مقدمه ای بر نظریه احتمالات

وحیدکریمی پور_ دانشکده فیزیک _ دانشگاه صنعتی شریف

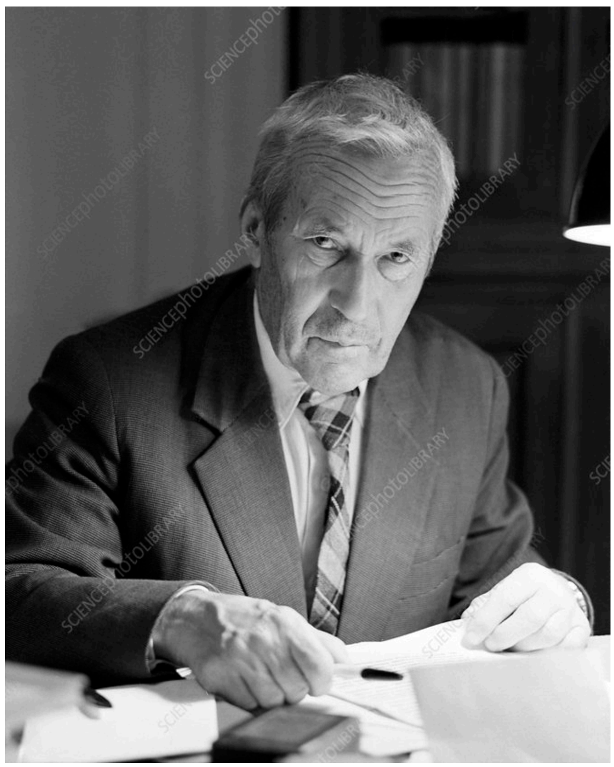
۲۷ مهر ۱۳۹۸

۱ مقدمه

مکانیک نیوتنی، مکانیک کوانتومی و الکترومغناطیس می توانند معادلات حاکم بر حرکت یک ذره را به طور دقیق تعیین کنند. امروزه اگر هم به صورت تحلیلی نتوانیم این معادلات را حل کنیم می توانیم با استفاده از رایانه و با روش های عددی و شبیه سازی این معادلات را حل کنیم. توانایی ما برای حل دقیق این معادلات وقتی که تعداد ذرات زیاد می شود به سرعت پایین می آید و برای تعداد خیلی زیادی از ذرات به سمت صفر میل می کند. ما نمی توانیم بگوییم که هر کدام از مولکولهای نیتروژن و اکسیژنی که در هوای اتاق هستند کجا هستند و با چه سرعتی حرکت می کنند و قرار است یک لحظه بعد کجا باشند. نمی توانیم بگوییم که هرکدام از مولکولهای درون یک لیوان آب کجا هستند، در چه جهتی قرار گرفته اند و با چه سرعتی دور خود می چرخندو هرکدام در هر لحظه با چند مولکول دیگر آب تماس دارند. نمی دانیم یک هسته رادیو اکتیو آیا در این لحظه از خود یک ذره آلفا صادر می کند یا یک لحظه بعد؟ نمی دانیم که در یک ساعت آینده دقیقا چند تا ذره آلفا از خود صادر می کند؟ با این وجود می خواهیم بدانیم که مولکول های آب و هوا رفتار میانگین شان چقدر است؟ چقدر سرعت دارند؟ چقدر به هم نزدیک می شوند؟ چقدر به دیواره ها فشار وارد می کنند؟ می خواهیم بدانیم که یک اتم رادیواکتیو به طور متوسط در یک ساعت چند ذره آلفا از خود تشعشع می کند؟ پاسخ این سوالها را با ترکیب قوانین اصلی فیزیک و قوانین آمار و احتمال می توانیم بیابیم. این علم جدید نامش مکانیک آماری است اگر چه روش هایی که در این علم یاد می گیریم تنها به درد مسایل فیزیک نمی خورد بلکه در طیف بزرگی از مسایل مختلف، از جمعیت شناسی و اقتصاد و بیمه گرفته تاْ ژنتیک و رشد بیماریها می توانیم از آنچه که در این علم جدید یاد می گیریم استفاده کنیم. در این درس با مقدمات آمار و

احتمال آشنا می شویم و در درس بعد نیز فرایندهای کاتوره ای یا تصادفی را یاد می گیریم. آنچه که در این دو درس یاد می گیریم برای شاخه های بسیار متنوعی از دانش های مختلف مفید خواهند بود.

۲ اصول نظریه احتمال



شکل ۱: آندره کولموگروف، ریاضیدان روس که نظریه احتمالات را بر مبنای اصول موضوع محکم بنا کرد و پیشرفت های زیادی نصیب این رشته کرد.

اصول موضوع نظریه احتمالات نخستین بار توسط اندره کولموگوروف [۱] ریاضیدان روس در سال **۱۹۳۳** در کتاب وی با عنوان «مبانی نظریه احتمال» [۲] به شکلی که امروزه می شناسیم تدوین شد. به طور ساده می توانیم این اصول را به صورت زیر تعریف کنیم. یک مجموعه دلخواه S

[۱] Andrei Kolmogorov

[۲] Foundations of Probability Ttheory

که به آن فضای نمونه [۳] و یک تابع f به صورت زیر،

$$P : S \longrightarrow R^{+} \tag{۱}$$

یک تابع احتمال تعریف می کنند هرگاه شرایط زیر برقرار باشند:

$$\begin{aligned} i)\ P(\phi) &= 0 \\ ii)\ P(S) &= 1 \\ iii)\ P(A\cup B) &= P(A)+P(B)-P(A\cap B) \qquad \forall\ A,\ B\subset S. \end{aligned} \tag{۲}$$

به هر زیرمجموعه از S یک رویداد [۴] می گوییم. بنابراین $A\cup B$ رویدادی است که یا A اتفاق می افتد یا B و $A\cap B$ رویدادی است که هم A اتفاق می افتد و هم B. به عنوان مثال اگر $S=\{1,2,3,4,5,6\}$ باشد، آنگاه $A=\{1,2,3,4\}$ رویدادی است که در آن یک شماره کمتر از 5 اتفاق می افتد و $B=\{2,4,6\}$ رویدادی است که در آن یک شماره زوج اتفاق می افتد. در این صورت $A\cup B=\{1,2,3,4,6\}$ رویدادی است که یا یک عدد کمتر از ۵ رخ دهد یا اینکه یک عدد زوج رخ دهد و $A\cap B=\{2,4\}$ رویدادی است که یک عدد زوج کمتر از ۵ اتفاق می افتد.

دو رویداد A و B را مجزا [۵] می خوانیم اگر داشته باشیم:

$$A\cap B=\phi \tag{۳}$$

برای چنین رویدادهایی داریم:

$$P(A\cup B)=P(A)+P(B). \tag{۴}$$

رویدادهای مجزا فقط با توجه به خصوصیات خود مجموعه ها مشخص می شوند و ربطی به تابع $P : S \longrightarrow R_{+}$ ندارند.

- مثال: دو رویداد $A=\{1,2,3\}$ و $B=\{4,5,6\}$ دو رویداد مجزا هستند. این دو رویداد با هر تابع احتمالی مجزا هستند.

رویدادهای مجزا البته با رویدادهای مستقل [۶] فرق دارند. دو رویداد A و B مستقل نامیده می شوند، اگر داشته باشیم:

$$P(A\cap B)=P(A)P(B). \tag{۵}$$

[۳] Sample Space
[۴] Event
[۵] Disjoint
[۶] Independent

بنابراین رویدادهای مستقل با توجه به تابع احتمال تعریف می شوند و نه فقط از روی خواص خود مجموعه ها.

احتمال شرطی

احتمال این که A رخ دهد مشروط بر اینکه B رخ داده باشد را با $P(A|B)$ نشان می دهیم. این احتمال را به صورت زیر تعریف می کنیم:

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \tag{۶}$$

دلیل شهودی این تعریف را نیز می توانیم با بازنویسی آن به صورت زیر بفهمیم:

$$P(A \cap B) = P(A|B)P(B). \tag{۷}$$

این رابطه در واقع بیان می کند که احتمال اینکه هم A روی دهد و هم B برابر است با احتمال اینکه نخست B روی دهد و سپس A به شرط B رخ دهد. با توجه به این رابطه برای دو رویداد مستقل خواهیم داشت:

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A) \tag{۸}$$

که همان چیزی است که انتظار داریم، به این معنا که رخ دادن B هیچ تاثیری در احتمال رخ دادن A ندارد. یکی دیگر از نتایج این رابطه هم این است که

$$P(S|B) = \frac{P(S \cup B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \tag{۹}$$

که معنای شهودی آن واضح است. هم چنین داریم

$$P(B|S) = \frac{P(B \cup S)}{P(S)} = \frac{P(B)}{P(S)} = \frac{P(B)}{1} = P(B). \tag{۱۰}$$

که بازهم معنای شهودی روشنی دارد. و بالاخره متمم یک رویداد A رویداد $\overline{A} = S - A$ است که به معنای این است گه هرگاه A روی دهد آنگاه $\overline{A}$ روی نمی دهد.

$$P(\overline{A}) = 1 - P(A) \tag{۱۱}$$

با چنین تعریفی می توانیم با استفاده از قوانین مربوط به مجموعه ها معنای اتحادهای زیر را بفهمیم:

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B} \tag{۱۲}$$

نه A رخ دهد و نه B.

$$\overline{(A \cap B)} = \overline{A} \cup \overline{B} \tag{۱۳}$$

هردو رویداد A و B باهم رخ ندهند (لااقل یکی از آنها رخ ندهد.)

تا کنون فقط خواص تابع احتمال را گفته ایم. اما نگفته ایم که این تابع بر مبنای چه ضابطه ای مشخص می شود. این که این تابع چگونه مشخص می شود فقط با تجربه مشخص می شود. رایج ترین راه این است که یک آزمایش را به تعداد بسیار زیادی تکرار کنیم و ببینیم که در چه درصدی از این آزمایش ها پیشامد A رخ داده است. به این ترتیب می توانیم بنویسیم:

$$P(A) := \lim_{N\longrightarrow\infty} \frac{N_A}{N}, \tag{۱۴}$$

که در آن N تعداد کل آزمایش ها و N_A تعداد کل دفعاتی است که پیشامد A رخ داده است. با توجه به حد مورد نظر که بی نهایت است مسلم است که هیچگاه نمی توان از نظر تجربی به مقدار دقیق تابع احتمال پی برد. در عمل آنچه که انجام می شود این است که در غیاب هر نوع ترجیحی فرض می کنیم که تمام اعضای یک زیرمجموعه (یا پیشامدِ) A هم احتمال اند و از هم مستقل اند. به این ترتیب تابع احتمال را به صورت زیر تعریف می کنیم:

$$P(A) := \frac{|A|}{|S|} \tag{۱۵}$$

که در آن $|A|$ و $|S|$ به ترتیب اندازه های زیر مجموعه A و مجموعه ی S هستند. به این ترتیب پیش نیاز تعریف تابع احتمال آن است که بتوانیم سازوکار مناسبی برای شمارش تعداد اعضای یک مجموعه پیدا کنیم. این کار در شاخه ای از ریاضیات انجام می شود که به آن ترکیبیات [۷] می گوییم. در زیر بخش بعد اشاره خیلی کوتاهی به این موضوع خواهیم کرد و تاکید هم می کنیم که برای بسیاری از مسایل احتمال که در مکانیک آماری به آنها می پردازیم، نیاز جدی به روش های پیشرفته ترکیبیاتی نداریم.

■ مثال: دو رویداد مجزای $A = \{1, 2, 3\}$ و $B = \{4, 5, 6\}$ را در نظر می گیریم. برای تابع احتمال

$$P : \{1, 2, 3, 4, 5, 6\} \longrightarrow \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\} \tag{۱۶}$$

[۷] Combinatorics

این دو رویداد مستقل نیستند زیرا

$$P(A\cap B)=P(\phi)=0, \qquad P(A)=\frac{1}{2}, \qquad P(B)=\frac{1}{2}. \tag{۱۷}$$

یعنی اینکه رابطه رابطه (۵) برآی آنها برقرار نیست. از نظر شهودی هم معلوم است که این دو اتفاق چرا مستقل نیستند چون که اگر A اتفاق افتاده باشد مطمئن خواهیم شد که B اتفاق نیفتاده است.

اما همان دو رویداد با تابع احتمالِ

$$P:\{1,2,3,4,5,6\}\longrightarrow\{0,0,0,\frac{1}{3},\frac{1}{3},\frac{1}{3}\} \tag{۱۸}$$

مستقل هستند، چون رابطه (۵) برای آنها برقرار است، دلیل شهودی اش هم این است که یکی از این رویدادها یعنی A اصلا اتفاق نمی افتد.

۳ بعضی قوانین ساده در ترکیبیات

مسایل احتمالات عموما با شمارش راه های ممکن چینش یک دسته اشیای (حقیقی یا مجازی) با توجه به قیدهای مشخص سروکار دارند. بنابراین دانستن بعضی از قوانین شمارش مفید است. کار خود را با یک شمارش ساده شروع می کنیم . سپس سعی می کنیم به تدریج مسایل کمی سخت تر را نیز حل کنیم و مهارت خود را در شمارش و در ترکیبیات بالا ببریم .

- مثال یک: تعداد N گوی داریم که رنگ همه آنها با هم فرق دارد. به چند طریق می توانیم این گوی ها را در یک ردیف کنار هم بچینیم؟ پاسخ این سوال را همه می دانیم: برا ی قرار دادن اولین گوی N انتخاب داریم، برای قرار دادن دومین گوی $N-1$ انتخاب و نهایتا برای قرار دادن آخرین گوی نیز یک انتخاب داریم. بنابراین تعداد کل راههایی که در اختیار داریم عبارت است از:

$$Q=N(N-1)(N-2)\cdots 2.1=N! \tag{۱۹}$$

- مثال دو: حال فرض کنیم که گوی ها از دو رنگ قرمز و آبی تشکیل شده اند. تعداد گوهای قرمز N_1 و تعداد گوی های آبی N_2 تاست. به چند طریق می توانیم این گوی ها را در یک ردیف کنار هم بچینیم؟ برای پاسخ دادن به این سوال می پرسیم به چند طریق می توانیم N_1 خانه از N_1+N_2 خانه را انتخاب کنیم و در آنها گوی های قرمز را قرار دهیم و در بقیه خانه ها گوی های آبی را قرار دهیم؟ پاسخ اش ساده است. برای قرار دادن اولین گوی قرمز N_1+N_2 انتخاب، برای دومی N_1+N_2-1 انتخاب و برای سومین گوی N_1+N_2-2

شکل ۲: : دانیل برنولی ریاضیدان و فیزیکدان هلندی (۱۷۰۰ــ۱۷۸۲) که یکی از پیشگامان نظریه ترکیبیات بوده است.

انتخاب داریم و به همین ترتیب تا اخرین گوی قرمز. بقیه گوی ها ی آبی را در جاهای باقی مانده قرار می دهیم. از آنجا که گوی های آبی همه مثل هم و گوی ها قرمز نیز همه مثل هم هستند برای جلوگیری از تکرار می بایست عدد بدست آمده را بر تعداد جایگشت های توپ ها از هر نوع تقسیم کنیم. بنابراین تعداد کل راه ها برابر است با

$$Q = \binom{N_1 + N_2}{N_1} = \frac{(N_1 + N_2)!}{N_1! N_2!} \tag{۲۰}$$

■ مثال سه: حال فرض کنید که تعداد K رنگ توپ داریم. N_1 نوع رنگ اول (قرمز)، N_2 توپ رنگ دوم (آبی)، و به همین ترتیب N_k توپ از رنگ k ام (سبز) داریم. در این حالت تعداد کل حالت های چیدن این توپ ها برابر است با:

$$Q = \frac{(N_1 + N_2 + \cdots N_K)!}{N_1! N_2! \cdots N_k!} \tag{۲۱}$$

- مثال چهار: یک قفسه کتاب داریم که دارای N خانه است. فرض کنید که در هر خانه تعداد دلخواهی کتاب جای می گیرد. به چند طریق می توانیم M کتاب را که همه باهم متفاوتند در این قفسه بچینیم؟

پاسخ این سوال ساده است. برای اولین کتاب N انتخاب داریم. برای دومین کتاب نیز همین تعداد انتخاب داریم، برای سومین کتاب نیز همین تعداد انتخاب داریم و الی اخر. بنابراین تعداد انتخاب هایمان برابر است با:

$$Q = N^M. \tag{۲۲}$$

- مثال پنج: حال فرض کنید که کتاب ها همه باهم یکسان هستند. در این صورت به چند طریق می توان کتاب ها را در این قفسه چید؟ برای پاسخ دادن به این سوال احتیاج به یک ابتکار ساده داریم.

می توانیم به این صورت به این مسئله نگاه کنیم: تعداد $N+M-1$ تا گوی داریم و می خواهیم $M-1$ تا از آنها را با رنگ سفید رنگ بزنیم. به چند طریق می توانیم این کار را انجام دهیم: پاسخ ساده است:

$$Q = \binom{N+M-1}{M-1} = \frac{(N+M-1)!}{N!(M-1)!} \tag{۲۳}$$

۴ متغیرهای تصادفی

یکی از مهم ترین ابداعات در نظریه احتمال، ابداع مفهوم متغیر تصادفی است. متغیر تصادفی به هر عضو از یک فضای نمونه یک عدد نسبت می دهد. علی الاصول شیوه این نسبت دادن هیچ محدودیتی ندارد. بنابراین اگر متغیر تصادفی را با X نشان دهیم، داریم:

$$X : S \longrightarrow R \tag{۲۴}$$

کاری که متغیر تصادفی می کند این است که رویدادها را با عدد جایگزین می کند. این کار باعث می شود ما بتوانیم به صورت معنا داری در باره متوسط این رویدادها صحبت کنیم که قبل از معرفی متغیر تصادفی نمی توانستیم. به عنوان مثلا سکه ای را در نظر بگیرید که دو روی شیر و خط دارد که معمولا در زبان انگلیسی با H و T [۸] نشان داده می شوند و هر کدام با احتمال $1/2$ رخ می دهند. هرگاه متغیر تصادفی زیر را تعریف کنیم و به شیر و خط نسبت دهیم:

$$Y : \{H, T\} \longrightarrow \{-1, 1\} \tag{۲۵}$$

[۸] Head and Tail

می توانیم از مقدار متوسط این مقدار تصادفی حرف بزنیم و بگوییم که مقدار متوسط آن برابر با صفر است و بنویسیم:

$$\langle Y\rangle = \sum_i p_i Y_i = 0. \tag{۲۶}$$

قبل از معرفی این متغیر تصادفی نمی توانستیم از متوسط شیر و خط صبحت کنیم اما معرفی این متغیر تصادفی به ما اجازه داده که به راحتی از چنین متوسطی حرف بزنیم و آن را حساب کنیم. به این ترتیب از این به بعد همواره از یک متغیر تصادفی حرف می زنیم که مقادیرش را از اعداد حقیقی انتخاب می کند. خود متغیر تصادفی یا مجموعه مقادیرش را با حروفی مثل X نشان می دهیم و مقادیری که این متغیر تصادفی می گیرد را با حروف کوچک نشان می دهیم. بنابراین داریم:

$$S=\{s_1,s_2,\cdots s_N\}\longrightarrow X=\{x_1,x_2,\cdots x_N\} \tag{۲۷}$$

در نتیجه تابع احتمال از به سوی اعداد نامنفی تعریف می شود

$$P: X\longrightarrow R^+. \tag{۲۸}$$

وقتی که متغیر تصادفی گسسته است داریم

$$P(x_i)\geq 0 \qquad \sum_i P(x_i)=1. \tag{۲۹}$$

و وقتی که این متغیر پیوسته است داریم

$$P(x)\geq 0 \qquad \int dx P(x)=1. \tag{۳۰}$$

در این جا معنای $P(x)dx$ احتمال این است که متغیر تصادفی مقدارش را در فاصله x تا $x+dx$ اختیار کند. بنابراین داریم

$$P(a\leq x\leq b)=\int_a^b dx P(x) \tag{۳۱}$$

- مثال ۱: یک هسته رادیواکتیو را در نظر بگیرید. این هسته اولین ذره ی α را در زمان T از خود ساطع می کند. در این جا T یک متغیر تصادفی است که مقادیرش را در بازه ی $[0,\ \infty)$ اختیار می کند. احتمال این که اولین ذره در زمان بین t و $t+dt$ ساطع شود را برابر با $P(t)dt$ می گیریم.

- مثال ۲: همین هسته رادیواکتیو را در نظر بگیرید که به مدت یکساعت مشاهده می شود. تعداد ذرات آلفایی که از خود ساطع می کند یک متغیر تصادفی K است که مقادیرش را از مجموعه ی گسسته $S_K=\{0,\ 1,\ 2,\ 3,\cdots\}$ اختیار می کند.

- مثال ۳: یک مجموعه خرگوش و روباه در یک مزرعه زندگی و زاد و ولد می کنند. روباه ها خرگوش ها را شکار می کنند و هر دو نیز در معرض مرگ طبیعی هستند. اگر تعداد خرگوش ها را در هر سال با $N_R(t)$ و تعداد روباه ها را در هر لحظه با $N_F(t)$ نشان دهیم آنگاه $N_R(t)$ و $N_X(t)$ متغیرهای تصادفی گسسته ای هستند که به صورت پیوسته با زمان تغییر می کنند. تغییر زمانی این متغیرهای تصادفی یک فرایند تصادفی یا استوکاستیک نامیده می شود. [۹] .

- تابع توزیع احتمال: [۱۰]

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

$$F(x) = \int_{-\infty}^{x} P(x')dx'. \tag{۳۲}$$

واضح است که $P(x) := \frac{dF(x)}{dx}$. معنای $F(x)$ از روی رابطه ی بالا روشن است:

$$F(x) = Prob(X \leq x). \tag{۳۳}$$

۱.۴ گشتاورهای یک متغیر تصادفی

یکی از مهمترین مشخصه های یک متغیرتصادفی و یک تابع توزیع گشتاور های [۱۱] آن است. ساده ترین گشتاور همان متوسط یک کمیت است که به صورت زیر تعریف می شود.

$$\langle X\rangle = \sum_i x_i P(x_i) \qquad \langle X\rangle = \int dx x P(x) \tag{۳۴}$$

معلوم است که توابع توزیع مختلف ممکن است یک متوسط داشته بشاند. به عنوان مثال هر تابع توزیعی که فرد باشد یعنی در رابطه $P(x) = -P(-x)$ صدق کند، متوسط اش صفر است. اما این توابع می توانند در گشتاورهای مرتبه بالاتر با هم فرق داشته باشند. این گشتاورها به صورت زر تعرف می شوند:

$$\langle X^k\rangle := \sum_i x_i^k P(x_i) \qquad \langle X^k\rangle := \int dx x^k P(x) \tag{۳۵}$$

---

[۹] Stochastic Process

[۱۰] Probability Distribution Function

[۱۱] Moments

از جمله مهم ترین گشتاورها گشتاور دوم است. این گشتاور کمک می کند که بفهمیم متغیر تصادفی چقدر افت و خیز حول مقدار میانگین دارد.

$$\langle X^2\rangle = \sum_i {x_i}^2 P(x_i) \qquad \langle X^2\rangle = \int dx x^2 P(x) \tag{۳۶}$$

برای محاسبه افت و خیز کمیت زیر را حساب می کنیم

$$\sigma_X := \sqrt{\langle X^2\rangle - \langle X\rangle^2} \tag{۳۷}$$

که با یک محاسبه ساده معلوم می شود برابر با کمیت زیر است:

$${\sigma_X}^2 = \langle (X - \langle X\rangle)^2\rangle \tag{۳۸}$$

عبارت اخیر که آن را واریانس [۱۲] نیز می نامند، نشان می دهد که چرا این کمیت نشان دهنده مقدار افت و خیز است. هر چه که σ_X بزرگ تر باشد به این معناست که متغیرتصادفی مقادیر دورتری را از مقدار متوسط خود اختیار کرده و بنابراین دچار افت و خیز بیشتری شده است.

۲.۴ تابع مولد مربوط به یک متغیر تصادفی

آیا باید گشتاورهای یک متغیر تصادفی را یک به یک حساب کرد یا اینکه راه ساده تری برای محاسبه همه آنها به صورت یک جا وجود دارد؟ خوشبختانه پاسخ این سوال مثبت است. برای چنین کاری کمیتی را محاسبه می کنیم که به آن تابع مولد [۱۳] مربوط به آن متغیر تصادفی می گوییم:

$$Z(t) := \sum_j e^{itx_j} P(x_j). \tag{۳۹}$$

اگر این تابع را محاسبه کنیم با مشتق گیری متوالی از ان می توانیم گشتاورهای متغیر تصادفی را یک به یک محاسبه کنیم. از همین رابطه بالا معلوم است که روابط زیر برقرارهستند:

$$Z(0) = 1 \qquad Z'(0) = i\langle X\rangle \qquad Z^{"}(0) = (i)^2\langle X^2\rangle, \tag{۴۰}$$

و به طور کلی

$$Z^(k)(0) = i^k\langle X^k\rangle. \tag{۴۱}$$

[۱۲] Variance

[۱۳] Generating Function

در بخش های بعدی با چند نمونه از توابع توزیع مشهور و مولدهای آنها آشنا خواهیم شد. اما قبل از آن باید به یک نکته اشاره کنیم و آن اینکه به لحاظ صوری می توانیم توابع توزیع گسسته و پیوسته را همگی در یک چارچوب واحد مطالعه کنیم. در واقع یک تابع توزیع گسسته را می توان به شکل زیر و با استفاده از تابع دلتای دیراک به یک تابع توزیع پیوسته تبدیل کرد:

$$P_X(x) = \sum_i P(x_i)\delta(x - x_i) \tag{۴۲}$$

حال که همه توابع توزیع را می توانیم تابع توزیع پیوسته در نظر بگیریم می توانیم تابع مولد را به شکل زیر تعریف کنیم و همین رابطه نیز نشان می دهد که چگونه تعریف مان از تابع مولد برای متغیرهای تصادفی گسسته به همان تعریف قبلی تبدیل می شود.

$$Z(t) := \int dx\, e^{itx} P(x) = \int dx e^{itx} \sum_j P(x_j)\delta(x - x_j) = \sum_j e^{itx_j} P(x_j) \tag{۴۳}$$

خیلی از خواص یک متغیرتصادفی به تابع توزیع آن بستگی دارد و بدون دانستن تابع توزیع نمی توان حرف زیادی در مورد رفتار متغیرتصادفی گفت. در این میان بعضی از قضایا وجود دارند که برای هر متغیر تصادفی مستقل از شکل تابع توزیع آنها برقرارند. این قضایا به دلیل کلیت شان اهمیت دارند. دو قضیه مهم از این نوع قضایایی هستند که نخستین بار پافنوتی چبیشف [۱۴] آن را طرح و ثابت کرده است. این دو قضیه را در بخش بعدی معرفی می کنیم.

۳.۴ دو قضیه مهم از چبیشف

برای ادامه بحث خود احتیاج به دو لم خیلی ساده در نظریه احتمال داریم. این لم ها دامنه کاربرد خیلی وسیعی دارند و یادگیری آنها اهمیت دارد.

- لم اول : نامساوی اول چبیشف ($Chebyshev\ inequlity$):

الف : فرض کنید که متغیرتصادفی X مقادیر مثبت $\{x_1, x_2, \cdots x_N\}$ را بااحتمالات $\{p_1, p_2, \cdots p_N\}$ اختیارمی کند . دراین صورت به ازای هرعدد مثبت α،

$$P(X \geq \alpha) \leq \frac{\overline{X}}{\alpha} \tag{۴۴}$$

که درآن $\overline{X}$ متوسط متغیر تصادفی X است.

[۱۴] Pafnuty Chebyshev

■ اثبات :

$$P(X \geq \alpha) = \sum_{x=\alpha}^{\infty} P(x) \leq \sum_{x=\alpha}^{\infty} \frac{x}{\alpha} P(x) \leq \frac{\overline{X}}{\alpha}. \tag{۴۵}$$

■ لم دوم : نامساوی دوم چبیشف ($Chebyshev\ inequlity$):

حال فرض کنید که متغیرتصادفی X مقادیر دلخواه مثبت یامنفی اختیارمی کند. دراین صورت به ازای هرعدد k

$$P((X - \overline{X})^2 \geq k^2\sigma_x^2) \leq \frac{1}{k^2}. \tag{۴۶}$$

اثبات : متغیر تصادفی $T = (X - \overline{X})^2$ را درنظرمی گیریم. این متغیرفقط مقادیر مثبت را اختیار می کند. ضمناً می دانیم که $\overline{T} = \sigma_x^2$.
از قسمت الف داریم:

$$P(T \geq \alpha) \leq \frac{\overline{T}}{\alpha}. \tag{۴۷}$$

هرگاه به جای α درنامساوی اخیر قراردهیم $k^2\sigma_x^2$ بدست می آوریم:

$$P((X - \overline{X})^2 \geq k^2\sigma_x^2) \leq \frac{\sigma_x^2}{k^2\sigma_x^2} = \frac{1}{k^2}. \tag{۴۸}$$

این نامساوی را به شکل زیر نیز می توان نوشت:

$$P(|X - \overline{X}| \geq k\sigma_x) \leq \frac{1}{k^2}, \tag{۴۹}$$

۵ تابعی از یک متغیر تصادفی

فرض کنید که X یک متغیر تصادفی با تابع توزیع $P_X(x)$ باشد، و Y یک تابع مشخص از متغیر X باشد مثلا $Y = X^2 + X + 3$. از انجا که مقادیر X تصادفی هستند، مقادیر Y نیز تصادفی خواهند بود. از خود می پرسیم که تابع توزیع احتمالات Y چیست و چگونه بدست می آید؟ نخست شکل () را در نظر می گیریم: واضح است که احتمال این که Y در فاصله نشان داده شده باشد، برابر است با احتمال اینکه X در فاصله های مربوطه قرار گیرد. بنابراین

$$P_Y(y)dy = P_X(x)dx \qquad \longrightarrow \qquad P_Y(y) = P_X(x)\frac{dx}{dy} \tag{۵۰}$$

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

$$P_Y(y)|dy| = P_X(x)|dx|, \tag{۵۱}$$

که مواردی مثل شکل () را نیز در بر بگیرد. ولی رابطه بالا نیز کامل نیست زیرا ممکن است که هر مقدار از Y متناظر با چندین مقدار از X باشد، شکل (). بنابراین رابطه کامل به شکل زیر است:

$$P_Y(y)|dy| = \sum_i P_X(x_i)|dx_i| \tag{۵۲}$$

که در آن جمع روی تمام x_i هایی است که در رابطه ی $Y = f(X_i)$ صدق می کنند. این رابطه را می توان به شکل فشرده تر و گویاتری نیز نوشت. کافی است که به این خاصیت از تابع دلتای دیراک توجه کنیم:

$$\delta(y - f(x)) = \sum_i \delta(x - x_i) \mid \frac{dx}{dy} \mid_{x=x_i} \tag{۵۳}$$

که در آن جمع روی تمام x_i هایی است که در رابطه $y = f(x_i)$ صدق می کنند. بنابراین

$$P_Y(y) = \sum_i P_X(x_i) \mid \frac{dx}{dy} \mid_{x=x_i} = \int P_X(x)\delta(y - f(x))dx \tag{۵۴}$$

که مثل این است که از تابع $\delta(y - f(x))$ متوسط گرفته ایم. بنابراین به طور کلی

$$P_Y(y) = \langle\delta(y - f(X))\rangle. \tag{۵۵}$$

۶ متغیرهای تصادفی چندتایی

فرض کنید که دو مهره تاس را باهم روی زمین بغلتانیم. احتمال اینکه مهره اول عدد ۴ و مهره دوم عدد ۳ را نشان دهد چقدر است؟ فرض کنید که به طور تصادفی کارنامه یک دانش آموز را نگاه کنیم. احتمال اینکه نمره علوم او بین ۱۵ و ۱۶ و نمره ریاضی او بین ۱۰ و ۱۲ باشد چقدر است؟ این ها نمونه هایی از متغیرهای تصادفی دوگانه هستند. یک مولکول از یک گاز را در نظر بگیرید. احتمال اینکه در یک لحظه معین مختصه های مکانی این ذره در بازه های $(x, x + dx)$ ، $(y, y + dy)$ و $(z, z + dz)$ باشد چقدر است؟ در اینجا با یک متغیر تصادفی سه گانه روبرو هستیم. به طور کلی می توان متغیرهای تصادفی چندگانه را تعریف کرد. برای سادگی خود را محدود به متغیرهای تصادفی دوگانه می کنیم ولی هر آنچه که می گوییم به سادگی تمام برای متغیرهای تصادفی چندگانه نیز برقرارند. تعمیم آنها خیلی سرراست و ساده است. نخست به معنای

تابع توزیع $P(x_i, y_j)$ توجه می کنیم: این تابع احتمال این را بیان می کند که متغیر تصادفی X مقدار x_i و متغیر تصادفی Y مقدار y_j را اختیار کند. اگر خیلی بخواهیم دقت و وسواس به خرج دهیم باید این تابع را به شکل $P_{X,Y}(x_i, y_j)$ بنویسیم. اما این نوع نوشتن ممکن است هر نوع ابهامی را رفع کند ولی به زحمت اش نمی ارزد. در عوض به خواننده اعتماد می کنیم که نوع متغیرهای تصادفی X, Y را در زمینه همه روابط می بیند و فقط وقتی ضرورت داشته باشد نوع متغیرهای تصادفی را می نویسیم. برای متغیرهای تصادفی طبیعتا رابطه بهنجارش به صورت زیر است:

$$\int dx \int dy P(x, y) = 1 \qquad (۵۶)$$

هرگاه روی مقادیر یک متغیر انتگرال بگیریم، تابعی که بدست می آید یک تابع توزیع احتمال برای متغیر تصادفی دیگر است. بنابراین

$$P_X(x) = \int dy P(x, y) \qquad P_Y(y) = \int dx P(x, y). \qquad (۵۷)$$

گشتاورها نیز به شکل زیر تعریف می شوند.

$$\langle X^m Y^n \rangle = \int dx \int dy x^m y^n P(x, y) \qquad (۵۸)$$

یکی از کمیت های مهم کوواریانس [۱۵] است که نشان می دهد دو متغیر تصادفی چقدر به هم وابسته هستند.

$$Cov(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle. \qquad (۵۹)$$

در واقع اگر این کمیت برابر با صفر باشد، به این معناست که افت و خیزهای این دو متغیر ربطی به یکدیگر ندارند.

۱.۶ تابعی از چند متغیر تصادفی

تابع $P(x, y)$ توزیع احتمال این را می دهد که یک دارت پرتاب شده روی نقطه (x, y) قرار بگیرد. (مثل هر متغیر پیوسته ای منظور در این جا این است که دارت در یک بازه در اطراف این نقطه قرار بنشیند.) از ما پرسیده اند که توزیع احتمال اینکه دارت به فاصله r از مبدا بنشیند چقدر است؟ این سوال را چگونه باید پاسخ دهیم؟ مسلم است که در این جا با یک متغیر تصادفی جدید یعنی R روبرو هستیم که وابسته به متغیرهای تصادفی قبلی یعنی (X, Y) است. هدف ما پیدا کردن این تابع توزیع جدید یعنی $F_R(r)$ از روی تابع توزیع $F_{X,Y}(x, y)$ است. کمی که فکر کنیم جواب زیر به ذهنمان می رسد:احتمال اینکه R مقدار r را اختیار کند، برابر با مجموع تمام احتمال هایی است که (x, y) اختیار شوند مشروط

[۱۵] Covariance

بر آنکه شر ط $r=\sqrt{x^2+y^2}$ برقرار باشد. بنابراین می توانیم بنویسیم:

$$F_R(r)=\int dx\int dyF_{X,Y}(x,y)\delta(r-\sqrt{x^2+y^2}) \tag{۶۰}$$

که می توان آن را به صورت یک متوسط هم نوشت:

$$F_R(r)=\langle\delta(r-\sqrt{X^2+Y^2})\rangle_{X,Y}. \tag{۶۱}$$

هر گاه به جای این مثال با یک مسئله کلی تر روبرو باشیم که در آن متغیر تصادفی جدیدی مثل R تابعی از متغیرهای تصادفی قبلی به صورت $R=h(X,Y)$ باشد، تابع توزیع جدید به صورت زیر محاسبه می شود:

$$P_R(r)=\int dx\int dy\delta(r-h(x,y))P(x,y)=\langle\delta(r-h(X,Y))\rangle_{X,Y}. \tag{۶۲}$$

۷ چند نمونه از توابع توزیع احتمال

یک تابع توزیع احتمال هیچ قیدی ندارد جز این که مقادیرش مثب باشد و جمع آنها نیز برابر با یک باشد. به عبارت بهتر هر تابعی که در شرطهای

$$P(x)\geq 0 \qquad \int dxP(x)=1 \tag{۶۳}$$

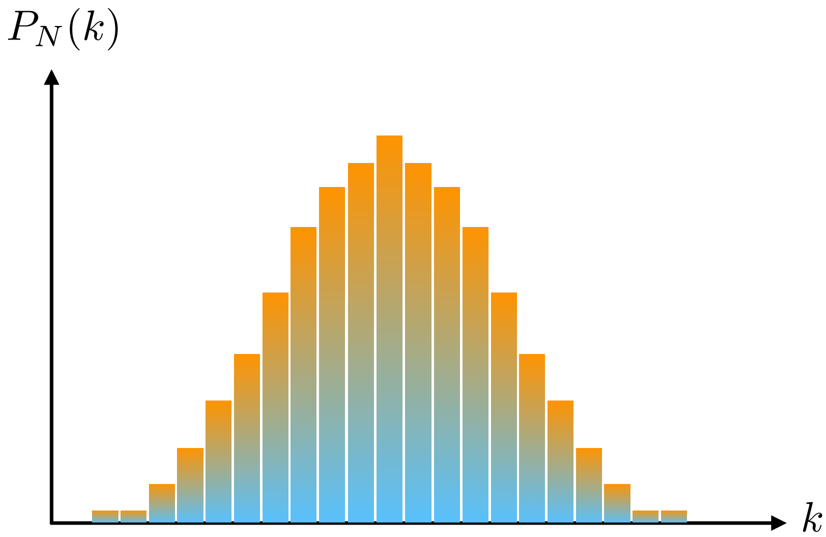
صدق کند یک تابع توزیع معتبر است. اما توابع توزیع مشهور آنهایی هستند که توصیف کننده یک فرایند واقعی تصادفی هستند و با ملاحظات شهودی تعریف می شوند. در این بخش چند تا از توابع توزیع شناخته شده را بررسی می کنیم.

۱.۷ تابع توزیع دو جمله ای

سکه ای را در نظر بگیرید که وقتی پرتاب می شود با احتمال p روی شیر و با احتمال q روی خط را نشان می دهد. یک سوال طبیعی و مهم این است: اگر سکه را N بار پرتاب کنیم، احتمال اینکه k بار روی شیر بیاید چقدر است؟ پاسخ این سوال با تابع توزیع دوجمله ای [۱۶] است.

$$P_N(k)=\binom{N}{k}p^k(1-p)^{N-k} \tag{۶۴}$$

[۱۶] Binomial Distribution



شکل ۳: تابع توزیع دوجمله ای

برای این توزیع تابع مولد برابر است با:

$$Z_K(t) = \sum_{k=0}^{\infty} e^{tk} \binom{n}{k} q^{N-k} = (p + e^t q)^N. \qquad (۶۵)$$

از این تابع مولد می توان کمیت های زیر را براحتی بدست آورد:

$$\langle K \rangle = Np, \qquad \sigma_K = \sqrt{Npq} \qquad (۶۶)$$

هرگاه به جای یک سکه چیزی داشته باشیم که بیش از دو حالت دلخواه مثلا سه حالت a, b, و c با احتمال های p, q و r از خود نشان دهد می توانیم بپرسیم در N بار آزمایش احتمال این سه پیشامد به ترتیب به تعداد K_a بار، K_b بار و K_c بار رخ بدهند چقدر است؟. پاسخ این سوال توزیع چندجمله ای است که برابر است با:

$$P_N(k_a, k_b, k_c) = \frac{N!}{k_a! k_b! k_c!} p^{k_a} q^{k_b} r^{k_c}. \qquad (۶۷)$$

البته بدیهی است که $K_a + K_b + K_c = N.$ تابع مولد چنین توزیعی برابر است با:

$$Z_{K_a,K_b}(t_a,t_b) = \sum_{k_a,k_b} e^{t_a k_a + t_b k_b} \frac{N!}{k_a! k_b! (N-k_a-k_b)!} p^{k_a} q^{k_b} r^{N-k_a-k_b} \quad (۶۸)$$

طرف راست را می توان ساده کرد که حاصل اش این است:

$$Z_{K_a,K_b}(t_a,t_b) = (e^{t_a} p + e^{t_b} q + r)^N. \quad (۶۹)$$

- تمرین: برای این تابع توزیع کمیت های زیر را حساب کنید:

$$\langle K_a K_b \rangle \qquad \langle K_a K_c \rangle \qquad \langle K_b K_c \rangle. \quad (۷۰)$$

بدیهی است که این تابع توزیع به تابع توزیع چندجمله ای [۱۷] تعمیم پیدا می کند.

توزیع دوجمله ای بیان می کند که در تعداد معینی از آزمایش که هر کدام دو حالت تصادفی دارند، چه تعداد از آزمایش ها به یک نتیجه معین رسیده اند.

هر گاه در تابع توزیع دوجمله ای حدهای خاصی از پارامترها را نگاه کنیم به توابع توزیع جدیدی می رسیم که برای توصیف فرایندهای تصادفی وسیعی اهمیت دارند و با نام های خاص نیز نامیده می شوند. در دو بخش بعدی این توابع توزیع را معرفی می کنیم.

۲.۷ تابع توزیع پوآسون

این بخش را با چند سوال آغاز می کنیم:

- شما صاحب یک فروشگاه هستید. به طور متوسط هر روز λ تا مشتری وارد فروشگاه شما می شوند. البته تعداد مشتری ها کم و زیاد می شود. گاهی اوقات تعداد مشتری ها بیشتر از این مقدار و بعضی از اوقات هم کمتر است. بعضی اوقات حتی ممکن است هیچ مشتری وارد فروشگاه شما نشود یا بعضی اوقات ممکن است ازدحام ایجاد شود. امروز از خود می پرسید که احتمال اینکه تعداد k تا مشتری به فروشگاه وارد شوند چقدر است؟ k یک عدد صحیح بزرگتر یا مساوی با صفر است. پاسخ این سوال توزیع پواسون است.

[۱۷] Multinomial Distribution

- مقداری ماده رادیواکتیو را در یک ظرف نگاه داشته اید. هر از گاهی یک ذره آلفا از این ماده ساطع می شود. می خواهید بدانید احتمال اینکه در t ثانیه k تا ذره آلفا ساطع شود چقدر است؟ پاسخ این سوال نیز توزیع پوآسون است.

- در یک روز بارانی ظرفی را زیر آسمان گذاشته اید. می خواهید بدانید احتمال اینکه پس از t ثانیه یک میلی متر آب باران در کف ظرف جمع شود، چقدر است؟ پاسخ این سوال نیز توزیع پوآسون است.

چه چیزی باعث می شود که پاسخ همه سوال های به ظاهر متفاوت بالا توزیع پوآسون باشد؟ دلیل اش این است که همه سوالهای بالا را می توان به سوالهایی در مورد یک توزیع دو جمله ای برگرداند. مثال فروشگاه را در نظر می گیریم. مدت زمان t را به زمان های خیلی کوچک $\epsilon = \frac{t}{N}$ تبدیل می کنیم. N را آنقدر بزرگ می گیریم که در هر فاصله زمانی ϵ به ندرت یک مشتری (و نه بیش از آن) وارد فروشگاه شود. به این ترتیب این فرایند را به یک فرایند دوجمله ای تبدیل کرده ایم که پیشامد اصلی اش دو حالت دارد: یک مشتری با احتمال $p << 1$ وارد فروشگاه می شود و با احتمال $1 - p$ هیچ کس وارد نمی شود. تعداد کل پیشامدها نیز برابر است با N و ما علاقمندیم بفهمیم احتمال اینکه k پیشامد به دلخواه ما باشد (در اینجا وارد شدن مشتری) چقدر است. این احتمال برابر است با:

$$P_N(k) = \frac{N!}{k!(N-k)!}p^k(1-p)^{N-k}. \tag{۷۱}$$

در این تابع توزیع p چقدر است؟ از آنجا که تعداد متوسط مشتری ها در یک روز برابر است با λ ، می دانیم که

$$Np = \lambda. \tag{۷۲}$$

بنابراین به جای p قرار می دهیم $p = \frac{\lambda}{N}$ و سپس به ازای هر k محدود (که مهم نیست چقدر بزرگ باشد) N را به سمت بی نهایت میل می دهیم. خواهیم داشت:

$$P_\lambda(k) = \frac{N(N-1)(N-2)\cdots(N-k+1)}{k!}p^k(1-\frac{\lambda}{N})^N \tag{۷۳}$$

وقتی که N را به سمت بی نهایت میل می دهیم این تابع تبدیل می شود به:

$$\begin{aligned} P_\lambda(k) \quad \longrightarrow \quad & \frac{N^k}{k!}p^k(1-\frac{\lambda}{N})^N \\ & \frac{\lambda^k}{k!}e^{-\lambda}. \end{aligned} \tag{۷۴}$$

متغیر تصادفی در این جا k است که می تواند مقادیر گسسته ی از 0 تا ∞ را اختیار کند. تابع مولد عبارت است از :

$$\begin{aligned}\tilde{P}_\lambda(t) &= \langle e^{t\mathcal{K}}\rangle = \sum_{n=0}^{\infty} t^k \frac{\lambda}{n!} e^{-\lambda} \\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!} = e^{\lambda(t-1)}.\end{aligned} \tag{۷۵}$$

از این رابطه نتیجه می گیریم:

$$\langle\mathcal{N}\rangle = \lambda, \qquad \langle\mathcal{N}^2\rangle = \lambda^2 + \lambda, \qquad \sigma_{\mathcal{N}} = \sqrt{\lambda}. \tag{۷۶}$$

■ مثال: یک هسته رادیواکتیو در یک ساعت به طور متوسط ۱۲ ذره ی گاما از خود تابش می کند. احتمال اینکه اولین ذره را پس از ۲۰ دقیقه ساطع کند چقدر است؟

حل: این ذره در مدت ۲۰ دقیقه به طور متوسط $\frac{12}{3} = 4$ ذره از خود ساطع می کند. بنابراین برای این فرایندِ پوآسون $\lambda = 4$ است. احتمال مورد نظر را به صورت حاصل ضرب دو احتمال دیگر می توان نوشت: یعنی

$$P = P_1 \times P_2 \tag{۷۷}$$

که در آن

$$\begin{aligned} P_1 &:= \text{احتمال اینکه در بیست دقیقه اول هسته هیچ ذره ای ساطع نکند} \\ P_2 &:= \text{احتمال اینکه در ۴۰ دقیقه بعدی یک یا بیشتر ذره ساطع کند.} \end{aligned} \tag{۷۸}$$

از فرایند پوآسون می دانیم که : $P_\lambda(n) = \frac{\lambda^n}{n!}e^{-\lambda}$ بنابراین با توجه به اینکه برای ۲۰ دقیقه اول مقدار λ یعنی متوسط تعداد ذرات ساطع شده در آن مدت برابر است با 4 و برای ۴۰ دقیقه بعدی تعداد متوسط ذرات ساطع شده در آن مدت برابر است با 8 بدست می آوریم:

$$P_1 = P_4(0) = e^{-4}, \qquad P_2 = 1 - P_8(0). \tag{۷۹}$$

در نتیجه احتمال P برابر می شود با:

$$P = e^{-4}(1 - e^{-8}) \approx 0.02. \tag{۸۰}$$

■ 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؟

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

$$P = P_{50}(40) \times P_{50}(60) = \frac{50^{40}}{40!} e^{-50} \times \frac{50^{60}}{60!} e^{-50} \quad (۸۱)$$

برای محاسبه این اعداد از تقریب استرلینگ استفاده می کنیم که بر مبنای آن داریم: $n! \approx \sqrt{2\pi n} n^n e^{-n}$. در نتیجه

$$40! \times 60! \approx 2\pi \times 10\sqrt{24} \times 40^{40} \times 50^{40} \times e^{-100}$$

پس از جایگذاری در رابطه قبلی خواهیم داشت:

$$\begin{aligned} P &\approx (\frac{5}{4})^{40} (\frac{5}{6})^{60} \times \frac{1}{2\pi \times 10 \times \sqrt{24}} \approx (\frac{25}{24})^{100} \times \frac{1}{20\pi\sqrt{24}} \\ &\approx (1 + \frac{100}{24}) \times \frac{1}{20\pi\sqrt{24}} \approx 0.016. \end{aligned} \quad (۸۲)$$

۳.۷ تابع توزیع گاووسی

تابع توزیع گاووسی برای یک متغیر پیوسته به صورت زیر تعریف می شود.

$$P(x) = A e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (۸۳)$$

در این تابع A یک ضریب بهنجارش است و σ پارامتری است که پهنای تابع توزیع را تعیین می کند. شکل (۴) این تابع توزیع را نشان می دهد. متغیر تصادفی X مقادیر از $-\infty$ تا $+\infty$ را اختیار می کند و متوسط این مقادیر نیز برابر است با x_0. با یک محاسبه ساده می توان نشان داد که ضریب بهنجارش برابر است با $A = \frac{1}{\sigma\sqrt{2\pi}}$ و در نتیجه تابع گاووسی به طور کامل برابر است با:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}. \quad (۸۴)$$

حال تابع مولد توزیع گاووسی را محاسبه می کنیم. داریم:

$$\tilde{P}(k) = \int e^{ikx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-x_0)^2} dx. \quad (۸۵)$$

برای محاسبه این تابع مولد از دو رابطه ی ساده که در آینده نیز به آنها نیاز خواهیم داشت استفاده می کنیم .این دو رابطه که خواننده براحتی می تواند درستی آنها را بیازماید عبارتند از:

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}ax^2}dx = \sqrt{\frac{2\pi}{a}}, \qquad \int_{-\infty}^{\infty} e^{-\frac{1}{2}ax^2+bx}dx = \sqrt{\frac{2\pi}{a}}e^{\frac{1}{2}\frac{b^2}{a}}. \tag{۸۶}$$

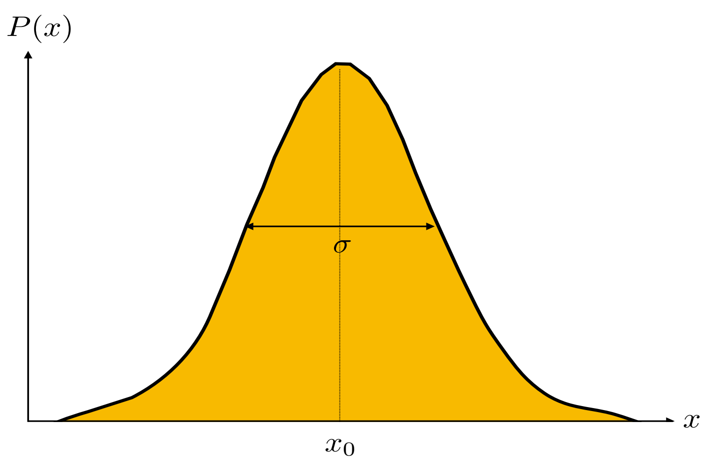
با استفاده از این دو رابطه تابع مولد بدست می آید:

$$\tilde{P}(k) = e^{-\frac{1}{2}\sigma^2k^2+ikx_0} \tag{۸۷}$$

و در نتیجه خواهیم داشت:

$$\langle X\rangle = x_0, \qquad \langle X^2\rangle - \langle X\rangle^2 = \sigma^2. \tag{۸۸}$$

پس پارامترهای x_0و و σ که در تعریف تابع توزیع گاووسی به کار رفته اند، به ترتیب نشان دهنده مقدار متوسط و واریانس متغیر تصادفی X هستند.



شکل ۴: تابع توزیع گاووسی

۸ ولگشت

معمولا بهترین توضیح مسئله ولگشت [۱۸] با مثال مرد مستی شروع می شود که اختیار گام های خود را از دست داده است و در بد ترین حالت با احتمال های مساوی یک گام به سوی خانه خود بر می دارد و یک گام ز خانه خود دور می شود. سوال این است که بعد از چند گام این شخص به خانه خود می رسد. می توان به حالت کلی تری فکر کرد که مرد مست چندان هم بی اختیار نیست بلکه با احتمال p یک گام به سوی خانه خود (که مثلا در سمت راست واقع است) برمی دارد و با احتمال $q = 1 - p$ یک گام به سمت چپ برمی دارد. سوال این است که بعد از N گام احتمال این که این شخص به نقطه ای با مختصه x رسیده باشد چقدر است؟ اسخ سوال خود را می توانیم با کمی دستکاری در تابع توزیع دو جمله ای پیدا کنیم.

در اینجا فرض می کنیم که طول گام های این شخص برابر با s است. تعداد کل گام های این شخص را با N و تعداد گام های رو به جلو را با K و تعداد گام های رو به عقب را با $N - K$ نشان می دهیم. در این صورت برای این که این شخص به نقطه x برسد می بایست شرایط زیر برقرار باشند:

$$(K - (N - K))s = s(2K - N)x. \tag{۸۹}$$

که معادل است با:

$$K = \frac{N + \frac{x}{s}}{2} \qquad N - K = \frac{N - \frac{x}{s}}{2}. \tag{۹۰}$$

بنابراین مسئله ولگشت تبدیل می شود به سوالی د رباره توزیع دو جمله ای: احتمال اینکه در N گام، این ولگرد K گام رو به جلو و $N - K$ گام رو به عقب بردارد به نحوی که شرایط (۸۹) برقرار باشد. به عبارت دیگر ما در اینجا با یک تغییر متغیر تصادفی روبرو هستیم. متغیر تصادفی اولیه N_+ است و متغیر تصادفی مورد علاقه ما در اینجا $X = (N_+ - N_-)s$ است. بنابراین داریم:

$$P_N(x) = P_N(K) = \frac{N!}{(K)!(N-K)!} p^K (q)^{N-K} \tag{۹۱}$$

و یا

$$P_N(x) = \frac{N!}{(\frac{N+\frac{x}{s}}{2})!(\frac{N-\frac{x}{s}}{2})!} p^{\frac{N+\frac{x}{s}}{2}} (1-p)^{\frac{N-\frac{x}{s}}{2}} \tag{۹۲}$$

اگر از خود بپرسیم که به طور متوسط بعد از N گام ولگرد به کجا رسیده است پاسخ اش ساده است:

$$\langle X \rangle = \langle (2K - N) \rangle s = \langle 2K - N \rangle s = (2p - 1)Ns. \tag{۹۳}$$

[۱۸] Random Walk

اگر $p = \frac{1}{2}$ باشد به نظر می رسد که ولگرد از نقطه اولش دور نشده است، ولی چنین تصوری درست نیست زیرا برای پاسخ درست به این سوال می بایست کمیت $\sqrt{\langle X^2\rangle}$ را که علامت جبری را در نظر نمی گیرد محاسبه کرد. می دانیم :

$$\langle X^2\rangle = s^2\langle(2K-N)^2\rangle = s^2\big[4\langle K^2\rangle - 4N\langle K\rangle + N^2\big]. \tag{۹۴}$$

با قرار دادن عبارت های زیر:

$$\langle K\rangle = Np, \qquad \langle K^2\rangle - \langle K\rangle^2 = Npq \tag{۹۵}$$

در عبارت بالا و پس از کمی محاسبه بدست می آوریم:

$$\langle X^2\rangle - \langle X\rangle^2 = (4pqN)s^2. \tag{۹۶}$$

به عبارت دیگر

$$\sqrt{\langle X^2\rangle - \langle X\rangle^2} = 2s\sqrt{pqN}. \tag{۹۷}$$

این رابطه به این معنی است که فبعد از N گام، فاصله متوسطی که ولگرد از مبداء مختصات دور می شود از مرتبه $\sqrt{N}$ است که نشان دهنده تاثیر مخرب ولگردی بر مسیریابی ولگرد است. به عنوان مثال این شخص بعد از برداشتن ۹۰۰ گام به طور متوسط تنها ۳۰ گام از نقطه مبدا دور شده است.

دلیل اهمیت مسئله ولگشت این است که بسیاری از مسایل دیگر چه در حوزه فیزیک و چه در حوزه های دیگر علوم قابل تبدیل به آن هستند. در طول این درس به تدریج با مثالهایی از این نوع در حوزه های مختلف آشنا می شویم.

۹ قضیه حد مرکزی

[۱۹] فرض کنید که $X_1, X_2, \cdots X_N$ متغیرهای تصادفی باشند با توابع توزیع $P_1, P_2, \cdots P_N$. حال متغیر تصادفی Y را به شکل زیر در نظر می گیریم:

$$Y := \frac{X_1 + X_2 + \cdots X_N}{N}. \tag{۹۸}$$

[۱۹] Central Limit Theorem

از خود می پرسیم که تابع توزیع متغیر تصادفی Y چیست؟ پاسخ این است که در حد N های بسیار بزرگ، تابع توزیع Y به سمت یک تابع توزیع گاووسی به شکل زیر میل می کند:

$$p_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}}e^{-\frac{1}{2\sigma_Y^2}(y-y_0)^2} \tag{۹۹}$$

که در آن

$$y_0 = \frac{1}{N}\sum_{i=1}^{N}\langle X_i\rangle, \qquad \sigma_Y := \frac{1}{N}\sum_{i=1}^{N}\sigma_i. \tag{۱۰۰}$$

اهمیت این قضیه در این است که مستقل از این که توابع توزیع P_1 تا P_N چگونه باشند همواره تابع توزیع نهایی گاووسی خواهد شد. البته این توابع توزیع می بایست شرایطی داشته باشند که در حین اثبات به آنها می پردازیم ولی نکته مهم این است که این عموم توابع توزیع چنین شرایطی را دارند و تنها توابع توزیع خیلی خاصی از این دایره بیرون می افتند. قبل از اثبات دقت کنید که هرگاه همه توابع توزیع یکسان باشند، روابط بالا به شکل زیر در می آیند:

$$y_0 = \langle X\rangle, \qquad \sigma_Y = \sigma_X. \tag{۱۰۱}$$

اثبات: می دانیم که

$$P_Y(y) = \langle\delta(y - \frac{X_1 + X_2 + \cdots X_N}{N})\rangle = \int dx_1dx_2\cdots dx_N\delta(y - \frac{x_1 + \cdots x_N}{N})P_1(x_1)P_2(x_2)\cdots P_N(x_N) \tag{۱۰۲}$$

بجای $P_Y(y)$ تابع مولد آن یعنی $\tilde{P}_Y(k)$ را حساب می کنیم:

$$\tilde{P}_Y(k) = \int e^{iky}P_Y(y)dy = \int e^{iky}dy\int dx_1dx_2\cdots dx_N\delta(y - \frac{x_1 + \cdots x_N}{N})\cdots dx_N \tag{۱۰۳}$$

حال روی y انتگرال می گیریم.

$$\tilde{P}_Y(k) = \int e^{ik(\frac{x_1+x_2+\cdots x_N}{N})}P_1(x_1)P_2(x_2)\cdots P_N(x_N)dx_1dx_2\cdots dx_N. \tag{۱۰۴}$$

انتگرال روی x_1 تا x_N از هم جدا می شوند هرکدام تابع مولد متغیر مربوط را به ازای $\frac{k}{N}$ بدست می دهند. بنابراین

$$\tilde{P}_Y(k) = \tilde{P}_1(\frac{k}{N})\tilde{P}_2(\frac{k}{N})\cdots\tilde{P}_N(\frac{k}{N}) \tag{۱۰۵}$$

اما می دانیم که تابع مولد یک متغیر بسط را دارد:

$$\tilde{P}_1(k) = 1 + ik\langle X_1\rangle + \frac{(ik)^2}{2}\langle X_1^2\rangle + O(k^3) \approx e^{ik\langle X_1\rangle + \frac{(ik)^2}{2}(\langle X_1^2\rangle - \langle X_1\rangle^2) + O(k^3)} \tag{۱۰۶}$$

بنابراین

$$\tilde{P}_i(\frac{k_i}{N}) = e^{\frac{ik}{N}\langle X_i\rangle + \frac{1}{2N^2}(ik)^2\sigma_{X_i}^2 + O(\frac{1}{N^3})} \tag{۱۰۷}$$

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

$$\tilde{P}_Y(\mathbf{k}) = e^{ik\left(\frac{\langle X_1\rangle + \langle X_2\rangle + \cdots \langle X_N\rangle}{N}\right) + (ik)^2\frac{1}{2}\left(\frac{\sigma_{X1}^2 + \cdots \sigma_{X_N}^2}{N}\right)} \tag{۱۰۸}$$

هرگاه تعریف کنیم:

$$y_0 := \frac{\langle X_1\rangle + \cdots \langle X_N\rangle}{N}, \tag{۱۰۹}$$

و

$$\sigma_Y^2 := \frac{\sigma_{X1}^2 + \cdots \sigma_{XN}^2}{N} \tag{۱۱۰}$$

و از توان های $\frac{1}{N^3}$ صرف نظر کنیم، به رابطه زیر می رسیم:

$$\tilde{P}_Y(k) = e^{iky_0 + \frac{1}{2}(ik)^2\sigma_Y^2} \tag{۱۱۱}$$

که به معنای آن است که

$$P_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}}e^{-\frac{1}{2}(y-y_0)^2}, \tag{۱۱۲}$$

یعنی اینکه تابع توزیع y واقعا یک تابع توزیع گاووسی است با متوسط y_0 و پهنا یا واریانس σ_Y.

۱۰ چند رابطه مفید در باره اعداد بزرگ

درمکانیک آماری تعداد ذرات یک سیستم بسیار بزرگ است. گاهی اوقات پیش می آید که می خواهیم مجموع جملات یک سری را که تعداد جملات آن بسیار زیاد است را حساب کنیم. یا پیش می آید که می خواهیم یک انتگرال را تخمین بزنیم. تحت شرایطی می توانیم تقریب

های بسیار خوبی برای این کمیت ها را که با اعداد بسیار بزرگ سرو کار دارند پیدا کنیم. در این بخش بعضی از این تکنیک ها را معرفی می کنیم.

- یک سری به صورت زیر در نظر بگیرید.

$$S = \sum_{k=1}^{N} E_k, \tag{۱۱۳}$$

که در آن E_k ها جملات مثبتی هستند. می خواهیم این عبارت را تخمین بزنیم. فرض کنید که E_{max} بزرگترین جمله بین E_k ها باشد. نخستین تخمینی که به آن توجه می کنیم این است که

$$E_{max} \leq S \leq NE_{max}. \tag{۱۱۴}$$

بنابراین بلافاصله یک حد بالا و پایین برای کمیت S بدست می آوریم. می توانیم تقریب خود را بهتر کنیم اگر اطلاعات بیشتری در باره جملات E_k داشته باشیم. فرض کنیم که این جملات به صورت زیر باشند، و این وضعیتی است که در بیشتر موارد به آن برمی خوریم:

$$E_k = e^{N\phi(k)}. \tag{۱۱۵}$$

حال می توانیم عبارت جمع S را برای N های بزرگ به صورت یک انتگرال بنویسیم:

$$S = \int_0^\infty e^{N\phi(x)} dx \tag{۱۱۶}$$

حالا می توانیم با تقریبی به نام تقریب نقطه زینی [۲۰] این انتگرال یا همان جمع اولیه را به صورت بهتری تخمین بزنیم: $\phi(x)$ تابعی به شکل زیر باشد که دارای یک ماکزیمم باشد . در این صورت واضح است که در حد N های بزرگ سهم عمده انتگرال از نقطه ماکزیمم تابع $\phi(x)$ و اطراف آن ناشی می شود. بنابراین $\phi(x)$ را حول نقطه ماکزیمم یعنی حول x_0 بسط می دهیم و بدست می آوریم:

$$I = \int dx e^{N\left(\phi(x_0) - \frac{1}{2}|\phi''(x_0)|(x-x_0)^2 + \cdots\right)} dx. \tag{۱۱۷}$$

دراین جا از این موضوع استفاده کرده ایم که $\phi''(x_0)$ منفی است چون نقطه ی x_0 ماکزیمم است. اگر از جملات مرتبه بالاتر از ۲ صرف نظر کنیم I تبدیل به یک انتگرال گاووسی می شود و براحتی قابل محاسبه است. در واقع با تغییر متغیر $\xi := x - x_0$ و توجه به این که محدوده انتگرال روی

[۲۰] Saddle Point Approximation

ξ را می توان به $(-\infty, \infty)$ گسترش داد بدست می آوریم:

$$I = e^{N\phi(x_0)} \int_{-\infty}^{\infty} d\xi e^{-\frac{1}{2}N|\phi''(x_0)|\xi^2} = e^{N\phi(x_0)} \sqrt{\frac{2\pi}{N|\phi''(x_0)|}} \quad (118)$$

حال سوال این است که با صرف نظر کردن از جملات مرتبه بالا تر چه چیزی را از دست داده ایم. در این جا می بایست از قضایایی که در مورد انتگرال های گاووسی می دانیم استفاده کنیم. فرض کنید که یک جمله دیگر را در بسط $\phi(x)$ حول نقطه x_0 نگاه داریم. در این صورت جمله متناسب با ξ^3 در انتگرال گیری صفر خواهد شد و جمله درجه چهارم با استفاده از قضیه ویک قابل محاسبه است. در نتیجه بدست می آوریم:

$$I = e^{N\phi(x_0)} \sqrt{\frac{2\pi}{N|\phi''(x_0)|}} \left(1 + N\beta\langle\xi^4\rangle\right) \quad (119)$$

اما از قضیه ویک (ضمیمه یک) می دانیم که $\langle\xi^4\rangle = 3\langle\xi^2\rangle^2$ و می دانیم که $\langle\xi^2\rangle \equiv \Delta = (N|\phi''(x_0)|)^{-1}$ بنابراین می بینیم که $\langle\xi^4\rangle \sim N^{-2}$. بنابر این جمله $N\beta\langle\xi^4\rangle$ متناسب با N^{-1} است و در حد N های بزرگ به سمت صفر میل می کند. .

- تقریب استرلینگ [۲۱] تقریب استرلینگ کاربرد زیادی در احتمالات و مکانیک آماری دارد زیرا برای N های بزرگ مقدار $N!$ را بدست می دهد. بنا بر این تقریب داریم:

$$N! \approx \sqrt{2\pi N} N^N e^{-N} \quad (120)$$

یا

$$\ln N! \approx N \ln N - N. \quad (121)$$

در عبارت دوم از جمله $\frac{1}{2}\ln N + \frac{1}{2}\ln 2N$ در مقابل بقیه جملات صرف نظر کرده ایم. نکته مهم در مورد تقریب استرلینگ آن است که لزومی ندارد N بسیار بزرگ باشد تا این تقریب خوب از آب درآید، حتی برای $N \sim 10$ نیز خطای این تقریب بسیار کم است. برای اثبات این تقریب از رابطه زیر شروع می کنیم:

$$\int_0^{\infty} e^{-t} t^N = N! \quad (122)$$

[۲۱] Stirling approximation

بنابراین می نویسیم:

$$N! = \int_0^\infty e^{-t+N\ln t}dt = \int_0^\infty e^{N(\ln t-\frac{1}{N})}dt$$
$$= \int_0^\infty e^{N\phi(t)}dt. \qquad (۱۲۳)$$

حال اخرین آنتگرال را با استفاده از تقریب نقطه زینی حساب می کنیم:

$$\phi(t)=\ln t-\frac{t}{N}, \qquad \phi'(t)=\frac{1}{t}-\frac{1}{N}, \qquad \phi'(N)=0, \qquad \phi''(N)<0. \qquad (۱۲۴)$$

در نتیجه بدست می آوریم

$$N! = e^{N\phi(N)}\sqrt{\frac{2\pi}{1/N}} = e^{N(\ln N-1)}\sqrt{2\pi N} \qquad (۱۲۵)$$

و از آنجا

$$N! \approx N^N e^{-N}\sqrt{2\pi N}. \qquad (۱۲۶)$$

۱۱ آنتروپی ، اطلاعات و بی نظمی

در ترمودینامیک با مفهوم آنتروپی و معنای ماکروسکوپی آن آشنا شدیم. در این درس با مفهوم میکروسکوپی آنتروپی آشنا می شویم و رابطه آن را با اطلاعات و بی نظمی توضیح می دهیم. نخست به مفهوم اطلاعات توجه می کنیم. چگونه می توان اطلاعات را به صورت یک مفهوم کمی فهمید؟ چگونه می توان میزان اطلاعات موجود در یک پیام را اندازه گرفت؟ به دو جمله خبری زیر که در یک روز گرم تابستان ممکن است به ما گفته شود توجه کنید:

۱) فردا آفتاب طلوع خواهد کرد.

۲) فردا باران خواهد بارید.

ما به جمله اول هیچ گونه واکنشی نشان نمی دهیم زیرا بیان امری است که قطعا به وقوع می پیوندد و گوینده این حرف هی چیزی به معلومات ما اضافه نمی کند. اما جمله دوم مقداری اطلاعات ارائه می کند زیرا بارش باران امری حتمی نیست و تنها محتمل است. فردا ممکن است باران

ببارد یا نبارد و گوینده این حرف با گفتن این جمله خبری از عدم یقین ما نسبت به این اتفاق کاسته است. بنابراین می توان پذیرفت که میزان اطلاعاتی که در یک جمله وجود دارد نسبت معکوس با احتمال وقوع آن داشته باشد. حال می بایست این دریافت شهودی را به صورت یک رابطه دقیق کمی دربیاوریم.

فرض کنید که آزمایش یا واقعه ای مثل X که نتایج یاپیشامدهای ممکن آن را بامجموعه $\{x_1, x_2, \cdots x_n\}$ نشان می دهیم اتفاق بیفتد و یک نتیجه معین مثل x_i حاصل شود. دراین صورت می توانیم بپرسیم که ما به عنوان ناظر یا مشاهده گر چه مقدار اطلاع حاصل کرده ایم، یا چه مقدار از عدم یقین ما نسبت به نتیجه های ممکن کاسته شده است. فرض ما این است که احتمالات وقوع یعنی $p(x_i)$ ها معلوم هستند. طبیعی است که با دانستن این احتمالات نمی توان یقینا گفت که چه پیشامدی رخ خواهد داد. میزان عدم یقینی که نسبت به نتیجه داریم و در نتیجه مقدار اطلاعی که از مشاهده خود دریافت می کنیم، طبیعتا تابعی از این احتمالات است. به عنوان مثال اگر داشته باشیم

$$P(x_1) = 1, \qquad P(x_i) = 0, \quad i = 2,\ 3,\ \cdots N, \tag{۱۲۷}$$

آنگاه نتیجه هر آزمایشی از قبل معلوم است و ما از مشاهده آزمایش هیچ اطلاعی حاصل نمی کنیم، زیرا از قبل و با محاسبه تحلیلی می توانیم بگوییم که همواره نتیجه x_1 حاصل خواهد شد. اما اگر داشته باشیم

$$P(x_i) = \frac{1}{N}, \tag{۱۲۸}$$

آنگاه هربار که آزمایش را انجام می دهیم یک نتیجه بدست می آید که به دانش ما اضافه می کند، دانشی که از قبل نداشتیم و نمی توانستیم با محاسبه ریاضی به آن برسیم. ازنظرشهودی هرچقدرکه پیشامدی که بوقوع پیوسته است محتمل تر بوده باشد اطلاعی که ماکسب کرده ایم کمتروهرچقدرکه آن پیشامد دورازانتظاربوده باشد تعجب ماازوقوع آن بیشتر و اطلاعی که ماکسب کرده ایم بیشتر خواهد بود. بنابراین اگر میزان اطلاع خودازوقوع پیشامد x_i را با h_i نشان دهیم می توانیم بگوییم که h_i می بایست نسبت معکوس با احتمال وقوع آن پیشامد یعنی p_i داشته باشد.

حال فرض کنید که یک آزمایش مرکب از دوواقعه مستقل (X, Y) شود که نتایج ممکن آن را با زوج های $\{(x_i, y_j),\ \ i = 1 \cdots m,\ j = 1 \cdots n\}$ نشان می دهیم. هرگاه احتمال وقوع x_i را با p_i واحتمال وقوع y_j را با q_j نشان دهیم احتمال هرپیشامد (x_i, y_j) برابرخواهد بود با $p_i q_j$ ومیزان اطلاعی که ازوقوع این پیشامد کسب می کنیم برابرخواهد بود با $h(p_i q_j)$. انتظارداریم که میزان اطلاع ما دراین مورد که دوپیشامد مستقل x_i و y_j رخ داده اند برابربا مجموع اطلاعاتی باشد که از وقوع پیشامد x_i به تنهایی و y_j به تنهایی کسب می کنیم بنابراین انتظارداریم که

$$h(p_i q_j) = h(p_i) + h(q_j). \tag{۱۲۹}$$

تنهاتابعی که شرط فوق رابرآورده کند و ضمناً نزولی باشد، تابع لگاریتم است بنابراین خواهیم داشت:

$$h(p_i) = \log_\alpha \frac{1}{p_i}, \tag{۱۳۰}$$

که درآن α ثابت است. ثابت α رامی توان با شرط بهنجارش تعیین کرد. قرارمی نهیم که میزان اطلاع کسب شده ما ازوقوع یک پدیدهٔ دوحالته متساوی الاحتمال برابربایک باشد، یعنی $h(1/2) = 1$. درنتیجه میزان ثابت α برابرمی شود با ۲.

اگر یک آزمایش X را N بارانجام دهیم به طور متوسط Np_i بارنتیجه x_i رخ خواهد داد و میزان اطلاع عی که درهربارکسب می کنیم برابرخواهد بود با $\log_2(\frac{1}{p_i})$. میزان اطلاعی که ما به طورمتوسط از وقوع نتایج آزمایش تصادفی X کسب می کنیم برابرخواهد بود با:

$$H(X) = -\frac{1}{N}\sum_x Np(x)\log_2 p(x) = -\sum_x p(x)\log_2 p(x). \tag{۱۳۱}$$

این تابع ، تابع آنتروپی یا تابع شانون نیزخوانده می شود. دقت کنید که تابع $p\log\frac{1}{p}$ درفاصله $p \in [0,1]$ یک تابع مثبت است بنابراین $H(X)$ یک تابع مثبت است.

- تمرین: با مراجعه به گوگل، فرکانس حروف انگلیسی را پیدا کرده و سپس تابع آنتروپی را برای آن پیدا کنید.

۱.۱۱ اطلاعات دو متغیرتصادفی

هرگاه دومتغیرتصادفی (X,Y) داشته باشیم که لزوماً ازهم مستقل نباشند تابع آنتروپی یا اطلاعات به طورطبیعی به شکل زیرتعریف می شود:

$$H(X,Y) := -\sum_{x,y} p(x,y)\log_2 p(x,y) \tag{۱۳۲}$$

درحالتی که دومتغیرتصادفی مستقل باشند یعنی $p(x,y) = p(x)q(y)$، رابطه بالابدست می دهد که $H(X,Y) = H(X) + H(Y)$.

این تعریف به همین شکل به بیش از دومتغیرتصادفی تعمیم می یابد به این معنا که تعریف می کنیم:

$$H(X,Y,Z) = -\sum_{x,y,z} p(x,y,z)\log_2 p(x,y,z). \tag{۱۳۳}$$

از یک زاویه مهم دیگر نیز می توانیم به مفهوم اطلاعات فکر کنیم. فرض کنیدکه هدف ما ارسال متن هایی است که تنهاازچهارحرف الفبا به نام های $A,\ B,\ C$ و D تشکیل شده است.یک روش برای ارسال این متن ها آن است که حرف های چهارگانه فوق را با بیت های 0 و 1 که درمخابرات دیجیتال معمول است، به ترتیب زیر کدکنیم .

$$\begin{aligned} A &\longrightarrow 00 \\ B &\longrightarrow 01 \\ C &\longrightarrow 10 \\ D &\longrightarrow 11. \end{aligned} \tag{۱۳۴}$$

دراین صورت به ازای هرحرف دوبیت مخابره کرده ایم. حال سوال این است که آیا می توانیم یک روش کد کردن به کارببریم که درآن طول به ازای هرحرف تعداد بیت هایی که به طور متوسط مخابره می کنیم کمتراز 2 باشد؟

فرض کنید که این حروف درمتن های یادشده بااحتمالات زیر ظاهرمی شوند:

$$P(A)=\frac{1}{8}\qquad P(B)=\frac{1}{8}\qquad P(C)=\frac{1}{4}\qquad P(D)=\frac{1}{2}. \tag{۱۳۵}$$

حال روش کدگذاری زیررا به کارمی بریم:

$$\begin{aligned} D &\longrightarrow 0 \\ C &\longrightarrow 10 \\ B &\longrightarrow 110 \\ A &\longrightarrow 111. \end{aligned} \tag{۱۳۶}$$

دراین روش کدگذاری برای بعضی ازحروف بیش از دوبیت به کاربرده ایم ولی اگر طول متوسط کدهایی را که برای حروف به کاربرده ایم محاسبه کنیم نتیجه جالب توجه خواهد بود. این طول متوسط برابراست با:

$$\langle l\rangle=\sum_{i=1}^{4} l_i\times p_i = 1\times\frac{1}{2}+2\times\frac{1}{4}+3\times\frac{1}{8}+3\times\frac{1}{8}=\frac{7}{4}. \tag{۱۳۷}$$

بنابراین با یک کدگذاری مناسب توانسته ایم طول متوسط رشته بیت هایی را که برای مخابره پیام بکاربرده ایم از ۲ به ۴/۷ تقلیل دهیم. ضمناً باید دقت کنیم که این نحوه کدگذاری هیچ نوع ابهامی درباره متنی که مخابره شده است دربرندارد وهررشته ای ازبیت ها به طوریکتا به متن اولیه

بازگشایی می شود. به عنوان مثال رشته زیر

$$0\,1\,0\,0\,0\,1\,0\,0\,0\,1\,1\,0\,1\,1\,1. \tag{۱۳۸}$$

بدون ابهام به متن زیرگشوده می شود ومتن دیگری برای بازگشایی آن قابل تصور نیست

$$D\ C\ D\ D\ C\ D\ D\ B\ A. \tag{۱۳۹}$$

این که چه نوع کد هایی یکتاگشاهستند موضوعی است که فعلاً موضوع بحث مانیست. ولی یک نکته‌مهم رابابد ذکرکنیم: هرگاه آنتروپی متغیرتصادفی $X=\{A,\ B,\ C,\ D\}$ را بااحتمالات ذکرشده حساب کنیم حاصل آن برابرخواهد بودبا:

$$\begin{aligned} H(X) &= \sum_{i=1}^{4} p_i \log_2(\frac{1}{p_i}) = \frac{1}{2}\times\log_2(2)+\frac{1}{4}\times\log_2(4)+\frac{1}{8}\times\log_2(8)+\frac{1}{8}\times\log_2(8) \\ &= \frac{1}{2}\times 1+\frac{1}{4}\times 2+\frac{1}{8}\times 3+\frac{1}{8}\times 3=\frac{7}{4}. \end{aligned} \tag{۱۴۰}$$

بنابراین دراین مثال خاص طول متوسط کدگزاری ای که به کاربردیم با میزان اطلاعات موجود درمتن برابراست. آیا این یک خصلت عمومی است؟ پاسخ این سوال مثبت است . فرض کنید که احتمالات به شکل زیر باشند.

$$P(A)=p_1, \qquad P(B)=p_2, \qquad P(C)=p_3, \qquad P(D)=p_4. \tag{۱۴۱}$$

در این صورت رشته بلندی از نتایج را در نظر بگیرید:

$$x_{i_1}x_{i_2}x_{i_3}x_{i_4}\cdots x_{i_N M} \tag{۱۴۲}$$

این رشته ها را به بلوک های به طول N که در آن N عدد بزرگی است تقسیم می کنیم. هر رشته به طول N می تواند به 4^N طریق تشکیل شود از $AAA\cdots A$ گرفته تا $DDD\cdots D$. هر بلوک N تایی که n_1 تا A دارد و n_2 تا B و ..... n_4 تا D دارد با احتمال $P=p_1^{n_1}p_2^{n_2}p_3^{n_3}p_4^{n_4}$ ظاهر می شود. برای چنین رشته ای کدی به کار می بریم که طول آن برابر باشد با $-\log P$ یعنی قرار می دهیم

$$l=-\log P=-\left(n_1\log p_1+n_2\log p_2+\cdots n_4\log p_4\right). \tag{۱۴۳}$$

طول متوسط رشته بیت ها چقدراست؟ این طول برابر است با:

$$\bar{l}=\sum P\times l\times\frac{N!}{n_1!n_2\cdots n_4!} \tag{۱۴۴}$$

که در آن $\frac{N!}{n_1!\cdots n_4!}$ تعداد تمام رشته هایی است که n_1 تا A دارند، و n_2 تا B دارند و الی اخر. بنابراین بدست می آوریم:

$$\bar{l} = -\sum_{n_1,n_2,n_3,n_4} p_1^{n_1}\cdots p_4^{n_4}\left(n_1\log p_1 + n_2\log p_2 + \cdots n_4\log p_4\right)\frac{N!}{n_1!\cdots n_4!} \tag{۱۴۵}$$

یک محاسبه ساده که به عهده خواننده است نشان می دهد که

$$\bar{l} = -N\left(p_1\log p_1 + p_2\log p_2 + \cdots p_4\log p_4\right) = NH(X). \tag{۱۴۶}$$

بنابراین $H(X) = \frac{\bar{l}}{N}$ و $H(X)$ واقعا تعداد بیت بر حرفی است که ما برای مخابره اطلاعات موجود در این رشته احتیاج داریم.

بنابراین می بینیم که آنچه را که به عنوان میزان اطلاعات تعریف کرده ایم یک معنای عملی و روشن دارد که مبنای فناوری اطلاعات نیز هست.

■ تمرین: فرض کنید که متنی که می خواهیم ذخیره کنیم از همان الفبای ساده چهارحرفی با همان احتمالات تشکیل شده است اما این بار می خواهیم حروف را به صورت دوتایی کد کنیم. ضمنا می دانیم که بین حروف یک همبستگی وجود دارد : این همبستگی به احتمالات زیر مشخص شده است:

$$P(x|x) = 5/8, \qquad P(y \neq x|x) = 1/8. \tag{۱۴۷}$$

الف: احتمالات مربوط به تمام حروف دوتایی را محاسبه کنید.

ب: حال حروف دوتایی را طوری کد کنید که بیشترین فشردگی حاصل شود. تعداد بیت های لازم برای ذخیره هر حرف به طور متوسط چقدر است؟ اگر این همبستگی وجود نداشت تعداد بیت های لازم برای ذخیره هر حرف چقدر می شد؟

ج: احتمالات مربوط به تمام حروف سه تایی را بدست بیاورید. فرض کنید که همبستگی ها فقط دوتایی است .

■ تمرین: فرض کنید که یک متن از حروف زیر با احتمالات نوشته شده تشکیل شده است. یک کد بهینه برای این حروف بنویسید به نحوی که هم طول متوسط پایین باشد و هم رشته ای صفر و یک ها به صورت یکتا به حروف نگاشته شود.

(۱۴۸)

$$P(A) = \frac{1}{32} \qquad P(B) = \frac{1}{32} \qquad P(C) = \frac{1}{16} \qquad P(D) = \frac{1}{8} \qquad P(E) = \frac{1}{4} \qquad P(F) = \frac{1}{2}$$

■ تمرین: فرض کنید که یک متن از حروف زیر با احتمالات نوشته شده تشکیل شده است. یک کد بهینه برای این حروف بنویسید به نحوی که هم طول متوسط پایین باشد و هم رشته ای صفر و یک ها به صورت یکتا به حروف نگاشته شود.

$$P(A)=\frac{1}{128}\qquad P(B)=\frac{1}{128}\qquad P(C)=\frac{1}{64}\qquad P(D)=\frac{1}{32}$$
$$P(E)=\frac{1}{16}\qquad P(F)=\frac{1}{8}\qquad P(G)=\frac{1}{4}\qquad P(H)=\frac{1}{2} \tag{۱۴۹}$$

۱۲ مسئله ها:

در مسئله های زیر یک مجموعه فضای نمونه مثل S در نظر بگیرید که $A,\ B,\cdots$ زیر مجموعه های آن (یا پیشامدهای ممکن) هستند.

■ نشان دهید که اگر $P(A)=0.9$ و $P(B)=0.8$ آنگاه $P(A\cap B)\geq 0.7.$ به طور کلی نشان دهید که نامساوی زیر موسوم به نامساوی بنفرونی [۲۲] برای هر دو پیشامدی برقرار است:

$$P(A\cap B)\geq P(A)+P(B)-1. \tag{۱۵۰}$$

■ اگر $A\subset B$ باشد، بدیهی نیست که $P(B)$ بزرگتر یا مساوی $P(A)$ باشد. این موضوع را می بایست با استفاده از اصول موضوع کولموگروف ثابت کرد. با استفاده از این اصول موضوع ثابت کنید که:

آنگاه نشان دهید

$$P(B)\geq P(A). \tag{۱۵۱}$$

■ فرض کنید که هر کودکی که متولد می شود با احتمال مساوی دختر یا پسر است. اگر خانواده ای دو کودک داشته باشد، احتمال این که هر دو کودک دختر باشند چقدر است اگر:

الف: دختر بزرگ تر دختر باشد،

ب: حداقل یکی از کودک ها دختر باشد.

[۲۲] Benferroni

- دو مهره تاس را می اندازیم. احتمال اینکه حداقل یکی از انها برابر با ۶ باشد، چقدر است؟ اگر دو وجه تاس با هم متفاوت باشند، احتمال اینکه حداقل یکی از آنها برابر با ۶ باشند چقدر است؟

- سه مهره تاس را می اندازیم. احتمال اینکه شماره دقیقا دو تا از تاس ها باهم برابر باشد چقدر است؟

- فرض کنید که در یک جامعه آماری تعداد زنان و مردان برابر است و 0.5 درصد از مردان و 0.25 درصد از زنان کور رنگ هستند. حال یک فرد کوررنگ را به صورت تصادفی انتخاب می کنیم. احتمال این که این شخص مرد باشد چقدر است؟

- شخص الف با شخص ب یک بازی تکراری را انجام می دهند. در هر بازی یک امتیاز مثبت به برنده تعلق می گیرد و به بازنده هم هیچ امتیازی داده نمی شود. این دو شخص تا جایی بازی می کنند که تفاوت امتیاز آنها به ۲ برسد. در این جا بازی تمام می شود. فرض کنید که در هر بار احتمال برد الف برابر با p و احتمال برد ب برابر با $1-p$ است. احتمال اینکه مجموع امتیازات آنها پس از پایان بازی برابر با $2n$ شود چقدر است؟ احتمال این که الف برنده بازی شود چقدر است؟

- در یک مهمانی n مرد شرکت کرده و در هنگام ورود کلاه هایشان را به جارختی آویزان کرده اند. در پایان مهمانی برق رفته و همه مردان به صورت تصادفی یک کلاه برداشته و رفته اند. نشان دهید احتمال اینکه هیچ کس کلاه خود را برنداشته باشد برابر است با:

$$P=\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\cdots\frac{(-1)^n}{n!}. \qquad (۱۵۲)$$

این احتمال را برای وقتی که تعداد میهمان ها به سمت بی نهایت میل کند حساب کنید.

- در مسئله قبلی X را متغیر تصادفی بگیرید که تعداد مردانی را نشان می دهد که کلاه خود را بر می دارند. $P_N(X)$ یعنی احتمال اینکه دقیقا X تا از میهمان ها کلاه خود را بردارند، را حساب کنید. سپس مقدار $\sigma_N(x)$ یعنی واریانس این متغیر تصادفی را برای وقتی که تعداد میهمان ها ۴ یا ۵ است، حساب کنید.

- فرض کنید که یک تاس دو بار انداخته می شود. مقادیر ممکن را برای متغیرهای تصادفی زیر تعیین کنید:

الف_ مقدار بیشینه ای که روی تاس ها ظاهر می شود.

ب_ مقدار کمینه ای که روی تاس ها ظاهر می شود.

پ_ جمع روی دو مقداری که روی تاس ها ظاهر می شود.

ت ــ مقدار اولین تاس منهای مقدار دومین تاس.

■ یک شرکت هواپیمایی می داند که پنج در صد مسافرانی که بلیط می خرند، موقع پرواز حاضر نمی شوند. [۲۳] به همین دلیل سیاست این شرکت این است که برای هواپیماهای خود که ظرفیت ۵۰ نفر دارند ۵۲ تا بلیط می فروشد. احتمال اینکه ب هر مسافری که هنگام پرواز حاضر می شود، یک صندلی داشته باشد (مسافری بدون صندلی نماند) چقدر است؟ [۲۴]

■ در یک بازی ورق چهار کارت به تصادف به شما داده می شود. احتمالات زیر را پیدا کنید:

۱ ــ احتمال اینکه دو آس و دو صورت (از هر نوعی) داشته باشید.

۲ ــ احتمال اینکه سه آس و یک صورت داشته باشید.

۳ــ احتمال اینکه حد اقل دو صورت داشته باشید.

۴ــ احتمال اینکه حداکثر سه صورت داشته باشید.

■ تابع توزیع یک متغیر تصادفی به نام x که مقادیرش را در فاصله $[-1,1]$ اختیار می کند یکنواخت است. در هرکدام از موارد زیر تابع توزیع y را بدست آورید:

$$a)\ y = x^2 \qquad b)\ y = \sin(x) \qquad c)\ y = \tan(x) \qquad d)\ y = \arcsin(x). \qquad (۱۵۳)$$

■ دو منشی ماشین نویس را در نظر بگیرید که آنها را A و B می نامیم. A در هر ۱۰ حرف مرتکب ۱ خطا و B در هر۱۰ حرف مرتکب ۲ خطا می شود. احتمال این را پیدا کنید که در تایپ ۴ حرف B دقیقا ۲ خطا بیشتر از A مرتکب شود.

■ شخصی از یک نردبان که N پله دارد بالا می رود. وی در هر گام یا دو پله و یا یک پله بالا می رود. کلا این شخص به چند طریق می تواند به بالای نردبان برسد. (وی حرکت خود را از روی زمین شروع می کند.) راهنمایی: می بایست یک رابطه تکرار بنویسید. برای مقادیر $N = 5, 6, 7$ تعداد راه ها را بنویسید.

[۲۳] Do not show up.

[۲۴] shows up

■ یک کمیته پنج نفره را از یک گروه ۱۵ نفره که دارای ۸ مرد و ۷ زن است انتخاب می کنیم.

۱ـ احتمال اینکه این کمیته دارای ۲ زن و ۳ مرد باشد چقدر ست؟

۲ـ احتمال اینکه این کمیته حداقل دارای یک زن باشد چقدر است؟

■ یک سکه با احتمال p شیر و با احتمال $1-p$ خط می آید.

الف: این سکه را به طور مرتب می اندازیم تا اینکه یک بار شیر بیاید. احتمال اینکه n بار سکه را بیندازیم تا یک بار شیر بیاید چقدر است؟

ب: این سکه را به طور مرتب می اندازیم تا اینکه r بار شیر بیاید. احتمال این که n بار سکه را بیندازیم تا r بار شیر بیاید چقدر است؟

■ درستی رابطه زیر را نشان دهید:

$$\lim_{n\longrightarrow\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k} = \frac{1}{2}. \qquad (۱۵۴)$$

راهنمایی: فرض کنید که X_n یک تابع توزیع پوآسون با مقدار متوسطِ n است. از قضیه حد مرکزی استفاده کنید و نشان دهید که

$$P(X_n \leq n) \longrightarrow \frac{1}{2}. \qquad (۱۵۵)$$

■ دو سکه مختلف در دست داریم. یک سکه به گونه ای است که وقتی آن را پرتاب می کنیم با احتمال p روی شیر را نشان می دهد و با احتمال $1-p$ روی خط را نشان می دهد. سکه دیگر شیر را با احتمال q و خط را با احتمال $1-q$ نشان می دهد. این دو سکه را به ترتیب سکه های از نوع p و q می خوانیم. شخصی به طور تصادفی یکی از این دو نوع سکه را به ما می دهد و ما آن را N بار به هوا پرتاب می کنیم و N_+ بار شیر می آید. احتمال اینکه سکه p را به ما داده باشد چقدر است؟ احتمال اینکه سکه از نوع q را به ما داده باشد چقدر است؟ برای وقتی که مقادیر پارامترها به ترتیب زیر هستند، این احتمال ها را حساب کنید:

الف:

$$p = 0.5, \qquad q = 0.6, \qquad N = 10, \qquad N_+ = 6, \qquad (۱۵۶)$$

ب:

$$p = 0.5, \qquad q = 0.6, \qquad N = 100, \qquad N_+ = 60. \qquad (۱۵۷)$$

■ مجموع سری زیر را با استفاده از تقریب نقطه زینی برای N های بزرگ تخمین بزنید:

$$S = \sum_{k=0}^{N} 2^N k N^k \tag{۱۵۸}$$

۱۳ ضمیمه: چند نکته در مورد انتگرال های گاووسی

انتگرال های یک متغیره گاووسی براحتی محاسبه می شوند. حدود همه انتگرال ها از منهای بی نهایت تا بعلاوه بی نهایت است. بنابراین در روابط زیر این حدود را نمی نویسیم.

$$\int e^{-\frac{a}{2}x^2} dx = \sqrt{\frac{2\pi}{a}}. \tag{۱۵۹}$$

$$\int e^{-\frac{a}{2}x^2+bx} dx = \sqrt{\frac{2\pi}{a}} e^{\frac{b^2}{2a}} \tag{۱۶۰}$$

با استفاده از این دو رابطه و تغییر مناسب نام پارامتر ها می توانیم تابع توزیع گاووسی را به صورت زیر بنویسیم:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}. \tag{۱۶۱}$$

و در نتیجه:

$$\int P(x)dx = 1 \qquad \int P(x)e^{bx}dx = e^{b^2\sigma^2}. \tag{۱۶۲}$$

با مشتق گیری از طرفین نسبت به پارامتر b و در پایان قرار دادن $b=0$ می توانیم گشتاورهای متوالی تابع توزیع گاووسی را بدست آوریم.

■ تمرین: نشان دهید که:

$$\langle x^2 \rangle = \sigma^2, \qquad \langle x^4 \rangle = 3\sigma^4 \tag{۱۶۳}$$