Statistical Physics-6

درس ششم: مقدمه‌ای بر نظریه‌ی احتمالات

Note Title 3/12/2009

- ☐ فضای احتمالات
- ☐ متغیر تصادفی
  - ☐ تابع توزیع ☐ گشتاورها و ممان‌ها ☐ تابع مولد
  - ☐ قضیه‌ی اول Chebyshev ☐ قضیه‌ی دوم Chebyshev
- ☐ چند تابع توزیع مهم
  - ☐ تابع توزیع دوجمله‌ای و چندجمله‌ای ☐ تابع توزیع پواسون ☐ تابع توزیع گاوسی
- ☐ تابع توزیع چند متغیر تصادفی
  - ☐ تابع توزیع چندجمله‌ای ☐ تابع توزیع گاوسی
- ☐ قضیه‌ی حد مرکزی
- ☐ چند رابطه‌ی مفید برای ابزار تقریب
- ☐ آنتروپی، اطلاعات و بی‌نظمی

متغیر تصادفی

آزمایشی را در نظر بگیرید که نتیجه‌ی آن یک عدد X باشد که در هر بار انجام آزمایش یکی از مقادیر متعلق به مجموعه‌ی

$$S_X = \{x_1, x_2, \dots x_n\} \subset R$$

را به طور تصادفی اختیار کند. در این صورت X را یک متغیر تصادفی Random variable می‌گوییم.

دلیل تصادفی بودن مقدار X آن است که ما دانش کامل و دقیقی راجع به شرایط اولیه‌ی آزمایش نداریم. ساده‌ترین مثال

از دید آزمایش و متغیر تصادفی، انداختن سکه است. هر بار که آزمایش را تکرار می‌کنیم، یک روی سکه یعنی شیر یا خط روی زمین می‌نشیند و ما به تناسب آن می‌توانیم بگوییم متغیر تصادفی X مقدار 0 یا 1 را اختیار کرده است. مثال دیگر انداختن تاس است که در این صورت $S_X = \{1, 2, 3, 4, 5, 6\}$.

- اگرچه مقدار دقیق X را با یک بار آزمایش نمی‌توانیم تعیین کنیم، اما می‌توانیم احتمال اینکه متغیر تصادفی X مقدار x_k را اختیار کند تعریف کنیم. این احتمال را با $P_k = P(X=x_k)$ نشان می‌دهیم. شرایط روی P_kها این‌ها هستند:

$$a) \quad 0 \leq P_k \leq 1$$

$$b) \quad \sum_k P_k = 1. \qquad (1)$$

معنای P_k یا احتمال به دست آمدن نتیجه x_k این است که اگر N بار آزمایش را انجام دهیم و در آن N_k بار نتیجه x_k به دست آید، آنگاه:

$$P_k = \lim_{N\to\infty} \frac{N_k}{N}. \qquad (2)$$

رابطه (2) هم معنای احتمال را روشن می‌کند و هم روشی تجربی را به دست می‌دهد که مطابق با آن می‌توانیم احتمالات مربوط به متغیر تصادفی را در یک آزمایش تعیین کنیم. از آنجا که هیچ وقت نمی‌توانیم بی‌نهایت بار یک آزمایش را تکرار کنیم، همواره می‌بایست این رابطه را به این صورت فهمید که در حد Nهای بزرگ با دقت خوبی برقرار است و هر چه که N بزرگتر باشد این تقریب بهتر است.

متغیر تصادفی X می‌تواند مقادیر خود را از یک مجموعه گسسته یا از یک مجموعه پیوسته $[a, b] \subset R$ اختیار کند. در این صورت رابطه (1) به شکل زیر خواهد بود:

$$a) \quad 0 \leq P(x)$$

$$b) \quad \int_a^b p(x)\,dx = 1. \qquad (3)$$

دقت کنید که شرط $p(x) \le 1$ اینجا دیگر لزوماً برقرار نیست. دلیل آن هم این است که با توجه به رابطه ② چیزی که برای متغیرهای پیوسته به شکل زیر در می‌آید:

$$p(x)\,dx := \lim_{\substack{N \to \infty \\ dx \to 0}} \frac{N(x)\,dx}{N} \qquad ④$$

که در آن $N(x)dx$ تعداد دفعاتی از آزمایش است که متغیر تصادفی مقداری بین x و $x+dx$ اختیار کند.

واضح است که $0 \le \frac{N(x)dx}{N} \le 1$ ولی الزامی ندارد که $\frac{N(x)}{N} \le 1$ باشد.

تابع $P(x)$ چگالی احتمال probability density خوانده می‌شود.

بدون نقض کلیت مجموعه S_x را برای یک متغیر تصادفی پیوسته به جای $[a,b]$، از بازه $(-\infty, \infty)$ در نظر گرفت.

مثال ۱)

یک هسته رادیواکتیو را در نظر بگیرید. این هسته اولین ذره α را در زمان T از خود ساطع می‌کند. در اینجا T یک متغیر تصادفی است که مقادیرش را در بازه $[0,\infty)$ اختیار می‌کند.

$p(t)\,dt$ = احتمال اینکه اولین ذره در زمان بین t و $t+dt$ ساطع شود.

مثال ۲) همین هسته رادیواکتیو را در نظر بگیرید که به مدت یک ساعت مشاهده می‌شود. تعداد ذرات αای که از خود ساطع می‌کند یک متغیر تصادفی K است که مقادیرش را از مجموعه گسسته $S_k = [0,1,2,3,\dots \infty]$ اختیار می‌کند.

• تابع توزیع احتمال probability distribution function

تابع توزیع احتمال $F(x)$ به صورت زیر تعریف می‌شود:

$$F(x) = \int_{-\infty}^{x} p(x')\,dx'$$

واضح است که $p(x) := \frac{dF}{dx}$. همچنین $F(x)$ لزوماً پیوسته است:

$$F(x) = \mathrm{Prob}(X \le x).$$

- هرگاه $f(x)$ یک تابع دلخواه از متغیر تصادفی x باشد، متوسط $f(x)$ به شکل زیر تعریف می‌شود:

$$\langle f(X)\rangle := \sum_k f(x_k)\, p(x_k).$$ برای متغیر گسسته

$$\langle f(X)\rangle := \int f(x)\, p(x)\, dx.$$ برای متغیر پیوسته

لذا از این به بعد همه‌ی رابطه‌ها را برحسب تابع توزیع پیوسته می‌نویسیم. خواننده خودش می‌تواند معادله‌ی مناسب را برای متغیرهای گسسته بنویسد. به علاوه با توجه به این که تابع توزیع برای متغیرهای گسسته را می‌توان برحسب تابع توزیع پیوسته نیز نوشت، به صورت زیر:

$$p(x) = \sum_k p(x_k)\,\delta(x - x_k).$$

متوسط‌های زیر خصوصاً مهم هستند:

$$\langle 1\rangle = \int p(x)\,dx = 1.$$

$$\langle X\rangle = \int x\, p(x)\, dx$$ متوسط

$$\langle X^2\rangle = \int x^2 p(x)\, dx$$ ممان درجه‌ی دوم

$$\langle X^k\rangle = \int x^k p(x)\, dx$$ ممان درجه‌ی k ام.

واضح است که اگر تمامی ممان‌ها را داشته باشیم می‌توانیم متوسط هر تابع دلخواهی که بسط تیلور دارد را حساب کنیم.

هم چنین کمیت زیر موسوم به واریانس بسیار مهم است زیرا میزان پراکندگی متغیر تصادفی را حول مقدار متوسط نشان می‌دهد:

$$\sigma_x := \sqrt{\langle X^2\rangle - \langle X\rangle^2}$$ (5)

- تمرین: نشان دهید که $\sigma_x = \sqrt{\langle (X - \langle X\rangle)^2\rangle}$.

- تابع مولد Generating function

تابع مولد، تابعی است که با استفاده از آن می‌توان همه‌ی ممان‌ها را به راحتی حساب کرد. این تابع به شکل زیر تعریف می‌شود:

$$\tilde{p}(k) := \int e^{ikx} p(x)\, dx \qquad (۶)$$

میتوان آن را به شکل ساده‌تر زیر نوشت، $\tilde{p}(k) := \langle e^{ikX} \rangle$. ←

$$\tilde{p}(k) = \left\langle \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} X^n \right\rangle = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle X^n \rangle .$$

رابطه با گشتاور مرتبه

$$\frac{1}{i^n} \frac{d^n}{dk^n} \tilde{p}(k) \Big|_{k=0} = \int x^n p(x)\, dx = \langle X^n \rangle .$$

• توضیح: برای متغیر گسسته و محدود حول مرکز که در واقع انتگرال را به سری تبدیل میکند میتوان تابع مولد را به شکل زیر تعریف کرد:

$$\tilde{p}(k) := \langle e^{kX} \rangle \longrightarrow \langle X^n \rangle := \frac{d^n}{dk^n} \tilde{p}(k) \Big|_{k=0} . \qquad (۷)$$

رابطه تابع چگالی احتمال و تابع مولد رابطه یک جفت تبدیل فوریه است: یعنی

$$p(x) = \frac{1}{2\pi} \int e^{-ikx} \tilde{p}(k)\, dk . \qquad (۸)$$

• قضیه اول Chebyshev.

هرگاه X یک متغیر تصادفی با مقادیر غیرمنفی و متناهی متوسط $\langle X \rangle$ باشد، آنگاه برای هر عدد حقیقی مثبت α نامساوی زیر برقرار است:

$$P(X \geq \alpha) \leq \frac{\langle X \rangle}{\alpha} \qquad (۹)$$

• اثبات:

$$\text{دلیل:} \quad P(X \geq \alpha) = \int_{\alpha}^{\infty} p(x)\, dx \leq \int_{\alpha}^{\infty} \frac{x}{\alpha} p(x)\, dx = \frac{1}{\alpha} \int_{\alpha}^{\infty} x\, p(x)\, dx$$

$$\le \frac{1}{\alpha}\int_0^\infty x\,p(x)\,dx = \frac{1}{\alpha}\langle x\rangle .$$

• قضیه دوم Chebyshev.

هرگاه X یک متغیر تصادفی دلخواه با مقدار متوسط $\langle x\rangle$ و واریانس σ_x باشد، آنگاه

(۱۱) $$P\left(|X-\langle x\rangle| \ge k\,\sigma_x\right) \le \frac{1}{k^2}.$$

متغیر تصادفی Y را به شکل زیر تعریف می‌کنیم.

$$Y := (X-\langle x\rangle)^2$$

واضح است Y یک متغیر مثبت است. همچنین می‌دانیم $\langle Y\rangle = \sigma_x^2$.

بنابر قضیه اول Chebyshev داریم:

$$P(Y \ge \alpha) \le \frac{\langle Y\rangle}{\alpha}$$

قرار می‌دهیم $\alpha = k^2\sigma_x^2$. $\leftarrow$

$$P\left((X-\langle x\rangle)^2 \ge k^2\sigma_x^2\right) \le \frac{1}{k^2}$$

و یا

(۱۲) $$P\left(|X-\langle x\rangle| \ge k\,\sigma_x\right) \le \frac{1}{k^2}.$$

این قضیه به خوبی معنای واریانس یا پراکندگی را نشان می‌دهد. به عنوان مثال می‌گوییم که احتمال اینکه یک متغیر تصادفی از مقدار متوسطش بیش از دو واریانس فاصله داشته باشد، از $\frac{1}{4}$ کمتر است.

■ چند تابع توزیع مهم. در این قسمت چند تابع توزیع مهم را معرفی می‌کنیم.

• تابع توزیع دوجمله‌ای Binomial Distribution

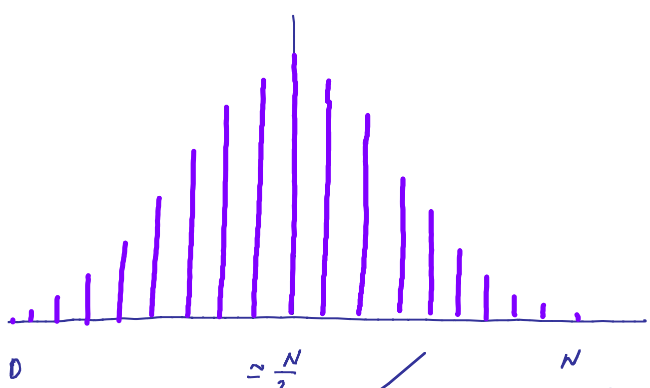
یک متغیر تصادفی در نظر بگیرید که در هر مرحله ۰ را یا ۱ را اختیار می‌کند. (مثل سکه‌ای که پرتاب می‌شود). فرض کنید

$$p(0)=p, \quad p(1)=q=1-p.$$

حال آزمایشی را در نظر بگیرید در آن N تا سکه از این نوع را پرتاب می‌کنیم (یا یک سکه را N بار پرتاب می‌کنیم). در این صورت تعداد دفعاتی که نتیجه ۰ ظاهر می‌شود یک متغیر تصادفی جدید مثل K است که مقادیرش از مجموعه $S_k=\{0,1,2,\dots N\}$ انتخاب می‌شود. احتمال این که k بار نتیجه ۰ ظاهر شود را با $P_N(k)$ نشان می‌دهیم. یک استدلال ساده ترکیباتی می‌گوید:

$$P_N(k)=\binom{N}{k}p^k q^{N-k}. \qquad (۱۳)$$

این تابع توزیع، تابع توزیع دوجمله‌ای مشهور است. شکل این تابع برای $p=q=1/2$ به صورت زیر است:



برای به دست آوردن ممان‌های این تابع توزیع، نخست تابع مولد را حساب می‌کنیم:

$$\tilde{P}_N(t)=\langle e^{tK}\rangle=\sum_{k=0}^{N}e^{tk}P_N(k)=\sum_{k=0}^{N}e^{tk}\binom{N}{k}p^k q^{N-k}$$

$$=\sum_{k=0}^{N}\binom{N}{k}(pe^t)^k q^{N-k}=(pe^t+q)^N. \qquad (۱۴)$$

که از آن به‌راحتی نتیجه می‌گیریم:

$$\langle k\rangle = \frac{d}{dt}\tilde{P}_N(t)\Big|_{t=0} = Np$$

$$\langle k^2\rangle = \frac{d^2}{dt^2}\tilde{P}_N(t)\Big|_{t=0} = N^2p^2 + Npq$$

$$\rightarrow \quad \sigma_k^2 := \langle k^2\rangle - \langle k\rangle^2 = Npq . \tag{15}$$

• یکی از کاربردهای تابع توزیع دوجمله‌ای توصیف گشت تصادفی یا ولگشت در یک بعد است. متحرکی را در نظر بگیرید که روی یک خط راست قدم می‌زند. این متحرک با احتمال p رو به جلو و با احتمال q رو به عقب قدم برمی‌دارد. طول هر کدام از قدم‌ها نیز برابر با 1 است. اگر این متحرک در ابتدا در مبدأ مختصات باشد، از خود می‌پرسیم که احتمال این که بعد از N قدم در نقطه‌ای به مختصات x باشد یعنی $P_N(x)$ چقدر است؟

در این جا X یک متغیر تصادفی است که مقادیرش را از مجموعهٔ $S_X = \{-N, -N+1, -N+2, \dots, 0, 1, \dots, N\}$ انتخاب می‌کند.

اگر این متحرک به نقطهٔ x رسیده باشد، تعداد قدم‌های رو به جلو را با N_+ و تعداد قدم‌های رو به عقب را با N_- نشان دهیم، داریم:

$$\begin{cases} N = N_+ + N_- \\ x = N_+ - N_- \end{cases} \longrightarrow \begin{cases} N_+ = \dfrac{N+x}{2} \\ N_- = \dfrac{N-x}{2} \end{cases}$$

بنابراین ←

$$P_N(x) = \binom{N}{N_+} p^{N_+} q^{N_-} = \binom{N}{\frac{N+x}{2}} p^{\frac{N+x}{2}} q^{\frac{N-x}{2}} \tag{16}$$

• تابع توزیع پواسون

فروشگاهی را در نظر بگیرید. در هر روز به طور متوسط تعداد λ مشتری به آن وارد می‌شود. از خود می‌پرسیم احتمال این

امروز، تعداد n تا مشتری به این فروشگاه وارد شوند چقدر است؟ مشابه این سوال در زمینه‌های مختلف پیش می‌آید:

مثال ۱) هسته‌های رادیواکتیوی را در نظر بگیرید که در کمیت به طور متوسط λ تا ذره از خود ساطع می‌کند. احتمال اینکه در یک ثانیه n ذره از خود ساطع کند چقدر است؟

مثال ۲) در یک روز بارندگی به هر سانتی‌متر مربع به طور متوسط λ تا قطره باران باریده است. احتمال اینکه به یک سانتی‌متر مربع n تا قطره باران دریافت کرده باشد چقدر است؟

همه این سوال‌ها پاسخ مشترکی دارند و این پاسخ تابع توزیع پواسون است که به صورت زیر بیان می‌شود:

$$P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda} .$$ ۱۷

برای فهم این رابطه مثال فروشگاه را در نظر بگیرید. یک روز را به تعداد بسیار زیادی بازه‌های کوچک تقسیم می‌کنیم. تعداد بازه‌ها را N می‌گیریم.

طول بازه را به قدری کوچک گرفته‌ایم که در هر بازه حداکثر یک مشتری بتواند وارد شود. احتمال ورود یک مشتری در این بازه زمانی را p، و احتمال وارد نشدن آن را با $q = 1-p$ نشان می‌دهیم. چون تعداد متوسط مشتریان وارد شده به فروشگاه برابر با λ است، داریم: $Np = \lambda$.

احتمال وارد شدن n مشتری به فروشگاه برابر است با:

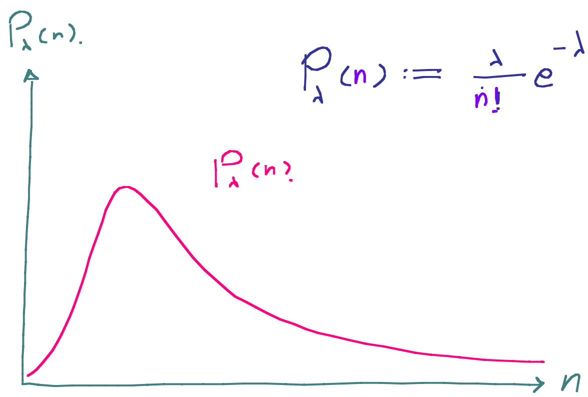
$$P_\lambda(n) = \lim_{\substack{N \to \infty \\ p \to 0 \\ Np = \lambda}} \binom{N}{n} p^n q^{N-n}$$

در این حد داریم :

$$P_N(n) = \frac{N!}{n!\,(N-n)!}\, p^n (1-p)^{N-n} \xrightarrow[p\to 0]{N\to\infty} \frac{N}{n!}\, p\, \left(1-\frac{\lambda}{N}\right)^{N-n} = \frac{\lambda}{n!}\, e^{-\lambda}$$

بنابراین تابع توزیع پواسون به شکل زیر تعریف می‌شود :

$$P_\lambda(n) := \frac{\lambda}{n!}\, e^{-\lambda}$$



متغیر تصادفی در این جا N است که مقادیر گسسته‌ای از ۰ تا ∞ را اختیار می‌کند. تابع مولد عبارت است از :

$$\tilde{P}_\lambda(t) = \langle e^{tN} \rangle = \sum_{k=0}^{\infty} e^{tn} \frac{\lambda}{n!} e^{-\lambda}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)}{n!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

که از این نتیجه می‌شود :

$$\langle N \rangle = \lambda, \quad \langle N^2 \rangle = \lambda^2 + \lambda$$

$$\rightarrow \quad \sigma = \sqrt{\lambda}.$$

• **مثال** : یک هسته رادیواکتیو در طی یک ساعت به طور متوسط ۱۲ ذره گاما از خود گسیل می‌کند. احتمال این که اولین ذره پس از ۲۰ دقیقه ساطع شود چقدر است؟

حل : این ذره در مدت ۲۰ دقیقه به طور متوسط $\frac{12}{3} = 4$ ذره از خود ساطع می‌کند. بنابراین معیار فراوانی پواسون $\lambda = 4$ است.

$P =$ احتمال اینکه اولین ذره را بعد از ۲۰ دقیقه ساطع کند $=$ احتمال اینکه تا ۲۰ دقیقه اصلاً ذره‌ای ساطع نکند $\times$ احتمال اینکه

در ۴۰ دقیقه بعدی (که λ برای آن 8 است) حداقل یک ذره ساطع کند $= P_1 \times P_2$.

از فرآیند پواسون می‌دانیم: $P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda}$ ← نمایی

$$P_1 = P_4(0) = e^{-4}$$

$$P_2 = 1 - P_8(0) = 1 - e^{-8} \quad \rightarrow \quad P = e^{-4}(1 - e^{-8}) \simeq 0.02.$$

- مثال: یک فروشگاه به طور متوسط در هر روز ۱۰۰ مشتری دارد. احتمال آنکه در نیمه اول روز ۴۰ مشتری و در نیمه دوم ۶۰ مشتری به فروشگاه وارد شوند چقدر است؟ فرض بر این است که مشتری‌ها به طور یکنواخت وارد می‌شوند.

حل: این فرآیند یک فرآیند پواسون است. در نیمه اول و دوم λ برابر با ۵۰ است. احتمال خواسته شده برابر است با:

$$P = P_{50}(40) \times P_{50}(60) = \frac{(50)^{40}}{40!} e^{-50} \times \frac{(50)^{60}}{60!} e^{-50}$$

برای محاسبه این اعداد: تقریب استرلینگ، استفاده کرده - برحسب آن: $n! \simeq \sqrt{2\pi n}\, n^n e^{-n}$

$$\rightarrow \quad 40! \times 60! \simeq 2\pi \times 10\sqrt{24}\; 40^{40} \times 50^{40}\; e^{-100}$$

پس کافیست جایگذاری در رابطه‌ی قبل ←

$$P \simeq \left(\frac{5}{4}\right)^{40} \left(\frac{5}{6}\right)^{60} \times \frac{1}{2\pi \times 10 \times \sqrt{24}} \simeq \left(\frac{25}{24}\right)^{100} \times \frac{1}{20\pi\sqrt{24}} \simeq \left(1 + \frac{100}{24}\right) \times \frac{1}{20\pi\sqrt{24}}$$

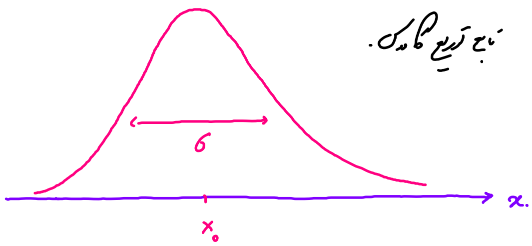
$$\rightarrow \quad P \simeq 0.016$$

- تابع توزیع گاوسی.

تابع توزیع گاوسی برای یک متغیر پیوسته X تعریف می‌شود که مقادیر خود را در حوزه $(-\infty, +\infty)$ اختیار می‌کند.

این تابع به شکل زیر است:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-x_0)^2}{\sigma^2}}.$$ ۱۸



تابع توزیع گاوسی.

حال تابع مولد توزیع گاوسی را محاسبه می‌کنیم:

$$\tilde{p}(k) = \int e^{ikx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-x_0)^2} dx.$$

برای محاسبه این تابع مولد از دو رابطه ساده که در آینده نیز به آنها نیاز خواهیم داشت استفاده می‌کنیم. این دو رابطه که خواننده به راحتی می‌تواند آنها را تحقیق کند عبارتند از:

$$\boxed{\int_{-\infty}^{\infty} e^{-\frac{1}{2}ax^2} dx = \sqrt{\frac{2\pi}{a}} \quad , \quad \int_{-\infty}^{\infty} e^{-\frac{1}{2}ax^2+bx} = \sqrt{\frac{2\pi}{a}}\, e^{\frac{1}{2}\frac{b^2}{a}}}$$ ۱۹

با استفاده از این دو رابطه تابع مولد بدست می‌آید:

$$\tilde{p}(k) = e^{-\frac{1}{2}\sigma^2 k^2 + ikx_0}.$$ ۲۰

با استفاده از این رابطه بدست می‌آوریم:

$$\langle x \rangle = x_0 \qquad \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2.$$ ۲۱

پس پارامترهای x_0 و σ که در تعریف تابع توزیع گاوسی به کار رفته‌اند به ترتیب نشان دهنده متوسط و واریانس متغیر تصادفی x هستند.

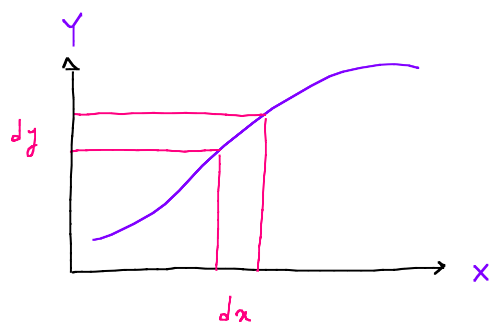
□ تغییر متغیرهای تصادفی

فرض کنید که X یک متغیر تصادفی با تابع توزیع $P_X(x)$ باشد و Y یک تابع مشخص از متغیر X باشد،

مثلاً $y = x^2+x+3$. از آنجا که مقادیر X تصادفی هستند، مقادیر Y نیز تصادفی خواهند بود.

از خود می‌پرسیم تابع توزیع احتمالات Y چیست و چگونه بدست می‌آید؟

نخست شکل زیر را در نظر بگیرید.



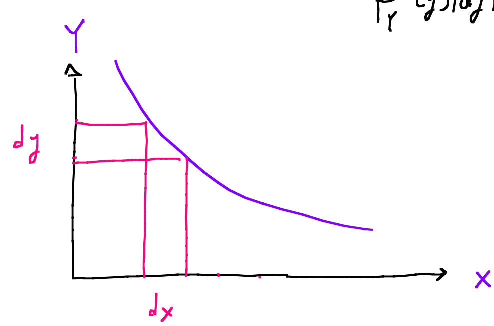
واضح است که احتمال اینکه y در فاصله‌ی نشان داده شده باشد برابر است با احتمال اینکه x در فاصله‌ی مربوط قرار بگیرد. بنابراین:

$$P_Y(y)\,dy = P_X(x)\,dx \longrightarrow P_Y(y) = P_X(x)\frac{dx}{dy}$$

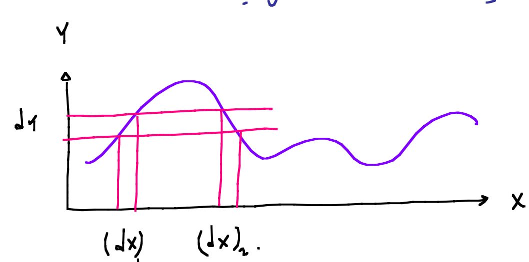
اما در این رابطه یک اشکال وجود دارد و آن این که ممکن است که $\frac{dx}{dy}$ منفی باشد. رابطه‌ی صحیح آن است که بنویسیم

$$P_Y(y)\,|dy| = P_X(x)\,|dx|$$

که حالتی مثل شکل زیر را نیز پوشش می‌دهد:



ولی رابطه‌ی بالا نیز کامل نیست، زیرا ممکن است هر نقطه‌ی Y متناظر با چند نقطه‌ی X باشد، (شکل زیر):



بنابراین رابطه حاصل به شکل زیر است :

$$P_Y(y)|dy| = \sum_i P_X(x_i)|dx_i|$$

که در آن جمع روی تمامی x_i هایی است که در رابطه $y = f(x_i)$ صدق می‌کنند. این رابطه را می‌توان به شکل فشرده‌تری نیز نوشت. کافی است که به این خاصیت از تابع دلتا رجوع کنیم:

$$\delta(y - f(x)) = \sum_i \delta(x - x_i) \left|\frac{dx}{dy}\right|_{x=x_i} \qquad ۲۲$$

که در آن جمع روی تمامی x_i هایی است که در رابطه $y = f(x_i)$ صدق می‌کنند. بنابراین ←

$$P_Y(y) = \sum_i P_X(x_i) \left|\frac{dx}{dy}\right|_{x=x_i} = \int P_X(x)\, \delta(y - f(x))\, dx \qquad ۲۳$$

که همین است که از تابع $\delta(y - f(x))$ متوسط گرفته‌ایم. بنابراین به طور خلاصه

$$\boxed{P_Y(y) = \langle \delta(y - f(X)) \rangle} \qquad ۲۴$$

- مثال: $P_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2}$, $y = x^2 - 4x + 4$. $P_Y(y)$ را بخواهیم
را بیابید.

حل:

$$y = x^2 - 4x + 4 = (x-2)^2 \longrightarrow x = 2 \pm \sqrt{y} . \qquad \left|\frac{dx}{dy}\right| = \frac{1}{2} y^{-1/2}$$

$$\rightarrow P_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(2+\sqrt{y})^2} \frac{1}{2\sqrt{y}} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(2-\sqrt{y})^2} \frac{1}{2\sqrt{y}}$$

$$\rightarrow P_Y(y) = \begin{cases} \frac{1}{2\sigma\sqrt{2\pi y}} \left\{ e^{-\frac{1}{2\sigma^2}(2+\sqrt{y})^2} + e^{-\frac{1}{2\sigma^2}(2-\sqrt{y})^2} \right. & y > 0 \\ 0 & y \leq 0 \end{cases}$$

- مثال. ولگشت با طول گام‌های متغیر. ولگردی را در نظر بگیرید که حسب اندازه هر گام او، که آن را با متغیر تصادفی ند

نشان دهید که تابع توزیع گاوسی زیر بهنجار است؛

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

اگر این متغیر N بار تکرار شود، تابع توزیع موقعیت آن را نسبت به مبدأ حرکت بدست آورید. فرض کنید حرکت از مبدأ شروع به حرکت کرده است.

حل: مکان گام Nام عبارت است از:

$$X = S_1 + S_2 + \cdots S_N$$

که در آن S_iها هر کدام یک متغیر تصادفی هستند. داریم:

$$P_X(x) = \langle \delta(x-(S_1+S_2+\cdots S_N)) \rangle = \int ds_1 \cdots ds_N \, \delta(x-(s_1+\cdots s_N)) P_S(s_1)\cdots P_S(s_N)$$

$$= \int ds_1 \cdots ds_N \, \delta(x-(s_1+\cdots s_N)) \frac{1}{(\sigma\sqrt{2\pi})^N} e^{-\frac{1}{2\sigma^2}(s_1^2+\cdots s_N^2)} \qquad ۲۵$$

اگر محاسبه انتگرال بالا که راحت نیست، به جای $P_X(x)$، مرحله به مرحله این یعنی $\tilde{P}_X(k)$ را حساب کنیم:

$$\tilde{P}_X(k) = \int e^{ikx} P_X(x)\,dx = \int e^{ik(s_1+\cdots s_N)} ds_1 \cdots ds_N \, e^{-\frac{1}{2\sigma^2}(s_1^2+\cdots s_N^2)}$$

$$= (\tilde{P}_S(k))^N = e^{-\frac{1}{2}\sigma^2 k^2 N} \qquad ۲۶$$

با استفاده از رابطه ۲-۱۸، نتیجه می‌گیریم

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

این رابطه نشان می‌دهد که تابع توزیع مسافت طی شده نیز یک تابع گاوسی با متوسط صفر ولی با واریانس $\sigma\sqrt{N}$ است.

دقت کنید که σ در واقع طول متوسط هر گام است، و رابطه فوق بیان می‌کند که مسافت متوسط طی شده برابر است با $\sqrt{N}$ در طول متوسط هر گام.

■ تابع توزیع چند متغیر تصادفی

ممکن است یک آزمایش داشته باشیم که نتیجه آن جمع چند متغیر تصادفی باشد. در این صورت با تابع توزیع چند متغیر تصادفی روبرو هستیم که آن را با

$P(x_1, x_2, \dots x_n)$ را در نظر بگیرید. برای این تابع توزیع روابط زیر برقرارند:

a) $$\int p(x_1, \dots x_n)\, dx_1\, dx_2 \dots dx_n = 1.$$ شرط بهنجارش

b) $$\langle f(x_1, x_2, \dots x_n) \rangle = \int f(x_1, \dots x_n)\, P(x_1, \dots x_n)\, dx_1 \dots dx_n.$$ متوسط

c) $$\tilde{P}(k_1, \dots k_n) := \langle e^{i(k_1 X_1 + \dots k_n X_n)} \rangle.$$ تابع مولد

$$\langle X_r^{m_1} \rangle = \frac{\partial^{m} \tilde{P}}{\partial (i k_r)^{m}} \Big|_{k=0} \quad , \quad \langle X_r^{m_1} X_s^{m_2} \rangle = \frac{\partial^{m_1+m_2}}{\partial (i k_r)^{m_1} \partial (i k_s)^{m_2}} \tilde{P} \Big|_{k=0}$$

d) $$P(x_1, \dots x_m) := \int p(x_1, \dots x_n)\, dx_{m+1}\, dx_{m+2} \dots dx_n.$$ تابع توزیع کاهش یافته

تابع توزیع کاهش یافته یا Reduced distribution function، احتمال این است که متغیرهای X_1 تا X_m مقادیر x_1 تا x_m را داشته باشند مستقل از مقادیری که متغیرهای دیگر دارند. می‌توان تابع توزیع کاهش یافته یک متغیره، ۲ متغیره، تا $n-1$ متغیره تعریف کرد.

- مثال: یک هسته رادیواکتیو را در نظر بگیرید که از خود ذرات α ساطع می‌کند. $P(t_1, t_2)$ یعنی احتمال این که ذره اول را در لحظه t_1 و ذره دوم را در لحظه t_2 ساطع کند، یک تابع توزیع دو متغیره است. دقت کنید که در اینجا t_1 و t_2 متغیرهای تصادفی پیوسته هستند و $P(t_1, t_2)$ را باید به صورت یک چگالی بفهمید.

- مثال: همان هسته را در نظر بگیرید که طی دو ساعت متوالی مشاهده می‌شود. $P(m,n)$ یعنی احتمال این که در ساعت اول m ذره و در ساعت دوم n ذره ساطع کند یک تابع توزیع دو متغیره است. در اینجا m , n متغیرهای گسسته هستند.

● مثال: تابع توزیع چندجمله‌ای Multinomial Distribution

این تابع توزیع، تعمیم سرراست توزیع دوجمله‌ای است. متغیری را در نظر بگیرید که m مقدار متفاوت به خود می‌گیرد، با احتمالات $p_1, p_2, \dots p_m$. (مثل تاسی که وقتی ریخته می‌شود، ۶ نتیجه متفاوت ممکن است به دست دهد.) حال آزمایش را N بار انجام می‌دهیم و از خود می‌پرسیم که احتمال این که k_1 بار نتیجه ۱، k_2 بار نتیجه ۲، ... k_m بار نتیجه m به دست آید چقدر است. واضح است که $k_1+k_2+\dots+k_m=N$. باز هم یک استدلال ترکیبی ساده می‌گوید:

$$P_N(k_1,k_2,\dots k_m) = \frac{N!}{k_1!\,k_2!\dots k_m!}\,p_1^{k_1}p_2^{k_2}\dots p_m^{k_m}.$$

● توابع همبستگی (Correlation function)

بهتر است که مفهوم همبستگی را در ساده‌ترین شکل آن برای یک تابع توزیع دو متغیره ببینیم. خواننده خود می‌تواند این مفاهیم را به چند متغیر تعمیم دهد.

یک تابع توزیع دو متغیره مثل $P_{XY}(x,y)$ در نظر می‌گیریم. توابع توزیع حاشیه‌ای عبارتند از:

$$P_X(x) = \int P_{XY}(x,y)\,dy \qquad \text{و} \qquad P_Y(y) = \int P_{XY}(x,y)\,dx$$

اگر $P_{XY}(x,y)=P_X(x)P_Y(y)$ باشد، دو متغیر تصادفی مستقل محسوب می‌شوند. برای این دو متغیر تصادفی داریم:

$$\langle XY\rangle = \int xy\,P_{xy}(x,y)\,dx\,dy = \int xy\,P_x(x)P_Y(y)\,dx\,dy = \langle X\rangle\langle Y\rangle.$$

بنابراین در حالت کلی تفاضل این دو کمیت معیار خوبی برای میزان وابستگی دو متغیر تصادفی است:

$$\sigma_{XY} := \langle XY\rangle - \langle X\rangle\langle Y\rangle.$$

• تابع توزیع چند متغیره گاوسی.

تابع توزیع گاوسی به شکل زیر به چند متغیر تعمیم می‌یابد:

$$p(x_1,x_2,\dots,x_n) = C e^{-\frac{1}{2}\sum_{ij} x_i A_{ij} x_j}$$

که در آن A_{ij} یک ماتریس متقارن و مثبت است. ماتریس مثبت ماتریسی است که همه ویژه‌مقادیر آن مثبت باشند. ردیف بالا را

می‌توان به شکل زیر نوشت:

$$p(x) = C\, e^{-\frac{1}{2}x^T A x}$$

که در آن C یک ثابت بهنجارش است. نخست ثابت C را حساب می‌کنیم.

$$\int dx_1\dots dx_n\, p(X) = C\int e^{-\frac{1}{2}x^T A x}\, DX$$

که در آن $DX := dx_1\dots dx_n$. این یک انتگرال زوج گاوسی است که یک تغییر متغیر مناسب انجام می‌دهیم. چون A متقارن است

یک ماتریس متعامد S آن را قطری می‌کند یعنی

$$A = S^T\Omega S\ ,\quad S^T S = I$$

که در آن Ω یک ماتریس قطری است که درایه‌های قطری آن ویژه‌مقادیر مثبت A و کمتر از صفر نیست است.

$$\Omega = \begin{pmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_n \end{pmatrix}$$

$$\rightarrow \int DX\, p(x) = C\int e^{-\frac{1}{2}x^T S^T \Omega S x}\, DX$$

قرار می‌دهیم: $y = Sx \leftarrow$ چون S ماتریس متعامد است دترمینان آن ± 1 است. از این رو داریم

$$DX = |\det S|\, DY = DY \longrightarrow$$

$$\int DX\, p(x) = C\int e^{-\frac{1}{2}y^T\Omega y} DY = C\int \prod_{i=1}^{N} e^{-\frac{1}{2}\omega_i y_i^2}\, dy_i = C\prod_{i=1}^{N}\left(\frac{2\pi}{\omega_i}\right)^{1/2}$$

$$= C\,(2\pi)^{N/2}\frac{1}{\sqrt{\det A}}\ .$$

بنابراین، برای بهنجارش توزیع احتمال می‌بایست داشته باشیم:

$$C = (\det A)^{1/2}(2\pi)^{-N/2}.$$

• تابع مولد داریم: $$p(x) = C\, e^{-\frac{1}{2} x^T A x} \leftarrow$$

$$\tilde{p}(k) = \int p(x)\, e^{ik\cdot x}\, Dx$$

$$= \int C\, e^{-\frac{1}{2} x^T A x + ik\cdot x}\, Dx$$

می‌توان با همان تغییر متغیری که در بالا توضیح داده شد، $\tilde{p}(k)$ را نیز محاسبه کرد: ←

$$\tilde{p}(k) = e^{\frac{1}{2}(ik)^T A^{-1} (ik)}.$$

که در آن A^{-1} وارون ماتریس A است و k برداری است که مشتق‌گیری نسبت به مؤلفه‌های آن توابع همبستگی را به دست می‌دهد.

$$\rightarrow \quad \langle x_i x_j \rangle = \frac{\partial^2}{\partial(ik_i)\,\partial(ik_j)}\, \tilde{p}(k) = (A^{-1})_{ij}$$

برای سادگی نمادگذاری از این به بعد $(A^{-1})_{ij}$ را با Δ_{ij} نمایش می‌دهیم. پیدا کردیم:

$$\langle x_i x_j \rangle = \Delta_{ij}$$

به خاطر تقارن وارونی است، تمام ممان‌های فرد صفر هستند، یعنی

$$\langle x_i \rangle = \langle x_i x_j x_k \rangle = \dots = 0.$$

حال دیگر ما می‌توانیم با استفاده از تابع مولد حساب کنیم. خواننده خود می‌تواند این محاسبه را انجام دهد. برای تابع چهارنقطه‌ای یعنی این تابع زیر نتیجه چنین است:

$$\langle x_i x_j x_k x_l \rangle = \langle x_i x_j \rangle \langle x_k x_l \rangle + \langle x_i x_k \rangle \langle x_j x_l \rangle + \langle x_i x_l \rangle \langle x_j x_k \rangle$$

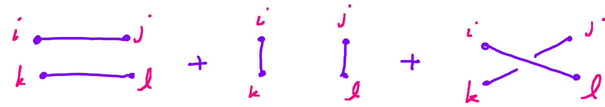
$$= \Delta_{ij} \Delta_{kl} + \Delta_{ik} \Delta_{jl} + \Delta_{il} \Delta_{jk}.$$

این نتیجه، یک نمونه از قضیهٔ کلی موسوم به قضیهٔ ویک Wick's theorem است. مطابق این قضیه هر تابع $2N$ نقطه‌ای به صورت مجموعی از تمام حاصل‌ضرب‌های ممکن از توابع دونقطه‌ای یعنی $\langle x_i x_j \rangle = \Delta_{ij}$ خواهد بود. دقت کنید

قضیه فقط برای توزیع گاوسی برقرار است.

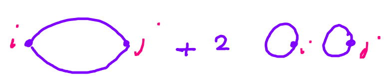
برای راحتی از نمودارهایی برای نشان دادن رابطه توزیع همبستگی استفاده کرده. $\langle x_i x_j \rangle = \Delta_{ij}$ را با  نشان می‌دهیم. در نتیجه رابطه قبلی به شکل زیر نوشته می‌شود.

$$\langle x_i x_j x_k x_l \rangle =$$



یا

$$\langle x_i^2 x_j^2 \rangle =$$



$$= \Delta_{ij}^2 + 2\Delta_{ii}\Delta_{jj}$$

یا

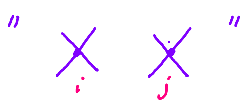
$$\langle x_i^4 x_j x_k \rangle =$$

$$= 3\,\Delta_{jk}\,\Delta_{ii}^2 + 12\,\Delta_{ij}\,\Delta_{ik}\,\Delta_{ii}$$

روشی که برای به دست آوردن دیاگرام‌های بالا به کار رفته است آن است که نقاط لازم با شاخص‌های مناسب در نظر گرفته شده و انتهای آن‌ها به انواع حالت‌های ممکن به هم متصل شده است (بنا بر قضیه Wick). ضرایب عددی نشان دهنده تعداد طرقی است که این اتصال‌ها را انجام شوند.

به عنوان آخرین مثال:

$$\langle x_i^4 x_j^4 \rangle =$$



$\rightarrow$ منظور از این عبارت " " آن است که تمام نقاط را به طرق مختلف مطابق با قضیه ویک به هم وصل کنیم. بنابراین

$$\langle x_i^4 x_j^4 \rangle =$$

$$= 72\,\Delta_{ii}\,\Delta_{ij}^2\,\Delta_{jj} + 9\,\Delta_{ii}^2\,\Delta_{jj}^2$$

• احتمال شرطی Conditional probability

احتمال شرطی $P(X|Y)$ به شکل زیر تعریف می‌شود:

$$p(x|y) := \frac{p(x,y)}{p(y)}$$

که معنی آن‌ها پس از بازنویسی آن به صورت $p(x,y)=p(x|y)p(y)$ می‌توان فهمید. واضح است که این تابع توزیع در حالت زیر صدق می‌کند:

$$\sum_x p(x|y) = 1.$$

تابع توزیع شرطی‌ها به صورت زیر برای چند متغیر تعمیم دارد:

$$p(x_1, x_2, \ldots x_m | x_{m+1}, x_{m+2}, \ldots x_n) = \frac{p(x_1, x_2, \ldots x_n)}{p(x_{m+1}, \ldots x_n)}$$

• مثال. دو سکه نظری را در شرایط خط‌ها به ترتیب با 0، 1 نشان می‌دهیم. احتمالات مربوط به آمدن این سکه‌ها به صورت زیر است:

سکه دوم \ سکه اول	0	1
0	$1/6$	$1/4$
1	$2/6$	$1/4$.

برای تک سکه داریم:

$$P_1(0) = 1/2 \qquad P_1(1) = 1/2$$

$$P_2(0) = 5/12 \qquad P_2(1) = 7/12$$

$$P_{1,2}(0|1) = \frac{P_{12}(0,1)}{P_2(1)} = \frac{2/6}{7/12} = \frac{4}{7}. \qquad P_{2,1}(0|1) = \frac{P_{21}(0,1)}{P_1(1)} = \frac{1/4}{1/2} = 1/2.$$

قضیه حد مرکزی Central limit theorem.

فرض کنید که $X_1, X_2, \ldots, X_N$ متغیرهای تصادفی باشند با تابع توزیع $P_1(x_1)$، $P_2(x_2)$، …، $P_N(x_N)$.

حال متغیر تصادفی Y را به شکل زیر در نظر می‌گیریم:

$$y := \frac{x_1+x_2+\cdots x_N}{N}$$

از خود می‌پرسیم که تابع توزیع متغیر تصادفی Y چیست؟ پاسخ این سوال این است که در حد N های بسیار بزرگ، تابع توزیع Y به سمت یک تابع گاوسی به شکل زیر میل می‌کند:

$$P_Y(y) = \frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma_y^2}(y-y_0)^2}$$

که در آن

$$y_0 := \frac{1}{N}\sum_{i=1}^{N}\langle x_i\rangle$$

و

$$\sigma_y := \frac{1}{N}\sum_{i=1}^{N}\sigma_{x_i}$$

اهمیت این قضیه در این است که مستقل از این که توابع توزیع P_1 تا P_N چگونه باشند همواره تابع توزیع نهایی یک گاوسی خواهد شد. البته شرایط محدودی باید بر روی این توابع توزیع وجود داشته باشند که در حین اثبات قضیه به آن‌ها برخورد می‌کنیم.

هرگاه متغیرهای تصادفی همگی یکسان باشند، دو رابطه بالا به شکل زیر در می‌آیند:

$$y_0 = \langle x\rangle \quad \text{و} \quad \sigma_y = \frac{1}{N}\sigma_x\,.$$

- اثبات: می‌دانیم که

$$P_Y(y) = \left\langle \delta\left(y-\frac{x_1+x_2+\cdots x_N}{N}\right)\right\rangle = \int dx_1\cdots dx_N\, \delta\left(y-\frac{x_1+\cdots x_N}{N}\right) P_1(x_1)\cdots P_N(x_N)$$

به جای $P_Y(y)$، تابع مولد آن یعنی $\tilde{P}_Y(k)$ را حساب می‌کنیم:

$$\rightarrow \tilde{P}_Y(k) \equiv \int e^{iky} P_Y(y)\,dy = \int e^{iky}\,dy \int dx_1\cdots dx_N\, \delta\left(y-\frac{x_1+\cdots x_N}{N}\right) P_1(x_1)\cdots P_N(x_N)$$

حال روی y انتگرال می‌گیریم. ←

$$\tilde{P}_Y(k) = \int e^{ik\left(\frac{x_1+x_2+\cdots x_N}{N}\right)} P_1(x_1)P_2(x_2)\cdots P_N(x_N)\, dx_1\cdots dx_N$$

انتگرال روی x_1 تا x_N از هم جدا می‌شوند هر کدام تابع مولد متغیر مربوط را به ازای $\frac{k}{N}$ می‌دهند. بنابراین

$$\tilde{P}_Y(k) = \tilde{P}_1\left(\frac{k}{N}\right)\tilde{P}_2\left(\frac{k}{N}\right)\cdots\tilde{P}_N\left(\frac{k}{N}\right).$$

اما برای هر تابع مولد تک متغیره بسط زیر را داریم:

$$\tilde{P}_1(k) = 1 + ik\langle x_1\rangle + \frac{(ik)^2}{2}\langle x_1^2\rangle + O(k^3) = e^{ik\langle x_1\rangle + \frac{(ik)^2}{2}\left(\langle x_1^2\rangle - \langle x_1\rangle^2\right) + O(k^3)}$$

بنابراین ←

$$\tilde{P}_i(k_i) = e^{i\frac{k}{N}\langle x_i\rangle + \frac{(ik)^2}{2N^2}\sigma_{x_i}^2 + O\left(\frac{1}{N^3}\right)}.$$

با ضرب کردن $\tilde{P}_i(k_i)$ ها در یکدیگر ←

$$\tilde{P}_Y(k) = e^{ik\left(\frac{\langle x_1\rangle+\cdots\langle x_N\rangle}{N}\right) + (ik)^2\frac{1}{2}\left(\frac{\sigma_{x_1}^2+\cdots\sigma_{x_N}^2}{N^2}\right) + O\left(\frac{1}{N^3}\right)}.$$

هرگاه تعریف کنیم:

$$y_0 := \frac{\langle x_1\rangle+\cdots\langle x_N\rangle}{N}$$

$$\sigma_y^2 := \frac{\sigma_{x_1}^2+\cdots+\sigma_{x_N}^2}{N^2}$$

و از توان $\frac{1}{N^3}$ به بعد صرف‌نظر کنیم به رابطهٔ زیر می‌رسیم

$$\tilde{P}_Y(k) = e^{iky_0 + \frac{1}{2}(ik)^2\sigma_y^2}$$

که به معنای آن است که

$$P_Y(y) = \frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{1}{2\sigma_y}(y-y_0)^2}$$

یعنی این که تابع توزیع y، در واقع یک تابع توزیع گاوسی است با متوسط y_0 و پهنای واریانس σ_y.

چند رابطه مفید درباره اعداد بزرگ

در مکانیک آماری تعداد ذرات یک سیستم بسیار بزرگ است. لذا گاهی اوقات پیش می‌آید که می‌خواهیم مجموع عبارات یک سری را که تعداد عبارات آن بسیار زیاد است حساب کنیم، یا پیش می‌آید که می‌خواهیم یک انتگرال را تخمین بزنیم. تحت شرایط معینی می‌توانیم تقریب‌های بسیار خوبی برای این کمیت‌ها که با اعداد بسیار بزرگ سر و کار دارند پیدا کنیم. در این بخش بعضی از این تکنیک‌ها را معرفی می‌کنیم.

- یک سری به صورت زیر در نظر بگیرید:

$$S = \sum_{i=1}^{N} E_i$$

که در آن E_i ها عبارات مثبت هستند. می‌خواهیم در حد Nهای بزرگ S را تخمین بزنیم. فرض کنید E_{max} بزرگترین جمله بین E_i ها باشد. در این صورت

$$E_{max} \leq S \leq N E_{max}$$

$$\rightarrow \quad \ln E_{max} \leq \ln S \leq \ln N + \ln E_{max}$$

$$\rightarrow \quad \frac{\ln E_{max}}{N} \leq \frac{\ln S}{N} \leq \frac{\ln N}{N} + \frac{\ln E_{max}}{N}$$

حال اگر $\frac{\ln E_{max}}{N}$ در حد Nهای بزرگ به یک مقدار ثابت میل کند، یعنی اگر

$$\lim_{N\to\infty} \frac{\ln E_{max}}{N} = \phi_{max}$$

از آنجایی که در حد $N \to \infty$ ، $\frac{\ln N}{N} \to 0$ ، نتیجه می‌گیریم که

$$\frac{\ln S}{N} = \phi_{max} \longrightarrow \ln S = N\phi_{max} \longrightarrow S = e^{N\phi_{max}} = E_{max}.$$

بنابراین در این حد، S برابر با بزرگترین جمله است و می‌توان از بقیه جملات صرف نظر کرد.

- مثال: حاصل جمع زیر را به ازای Nهای بزرگ پیدا کنید.

$$S = \sum_{n=1}^{N} \frac{1}{n} \, 2^{n}$$

$$I = \int_{-\infty}^{\infty} dx \; e^{N\left\{\phi(x_0) - \frac{1}{2}|\phi''(x_0)|(x-x_0)^2 + \alpha(x-x_0)^3 + \beta(x-x_0)^4\right\}}$$

که در آن $\alpha = \frac{1}{3!}\phi'''(x_0)$ و $\beta = \frac{1}{4!}\phi^{iv}(x_0)$.

I را به شکل زیر می‌توانیم بنویسیم $\leftarrow$

$$I = e^{N\phi(x_0)} \int e^{-\frac{1}{2}N|\phi''(x_0)|(x-x_0)^2} \left\{ 1 + N\alpha(x-x_0)^3 + N\beta(x-x_0)^4 + \cdots \right\} dx$$

با تغییر متغیر $\xi = x - x_0 \leftarrow$

$$I = e^{N\phi(x_0)} \int e^{-\frac{1}{2}N|\phi''(x_0)|^2 \xi^2} \left\{ 1 + N\alpha\xi^3 + N\beta\xi^4 + \cdots \right\}.$$

حل:

$$\rightarrow S = \sum_{n=1}^{N} e^{-\ln n + n\ln 2} = \sum_{n=1}^{N} e^{\phi(n)}$$

ابتدا جمله ماکزیمم را پیدا کنیم.

$$\phi(n) = -\ln n + n\ln 2 \rightarrow \phi'(n) = \frac{-1}{n} + \ln 2 = 0 \rightarrow n = \frac{1}{\ln 2}$$

$\phi''(n) = \frac{1}{n^2} > 0 \rightarrow$ نقطهٔ $n = \frac{1}{\ln 2}$ یک نقطهٔ مینیمم است

نقطهٔ ماکزیمم حالا N یعنی آخرین نقطه است. پس $\leftarrow$

$$\boxed{S \simeq \frac{1}{N} 2^N.}$$

• مثال: حاصل جمع زیر را برای Nهای بزرگ پیدا کنید.

$$S = \sum_{n=1}^{N} n^4\, 2^{-3n}$$

حل: $\phi(n) = 4\ln n - 3n\ln 2$

$$S = \sum_{n=1}^{N} e^{4\ln n - 3n\ln 2} = \sum_{n=1}^{N} e^{\phi(n)}$$

$$\rightarrow \phi'(n) = \frac{4}{n} - 3\ln 2 \qquad \phi'(n_*) = 0 \rightarrow n_* = \frac{4}{3\ln 2}$$

$\phi''(n_*) < 0 \rightarrow$ $\phi(n^*)$ ماکزیمم است.

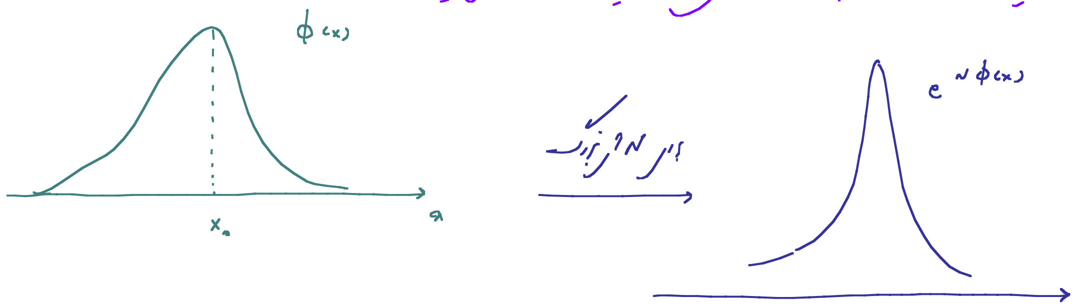
$$\boxed{\phi(n^*) = \left(\frac{4}{3\ln 2}\right)^4 e^{-4}}$$

• تقریب نقطه زینی Saddle point approximation.

$$I = \int e^{N\phi(x)}\, dx$$

می‌خواهیم انتگرال به شکل زیر را تخمین بزنیم که در آن N یک عدد خیلی بزرگ است.

فرض کنید $\phi(x)$ تابعی باشد که دارای یک ماکزیمم باشد. (شکل زیر).



شکل‌ها به وضوح نشان می‌دهند که برای N های بزرگ، بیشترین سهم انتگرال از نقطه ماکزیمم و اطراف آن ناشی می‌شود. بنابراین

$\phi(x)$ را حول نقطه ماکزیمم یعنی حول x_0 بسط داده و بدست می‌آوریم:

$$I = \int dx\; e^{N\left\{\phi(x_0) - \frac{1}{2}|\phi''(x_0)|\,(x-x_0)^2 + \dots\right\}}\, dx$$

در این جا از این موضوع استفاده کرده‌ایم که $\phi''(x_0)$ منفی است چون نقطه x_0 ماکزیمم است. اگر از جملات مرتبه بالاتر از ۲ صرف نظر کنیم

I تبدیل به یک انتگرال گاوسی شده و بر حسب تابع قابل محاسبه است: در واقع با تغییر متغیر $\xi := x - x_0$ و توجه به اینکه محدوده انتگرال‌گیری را به $(-\infty, \infty)$

$$I = e^{N\phi(x_0)} \int_{-\infty}^{\infty} d\xi \, e^{-\frac{1}{2}N|\phi''(x_0)|\xi^2}$$ ← گسترش دادیم

$$\longrightarrow \quad \boxed{I \simeq e^{N\phi(x_0)} \sqrt{\frac{2\pi}{N|\phi''(x_0)|}}}$$

حال سوال این است که با صرف نظر کردن از جملات مرتبه بالاتر چه چیزی را از دست داده‌ایم. در این جا بایستی از قضیه‌ی ویک در انتگرال‌های گاوسی استفاده کنیم. فرض کنید یک جمله‌ی مرتبه سه در $\phi(x)$ حول نقطه x_0 نگه داریم: در این صورت ←

جمله‌ی متناسب با ξ^3 در انتگرال‌گیری صفر خواهد شد و جمله‌ی بعدی با استفاده از قضیه‌ی ویک قابل محاسبه است. ←

$$I = e^{N\phi(x_0)} \sqrt{\frac{2\pi}{N|\phi''(x_0)|}} \left\{ 1 + N\beta \langle \xi^4 \rangle \right\}$$

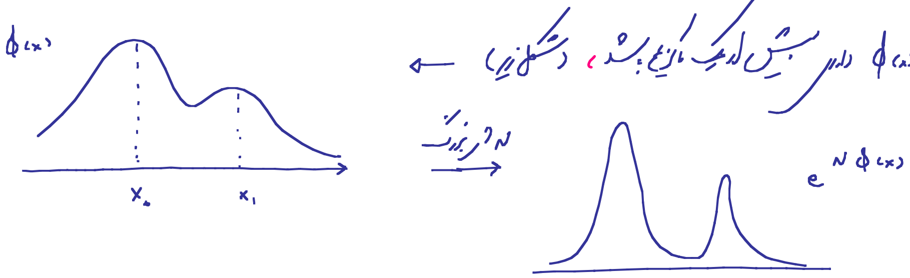
اما بنابر قضیه‌ی ویک می‌دانیم که

$$\langle \xi^4 \rangle = 3 \langle \xi^2 \rangle^2$$

و می‌دانیم که $\langle \xi^2 \rangle \equiv \Delta = (N|\phi''(x_0)|)^{-1}$ بنابراین ← $\langle \xi^4 \rangle \sim N^{-2}$

بنابراین جمله‌ی $N\beta\langle\xi^4\rangle$ متناسب با N^{-1} است و در حد N های بزرگ به سمت صفر میل می‌کند.

حال فرض کنید تابع $\phi(x)$ دارای بیش از یک ماکزیمم باشد، در شکل زیر ←



در این صورت I مجموع دو انتگرال گاوسی خواهد شد ←

$$I \simeq e^{-N\phi(x_0)}\sqrt{\frac{2\pi}{N|\phi''(x_0)|}} + e^{-N\phi(x_1)}\sqrt{\frac{2\pi}{N|\phi''(x_1)|}} .$$

• تقریب استرلینگ Stirling Approximation

تقریب استرلینگ را برای فاکتوریل‌ها در حالات ترمودینامیکی داریم زیرا عدد N های بزرگ سروکار داریم ! N ها بدست می‌دهد. بنابراین تقریب داریم

$$N! \simeq \sqrt{2\pi N}\; N^N e^{-N}$$

یا

$$\ln N! = N\ln N - N$$

در عبارت دوم که جمله‌ی $\frac{1}{2}\ln N + \frac{1}{2}\ln 2\pi$ در مقابل بقیه جملات نیز صرف‌نظر کرده‌ایم.

نکته‌ی مهم در مورد تقریب استرلینگ این است که لزومی ندارد N بسیار بزرگ باشد تا این تقریب خوب باشد. حتی برای $N=10$ نیز خطای این تقریب بسیار کم است.

• اثبات تقریب استرلینگ: می‌دانیم

$$\int_0^\infty e^{-t}\, t^N\, dt = N!$$

$$\rightarrow N! = \int_0^\infty e^{-t+N\ln t}\, dt = \int_0^\infty e^{N\{\ln t - \frac{t}{N}\}}\, dt = \int_0^\infty e^{N\phi(t)}\, dt$$

حال انتگرال مربوط به e را با استفاده از تقریب نقطه زینی حساب می‌کنیم. $\leftarrow$

$$\phi(t) = \ln t - \frac{t}{N} \qquad \phi'(t) = \frac{1}{t} - \frac{1}{N} \qquad \phi'(N) = 0 \qquad \phi''(N) = \frac{-1}{N^2} < 0$$

$$\rightarrow N! = e^{N\phi(N)}\sqrt{\frac{2\pi}{N\,{}^{1}/_{N^2}}} = e^{N[\ln N - 1]}\sqrt{2\pi N}$$

$$\rightarrow N! = N^N e^{-N}\sqrt{2\pi N}$$

آنتروپی، اطلاعات و بی‌نظمی

در ترمودینامیک با مفهوم آنتروپی و معنای ماکروسکوپی آن آشنا شدیم. در این درس با مفهوم میکروسکوپی آنتروپی آشنا می‌شویم و رابطه آن را با اطلاعات و بی‌نظمی توضیح می‌دهیم. نخست به مفهوم اطلاعات توجه می‌کنیم. چگونه میزان اطلاعات را به صورت یک مفهوم کمّی می‌فهمید؟ چگونه میزان اطلاعات را در یک پیام اندازه گرفت؟ به دو جمله خبری زیر که در یک روز گرم تابستان ممکن است بشنوید، با دقت توجه کنید:

1) فردا آفتاب طلوع خواهد کرد.

2) فردا باران خواهد بارید.

با جمله اول هیچ چیز تازه‌ای به کسی نشان نمی‌دهیم زیرا بیان امری است که حتماً به وقوع می‌پیوندد و گوینده این حرف هیچ چیزی به معلومات ما اضافه نکرده است. اما جمله دوم به ما مقداری اطلاعات ارائه می‌دهد، زیرا بارش باران یک امر حتمی‌الوقوع نیست. فردا ممکن است باران ببارد و یا نبارد و گوینده این حرف، با جمله خبری خود از عدم تعیین‌پذیری این امر می‌کاهد، بنابراین به ما مقداری اطلاع می‌دهد. اگر این جمله خبری در زمستان گفته شود مقدار کمتری اطلاع در بر دارد تا زمانی که در وسط تابستان و گرمای تیر ماه گفته شود. بنابراین می‌توان دریافت که میزان اطلاعاتی که در یک جمله خبری وجود دارد با نسبت احتمال وقوع آن کاهش می‌یابد. حال می‌خواهیم این دریافت شهودی را به صورت یک رابطه دقیق کمّی دریابیم.

فرض کنید که X یک متغیر تصادفی است که مقادیر $\{x_1, x_2, \dots, x_n\}$ را با احتمالات $\{p_1, p_2, \dots, p_n\}$ اختیار می‌کند. اگر کسی بگوید که در آزمایش مربوطه نتیجه x_i رخ داده است، مقدار اطلاعی را که داده است با h_i نشان می‌دهیم. بنابر آنچه که گفته شد h_i تابعی از خود x_i نیست بلکه تابعی است از احتمال وقوع آن یعنی p_i. بنابراین قرار می‌دهیم:

$$h_i = h(p_i)$$

هدف ما آن است که شکل تابع h را پیدا کنیم. می‌دانیم که h می‌بایست یک تابع نزولی باشد. اما هر تابع نزولی را

را باید انتخاب کنیم. برای آنکه فرم تابع h را بیابیم به یک نکته توجه می‌کنیم. فرض کنید به شخص مورد نظر اطلاعاتی در مورد دو متغیر تصادفی مستقل به ما ارائه کنند. دو متغیر تصادفی مستقل X و Y را در نظر بگیرید که مقادیر $\{x_i\}$ و $\{y_j\}$ را با احتمالات p_i و q_j اختیار کنند.

در این صورت آگاهی شخص، با توجه به اینکه در آزمایش مربوط به این دو متغیر، مقدار (x_i, y_j) رخ داده است، افزایش یافته و اطلاعات ارائه شده توسط وی مجموع اطلاعات ارائه شده برای وقوع x_i و y_j باشد. یعنی

$$h_{ij} = h_i + h_j$$

یا

$$h(p_i q_j) = h(p_i) + h(q_j)$$

رابطه فوق، تابع h را تعیین می‌کند ←

$$h(p) = \alpha \log p$$

که در آن α یک عدد منفی است، زیرا h باید بر حسب p نزولی باشد. تعیین مقدار α در واقع واحد سنجش اطلاعات را معین خواهد کرد. برای این کار به ساده‌ترین متغیر تصادفی، یعنی متغیر $X = \{0,1\}$ با $P_0 = P_1 = 1/2$ توجه کنیم.

همچنین آزمایش قرار می‌دهیم که کسی که می‌گوید مقدار 0 یا 1 اتفاق افتاده است یک واحد اطلاعات به ما داده است. در نتیجه ←

$$1 = \alpha \log 1/2$$

که معنیش این است که اگر log را در پایه ۲ بگیریم، α برابر ۱- خواهد شد. ← $h(p) = -\log_2 p.$

واحد اطلاعات را بیت bit می‌نامیم.

از یک زاویه دیگر هم می‌توانیم به این مسئله نگاه کنیم.

تا زمانی که در مورد یک نتیجه مشخص مثل x_i حرف نزنیم. اگر بخواهیم میزان اطلاعاتی که این شخص با پیشگویی خود از نتیجه آزمایش به ما می‌دهد تعیین کنیم، چه باید بکنیم؟ نتیجه x_i با احتمال p_i رخ می‌دهد و میزان اطلاع دریافت شده برابر است با $h(p_i)$. بنابراین

بطور خلاصه اطلاعات به طور میانگین برابر است با: $H = \sum_i p_i h(p_i) = -\sum_i p_i \log p_i$

معمولاً H را بصورت H(X) می‌نویسیم که تاکید کنیم H تابعی از احتمالات مربوط به متغیر تصادفی X است.

$$H(X) = -\sum_i p_i \log p_i$$

H تابع آنتروپی شانون Shannon خوانده می‌شود.

از یک زاویه دیگر نیز می‌توان به بحث بالا نگاه کرد. H(X) میزان عدم تعیینی را نشان می‌دهد که ما قبل از انجام آزمایش و حصول یک نتیجه معین داریم. بعد از حصول نتیجه، عدم تعیین ما به صفر می‌رسد زیرا یک نتیجه معین حاصل می‌شود. بنابراین H(X) هم میزان عدم تعیین را در متغیر تصادفی X نشان می‌دهد و هم میزان اطلاعاتی را که، پس از انجام آزمایش و حصول نتیجه پیدا می‌کنیم. هر چه نتایج قابل حصول پراکنده‌تر (بی‌نظم‌تر) باشد H(X) یعنی عدم تعیین ما درباره نتایج حاصل شده بیشتر است. به ۲ مثال زیر نگاه کنید:

$p_i = 1 , \quad p_{j\neq i} = 0$

$H(X) = 0.$

$p_i = \frac{1}{N} \quad \forall i$

$H(X) = \log_2 N$

خواننده می‌تواند ثابت کند که H(X) مقادیر بین دو توزیع بالا (گذشته شده) اختیار می‌کند. در واقع به طور عمومی می‌توان نشان داد که:

$$0 \le H(X) \le \log_2 N .$$

H(X) میزان بی‌نظمی در پراکندگی x_i ها است. در درس آینده نشان خواهیم داد که H(X) در واقع تعریف میکروسکوپی آنتروپی است. برای یک سیستم ماکروسکوپی x_i ها در واقع میکرو حالت‌های ممکن برای یک سیستم‌اند و p_i احتمال آن است که سیستم در میکرو حالت x_i باشد. قبل از پرداختن به این مطلب می‌خواهیم یک بار دیگر نگاه دقیق‌تری به مفهوم H(X)

به عنوان اندازه اطلاعات کمینه. فرض کنید که یک آزمایش تصادفی با چهار نتیجه $\{A, B, C, D\}$ به طور مکرر انجام می‌شود. در یک طرف خط، جایی که آزمایش انجام می‌شود، شخصی به اسم آلیس قرار دارد و می‌خواهد نتایج آزمایش را در یک دنباله از ۰ و ۱ برای شخص دیگری به اسم باب که در جای دیگری قرار دارد فاش کند. آلیس و باب هر دو از توزیع احتمال نتایج آگاه هستند ولی نمی‌دانند که در هر بار آزمایش چه نتیجه‌ای حاصل می‌شود. دو حالت حدی را در نظر بگیرید:

الف) احتمالات به صورت زیر هستند: $P(A) = 1, \quad P(B) = P(B) = P(C) = 0.$

در این صورت رشته‌ای که حاصل می‌شود به شکل زیر است:

AAAA···AAA····

در این صورت آلیس هیچ رمز ۰ و ۱ لازم نیست به باب فاش کند. کافی است که حرف اول را فاش کند و باب گیرنده بداند که تا آخر رشته حرف A تکرار می‌شود.

ب) احتمالات به صورت زیر هستند: $P(A) = P(B) = P(C) = P(D) = \frac{1}{4}$.

در این حالت، همه نتایج با احتمال یکسان رخ می‌دهند و رشته نتایج مورد نظر به شکل زیر است:

ABDDCABCADCBAABBCCDDABCDA····

در این صورت آلیس می‌بایست همه حروف را یک به یک فاش کند و برای این کار می‌تواند کدگذاری زیر را به کار ببرد:

$A \rightarrow 00 \qquad B \rightarrow 01 \qquad C \rightarrow 10 \qquad D \rightarrow 11$

بنابراین به ازای هر حرف ۲ بیت (اصطلاحی که برای ۰ و ۱ به کار می‌بریم) لازم است.

در مورد رمزدهی بالا تعداد بیت برای حرف که به کار برده می‌شود برابر است با $H(X)$.

حال حالت دیگری را در نظر بگیرید: $P(A)=\frac{1}{2} \quad P(B)=\frac{1}{4} \quad P(C)=\frac{1}{8} \quad P(D)=\frac{1}{8}$.

در این حالت هم تعداد بیت برای هر حرف می‌بایست توسط آلیس به کار برده شود. آلیس می‌تواند که کدهای زیر را به کار ببرد:

$$A \longrightarrow 0 \qquad B \longrightarrow 10 \qquad C \longrightarrow 110 \qquad D \longrightarrow 111$$

$n =$ تعداد متوسط بیت‌هایی که برای مخابره اطلاعات به کار می‌رود به ازای هر حرف $= 1\times P_A + 2\times P_B + 3\times P_C + 3\times P_D$

$$= 1\times\frac{1}{2} + 2\times\frac{1}{4} + 3\times\frac{1}{8} + 3\times\frac{1}{8}$$

$$= \frac{7}{4}.$$

بنابراین در این حالت می‌توان با کمتر از ۲ بیت به ازای هر حرف اطلاعات موجود در رشته را مخابره کرد. نکته جالب و مهم این است که این عدد در واقع همان $H(X)$ است. در واقع

$$H(X) = -\sum_i p_i \log_2 p_i = -\left\{\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{8}\log\frac{1}{8} + \frac{1}{8}\log\frac{1}{8}\right\}$$

$$= \frac{7}{4}.$$

آیا این نتیجه در حالت کلی نیز صادق است؟ پاسخ این سوال مثبت است. فرض کنید احتمالات به شکل زیر باشند:

$$P(A)=p_1, \quad P(B)=p_2, \quad P(C)=p_3, \quad P(D)=p_4.$$

در این صورت رشته‌ای طولانی از نتایج را در نظر بگیرید:

$$X_{i_1}X_{i_2}X_{i_3}X_{i_4}X_{i_5}\cdots\cdots X_{i_M}$$

این رشته را به بلوک‌هایی به طول N که در آن N عدد بزرگی است تقسیم می‌کنیم.

$$\underbrace{X_{i_1}X_{i_2}X_{i_3}X_{i_4}X_{i_5}\cdots X_{i_N}}_{N}\cdots\cdots X_{i_M}$$

هر رشته به طول N، می‌تواند به 4^N طریق تشکیل شود، از $AA\cdots A$ گرفته تا $DD\cdots D$.

احتمال هر بلوک N تایی که n_1 تا A، n_2 تا B، n_3 تا C و n_4 تا D دارد برابر است با:

$$P = P_1^{n_1} P_2^{n_2} P_3^{n_3} P_4^{n_4}$$

برای چنین رشته‌ای کد دودویی‌ای با طول که طول آن برابر است با: $-\log P$ ، (در کد حالت ب) یعنی

$$l = -\log P = -(n_1 \log P_1 + n_2 \log P_2 + n_3 \log P_3 + n_4 \log P_4).$$

طول متوسط رشته‌ی بیت‌ها چقدر است؟ ←

$$\bar{l} = \sum P\, l\, \frac{N!}{n_1!\, n_2!\, n_3!\, n_4!}$$

عدد $\frac{N!}{n_1!\, n_2! \dots n_4!}$ تعداد(تمام) رشته‌هایی است که n_1 تا A، n_2 تا B، .... دارند. بنابراین ←

$$\bar{l} = \sum_{n_1,n_2,n_3,n_4} P_1^{n_1} P_2^{n_2} P_3^{n_3} P_4^{n_4} \left(-n_1 \log P_1 - n_2 \log P_2 - n_3 \log P_3 - n_4 \log P_4\right) \frac{N!}{n_1! \dots n_4!}$$

با یک محاسبه ساده که تمرین برای شما است انجام دهید ←

$$\bar{l} = -N\left(P_1 \log P_1 + P_2 \log P_2 + \dots + P_4 \log P_4\right) = N H(X).$$

بنابراین $\frac{\bar{l}}{N} = H(X)$ و $H(X)$ دقیقاً تعداد بیت‌هایی است که برای فشرده‌سازی اطلاعات هر حرف در این رشته احتیاج داریم.

بنابراین $H(X)$ به معنای بسیار دقیق نشان دهنده اطلاعات مربوط به یک آزمایش است، خواه این آزمایش نتایج یک آزمایش تصادفی باشد خواه آزمایش یک حرفی باشد که یک متن را تشکیل داده است. این کمیت نشان دهنده عدم تعیینی است که نسبت به نتایج ممکن داریم و در عین حال نشان دهنده به دست آوردن اطلاعات بعد از انجام آزمایش است.