کد درس: ۱۸/۰۶۲J و ۶/۰۴۲J مقطع آموزشی: کارشناسی	بسمه تعالی عنوان درس: **ریاضیات** **برای علوم کامپیوتر**	دوره‌های آزاد رایانه‌ای SBU-MIT OCW Joint Project
استاد مدرس MIT : پروفسور آلبرت مِیر و پروفسور رونیت روبینفلد استاد مترجم SBU: دکتر چنگیز اصلاح‌چی		معاونت فناوری اطلاعات وارتباطات پروژه مشترک دانشگاه شهید بهشتی و دانشگاه MIT

فصل چهاردهم

امید ریاضی از دست رفته؟

در یادداشت‌های پیشین، دیدیم که میانگین یک کمیت تصادفی بوسیله درک ریاضی از امید یک متغیر تصادفی بیان می‌شود و امیدهایی را برای چند نوع متغیر تصادفی محاسبه کردیم. اینک دو چیز که امیدها را بسیار مفید می‌سازد را خواهیم دید. اول، یکبار که بدانید امید از چه قرار است، می‌توانید همچنین برخی از کران‌ها را برای احتمال اینکه از امید دور باشید، بدست آوردید– یعنی که، می‌توانید نشان دهید که اشیاء واقعاً مرموز محتمل نیستند. اینکه یک کرانه را به چه مناسبت می‌توانید بدست آورید بستگی به آن چیزی دارد که درباره توزیع می‌دانید، ولی نگران نباشید، حتی اگر تقریباً هیچ چیز را ندانید، هنوز هم می‌توانید چیزی نسبتاً جالب توجه بگوئید.

۱. خطی بودن امید

۱. ۱ امید یک حاصل جمع

مقادیر انتظار از قانونی ساده و بسیار مفید به نام خطی بودن امید پیروی می‌کنند.

ساده‌ترین شکل آن می‌گوید که مقدار انتظار حاصل‌جمع متغیرهای تصادفی عبارت از مجموع مقادیر انتظار متغیرهاست.

قضیه ۱. ۱ برای هر متغیر تصادفی R_1 و R_2،

$$E[R_1+R_2]=E[R_1]+E[R_2].$$

برهان. فرض کنید $T:=R_1+R_2$ این برهان به صورت ساده با تنظیم دوباره عبارت‌ها از تعریف $E[T]$ بدست می‌آید.

$$\begin{aligned} E[R_1+R_2] &:= E[T] \\ &:= \sum_{s\in S} T(S).\Pr\{S\} \\ &= \sum_{s\in S} \left(R_1(S)+R_2(S)\right).\Pr\{S\} && \text{(تعریف ۷)} \\ &= \sum_{s\in S} R_1(S)\Pr\{S\} + \sum_{s\in S} R_2(S)\Pr\{S\} && \text{تنظیم دوبارهٔ عبارت‌ها} \\ &= E[R_1]+E[R_2] \end{aligned}$$

□

به همین سان، داریم

لم ۲. ۱ برای هر مقدار تصادفی R و ثابت $a\in R$،

$$E[aR]=aE[R].$$

برهان به سادگی از تعریف انتظار حاصل می‌شود که آن را حذف می‌کنیم.

با ترکیب کردن قضیه ۱. ۱ و لم ۲. ۱ نتیجه می‌گیریم

قضیه ۳. ۱ (خطی بودن امید). برای تمام متغیرهای تصادفی R_2 و R_1 و ثابت‌هـای $a_1,a_2\in \mathbb{R}$ داریم.

$$E[a_1R_1+a_2R_2]=a_1E[R_1]+a_2E[R_2].$$

به سخن دیگر، امید یک تابع خطی است قانون و برهانش را می‌توان گسترش داد تا بیش از دو متغیر تصادفی را پوشش دهد:

قضیه فرعی ٤. ۱ برای هر متغیرهای تصادفی $R_1,\ldots,R_k$ و ثابت‌های $a_1\ldots a_k$ از داریم

$$E\left[\sum_{i=1}^{k} a_i \quad R_i\right] = \sum_{i=1}^{k} a_i\, E\left[R_i\right]. \qquad a_i \in R$$

موضوع مهمی که درباره خطی بودن امید در کار است این است که هیچ استقلالی الزامی نیست. واقعاً موضوع مفیدی است، برای اینکه کارکردن با استقلال رنج‌آور است و اغلب نیازمندیم با متغیرهای تصادفی کار کنیم که مستقل نیستند.

۲. ۱ مقدار مورد انتظار دو تاس

مقدار مورد انتظار حاصل‌جمع دو تاس سالم چقدر است؟

فرض کنید متغیر تصادفی R_1 عدد اولین تاس باشد و در نظر بگیرید R_2 شماره دومین تاس باشد. پیش از این متوجه شدیم که مقدار انتظار یک تاس ۳.۵ است.

می‌توانیم مقدار انتظار حاصل‌جمع را با استفاده از خطی بودن امید پیدا کنیم:

$$E\left[R_1 + R_2\right] = E\left[R_1\right] + E\left[R_2\right] = ۳.۵ + ۳.۵ = ۷.$$

توجه داشته باشید که الزامی نداشتیم فرض کنیم که هر دو تاس مستقل باشند. حاصل‌جمع مورد انتظار دو تاس ۷ است، حتی اگر به هم متصل باشند.[1]

[1] ولی هر تاس پس از اتصال باید درست باشد.

اثبات اینکه حاصل‌جمع مورد انتظار ۷ می‌شود با نمودار درختی سخت خواهد بودو ۳۶ حالت موجود است و اگر تاسها را متصل در نظر بگیریم، این کار به مثابه کابوس خواهد بود!

۳. ۱ مسئله بررسی– کلاه

در ضیافت شامی n مرد کلاه هایشان را آویزان می‌کنند. کلاه‌ها در مدت شام در هم شده‌اند، بنابراین از آن پس هر مرد یک کلاه تصادفی دریافت می‌کند.

بویژه، هر مرد به احتمال $1/n$ کلاه خودش را می‌گیرد. تعداد مورد انتظار مردانی که کلاه خودشان را دریافت می‌کنند چقدر است؟

بدون خطی بودن امید، پاسخ دادن به چنین پرسشی بسیاری مشکل خواهد بود. می‌توانیم به شکل زیر عمل کنیم. فرض کنید متغیر تصادفی R تعداد مردانی باشد که کلاه خودشان را دریافت می‌کنند. می‌خواهیم $E[R]$ را محاسبه کنیم. با توجه به تعریف امید، داریم:

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

حالا توی در سر افتادیم، زیرا برآورد $\Pr\{R=k\}$ یک آشفتگی است و بعد از آن باید این آشفتگی‌ها را در یک عمل جمع جایگزین کنیم. وانگهی، برای داشتن امید، به داشتن احتمال تمام جایگشت کلاه‌ها نیاز خواهیم داشت. برای مثال، ممکن است فرض کنیم که همه جایگشت‌های کلاه ها متساوی الاحتمال هستند.

حالا بیائید سعی کنیم از خطی بودن امید استفاده کنیم. مثل قبل، فرض کنید متغیر تصادفی R تعداد مردانی باشد که کلاه خودشان را دریافت می‌کنند. ترفند همانا تأیین R به عنوان

حاصل‌جمع متغیرهای شاخص است. بویژه، فرض کنید R_i شاخصی برای رویدادی باشد که iامین مرد کلاه خودش را بدست می‌آورد. یعنی اینکه، $R_i = ۱$ رویدادی است که کلاه خودش را دریافت می‌کند و $R_i = ۰$ رویدادی است که کلاه اشتباهی گیرش می‌آید. تعداد مردانی که کلاه خودشان را دریافت می‌کنند عبارت از حاصل‌جمع این شاخص‌ها است:

$$R = R_۱ + R_۲ + \ldots, R_n$$

این متغیرهای شاخص مستقل نیستند. برای مثال اگر $n - ۱$ نفر از مردان کلاه‌های خودشان را دریافت کنند، پس آخرین مرد مطمئن است که کلاه خودش را دریافت می‌کند. ولی از آن‌رو که خیال داریم از خطی بودن امید استفاده کنیم، درباره استقلال نگرانی نداریم!

بیائید مقدار امید هر دو طرف معادله بالا را برداریم و خطی بودن امید را به کار ببریم:

$$\begin{aligned} E[R] &= E[R_۱ + R_۲ + \ldots + R_n] \\ &= E[R_۱] + E[R_۲] + \ldots + E[R_n] \end{aligned}$$

از آنجا که R_iها متغیرهای شاخص هستند، $E[R_i] = \Pr\{R_i\}$ و از آنجا که احتمال دریافت کلاه درست برای همه مردان یکسان است پس این احتمال $\frac{۱}{n}$ است. با در کنار هم قراردادن همه اینها، داریم:

$$\begin{aligned} E[R] &= E[R_۱] + E[R_۲] + \ldots + E[R_n] \\ &= \Pr\{R_۱ = ۱\} + \Pr\{R_۲ = ۱\} + \ldots \Pr\{R_n = ۱\} \\ &= n.\frac{۱}{n} = ۱. \end{aligned}$$

بنابراین باید انتظار داشته باشیم که به طور میانگین یک مرد کلاه خودش را پس بگیرد!

توجه کنید که در نظر نگرفتیم که همه جایگشت‌های کلاه‌ها متساوی الاحتمال هستند یا حتی اینکه تمام جایگشت‌ها ممکن هستند. فقط نیاز داشتیم بدانیم که هر مرد با احتمال $1/n$ کلاه خود را دریافت می‌کند.

"همه جا امید وجود دارد"

٤. ١ امید توزیع دوجمله‌ای

فرض کنید که مستقلاً n سکه غیر منصف را پرتاب کنیم، هر کدام با احتمال p با سر پائین می‌آید. تعداد انتظار آن که سر بیاید چقدر است؟

فرض کنید $H_{n,p}$ تعداد سرها بعد از پرتاب‌ها باشد. آنگاه $H_{n,p}$ توزیع دوجمله‌ای با متغیرهای n و p است. حالا فرض کنید I_k شاخص برای kامین سکه‌ای باشد که سر می‌آیند. از آن‌رو که I_k یک متغیر شاخص با احتمال p برای مقدار ۱ است، می‌دانیم که

$$E[I_k] = p.$$

ولی

$$H_{n,p} = \sum_{k=1}^{n} I_k,$$

بنابراین با توجه به خطی بودن

$$E\left[H_{n,p}\right] = E\left[\sum_{k=1}^{n} I_k\right] = \sum_{k=1}^{n}[I_k] = \sum_{k=1}^{n} p = pn.$$

یعنی اینکه، امید یک متغیر دو جمله‌ای با پارامترهای (n,p)، np است.

۲. مسئله جمع کننده کوپن

هربار که در فروشگاه تاکوبل غذای کودک می‌خرم، به گرمی یک عدد ماشین مینیاتوری "موشکی مسابقه‌ای" به همراه یک دستگاه پرتابی که مرا قادر می‌سازد وسیله نقلیه تازه‌ام را بر روی لبه هر میز یا سطح صاف با شتاب بالا به حرکت در آورم، به من اعطا می‌شود حقیقتاً، خوشحالی‌ام هیچ حد و مرزی ندارد.

تعداد n نوع مختلف ماشین موشک مسابقه (آبی، سبز، قرمز، خاکستری و غیره) وجود دارند. نوع ماشینی که هر روز توسط آن خانم مهربان در ثبت دفتر تاکوبل به من اعطا می‌شود به نظر می‌رسد مستقلاً و به شکل تصادفی انتخاب شده‌اند. تعداد مورد انتظار غذای بچه‌ای که باید به منظور رسیدن به یکی از هر نوع ماشین موشکی مسابقه بخرم چقدر است؟

این پرسش ریاضی به اشکال بسیاری چهره می‌نماید، برای مثال، تعداد مورد انتظار از افرادی که باید گرد آوری کنید تا اینکه حداقل یک نفر را با هر روز تولد ممکن پیدا کنید چقدر است؟

اینجا، در عوض جمع‌آوری ماشین‌های مسابقه‌ای موشکی دارید روز تولد جمع‌آوری می‌کنید.

پرسش کلی به صورت عمومی، پس از هر تفسیر دیگری، مسئله جمع‌کننده کوپن نامیده می‌شود.

۱. ۲ یک راه حل با استفاده از خطی بودن امید

خطی بودن امید به نحوی شبیه استقراء و اصل لانه کبوتر است؛ نظریه ساده لوحانه‌ای است که می‌توان آن را در همه انواع روش‌های هوشمندانه بکار برد. برای مثال می‌توانیم از خطی بودن امید

در راهی هوشمندانه برای حل مسئله جمع‌کننده کوپن استفاده کنیم. فرض کنید پنج نوع مختلف موشک مسابقه وجود دارند و من این دنباله را دریافت می‌کنم:

آبی سبز سبز قرمز آبی نارنجی آبی نارنجی خاکستری

بیائید این دنباله را به ۵ پاره‌خط تقسیم‌بندی کنیم:

آبی | سبز | سبز قرمز | آبی نارنجی | آبی نارنجی خاکستری

×۰ ×۱ ×۲ ×۳ ×۴

قانون این است که هرگاه یک نوع جدید ماشین گیرم می‌آید پاره‌خط پایان می‌پذیرد. برای مثال، وقتی که برای اولین‌بار ماشین قرمز دریافت می‌کنم پاره‌خط میانی به انتها می‌رسد. در این راه، می‌توانم مسئله جمع‌آوری هر نوع ماشین را به ۵ مرحله تقسیم کنم. سپس می‌توانم هر مرحله را به تنهایی تجزیه کرده و با استفاده از خطی بودن امید نتایج را جمع‌بندی کنم.

بیائید به مورد کلی که در حال جمع‌آوری موشک‌های مسابقه هستم، برگردیم. فرض کنید X_k طول kامین پاره‌خط باشد. تعداد کل غذای بچه را که باید بخرم تا همه n موشک مسابقه را بدست آورم عبارت از حاصل‌جمع طول‌های کلیه این پاره‌خط‌هاست:

$$T = X_0 + X_1 + \ldots + X_{n-1}$$

حالا بیائید توجه خود را از X_k به طول kامین پاره خط متمرکز کنیم. در آغاز پاره‌خط kام، من k نوع مختلف ماشین دارم و پاره‌خط هنگامی به پایان می‌رسد که من نوعی ماشین جدید گیرم بیاید. وقتی که k نوع ماشین برنده شدم، هر غذای بچه حاوی نوعی است که قبلاً به احتمال k/n داشته‌ام. بنابراین، هر غذا حاوی نوعی جدید از ماشین به احتمال

$1-k/n=(n-k)/n$ است. بنابراین، تعداد مورد انتظار غذا تا اینکه ماشین جدیدی به دست آورم $n/(n-k)$ است باشد "زمان مناسب برای شکست"، همان فورمولی که فصل پیش به آن رسیدیم. داریم:

$$E[X_k]=\frac{n}{n-k}$$

خطی بودن امید، به همراه این نگرش، مسئله جمع‌کننده کوپن را حل می‌کند:

$$\begin{aligned} E[T] &= E[X_0+X_1+\ldots+X_{n-1}] \\ &= E[X_0]+E[X_1]+\ldots+E[X_{n-1}] \\ &= \frac{n}{n-0}+\frac{n}{n-1}+\ldots+\frac{n}{3}+\frac{n}{2}+\frac{n}{1} \\ &= n\left(\frac{1}{n}+\frac{1}{n-1}+\ldots+\frac{1}{3}+\frac{1}{2}+\frac{1}{1}\right) \\ &= nH_n \end{aligned}$$

جمع‌بندی سطر یکی به آخر حاصل توافقی با عبارت‌هایی در ترتیب معکوس است. همانطور که شاید به خاطر داشته باشید، از این حاصل‌جمع به صورت H_n یاد می‌شود و به طور تقریبی عبارت از I_n است.

بیائید با استفاده از راه‌حل عام به پرسش‌های خاص پاسخ دهیم. برای مثال، تعداد مورد انتظار پرتاب تاس که ریخته شود تا تمام اعداد ۱ تا ۶ دیده شود عبارت است از:

$$6H_6=14.7\ldots$$

و تعداد افراد مورد انتظار مردمی که شما باید گردآوری کنید تا حداقل یک فرد را با هر روز تولـد ممکن بیابید عبارت است از:

$$۳۶۵H_{۳۶۵} = ۲۳۶۴.۶\ldots$$

۳. امید شرطی

درست شبیه احتمالات رویدادها، امیدها نیز می‌توانند به رویدادی مشروط شوند.

تعریف ۱. ۳. امید شرطی یک متغیر تصادفی R بـه فـرض رویـداد A، $E\left[R\middle|A\right]$، را تعریـف می‌کنیم:

$$E\left[R\middle|A\right] := \sum_r r.\Pr\left\{R = r\middle|A\right\}.$$

به سخن دیگر، این همان مقدار مورد انتظار متغیر R است که بر رویداد A مشروط شده است.

مثال ۲. ۳ فرض کنید D پی‌آمد پرتاب یک تاس صحیح باشد. $E\left[D\middle|D \ge ۴\right]$ چیست؟

$$\sum_{i=۱}^{۶} i.\Pr\left\{D = i\middle|D \ge ۴\right\} = ۱.۰ + ۲.۰ + ۳.۰ + ۴.\frac{۱}{۳} + ۵.\frac{۱}{۳} + ۶.\frac{۱}{۳} = ۵.$$

دیدن اینکه قانون‌ها برای امید به امید شرطی گسترش می‌یابد آسان است. برای مثال، امید شـرطی نیز خطی خواهد شد.

قضیه ۳. ۳ برای هر دو متغیر تصادفی $R_۱$ و $R_۲$، ثابت‌های $a_۱, a_۲ \in \mathbb{R}$ و رویداد A،داریم

$$E\left[a_۱R_۱ + a_۲R_۲\middle|A\right] = a_۱E\left[R_۱\middle|A\right] + a_۲E\left[R_۲\middle|A\right].$$

یک مزیت واقعی امید شرطی این است که به ما امکـان مـی‌دهـد محاسـبات پیچیـده امیـد را بـه مواردی ساده‌تر تقسیم کنیم.

قضیه ٤. ۳ (قانون امید کل). اگر فضای نمونه اجتماع رویدادهای دو بـه دو مجـزای $A_1, A_2 \ldots$ باشد، آنگاه

$$\mathrm{E}[R] = \sum_i \mathrm{E}[R \mid A_i] \Pr\{A_i\}.$$

برهان.

$$\begin{aligned}
E[R] &= \sum_r r.\Pr\{R=r\} \\
&= \sum_r r.\sum_i \Pr\{R=r|A_i\}\Pr\{A_i\} && \text{(قانون احتمال کل)} \\
&= \sum_r \sum_i r.\Pr\{R=r|A_i\}\Pr\{A_i\} && (\text{توزیع ثابت } r) \\
&= \sum_i \sum_r r.\Pr\{R=r|A_i\}\Pr\{A_i\} && \text{(تعویض ترتیب جمع‌بندی)} \\
&= \sum_i Pr\{A_i\}\sum_r r.\Pr\{R=r|A_i\} && (\{Ai\}\ p \text{ عامل ثابت}) \\
&= \sum_i \Pr\{A_i\} E\left[R|A_i\right]. && \text{(تعریف ۱. ۳)}
\end{aligned}$$

□

مثال ٥. ۳ نیمی از مردم جهان مذکر و نیمی دیگر مؤنـث هسـتند. طـول قـد یـک مـذکر تصـادفاً انتخاب شده ″۱۱ ′۵ است، در حالی که طول قد یک مؤنث کـه بطـور تصـادفی انتخـاب شـده، ″۵ ′۵ است. طول قد مورد انتظار یک فرد تصادفاً انتخاب شده، چقدر است؟

فرض کنید $H(P)$ بلندی قد شخص تصادفی P باشد. رویدادهای مذکر "$M := P$" و مؤنث "$F := P$" است فضای نمونه را به دو قسمت تقسیم می‌کنند پس

$$E[H] = E[H|M]\Pr\{M\} + E[H|F]\Pr\{F\}$$

$$= ۵/۱۱.\frac{۱}{۲} + ۵/۵.\frac{۱}{۲}$$

$$= ۵/۸.$$

در بخش بعدی خواهیم دید که قانون امید کل بیشتر از آن قدرت دارد که کسی فکرش را بکند.

٤. مقدار انتظار حاصل ضرب

١. ٤ حاصل‌ضرب امیدهای مستقل

مشخص کرده‌ایم که امید یک حاصل‌جمع عبارت از حاصل‌جمع امیدهاست. این روال همیشه برای حاصل‌ضرب‌ها برقرار نیست. در کل، امید یک حاصل‌ضرب نیازی ندارد با حاصل‌ضرب امیدها برابر باشد. ولی در یک مورد خاص مهم حقیقت دارد، مثلاً، وقتی که متغیرهای تصادفی مستقل باشند.

لم ١. ٤ اگر $R_۱$ و $R_۲$ متغیرهای تصادفی مستقل باشند، آنگاه

$$E[R_۱|R_۲ = a] = E[R_۱].$$

این لم بلافاصله از تعریف ١. ٣ متعلق به امید شرطی و این حقیقت که $\Pr\{R_۱ = r\} = \Pr\{R_۱ = r|R_۲ = a\}$ تبعیت می‌کند.

قضیه ٢. ٤ برای هر دو متغیر تصادفی مستقل $R_۱, R_۲$،

$$E[R_1.R_2]=E[R_1].E[R_2].$$

برهان. ما قانون امید کل را با مشروط کردن مقدار R_1 به کار می‌گیریم.

(قضیه ٤. ۳)
$$E[R_1.R_2]=\sum_{r \text{ در برد } R} E\left[R_1.R_2 \,|\, R_1=r\right].\Pr\{R_1=r\}$$

$$=\sum_{r} E\left[r.R_2 \,|\, R_1=r\right].\Pr\{R_1=r\}$$

(لم ۲. ۱)
$$=\sum_{r} r.E\left[R_2 \,|\, R_1=r\right].\Pr\{R_1=r\}$$

(لم ۱. ٤)
$$=\sum_{r} r.E[R_2].\Pr\{R_1=r\}$$

$$=E[R_2]\sum_{r} r.\Pr\{R_1=r\}$$

$$=E[R_2].E[R_1].$$ □

قضیه ۲. ٤ به شکل روتین به مجموعه‌ای از متغیرهای مستقل گسترش می‌یابد.

قضیه فرعی ۳. ٤ اگر که متغیرهای تصادفی $R_1, R_2 \ldots R_K$ مستقل باشند،

آنگاه

$$E\left(\prod_{i=1}^{k} R_i\right)=\prod_{i=1}^{k} E(R_i)$$

۲. ٤ حاصل‌ضرب دو تاس

فرض کنید دو تاس مستقل، درست را پرتاب می‌کنیم و اعدادی که می‌آیند را ضرب می‌کنیم.

مقدار انتظار این حاصل‌ضرب چقدر است؟

فرض کنید متغیرهای تصادفی $R_۱$ و $R_۲$ اعدادی باشند که روی هر دو تاس نشان داده شده‌اند. می‌توانیم مقدار مورد انتظار را به شکل زیر محاسبه کنیم:

$$E[R_۱.R_۲]=E[R_۱].E[R_۲]=۳.۵\ .\ ۳.۵=۱۲.۲۵ \qquad (۱)$$

در اینجا اولین تساوی با قضیه ۲. ٤ حاصل می‌شود زیرا تاس‌ها مستقل هستند.

حالا فرض کنید که دو تاس مستقل نباشند؛ در واقع، در نظر آورید که تاس دوم همیشه به همان میزان اولی است. در این حالت حاصل‌ضرب امیدها با امید حاصل‌ضرب برابر نخواهد بود.

به منظور مشخص نمودن این، فرض کنید متغیرهای تصادفی $R_۱$ و $R_۲$ اعداد نمایش داده شده روی دو تاس باشند. می‌توانیم مقدار انتظار حاصل‌ضرب را بدون قضیه ۲. ٤ به شکل زیر محاسبه کنیم:

$$\begin{aligned}
E[R_۱.R_۲] &= E\left[R_۱^۲\right] \\
&= \sum_{i=۱}^{۶} i^۲.\Pr\left\{R_۱^۲=i^۲\right\} \\
&= \sum_{i=۱}^{۶} i^۲.\Pr\{R_۱=i\} \\
&= \frac{۱^۲}{۶}+\frac{۲^۲}{۶}+\frac{۳^۲}{۶}+\frac{۴^۲}{۶}+\frac{۵^۲}{۶}+\frac{۶^۲}{۶} \\
&= \frac{۹۱}{۶} \\
&\neq ۱۲.۲۵ \\
&= E[R_۱].E[R_۲].
\end{aligned}$$

از

۵. امید میانگین

مثال‌های بسیاری از متغیرهای تصادفی را دیده‌ایم که هیچ گاه یک مقداری به اندازه میانگین خـود را نمی‌گیرند. ولی تجربه می‌گوید که می‌توانیم انتظار داشته باشیم مقـادیر یـک متغیـر نزدیـک بـه میانگین آن باشد از این‌رو است که میانگین همچنین "امید" نامیده می‌شود. به سخن دیگر، مقادیر یک متغیر تصادفی احتمالاً خیلی زیاد از میانگین منحرف نمی‌شود. کمی از نتایج پایه‌ای درباره این عنوان مرکزی انحراف از میانگین را شرح خواهیم داد و خاطر نشان خواهیم کرد که چگونـه ایـن نتایج برای آزمودن فرضیه‌ها و تخمین زدن با نمونه سازی به کار می‌روند.

در این یادداشت‌ها دو نتیجه را گسترش خواهیم داد. اول قضـیه مـارکوف اسـت، کـه یـک کرانـه بالایی ساده، ولـی نوعـاً زمخـت، بـرای ایـن احتمـال کـه مقـدار یـک متغیـر تصـادفی بیشـتر از حاصل‌ضرب معینی از میانگین آن است را فراهم می‌کند. نتیجه‌گیری مارکوف صدق می‌کنـد اگـر که هیچ چیز درباره یک متغیر تصادفی ندانیم مگر آن که بدانیم میانگین آن چیست و اینکه مقـادیر آن غیر منفی هستند. از این رو، قضیه مارکوف خیلی کلی اسـت، ولـی بـاز هـم ضـعیف‌تـر از آن نتایجی است که اطلاعات بیشتری درباره توزیع متغیر به محاسبه در آید.

دربسیاری از موقعیت‌ها، نه تنها میانگین را می‌شناسیم، بلکه کمیت عـددی دیگـری کـه واریـانس متغیر تصادفی نامیده می‌شود را هم می‌شناسیم. دومین نتیجه‌گیری پایـه‌ای مـا قضـیه چبـی شـف است، که قضیه مارکوف و اطلاعات درباره واریانس را ترکیب می‌کند تا کرانه‌های بهتری بدسـت دهد. همچنین خصوصیات واریانس و راه‌های محاسبه آن را بررسی خواهیم کرد.

٦. قضیه مارکوف

قضیه مارکوف یک برآورد کلی زمخت از احتمال اینکه یک متغیر تصادفی مقداری بسیار بزرگتر از میانگین خود حمل می‌کند را به دست می‌دهد.

ایده پس زمینه قضیه مارکوف را می‌توان با مثالی ساده از ضریب هوشی، IQ، توضیح داد. IQ اختراع شد تا اینکه میانگین (حد متوسط) اندازه‌گیری IQ، ۱۰۰ باشد. حالا تنها از این واقعیت می‌توانیم نتیجه بگیریم که حداکثر ۱/۲ جمعیت می‌توانند IQی ۲۰۰ یا بیشتر داشته باشند، چون که اگر بیشتر از نیمی از جمعیت IQی ۲۰۰ داشتند، آنگاه حد متوسط می‌بایست بیشتر از $(۱/۲)۲۰۰=۱۰۰$ در نظر گرفته می‌شد و این ناقض آن واقعیت می‌شود که حدمتوسط ۱۰۰ بود. بنابراین احتمال اینکه یک شخص IQی ۲۰۰ یا بیشتر داشته باشد حداکثر ۱/۲ است. البته این نتیجه‌گیری خیلی محکمی نیست! در واقع، هیچ IQی بالاتر از ۲۰۰ تاکنون به ثبت نرسیده است. ولی تا حدود ۱۵۰ یا بیشتر ثبت شده است. IQی بالای ۱۵۰ مطمئناً به ثبت رسیده است، پس باز هم کسر بسیار کوچکی از مردم واقعاً IQیی که بالا باشد را دارا هستند. ولی اگر چه این نتیجه‌گیری‌ها درباره IQ ضعیف هستند، عملاً قوی‌ترین امکان نتیجه‌گیری‌های کلی هستند که می‌توان درباره متغیر تصادفی غیرمنفی با استفاده از تنها واقعیتی که میانگین آن ۱۰۰ باشد به دست آورد. برای مثال، اگر یک متغیر تصادفی با احتمال ۱/۲ مقدار ۲۰۰ را انتخاب کنیم و با احتمال ۱/۲ مقدار ۰ را بگیرد میانگین آن ۱۰۰ می‌شود و احتمال اینکه مقدار ۲۰۰ یا بیشتر رخ دهد واقعاً

۱/۲ است. بنابراین نمی‌توانیم امید داشته باشیم که یک کرانه بالایی بهتر برای احتمال رخـداد ۲۰۰ بیشتر از ۱/۲ بدست آوریم.

قضیه ۱. ٦ (قضیه مارکوف) اگر R یک متغیر تصادفی غیرمنفی باشد، آنگاه بـرای تمـام مقـادیر مثبت x داریم:

$$\Pr\{R \geq x\} \leq \frac{E[R]}{x}.$$

برهان. نشان خواهیم داد که $E[R] \geq x\Pr\{R \geq x\}$ و با تقسیم هر دو طرف بر x نتیجه مورد نظر را بدست می‌آوریم.

بنابراین فرض کنید I_x متغیر شاخص رویـداد $[R \geq x]$ باشـد و متغیـر تصـادفی xI_x را در نظـر بگیرید. توجه داشته باشید که

$$R \geq xI_x,$$

چونکه در هر نقطه نمونه، (w)،

- اگر $R(w) \geq x$ پس $R(w) \geq x = x.1 = xI_x(w)$ و
- اگر $R(w) < x$ پس $R(w) \geq = x.0 = xI_x(w)$.

بنابراین،

$$E[R] \geq E[xI_x] \qquad \text{(از آن‌رو که } R \geq xI_x\text{)}$$

$$= xE[I_x] \qquad \text{(خطی بودنِ } E\text{)}$$

$$= x\Pr\{I_x = 1\}$$

$$= x\Pr\{R \ge x\}. \qquad (\text{تعریف } I_x)$$

□

قضیه مارکوف اغلب به شکلی جایگزین، عنوان می‌شود، جایگزینی از آن در زیر بـه عنـوان یـک قضیه فرعی بیان شده است.

قضیه فرعی ۲. ٦ اگر R یک متغیر تصادفی غیرمنفی باشد، آنگاه برای هر مقدار $c \ge ۱$

$$\Pr\{R \ge c.E[R]\} \le \frac{۱}{c}.$$

برهان. در قضیه مارکوف قرار دهید $x = c.E[R]$ □

۱. ٦ مثال‌هایی از قضیه مارکوف

در نظر بگیرید که n مرد به یک ضیافت شام می‌روند و کلاه‌هایشان را آویزان می‌کنند. در پایان شب، کلاه‌ها تصادفاً جابه‌جا شده به صاحبانشان عودت داده شد، بنابراین هر مرد کلاه خود را به احتمال $۱/n$ می‌گیرد، در یادداشت‌های قبلی از خطی بودن امید برای نشان دادن اینکه $E[R] = ۱$ استفاده کردیم. با توجه به قضیه مارکوف احتمال اینکه حداقل x مرد کلاه درست را بدست آورند عبارت است از:

$$\Pr\{R \ge x\} \le \frac{E[R]}{x} = \frac{۱}{x}.$$

برای مثال، بهتر از یک بخت ۲۰٪ وجود ندارد که هر مردی کلاه درست را بگیرد صرف نظر از تعداد افرادی که در ضیافت شام حضور دارند.

پیش‌غذای چینی مسئله بسیار مشابهی است. در این مورد n نفر در حال خوردن پیش‌غذای چینی هستند که روی یک سینی مدور گردان قرار دارد. سپس یک نفر سینی را می‌چرخاند تا هر کدام یک پیش غذای تصادفی گیرش بیاید. احتمال اینکه هر یک همان پیش غذای قبلی را دریافت کند چقدر است؟ تعداد n سمت و سوی متساوی الاحتمال برای توقف پس از چرخش سینی وجود دارند. هر نفر پیش‌غذای صحیح را درست در یکی از این n جهت‌ها دریافت می‌کند. بنابراین پاسخ صحیح عبارت است از $1/n$.

ولی از قضیه مارکوف چه احتمالی بدست می‌آوریم؟ فرض کنید متغیر تصادفی R تعداد آدم‌هایی باشد که پیش‌غذای صحیح را دریافت می‌کنند. می‌توانید نشان بدهید که $E[R]=1$. با به کارگیری قضیه مارکوف متوجه می‌شویم:

$$\Pr\{R \geq n\} \leq \frac{E[R]}{n} = \frac{1}{n}.$$

بنابراین برای مسئله پیش‌غذای چینی، قضیه مارکوف استوار است!

از سوی دیگر، قضیه مارکوف همان کران $1/n$ را برای این احتمال که در مسئله کلاه، کلاهایشان را دریافت کنند، بدست می‌دهد. ولی در واقع، احتمال این رویداد $1/(n!)$ است. بنابراین برای مسئله کلاه، قضیه مارکوف کران بالایی که خیلی از واقعیت دور است را بدست می‌دهد.

۲. ٦ قضیه مارکوف برای متغیرهای کران‌دار

فرض کنید خبردار می‌شویم که حدمتوسط IQ در میان دانشجویان MIT (که به هر حال، درست نیست) ۱۵۰ است. درباره احتمال اینکه یک دانشجوی MIT IQ بیشتر از ۲۰۰ داشته باشد چه

می‌توانیم بگوئیم؟ قضیه مارکوف بلافاصله به ما می‌گوید که بیش از ۱۵۰/۲۰۰ یا ۳/۴ دانشجویان نمی‌توانند چنین IQ بالایی داشته باشند. در اینجا به سادگی قضیه مارکوف را به یک متغیر تصادفی R، هم ارز با IQی یک دانشجوی تصادفی MIT به کار بردیم تا نتیجه‌گیری کنیم که:

$$\Pr\{R > ۲۰۰\} \leq \frac{E[R]}{۲۰۰} = \frac{۱۵۰}{۲۰۰} = \frac{۳}{۴}.$$

ولی بیائید یک واقعیت دیگر را مشاهده کنیم (که شاید صحیح باشد): هیچ دانشجوی MIT یک IQی کمتر از ۱۰۰ ندارد. این به آن معنی است که اگر فرض کنیم $T ::= R - ۱۰۰$، آنگاه T غیرمنفی است و بنابراین می‌توانیم قضیه مارکوف را برای T به کار ببریم و نتیجه بگیریم که:

$$\Pr\{R > ۲۰۰\} = \Pr\{T > ۱۰۰\} \leq \frac{E[T]}{۱۰۰} = \frac{۵۰}{۱۰۰} = \frac{۱}{۲}.$$

بنابراین فقط نیمی، نه $\frac{۳}{۴}$، از دانشجویان می‌توانند به آن میزان که فکر می‌کنند اعجاب‌آور باشند. کمی هم آرامش خوب است!

به طور کلی‌تر، با به کار بردن قضیه مارکوف به $R - L$ به جای R می‌توانیم کرانه‌های بهتری با توجه به کران پائینی $L > ۰$ از R بدست آوریم.

به طور هم‌سان هم، اگر یک کران بالایی، u، برای یک متغیر تصادفی، S، داشته باشیم، آنگاه $u - S$ یک متغیر تصادفی غیرمنفی خواهد بود و به کار بردن قضیه مارکوف به $u - S$ به ما اجازه خواهد داد احتمال اینکه S خیلی از امید خود کمتر باشد را محدود کنیم.

۷. قضیه چپی شف

ما تفسیرهایی از قضیه مارکوف برای احتمال انحراف از (امید) میانگین در دست داریم، ولی غالباً کرانه‌هایی را می‌خواهیم که برای فاصله گرفتن از میانگین در هر جهتی به کار رود، یعنی که، کرانه‌هایی بر این احتمال که $|R-E[R]|$ بزرگ باشد.

کمی آشفته است که مستقیماً قضیه مارکوف را در این مسئله به کار بریم، زیرا در کل محاسبه $E\big[|R-E[R]|\big]$ آسان نیست. با این وجود، از آن‌رو که $|R|$ و به همین ترتیب $|R|^k$ برای هر R متغیر غیرمنفی است، عدم تساوی مارکوف هم برای رویداد $\big[|R|^k \ge x^k\big]$ به کار برده می‌شود. ولی این رویداد با رویداد $\big[|R| \ge x\big]$ هم‌ارز است، بنابراین داریم:

لم ۱. ۷ برای هر متغیر تصادفی R و هر عدد صحیح مثبتی k و هر $x>0$،

$$\Pr\{|R| \ge x\} \le \frac{E\big[|R|^k\big]}{x^k}.$$

حالت ویژه این لم برای $k=2$ را می‌توان برای محدود کردن متغیر تصادفی $|R-E[R]|$، به کار برد که انحراف R از میانگین خود را اندازه‌گیری می‌کند. مثلاً

$$\Pr\{|R-E[R]| \ge x\} = \Pr\{(R-E[R])^2 \ge x^2\} \le \frac{E\big[(R-E[R]^2)\big]}{x^2} \qquad (2)$$

جایی که در آن نامساوی در (۲) با به کار گرفتن لم ۱. ۷ از متغیر تصادفی غیرمنفی، $(R-E[R])^2$، تبعیت می‌کند. با فرض بر اینکه کمیت $E\big[(R-E[R])^2\big]$ متناهی است، می‌توانیم

نتیجه بگیریم که احتمال اینکه R از میانگین خود با بیشتر از x منحرف شود عبارت است از $O\left(1/x^2\right)$.

تعریف ۷.۲. واریانس یک متغیر تصادفی، R، $\operatorname{var}[R]$، عبارت است از:

$$\operatorname{var}[R] ::= E\left[\left(R-E[R]\right)^2\right].$$

بنابراین می‌توانیم دوباره (۲) را به شکل زیر عنوان کنیم

قضیه ۷.۳ (چبی شف). فرض کنید R یک متغیر تصادفی باشد و فرض کنید x یک عدد مثبت حقیقی باشد. آنگاه

$$\Pr\left\{\left|R-E[R]\right| \geq x\right\} \leq \frac{\operatorname{var}[R]}{x^2}.$$

گزاره $E\left[\left(R-E[R]\right)^2\right]$ برای واریانس کمی مرموز است. بهترین شیوه حل آن تمرین از خلال آن از درون خارج است درونی‌ترین گزاره $R-E[R]$، به طور مشخص انحراف R از میانگین خود است. این را از راه به توان دو رساندن، بدست می‌آوریم، $\left(R-E[R]\right)^2$. این یک متغیر تصادفی است که نزدیک به ۰ است وقتی که R نزدیک به میانگین باشد و یک عدد بزرگ مثبت است وقتی که R خیلی از میانگین منحرف می‌شود. بنابراین اگر R همیشه نزدیک به میانگین باشد در آن صورت واریانس کوچک خواهد بود و اگر R غالباً دور از میانگین باشد، واریانس بزرگ خواهد بود.

۷.۱ واریانس در دو بازی قمار

هنگامی که دو بازی قمار زیر را با هم مقایسه کنیم ارتباط واریانس علنی است.

بازی A: با احتمال ۲/۳ دو دلار (۲$) می‌بریم و به احتمال ۱/۳ یک دلار (۱$) می‌بازیم.

بازی B: با احتمال ۲/۳، ۱۰۰۲$ دلار می‌بریم و با احتمال ۱/۳، ۲۰۰۱$ دلار می‌بازیم.

از نظر مالی کدام بازی بهتر است؟ همان احتمال ۲/۳ را داریم، تا در هر بازی برنده شویم، ولی آن همه داستان را نمی‌گوید. درباره امید برای هر بازی چطور؟

فرض کنید متغیرهای A و B نتیجه‌نهایی هر دو بازی باشد. برای مثال A با احتمال ۲/۳ ۲ و با احتمال ۱/۳، ۱− می‌شود می‌توانیم نتیجه نهایی امید برای هر بازی را به شکل زیر محاسبه کنیم:

$$E[A]=۲.\frac{۲}{۳}+(-۱).\frac{۱}{۳}=۱,$$

$$E[B]=۱۰۰۲.\frac{۲}{۳}+(-۲۰۰۱).\frac{۱}{۳}=۱.$$

نتیجه نهایی برای هر دو بازی همان است، ولی به طور آشکار بسیار متفاوت هستند!

این تفاوت در مقدار امید آنها آشکار نیست، ولی بوسیله واریانس دیده می‌شود. می‌توانیم $\mathrm{var}|A|$ را با کار کردن "از درون خارج" به شکل زیر محاسبه کنیم:

$$A-E[A]=\begin{cases} ۱ & \text{با احتمال } \frac{۲}{۳} \\ -۲ & \text{بااحتمال } \frac{۱}{۲} \end{cases}$$

$$(A-E[A])^۲=\begin{cases} ۱ & \text{با احتمال } \frac{۲}{۳} \\ ۴ & \text{بااحتمال } \frac{۱}{۲} \end{cases}$$

$$E\left[(A-E[A])^۲\right]=۱.\frac{۲}{۳}+۴.\frac{۱}{۳}=۲$$

$$\mathrm{var}[B] = ۲.۰۰۴,۰۰۲.$$

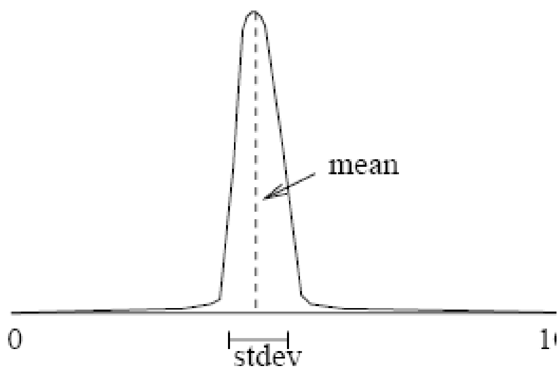
واریانس بازی A ,۲ است و واریانس بازی B بیش از دو میلیون! بطـور حسـی، معنـی‌اش ایـن است که نتیجه نهایی بازی A معمولاً نزدیک به مقدار مـورد انتظـار ۱$ دلار اسـت، ولـی نتیجـه نهایی بازی B می‌تواند خیلی دور از این مقدار مورد انتظار شود. واریانس بالا اغلب با خطر بـالا همراه می‌شود. برای مثال، در ده دور بازی A امیدواریم ۱۰$ دلار ببریم، ولی تصوراً می‌توانستیم ۱۰$ در عوض ببازیم. از سوی دیگر، در ده دور از بازی B همچنین امیدواریم ۱۰$ دلار برنـده شویم، ولی عملاً ممکن بود بیشتر از ۲۰,۰۰۰$ دلار ببازیم!

۲. ۷ انحراف از معیار

به دلیل تعریف خود در عبارت‌های مجذور (مربع) یک متغیر تصادفی، واریـانس متغیـر تصـادفی ممکن است خیلی بیشتر از یک انحراف از میانگین باشد. برای مثال، در بازی B بالا، انحـراف از میانگین در یک پی‌آمد ۱۰۰۱ است و در دیگری ۲۰۰۲ اسـت ولـی واریـانس آن مقـدار مهلـک ۲,۰۰۴,۰۰۲ است. از نقطه نظر تجزیه ابهادی، "اتحادهای" واریانس اشـتباه هسـتند: اگـر متغیـر تصادفی به صورت دلار باشد، آنگاه امید هم به شکل دلار است، ولی واریانس به صورت مجـذور دلار است. به این دلیل، اغلب متغیرهای تصـادفی را بـا اسـتفاده از انحـراف اسـتاندارد بـه جـای واریانس توضیح می‌دهند.

تعریف ۷. ٤. انحراف استاندارد σ_R از یک متغیر تصادفی R جذر واریانس است:

$$\sigma_R := \sqrt{\mathrm{var}[R]} = \sqrt{E\left[\left(R - E[R]\right)^{۲}\right]}.$$



تصویر ۱: انحراف استاندارد یک توزیع را نشان می‌دهد که "قسمت اصلی" آن چقدر گسترده است.

بنابراین انحراف استاندارد جذر انحراف از میانگین است، که دارای همان واحدها– در مثال ما دلار– را به عنوان متغیر تصادفی اصلی و به عنوان میانگین است. بطور حسی، کارش اندازه-گیری "انحراف از (حد متوسط) میانگین" است، زیرا که می‌توانیم به جذر خارج به عنوان فسخ مربع در داخل فکر کنیم.

مثال ۵. ۷ انحراف استاندارد نتیجه نهایی بازی B عبارت است از:

$$\sigma B=\sqrt{\operatorname{var}[B]}=\sqrt{۲,۰۰۴,۰۰۲}\approx ۱۴۱۶ \ .$$

متغیر تصادفی B عملاً از میانگین با مقدار مثبت ۱۰۰۱ یا مقدار منفی ۲۰۰۲ منحرف می‌شود؛ انحراف استاندارد ۱۴۱۶ این موقعیت را به خوبی شرح می‌دهد.

به صورت حسی انحراف استاندارد " گستردگی" " قسمت اصلی" توزیع گراف را اندازه‌گیری می‌کند، همان طور که در تصویر ۱ به شکل درآمده است.

یک فورمول‌بندی مجدد ساده مفید از قضیه چبی شف در عبارت‌های انحراف استاندارد وجود دارد.

قضیه فرعی ٦. ۷ فرض کنید R متغیر تصادفی است و c یک عدد حقیقی مثبت باشد.

$$\Pr\{|R-E[R]\geq c\sigma_R\}\leq\frac{۱}{c^{۲}}$$

در اینجا بوضوح می‌بینیم که چطور مقادیر "احتمال" R در یک (σR)○ منطقه اندازه‌گیری شده پیرامون $E[R]$، با تأکید بر اینکه انحراف استاندارد اندازه می‌گیرد که توزیع R پیرامون میانگین آن چطور گسترده می‌شود.

$$\Pr\{|R-E[R]|\geq c\sigma_R\}\leq\frac{\mathrm{var}[R]}{(c\sigma_R)^{۲}}=\frac{\sigma_R^{۲}}{(c\sigma_R)^{۲}}=\frac{۱}{c^{۲}}.$$ □

۳. ۷ مثال IQ

فرض کنید که، در مجموع میانگین IQ ۱۰۰ باشد همچنین انحراف استاندارد IQ که ۱۰ است را می‌دانیم. یک IQ که ۲۰۰ یا بیشتر باشد چقدر نایاب است؟

فرض کنید متغیر تصادفی، R، IQ فردی که به تصادف انتخاب شده است، باشد. بنابراین داریم $\sigma_R=۱۰$، $E[R]=۱۰۰$ و R غیرمنفی است. می‌خواهیم $\Pr\{R\geq ۲۰۰\}$ را محاسبه کنیم.

قبلاً دیده‌ایم که قضیه ۱. ٦ مارکوف کران زمخت بدست می‌دهد، مثلاً،

$$\Pr\{R\geq ۲۰۰\}\leq\frac{۱}{۲}.$$

اینک قضیه چبی شف را برای همان مسئله به کار می‌بریم:

$$\Pr\{R \geq ۲۰۰\} = \Pr\{|R - ۱۰۰| \geq ۱۰۰\} \leq \frac{\operatorname{var}[R]}{۱۰۰^۲} = \frac{۱۰^۲}{۱۰۰^۲} = \frac{۱}{۱۰۰}.$$

گام اول بیان احتمال مطلوب به شکل مورد نیاز قضیه چبی شف است؛ و تساوی صادق است زیرا R غیرمنفی است. قضیه چبی شف آنگاه عدم تساوی را بدست می‌آورد.

بنابراین قضیه چبی شف دلالت می‌کند که حداکثر یک شخص در میان صد نفر IQ ۲۰۰ یا بیشتر دارد. با استفاده از اطلاعات افزوده، مثلاً واریانس R، کران بسیار بهتری از آنچه با دانستن امید می‌توانستیم بدست آوریم.

۸. خصوصیات واریانس

۸.۱ چرا واریانس؟

تعریف واریانس R به عنوان $E\left[\left(R - E[R]\right)^۲\right]$ چه بسا اختیاری‌تر به نظر برسد. واریانس حدمتوسط مجذور انحراف از میانگین است. به این دلیل، برخی اوقات به واریانس "انحراف مجذور میانگین" می‌گویند. ولی چرا جذرگیری را به زحمت بیاندازیم؟ چرا به سادگی حدمتوسط انحراف از میانگین را محاسبه نکنیم؟ یعنی اینکه، چرا واریانس $E\left[R - E[R]\right]$ را تعیین نکنیم؟

مسئله این تعریف این است که انحراف‌های مثبت و منفی از میانگین دقیقاً یکدیگر را خنثی می‌کنند. با توجه به خطی بودن امید، داریم:

$$E\left[R - E[R]\ \right] = E[R] - E\left[E[R]\right]\ .$$

از آن‌رو که $E[R]$ یک ثابت است، مقدار انتظار آن خودش است. بنابراین

$$E\left[R-E[R]\right]=E[R].\;\; E[R]=0.$$

با توجه به این تعریف، هر متغیر تصادفی واریانس صفر دارد. که مفید نیست! از آنجا که مجذور کران در تعریف، هر دو انحراف‌های مثبت و منفی از میانگین را مثبت می‌کند؛ انحراف‌های مثبت و منفی یکدگر را از بین نمی‌برند.

البته، همچنین می‌توانستیم از بین رفتن انحراف‌های مثبت و منفی با گرفتن یک قدر مطلق جلوگیری کنیم. یعنی اینکه، می‌توانستیم تعریف کنیم که واریانس $E\left[\left|R-E[R]\right|\right]$ باشد. هیچ دلیل منطقی وجود ندارد که از این تعریف استفاده نباید کرد. با این وجود تعریف قراردادی واریانس مقداری خصوصیات ریاضی ارزشمند دارد که تعریف قدر مطلق ندارد.

در قسمت‌های زیر این خصوصیات را شرح می‌دهیم و از آنها برای تعیین واریانس تعدادی از توزیع‌های احتمالی مهم استفاده می‌کنیم.

۲. ۸ یک تعریف جایگزین واریانس

یک راه هم‌ارز برای تعیین واریانس یک متغیر تصادفی که کمتر مورد توجه است وجود دارد، ولی اغلب برای استفاده در محاسبات و برهان‌ها آسان‌تر است:

قضیه ۱. ۸ برای هر متغیر تصادفی، R.

$$\operatorname{var}[R]=E\left[R^2\right]-E^2[R],$$

در اینجا از علامت‌گذاری $E^{۲}[R]$ به عنوان کوتاه‌نویسی $(E[R])^{۲}$ استفاده می‌کنیم. بـه خـاطر داشته باشید که $E[R^{۲}]$ بـه طـور کلـی بـا $E^{۲}[R]$ برابـر نیسـت. مـی‌دانـیم کـه مقـدار امیـد حاصل‌ضرب، حاصل‌ضرب مقادیر امید متغیرهای مستقل است، ولی در کل اینطـور نیسـت. و R مستقل از خودش نیست مگر که ثابت باشد.

برهان. فرض کنید $\mu = E[R]$: پس

$$\begin{aligned}
\mathrm{var}[R] &= E\left[\left(R - E[R]\right)^{۲}\right] && \text{(تعریف ۲. ۷ واریانس)}\\
&= E\left[\left(R-\mu\right)^{۲}\right] && (\text{تعریف } \mu)\\
&= E\left[R^{۲} - ۲\mu R + \mu^{۲}\right] \\
&= E\left[R^{۲}\right] - ۲\mu E[R] + \mu^{۲} && \text{(خطی بودن امید)}\\
&= E\left[R^{۲}\right] - ۲\mu^{۲} + \mu^{۲} && (\text{تعریف } \mu)\\
&= E\left[R^{۲}\right] - \mu^{۲} \\
&= E\left[R^{۲}\right] - E^{۲}[R] \ . && (\text{تعریف } \mu)
\end{aligned}$$

□

برای مثال، اگر B یک متغیر برنولی باشد جائی که $p := \Pr\{B = ۱\}$ آنگاه

$$\mathrm{var}[B] = P - P^{۲} = P(۱-P) \ . \qquad (۳)$$

برهان. از آن‌رو که B فقط مقادیر ۱ و ۰ را حمل می‌کند، داریم $E[B]=P.1+(1-P).0=P$. از آن‌رو که $B=B^2$ همچنین داریم $E\left[B^2\right]=P$ بنابراین (۳) بلافاصله از (۱. ۸) بدست می‌آید. □

۱. ۲. ۸ واریانس صفر

چه موقع یک متغیر تصادفی، R، واریانس صفر دارد؟... وقتی که متغیر تصادفی هرگز از میانگین منحرف نمی‌شود!

لم ۲. ۸ واریانس یک متغیر تصادفی ، R، صفر است اگر و فقط اگر

$$\Pr\{R=E[R]\}=1$$

بنابراین گفتن اینکه $\operatorname{var}[R]=0$ تقریباً مثل این است که بگوئیم R ثابت است.مثلاً آن مقدار ثابت را برابر با امید بر روی تمام اعضای فضای نمونه با احتمال غیر صفر حمل می‌کند.(می‌تواند هر مقدار متناهی روی اعضای فضای نمونه با احتمال صفر بدون اثرگذاری بر واریانس را حمل کند.)

برهان. با توجه به تعریف واریانس

$$\operatorname{var}[R]=0 \qquad \text{اگر و فقط اگر} \qquad E\left[\left(R-E|R|\right)^2\right]=0$$

گزاره داخلی سمت راست ، $\left(R-E[R]\right)^{۲}$، همیشه غیرمنفی است (بدلیل مربع). به عنوان نتیجه‌گیری، $E\left[\left(R-E[R]\right)^{۲}\right]=۰$ اگر و فقط اگر $\Pr\left\{\left(R-E[R]\right)^{۲}\neq ۰\right\}$ صفر باشد، که به این معنی است که بگوئیم $\Pr\left\{\left(R-E[R]\right)^{۲}=۰\right\}$ یک است. یعنی که،

$$\operatorname{var}[R]=۰ \quad \text{اگر و فقط اگر} \quad \Pr\left\{\left(R-E[R]\right)^{۲}=۰\right\}=۱$$

ولی $\left(R-E[R]\right)^{۲}=۰$ و $R=E[R]$ توصیف‌های متفاوتی از یک رویداد هستند.

بنابراین،

$$\operatorname{var}[R]=۰ \quad \text{اگر و فقط اگر} \quad \Pr\{R=E[R]\}=۱ \ .$$

□

۲. ۲. ۸ کارکردن با ثابت‌ها

قضیه زیر نشان می‌دهد که واریانس یک متغیر تصادفی چگونه وقتی که توسط یک ثابت ضرب می‌شود یا تغیر مکان می‌دهد، تغیر می‌یابد.

قضیه ۳. ۸ فرض کنید R یک متغیر تصادفی و a و b ثابت باشند.

آن‌گاه

$$\operatorname{var}[aR+b]=a^{۲}\operatorname{var}[R] \ . \qquad (٤)$$

این قضیه دو نکته را ایجاب می‌کند. اول اینکه افزودن یک ثابت b به متغیر تصادفی بر واریانس اثر نمی‌گذارد. دوم، ضرب یک متغیر تصادفی در یک ثابت واریانس را به اندازه مربع آن ثابت تغییر می‌دهد.

برهان. از سمت چپ (٤) را به سمت راست می‌رسیم. اولین گام گسترش $\operatorname{var}[aR+b]$ با استفاده از تعریف جایگزین واریانس است.

$$\operatorname{var}[aR+b]=E\left[(aR+b)^{۲}\ \right]-E^{۲}[aR+b]$$

ما بر روی عبارت اول کار خواهیم کرد و بعد روی عبارت دوم. برای عبارت اول، یادتان باشد که با توجه به خطی بودن امید،

$$E\left[(aR+b)^{۲}\right]=E\left[a^{۲}R^{۲}+۲abR+b^{۲}\right]=a^{۲}E\left[R^{۲}\right]+۲abE[R]+b^{۲} \qquad (۵)$$

همینطور هم برای عبارت دوم:

$$E^{۲}[aR+b]=(aE[r]+b)^{۲}=a^{۲}E^{۲}[R]+۲abE[R]+b^{۲} \qquad (٦)$$

سرانجام، عبارت گسترش یافته دوم را از عبارت اول کم می‌کنیم.

$$\operatorname{var}[aR+b]=E\left[(aR+b)^{۲}\right]-E^{۲}[aR+b] \qquad \text{(قضیه ۱. ۸)}$$

$$=a^{۲}E\left[R^{۲}\right]+۲abE[R]+b^{۲}-$$

$$\left(a^{۲}E^{۲}\left[R^{۲}\right]+۲abE[R]+b^{۲}\right) \qquad \text{(با (۵) و (٦))}$$

$$=a^{۲}\left(E\left[R^{۲}\right]-a^{۲}E^{۲}[R]\right.$$

$$= a^{2}\left(E\left[R^{2}\right] - E^{2}\left[R\right]\right)$$

$$= a^{2}\,\mathrm{var}\left[R\right]. \qquad \text{(قضیه ۱. ۸)}$$

□

یک قانون مشابه برای انحراف استاندارد وقتی که یک متغیر تصادفی توسط یک ثابت تصحیح شده باشد صدق می‌کند. به خاطر بیاورید که انحراف استاندارد جذر واریانس است.

بنابراین، افزودن یک ثابت b به یک متغیر تصادفی انحراف استاندارد را تغییر نمی‌دهد. ضرب کردن یک متغیر تصادفی در ثابت a انحراف استاندارد را در a ضرب می‌کند. بنابراین داریم

قضیه فرعی ۸.٤. انحراف استاندارد $aR+b$ برابر است با a برابر با انحراف استاندارد R.

۳. ۸ واریانس حاصل‌جمع

پیش از این، ادعا کردیم که برای متغیرهای تصادفی دوبدو مستقل، واریانس حاصل‌جمع عبارت است از حاصل‌جمع واریانس‌هاست.

شرط استقلال دوبدو ضروری است. اگر استقلال را نادیده بگیریم، نتیجه می‌گرفتیم که $\mathrm{var}[R+R] = \mathrm{var}[R] + \mathrm{var}[R]$. با این وجود؛ با توجه به قضیه ۳. ۸ سمت چپ برابر با ۴برابر $\mathrm{var}[R]$ است در حالی که سمت راست عبارت است از $\mathrm{var}[R]$. ۲. تساوی‌اش برابر است که $\mathrm{var}[R] = 0$،که، با توجه به لم ۲. ۸ فقط در صورتی درست است که R ثابت باشد.

با این همه استقلال متغیرها با هم ضروری نیست: و فقط استقلال دو به دو مورد نیاز خواهد بود. دانستن این مسئله مفید است زیرا موقعیت‌های مهمی من‌جمله متغیرها وجود دارند که دو به دو

مستقل هستند ولی همگی مستقل نیستند. تلاقی کردن روزهای تولد مثالی از این دست است، همانگونه که در زیر متوجه خواهیم شد.

قضیه ۸.۵ [خاصیت جمع واریانس دو به دو مستقل] اگر $R_1, R_2, \ldots, R_n$ متغیرهای تصادفی دوبدو مستقل باشند، آنگاه

$$\operatorname{var}[R_1 + R_2 + \ldots + R_n] = \operatorname{var}[R_1] + \operatorname{var}[R_2] + \ldots + \operatorname{var}[R_n] \ .$$

برهان. با توجه به خطی بودن امید، داریم

$$E\left[\left(\sum_{i=1}^{n} R_i\right)^2\right] = E\left[\sum_{i=1}^{n} \sum_{j=1}^{n} R_i \quad R_j\right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[R_i \quad R_j\right] \qquad \text{(خطی بودن)}$$

$$= \sum_{1 \le i \ne j \le n} E[R_i] E\left[R_j\right] + \sum_{i=1}^{n} E\left[R_i^2\right] \qquad \text{(۷) (استقلال دوبه‌دو)}$$

در (۷) از این واقعیت استفاده می‌کنیم که امید حاصل‌ضرب دو متغیر مستقل حاصل‌ضرب امیدهای آنهاست.

همچنین،

$$E^2\left[\sum_{i=1}^{n} R_i\right] = \left(E\left[\sum_{i=1}^{n} R_i\right]\right)^2$$

$$= \left(\sum_{i=1}^{n} E[R_i]\right)^2 \qquad \text{(خطی بودن)}$$

$$=\sum_{i=1}^{n}\ \sum_{j=1}^{n}E[R_i]E[R_j]$$

$$=\sum_{1\le i\ne j\le n}E[R_i]\ \ E[R_j]+\sum_{i=1}^{n}E^2[R_i]\ . \qquad (8)$$

بنابراین

$$\operatorname{var}\left[\left(\sum_{i=1}^{n}R_i\right)\right]=E\left[\left(\sum_{i=1}^{n}R_i\right)^2\right]-E^2\left[\sum_{i=1}^{n}R_i\right] \qquad \text{(قضیه ۱. ۸)}$$

$$=\sum_{1\le i\ne j\le n}E[R_i]+\sum_{i=1}E\left[R_i^2\right]-$$

$$\left(\sum_{1\le i\ne j\le n}E[R_i]\ \ E[R_j]+\sum_{i=1}^{n}E^2[R_i]\right) \qquad \text{((۷) و (۸))}$$

$$=\sum_{i=1}^{n}E\left[R_i^2\right]-\sum_{i=1}^{n}E^2[R_i]$$

$$=\sum_{i=1}^{n}\left(E\left[R_i^2\right]-E^2[R_i]\right) \qquad \text{(با یکی کردن حاصل جمع‌ها)}$$

$$=\sum_{i=1}^{n}\operatorname{var}[R_i]\ . \qquad \text{(قضیه ۱. ۸)}$$

و قضیه ثابت می‌شود. □

اینک یک راه ساده محاسبه واریانسی یک متغیر $H_{n,p}$ داریم که یک توزیع دوجمله‌ای با متغیرهای n و p است. می‌دانیم $H_{n,p}=\sum_{k=1}^{n}I_k$ که I_kها متغیرهای تصادفی مستقل با

مقادیر ۰,۱ هستند. متقابلاً مستقل هستند. $\Pr\{I_K = ۱\} = p$ واریانس $P(۱-P)$ است، بنابراین با توجه به خطی بودن واریانس، داریم.

لم (واریانس توزیع دوجمله‌ای).

$$\operatorname{var}\left[H_{n,p}\right] = n \quad \operatorname{var}\left[I_K\right] = np(۱-p). \tag{۹}$$

۹. تخمین نمونه‌گیری تصادفی

۹ .۱ تخمین ترجیحات رای‌گیری با استفاده از قضیه چی‌شف

در محاسبه ترجیحات رأی‌گیری در یادداشت‌های فصل ۱۳، از کران‌های توزیع دوجمله‌ای برای تعیین سطوح اعتماد برای رای گرفتن از ترجیحات رأی‌دهندگان کلینتون در مقابل جولیانی استفاده کردیم. حالا که واریانس توزیع دوجمله‌ای را می‌شناسیم، می‌توانیم از قضیه چبی شف به عنوان راه حلی جایگزین برای محاسبه میزان رأی‌گیری استفاده کنیم.

ترتیب به همان گونه است که در یادداشت‌های فصل ۱۳ بود: ما n رأی دهنده بطور تصادفی را جمع‌آوری خواهیم کرد و فرض می‌کنیم S_n عدد کلی نمونه ما باشد که کلینتون را ترجیح داده‌اند. S_n/n را به عنوان تخمین خود از کسر فعلی p، متعلق به رأی‌دهندگانی که کلینتون را ترجیح می‌دهند استفاده می‌کنیم. می‌خواهیم n را انتخاب کنیم تا اینکه محاسبه تخمین ما در میان ۰.۰۴ متعلق به p حداقل ۹۵٪ زمان باشد.

از آنجا که S_n به صورت دوجمله‌ای توزیع شده است، بنابراین از (۹) داریم

$$\operatorname{var}\left[S_n\right] = n\big(p(۱-p)\big) \le n\frac{۱}{۴}.$$

کران ۱/۴ از این واقعیت ساده بدست آمد که ماکسیم $p(۱-p)$ وقتی حاصل می‌شود که $p=۱-p$، یعنی، هنگامی که $p=۱/۲$.

بعد، واریانس S_n/n را محدود می‌کنیم:

$$\operatorname{var}\left[\frac{S_n}{n}\right]=\left(\frac{۱}{n}\right)^{۲}\operatorname{var}\left[S_n\right] \qquad \text{(در (٤))}$$

$$\leq\left(\frac{۱}{n}\right)^{۲} n\frac{۱}{۴} \qquad \text{(در (۹))}$$

$$=\frac{۱}{۴n}. \qquad (۱۰)$$

حالا از چبی شف و (۱۰) داریم:

$$\Pr\left\{\left|\frac{S_n}{n}-P\right|\geq ۰.۰۴\right\}\leq\frac{\operatorname{var}\left[S_n/n\right]}{(۰.۰۴)^{۲}}=\frac{۱}{۴n(۰.۰۴)^{۲}}=\frac{۱۵۶/۲۵}{n}\ . \qquad (۱۱)$$

برای اینکه تخمین خود را با ۹۵٪ اعتماد بر پا کنیم، می‌خواهیم جهت سمت راست (۱۱) حداکثر $\frac{۱}{۲۰}$ باشد. بنابراین n را انتخاب می کنیم تا اینکه

$$\frac{۱۵۶.۲۵}{n}\leq\frac{۱}{۲۰},$$

یعنی اینکه،

$$n\geq ۳,۱۲۵.$$

ممکن است به یاد داشته باشید که در یادداشت‌های ۱۳ محاسبه کردیم که عملاً کافی بود فقط ۶۶۴ رأی‌دهنده را جمع کرد– بسیار کمتر از آنچه ۳,۱۲۵ با استفاده از قضیه چبی شف استخراج

کردیم. بنابراین کرانه چبی شف تقریباً به همان خوبی (کارآمدی) کرانه‌ای که قبلاً بدست آوردیم، نیست. این مسئله جای تعجب ندارد. در به کارگیری قضیه چبی شف فقط از یک کران برای واریانس S_n استفاده کردیم. در یادداشت‌های ۱۳ از سوی دیگر، از این واقعیت که متغیر تصادفی S_n دوجمله‌ای بود استفاده کردیم (با متغیر معلوم ، n، و متغیر نامعلوم p). معنی‌دار است که اطلاعات مفصل‌تر درباره یک توزیع به کرانه‌های بهتر می‌انجامد. ولی حتی اگر چه آن کران خیلی مناسب نباشد، این مثال بخوبی یک راه محاسبه قضیه چبی شف را که بطور گسترده‌ای به تخمین‌های دوجمله‌ای کاربرد دارد را به تصویر می‌کشد.

۲. ۹ دوباره روز تولد

حالت‌های مهمی وجود دارند که توزیع‌های مربوط دوجمله‌ای نیستند زیرا در خصوصیات استقلال مثال ترجیح رأی‌دهندگان صدق نمی‌کند. در این حالت‌ها، روش‌های تخمین مبنی بر کران چبی شف ممکن است بهترین راه حل باشد.

تلاقی روز تولد یک چنین مثالی است.

قبلاً دیده‌ایم که در یک کلاس صد نفری یا بیشتر، احتمال بالایی وجود دارد که حداقل دو دانشجو یک روز تولد یکسان داشته باشند. همچنین به سادگی می‌توانیم عدد مورد انتظار تعداد دانشجویان با روز تولد یکسان را محاسبه کنیم. ولی آیا این متشابهاً تعداد جفت‌های یکسان در یک گروه نوعی عملاً نزدیک به تعداد مورد انتظار خواهد بود؟ می‌توانیم همان راه حل را برای

پاسخ دادن به این پرسش همان طور که در تخمین ترجیح رأی‌دهندگان انجام دادیم، به کار بگیریم.

ولی توجه داشته باشید که برخورداری از روز تولدهای یکسان بین جفت‌های دانشجویان به معنی رویدادهای مستقل نیست. برای مثال، با علم بر اینکه آلیس و باب روز تولد یکسان دارند و همینطور تد و آلیس روز تولدهای یکسان دارند آشکارا دلالت می‌کند بر اینکه باب و تد روز تولدهای یکسان دارند. از سوی دیگر، با علم بر اینکه آلیس و باب روز تولد یکسان دارند، چیزی درباره اینکه آیا آلیس و کارول روز تولد یکسان داشته باشند به ما نمی‌گوید، مثلاً این دو رویداد واقعاً مستقل هستند.

بنابراین حتی اگر چه رویدادهای جفت‌های مختلف دانشجویان با روز تولد یکسان مستقل نیستند، در واقع حتی نه سه– راه مستقل، آنها دو به دو مستقل هستند.

این به ما اجازه می‌دهد همان استدلال تلاقی روز تولد را همان طور که برای ترجیح رأی‌دهنده استفاده کردیم، به کار ببریم. مثلاً فرض کنید $B_۱, B_۲, \ldots, B_n$ روز تولد n شخص که بطور مستقل انتخاب شده‌اند باشد. و فرض کنید $E_{i,j}$ متغیر شاخص برای رویدادی باشد که iامین و jامین فرد منتخب روز تولد یکسان را داشته باشند، یعنی اینکه رویداد $[B_i = B_j]$ برای ساده سازی، در نظر می‌گیریم که برای $i \neq j$، احتمال[۲] اینکه $B_i = B_j$، $\frac{۱}{۳۶۵}$ است. بنابراین B_iها متغیرهای

۲. در آمریکا روزهای تولد پائیزی عمومی‌تر از روز تولدهای زمستانی است.بنابراین $\Pr\{Bi=Bj\}$ عملاً کمی بزرگتر از ۱/۳۶۵ است.

مستقل هستند و از این‌رو $E_{i,j}$ها متغیرهای دو به دو مستقل هستند، که همه آن چیزی است که نیاز خواهیم داشت.

فرض کنید M_n تعداد جفت‌های یکسان روزهای تولد نزد n گزینه باشد یعنی که،

$$M_n ::= \sum_{۱\le i<j\le n} E_{i_j} \quad . \qquad (۱۲)$$

بنابراین با توجه به خطی بودن امید

$$E[M_n] = E\left[\sum_{۱\le i<j\le n} E_{i_j}\right] = \sum_{۱\le i<j\le n} E\left[E_{i_j}\right] = \binom{n}{۲}.\frac{۱}{۳۶۵}.$$

همچنین با توجه به قضیه ۵. ۸ واریانس متغیرهای دو به دو مستقل جمعی است، بنابراین

$$\mathrm{var}[M_n] = \mathrm{var}\left[\sum_{۱\le i<j\le n} E_{i_j}\right] = \sum_{۱\le i<j\le n} \mathrm{var}\left[E_{i_j}\right] = \binom{n}{۲}.\frac{۱}{۳۶۵}\left(۱-\frac{۱}{۳۶۵}\right).$$

اینک برای کلاسی با ۱۰۰ دانشجو، داریم $E[M_{۱۰۰}] \approx ۱۴$ و $۱۴ > (۱-۱/۳۶۵)۱۴ > \mathrm{var}[M_{۱۰۰}]$

بنابراین با توجه به قضیه چبی شف

$$\Pr\{|M_{۱۰۰} - ۱۴| \ge x\} < \frac{۱۴}{x^۲}.$$

با فرض بر اینکه $x = ۶$ نتیجه می‌گیریم که شانسی بهتر از ۵۰٪ وجود دارد که در یک کلاس ۱۰۰ نفری، تعداد جفت‌های دانشجویان با یک روز تولد در میان ۸ و ۲۰ خواهد بود.

۱۰. نمونه‌گیری دو به دو مستقل

استدلالی که در بالا برای تجزیه جمع‌آوری رأی و روز تولد یکسان استفاده کردیم خیلی شبیه به هم است. ما آن را به نرمی به شکلی کلی‌تر با یک نتیجه‌گیری اساسی که آن را قضیه نمونه‌گیری دو به دو مستقل می‌نامیم بیان می‌کنیم. بویژه، نیازی نداریم که خودمان را به حاصل‌جمع‌های متغیرهای مقدار صفر– یک محدود کنیم یا به متغیرهایی که دارای توزیع یکسان باشند. برای ساده سازی، قضیه مورد نظر را برای متغیرهای دو به دو مستقل با توزیع‌های مختلف ممکن ولی با میانگین و واریانس یکسان شرح می‌دهیم.

قضیه (نمونه‌گیری دو به دو مستقل). فرض کنید $G_1,\dots,G_n$ متغیرهای دو بدو مستقل با میانگین μ و انحراف معیار σ باشند. تعریف کنید

$$S_n ::= \sum_{i=1}^{n} G_i \tag{۱۳}$$

آن‌گاه

$$\Pr\left\{\left|\frac{S_n}{n}-\mu\right| \geq x\right\} \leq \frac{1}{n}\left(\frac{\sigma}{x}\right)^2.$$

برهان. در ابتدا مشاهده می‌کنیم که امید S_n/n عبارت است از μ:

$$E\left[\frac{S_n}{n}\right] = E\left[\frac{\sum_{i=1}^{n} G_i}{n}\right] \qquad (\text{تعریف } S_n)$$

(خطی بودن امید)
$$= \frac{\sum_{i=1}^{n} E[G_i]}{n}$$

$$= \frac{\sum_{i=1}^{n} \mu}{n}$$

$$= \frac{n\mu}{n} = \mu.$$

دومین خصوصیت مهم S_n / n آن است که واریانس آن مساوی با واریانس G_i تقسیم بر n است:

(در ٤)
$$\operatorname{var}\left[\frac{S_n}{n}\right] = \left(\frac{1}{n}\right)^2 \operatorname{var}[S_n]$$

(تعریف Sn)
$$= \frac{1}{n^2} \operatorname{var}\left[\sum_{i=1}^{n} Gi\right]$$

(جمعی بودن استقلال دو به دو)
$$= \frac{1}{n^2} \sum_{i=1}^{n} \operatorname{var}[Gi]$$

$$= \frac{1}{n^2} . n\sigma^2 = \frac{\sigma^2}{n}. \qquad (14)$$

این برای به کار بردن کرانه چبی شف و نتیجه‌گیری کافی است:

(کرانه چبی شف)
$$\Pr\left\{\left|\frac{S_n}{n} - \mu\right| \ge x\right\} \le \frac{\operatorname{var}[S_n / n]}{x^2}$$

(در ١٤)
$$= \frac{\sigma^2 / n}{x^2}$$

$$= \frac{1}{n}\left(\frac{\sigma}{x}\right)^2. \qquad \square$$

قضیه نمونه‌گیری دو به دو مستقل یک اظهار مشخص درباره اینکه چطور حد وسط نمونه‌های مستقل از متغیرهای تصادفی به میانگین نزدیک می‌شوند. بویژه، نشان می‌دهد که با انتخاب یک نمونه به اندازه کافی بزرگ ، n، می‌توانیم به صورت اختیاری محاسبات صحیح میانگین را با اعتماد نزدیک ۱۰۰٪ بدست آوریم.