای مدل برای توصیف آزمایش باشیم. بار دیگر به معادله (۱.۵) توجه کنید. می‌توان کمیت‌های زیر را تعریف کنیم:

$$A_{ij} = \frac{X_j(x_i)}{\sigma_i}$$

http://facultymembers.sbu.ac.ir/movahed/

$$b_i = \frac{y_i}{\sigma_i} \tag{۱.۱۳}$$

که $\mathbf{A}$ ماتریسی از مرتبه $N \times M$ می‌باشد و ماتریس طرح[۹] نامیده می‌شود. $\mathbf{b}$ بردار سطری از مرتبه N بوده و $\mathbf{a}$ که دسته پارامترها را مشخص می‌کند، برداری ستونی از مرتبه M می‌باشد. بنابراین ماتریس طرح به صورت زیر خواهد بود:

$$\begin{pmatrix} \frac{X_1(x_1)}{\sigma_1} & \frac{X_2(x_2)}{\sigma_1} & . & . & . & \frac{X_M(x_1)}{\sigma_1} \\ . & . & & & & . \\ . & & & . & & . \\ . & & & & . & . \\ \frac{X_1(x_N)}{\sigma_N} & \frac{X_2(x_N)}{\sigma_N} & . & . & . & \frac{X_M(x_N)}{\sigma_N} \end{pmatrix} \tag{۱.۱۴}$$

بر اساس تخمین‌گر بیشینه احتمال دسته پارامترهای مناسب با کمینه کردن معادله زیر بدست می آیند:

$$Min \parallel b - Aa \parallel \tag{۱.۱۵}$$

کمینه کردن معادله (۱.۱۵) به طور کلی با کمک دو رهیافت زیر یعنی[?]:

الف : با کمک توابع نرمال[۱۰]

ب : با کمک تجزیه مقادیر تکین[۱۱]

بر اساس رهیافت الف، وقتی معادله (۱.۵) کمینه می‌شود که مشتق χ^2 نسبت به تمام پارامترها $\{a_i\}$ صفر شود. این شرط باعث تولید M معادله کوپل، می‌شود:

$$0 = \frac{\partial \chi^2}{\partial a_k} = \sum_{i=1}^{N} \left[y_i - \sum_{j=1}^{M} a_j X_j(x_i) \right] \frac{X_k(x_i)}{\sigma_i^2}, \qquad k = 1, ..., M \tag{۱.۱۶}$$

معادله (۱.۱۶) را می‌توان به صورت یک معادله ماتریسی نوشت:

$$\sum_{j=1}^{M} \alpha_{kj} a_j = \beta_k \tag{۱.۱۷}$$

---

[۹] Design Matrix

[۱۰] Normal equations

[۱۱] Singularity values decomposion

http://facultymembers.sbu.ac.ir/movahed/

که

$$\alpha_{kj} = \sum_{i=1}^{N} \frac{X_j(x_i)X_k(x_i)}{\sigma_i^2} \tag{1.18}$$

و

$$\beta_k = \sum_{i=1}^{N} \frac{y_i X_k(x_i)}{\sigma_i^2} \tag{1.19}$$

معادلات (۱.۱۸) و (۱.۱۹) به معادلات نرمال معروف هستند. نهایتاً دسته پارامترها $\{\mathbf{a}\}$ به صورت زیر قابل محاسبه است:

$$a_j = \sum_{k=1}^{M} [\alpha_{ik}]^{-1} \beta_k = \sum_{i=1}^{N} \frac{y_i X_k(x_i)}{\sigma_i^2} \tag{1.20}$$

که $c_{ik} \equiv [\alpha_{ik}]^{-1}$ و انحراف معیار میانگین مربوط به هریک از پارامترها a_i، برابر است با:

$$\begin{aligned} \sigma^2(a_i) &= \sum_{j=1}^{N} \sigma_j^2 \left(\frac{\partial a_i}{\partial y_j}\right)^2 \\ &= \sum_{k=1}^{M}\sum_{l=1}^{M} c_{ik}c_{il} \left[\sum_{j=1}^{N} \frac{X_k(x_j)X_l(x_j)}{\sigma_j^2}\right] \end{aligned} \tag{1.21}$$

کمیت موجود در براکت، ماتریس معکوس $[\mathbf{c}]$ می‌باشد. بنابراین معادله (۱.۲۱) برابر است با:

$$\sigma(a_i)^2 = c_{ii} \tag{1.22}$$

به زبانی دیگر عناصر قطری $[\mathbf{c}]$ همان انحراف معیار میانگین پارامترها می‌باشند. و عناصر غیر قطری کئوواریانس a_i و a_j می‌باشد. در برخی موارد معادلات نرمال ((۱.۱۸) و (۱.۱۹)) خیلی به مقدار تکین نزدیک می‌شوند در این صورت به تکینکی‌ها برخورد کرده و محاسبه متوقف می‌شود. در این حالت محاسبه با روش "ب" یعنی تجزیه مقادیر تکین دنبال می‌شود که توصیف آن در مراجع [?، ?] آمده است.

۱.۱.۱ حدهای تطابق روی پارامترهای مدل

بردار ستونی $\mathbf{a}$ دارای انحراف معیاری است که توسط معادله (۱.۲۱) داده می‌شود. اکنون قصد داریم نگاه دیگری به مفهوم عدم قطعیت در این پارامترها داشته باشیم. در ضمن روش تعیین کمّی حدهای تطابق[۱۲]

[۱۲] Confidence Levels

http://facultymembers.sbu.ac.ir/movahed/

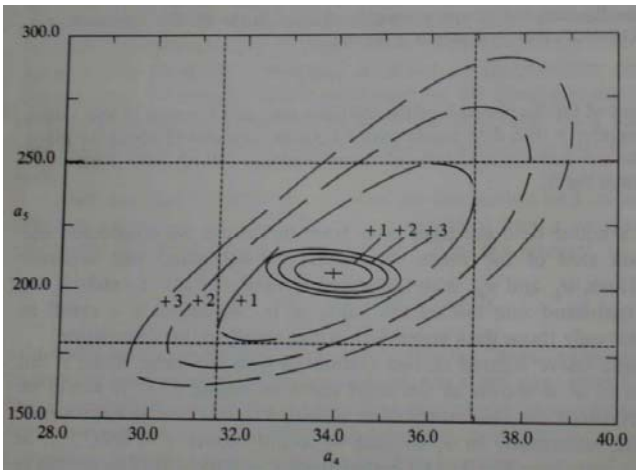
را بیان می‌کنیم. هنگامی که یک آزمایش‌گر دسته‌ای از داده‌ها را بدست می‌آورد بر اساس آنها پارامترهای مدل تعیین می‌شوند. از آنجا که این دسته داده‌ها تنها دسته اندازه‌گیری شده به شمار نمی‌آید، پارامترهای تعیین شده دارای عدم تعیین می‌باشند. اگر یک دسته اندازه‌گیری را با D_i مشخص کنیم دسته پارامترهای مرتبط با این اندازه‌گیری را $\mathbf{a_i}$ می‌نامیم. مقدار دقیق پارامترهای مدل یعنی $\mathbf{a_{true}}$ اساساً مخفی خواهد بود و با تعیین عدم تعیّن آنها میتوان فقط با احتمال ناحیه وجود پارامترها را تعیین کرد. اکنون این سئوال مطرح می‌شود که توزیع احتمال $\mathbf{a_i} - \mathbf{a_{true}}$ بدون اینکه مقدار دقیق $\mathbf{a_{true}}$ معلوم باشد چقدر است؟ توزیع احتمال در واقع تابعی در فضای M پارامتری می‌باشد. یک ناحیه تطابق در حقیقت یک ناحیه از فضای M بُعدی است که شامل درصد معینی از توزیع احتمال کلی می‌باشد. مثلاً وقتی‌ناحیه‌ای با ۶۸% تراز تطابق مشخص می‌کنیم یعنی با احتمال ۶۸% مقدار واقعی در این ناحیه قرار دارد. بنابراین در یک بُعد، یک بازه و در بُعدهای بالاتر با بیضی یا بیضی‌گون روبرو هستیم. قبل از اینکه به صورت تحلیلی به تعیین تراز تطابق بپردازیم به صورت کیفی به توضیح نواحی با $\chi^۲$ ثابت به عنوان محک تعیین نواحی تراز تطابق می‌پردازیم.

همانطور که قبلاً نیز مطرح شد، فرض کنید که به ازاء $\mathbf{a_0}$ کمیت $\chi^۲$ به مقدار کمینه یعنی $\chi^۲_{min}$ برسد. بنابراین هر میزان انحراف در پارامترها، مقدار $\chi^۲$ را از مقدار کمینه‌اش منحرف کرده و افزایش می‌دهد. دسته پارامترهای $\mathbf{a}$ که به ازاء تغییر معینی در مقدار $\chi^۲$ به صورت $\Delta\chi^۲ = \chi^۲ - \chi^۲_{min}$ بدست می‌آیند، ناحیه‌ای در فضای M بُعدی جاروب می‌کنند. بنابراین ارتباط تنگاتنگی بین میزان انحراف $\chi^۲$ از مقدار کمینه‌اش و ناحیه جاروب شده در فضای M بُعدی پارامترهای مدل وجود دارد.

در فضای تک پارامتری تراز تطابق خیلی سرراست قابل محاسبه است. در حقیقت با در نظر گرفتن تابع توزیع گوسی، احتمال‌های مختلف برای فضای تک پارامتری به صورت زیر خواهد بود:

$$\begin{aligned} ۶۸.۳\% &= \int_{\bar{x}-\sigma}^{\bar{x}+\sigma} \frac{e^{-\frac{x-\bar{x}}{۲\sigma^۲}}}{\sqrt{۲\pi\sigma^۲}} \\ ۹۵.۴\% &= \int_{\bar{x}-۲\sigma}^{\bar{x}+۲\sigma} \frac{e^{-\frac{x-\bar{x}}{۲\sigma^۲}}}{\sqrt{۲\pi\sigma^۲}} \\ ۹۹.۷۳\% &= \int_{\bar{x}-۳\sigma}^{\bar{x}+۳\sigma} \frac{e^{-\frac{x-\bar{x}}{۲\sigma^۲}}}{\sqrt{۲\pi\sigma^۲}} \end{aligned} \tag{۱.۲۳}$$

http://facultymembers.sbu.ac.ir/movahed/



شکل ۱ — ۲: حدود تطابق

هنگامی که دو یا چند پارامتر داشته باشیم و در ضمن بین این پارامترها نیز همبستگی‌هایی موجود باشد، در آن صورت نه تنها تعریف احتمال تراز تطابق به سادگی حالت‌های فوق نخواهد بود بلکه انتخاب نواحی جاروب شده در فضای پارامترها بایستی به دقت انجام گیرد. فرض کنید در یک آزمایش، مدل دارای ۵ پارامتر بوده که پارامترهای a_1°، a_2°، a_3°، a_4° و a_5° مقدار χ^2 را کمینه کنند. شکل (B—۲) ترازهای تطابق را در فضای (a_4, a_5) نشان می‌دهد. مرزهای کوچکتر که با خطوط ممتد نشان داده شده است، بدین صورت بدست آمده است که پارامترهای $a_1 = a_1^\circ$، $a_2 = a_2^\circ$ و $a_3 = a_3^\circ$ ثابت در نظر گرفته شده‌اند و فضای پارامترهای باقیمانده جاروب شده و نواحی $\Delta\chi^2 = 1, 2, 3$ مشخص شده است. کانتورهای بزرگتر که با خط چین بلند نشان داده شده است، بدین صورت بدست آمده‌اند که برای هر جفت (a_4, a_5) اجازه داده شده است که بقیه پارامترها تغییر کرده و مقدار χ^2 کمینه کند و دوباره نواحی $\Delta\chi^2 = 1, 2, 3$ مشخص شده است. حال دو سوال زیر مطرح است :

۱ — چه مقادیری برای $\Delta\chi^2$ با نواحی 1σ، 2σ و 3σ همخوانی دارد؟

۲ — کدام یک از رهیافت‌های فوق برای تعیین نواحی تراز تطابق انتخاب می‌شود؟

اجازه دهید در همین جا تکلیف مرزها با خطوط ممتد را روشن کنیم. در صورتی که هیچ همبستگی بین پارامترهای مدل وجود نداشته باشد، دو حالت بالا نتایج یکسان بدست می‌دهند. اما در حضور هر گونه همبستگی نمی‌توان انتظار داشت که مرزهای کوچکتر ناحیه مورد علاقه ما را نشان دهند زیرا به ازاء هر

http://facultymembers.sbu.ac.ir/movahed/

تغییر در ν پارامتر اول مقدار اولیه $M-\nu$ پارامتر باقیمانده نمی‌تواند χ^2 را به صورت موضعی کمینه کند. بنابراین نواحی با خط چین‌های بلند در هر وضعیت، روش صحیح برای یافتن نواحی مورد علاقه تلقی می‌شود. غالباً ما علاقه‌مند به تعیین نواحی تطابق در فضای یک، دو و حداکثر سه بُعدی به جای M بُعدی هستیم. که هرکدام مطابق با تصویر تابع توزیع احتمال M بُعدی در بُعدهای پایین‌تر می‌باشد. ناحیه تراز تطابق در زیر فضای ν بُعدی از فضای M بُعدی، تصویر ناحیه M بُعدی محدود شده توسط مقدار معین $\delta\chi^2$ در زیر فضای ν بُعدی است. اکنون اجازه دهید به صورت تحلیلی به ارتباط بین نواحی محدود شده توسط $\Delta\chi^2 = const$ و احتمال تراز تطابق بپردازیم. شاید این سوال مطرح شود که با توجه به وجود ارتباط تحلیلی، چرا ما می‌خواهیم از تغییر χ^2 از مقدار کمینه‌اش، به نواحی تراز تطابق پی‌ببریم. دلیل اصلی این است که کمینه کردن χ^2 روشی کاربردی برای تخمین‌پارامترهای مدل است حتی اگر خطاهای اندازه‌گیری نرمال نباشد. قضایای پایه‌ای زیر وقتی که خطاهای اندازه‌گیری به صورت نرمال توزیع شده باشند صادق هستند:

قضیه ۱ — اگر $\mathbf{a}$ دسته‌ای از پارامترها باشند که مقدار $\chi^2 > \chi^2_{min}$ را بدست دهد تابع توزیع احتمال $\delta\mathbf{a} \equiv \mathbf{a} - \mathbf{a_0}$ به صورت زیر می‌باشد:

$$P(\delta a)da_1...da_M \sim \exp\{-\frac{1}{2}\delta a.[\alpha].\delta a\} \qquad (۱.۲۴)$$

قضیه ۲ — تابع توزیع کمیت $\delta\chi^2 = \chi^2(\mathbf{a}) - \chi^2(\mathbf{a_0})$ مانند تابع توزیع χ^2 (معادله (۱.۷)) با M درجه آزادی می‌باشد.

فرض کنید در یک مسئله تعداد درجات آزادی برابر با $\nu = 1$ باشد. می‌خواهیم تراز تطابق را برای پارامتر آزاد a_1 تعیین کنیم. با توجه به قضیه ۲ داریم:

$$\begin{aligned}\chi^2 &= -2\ln(P(a_1, a_2, ..., a_M)) + 2\sum_{i=1}^{N}\ln(\sigma_i\sqrt{2\pi}) \\ &= \frac{(a_1 - a_1^\circ)^2}{\sigma(a_1)^2} + const \qquad (۱.۲۵)\end{aligned}$$

اگر به ازاء هر a_1، بقیه پارامترها تغییر کرده و χ^2 را کمینه کنند آنگاه می‌توان نوشت:

$$\chi^2 = \chi^2_{min} + \frac{\partial^2\chi^2}{\partial a_1^2}\frac{(a_1 - a_1^\circ)^2}{2}$$

http://facultymembers.sbu.ac.ir/movahed/

$$= \quad \chi^2_{min} + \frac{(a_1 - a_1^0)^2}{\sigma(a_1)^2} \tag{۱.۲۶}$$

اگر $a_1 = a_1^0 + n\sigma(a_1)$ باشد در آن صورت $\Delta\chi^2 = n^2$ بنابراین در فضای تک پارامتری تغییرات مقدار χ^2 از مقدار کمینه به صورت n^2 نواحی تراز تطابق برابر با $n\sigma$ را برای پارامتر a_1 مشخص می‌کند. در حالت کلی تابع توزیع دسته پارامترهای مدل برابر است با:

$$P(a_1, a_2, ..., a_M) = \frac{(2\pi)^{M/2}}{\left(det[\alpha]^{-1}\right)^{1/2}} \times \exp\{\frac{-1}{2}\sum_{i,j}^{M} \delta a_i.\alpha_{ij}.\delta a_j\} \tag{۱.۲۷}$$

و

$$\Delta\chi^2 = \sum_{ij}^{M} \delta a_i.\alpha_{ij}.\delta a_j \tag{۱.۲۸}$$

اکنون سوال این است که چگونه می‌توان $\Delta\chi^2(P)$ را تعیین کرد که P احتمال اینکه در فضای M بُعدی ν پارامتر اول در بیضی‌گون تراز تطابق قرار می گیرد، مشخص می‌کند؟ این احتمال مربوط به حجم بیضی‌گون ν بُعدی است یعنی:

$$\int \frac{(2\pi)^{\nu/2}}{\left(det[\alpha]^{-1}\right)^{1/2}} \times \exp\{-\frac{1}{2}\sum_{i,j}^{\nu} \delta a_i.\alpha_{ij}.\delta a_j\} d\delta a_1 ... d\delta a_\nu = P(\frac{\nu}{2}, \frac{\Delta\chi^2}{2}) \tag{۱.۲۹}$$

که سمت راست معادله (۱.۲۹) به صورت زیر است:

$$P(\frac{\nu}{2}, \frac{\Delta\chi^2}{2}) = \frac{1}{\Gamma(\nu/2)} \int_0^{\Delta\chi^2/2} t^{\nu/2-1} e^{-t} dt \tag{۱.۳۰}$$

بنابراین در فضای ν بُعدی برای بدست آوردن مرزهای تراز تطابق متناظر با احتمال P می‌توان با استفاده از معادله (۱.۳۰) مقدار $\Delta\chi^2$ را بدست آورد. در جدول (B—۱) مقادیر مربوط به $\Delta\chi^2$ برای تعداد درجات آزادی مختلف داده شده است [?]. به طور خلاصه می‌توان گفت هنگامی که بخواهیم نواحی تراز تطابق را در فضای ν بُعدی مشخص کنیم ابتدا دسته پارامترهایی که کمینه مطلق χ^2_{min} را یا بیشینه مطلق تابع درست نمایی را بدست می‌دهند، تعیین کرده و سپس ν تا از M پارامترهایی که می‌خواهیم منحنی تراز تطابق را برای آنها تعیین کنیم ثابت نگه داشته و بر روی فضای بقیه پارامترها حرکت کرده و کمینه موضعی χ^2_{loc} را بدست می آوریم. از طرف دیگر برای احتمال‌های مثلاً ۶۸.۳%، ۹۵.۴% و غیره

http://facultymembers.sbu.ac.ir/movahed/

جدول ۱—۱: میزان تغییر در کمیت کمترین مربعات برای ایجاد نواحی احتمال تراز تطابق

M	۱	۲	۳	۴	۵	۶
$p\%$	$\Delta\chi^۲$	$\Delta\chi^۲$	$\Delta\chi^۲$	$\Delta\chi^۲$	$\Delta\chi^۲$	$\Delta\chi^۲$
۶۸.۳%	۱.۰۰	۲.۳۰	۳.۵۳	۴.۷۲	۵.۸۹	۷.۰۴
۹۰%	۲.۷۱	۴.۶۱	۶.۲۵	۷.۷۸	۹.۲۴	۱۰.۶
۹۵.۴%	۴.۰۰	۶.۱۷	۸.۰۲	۹.۷۰	۱۱.۳۰	۱۲.۸۰
۹۹%	۶.۶۳	۹.۲۱	۱۱.۳۰	۱۳.۳۰	۱۵.۱۰	۱۶.۸۰
۹۹.۷۳%	۹.۰۰	۱۱.۸۰	۱۴.۲۰	۱۶.۳۰	۱۸.۲۰	۲۰.۱۰
۹۹.۹۹%	۱۵.۱۰	۱۸.۴۰	۲۱.۱۰	۲۳.۵۰	۲۵.۷۰	۲۷.۸۰

مقدار $\Delta\chi^۲ = \chi^۲_{loc} - \chi^۲_{min}$ را بدست آورده و مرزهای تراز تطابق را مشخص می‌کنیم. شکل (B—۳) به صورت طرح‌وار این منحنی‌ها را در فضای دو بُعدی نشان می‌دهد.

در قسمت قبل دیدیم که روش $\chi^۲$ ابزار قدرتمندی برای تعیین پارامترهای بهینه مدل به حساب می‌آید. از آنجا که اساس این روش بر بیشینه کردن احتمال می‌باشد بهتر است به استفاده مستقیم از تخمینگر بیشینه احتمال (MLE) اشاره‌ای داشته باشیم. استفاده مستقیم از MLE در حقیقت به دو دلیل زیر بر استفاده از $\chi^۲$ ارجحیّت دارد:

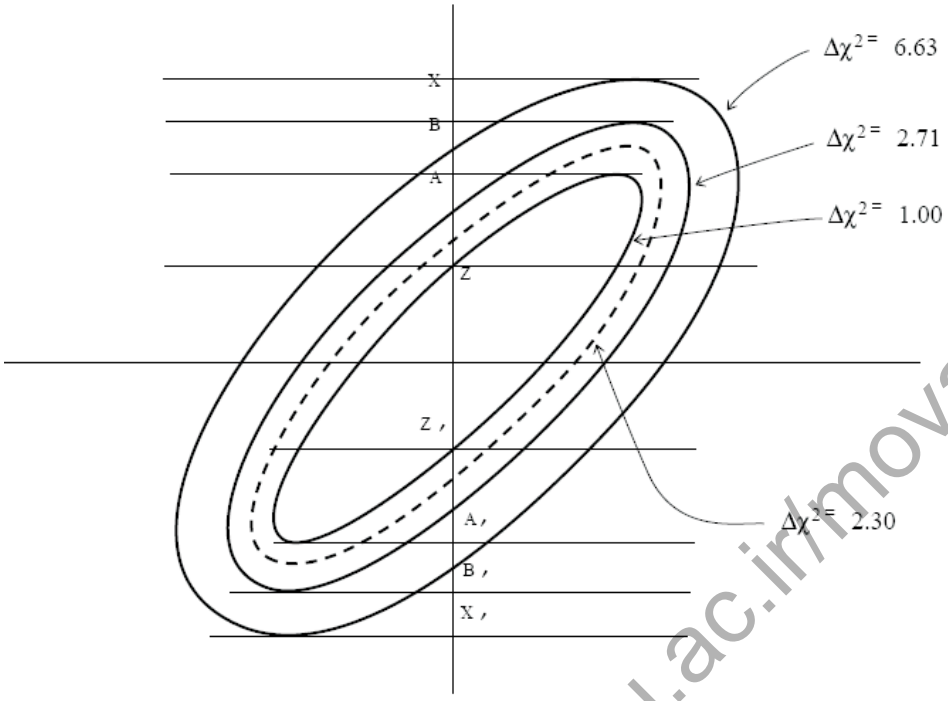
۱— کم بودن داده‌های بدست آمده از آزمایش باعث می‌شود، شرط گوسی بودن تابع توزیع حول مقدار متوسط برآورده نشود.

۲— در بیشتر مواقع به ازاء هر مقدار اندازه‌گیری شده شکل تابع توزیع لزوماً یکسان نمی‌باشد. بنابراین مستقیماً از MLE استفاده می‌کنیم:

$$L(a_۱, ..., a_M) = \prod_{i=۱}^{N} P_i \qquad (۱.۳۱)$$

و با بیشینه کردن آن پارامترهای مدل بدست خواهند آمد.

http://facultymembers.sbu.ac.ir/movahed/



شکل ۱ــ۳: طرح وار منحنی تراز تطابق

۲.۱.۱ روش تخمینگر بیشینه احتمال انتگرال‌گیری شده و حذف پارامترهای اضافی

در علوم تجربه‌ای اطلاعات غالباً از برازش نتایج حاصل از مشاهده و یک مدل تئوری که ممکن است در حالت کلی نسبتاً پیچیده باشد استخراج می‌شود. مهمترین شیوه برای تخمین این پارامترها استفاده از تخمین‌گر بیشینه احتمال است. در مواقعی که شرط لازم برای استفاده از تابع توزیع گوسی برای هر اندازه‌گیری فراهم شود این روش به استفاده مستقیم از محک χ^2 می‌انجامد و یا وقتی که عدم قطعیت اندازه‌گیری‌ها یکسان باشد به روش کمترین مربعات می‌رسیم.

پارامترهای مدل نقش عمده و مهمی در فهم ما از فرآیندهای فیزیکی ایفا می‌کنند. ولی آنچه که غالباً اتفاق می‌افتد این است که تمام پارامترهای آزاد نیازمند به تعیین دقیق نمی‌باشند، که اصطلاحاً به آنها پارامترهای اضافی[۱۳] گفته می‌شود معمولاً با انتگرال‌گیری روی آنها، اثرشان حذف می‌شود[۱۴] . وجود عدم قطعیت در پارامترهای مدل باعث عدم تعیّن در مدل ارائه شده می‌شود. بنابراین یافتن روش بهینه برای

[۱۳] Nuicence Parameter

[۱۴] Marginalized Likelihood Estimator

http://facultymembers.sbu.ac.ir/movahed/

تخمین پارامترهای مورد علاقه هنگامی که پارامترهای غیر ضروری وجود دارند جایگاه ویژه‌ای دارد. فرض کنید تابع تخمینگر بیشینه در حالت کلی به صورت زیر باشد:

$$L(a, \lambda|(x, y)) \tag{۱.۳۲}$$

که $(\mathbf{x}, \mathbf{y})$ مقادیری که مستقیماً از اندازه‌گیری بدست آمده می‌باشد. و $\mathbf{a}$ و λ به ترتیب پارامترهای مورد علاقه و اضافی مدل می‌باشند. هدف در اینجا یافتن مقادیر بهینه $\mathbf{a}$ می‌باشد بدون اینکه مقدار دقیق λ را داشته باشیم. روش معمول در استخراج مقدار درست $\mathbf{a}$ این است که عدم تعیین در پارامترهای غیر ضروری را در نظر نگیریم و کافی است مقدار معینی برای آنها با توجه به اطلاعاتمان از روش‌های دیگر اتخاذ کرده و در تابع تخمینگر جاگذاری کنیم:

$$L(a|(x, y)) = \prod_i P_i(a, \lambda = \bar{\lambda}|(x, y)) \tag{۱.۳۳}$$

واضحاً مشخص است که اگر پارامترهای غیر ضروری خود دارای عدم قطعیت باشند در آن صورت پارامترهای تخمین زده شده غیر دقیق خواهد بود[? ،? ،? ،? ،?]. برای بهینه کردن تخمین پارامترهای ضروری در حضور پارامترهای غیر ضروری از رهیافت MRL استفاده می‌شود. این روش در ادامه کارهای Butler 1988 پیشنهاد شده است و توسط بقیه تکمیل شده است [? ،? ،? ،? ،? ،?]. در این روش اثر پارامترهای غیر ضروری با انتگرال‌گیری روی آنها از بین می‌رود یعنی:

$$L(a|(x, y)) = \int L(a, \lambda|(x, y))d\lambda \tag{۱.۳۴}$$

۳.۱.۱ روش تخمینگر بیشینه احتمال انتگرال‌گیری شده وزن‌دار

استخراج دسته پارامترهای ضروری مدل با استفاده از داده‌های حاصل ازاندازه‌گیری $(\mathbf{x}, \mathbf{y})$ لزوماً بازه محدودی برای پارامترهای مدل به دست نمی‌دهد. این بدان معناست که اطلاعات اولیه بیشتری برای تخمین زدن نیاز داریم. در این وضعیت یک تابع توزیع اولیه[۱۵] برای پارامترهای دلخواه در صورت وجود

[۱۵] prior

http://facultymembers.shu.ac.ir/nowaned/

در نظر گرفته می‌شود.

$$L^B(a|(x,y)) = L(a|(x,y))\Pi(a) \tag{۱.۳۵}$$

و یا در حالت کلی

$$L^B_M(a|(x,y)) = \int L(a,\lambda|(x,y))\Pi(a)\Pi(\lambda)d\lambda \tag{۱.۳۶}$$

در معادله (۱.۳۶) کمیت‌های $\Pi(\lambda)$ و $\Pi(\mathbf{a})$ به ترتیب تابع احتمال اولیه پارامترهای غیر ضروری و ضروری می‌باشند. اکنون از طرفی مطمئن خواهیم بود که اثر پارامترهای اضافی با انتگرال گیری ازبین رفته است و از طرف دیگر با حضور تابع احتمال اولیه که از اطلاعات اولیه بدست آمده است بازه محدودتری را برای پارامترهای اولیه بدست می آوریم.

http://facultymembers.sbu.ac.ir/novahed/