آغاز فصل سیزدهم

متغیرهای تصادفی، توزیع(ها) و امید ریاضی

۱. متغیرهای تصادفی

از احتمال برای نمونه سازی متغیر تجارب، بازی‌ها و آزمایش‌های پزشکی استفاده کردهایم. بطور کلی، کوشیده‌ایم احتمالات رویدادها را محاسبه کنیم. پرسیده‌ایم، برای مثال، چقدر احتمال دارد که در مسابقه مانتی هال برنده شوید؟ احتمال این رویداد که باران ببارد، مفروض براینکه مرد هواشناس چترش را امروز آورده باشد، چقدر است؟ چقدر احتمال این رویداد وجود دارد که شما یک بیماری نادر داشته باشید، فرض براینکه آزمایش مثبت باشد؟

ولی هر کسی می‌تواند سئوال‌های کلی‌تری بپرسد. به چه شدت باران خواهد بارید؟

این بیماری تا چند مدت ادامه خواهد یافت؟ چه میزان بازنده خواهم شد در صورتی که تمام روز با ۶.۰۴۲ بازی کنم؟ این پرسش‌ها اساساً متفاوت هستند و به شکل عبارت‌های رویداد به آسانی جمله‌بندی نشده‌اند. مسئله این است که یک رویداد خواه ناخواه اتفاق می‌افتد: یا می‌برید یا می‌بازید، یا می‌بارد یا نمی‌بارد، یا بیمار هستید یا بیمار نیستید. ولی این پرسش‌ها درباره موضوع درجه‌بندی مطرح می‌شوند: چه‌قدر، به چه‌شدت، چه‌مدت؟ برای نزدیک شدن به این پرسش‌ها، به ابزار ریاضی جدیدی نیاز داریم.

۱. ۱ تعریف

بیائید با یک مثال شروع کنیم. آزمایش بالا انداختن سه سکه مستقل، و غیر منصف را در نظر بگیرید. فرض کنید C تعداد سرهایی باشد که ظاهر می‌شوند. فرض کنید $M=۱$ اگر که سکه‌ها همگی شیر یا خط باشند و فرض کنید $M=۰$ در غیر اینصورت باشد. اینک هر پی‌آمدی از پرتاب سه سکه محتملاً مقادیری برای C و M را معین می‌کند. برای مثال، اگر شیر، خط، شیر، خط دهد، پس $C=۲$ و $M=۰$. اگر خط، خط، خط رخ دهد، پس $C=۰$ و $M=۱$. در نتیجه، C تعداد شیرها را می‌شمارد و M مقرر می‌دارد که آیا سکه‌ها با هم یکسان هستند.

از آنرو که هر پی‌آمدی محتملاً C و M را معین می‌کند، می‌توانیم آن‌ها را به صورت تابع‌هایی که پیشامدها را به اعداد نگاشت می‌کنند در نظر بگیریم. برای این آزمایش، فضای نمونه عبارت است از:

$$S=\{HHH,HHT,HTH,HTT,THH,THT,TTT\}$$

اینک C تابعی است که هر پیشامد در فضای نمونه را به یک عدد به شکل زیر می‌نگارد:

$$C(HHH)=۳ \qquad C(THH)=۲$$
$$C(HHT)=۲ \qquad C(THT)=۱$$
$$C(HTH)=۲ \qquad C(TTH)=۱$$
$$C(HTT)=۱ \qquad C(TTT)=۰$$

به طور مشابه، M تابعی است که هر پی‌آمدی را به روشی دیگر می‌نگارد:

$$M(HHH)=1 \qquad M(THH)=0$$

$$M(HHT)=0 \qquad M(THT)=0$$

$$M(HTH)=0 \qquad M(TTH)=0$$

$$M(HTT)=0 \qquad M(TTT)=1$$

تابع‌های C و M مثال‌هایی از متغیرهای تصادفی هستند. به طور کلی، یک متغیر تصادفی تابعی است که دامنه آن فضای نمونه است. (هم دامنه می‌تواند هر چیزی باشد، ولی ما معمولاً زیر مجموعه‌ای از اعداد حقیقی را استفاده خواهیم کرد. توجه داشته باشید که "متغیر تصادفی" اسمی بی مسمی است، متغیرهای تصادفی عملاً توابع هستند!)

٢. ١ متغیرهای تصادفی شاخص

یک متغیرهای تصادفی شاخص (یا به سادگی شاخص، یا متغیر تصادفی برنولی)

نوعی متغیر تصادفی است که هر پی‌آمدی را به ٠ یا ١ می‌نگارد. متغیر تصادفی M یک مثال است. اگر هر سه سکه یکسان باشند آنگاه $M=1$ و در غیر اینصورت $M=0$.

متغیرهای تصادفی شاخص تقریباً با رویدادها مرتبط هستند. بویژه، یک شاخص، فضای نمونه را به پی‌آمدهایی که به ١ نگاشت شده‌اند و آن پیشامدهایی که به ٠ نگاشت شده‌اند قسمت‌بندی می‌کند. برای مثال، شاخص M فضای نمونه را به دو تکه به شکل زیر تقسیم می‌کند:

$$\underbrace{HHH \quad TTT}_{M=1} \quad \underbrace{HHT \quad HTH \quad HTT \quad THH\, THT \quad TTH}_{M=0}$$

به همان روش، یک رویداد فضای نمونه را به آن پی‌آمدهای موجود در رویداد و آن پیشامدهایی که در رویداد نیستند تقسیم می‌کند. بنابراین، هر رویدادی طبیعتاً با یک متغیر تصادفی شاخص خاص مرتبط می‌شود و برعکس: یک شاخص برای یک رویداد E عبارت است از متغیر تصادفی شاخصی که برای تمام پیشامدهای موجود در E ، ۱ است و برای همه پیشامدهای غیر موجود در E، ۰ باشد. بر این اساس، M یک متغیر تصادفی شاخص برای رویدادی است که همه سه سکه‌ها با هم یکسان باشند.

۳. ۱ متغیرهای تصادفی و رویدادها

رابطه نیرومند خوبی میان رویدادها و متغیرهای تصادفی کلی‌تر وجود دارد. یک متغیر تصادفی که چندین مقدار را در بر می‌گیرد فضای نمونه را به چندین تکه قسمت می‌کند.

برای مثال، C به صورت زیر فضای نمونه را قسمت‌بندی می‌کند

$$\underbrace{TTT}_{C=0} \quad \underbrace{TTH \quad THT \quad HTT}_{C=1} \quad \underbrace{THH \quad HTH \quad HHT}_{C=2} \quad \underbrace{HHH}_{C=3}$$

هر قسمت زیر مجموعه‌ای از فضای نمونه است و بنابراین یک رویداد به حساب می‌آید. بر این اساس می‌توانیم یک تساوی یا عدم تساوی از یک متغیر تصادفی را به عنوان یک رویداد در نظر بگیریم. برای مثال، رویدادی که $C=2$ تشکیل می‌شود از پیشامدهای THH, HTH و HHT. رویداد $C \leq 1$ از پیشامدهای THT و TTH و TTT و HTT تشکیل می‌شود.

به اندازه کافی طبیعی، می‌توانیم درباره احتمال رویدادهایی که با معادله معین شده از متغیر تصادفی است صحبت کنیم. برای مثال:

$$\begin{aligned}\Pr\{C=2\} &= \Pr\{THH\}+\Pr\{HTH\}+\Pr\{HHT\}\\ &= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}\\ &= \frac{3}{8}.\end{aligned}$$

مثالی دیگری از این دست:

$$\begin{aligned}\Pr\{M=1\} &= \Pr\{TTT\}+\Pr\{HHH\}\\ &= \frac{1}{8}+\frac{1}{8}\\ &= \frac{1}{4}\end{aligned}$$

٤. ١ احتمال مشروط

درهم کردن احتمالات مشروط و رویدادهایی که متضمن متغیرهای تصادفی‌اند هیچ مشکل تازه‌ای به بار نمی‌آورد. برای مثال، $\{C\geq 2 \mid M=0\}$ احتمالی است که حداقل دو سکه با سر می‌آیند $(C\geq 2)$ مفروض برآنکه هر سه سکه یک نوع نباشند $(M=0)$. می‌توانیم با استفاده از تعریف احتمال مشروط این احتمال را محاسبه کنیم:

$$\Pr\{C\geq 2 \mid M=0\}=\frac{\Pr\{[C\geq 2]\cap[M=0]\}}{\Pr\{M=0\}}$$

$$=\frac{PR\{\{THH,HTH,HHT\}\}}{\Pr\{\{THH,HTH,HHT,HTT,THT,TTH\}\}}$$

$$=\frac{\frac{۳}{۸}}{\frac{۶}{۸}}$$

$$=\frac{۱}{۲}.$$

گزاره $[M=۰]\bigcap[c\geq۲]$ موجود در خط اول ممکن است نامأنوس به نظر برسد؛ مجموعه عمل $\bigcap$ میان تساوی و یک عدم تساوی چیست؟ ولی آن را به خاطر بیاورید، در این زمینه $[c\geq۲]$ و $[M=۰]$ رویداد هستند، به عنوان مثال، مجموعه‌هایی از پی‌آمدها.

۵. ۱ استقلال

مفهوم استقلال به خوبی از رویدادها به متغیرهای تصادفی منتقل می‌شود. متغیرهای تصادفی $R_۱$ و $R_۲$ مستقل هستند اگر برای تمام $x_۱$ در هم‌دامنه $R_۱$ و $x_۲$ در هم‌دامنه $R_۲$ داشته باشیم:

$$\Pr\{[R_۱=x_۱]\bigcap[R_۲=x_۲]\}=\Pr\{R_۱=x_۱\}.\Pr\{R_۲=x_۲\}.$$

به مانند رویدادها، می‌توانیم استقلال را برای متغیرهای تصادفی به یک روش هم‌ارز و شاید باطنی‌تر فورمول‌بندی کنیم: متغیرهای تصادفی $R_۱$ و $R_۲$ مستقل هستند اگر برای تمام $x_۱$ و $x_۲$ها به ترتیب در هم‌دامنه $R_۱$ و $R_۲$، بطوری که $\Pr\{R_۲=x_۲\}>۰$ داشته باشیم:

$$\Pr\{R_۱=x_۱\mid R_۲=x_۲\}=\Pr\{R۱=x۱\}.$$

به شکل کلمات: احتمال اینکه $R_۱$ مقداری ویژه را به عهده بگیرد تحت تأثیر $R_۲$ نیست.

به عنوان یک مثال، آیا C و M مستقل هستند؟ از روی حس باطنی پاسخ باید "نه" باشد. تعداد سرها، C ، کاملاً معین می‌کند که چه بسا سه سکه یکسان هستند؛ یعنی اینکه چه بسا $M=1$ باشد. ولی برای تعیین این حس درونی مقادیر $x_1, x_2 \in R$ پیدا کنیم به طوری که:

$$\Pr\{[C=x_1]\cap[M=x_2]\} \neq \Pr\{C=x_1\}.\Pr\{M=x_2\}.$$

یک انتخاب مناسب، $x_1=2$ و $x_2=1$ است. در این حالت، داریم:

$$\Pr\{[C=2]\cap[M=1]\}=0 \text{ ولی } \Pr\{C=2\}.\Pr\{M=1\}=\frac{3}{8}.\frac{1}{4}\neq 0.$$

اولین احتمال صفر است زیرا هرگز دقیقاً دو تا سر نداشتیم $(C=2)$ وقتی که هر سه سکه با هم یکسان باشند $(M=1)$. دو احتمال دیگر پیش از این محاسبه شدند.

مفهوم استقلال به مجموعه‌ای از متغیرهای تصادفی که در پائین آمده تعمیم پیدا می‌کند. متغیرهای تصادفی $R_1, R_2, \ldots, R_n$ مستقل هستند اگر برای همه $x_1, x_2, \ldots, x_n$ به ترتیب در هم‌دامنه‌های $R_1, R_2, \ldots, R_n$ داشته باشیم:

$$\Pr\{[R_1=x_1]\cap[R_2=x_2]\cap\ldots\cap[R_n=x_n]\}$$
$$=\Pr\{R_1=x_1\}.\Pr\{R_2=x_2\}\ldots\Pr\{R_n=x_n\}.$$

یک نتیجه تعریف استقلال این است که احتمال اینکه هر زیر مجموعه متغیرها با یک مجموعه ویژه مقادیر برابر است با حاصل‌ضرب احتمالات متغیرهای مفردی که مقادیرشان را به خود می‌گیرد. با این حساب، برای مثال، اگر $R_1, R_2, \ldots, R_{100}$ متغیرهای تصادفی مستقل باشند، از آن نتیجه می‌گیریم که :

$$\Pr\{[R_۱ = ۷]\bigcap[R_۷ = ۹.۱]\bigcap[R_{۲۳} = \pi]\} = \Pr\{R_۱ = ۷\}.\Pr\{R_۷ = ۹.۱\}.\Pr\{R_{۲۳} = \pi\}.$$

۲. اصل روز تولد

در یک تالار سخنرانی ۱۰۰ دانشجو حضور دارند. احتمال اینکه دو نفر یک روز تولد داشته باشند چه قدر است؟ ممکن است در حدود ۱/۳ ؟ بیائید بررسی کنیم! در خلال تجزیه خود از دو متغیر زیر استفاده خواهیم کرد:

- فرض کنید n تعداد آدم های گروه باشد.
- فرض کنید d تعداد روزهای سال باشد.

وانگهی، اینطور تصور می‌کنیم که روزهای تولد به صورت یک نواختی- توزیع شده‌اند و متغیرهای تصادفی مستقل باشند. این تصور واقعاً در جهان خارج معتبر نیست، از آن‌رو که فرزندان بسیاری در وقت معینی از سال به دنیا آمده‌اند و روز تولد دو قلوها به طور واضح مستقل نیست. با این وجود، تجزیه ما از این مسئله به کار بسیاری از وضعیت‌های علوم کامپیوتر که توسط دو قلوها، روز کبیسه و به هر حال روزهای رمانتیک، غیر اثر گذارند، می‌آیند، بنابراین روی آن پیچیدگی‌ها زندگی نمی‌کنیم.

فضای نمونه برای این آزمایش متشکل است از کلیه راه‌های واگذاری روز تولد به افراد گروه است. تعداد d^n از چنین واگذاری‌ها وجود دارد، از آن جا که اولین شخص می‌تواند d روز تولد مختلف داشته باشد، دومین شخص می‌تواند d روز تولد متفاوت داشته باشد و به همین

ترتیب. وانگهی، هر چنین واگذاری متساویاً با توجه به تصور ما که روزهای تولد به طور یکنواخت و مستقل توزیع شده‌اند محتمل است، بنابراین فضای نمونه یکنواخت است.

فرض کنید D رویدادی است که هر کسی روز تولد جداگانه‌ای داشته باشد. این متمم رویدادی است که به آن علاقه‌مندیم، ولی احتمال D برای ارزیابی آسان‌تر است.

بعداً می‌توانیم از این واقعیت که $\Pr\{\overline{D}\}=۱-\Pr\{D\}$ برای محاسبه احتمالی که واقعاً طالب آنیم استفاده کنیم. به هر حال، رویداد D از $d(d-۱)(d-۲)\ldots(d-n+۱)$ تعدا پیامد تشکیل می‌شود، از آن‌رو که می‌توانیم روز تولد اولین نفر را یکی از d روز انتخاب کنیم، روز تولد نفر دوم را در $d-۱$ روش و به همین ترتیب. بنابراین، احتمال اینکه هر شخصی روز تولد متفاوتی داشته باشد عبارت است از:

$$\Pr\{D\}=\frac{d(d-۱)(d-۲)\ldots(d-n+۱)}{d^n}$$

برای $n=۱۰۰$ این احتمال به طور خارق‌العاده‌ای کوچک است– کمتر از یک در میلیون! اگر ۱۰۰ نفر در یک اتاق باشند، تقریباً حتماً دو نفر هستند که در یک روز تولد سهیم باشند. بیائید برای بازنویسی طرف صحیح معادله پیشین به شکلی روشن بنیانه‌تر از تقریب استفاده کنیم:

$$\begin{aligned}\Pr\{D\}&=\left(۱-\frac{۰}{d}\right)\left(۱-\frac{۱}{d}\right)\left(۱-\frac{۲}{d}\right)\ldots\left(۱-\frac{n-۱}{d}\right)\\&\approx e^{۰}.e^{-۱/d}.e^{-۲/d}\ldots e^{-(n-۱)/d}\\&=e^{\frac{-n(n-۱)}{۲d}}.\end{aligned}$$

در مرحله اول، هر عبارت را در صورت کسر با یک عبارت d در مخرج کسر جفت می‌کنیم.

بعد، از تقریب $e^{-x} \approx ۱-x$ استفاده می‌کنیم که کاملاً درست است وقتی که x کوچک باشد.[۱]

در آخرین مرحله، با استفاده از فورمول آشنای $۰+۱+۲+\ldots+(n-۱)=n(n-۱)/۲$ توان‌ها را ترکیب می‌کنیم.

توان موجود در گزاره پایانی بالا ۱− است وقتی که $n \approx \sqrt{۲d}$. این شدیداً نکته سر به سر شدن است، جایی که احتمال اینکه دو نفر در یک روز تولد سهیم باشند برابر ۱/۲ است.

این به قانونی به نام اصل روز تولد می‌انجامد، که در بسیاری زمینه‌های علوم رایانه مفید است:

اگر d روز در یک سال باشد و $\sqrt{۲d}$ شخص در یک اتاق، آنگاه احتمال اینکه دو نفر در یک روز تولد سهیم شوند در حدود $۱-۱/e \approx ۰.۶۳۲$ است.

برای مثال، این اصل می‌گوید که اگر $\sqrt{۲۳۶۵} \approx ۲۷$ شخص در یک اتاق داشته باشید، آنگاه احتمال اینکه دو نفر در یک روز تولد سهیم باشند در حدود ۰.۶۳۲ است. احتمال فعلی در حدود ۰.۶۲۶ است، بنابراین تقریب کاملاً خوب است.

اصل روز تولد یک قاعده عملی با بسیاری از کاربردهای شگفت‌انگیز است. برای مثال، سیستم‌های رمزنگاری و طرح اولیه امضای الکترونیکی باید علیه "یورش‌های روز تولد" شدت پیدا کنند. اصل مورد نظر همچنین به ما می‌گوید که چه تعداد آیتم را می‌توان در یک جدول در هم‌ریخته (مخلوط) جای داد قبلی از آنکه افراد شروع به برخورد بکنند.

---

۱. این تقریب بوسیله کوتاه سازی سری تیلور $e^{-x} = ۱ - x + x^{۲}/۲! - x^{۳}/۳! + \ldots$ بدست می‌آید.

۳. توزیع احتمال

یک متغیر تصادفی عبارت بود از تابعی که دامنه آن فضای نمونه یک آزمایش باشد. اغلب، با این وجود، متغیرهای تصادفی اساساً با همان خصوصیات به صورت کامل در آزمایش‌های مختلف به نمایش در می‌آیند. برای مثال، تعدادی متغیر تصادفی که در رأی– گیری مطرح می‌شوند، در آزمایش‌های اولیه و در پرتاب سکه تماماً در تعدادی خصوصیات مشترک سهیم هستند. اگر می‌توانستیم چنین متغیرهای تصادفی را به صورت مجرد مطالعه کنیم، جدا از هر گونه آزمایش ویژه، آنگاه نتیجه‌گیری ما برای تمام آزمایش‌ها در جایی که آن نوع متغیر تصادفی رخ داده است به کار می‌رود. چنین نتیجه‌گیری‌های کلی بسیار سودمند خواهد بود. چندین ابزار وجود دارد که خصوصیات اساسی یک متغیر تصادفی را در بر می‌گیرد، ولی جزئیات دیگر آزمایش گردآوری شده را در پشت سر به جا می‌گذارد.

{تابع چگالی احتمال} (Pdf) برای یک متغیر تصادفی R با هم‌دامنه V تابعی است که بوسیله $PDF_R : V \to [\text{۰,۱}]$ تعیین می‌شود:

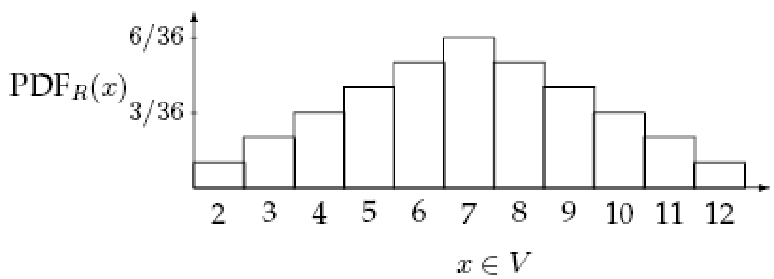
$$PDF_R(x) = \Pr\{R = x\}$$

نتیجه این تعریف این است که

$$\sum_{x \in V} PDF_R(x) = \text{۱}$$

از آن‌رو که متغیر تصادفی همیشه دقیقاً مقدار یک را روی مجموعه V می‌گیرد.

به عنوان مثال، بیائید به آزمایش پرتاب دو تاس مستقل، منصف برگردیم. مثل قبل، فرض کنید T مجموع دو پرتاب باشد. این متغیر تصادفی مقادیری در مجموعه $V=\{۲,۳,...,۱۲\}$ را پوشش می‌دهد. طرحی از تابع چگالی احتمال در پائین نشان داده می‌شود:

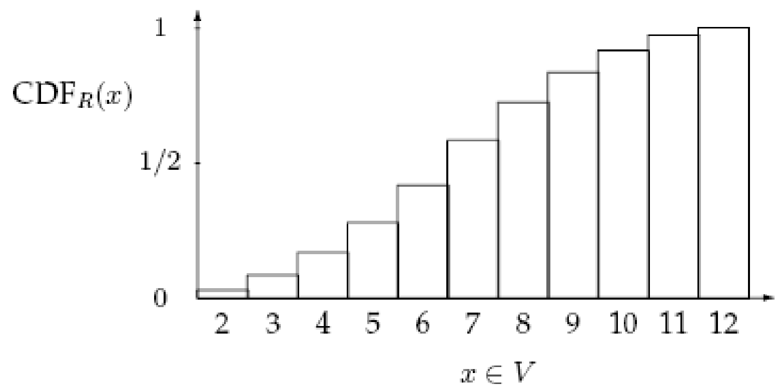


مجموع میانی مشخص می‌کند که حاصل‌جمع‌های نزدیک به ۷ محتمل‌ترین هستند. کل مساحت تمام مستطیل‌ها ۱ است زیرا که تاس‌ها باید دقیقاً یکی از حاصل‌جمع‌ها را در $V=\{۲,۳,...,۱۲\}$ پوشش بدهند.

یک ایده تقریباً- مربوط عبارت است از تابع توزیع جمع شونده (cdf) برای متغیر تصادفی R. این یک تابع است $CDF_R : V \to [۰,۱]$ که توسط ضابطه زیر مشخص می‌شود:

$$CDF_R(x)=\Pr\{R \le x\}$$

به عنوان یک مثال، تابع توزیع جمع‌شونده برای متغیر تصادفی T در پائین نمایش داده می‌شود:



ارتفاع، مستطیل iام در تابع توزیع جمع‌شونده برابر است با حاصل‌جمع مستطیل‌های سمت چپ i در تابع چگالی احتمال. که از تعریف pdf و cdf تبعیت می‌کند:

$$\begin{aligned} CDF_R(x) &= \Pr\{R \le x\} \\ &= \sum_{y \le x} \Pr\{R = y\} \\ &= \sum_{y \le x} PDF_R(y) \end{aligned}$$

بطور خلاصه، $PDF_R(x)$ احتمال اینکه $R = x$ را اندازه‌گیری می‌کند و $CDF_R(x)$ احتمال اینکه $R \le x$ را اندازه‌گیری می‌کند. هر دوی PDF_R و CDF_R همان اطلاعات یکسان درباره متغیر تصادفی R را تسخیر می‌کنند– می‌توانید یکی را از دیگری بدست آورید– ولی پاره‌ای اوقات یکی از آنها مناسب‌تر است. نکته کلیدی در اینجا این است که هیچکدام نه تابع چگالی احتمال نه تابع توزیع جمع شونده فضای نمونه یک آزمایش را متضمن نمی‌شوند. بنابراین، از خلال این تابع‌ها، بدون رجوع به آزمایشی ویژه می‌توانیم متغیرهای تصادفی را مطالعه کنیم.

اینک به سه توزیع مهم و برخی کاربری‌ها نگاه خواهیم کرد.

۱. ۳ توزیع برنولی

متغیرهای تصادفی شاخص به دلیل پیوند نزدیکشان با رویدادها شاید عمومی‌ترین نوع باشند. تابع چگالی احتمال یک متغیر تصادفی شاخص B همیشه عبارت است از

$$PDF_B(۰) = p$$

$$PDF_B(۱) = ۱ - p$$

جایی که $۰ \leq p \leq ۱$. تابع توزیع جمع‌شونده متناظر عبارت است از:

$$CDF_B(۰) = p$$

$$CDF_B(۱) = ۱$$

۲. ۳ توزیع یکنواخت

تابع متغیری که هر مقدار ممکن را با احتمال یکسان حمل کند یکنواخت نامیده می‌شود. برای مثال، تابع چگالی احتمال یک متغیر تصادفی U که در مجموعه $\{۱,۲...N\}$ یکنواخت باشد عبارت است از:

$$PDF_U(k) = \frac{۱}{N}$$

و تابع توزیع جمع‌شونده چنین است:

$$CDF_U(k) = \frac{k}{N}$$

توزیع‌های یکنواخت تمام مدت مطرح می‌شوند. برای مثال، احتمال عدد روشده روی یک تاس سالم در مجموعه $\{۱,۲,...,۶\}$ یکنواخت است.

۳. ۳ بازی اعداد

بیائید بازی کنیم! من دو تا پاکت دارم. هر کدام حاوی یک عدد صحیح در بین اعداد ۰,۱,...,۱۰۰ است و اعداد از هم جدا هستند. برای برنده شدن در بازی، باید تشخیص بدهید کدام پاکت دارای عدد بزرگتر است. برای اینکه شانس نبرد را به شما بدهم، اجازه می‌دهم که یک نگاه یواشکی به پاکتی که تصادفاً انتخاب شده بی‌اندازید.

آیا می‌توانید راه بردی اختراع کنید تا شانسی بهتر از ۵۰٪ پیروزی به شما بدهد؟

برای مثال، می‌توانید همین طوری یک پاکت را تصادفاً انتخاب کنید و حدس بزنید که حاوی عدد بزرگ‌تر است. ولی این راه برد (استراتژی) فقط در ۵۰٪ اوقات برنده است. چالش شما باید بهتر باشد.

بنابراین می‌توانید تلاش کنید هوشیارتر باشید. فرض کنید دزدکی به پاکت سمت چپ نگاه می‌کنید و عدد ۱۲ را می‌بینید. از آن‌رو که ۱۲ عدد کوچکی است، ممکن است حدس بزنید که عدد دیگر بزرگتر است.

ولی شاید من به نوعی حقه باز باشم و اعداد کوچکی را در هر دو پاکت بگذارم. پس حدستان چه بسا که خیلی خوب نباشد! اینجا یک نکته مهم این است که اعداد موجود در پاکت‌ها ممکن است تصادفی نباشند. در حال گزینش اعداد هستم و دارم آنها را به روشی انتخاب می‌کنم که فکر می‌کنم راه برد حدس زدن تان را شکست می‌دهم. من فقط از عمل تصادفی کردن برای انتخاب اعداد استفاده خواهم کرد اگر که به هدف پایانی‌ام کمک کند: باعث باخت شما بشود!

۱. ۳. ۳ حس باطنی در آنسوی راه‌برد برنده شدن

به طور شگفت‌آوری، راه‌بردی وجود دارد که در بیش از ۵۰٪ وقت‌ها پیروز می‌شود، صرف نظر از اینکه چه اعدادی را درون پاکت‌ها بگذاریم.

فرض کنید که به نحوی یک عدد x در میان اعداد پائینی و بالایی من را می‌دانید. حالا یک پاکت را بر می‌دارید و این یا آن عدد را می‌بینید. آن عدد از x بزرگتر است، سپس زیر چشمی به دنبال عدد بالایی هستید. اگر از x کوچکتر باشد، سپس زیر چشمی به دنبال عدد پائینی هستید، به سخن دیگر اگر یک عدد x میان اعداد پائینی و بالایی مرا بدانید، پس مطمئن هستید که بازی را می‌برید.

تنها عیب این راه‌برد درخشان این است که عدد x را نمی‌دانید. اوه بسیار خوب. ولی اگر تلاش کنید حدس بزنید چه می‌شود؟ کمی احتمال وجود دارد که به درستی حدس بزنید. در این حالت، در ۱۰۰٪ زمان برنده می‌شوید. از سوی دیگر اگر به نادرستی حدس بزنید، وضعتان بدتر از قبل نیست؛ همچنان از ۵۰٪ شانس برد برخوردار هستید. با ترکیب این دو حالت، شانس کلی‌تان برای برنده شدن بهتر از ۵۰٪ است!

بگو مگوهای غیر رسمی درباره احتمال، مثل این یکی، اغلب پذیرفتنی به نظر می‌رسند، ولی تحت بررسی از نزدیک درست از آب در نمی‌آیند. در مقابل، این بحث و جدل کاملاً غیرقابل قبول به نظر می‌رسد– ولی در واقع صحیح است!

۲. ۳. ۳ تحلیل راه‌برد برنده شدن

به طور کلی، فرض کنید که می‌توانم اعدادی را از مجموعه $\{0,1,\ldots,n\}$ انتخاب کنم. عدد پائینی را L و عدد بالایی را H بنامید.

هدفتان حدس زدن یک عدد x میان L و H است. برای اجتناب از حالت‌های متساوی به شکل تصادفی x را از میان اعداد نیم– صحیح انتخاب می‌کنید:

$$\left\{\frac{1}{2},1\frac{1}{2},2\frac{1}{2},\ldots,n-\frac{1}{2}\right\}$$

ولی از چه توزیع احتمالی باید استفاده کنید؟

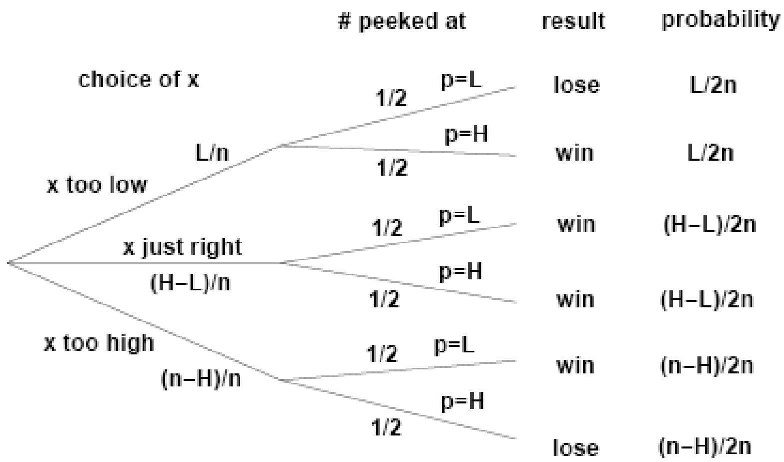
توزیع یکنواخت به این نتیجه مطلوب می‌رساند که بهترین شرط‌بندی شماست. یک توجیه غیر رسمی این است که اگر تصور کنم که شما به طور غیر محتمل عددی را بردارید– فرضاً $50\frac{1}{2}$– پس همیشه ۵۰ و ۵۱ را در پاکت‌ها قرار می‌دادم.

آنگاه به طور غیر محتمل عدد x میان L و H انتخاب می‌کردید و شانس کمتری برای برنده شدن خواهید داشت.

بعد از اینکه عدد x را انتخاب کردید، پواشکی درون یک پاکت را می‌بینید و یک عدد P را می‌بینید. اگر $P > x$ پس حدس می‌زنید که دارید به عدد بزرگ‌تر نگاه می‌کنید. اگر $P < x$ پس حدس می‌زنید که عدد دیگر بزرگ‌تر است.

همه آنچه باقی می‌ماند این است که تشخیص احتمال موفقیت این روش را بدهیم. می‌توانیم این کار را با شیوه معمولی چهار– مرحله‌ای و نمودار درختی انجام دهیم.

مرحله ۱: فضای نمونه را پیدا کنید. شما یا x را خیلی کم $(<L)$ خیلی زیاد $(>H)$ یا به درستی $(L<x<H)$ انتخاب می‌کنید. سپس یواشکی به عدد پائینی $(P=L)$ یا به عدد بالایی $(P=H)$ نگاه می‌کنید. این جریان در مجموع شش پی‌آمد ممکن را فراهم می‌سازد.



مرحله ۲: رویدادهای منفعت را معین کنید. چهار پی‌آمد در رویدادی که شما برنده می‌شوید در نمودار درختی علامت‌گذاری شده‌اند.

مرحله ۳: احتمالات پی‌آمد را اختصاص دهید. ابتدا، احتمالات یال را مشخص می‌کنیم. حدس x شما با احتمال L/n خیلی پائین، با احتمال $(n-H)$ خیلی بالا و با احتمال $(H-L)/n$ درست و به اندازه است. بعد، با احتمال برابر یواشکی به عدد پائینی یا عدد بالایی نگاه می‌کنید. ضرب کردن در طول مسیر ریشه– به– برگ احتمالات پی‌آمد را بدست می‌دهد.

مرحله ۴: احتمالات رویداد را محاسبه کنید. احتمال رویدادی که برنده می‌شوید عبارت از احتمالات چهار پی‌آمد موجود در آن رویداد است:

$$\Pr\{\text{برنده شدن}\}=\frac{L}{۲n}+\frac{H-L}{۲n}+\frac{H-L}{۲n}+\frac{n-H}{۲n}$$

$$= \frac{1}{2} + \frac{H-L}{2n}$$

$$\geq \frac{1}{2} + \frac{1}{2n}$$

آخرین عدم تساوی بر این واقعیت متکی است که عدد بالاتر H حداقل ۱واحد بزرگتر از عدد پائینی L است از آن‌رو که باید از هم متمایز باشند.

کاملاً مطمئن، با این راه‌برد بیش از نیمی از وقت برنده می‌شوید، صرف‌نظر از اعدادی که در پاکت‌ها باشند! برای مثال، اگر عددهایی در برد ۰,۱,...,۱۰۰ را انتخاب کنم، آنگاه حداقل به احتمال $\frac{1}{2} + \frac{1}{200} = 50/5\%$ برنده می‌شوید. حتی بهتر، اگر که مجاز باشم اعداد بین ۰,...,۱۰ را انتخاب کنم، آنگاه احتمال برد شما تا ۵۵٪ بالا می‌رود. بر اساس معیارهای لاس و گاس، آنها اعداد فرد بزرگ هستند!

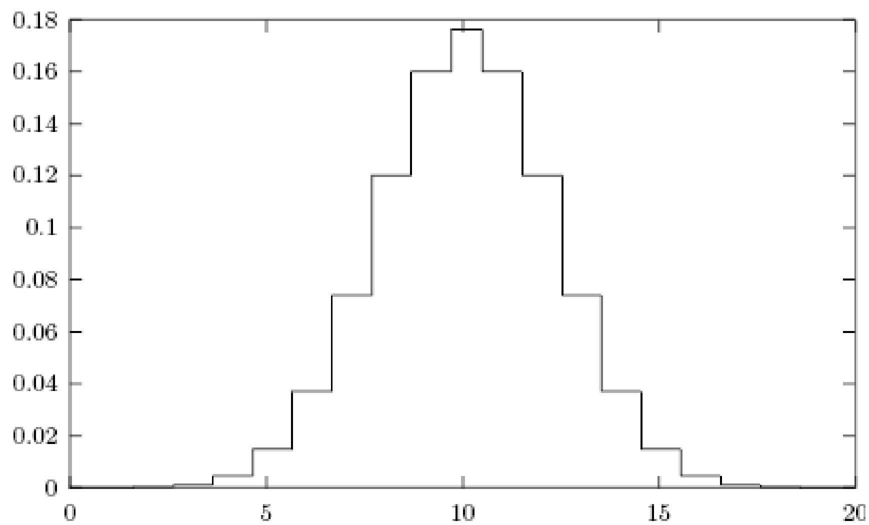
٤. ۳ توزیع دو جمله‌ای

نسبت به توزیع‌های پیچیده‌تر، توزیع دو جمله‌ای بطور حتم مهم‌ترین توزیع در علوم رایانه است. مثال استاندارد متغیر تصادفی با یک توزیع دو جمله‌ای تعداد سرهایی است که در n پرتاب مستقل از یک سکه ظاهر می‌شوند؛ این متغیر تصادفی را H بنامید. اگر سکه درست باشد، آنگاه H دارای یک تابع چگالی دو جمله‌ای منصف به شکل زیر است:

$$PDF_H(k) = \binom{n}{k} 2^{-n}$$

زیرا تعداد دنباله‌ای nتایی که دقیقاً k سر در آنها رخ داده است $\binom{n}{k}$ است که هر چنین دنباله‌ای نیز با احتمال ۲^{-n} رخ می‌دهد.

در اینجا طرحی از تابع چگالی احتمال منصف $PDF_H(k)$ متناظر با $n=۲۰$ پرتاب سکه وجود دارد. محتمل‌ترین پی‌آمد عبارت از $k=۱۰$ سر است و احتمال به سرعت برای مقادیر بزرگتر و کوچکتر k نزول پیدا می‌کند. این مناطق سقوط به طرف چپ و راست بر روی برآمدگی را غالباً دنباله‌های توزیع می‌نامند.



تعداد بی‌شماری از تحلیل‌های علوم رایانه وجود دارند تا ثابت کنند که دنباله‌های دو جمله‌ای و توزیع‌های مشابه بسیار کوچکند. در متن یک مسئله، نوعاً به این معناست که شرایطی با احتمال بسیار کوچک اتفاق می‌افتد که باعث می‌شود یک سرویس دهنده یا خط ارتباطی افزودن باری

بیش از حد تحمل را بدوش کشد باشد یا یک الگوریتم تصادفی بصورت نمایی زمان صرف کند یا ایجاد کننده نتیجه غلط باشد.

۱. ٤. ۳ توزیع دو جمله‌ای عام

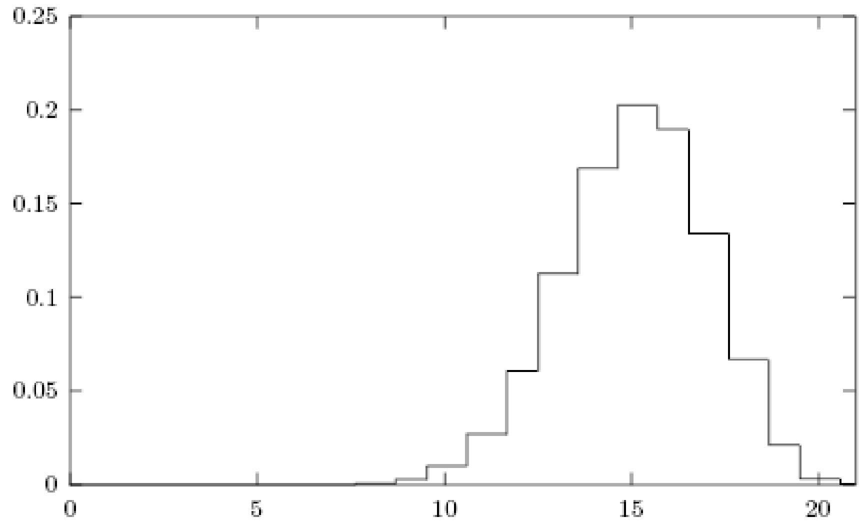
حالا فرض کنید J تعداد سرهایی باشد که در پرتاب n سکه مستقل به دست می‌آید، که هر کدامشان با احتمال p رخ می‌دهد. سپس J یک تابع چگالی دو جمله‌ای عام دارد:

$$PDF_J(k)\binom{n}{k}p^k(۱-p)^{n-k}$$

مثل قبل، تعداد $\binom{n}{k}$ دنباله با k سر و $n-k$ سر وجود دارند، ولی حالا احتمال هر چنین دنباله‌ای عبارت است از $p^k(۱-p)^{n-k}$.

به عنوان یک مثال، طرح پائین تابع احتمال چگالی $PDF_J(k)$ متناظر با پرتاب $n=۲۰$ سکه مستقل که با احتمال $p=۰/۷۵$ سر می‌آید را نشان می‌دهد.

گراف نشان می‌دهد که ما به احتمال زیاد حدود $k=۱۵$ سر بدست می‌آوریم، همانطور که ممکن است توقع داشته باشید. یکبار دیگر، احتمال سریعاً برای مقادیر بزرگ‌تر و کوچک‌تر k تنزل پیدا می‌کند.



۲. ٤. ۳ تقریب تابع چگالی دو جمله‌ای

یک فورمول شکل بسته– تقریبی برای تابع چگالی دو جمله‌ای عام وجود دارد، اگر چه کمی ثقیل است. ابتدا، به یک تقریب برای عبارت کلیدی در فورمول دقیق $\binom{n}{k}$ نیازمندیم. برای راحتی، بیائید k را با an عوض کنیم در جایی که a عددی میان ۰ و ۱ است. سپس، با توجه به فورمول استرلینگ، می‌یابیم که:

$$\binom{n}{an} \leq \frac{۲^{nH(a)}}{\sqrt{۲\pi a(۱-a)^n}}$$

که $H(a)$ تابع آنتروپی معروف است:

$$H(a) ::= a\log_{۲}\frac{۱}{a} + (۱-a)\log_{۲}\frac{۱}{۱-a}$$

این کران بالا بر $\binom{n}{an}$ بسیار محکم است و به شکل تقریبی عالی سرویس می‌دهد. حالا بیائید این فورمول را به درون تابع چگالی دو جمله‌ای فرو کنیم. احتمال رخداد an سر در n پرتاب سکه‌ای که هر سر با احتمال p رخ می‌دهد عبارت است از:

$$PDF_J(an) \leq \frac{۲^{nH(a)}}{\sqrt{۲\pi a(۱-a)^n}} . p^{an}(۱-p)^{(۱-a)n} \qquad (۱)$$

این فورمول مثل یک کفش چوبی زشت است، ولی بسیار سودمند است. برای مثال، فرض کند یک سکه منصف را n دفعه پرتاب می‌کنیم. احتمال اینکه دقیقاً $\frac{۱}{۲}n$ سر گیرمان بیاید چقدر است؟ فرو بردن $a = ۱/۲$ و $p = ۱/۲$ به درون این فورمول نتیجه می‌دهد:

$$PDF_J(an) \leq \frac{۲^{nH(\frac{۱}{۲})}}{\sqrt{۲\pi(۱/۲)(۱-(۱/۲))n}}$$

$$= \sqrt{\frac{۲}{\pi n}}$$

بنابراین برای مثال اگر یک سکه منصف را ۱۰۰ دفعه بالا بیاندازیم، دقیقاً احتمال بدست آوردن ۵۰ سر در حدود $۱/\sqrt{۵۰} \approx ۰.۰۷۹$ یا در حدود ۸ درصد است.

۵. ۳ تقریب تابع توزیع جمع‌شونده دو جمله‌ای

فرض کنید سکه‌ای را که احتمال رخداد سر p باشد در اختیار داریم. مثل قبل، فرض کنید متغیر تصادفی J تعداد سرهایی باشد که در n پرتاب مستقل اتفاق می‌افتد.

آنگاه احتمال بدست آوردن حداکثر k سر توسط تابع توزیع جمع‌شونده دو جمله‌ای فراهم می‌شود:

$$CDF_J\left(k\right)=pr\left\{J\le k\right\}$$

$$=\sum_{i=\circ}^{k}PDF_J\left(i\right)$$

$$=\sum_{i=\circ}^{K}\binom{n}{i}p^i\left(\text{۱}-p\right)^{n-i}$$

برای ارزیابی این گزاره برحسب بزرگی n,k، مقدار زحمت بیشتری باید متحمل شویم، بنابراین یک تقریب، واقعاً سودمند خواهد بود. یکبار دیگر، می‌توانیم فرض کنیم $k=an$؛ یعنی اینکه، بجای اینکه به تعداد قطعی سرها ، k، فکر کنیم، به کسری از پرتاب‌هایی که سر باشند ، (a)، توجه می‌کنیم. تقریب زیر مشروط بر اینکه $a<p$ باشد صادق است:

$$CDF_J\left(an\right)\le\frac{\text{۱}-a}{\text{۱}-a/p}.PDF_J\left(an\right)$$

$$\le\frac{\text{۱}-a}{\text{۱}-a/p}.\frac{\text{۲}^{nH(a)}}{\sqrt{\text{۲}\pi a\left(\text{۱}-a\right)^n}}.p^{an}\left(\text{۱}-p\right)^{(\text{۱}-a)n}$$

در مرحله اول، با توجه به حاصل‌جمع هندسی کرانه بالا را پیدا کرده و فورمول مورد نظر را برای حاصل‌جمع دنباله‌های هندسی به کار می‌بریم. (جزئیات متأثر کننده‌اند و حذف می‌شوند.) آنگاه فورمول تقریبی (۱) برای $PDF_J\ (an)$ را از بخش قبلی بکار می‌گیریم.

برای ارزیابی کردن این فورمول برای مقادیر معین a، p و n باید کلیدهای زیادی را روی یک ماشین حساب فشار دهید. (حتی محاسبه $H(a)$ مقدار نسبتاً خوبی کار است!) ولی برای n بزرگ، ارزیابی تابع توزیع جمع‌شونده بطور گسترده‌ای کاری بیشتری می‌طلبد! بنابراین به هدیه دعای قبل از غذا پیش از آنکه از دهان خارج شوند نگاه نکنید.

به عنوان یک مثال، احتمال اتفاق به حداکثر ۲۵ سر در ۱۰۰ پرتاب یک سکه منصف با قراردادن $n=۱۰۰, a=۱/۴, p=۱/۲$ بدست آورده می‌شود:

$$CDF_j\left(\frac{n}{۴}\right) \le \frac{۱-(۱/۴)}{۱-(۱/۴)/(۱/۲)}.PDF_J\left(\frac{n}{۴}\right) \le \frac{۳}{۲}.۱.۹۱۳.۱۰^{-۷}$$

این بیان می‌کند که اتفاق ۲۵ بار سر یا کمتر تا حد زیادی غیر محتمل است، که با قضیه قبلی ما که دنباله‌های توزیع چند جمله‌ای بسیار کوچکند در هماهنگ است. در واقع، توجه داشته باشید که احتمال بالا انداختن ۲۵ سر یا کمتر فقط ۵۰٪ بیشتر از احتمال دقیقاً ۲۵ بار سرآوردن است. به این ترتیب، احتمال برآوردن دقیقاً ۲۵ سر دو برابر اندازه احتمال برآوردن هر عددی میان ۰ و ۲۴ است!

پیش آگهی: کران بالای روی $CDF_J(an)$ فقط در صورتی صادق است که $a < p$. اگر این مورد مسئله شما نباشد، پس بکوشید به عبارت‌های متمم فکر کنید؛ یعنی اینکه، به تعداد دم‌هایی که آمده به جای تعداد سرها نگاه کنید. در مثال ما، احتمال آوردن ۷۵ سر یا بیشتر همان احتمال ۲۵ دُم یا کمتر آوردن است. با توجه به تجزیه بالا، این هم تا حد زیادی کوچک است.

٦. ٣ رأی گیری

فرض کنید می‌خواهیم کسر جمعیت رأی دهنده $U.S$ (آمریکا) که هیلاری کلینتیون را به رودی جولیانی ترجیح می‌دهند را در انتخابات ریاست جمهوری سال تخمین بزنیم. در نظر بگیرید این کسر نامعلوم باشد. بیائید فرض کنیم تعدادی فرآیند تصادفی داریم– مثلاً پرتاب دارت به فهرست ثبت رأی دهنده باشد– که هر رأی– دهنده را با احتمال متساوی انتخاب می‌کند. می‌توانیم متغیر برنولی را معین کنیم، K، با توجه به این قانون که $K=١$ اگر رأی دهنده تصادفی بیشترین ترجیحش کلینتون باشد و در غیر این صورت $K=٠$.

اینک برای محاسبه p ، یک عدد بزرگ، n، از رأی دهندگان را به تصادف بر می‌داریم و کسر کسانی که به کلینتون لطف دارند را محاسبه می‌کنیم. یعنی اینکه، متغیرهای $K_١,K_٢,\ldots,$ را تعیین می‌کنیم، جایی که K_i به این معنی است که متغیر شاخصی است برای رویدادی که iامین نفر تا این رأی دهنده انتخاب کلینتون را ترجیح می‌دهد.

از آن‌رو که گزینه‌های ما مستقلاً ساخته می‌شوند، K_iها مستقل هستند. بنابراین به طور رسمی، فرآیند محاسبه خود را به سادگی با تصور اینکه متغیرهای مستقل برنولی $K_١,K_٢,\ldots$ را داریم که هر کدام با احتمال ، p، ١ می‌شود، انجام می‌دهیم حالا فرض کنید S_n حاصل‌جمع آنها باشد، یعنی که،

$$S_n ::= \sum_{i=١}^{n} k_i \qquad (٢)$$

بنابراین S_n دارای توزیع دو جمله‌ای با متغیر معلوم n و متغیر نامعلوم p است.

متغیر S_n/n کسر رأی دهندگان در نمونه ما که به کلینتون لطف دارند را توصیف می‌کند. می‌توانیم انتظار داشته باشیم که S_n/n چیزی شبیه p باشد. از مقدار نمونه S_n/n به عنوان تخمین آماری p استفاده خواهیم کرد. بخصوص، تصور کنید می‌خواهیم که بر آوردمان با فاصله حداکثر ۰.۰۴ از p در حداقل ۹۵٪ اوقات باشد. مثلاً می‌خواهیم

$$pr\left\{\left|\frac{S_n}{n}-p\right|\leq ۰.۰۴\right\}\geq ۰.۹۵$$

فرض کنید ε حاشیه خطایی باشد که می‌توانیم ندیده بگیریم و فرض کنید δ احتمالی باشد که نتیجه ما خارج از این حاشیه قرار می‌گیرد. ما علاقه‌مند به داشتن ۰.۰۴=ε و $\delta\leq$۰.۰۵ هستیم، ولی اشتقاق در صورتی که وارد کردن این مقادیر را تا به پایان به تعویق بی‌اندازیم روشن‌تر خواهد بود.

می‌خواهیم تعداد ،n، زمان‌هایی که باید رای دهندگان را فرا بخوانیم تا اینکه مقدار S_n/n، متعلق به تخمین ما، با حداقل احتمال ۱-δ به فاصله حداکثر ε از کسر موجود به ملتی که به کلینتون لطف دارند، باشد.

δ را تعریف می‌کنیم، احتمال اینکه رأی‌گیری ما با بیش از حاشیه اشتباه ε به شکل زیر مشخص شود:

$$\delta=\Pr\underbrace{\left\{\frac{S_n}{n}\leq p-\varepsilon\right\}}_{\text{بسیاری در نمونه "جولیانی" را ترجیح می‌دهند}}+\Pr\underbrace{\left\{\frac{S_n}{n}\geq p+\varepsilon\right\}}_{\text{بسیاری در نمونه "کلنبتون" را ترجیح می‌دهند}}$$

$$=\Pr\left\{S_n\leq(p-\varepsilon)n\right\}+\Pr\left\{S_n\geq(p+\varepsilon)n\right\}$$

حالا

$$CDF_{S_n}\left(\left(p-\varepsilon\right)n\right) ::= \Pr\left\{S_n \le \left(p-\varepsilon\right)n\right\}$$

همچنین،

$$\Pr\left\{S_n \ge \left(p+\varepsilon\right)n\right\} = \Pr\left\{n-S_n \le \left(\left(۱-p\right)-\varepsilon\right)n\right\}.$$

ولی $Tn ::= n-S_n$ به سادگی تعداد رای دهندگان در نمونه است که جولیانی را ترجیح می‌دهند، که حاصل‌جمعی از متغیرهای تصادفی برنولی با متغیر p - ۱ است و بنابراین

$$\Pr\{\ Tn \le \left(\left(۱-p\right)\varepsilon\right) = CDF_{Tn}\left(\left(\left(۱-p\right)-\varepsilon\right)n\right).$$

از این جهت

$$\delta = CDF_{S_n}\left(\left(p-\varepsilon\right)n\right) + CDF_{Tn}\left(\left(\left(۱-p\right)-\varepsilon\right)n\right). \tag{۳}$$

بنابراین می‌توانیم یافتن یک ترتیب خوب از اندازه نمونه مورد نیاز را به کرانه‌های مناسب بر دو توزیع تجمعی دو جمله‌ای با متغیراتی به ترتیب p و p - ۱ کاهش دهیم.

استفاده کرانه برای تابع توزیع چگالی تجمعی به ما اجازه می‌دهد گزاره (۳) را برحسب عبارت‌های n،ε و p را محاسبه کنیم. مسئله این است که این کرانه شامل p، کسر آمریکائی‌هایی خواهد شد که کلینتون را ترجیح می‌دهند. این همان عدد نامعلومی است که داریم سعی می‌کنیم با رأی‌گیری مشخص کنیم! خوشبختانه، یک راه ساده برای بیرون رفتن از این دور وجود دارد. از آن‌رو که (۳) در p متقارن است، یک نقطه عطف وقتی که $p=۱/۲$ دارد و این نقطه عطف، در واقع، ماکسیمم آن است:

واقعیت BF/: برای تمام n، ε، ماکسیمم مقدار برای δ در معادله (۳) وقتی است که $p=۱/۲$.

به سخن دیگر، دنباله‌های دو جمله‌ای به آرامی وقتی که $P=۱/۲$ باشد تنزل پیدا می‌کنند. با استفاده از این واقعیت و وارد کردن به درون معادله‌هایی برای $CDF_{Tn}\left(\left(P-\in\right)n\right)$ و $CDF_{Tn}\left(\left(\left(۱-P\right)-\in\right)n\right)$ فرضیه زیر بدستمان می‌رسد:

قضیه ۱. ۳ (نمونه‌گیری دو جمله‌ای) BF. فرض کنید $K_۱, K_۲, \ldots,$ دنباله‌ای از متغیرهای تصادفی مستقل با مقادیر ۰,۱ با امید یکسان p باشند فرض کنید

$$S_n ::= \sum_{i=۱}^{n} K_i .$$

سپس، برای $۱/۲ > \varepsilon > ۰$

$$\Pr\left\{\left|\frac{S_n}{n}-p\right| \ge \varepsilon\right\} \le \frac{۱+۲\varepsilon}{۲\varepsilon}.\frac{۲^{-n\left(۱-H\left(\left(۱/۲\right)-\varepsilon\right)\right)}}{\sqrt{۲\pi\left(۱/۴-\varepsilon^{۲}\right)n}} \tag{٤}$$

می‌خواهیم $\varepsilon=۰.۰۴$، بنابراین وارد کردن به (٤) می‌دهد

$$\delta \le ۱۳.۵\frac{۲^{-n(۰.۰۰۴۶۲)}}{۱.۲۴۹۲\sqrt{n}} \tag{٥}$$

جایی که δ احتمال این است که محاسبه ما درون ε از p نیست. می‌خواهیم به اندازه کافی نمونه از مردم انتخاب کنیم تا اینکه $\delta \le ۰.۰۵$. آسان‌ترین راه برای یافتن اندازه نمونه لازم n

وارد کردن مقادیر برای n است تا کوچک‌ترین شان جائیکه در سمت راست (۵) عبارت از ۰.۰۵≥ است را پیدا نمود:

آراء مردم $n =$	کرانه بالایی بر احتمال رأی اشتباه است		
500	9.7%		
600	6.4%		
623	5.9%		
650	5.3%		
664	5.0%	$\leftarrow$	our poll size
700	4.3%		

بنابراین در ۹۵٪ اوقات، رأی‌گیری از ۶۶۴ نفر نتیجه خواهد داد کسری را که درون ۰.۰۴ از کسر فعلی رأی دهندگان است که کلینتون را ترجیح می‌دهند.

یک نکته قابل توجه این است که جمعیت کشور هیچ تأثیری بر حجم آراء مأخوذه ندارد. خواه در کشور هزار نفر یا یک میلیارد وجود داشته باشد، رأی دادن چند صد نفر کافی است!

این شیوه محاسبه (تخمین) با نمونه‌گیری از یک کمیت— در این مثال رأی ترجیح داده شده— تکنیکی است که آشکارا می‌تواند برای برآورد بسیاری از کمیت‌های نامعلوم دیگر مفید واقع شود.

مسئله ۱. توضیح نمونه‌گیری برای هیئت منصفه

همین تازگی نشان داده‌ایم که صرفاً نمونه‌گیری از ۶۶۲ رأی دهنده این کسر را نتیجه خواهد داد که ۹۵٪ اوقات، درون ۰.۰۴ کسر فعلی رأی دهندگانی که کلینتون را ترجیح می‌دهند قرار دارد. حجم واقعی جمعیت رأی دهنده (۱۰ها میلیون) ملاحظه نشد زیرا اهمیت نداشت.

فرض کنید به عنوان یک کارشناس در یک دادگاه می‌خواهید خدمت کنید. چگونه توضیح خواهید داد که تعداد واجدین شرایط رأی دادن بستگی به تعداد جمعیت ندارد؟

٤. سطوح اعتماد

فرض کنید یک ناظر انتخابات نمونه‌ای از ۶۶۲ رأی دهنده تصادفی را به کار می‌برد تا کسر رأی دهندگان به کلینتون را محاسبه کند و ناظر انتخابات در می‌یابد که ۳۶۴تای آنها کلینتون را ترجیح می‌دهند. اغواگرانه ولی کثیف است، تا اقرار به این شود که:

قضیه اشتباه. به احتمال ۰.۹۵، کسر p، رأی دهندگانی که کلینتون را ترجیح می‌دهند ۳۶۴/۶۶۲±۰.۰۴ است. از آن‌رو که $\frac{۳۶۴}{۶۶۲}-۰/۰۴>۰/۵$، ۹۵٪ شانس وجود دارد که بیش از نیمی از رأی دهنده‌گان کلینتون را ترجیح بدهند.

آنچه که درباره این گفته قابل اعتراض است این است که درباره احتمال یا "بخت" حرف می‌زند که یک جهان واقعی حقیقت دارد، مثلاً اینکه در کسر فعلی p، مربوط به رأی دهندگانی که کلیبتون را ترجیح می‌دهند بیشتر از ۰.۵۰ است. ولی p عبارت است از چیزی که وجود دارد و به سادگی معنایی ندارد که درباره احتمال اینکه چیزی دیگری است حرفی زده شود. برای مثال، فرض کنید p عملاً ۰.۴۹ است، پس پرسیدن درباره احتمال اینکه حداکثر ۰/۰۴ از ۳۶۴/۶۶۲ فاصله دارد بی‌معنی است— به سادگی قرار ندارد

خلاصه‌ای دقیق‌تر از آنچه کامل کرده‌ایم به این ترتیب است:

یک راه و روش احتمالی برای محاسبه مقدار کسر واقعی، p را شرح داده‌ایم. احتمال اینکه روش محاسبه ما به یک مقدار با فاصله حداکثر ۰/۰۴ از p منجر می‌شود عبارت است از ۰.۹۵.

کمی دهان پرکن است، بنابراین عبارت سازی ویژه، کمی کثیف، به طور مشترک استفاده می‌شود. ناظر انتخاباتی نتیجه‌گیری‌اش را با گفتن این توضیح می‌دهد که

در ۹۵٪ سطح اعتماد، کسر رأی دهندگانی که کلینتون را ترجیح می‌دهند عبارت است از $۳۶۴/۶۶۲ \pm ۰.۰۴$.

بخاطر سپردن اینکه سطوح اعتماد که به نتایج محاسبه روش‌ها در مقادیر جهان واقعی بر می‌گردد، مهم است. مورد محاسبه قرار دادن کمیت جهان واقعی نوعاً نامعلوم است، ولی ثابت است؛ یک متغیر تصادفی نیست، بنابراین بی‌معنی می‌نماید که درباره احتمال اینکه خواصی را دارا باشد حرف زده شود.

۵. امید ریاضی

انتظار یا امید یک متغیر تصادفی عددی است که درباره رفتار متغیر حرف زیادی به شما می‌زند. بطور برجسته، انتظار عبارت است از میانگین مقدار، جایی که هر مقدار بر اساس احتمالی که با آن می‌آید وزن داده می‌شود. به طور رسمی، مقدار مورد انتظار یک متغیر تصادفی R که بر فضای نمونه S تعریف می‌شود عبارت است از:

$$E[R] = \sum_{w \in S} R(w) \Pr\{w\}$$

برای سپاسگزاری از معنایش، در نظر بگیرید که S عبارت از مجموعه دانشجویان یک کلاس است، ما به صورت یکنواختی تصادفاً یک دانشجو را انتخاب کرده‌ایم. فرض کنید R نمره امتحانی دانشجوی منتخب باشد. آنگاه $E[R]$ درست میانگین کلاس است– اولین چیزی که

هرکسی بعد از اینکه ورقه امتحان‌شان را پس دادند می‌خواهد بداند! به همان روش، انتظار معمولاً اولین چیزی است که یکی بخواهد درباره هر متغیر تصادفی تعیین کند.

بیائید با یک مثال تمرین کنیم. فرض کنید R عددی باشد که بر روی یک تاس منصف، شش وجهی واقع باشد. سپس مقدار انتظار R عبارت است از:

$$E[R] = \sum_{k=۱}^{۶} k.\frac{۱}{۶}$$

$$= ۱.\frac{۱}{۶} + ۲.\frac{۱}{۶} + ۳.\frac{۱}{۶} + ۴.\frac{۱}{۶} + ۵.\frac{۱}{۶} + ۶.\frac{۱}{۶}$$

$$= \frac{۷}{۲}$$

این محاسبه نشان می‌دهد که نام "مقدار انتظار" کمی گمراه کننده است؛ متغیر تصادفی چه بسا هرگز آن مقدار را عملاً حمل نکند. بر روی یک تاس عادی نمی‌توانید $۳\frac{۱}{۳}$ بیاورید!

۱. ۵ تعریف‌های هم‌ارز انتظار

تعدادی راه‌های دیگر برای نگارش تعریف انتظار وجود دارند. پاره‌ای اوقات استفاده از یکی از این فورمول بندی‌های دیگر می‌تواند محاسبه انتظار را آسان‌تر کند. یک گزینه گرد آوردن تمام پی‌آمدهایی است که متغیر تصادفی همان مقدار را حمل می‌کند.

فرضیه ۱. ۵

$$E[R] = \sum_{x \text{ در برد } R} x.\Pr\{R = x\}$$

برهان. سمت چپ را به راست منتقل خواهیم کرد. فرض کنید $[R=x]$ رویدادی باشد که $R=x$.

$$
\begin{aligned}
E[R] &= \sum_{w\in S} R(w)\Pr\{w\} \\
&= \sum_{x\ \text{دربرد}\ R}\ \sum_{w\in[R=x]} R(w)\Pr\{w\} \\
&= \sum_{x\ \text{دربرد}\ R}\ \sum_{w\in[R=x]} x\Pr\{w\} \\
&= \sum_{x\ \text{دربرد}\ R} \left(x.\sum_{w\in(R=x)} \Pr\{w\} \right) \\
&= \sum_{x\ \text{دربرد}\ R} x.\Pr\{R=x\}
\end{aligned}
$$

در خط دوم، ما یک حاصل‌جمع را به دو قسمت تبدیل می‌کنیم. حاصل‌جمع بیرونی تمام مقادیر ممکن x که متغیر تصادفی حمل می‌کند را در بر می‌گیرد و حاصل‌جمع درونی تمام پی‌آمدهایی که آن مقدار را حمل می‌کند در بر می‌گیرد. از اینرو، هنوز هم هر پی‌آمد موجود در فضای نمونه را دقیقاً یک بار محاسبه می‌کنیم. در آخرین سطر، از تعریف احتمال رویداد $[R=x]$ استفاده می‌کنیم. $\square$

قضیه فرعی ۲. ۵: اگر R یک متغیر تصادفی مقدار – طبیعی باشد آنگاه:

$$E[R]=\sum_{i=0}^{\infty} i\Pr\{R=i\}$$

وقتی که یک متغیر تصادفی را که فقط مقادیر طبیعی را دریافت می‌کند یعنی $N=\{۰,۱,۲,\ldots\}$، را در نظر می‌گیرید، راهی دیگری برای نوشتن مقدار انتظار وجود دارد:

قضیه ۳. ۵: اگر R یک متغیر تصادفی مقدر– طبیعی باشد، آنگاه:

$$E[R]=\sum_{i=۰}^{\infty}\Pr\{R>i\}$$

برهان. حاصل‌جمع را در نظر بگیرید:

$$\begin{aligned}
&\Pr\{R=۱\}+\Pr\{R=۲\}+\Pr\{R=۳\}+\ldots\\
&\qquad\qquad\quad +\Pr\{R=۲\}+\Pr\{R=۳\}+\ldots\\
&\qquad\qquad\qquad\qquad\quad +\Pr\{R=۳\}+\ldots\\
&\qquad\qquad\qquad\qquad\qquad\qquad +\ldots
\end{aligned}$$

حاصل‌جمع ستونها برابر هستند با $۱.\Pr\{R=۱\},۲.\Pr\{R=۲\},۳.\Pr\{R=۳\}$ و غیره

بنابر این کل مجموع برابر است با: $\sum_{i=۰}^{\infty} i.\Pr\{R=i\}=E[R]$

اینجا می‌خواهیم از قضیه فرعی ۲. ۵ استفاده کنیم. از سوی دیگر، حاصل‌جمع سطرها برابراند با $\Pr\{R>۰\},\Pr\{R>۱\},\Pr\{R>۲\}$ و الی آخر. بنابراین، کل حاصل‌جمع برابر است با:

$$\sum_{i=۰}^{\infty}\Pr\{R>i\}$$

این دو گزاره برای کل حاصل‌جمع باید مساوی باشد، که قضیه را ثابت می‌کند. □

۲. ۵ مقدار انتظار یک متغیر شاخص

مقدار انتظار یک متغیر تصادفی شاخص برای یک رویداد درست همان احتمال رویداد است. (یادتان باشد که یک متغیر تصادفی I_A متغیر تصادفی شاخص برای رویداد A است، اگر $I_A = ۱$ وقتی که A رخ بدهد و در غیر این صورت $I_A = ۰$.)

لم ٤. ۵ اگر I_A متغیر تصادفی شاخص برای رویداد A باشد، آنگاه

$$E[I_A] = \Pr\{A\}.$$

برهان.

$$\begin{aligned} E[I_A] &= ۱.\Pr\{I_A = ۱\} + ۰.\Pr\{I_A = ۰\} \\ &= \Pr\{I_A = ۱\} \\ &= \Pr\{A\}. && \{\text{تعریف } I_A.\} \end{aligned}$$

□

برای مثال، اگر A رویدادی باشد که یک سکه که با احتمال p با سر فرود آید آنگاه:

$$E[I_A] = \Pr\{I_A = ۱\} = p$$

۳. ۵ میانگین زمان تا شکست

بیائید به مسئله‌ای نگاه کنیم که یکی از این تعریف‌های جایگزین مقدار انتظار بطور ویژه‌ای مفید است. یک برنامه رایانه‌ای در پایان هر ساعت استفاده به احتمال p داغان می‌شود، اگر قبلاً داغان نشده باشد. زمان انتظار تا اینکه برنامه داغان بشود چقدر است؟

اگر فرض کنیم R تعداد ساعاتی برای داغان شدن باشد پس پاسخ مسئله ما $E[R]$ است. این متغیر، مقدار طبیعی است، بنابراین می‌توانیم از این فورمول استفاده کنیم:

$$E[R] = \sum_{i=0}^{\infty} \Pr\{R > i\}$$

وقتی $R > i$ است که سیستم بعد از i ساعت هنوز داغان نشده باشد، که با احتمال $(1-p)^i$ رخ می‌دهد.از داخل کردن این به درون فورمول فوق بدست می‌آید:

$$E[R] = \sum_{i=0}^{\infty} (1-p)^i$$

$$= \frac{1}{1-(1-p)}$$

$$= \frac{1}{p}$$

دومین شکل در خط دوم، از فورمول حاصل‌جمع یک دنباله هندسی نامتناهی بدست می‌آید جایی که نسبت عبارت‌های متوالی $1-p$ است.

بنابراین، برای مثال، اگر ۱٪ شانس برای اینکه برنامه در پایان هر ساعت نابود شود وجود داشته باشد، آنگاه زمان انتظار تا اینکه برنامه از بین برود عبارت است از $\frac{1}{0.01} = 100$ ساعت. اصل کلی در اینجا ارزش خوب به خاطر سپردن است:

اگر یک سیستم در هر مرحله زمانی با احتمال p شکست بخورد، پس تعداد مراحل انتظار تا به اولین شکست برسد عبارت از $1/p$ است.

۱. ۳. ۵ یک فرزند دختر به دنیا آوردن

یک زوج واقعاً یک فرزند دختر می‌خواهند. یک شانس ۵۰٪ وجود دارد که هر بچه‌ای که دارند دختر باشد و جنس فرزندانشان مستقل است. اگر آن زوج اصرار بورزد فرزند داشته باشند تا اینکه دختر گیرشان بیاید، در آن صورت نخست باید انتظار چند پسر بچه را داشته باشند؟

این واقعاً نمونه دیگری از مسئله پیشین است. پرسش "چند ساعت تا اینکه برنامه داغان شود؟" از نظر ریاضی شبیه همان پرسش "آن زوج باید چندین فرزند داشته باشند تا دختر گیرشان بیاید؟" است. در این مورد، یک ویرانی متناظر با داشتن یک دختر است، پس باید p را $\frac{۱}{۲}$ قرار دهیم. با توجه به تجزیه قبلی آن زوج باید بعد از برخورداری از ۲ فرزند منتظر یک دختر باشند. از آن رو که آخرین اینها دختر خواهد شد، آنها باید فقط منتظر یک پسر باشند.

چیزی برای فکر کردن: اگر هر زوج از ،راه برد، داشتن فرزند تا اینکه دختر گیرشان بیاید تبعیت کنند، احتمالاً چه اتفاقی برای کسر دخترانی که در این دنیا متولد شده‌اند می‌افتد؟