« بسم الله الرحمن الرحیم »

No: جلسه‌ی ۳

Date: ۱۴۰۱/۱۱/۲۸

دفاع احتمالاتی

« ترمودینامیک چی؟ »

(خلاصه‌ی مبحث قبل)

✓ ما می‌دونیم که اصل کار ترمودینامیک بررسی احتمال است.

چه طور به یک رویداد در فیزیک یک احتمال نسبت بدیم؟

$$\begin{cases} p(A) \geq 0 \\ \sum p = 1 \end{cases}$$

① راه حل تکرار ② راه و روش بیز

مثال = شتاب‌دهنده‌ی LHC ← دوتا پروتون رو می‌کوبیم به هم ببینیم چیزهای تولید میشه. کلی انرژی لازم داره. انرژی الکتریکی نیاز داره. نمیشه مدام این کار رو تکرار کرد.

روش منطقی تر روش بیز: بر پایه‌ی اینه، از قبل که بریم آزمایش، برای آن بریم. بریم ببینیم شرایط آزمایش چی بوده، مثلاً قبلاً شکست خوردیم ببینیم چی بوده، از چه جنسی و ....

اطلاعات رو حدس میزنیم و با داده‌ها مقایسه می‌کنیم و تعمیم میدیم. prior → posterior

نیاز داریم برای احتمال، احتمال شرطی

$$p(A,B) = p(A|B)\,p(B) = p(B|A)\,p(A)$$ → قضیه‌ی احتمال معکوس

$$p(A|B) = \frac{p(A|B)\,p(B)}{p(A)}$$ → حالا ببینیم چه طور ازش استفاده میشه

فرض بشه یک تفسیر یا مدل داشته باشیم. M ← تفسیر

و یک مجموعه داده داشته باشیم. D ← داده

مثال = در آزمایشگاه میخوایم روی اتم H کار کنیم. مدل خودمون مدل بور هست. داده‌هامون داده‌های رادرفورد است.

پس یک سری داده داریم که مدل نوشتیم و میخوایم ربط بدیم که مدل به داده‌ها می‌خوره یا نه.

رویداد نوشتن مدل B ← M

رویداد داده A ← D

$$p(M|D) = \frac{p(D|M)\,p(M)}{p(D)}$$

$P(M|D)$ → احتمال درست بودن مدل با شرط درست بودن داده posterior

$P(D)$ → احتمال درست بودن داده

$P(M)$ → احتمال درست بودن مدل - بدون اینکه به داده‌ها نگاه کنیم به مدل (درستیابی) prior

$P(D|M)$ → احتمال درست بودن داده با شرط درست بودن مدل likelihood

ما یک مدلی نوشتیم یک سری داده هم به دست آوردیم. خب حالا از کجا معلوم که این داده‌ها برای همین مدل هست؟ پس فرض می‌کنیم که مثلاً یک خط طیفی از یک اتم دیدیم. بعد بریم ببینیم که هر تئوری می‌تونیم اون خط رو پیش‌بینی کنیم یا نه. که این میشه ← posterior

اون چیزی که ما در فیزیک می‌خوایم $P(M|D)$ است. و $P(D|M)$ رو می‌نویسیم. $P(M)$ هم می‌نویسیم ضرب می‌کنیم ولی $P(D)$ هم که نداریم و نمیشه به دست آورد.

خب پس $P(D)$ رو چطور به دست بیاریم؟ ما میدونیم جمع همه احتمالات ۱ است. $\Sigma P = 1$

ما جمع می‌کنیم و جواب ۱ نیست - یک ضریبی داریم که همون $P(D)$ ماست. بخش مهم قضیه بیز.

به صورت کلی هرچه هستی از فیزیک کار فیزیکدانان همین است: ما مدلمون بهترین پارامترهارو به حساب میاریم.

مثال می‌خوایم تابعی از جرم الکترون به دست بیاریم. برای درس زدن prior هم میدونیم جرم الکترون از 1eV بیشتر نیست. پس میریم و هر تئوری می‌بینیم چه شرط‌هایی برای پارامترمون هست.

پس میریم مدلی رو به دست میاریم که با بیشترین احتمال درست باشه. این میشه بهترین مدل.

پس پارامتری رو پیدا می‌کنیم که posterior رو ماکسیمم کنه. یعنی درست‌ترین باشه $P(M|D)$

No ... ۲ ... جلسه‌ی

Date ۱۴۰۱/۱۱/۲۵

(مفاهیم مقدماتی)

پس اون چه که ما می‌خوایم یاد بگیریم $p(D|M)$ یک احتمال که توش پارامتری داریم.

برای یک رویداد می‌خوایم یک احتمال بنویسیم که بهش می‌گیم تابع توزیع احتمال. ← $p(D|M)$

احتمال‌ها، متغیرها دو نوع هستند یا پیوسته هستند یا گسسته هستند.

احتمال هایی که برای متغیرهای پیوسته می‌نویسیم بهش می‌گویند توزیع احتمال ← PDF و $p(x)$.

احتمال هایی که برای متغیرهای گسسته می‌نویسیم بهش می‌گویند تابع جرم احتمال ← PMF و $p(x)$.

برای همین ما دنبال احتمال رخ دادن یک رویداد هستیم. یعنی $p(D|M)$ برای یک مسئله‌ی فیزیکی می‌خوایم تابع توزیع احتمالش رو بنویسیم.

برای توزیع گسسته، احتمال گسسته بیشتر فرمول داریم. $S=\{x_1, x_2, \dots, x_n\}$

برای پیوسته این که یک مجموعه‌ی پیوسته است ـ از صفر تا بی‌نهایت. $S=[0, \infty)$

مثال = اندازه‌گیری قد آدم‌ها کرده باشیم x ـ ولی چون خطا داره میشه $x \pm dx$. پس بهتره به جای عدد در باره‌ی بازه حرف بزنیم. بازه‌ی $x \pm dx$.

پیوسته → $p(x)\,dx$ → $(x-dx), x+dx)$

چون پیوسته است جمع آنها میشه انتگرال → $\underset{\text{کل بازه}}{\int} p(x)\,d(x) = 1$ روی کل بازه

$\int p(x)\,d(x) = 1$ ← بهنجار است.

اگر بهنجار نباشه یعنی انتگرالش = ۰ نیست، و عدد دیگری است که متناهی است. این جور مواقع ما به این تابع احتمال نفس، و با یک $p^*(x)$ نشون میدن.

$A = \int p^*(x)\,d(x)$

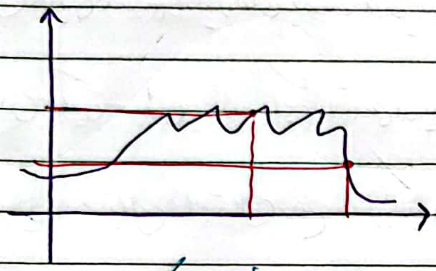
اون وقت ما میتونیم احتمال بهنجار شده رو تعریف کنیم.

$$p(x) = \frac{p^*(x)}{\int p^*(x)\,d(x)} = \frac{p^*(x)}{A} \rightarrow$$

بهنجار شده ⇐ $\int p(x)\,d(x) = 1$

اگر کنجکاو نبودیم انتگرال می‌گیریم ببینیم چنده و تقسیم بر احتمال می‌کنیم.

ولی اگر نامتناهی باشه جوابش میشه صفر پس بهتر هست که متناهی باشه. حالا فرض کنیم که تابع احتمال رو بدست آوردیم. انتگرالش ۱ است و بر ده‌ها است. سوال اینه = بعد از اینکه ما تابع توزیع احتمال رو نوشتیم ازش چه استفاده‌ای کنیم؟! بر می‌گردیم به تعریف تابع توزیع احتمال $p(x)$ اینکه آزمایش رو انجام بدیم و عدد رو بدست بیاریم که به نوعی رفتیم محور x عددی.



شکل تابع توزیع چاقه یا چه احتمالی چه عددی بدست میاد.

مثال = اگر ۱۰۰ بار آزمایش انجام بدیم. ٪۶۰ دفعاتش عدد ۵ بدست بیاریم، ۴۰ دفعاتش ... انتگرال ۱ است.

پس از تابع توزیع احتمال می‌خوایم. دنبال چه اعدادی در آزمایش بگردیم. یعنی میتونیم در چه رنجی دنبال عدد بگردیم.

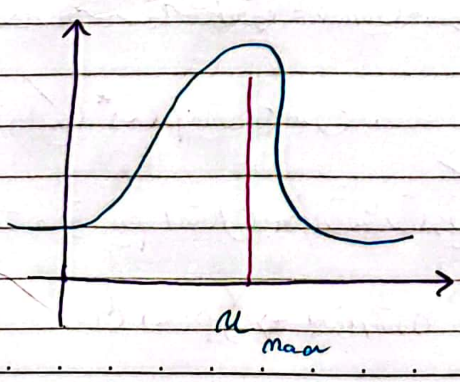
اینه دقیقا تو آزمایش ما سوم عدد رو پیدا کنیم یک مسئله‌ی دیگه است. میایم یه سری تابع آماری معرفی می‌کنیم.

① میانگین $= E[x] = \int x\, p(x)\, dx$

تعریف میانگین = هر عددی با چه احتمالی اتفاق می‌افته و همه رو با هم جمع می‌کنیم. یعنی هر عدد رو در احتمال اتفاق افتادنش ضرب می‌کنیم میشه میانگین تابع توزیع ما.

یه دامنه‌ی دیگه هم داریم به نام x_{max} یعنی جایی که در آن $p(x)$ بزرگترین و بیشترین است. یعنی نقطه‌ای که احتمالش از همه بیشتر است.

مثال =



No ۲ جلسه

Date ۱۴۰۱/۱۱/۲۶ رفتارهای مغاری

✓ بوشودکر اولی قله‌های زیادی بود و با پیدا کردن x_{max} نمی‌شد همه‌ی داده‌ها و اطلاعات رو به دست آورد. برای چنین نمودارهایی از میانگین استفاده کنیم بهتر است.

۲ واریانس ← انحراف از معیار میانگین

تمرین: می‌خواهیم ببینیم از عددی که گزارش شده چه قدر دور شده. (هر عددی رو - میانگینش کنیم و همه رو با هم جمع کنیم.

$$\int (x-E[x])\,p(x)\,dx$$

فرض است اختلاف از میانگین را به دست بیاوریم، ولی مشکل این است که با این عبارت جواب صفر میشه.

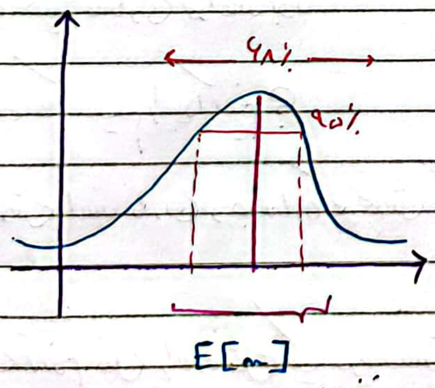
$$= \underbrace{\int x\,p(x)\,dx}_{E[x]} - \underbrace{\int E[x]\,p(x)\,dx}_{1} = \boxed{E[x] - E[x] = 0}$$

پس اینطوری به ما جواب نمیده. میایم اینو مثبت می‌کنیم. مثلاً | | قدر مطلق می‌گیریم یا به توان ۲ می‌رسونیم. که به توان ۲ می‌رسونیم جذر می‌گیریم.

$|x-E[x]| \to$ قدر مطلق

$$\to \sqrt{\int (x-E[x])^2\,p(x)\,dx}$$

مثال:



$<$ عدد درست $< E[x]+\sigma$

$E[x]-\sigma$

$\sigma \leftarrow$ سیگما

هر چه قدر آزمایش دقیق‌تر باشه، عددها به عدد واقعی نزدیک‌تر است. به σ نشون می‌دیم.

1σ ۶۸٪ داده‌ها

تقسیم داده‌هاش عددش σ است ←

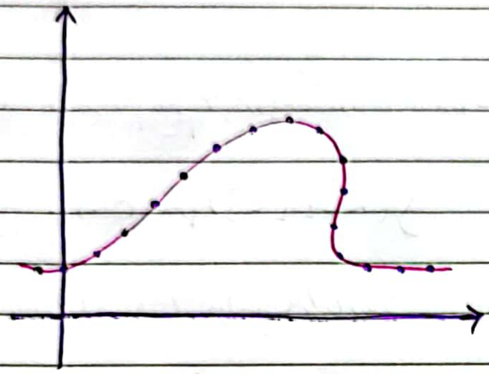
2σ ۹۵٪ داده‌ها

3σ ۹۹.۷٪

وقتی یک کمیتی با دقت 2σ اندازه‌گیری شده، یعنی ۹۵٪ احتمالش همین بوده که ما به دست آوردیم برای این بازه.

مشکل اصلی این هست که ما از آنالیتیکی نمی‌تونیم به کفایت کدر را اندازه بگیریم. برای همین به χ^2 نیاز هست.

ما باید بتونیم این تابع توزیع رو با یک تعداد تخمین بزنیم.



به جای تابع پیوسته با این نقاط کار کنیم.

اگر نقاط خیلی خوب انتخاب شده باشن ما میتونیم از اون نقاط استفاده کنیم به جای تابع توزیع.

مثال: برای میانگین به جای اینکه انتگرال بگیریم ← $E[x] = \int p(x)\, dx$

می‌نویسیم → $\approx \sum_{i=1}^{n} x_i\, p(x_i)$

برای اینکه بتونیم از تابع توزیع استفاده کنیم باید تخمینش بزنیم با یک سری کد.

آیا $E[x]$ با نتیجه‌ی بالا مساوی است؟

خیر در حالت کلی مساوی نیستند. ما اگر n ← به سمت ∞ میل کنند مساوی اند. ولی زمانی که $n \to \infty$ نره، خیر مساوی نخواهند بود.

چجوری ما لیستیم رو پیدا کنیم؟

یک روش استانداردی وجود داره که کمک می‌کنه: MCMC.

Markov chain monte carlo

↑ یک حرکت رندوم هست.

حرکت رندوم مثل راه رفتن یک فرد مست هست. یک فردی که نمیدونه کجاس و چیکاره.

گام‌های خطیم را انتخاب کنیم.

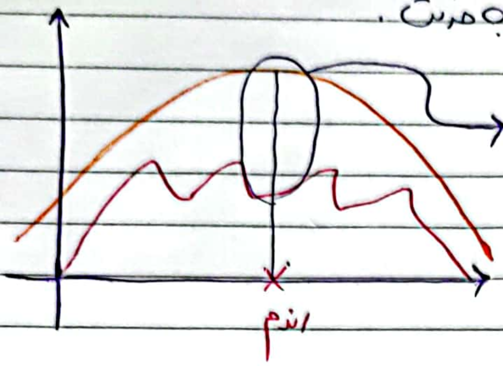
فرض شده گام‌های خطیم برای یک تابع توزیع چند تا کد تخمین بزنیم.

No ...... جلسه ی ۲ ......

Date ۱۴۰۰/۱۱/۲۵

رضا طلع مقاری(۱)

یک تابع گوسی می کشیم. بعد رندوم انتخاب می کنیم. و شروع می کنیم به حرکت.



اگه این بیشتر بود یعنی باید بریم پایین تر اگه کمتر بود

یعنی همون جا بایستیم و چک کنیم ببینیم اینجا هستیم

ماکسیمم هست یا نه؟