جلسه‌ی هشتم آمار و احتمال مهندسی - دکتر شریفی زارچی - دانشگاه شریف

مرور

متغیر تصادفی گسسته Discrete Random Variable

که الگوی خاصی پیروی می‌کنند ← توزیع (Distribution) ← PMF / CDF خاصی دارند

توزیع برنولی: پرتاب سکه

$$\mathbb{P}_{(X=x)} \begin{cases} p \\ 1-p \end{cases} \qquad X \sim Ber(p) \Rightarrow \begin{cases} E(X) = p \\ Var(X) = pq = p(1-p) \end{cases}$$

توزیع دوجمله‌ای: n بار پرتاب سکه (هر پرتاب مستقل)

$$PMF_{(n=k, X=n)} = \binom{n}{k} p^k (1-p)^{n-k}$$

$$X \sim Binomial(n,p) \Rightarrow \begin{cases} E(X) = np \\ Var = npq \end{cases}$$

توزیع هندسی

کنجکاوی: در یک دهکده دوست دارن دختر داشته باشن. با احتمال p بچه دختر میشه و با 1-p پسر. هر خانواده اونقدر بچه دار میشن تا اولین دختر به دنیا بیاد. تعداد فرزندان هر خانواده به طور متوسط چند تاست؟

ارسال بسته بین دو سوییچ. با احتمال p موفق و با 1-p شکست. به طور متوسط چند بار بسته ارسال می‌شود تا یک بار بسته سالم ارسال بشه؟

به سوالات مثل بالا می‌گوییم توزیع هندسی - Geometric با پارامتر تصادفی p

$X \sim Geometric(p)$

$$PMF_{(X=k)} = (1-p)^{k-1} p$$

k-1 بار شکست و بار نهایی موفقیت

$$E(X) = \sum_{k=1}^{\infty} k\,\mathbb{P}(X=k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p = \frac{1}{p}$$

اثبات

$$= 1 \times (1-p)^0 p + 2 \times (1-p)^1 p + 3 \times (1-p)^2 p + \dots$$

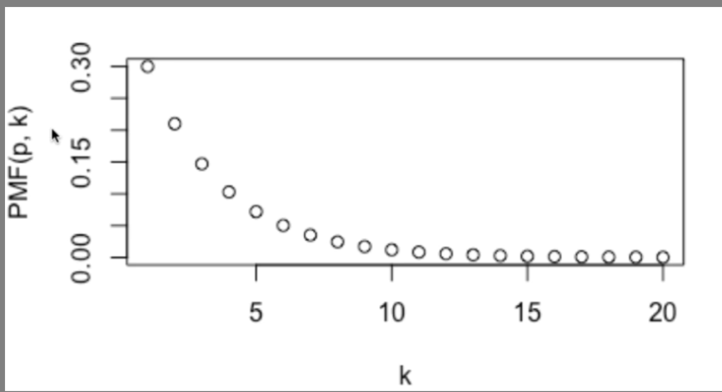
$$= [p + (1-p)p + (1-p)^2 p + \dots] + [(1-p)p + (1-p)^2 p + \dots] + [(1-p)^2 p + (1-p)^3 p + \dots] + \dots$$

$$= \frac{p}{p} + \frac{p(1-p)}{p} + \frac{(1-p)^2 p}{p} + \dots = \frac{p + p(1-p) + p(1-p)^2 + \dots}{p} = \frac{\frac{p}{1-(1-p)}}{p} = \frac{\frac{p}{p}}{p} = \frac{1}{p}$$

از نظر شهودی یعنی وقتی احتمال موفقیت 1 هست پس باید 1 بار تلاش کنیم موفق می‌شویم. زمانی که احتمال موفقیت $\frac{1}{2}$ است یعنی از 2 بار تلاش یکی منجر به موفقیت می‌شود پس 2 بار تلاش می‌کنیم و ...

با انتگرال گیری ثابت میشه:

$$Var(X) = \sum_{k=1}^{\infty} (1-p)^{k-1} p \left(k-\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$



شکل کلی توزیع هندسی این مدلیه با عدد p های متفاوت فقط شیب و اعداد تغییر می‌کنن و نمای کلی نمودار همینه.

نکته

تعاریف متفاوتی برای توزیع هندسی داریم. یک تعریف دیگه، تعداد شکست‌ها تا قبل از اولین موفقیت هست که در هر مرجع یا هر مقاله ممکنه تعریف توزیع هندسی متفاوت باشه پس همواره باید بهش دقت کرد.

توزیع فوق هندسی (Hyper Geometric Distribution)

کنجکاوی: در یک دهکده بیماری اومده. از بین کل اعضای دهکده 20 نفر بیمار شدن. در دهکده 30 تا مرد و 20 تا خانوم داریم. و بیماری ربطی به جنسیت نداره. چند تا از آقایون بیمار هستند؟

$X =$ تعداد آقایان بیمار

بیمار $= 20$

کل جمعیت $= 50$

30 آقا ، 20 خانوم

$$PMF(X=k) = ?$$

تست دارو ناشی از توزیع. 100 بیمار داریم که به دو قسمت تقسیم می‌کنیم. به یک دسته داروی اصلی رو می‌دیم و به یک دسته داروی بی‌اثر. 25 نفر از دسته اول و 16 نفر از دسته دوم خوب شدن. آیا دارو قابل اعتماد هست؟

N_1 توپ قرمز و N_2 توپ آبی. n توپ را بیرون می‌آوریم. احتمال اینکه x توپ قرمز باشد؟

$$PMF_{(X=x)} = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}$$

N: تعداد کل توپ‌ها $N = N_1 + N_2$

$X = X_1 + X_2 + \dots + X_n$: این احتمال‌ها از هم مستقل نیستند.

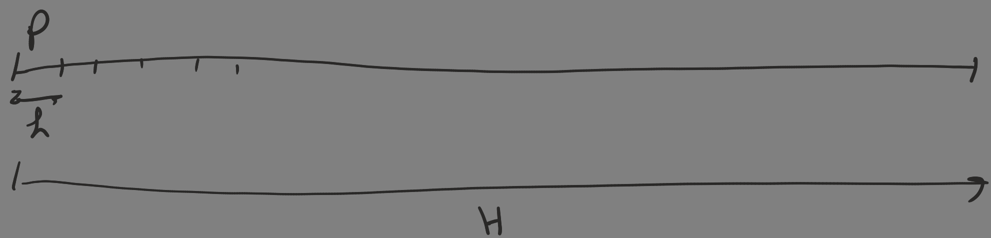
$$E(X) = E(X_1) + E(X_2) + \dots + E(X_n) = n\left(\frac{N_1}{N}\right)$$

احتمال قرمز بودن توپ که به طور طبیعی همواره برابر تعداد کل توپ‌های قرمز تقسیم بر کل تعداد توپ هاست

$$Var(X) = n\frac{N_1}{N}\,\frac{N_2}{N}\,\frac{N-n}{N-1}$$

توزیع پواسون (Poisson Distribution)

رویدادی که اتفاق افتادنش در طول زمان رخ می‌دهد. مثل شمارش تعداد مشتری‌های وارد بانک شده در یک ساعت. در صورتی که بدونیم ورود افراد مستقل هستند.



بازه‌ی زمانی اصلی H رو به قسمت‌های مساوی برابر h تقسیم می‌کنیم به طوری که در هر h فقط یک نفر وارد بانک بشه و ورود افراد در هر بازه‌ی h از هم مستقل باشند.

حالا واضحه که دقیقاً مسئله تبدیل شد به توزیع باینومیال

$$PMF_{(X=k)} = \binom{H/h}{k} p^k (1-p)^{H/h - k}$$

if $\lambda = \frac{pH}{h} \rightarrow PMF_{(X=k)} = \binom{H/h}{k}\left(\frac{\lambda h}{H}\right)^k \left(1-\frac{\lambda h}{H}\right)^{\frac{H}{h} - k}$

λ: نرخ متوسط ورود افراد به بانک

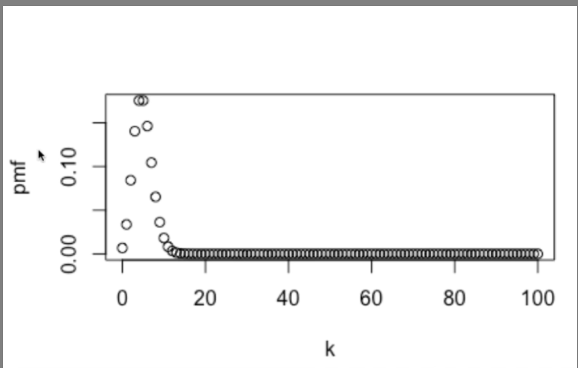
در واقع λ امید ریاضی ماست.

$$\text{if } n = \frac{H}{h}, \quad \lim_{n \to \infty} PMF = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \boxed{\frac{e^{-\lambda} \lambda^k}{k!}}$$

خب به این معناست که ما h های خودمون رو بسیار کوچک می‌گیریم. یعنی به بازه‌های زمانی خیلی خیلی کوچک آزمایش رو انجام می‌دیم.

در واقع این تابع به ما احتمال این رو میده که در بازه‌ی زمانی H با چه احتمالی $\underline{k}$ نفر وارد بانک شدن.

نمایش یک توزیع پواسون به صورت کلی به شکل روبه‌رو هست.



$$Var(X) = \lambda$$

متغیرهای تصادفی پیوسته

درست مثل گسسته هستن فقط به جای Σ از انتگرال استفاده می‌کنیم. و به جای PMF یک تابع چگالی احتمال داریم PDF که اینم همون رابطه‌ی PMF رو داره فقط به جای Σ از انتگرال استفاده شده.