هوالحق

# Stochastic Processes

## Note 1

مهدی جعفری سیاوشانی

دانشگاه صنعتی شریف

* مطالب این درس گفتار:

- پیشگفتار
- مباحث درس
- مروری بر نظریه احتمالات
  - اصول موضوعه احتمالات
  - وقایع و آزمایش های مستقل
  - متغیرهای تصادفی
  - امید ریاضی
  - گشتاورهای یک متغیر تصادفی
  - امید ریاضی شرطی
  - تابع مولد گشتاور
  - نامساوی های پایه در نظریه احتمال
  - قوانین اعداد بزرگ
  - قضیه حد مرکزی
  - همگرایی متغیرهای تصادفی

* در این درس می‌خواهیم در مورد فرایندهای تصادفی صحبت کنیم.

قاعدتاً شما قبل از این کلاس درس مقدماتی درباره احتمالات گذرانده اید. اما عموماً درس های مقدماتی احتمالات بیشتر در مورد حل کردن مسایلی که از قبل تعریف شده اند (well posed problems) صحبت می‌کنند. اما نکته مهم و اصلی در تئوری احتمالات و (فرایند های تصادفی) حل کردن well posed problems نیست بلکه کار دشوار یافتن مدل مناسب برای مسئله‌ای که در عمل و دنیای واقعی با آن سروکار داریم است. در اینجا منظور از دنیای واقعی مفهومی مقابل تئوری است.

- مبحث احتمالات چگونه در عمل شروع شد؟

- شروع احتمالات از بازی های مبتنی بر شانس بوده است. در نتیجه این بازی ها، آدم ها از نظر شهودی احساسی در مورد احتمالات به درست می‌آوردند و برایشان وقایع پر احتمال و کم احتمال معنی پیدا می‌کند.
- یکی از مسایلی که باعث می‌شود بررسی بازی های مبتنی بر شانس راحت تر از موارد دیگر که عدم قطعیت در آنها وجود دارد باشد این است که این بازی ها در عمل تکرار پذیر هستند. یعنی یک بازی (آزمایش) یکسان بارها تکرار می‌شود (ریختن تاس، پرتاب سکه، چرخاندن میز و ...) و به این ترتیب مفاهیمی مثل فرکانس نسبی، وقایع کم احتمال (odds) و پر احتمال و ... کم کم معنی پیدا می‌کنند. (مهم است که یک آزمایش به صورت مستقل بارها تکرار می‌شود).
- بسیاری از تصمیم گیری ها در زندگی شامل عدم قطعیت هستند. بنابراین در زندگی واقعی ما در مورد وقایعی که ممکن است رخ دهند حدس هایی می‌زنیم. در طول زندگی و این فرایندهای تصمیم گیری یاد می‌گیریم که به گزینه های مختلف احتمال وقوع نسبت دهیم (البته این مفهوم با مفهوم دقیق ریاضی احتمال که در ادامه در مورد آن صحبت می‌کنیم لزوماً یکی نیست) و بر اساس آن تصمیم گیری انجام دهیم.

- چند نکته مهم در مورد احتمالات

- مثال ساده پرتاب سکه را در نظر بگیرید (گاهی برای درک مسایل پیچیده خوب است با مثال های خیلی ساده شروع کنیم). در پرتاب سکه، خروجی آزمایش به سرعت و جهت گیری اولیه سکه، به مقاومت هوا، به سطحی که سکه به آن می‌خورد، به جنس و شکل سکه و ... بستگی دارد. با دانستن این عوامل و استفاده از معادلات مناسب (در اینجا معادلات مکانیک نیوتونی) هیچ پدیده تصادفی ای در این

آزمایشی وجود ندارد. با این وجود می‌شود بحث کرد که مکانیک ما تقارن دارد و ... و مدل احتمالاتی برای این آزمایش ارایه کرد که در عمل پاسخ برخی سوالات ما را بدهد.

به عنوان مثال دیگر مبحث فشرده‌سازی را در نظر بگیرید. در اینجا برای فشرده‌سازی اطلاعات از مدل‌های احتمالاتی برای داده استفاده می‌کنیم. نکته مهم اینست که معمولا داده‌هایی که با آنها سروکار داریم (مثلا متن یک کتاب، یک عکس، فایل یک سخنرانی و ...) به هیچ وجه تصادفی نیستند. ولی برای فشرده کردن داده‌ها را با یک مدل احتمالاتی و تصادفی (با ویژگی‌های معین) "مدل" می‌کنیم و این مدل برای کاربرد فشرده‌سازی جواب می‌دهد در صورتیکه داده‌ها واقعا تصادفی نیست.

- در مورد مسایل دیگر هم با چنین مسئله‌ای روبرو هستیم. مثلا اگر علاقمند باشیم وضعیت هوا را در دو سال آینده پیش‌بینی کنیم یا احتمال اینکه رشد اقتصادی ایران در پنج سال بعد بالای 7 درصد باشد را محاسبه کنیم، نکته اصلی اینست که این پدیده‌ها را چگونه مدل کنیم. در اینجا فهم تئوری احتمالات و ابزارهای ریاضی مربوطه کمکی به این مدل‌سازی نمی‌کنند و خواننده مدل‌سازی برای یک مسئله، خارج از تئوری احتمالات است.

نکته بسیار مهم اینست که وقتی برای مسئله‌ای مدل غیر مناسبی به کار ببریم جواب‌های اشتباهی به دست می‌آوریم که ممکن است به فاجعه منجر شوند (مثلا وقوع احتمال یک زلزله شدید در یک منطقه را کم بدست می‌آوریم و در طراحی ساختمانی که قرار است ساخته شود مقاوم‌سازی‌های لازم را انجام نمی‌دهیم). در اینجا محاسبات ما و ریاضیاتی که به کار برده‌ایم غلط نیستند بلکه مدل ما مدل خوبی نیست.

- سوال: چگونه یک مدل احتمالاتی برای مسایل دنیای واقعی ارایه کنیم؟

- این کار سختی است که نیاز به دانش و تجربه زیادی دارد.
- اگر مدل ما مشخص باشد استفاده از تئوری احتمالات و ابزارهای ریاضی آن روشن هستند (هرچند که ممکن است سخت باشند).
- برای شروع با مدل‌های خیلی ساده شده شروع می‌کنیم. در طول زمان مدل‌های استاندارد معرفی شده‌اند که می‌توان از آنها برای شروع کار استفاده کرد. معمولا نتایج تحلیل‌های ریاضی این مدل‌ها به ما در فهم بیشتر مسئله اصلی کمک می‌کنند. به این ترتیب در چندین مرحله می‌توان مدل

ارائه شده را به تدریج کامل‌تر کنیم تا به فهم بهتر و دقیق‌تری از مسئله مورد نظرمان برسیم.

- در عمل هیچ مدلی کامل نیست.
- Alfred North whitehead: Seek simplicity and distrust it!

- مبحث احتمالات به صورت عام و فرایندهای تصادفی به صورت خاص مبحث بسیار گسترده‌ای است که حتی آشنایی با آن به چندین درس احتیاج دارد. در این کلاس هدف ما اینست که برخی مفاهیم و ابزارهای پایه‌ای فرایندهای تصادفی را معرفی کرده به طوریکه دانشجویان بتوانند با استفاده از این دانش و مهارت‌هایی که کسب کرده‌اند موضوعات مربوط به مسئله خودشان را عمیق‌تر دنبال کنند.

- مباحثی که در این درس درباره آن صحبت می‌کنیم عبارتند از:

(3) • مروری بر احتمالات (اصول موضوعه آمار و احتمال، برخی نامساوی‌های مفید احتمالاتی، قوانین اعداد بزرگ، همگرایی)

(6) • معرفی فرایند تصادفی و مفاهیم پایه مربوط به آن (فرایندهای ایستا (Stationary)، تحلیل مبتنی بر فرایندهای تصادفی سیستم‌ها، طیف توان (Power spectrum) فرایندها و ergodicity)

(2) • فرایندهای تصادفی پواسن (Poisson)

(4) • زنجیره مارکف (Markov chain) با تعداد متناهی حالت

(6) • تئوری تخمین (estimation theory)

(2) • تئوری پیشبینی (prediction theory)

(1) • مدل‌های گرافیکی و hidden markov models (HMM)

(1) • وگشت (random walk)

(2) • مارتینگل‌ها (Martingales)

- برای شروع سوال مهم اینست که چگونه یک مدل و تئوری احتمالات ارایه کنیم که از منظر ریاضی سازگار باشد.

در حال حاضر تئوری و مجموعه اصول موضوعه‌ای که کولموگروف در سال 1933 ارایه کرد مورد پذیرش همگان است.

این اصول موضوعه در واقع تئوری احتمالات را بخشی از تئوری عمومی‌تر measure theory قرار می‌دهند.

برای رفع پارادوکس هایی که خصوصا برای تعریف احتمال بر روی فضاهای پیوسته ممکن است ایجاد شوند، استفاده از این اصول موضوعه لازم است. اما برای مباحث این درس خیلی به مباحث عمیق‌تر measure theory نیاز نداریم و وارد آن نمی‌شویم.

* مروری بر احتمالات

- اصول موضوعه احتمالات

یک مدل احتمالاتی شامل سه عنصر است:

① فضای نمونه (sample space) $\Omega$ که مجموعه همه خروجی های ممکن یک آزمایش تصادفی است.

② یک مجموعه $F$ از وقایع (event) که هر واقعه طبق تعریف یک زیر مجموعه از $\Omega$ است.

$$\forall A \in F \Rightarrow A \subseteq \Omega$$

③ یک تابع احتمال $\mathbb{P}$ که به هر واقعه یک احتمال (عددی در بازه $[0,1]$) نسبت می‌دهد.

- مجموعه وقایع $F$ خواص زیر را دارد:

① $\Omega$ یک واقعه است: $\Omega \in F$.

② اگر $A_1, A_2, \ldots$ یک دنباله از وقایع باشند آنگاه:

$$\bigcup_{i=1}^{\infty} A_i \in F$$

③ اگر $A \in F$ آنگاه $A^c \in F$.

• توجه کنید که لزوما هر زیر مجموعه $\Omega$ یک واقعه نیست و احتمال را فقط برای وقایع تعریف می‌کنیم (به اصطلاح برای زیر مجموعه‌های خوب یا بدرد بخور $\Omega$ که مجموعه $F$ باشند).

به مجموعه $F$ که خواص بالا را داشته باشد یک میدان سیگما ($\sigma$-field) می‌گوییم.

• زیر مجموعه هایی از $\Omega$ که یک واقعه نیستند زیر مجموعه های عجیبی هستند و تقریبا در همه کاربردها کاری به آنها نداریم.

• مواقعی که $\Omega$ مجموعه‌ای متناهی یا شمارا است معمولا $F$ را به صورت $F = 2^{\Omega}$ در نظر می‌گیریم.

- نتایجی که از تعریف بالا به دست می‌آیند:

- $\emptyset = \Omega^c \Rightarrow \emptyset \in F$

- اگر همه اعضای $\Omega$ را به عنوان وقایع تک عضوی در $F$ داشته باشیم، در این صورت همه زیرمجموعه‌های متناهی و شمارای $\Omega$ واقعه هستند (عضوی از $F$ هستند).

- اگر $A_1, A_2, \dots \in F$ آنگاه $A_1^c, A_2^c, \dots \in F$ بنابراین:

$$\bigcup_i A_i^c \in F \Rightarrow \left[\bigcup_i A_i^c\right]^c \in F \Rightarrow \bigcap_i A_i \in F$$

با استفاده از قانون دمورگان

بنابراین تعداد شمارای اشتراک وقایع نیز یک واقعه است.

- در نتیجه هر ترکیب شمارایی از اجتماع و اشتراک وقایع خواص یک واقعه است.

- تابع احتمال $\mathbb{P}$ که بر روی مجموعه وقایع $F$ تعریف می‌شود ($\mathbb{P}: F \to [0,1]$) در اصول موضوعه زیر صدق می‌کند:

① $\mathbb{P}[\Omega] = 1$. به $\Omega$ واقعه حتمی می‌گوییم.

② اگر $A \in F$ آنگاه $\mathbb{P}[A] \geq 0$

③ اگر $A_1, A_2, \dots$ وقایع disjoint باشند ($\forall i,j: A_i \cap A_j = \emptyset$) آنگاه:

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i] = \lim_{m\to\infty} \sum_{i=1}^{m} \mathbb{P}[A_i]$$

- نکته جالب توجه این است که همین تعداد اندک اصول موضوعه برای ساختن یک تئوری خوب احتمال کافی است.

- چند نتیجه زیر را می‌توانیم به سادگی بدست بیاوریم:

- $A \,\&\, A^c \in F \Rightarrow A \cap A^c = \emptyset \Rightarrow \mathbb{P}[\underbrace{A \cup A^c}_{=\Omega}] = \mathbb{P}[A] + \mathbb{P}[A^c]$

$$\Rightarrow \boxed{\mathbb{P}[A^c] = 1 - \mathbb{P}[A]}$$

- $\Rightarrow \boxed{\mathbb{P}[\emptyset] = 1 - \mathbb{P}[\Omega] = 0}$

- $A \,\&\, B \in F$ and $A \subseteq B$

$$\Rightarrow \mathbb{P}[B] = \mathbb{P}[B \cap \overbrace{(A \cup A^c)}^{=\Omega}] = \mathbb{P}[\overbrace{(B \cap A)}^{\text{disjoint}} \cup \overbrace{(B \cap A^c)}^{\text{disjoint}}]$$

$$= \mathbb{P}[A] + \underbrace{\mathbb{P}[B \cap A^c]}_{\geq 0} \Rightarrow \boxed{\mathbb{P}[A] \leq \mathbb{P}[B] \leq 1}$$

• Union Bound :

برای تعداد متناهی یا شمارا از وقایع $A_i$ داریم:

$$\mathbb{P}\left[\sum_i A_i\right] \leq \sum_i \mathbb{P}[A_i]$$


$$\Rightarrow \quad A_1 \cup A_2 = A_1 \cup (A_1^c \cap A_2)$$

$$\Rightarrow \quad \mathbb{P}[A_1 \cup A_2] = \mathbb{P}[A_1] + \mathbb{P}[A_1^c \cap A_2]$$

$$\leq \mathbb{P}[A_1] + \mathbb{P}[A_2]$$

- در بسیاری موارد به صورت مستقیم به اصول موضوعه احتمال نیاز پیدا نمی‌کنیم (هر چند که همیشه باید مد نظرمان باشند). بیشترین دقت را باید در مواردی انجام دهیم که با اعمالی که بر بی‌نهایت واقعه اثر می‌کنند (مانند جمع یا دوگیری) سروکار دارند.

- در درس‌های مقدماتی احتمال تاکید زیادی بر متغیرهای تصادفی و امید ریاضی وجود دارد. اما اصول موضوعه "وقایع" و "احتمال نسبت داده شده به آنها" را به عنوان مفاهیم بنیادی معرفی می‌کند.

- وقایع و آزمایش‌های مستقل

• دو واقعه $A_1$ و $A_2$ مستقل هستند اگر $\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1] \cdot \mathbb{P}[A_2]$

• مدل توامان (Combined model):

فرض کنید دو مدل احتمالی با فضاهای نمونه $\Omega_1$ و $\Omega_2$ داشته باشیم. مدل توامان این دو آزمایش تصادفی به این صورت تعریف می‌شود که ابتدا فضای نمونه $\Omega_c = \Omega_1 \times \Omega_2$ را برای مدل توامان تشکیل می‌دهیم. سپس برای هر واقعه $A \times B \subseteq \Omega_c$ احتمال را به صورت زیر تعریف می‌کنیم:

$$\mathbb{P}_c[A \times B] = \mathbb{P}_1[A] \cdot \mathbb{P}_2[B]$$

دو آزمایش تصادفی اولیه را در این مدل توامان مستقل می‌گوییم و به این ترتیب استقلال دو مدل تصادفی را تعریف می‌کنیم.

به سادگی می‌توان بررسی کرد که اگر اصول موضوعه آمار و احتمال برای هر یک از مدل‌های ابتدایی برقرار باشد، برای مدل توامان نیز برقرار است.

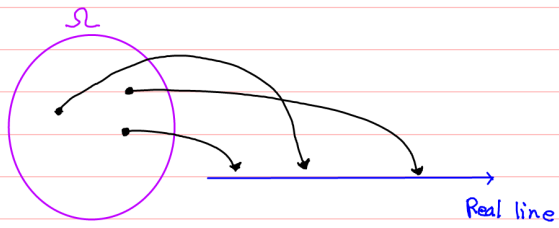
به این ترتیب با ترکیب مدل‌های ساده می‌توانیم مدل‌های پیچیده‌تر بسازیم.

- متغیرهای تصادفی

یک متغیر تصادفی $X$ (یا به بیان دقیق‌تر $X(\omega)$) یک تابع از فضای نمونه $\Omega$ به مجموعه اعداد حقیقی است. این تابع باید این خاصیت را داشته باشد که

$$\{\omega : X(\omega) \leq a\}$$

به ازای هر $a \in \mathbb{R}$ یک واقعه باشد.




- اگر بخواهیم از شکل استفاده کنیم داریم:

به بیان دیگر یک متغیر تصادفی فضای نمونه $\Omega$ را اندازه‌گیری (measure) می‌کند.

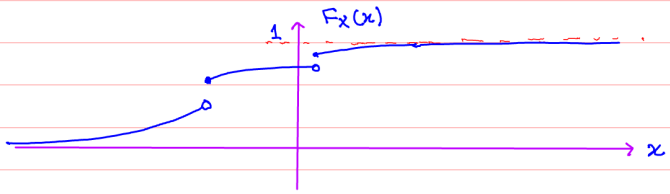
- اگر چندین متغیر تصادفی به طور همزمان بر روی یک آزمایش تصادفی تعریف شوند آنگاه باید

$$\{\omega : X_1(\omega) \leq a_1, \dots, X_n(\omega) \leq a_n\}$$

به ازای هر $a_1, \dots, a_n \in \mathbb{R}$ یک واقعه باشد (در مورد چند متغیر تصادفی در ادامه بیشتر صحبت می‌کنیم).

- برای هر متغیر تصادفی $X$ تابع توزیع (تجمعی) $F_X(x)$ به صورت زیر تعریف می‌شود

$$F_X(x) = \mathbb{P}[X(\omega) \leq x]$$




- برخی از خواص تابع توزیع تجمعی $F_X(x)$:

- این تابع یک تابع غیرنزولی است که از 0 تا 1 به خود مقدار می‌گیرد.
- $$\lim_{x \to -\infty} F_X(x) = 0$$
- $$\lim_{x \to \infty} F_X(x) = 1$$
- تابع $F_X(x)$ از سمت راست پیوسته است:

$$\lim_{x \to 0^+} F_X(x+\varepsilon) = F_X(x) \quad : \forall x \in \mathbb{R}$$

- تعریف متغیر تصادفی گسسته و پیوسته:

• اگر $X$ اعضای فضای نمونه را فقط به تعداد متناهی یا شمارا مقدار بگیرد در این صورت به آن متغیر تصادفی گسسته می‌گوییم. در این حالت برای توصیف آن می‌توان از تابع جرمی احتمال (Probability mass function) استفاده کرد که به صورت زیر تعریف می‌شود:

$$P_X(x) = \mathbb{P}[X=x]$$

در این صورت تابع $F_X(x)$ تابعی پیوسته نیست و در نقاطی که $X$ به خود مقداری را با احتمال ناصفر می‌گیرد پرش دارد. در این حالت $F_X(x)$ یک تابع پلکانی است.

• یک متغیر تصادفی <u>پیوسته</u> نامیده می‌شود اگر که یک تابع $f_X(x)$ وجود داشته باشد که برای هر $x \in \mathbb{R}$ تابع چگالی احتمال را بتوان به صورت زیر نوشت:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

اگر چنین تابعی وجود داشته باشد به آن تابع چگالی احتمال $X$ می‌گوییم.

• اگر $\frac{dF_X(x)}{dx}$ در همه نقاط موجود و متناهی باشد در این صورت $X$ را یک <u>متغیر تصادفی پیوسته</u> می‌نامند. در این صورت تابع چگالی احتمال برای متغیر $X$ به صورت زیر تعریف می‌شود:

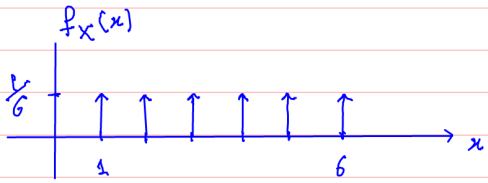
$$f_X(x) = \frac{dF_X(x)}{dx}$$

تعبیر دیگری که می‌توان برای چگالی احتمال داد اینست که

$$f_X(x) = \lim_{\delta \to 0} \frac{\mathbb{P}[x \leq X \leq x+\delta]}{\delta}$$

- نکته: گاهی برای اینکه متغیرهای تصادفی گسسته یا متغیرهای تصادفی‌ای که ترکیب گسسته و پیوسته هستند را یکجا در نظر بگیریم از تابع ضربه در چگالی احتمال این متغیرها استفاده می‌کنیم.

مثلا برای یک تاس داریم:




$$f_X(x) = \sum_{i=1}^{6} \frac{1}{6}\,\delta(x-i)$$

- اثبات اینکه $F_X(x)$ تابعی از سمت راست پیوسته است.

• برای متغیر دلخواه $X$ و عدد حقیقی $x$ و هر عدد صحیح $n \geq 1$ وقایع زیر را تعریف می‌کنیم:

$$A_n = \{\omega : X(\omega) > x + \tfrac{1}{n}\}$$

می‌دانیم برای هر $x$: $x+\frac{1}{n} \leq x+\frac{1}{n-1}$ در نتیجه

$$A_1 \subseteq A_2 \subseteq \dots$$

تعریف می‌کنیم: $B_1 = A_1$ , $B_2 = A_2 - A_1$ , $B_k = A_k - A_{k-1}$

در این صورت $B_k$ ها اشتراکی با یکدیگر ندارند (disjoint هستند).

حال داریم:

$$\mathbb{P}[\bigcup_{n\geq 1} A_n] = \mathbb{P}[\bigcup_{k\geq 1} B_k] = \sum_{k=1}^{\infty} \mathbb{P}[B_k]$$

$$= \lim_{k\to\infty} \sum_{l=1}^{k} \mathbb{P}[B_l]$$

$$= \lim_{k\to\infty} \left(\mathbb{P}[A_1] + \sum_{l=2}^{k} \mathbb{P}[A_l] - \mathbb{P}[A_{l-1}]\right)$$

$$= \lim_{k\to\infty} \mathbb{P}[A_k]$$

• با توجه به شعار مبنی که گرده‌ایم داریم

$$A_n \subseteq \underbrace{\{\omega : X(\omega) > x\}}_{\triangleq A_\infty}$$

و $A_\infty$ کوچکترین مجموعه‌ای است که چنین خاصیتی دارد. بنابراین

$$\{\omega : X(\omega) > x\} = \bigcup_{n=1}^{\infty} A_n$$

در نتیجه با استفاده از قسمت قبل داریم

$$\mathbb{P}[X > x] = \mathbb{P}[\bigcup_{n=1}^{\infty} A_n] = \lim_{n\to\infty} \mathbb{P}[A_n] = \lim_{n\to\infty} \mathbb{P}\{X > x + \tfrac{1}{n}\}$$

• می‌نویسیم:

$$\mathbb{P}[X \leq x] = \lim_{n\to\infty} \mathbb{P}[\underbrace{X \leq x + \tfrac{1}{n}}_{A_n^c}]$$

یا به عبارتی

$$\lim_{n\to\infty} F_X(x + \tfrac{1}{n}) = F_X(x)$$

چون دنباله $x_n = x + 1/n$ یک دنباله یکنوا افت کاهشی به سمت $x$ است پس حد از سمت راست $F_X(x)$ وجود دارد و برابر خود $F_X(x)$ است.

- امید ریاضی

امید ریاضی متغیر تصادفی X را با $\mathbb{E}[X]$ نشان می دهیم. در واقع امید ریاضی یک متغیر تصادفی را می توان به عنوان بهترین تقریبی در نظر بگیریم اگر که بخواهیم X را فقط با یک عدد تقریب بزنیم.

برای متغیرهای گسسته امید ریاضی به صورت زیر تعریف می شود:

$$\mathbb{E}[X] = \sum_i a_i P_X(a_i)$$

در صورتی که این جمع وجود داشته باشد می گوییم امید ریاضی X وجود دارد.

برای متغیرهای پیوسته امید ریاضی به صورت زیر تعریف می شود:

$$\mathbb{E}[X] = \int x f_X(x)\, dx$$

- برای متغیر تصادفی دلخواه و نامنفی X می توانیم امید ریاضی X را به صورت زیر هم بنویسیم:

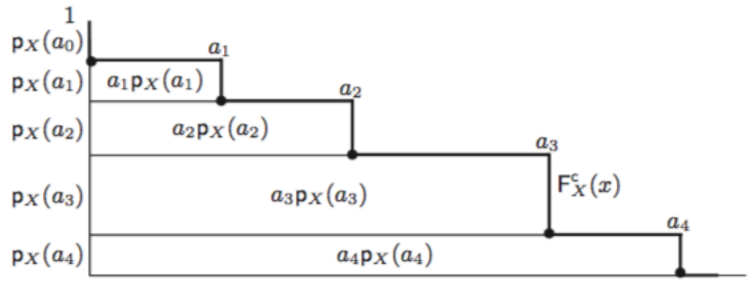
$$\mathbb{E}[X] = \int_0^\infty F_X^c(x)\, dx$$

که:

$$F_X^c(x) = 1 - F_X(x) = \mathbb{P}[X > x]$$

اگر انتگرال فوق متناهی باشد می گوییم که امید ریاضی وجود دارد و در غیر این صورت می گوییم وجود ندارد یا متناهی نیست.

شکل زیر به طور شهودی رابطه بالا را برای یک مثال خاص و متغیرهای تصادفی گسسته نشان می دهد:




The figure shows the complementary CDF $F_X^c$ of a non-negative discrete rv $X$. For this example, $X$ takes on five possible values, $0 = a_0 < a_1 < a_2 < a_3 < a_4$. Thus $F_X^c(x) = \Pr\{X > x\} = 1 - p_X(a_0)$ for $x < a_1$. For $a_1 \le x < a_2$, $\Pr\{X > x\} = 1 - p_X(a_0) - p_X(a_1)$. Similar drops occur in $\Pr\{X > x\}$ as $x$ reaches $a_2$, $a_3$, and $a_4$. From (1.28), $\mathsf{E}(X)$ is $\sum_i a_i p_X(a_i)$, which is the sum of the rectangles in the figure. This is also the area under the curve $F_X^c(x)$, i.e., $\int_0^\infty F_X^c(x)\, dx$. It can be seen that this argument applies to any non-negative rv, thus verifying (1.30).

- تعریف امید ریاضی در حالت کلی :

در حالت کلی امید ریاضی یک متغیر تصادفی دلخواه را با رابطه زیر تعریف می کنیم :

$$\mathbb{E}[X] = -\int_{-\infty}^{0} F_X(x)\,dx + \int_{0}^{\infty} F_X^c(x)\,dx$$

$$= \int_{-\infty}^{+\infty} [u(x) - F_X(x)]\,dx$$

که

$$u(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- متغیر های تصادفی به عنوان تابعی از متغیرهای تصادفی دیگر

اگر $h: \mathbb{R} \to \mathbb{R}$ یک تابع باشد و $X$ یک متغیر تصادفی، در این صورت $Y = h(X)$ هم یک متغیر تصادفی است که پیشامد $\omega$ را به تابع ترکیبی $h(X(\omega))$ می برد.

با استفاده از رابطه بالا تابع توزیع تجمعی $Y$ را می توان حساب کرد و امید ریاضی $Y$ را از روی آن بدست آورد.

ولی می توان نشان داد که در حالت گسسته امید ریاضی $Y$ برابر است با:

$$\mathbb{E}[Y] = \sum_{x} h(x)\, P_X(x)$$

و در حالت پیوسته برابر است با:

$$\mathbb{E}[Y] = \int_{-\infty}^{+\infty} h(x)\, f_X(x)\,dx$$

با استفاده از انتگرال Stieltjes دو عبارت بالا را می توان به طور همزمان به صورت زیر نوشت. یعنی برای متغیر تصادفی دلخواه $X$ داریم:

$$\mathbb{E}[Y] = \int_{-\infty}^{+\infty} h(x)\, dF_X(x)$$

در این درس در بیشتر مواقع از عبارت بالا به عنوان مختصر نویسی برای عبارت های بالا مربوط به متغیر گسسته و پیوسته استفاده می کنیم.

More specifically, the Riemann–Stieltjes integral, abbreviated here as the Stieltjes integral, is denoted as $\int_a^b h(x)dF_X(x)$. This integral is defined as the limit of a generalized Riemann sum, $\lim_{\delta\to 0}\sum_n h(x_n)[F(y_n) - F(y_{n-1})]$, where $\{y_n; n \geq 1\}$ is a sequence of increasing numbers from $a$ to $b$ satisfying $y_n - y_{n-1} \leq \delta$ and $y_{n-1} < x_n \leq y_n$ for all $n$. The Stieltjes integral is defined to exist over finite limits if the limit exists and is independent of the choices of $\{y_n\}$ and $\{x_n\}$ as $\delta \to 0$. It exists over infinite limits if it exists over finite lengths and a limit over the integration limits can be taken. See Rudin [24] for an excellent elementary treatment of Stieltjes integration, and see Exercise 1.15 for some examples.

- گشتاورها و گشتاورهای مرکزی یک متغیر تصادفی (moments and central moments):

گشتاور nام متغیر تصادفی X طبق تعریف برابر است با: $\mathbb{E}[X^n]$

گشتاور مرکزی nام متغیر تصادفی X برابر است با: $\mathbb{E}[(X-\mu_X)^n]$ که $\mu_X = \mathbb{E}[X]$

گشتاور مرکزی دوم یک متغیر تصادفی را واریانس می‌گویند: $Var[X] = \mathbb{E}[(X-\mu_X)^2]$

- یک رابطه مهم بین امید ریاضی و واریانس (یا بین امید ریاضی و انحراف معیار) اینست که

$$\mathbb{E}[(X-\alpha)^2]$$

نسبت به $\alpha$ در نقطه $\alpha = \mathbb{E}[X]$ کمینه می‌شود. در واقع می‌توان $\sqrt{\mathbb{E}[(X-Y)^2]}$ را به عنوان فاصله بین دو متغیر تصادفی در نظر گرفت (در فضای برداری متغیرهای تصادفی).

- امید ریاضی یک عملگر خطی است:

مستقل از اینکه متغیرهای تصادفی $X_1 \dots X_n$ مستقل یا وابسته باشند داریم ($a_i \in \mathbb{R}$ , $b \in \mathbb{R}$):

$$\mathbb{E}[a_1 X_1 + \dots + a_n X_n + b] = a_1 \mathbb{E}[X_1] + \dots + a_n \mathbb{E}[X_n] + b$$

- اگر X و Y دو متغیر تصادفی باشند و تعریف کنیم $Z = X+Y$، آنگاه تابع توزیع تجمعی Z را می‌توانیم به صورت زیر محاسبه کنیم:

$$F_Z(z) = \int_{-\infty}^{+\infty} F_X(z-y)\, dF_Y(y) = \int_{-\infty}^{+\infty} F_Y(z-x)\, dF_X(x)$$

اگر X و Y هر دو تابع چگالی احتمال داشته باشند، عبارت آشنای کانولوشن را برای محاسبه $f_Z(z)$ داریم:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\, dy = \int_{-\infty}^{+\infty} f_Y(z-x) f_X(x)\, dx$$

- امید ریاضی شرطی :

• X و Y هر دو متغیرهای تصادفی گسسته : اگر $P_Y(y)>0$ باشد، امید ریاضی X به شرط واقعه $\{Y=y\}$ به صورت زیر تعریف می‌شود:

$$E[X \mid Y=y] = \sum_x x\, P_{X|Y}(x|y)$$

• اگر X متغیر تصادفی دلخواه و Y متغیر تصادفی گسسته باشد و اگر $P_Y(y)>0$ تابع توزیع تجمعی X به شرط $\{Y=y\}$ برابر است با :

$$F_{X|Y}(x|y) = \frac{P[X \leq x,\ Y=y]}{P[Y=y]}$$

بنابراین امید ریاضی شرطی X به شرط $Y=y$ را به صورت زیر تعریف می‌کنیم :

$$E[X \mid Y=y] = -\int_{-\infty}^{0} F_{X|Y}(x|y)\,dx + \int_{0}^{\infty} F^c_{X|Y}(x|y)\,dx$$

• اگر X و Y هر دو پیوسته باشند امید ریاضی شرطی را می‌توانیم برحسب توابع چگالی احتمال تعریف کنیم.

به عبارت دیگر داریم :

$$E[X \mid Y=y] = \int_{-\infty}^{+\infty} x\, f_{X|Y}(x|y)\,dx$$

• نکته: می‌توان امید ریاضی شرطی را به عنوان تابعی از y در نظر گرفت مثل

$$g(y) = E[X \mid Y=y]$$

و بعد با استفاده از این تابع متغیر تصادفی جدید $g(Y)$ را تعریف کرد. این متغیر تصادفی جدید را به صورت زیر نشان می‌دهند :

$$E[X \mid Y]$$

- قضیه امید ریاضی کل (Total expectation) :

در حالت کلی این قضیه بیان می‌کند که

$$E[X] = \underset{Y}{E}\Big[\ \underset{X|Y}{E}[X \mid Y]\Big]$$

در حالت کلی قضیه فوق جزییات ترتیبی دارد که باید به آنها دقت کرد (مثل مسئله همگرایی و...)، ولی برای حالتی که X و Y هر دو گسسته یا پیوسته باشند می‌توان رابطه بالا را به صورت زیر بیان کرد.

• X و Y هر دو گسسته:

$$\mathbb{E}[X] = \mathbb{E}[\,\mathbb{E}[X|Y]] = \sum_y P_Y(y)\,\mathbb{E}[X|Y=y]$$

$$= \sum_y P_Y(y) \sum_x x P_{X|Y}(x|y)$$

• X و Y هر دو پیوسته:

$$\mathbb{E}[X] = \mathbb{E}[\,\mathbb{E}[X|Y]] = \int_{-\infty}^{+\infty} f_Y(y)\,\mathbb{E}[X|Y=y]\,dy$$

$$= \int_{-\infty}^{+\infty} f_Y(y) \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\,dx\,dy$$

- تابع مولد گشتاور (Moment generating function):

تابع مولد گشتاور متغیر تصادفی X به صورت زیر تعریف می‌شود ($r \in \mathbb{R}$):

$$\phi_X(r) = \mathbb{E}[e^{rX}] = \int_{-\infty}^{+\infty} e^{rx}\,dF_X(x)$$

$$= \int_{-\infty}^{0} e^{rx}\,dF_X(x) + \int_{0}^{+\infty} e^{rx}\,dF_X(x)$$

این دو انتگرال همیشه در نقطه $r=0$ وجود دارد و مقدارش برابر 1 است.

فرض $r_+(X)$ sup r هایی باشد که انتگرال دوم وجود دارد. داریم: $0 \leq r_+(X) \leq \infty$.

در این صورت انتگرال دوم برای $r < r_+(X)$ وجود دارد. به طریق مشابه $r_-(X)$ را inf r هایی تعریف می‌کنیم که انتگرال اول جواب داشته باشد: $-\infty \leq r_-(X) \leq 0$.

اگر $\phi_X(r)$ در یک بازه باز حول صفر وجود داشته باشد (یعنی $r_- < 0 < r_+$) آنگاه مشتقات هر مرتبه‌ای از $\phi_X(r)$ وجود دارند و داریم:

$$\frac{d^k \phi_X(r)}{dr^k} = \int_{-\infty}^{+\infty} x^k e^{rx}\,dF_X(x) \quad \Longrightarrow \quad \left.\frac{d^k \phi_X(r)}{dr^k}\right|_{r=0} = \mathbb{E}[X^k]$$

- تابع مولد گشتاور جمع چند متغیر تصادفی مستقل ($S_n = X_1 + \cdots + X_n$):

$$\phi_{S_n}(r) = \mathbb{E}[e^{rS_n}] = \mathbb{E}[e^{r(X_1+\cdots+X_n)}]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n} \exp(rX_i)\right] = \prod_{i=1}^{n} \phi_{X_i}(r)$$

با استفاده از استقلال $X_i$ ها

اگر $X_1, \ldots, X_n$ متغیر های تصادفی iid باشند داریم:

$$\phi_{S_n}(r) = [\phi_X(r)]^n$$

این رابطه بیان می‌کند توابعی که $\phi_X$ و $\phi_{S_n}$ وجود دارند یکسان هستند.

- نامساوی های پایه در نظریه احتمال:

نامساوی ها نقش مهمی در احتمالات بازی می‌کنند زیرا که بسیاری از مدل ها پیچیده‌تر از آن هستند که بتوان جواب دقیق برای شان پیدا کرد. از طرف دیگر قضایای مفید بسیاری در مورد رفتارهای حدی صعبت می‌کنند نه جواب های دقیق. در ادامه نامساوی مارکف (Markov) و چبیشف (Chebyshev) و چرنف (Chernoff) صحبت می‌کنیم.

• نامساوی مارکف:

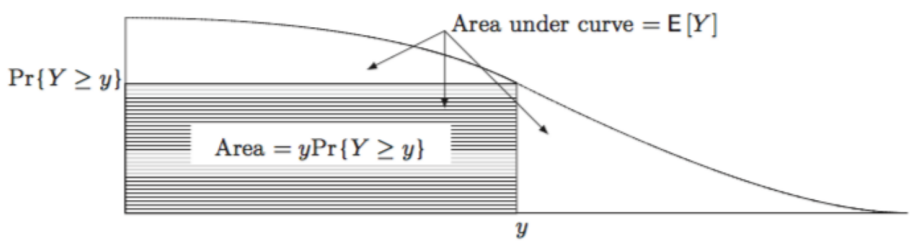
این نامساوی یکی از ساده‌ترین و پایه‌ای‌ترین نامساوی های احتمالاتی است.

اگر متغیر نامنفی $Y$ امید ریاضی $\mathbb{E}[Y]$ داشته باشد آنگاه برای هر $\alpha > 0$ داریم:

$$\mathbb{P}[Y \geq \alpha] \leq \frac{\mathbb{E}[Y]}{\alpha}$$

(برای اثبات می‌توان از این نکته استفاده کرد که: $\alpha \cdot 1\{y \geq \alpha\} \leq y$)

شکل زیر نیز روش دیگری برای درک این نامساوی است.




Demonstration that $y\Pr\{Y \geq y\} \leq \mathsf{E}[Y]$. By letting $y \to \infty$, it can also be seen that the shaded area becomes a negligible portion of the area $\mathsf{E}[Y]$, so that $\lim_{y\to\infty} y\Pr\{Y > y\} = 0$ if $\mathsf{E}[Y] \leq \infty$.

• نامساوی چبیشف

حال با استفاده از نامساوی مارکف نامساوی دیگری به دست می آوریم.

فرض کنید $Z$ متغیر تصادفی دلخواهی با امید ریاضی و واریانس متناهی باشد. متغیر نامنفی $Y$ را به صورت $Y=(Z-\mathbb{E}[Z])^2$ تعریف می کنیم. بنابراین $\mathbb{E}[Y]=\sigma_Z^2$ پس با استفاده از نامساوی مارکف داریم:

$$\mathbb{P}[(Z-\mathbb{E}[Z])^2 \geq k] \leq \frac{\sigma_Z^2}{k} \quad : \quad \forall\, k>0$$

رابطه بالا را می توان به شکل زیر بازنویسی کرد:

$$\mathbb{P}[\,|Z-\mu_Z| \geq \varepsilon\,] \leq \frac{\sigma_Z^2}{\varepsilon^2} \quad : \quad \forall\, \varepsilon>0$$

• چند نکته:

Note that the Markov inequality bounds just the upper tail of the CDF and applies only to non-negative rv s, whereas the Chebyshev inequality bounds both tails of the CDF. The more important differences, however, are that the Chebyshev bound requires a finite variance and approaches zero inversely with the square of the distance from the mean, whereas the Markov bound does not require a finite variance and goes to zero inversely with the distance from 0 (and thus asymptotically with distance from the mean).

• کران چرنف:

دیدیم که نامساوی چبیشف از اعمال نامساوی مارکف بر متغیر تصادفی $(Z-\mu_Z)^2$ به دست آمد.

به طریق مشابهی کران چرنف هم از اعمال نامساوی مارکف بر متغیر تصادفی $e^{rZ}$ برای یک $r$ داده شده به دست می آید.

کران چرنف به صورت نمایی با افزایش فاصله از میانگین به سمت صفر میل می کند.

برای یک متغیر دلخواه $Z$ فرض کنید $I(Z)$ بازه‌ای باشد که تابع مولد گشتاور $Z$ به آن تعریف شده باشد.

با قرار دادن $Y=e^{rZ}$ برای $r\in I(Z)$ و استفاده از نامساوی مارکف داریم:

$$\mathbb{P}[\,e^{rZ} \geq \alpha\,] \leq \frac{\phi_Z(r)}{\alpha} \quad : \quad \forall\, \alpha>0$$

برای اینکه رابطه‌ای معنادارتر بیان کنیم قرار می دهیم $\alpha=e^{r\beta}$. بنابراین برای هر $\beta\in\mathbb{R}$ داریم:

$$\mathbb{P}[\,Z \geq \beta\,] \leq \phi_Z(r)\, e^{-r\beta} \quad : \quad \forall\, r>0,\ r\in I(Z)$$

$$\mathbb{P}[\,Z \leq \beta\,] \leq \phi_Z(r)\, e^{-r\beta} \quad : \quad \forall\, r<0,\ r\in I(Z)$$

(به ازای $r=0$ هر دو کران فوق نتیجه‌ای بدیهی می دهند).

در حالت کلی برای بدست آوردن بهترین کران باید عبارات فوق را نسبت به r بهینه کنیم.

• یکی از کاربردهای مهم کران چرنف برای جمع متغیرهای iid است. فرض کنید

$$S_n = X_1 + \cdots + X_n \quad , \quad \phi_{X_i}(r) = \phi_X(r)$$

$$\Rightarrow \phi_{S_n}(r) = [\phi_X(r)]^n$$

بنابراین روابط بالا را می‌توان به صورت زیر نوشت:

$$\mathbb{P}[S_n \geq na] \leq [\phi_X(r)]^n e^{-rna} \quad : \quad r>0 , r \in I(X)$$

$$\mathbb{P}[S_n \leq na] \leq [\phi_X(r)]^n e^{-rna} \quad : \quad r<0 , r \in I(X)$$

**Lemma 1.6.1** *Let $\{X_i;\ i \geq 1\}$ be IID rv s and let $S_n = X_1 + \cdots + X_n$ for each $n \geq 1$. Let $I(X)$ be the interval over which the semi-invariant MGF $\gamma_X(r)$ is finite and assume 0 is in the interior of $I(X)$. Then, for each $n \geq 1$ and each real number $a$,*

$$\Pr\{S_n \geq na\} \leq \exp(n\mu_X(a)) \qquad \text{where } \mu_X(a) = \inf_{r\geq 0,\, r\in I(X)} \gamma_X(r) - ra. \qquad (1.66)$$

*Furthermore, $\mu_X(a) < 0$ for $a > \overline{X}$ and $\mu_X(a) = 0$ for $a \leq \overline{X}$.*

• در عمل معمولا از بایندهای کمتر tight ولی راحت‌تر زیر استفاده می‌کنیم:

$$\mathbb{P}[S_n \geq (1+\delta)\mathbb{E}[S_n]] \leq \exp\left(-\frac{\delta^2 \mathbb{E}[S_n]}{3}\right) , \quad 0<\delta<1$$

$$\mathbb{P}[S_n \leq (1-\delta)\mathbb{E}[S_n]] \leq \exp\left(-\frac{\delta^2 \mathbb{E}[S_n]}{2}\right) , \quad 0<\delta<1$$

- قوانین اعداد بزرگ

قوانین اعداد بزرگ به مجموعه‌ای از نتایج در تئوری احتمالات می‌گویند که رفتار میانگین حسابی (arithmetic average) n متغیر تصادفی را برای n های بزرگ توصیف می‌کند.

برای n متغیر تصادفی $X_1$ و ... و $X_n$ میانگین حسابی آنها به صورت زیر تعریف می‌شود:

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

از آنجایی که در هر تکرار یک آزمایش تصادفی مقدار نمونه (Sample Value) این متغیر تصادفی میانگین حسابی مقادیر متغیرهای $X_1$ ... $X_n$ در آن آزمایش است به این متغیر تصادفی میانگین نمونه (Sample avg.) هم گفته می‌شود.

اگر $X_1$ و ... و $X_n$ را به عنوان متغیرهای متوالی‌ای در زمان در نظر بگیریم، این میانگین نمونه، میانگین زمانی نیز نامیده می‌شود.

تحت شرایط نسبتا عمومی، انحراف معیار میانگین نمونه با افزایش n به سمت 0 میل می‌کند. به این معنی که میانگین نمونه به امید ریاضی میل می‌کند.

این نتایج نقش محوری در مبحث فرایندهای تصادفی دارند زیرا که اجازه می‌دهند متوسط زمانی (یعنی میانگین زمانی بر روی یک مسیر نمونه (Sample path)) را به متوسط گروهی (ensemble avg.) (یعنی امید ریاضی یک فرایند در یک زمان خاص) مربوط کنیم.

در اینجا در مورد قانون ضعیف اعداد بزرگ (weak law of large number) صحبت می‌کنیم.

همچنین اشاره مختصری به قانون قوی اعداد بزرگ می‌کنیم. در مورد قضیه حد مرکزی هم صحبت خواهیم کرد. هم به علت اینکه فهم ما از قانون اعداد بزرگ را افزایش می‌دهد و هم به علت اهمیتی که مستقلا دارد.

- قانون ضعیف اعداد بزرگ با واریانس متناهی:

فرض کنید $X_1$ و ... و $X_n$ متغیرهای iid باشند که امید ریاضی $\mu$ و واریانس $\sigma^2$ دارند. تعریف می‌کنیم:

$$S_n = X_1 + \cdots + X_n$$

و میانگین نمونه را به صورت $\frac{S_n}{n}$ در نظر می‌گیریم. چون $X_i$ ها مستقل از یکدیگر هستند، داریم:

$$\sigma^2_{S_n} = n\sigma^2$$

بنابراین برای واریانس $\frac{S_n}{n}$ داریم:

$$Var\left[\frac{S_n}{n}\right] = \mathbb{E}\left[\left(\frac{S_n - n\mu}{n}\right)^2\right] = \frac{1}{n^2}\,\mathbb{E}\left[(S_n - n\mu)^2\right] = \frac{\sigma^2}{n}$$

همانطورکه می بینیم واریانس میانگین نمونه $\frac{S_n}{n}$ با افزایش n به سمت ۰ می رود.

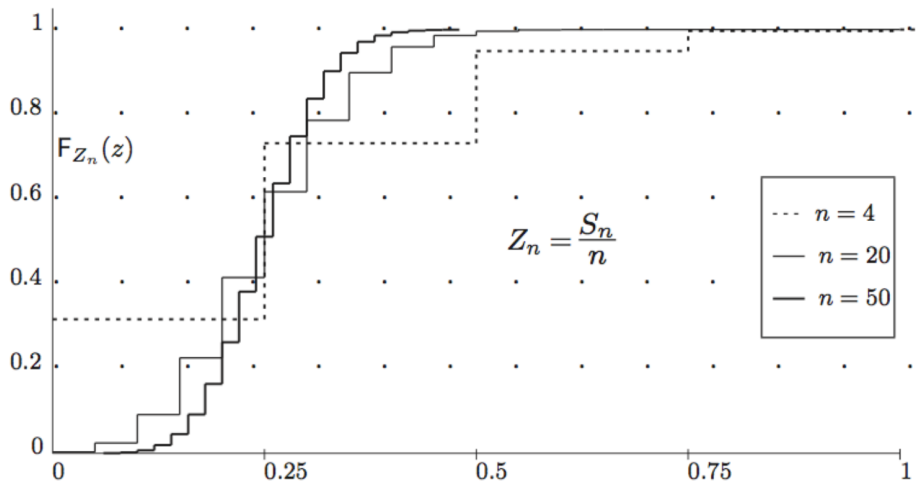
(توجه کنید که واریانس $S_n$ با افزایش n به سمت ∞ می رود ولی واریانس $\frac{S_n}{n}$ به سمت ۰ می رود).

از عبارت بالا نتیجه می گیریم:

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = 0$$

در این شرایط می گوییم $\frac{S_n}{n}$ به صورت mean square به $\mu$ (امید ریاضی $X_i$ ها) میل می کند.

شکل زیر نحوه کاهش واریانس $\frac{S_n}{n}$ با افزایش n را نشان می دهد:




The same CDF as Figure 1.5, scaled differently to give the CDF of the sample average $Z_n$. It can be visualized that as $n$ increases, the CDF of $Z_n$ becomes increasingly close to a unit step at the mean, 0.25, of the variables $X$ being summed.

همگرایی به صورت mean square بیان می کند که میانگین نمونه $Z_n = \frac{S_n}{n}$ با امید ریاضی $\mathbb{E}[X_i]=\mu$ به اندازه یک متغیر تصادفی اختلاف دارد که با افزایش n واریانس آن به سمت ۰ می رود.

در اینجا درباره همگرایی یک دنباله از متغیرهای تصادفی (یک دنباله از توابع) که به یک عدد ثابت میل می کنند صحبت می کنیم که مسئله پیچیده تری از همگرایی یک دنباله از اعداد به یک عدد ثابت است. در ادامه این نوشتار در این باره بیشتر صحبت خواهیم کرد.

قانون اعداد بزرگ این ایده مرکزی را به طریق بنیادی تری برجسته می کند.

فرض کنید نامساوی چبیشف را به میانگین نمونه $\frac{S_n}{n}$ اعمال کنیم:

$$\mathbb{P}\left[\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right] \leq \frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \qquad (1.74)$$

این عبارت کران بالایی برای احتمال اینکه $\frac{S_n}{n}$ بیشتر از $\varepsilon$ با $\mu$ متفاوت باشد بیان می‌کند.

شکل قبل که تابع توزیع تجمعی $\frac{S_n}{n}$ را برای چندین $n$ رسم کرده این مسئله را نشان می‌دهد. شکل فوق بیان می‌کند که:

$$\lim_{n\to\infty} F_{S_n/n}(z)=0 \quad : \quad z<\mu$$

$$\lim_{n\to\infty} F_{S_n/n}(z)=1 \quad : \quad z>\mu$$

قضیه زیر که به قانون ضعیف اعداد بزرگ (WLLN) مشهور است این مسئله را به نحو روشن‌تری بیان می‌کند.

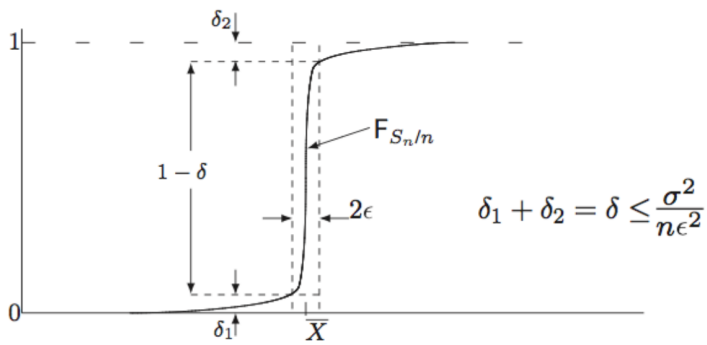
- قضیه (قانون ضعیف اعداد بزرگ با واریانس متناهی):

برای هر $n\geq 1$ فرض کنید $S_n = X_1+\cdots+X_n$ جمع $n$ متغیر تصادفی iid با واریانس متناهی باشد. در این صورت:

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right]=0 \quad : \quad \forall\, \varepsilon>0$$

برای اثبات می‌توان از نامساوی چبیشف که در بالا نوشته شده استفاده کرد.

البته شکل زیر نیز به فهم بهتر مطلب کمک می‌کند که چرا CDF $\frac{S_n}{n}$ به تابع پله میل می‌کند.




Approximation of the CDF $F_{S_n/n}$ of a sample average by a step function at the mean. From (1.74), the probability $\delta$ that $S_n/n$ differs from $\overline{X}$ by more than $\epsilon$ (i.e., $\Pr\{|S_n/n-\overline{X}|\geq\epsilon\}$) is at most $\sigma^2/n\epsilon^2$. The complementary event, where $|S_n/n-\overline{X}|<\epsilon$, has probability $1-\delta\geq 1-\sigma^2/n\epsilon^2$. This means that we can construct a rectangle of width $2\epsilon$ centered on $\overline{X}$ and of height $1-\delta$ such that $F_{S_n/n}$ enters the rectangle at the lower left (say at $(\overline{X}-\epsilon,\delta_1)$) and exits at the upper right, say at $(\overline{X}+\epsilon, 1-\delta_2)$). Now visualize increasing $n$ while holding $\epsilon$ fixed. In the limit, $1-\delta\to 1$ so $\Pr\{|S_n/n-\overline{X}|\geq\epsilon\}\to 0$. Since this is true for every $\epsilon>0$ (usually with slower convergence as $\epsilon$ gets smaller), $F_{S_n/n}(z)$ approaches 0 for every $z<\overline{X}$ and approaches 1 for every $z>\overline{X}$, i.e., $F_{S_n/n}$ approaches a unit step at $\overline{X}$. Note that there are two 'fudge factors' here, $\epsilon$ and $\delta$ and, since we are approximating an entire CDF, neither can be omitted, except by directly going to a limit as $n\to\infty$.

$$F_{S_n/n}(\mu+\varepsilon) - F_{S_n/n}(\mu-\varepsilon) = \mathbb{P}\left[\left|\frac{S_n}{n}-\mu\right|<\varepsilon\right] = 1-\delta \geqslant 1-\frac{\sigma^2}{n\varepsilon^2}$$

به همگرایی‌ای که در قضیه بالا مطرح شده است همگرایی در احتمال (in probability) می‌گوییم.

یک نکته مهم در مورد رابطه (1.74) اینست که کران فوق خیلی شل (loose) است. در واقع اگر تابع مولد گشتاور $X_i$ در یک بازه حول 0 وجود داشته باشد می‌توان نامساوی چبیشف را توسط کران چرنف خیلی بهبود بخشید.

$$\mathbb{P}\left[\left|\frac{S_n}{n}-\mu\right|\geqslant\varepsilon\right] \leqslant \underbrace{\mathbb{P}\left[\frac{S_n}{n}\geqslant\mu+\varepsilon\right]}_{} + \underbrace{\mathbb{P}\left[-\frac{S_n}{n}\geqslant-\mu+\varepsilon\right]}_{}$$

به صورت نمایی نسبت به $n$ کوچک می‌شوند.

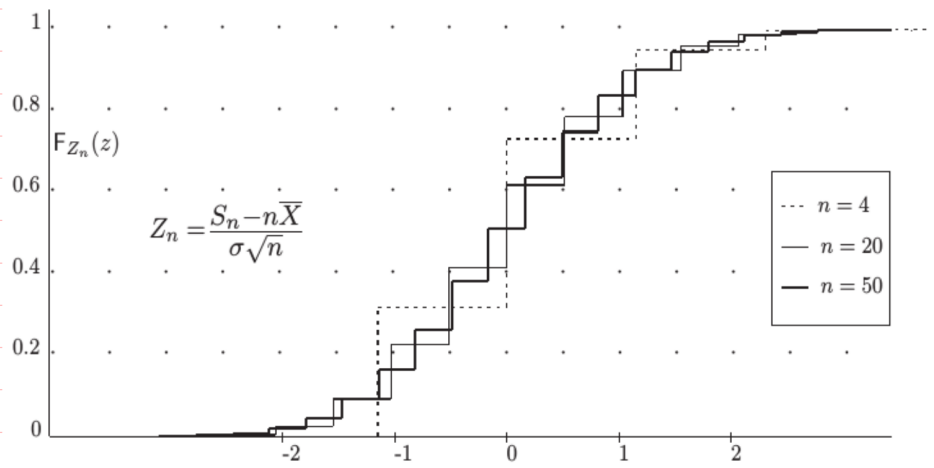
- قضیه حد مرکزی (Central limit theorem):

قانون ضعیف اعداد بزرگ بیان می‌کند که با احتمال زیاد $\frac{S_n}{n}$ نزدیک $\mu$ است برای $n$ های بزرگ. همانطور که در شکل های قبل دیدیم تابع توزیع تجمعی $S_n/n$ با افزایش $n$ حول $\mu$ فشرده و فشرده‌تر می‌شود تا در نهایت در حد $n\to\infty$ به صورت یک تابع پله در بیاید.

حال اگر متغیر $\frac{S_n}{n}$ را به یک متغیر با امید ریاضی 0 و واریانس 1 نرمالیزه کنیم متغیر نرمالیزه شده زیر را داریم:

$$Z_n = \frac{\frac{S_n}{n}-\mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

تابع توزیع تجمعی $Z_n$ برای همان مقادیر $n$ که در شکل های قبلی استفاده شده است را در شکل زیر می‌بینیم. تفاوت این شکل با شکل قبل در این است که محور افقی آن مقیاس شده است.




The same CDFs as Figure 1.5 normalized to 0 mean and unit standard deviation, i.e., the CDFs of $Z_n = (S_n/n - \overline{X})\sqrt{n}/\sigma$ for $n = 4, 20, 50$. Note that as $n$ increases, the CDF of $Z_n$ slowly starts to resemble the normal CDF.

با مشاهده شکل قبل، به نظر می‌آید که تابع توزیع تجمعی متغیر نرمالیزه شده $Z_n$ به یک توزیع خاص میل می‌کند.

این مسئله در قضیه زیر بیان شده است.

- قضیه حد مرکزی (Central Limit Theorem):

فرض کنید $X_1$ و ... و $X_n$ متغیرهای تصادفی iid با امید ریاضی متناهی $\mu$ و واریانس متناهی $\sigma^2$ باشند. برای هر $z \in \mathbb{R}$ داریم:

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right] = \Phi(z)$$

که $\Phi(z)$ تابع توزیع تجمعی توزیع نرمال است:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

قضیه بالا بیان می‌کند که:

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z) \quad : \forall z \in \mathbb{R}$$

به این نوع همگرایی، همگرایی در توزیع (Convergence in dist.) گفته می‌شود، چرا که به جای اینکه دنباله‌ای از متغیرهای تصادفی همگرا شوند دنباله‌ای از توابع توزیع تجمعی همگرا می‌شوند.

چند نکته:

- قضیه حد مرکزی در مورد نحوه میل کردن تابع توزیع تجمعی $\frac{S_n}{n}$ به تابع پله بیشتر اطلاعات می‌دهد. در واقع با افزایش $n$، $\frac{S_n}{n}$ نزدیک به توزیع $N(\mu, \frac{\sigma^2}{n})$ می‌شود.
- لغت "مرکزی" در نام قضیه اشاره به این دارد که برخلاف کران چرنف در این قضیه احتمال اینکه میانگین نمونه به اندازه مقداری که به صورت $\frac{1}{\sqrt{n}}$ مقیاس می‌شود از امید ریاضی فاصله داشته باشد محاسبه می‌گردد. در کران چرنف این مقدار مقاومت یک عدد ثابت است که به کران بالایی برای احتمال منتهی می‌شود که به صورت نمایی نسبت به $n$ کاهش می‌یابد.
- قضیه مرکزی توضیح می‌دهد که چرا در بسیاری از مسایل فرض ضمنی اینکه متغیر خاصی توزیع نرمال داشته باشد، به نتایج معقولی منتهی می‌شود.
- باید دقت کرد که در مورد قضیه حد مرکزی برداشت بیشتر از چیزی که ممکن است نکنیم. به عنوان مثال متغیر $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ لزوما تابع چگالی احتمالی که با تقریب گاوسی باشد ندارد.

اگر متغیرهای $X_i$ گسسته باشند این مجموع نرمالیزه شده یک متغیر تصادفی گسسته است و تابع چگالی احتمال برای آن وجود ندارد. ولی به علت انتگرال گیری اثر این گسستگی ها محو می‌شود و تابع توزیع تجمعی این متغیر به تابع توزیع تجمعی نرمال میل می‌کند.

- همگرایی متغیرهای تصادفی:

تا اینجا در قانون ضعیف اعداد بزرگ و قضیه حد مرکزی دو نوع همگرایی دنباله‌ای از متغیرهای تصادفی را دیده‌ایم. دنباله‌ای از متغیرهای تصادفی عملا دنباله‌ای از توابع هستند که بر روی فضای نمونه تعریف شده‌اند و همگرایی آنها پیچیده‌تر از همگرایی دنباله‌ای از اعداد است.

در ادامه چهار نوع همگرایی برای متغیرهای تصادفی تعریف می‌کنیم که عبارتند از:

همگرایی در توزیع، همگرایی در احتمال، همگرایی با معیار mean square و همگرایی با احتمال 1.

• تعریف: همگرایی در توزیع (Convergence in dist.)

می‌گوییم یک دنباله از متغیرهای تصادفی $X_1, \dots, X_n$ در توزیع به متغیر تصادفی $X$ همگرا می‌شوند اگر که:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

به ازای همه $x$ هایی که $F_X(x)$ پیوسته باشد.

در قضیه حد مرکزی متغیرهایی که در توزیع همگرا می‌شوند $\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}}, n\geq 1\right\}$ هستند که به توزیع نرمال همگرا می‌شوند.

دقت کنید که این همگرایی به این معنی نیست که متغیرهای تصادفی دنباله به یک متغیر خاص میل می‌کنند بلکه فقط توزیع آنها به یک توزیع خاص میل می‌کند.

به عنوان مثال دنباله‌ای از متغیرهای تصادفی iid $Y_1, \dots, Y_n$ را با تابع توزیع تجمعی $F_Y$ در نظر بگیرید. برای هر $n\geq 1$ تعریف می‌کنیم:

$$Z_n = Y_n + \frac{1}{n}$$

به راحتی می‌توان بررسی کرد که دنباله متغیرهای $Z_n$ در توزیع به $Y$ میل می‌کنند. هر چند که اگر فرض

کنیم $Y$ از $Z_i$ ها مستقل باشد، آنگاه $Z_n \sim Y$ واریانس $2\sigma_Y^2$ دارد. بنابراین $Z_n$ با افزایش $n$ به $Y$ نزدیک نمی‌شود. همچنین $Z_n - Z_m$ نیز با افزایش $m$ و $n$ کوچک نمی‌شود.

$$\frac{S_n - n\mu}{\sigma\sqrt{n}}$$

حرف مشابهی نیز برای قضیه حد مرکزی می‌توان زد که هر کدام از جملات دنباله حتی با افزایش $n$ به هم نزدیک نمی‌شوند و به متغیر خاصی هم میل نمی‌کنند.

- تعریف: همگرایی در احتمال (Convergence in prob.)

می‌گوییم یک دنباله از متغیرهای تصادفی $X_1, \dots, X_n$ در احتمال به متغیر $X$ همگرا می‌شوند اگر که:

$$\lim_{n\to\infty} \mathbb{P}[\,|X_n - X| > \varepsilon\,] = 0 \quad : \forall \varepsilon > 0$$

همگرایی در احتمال نسبت به همگرایی در توزیع اطلاعات بیشتری در مورد دنباله متغیرهای تصادفی می‌دهد. به عنوان مثال اگر $\{X_n \,,\, n \geq 1\}$ در احتمال به $0$ میل کند آنگاه برای $m$ و $n$ به اندازه کافی بزرگ $X_m$ و $X_n$ با احتمال زیاد نزدیک $0$ هستند و بنابراین نزدیک یکدیگر هستند:

$$\lim_{m\to\infty, n\to\infty} \mathbb{P}[\,|X_m - X_n| > \varepsilon\,] = 0 \quad : \forall \varepsilon > 0$$

به بیان دقیق‌تر داریم: $|X_m - X_n| \leq |X_m - 0| + |X_n - 0|$

بنابراین:

$$\begin{aligned}
\mathbb{P}[\,|X_m - X_n| > \varepsilon\,] &\leq \mathbb{P}[\,|X_m - 0| + |X_n - 0| > \varepsilon\,] \\
&\leq \mathbb{P}[\,|X_m - 0| > \tfrac{\varepsilon}{2} \ \text{ or } \ |X_n - 0| > \tfrac{\varepsilon}{2}\,] \\
&\leq \mathbb{P}[\,|X_m - 0| > \tfrac{\varepsilon}{2}\,] + \mathbb{P}[\,|X_n - 0| > \tfrac{\varepsilon}{2}\,] \\
&\xrightarrow{n\to\infty} 0
\end{aligned}$$

- تعریف: همگرایی با معیار mean square:

می‌گوییم یک دنباله از متغیرهای تصادفی $X_1, \dots, X_n$ با معیار MS به متغیر $X$ همگرا می‌شوند اگر:

$$\lim_{n\to\infty} \mathbb{E}[\,(Z_n - Z)^2\,] = 0$$

توجه کنید که همگرایی با معیار MS همگرایی در احتمال را نتیجه می‌دهد.

• تعریف: همگرایی با احتمال 1 ( Convergence with prob. 1 ):

به این نوع همگرایی almost surely و almost everywhere هم گفته می‌شود.

فرض کنید $X_1$ و ... و $X_n$ دنباله‌ای از متغیرهای تصادفی بر روی فضای نمونه $\Omega$ باشند و

$X$ نیز متغیر تصادفی دیگری بر روی $\Omega$ باشد.

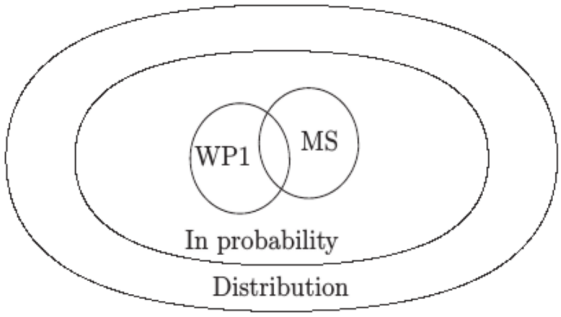
می‌گوییم $\{X_n, n \geq 1\}$ به $X$ با احتمال 1 همگرا می‌شود اگر:

$$\mathbb{P}\left[\omega \in \Omega : \lim_{n \to \infty} Z_n(\omega) = Z(\omega)\right] = 1$$

به ازای هر $\omega \in \Omega$ یک مسیر نمونه ( Sample path ) که دنباله‌ای از اعداد است داریم. این دنباله از اعداد ممکن است به عدد $Z(\omega)$ همگرا شود یا به عدد دیگری همگرا شود یا اصلا همگرا نشود.

همگرایی با احتمال 1 می‌گوید که مجموعه‌ای از پیشامدهای $\omega$ که مسیر نمونه $Z_n(\omega)$ به $Z(\omega)$ میل می‌کند، واقعه‌ای با احتمال 1 است.

شکل زیر رابطه بین انواع همگرایی برای متغیرهای تصادفی را نشان می‌دهد:




Relationship between different kinds of convergence: Convergence in distribution is the most general and is implied by all the others. Convergence in probability is the next most general and is implied both by convergence with probability 1 (WP1) and by mean-square (MS) convergence, neither of which implies the other.

مثال زیر دنباله‌ای از متغیرهای تصادفی تعریف می‌کند که در احتمال همگرا می‌شوند ولی با احتمال 1 همگرا نمی‌شوند.

**Example 1.7.9** Consider a sequence $\{Y_n;\ n \geq 1\}$ of rv s for which the sample paths constitute the following slight variation of the sequence of real numbers in Figure 1.16. In particular, as illustrated in Figure 1.17, the non-zero term at $n = 5^j$ in Figure 1.16 is replaced by a non-zero term at a randomly chosen $n$ in the interval[34] $[5^j, 5^{j+1})$.

Since each sample path contains a single one in each segment $[5^j, 5^{j+1})$, and contains zeroes elsewhere, none of the sample paths converges. In other words, $\Pr\{\omega : \lim Y_n(\omega) = 0\} = 0$. On the other hand, $\Pr\{Y_n = 0\} = 1 - 5^{-j}$ for $5^j \leq n < 5^{j+1}$, so $\lim_{n\to\infty} \Pr\{Y_n = 0\} = 1$.

Thus this sequence of rv s converges to 0 in probability, but does not converge to 0 WP1. This sequence also converges in mean square and (since it converges in probability) in distribution. Thus we have shown (by example) that convergence WP1 is not implied by any of the other types of convergence we have discussed. We will show in Section 5.2 that convergence WP1 does imply convergence in probability and in distribution but not in mean square (as illustrated in Figure 1.14).

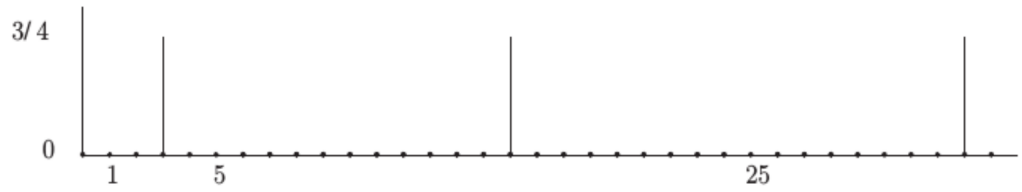



**Figure 1.17** Illustration of a sample path of a sequence of rv s $\{Y_n; n \geq 0\}$ where, for each $j \geq 0$, $Y_n = 1$ for an equiprobable choice of $n \in [5^j, 5^{j+1})$ and $Y_n = 0$ otherwise.