Discrete Distributions

$$n = \binom{m}{2}$$

(x, y) (y, z)

$$p = \frac{1}{365}$$

$$\lambda = np = \binom{m}{2} \frac{1}{365}$$

$$P\{X=0\} = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda} < \frac{1}{2} \implies \boxed{m \geq 23}$$

1. Bernoulli
2. Binomial
3. Poisson

توزیع برنولی، دوجمله‌ای، پواسون

توزیع فوق هندسی (Hypergeometrical)

$\boxed{x \leq K}$ $\boxed{n - x \leq N - K}$ $\boxed{x \leq n}$

$$P_X(x) = \begin{cases} \dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} & \max\{0, n-N+K\} \leq x \leq \min\{n, K\} \\ 0 & \text{otherwise} \end{cases}$$

N: تعداد کارت ها

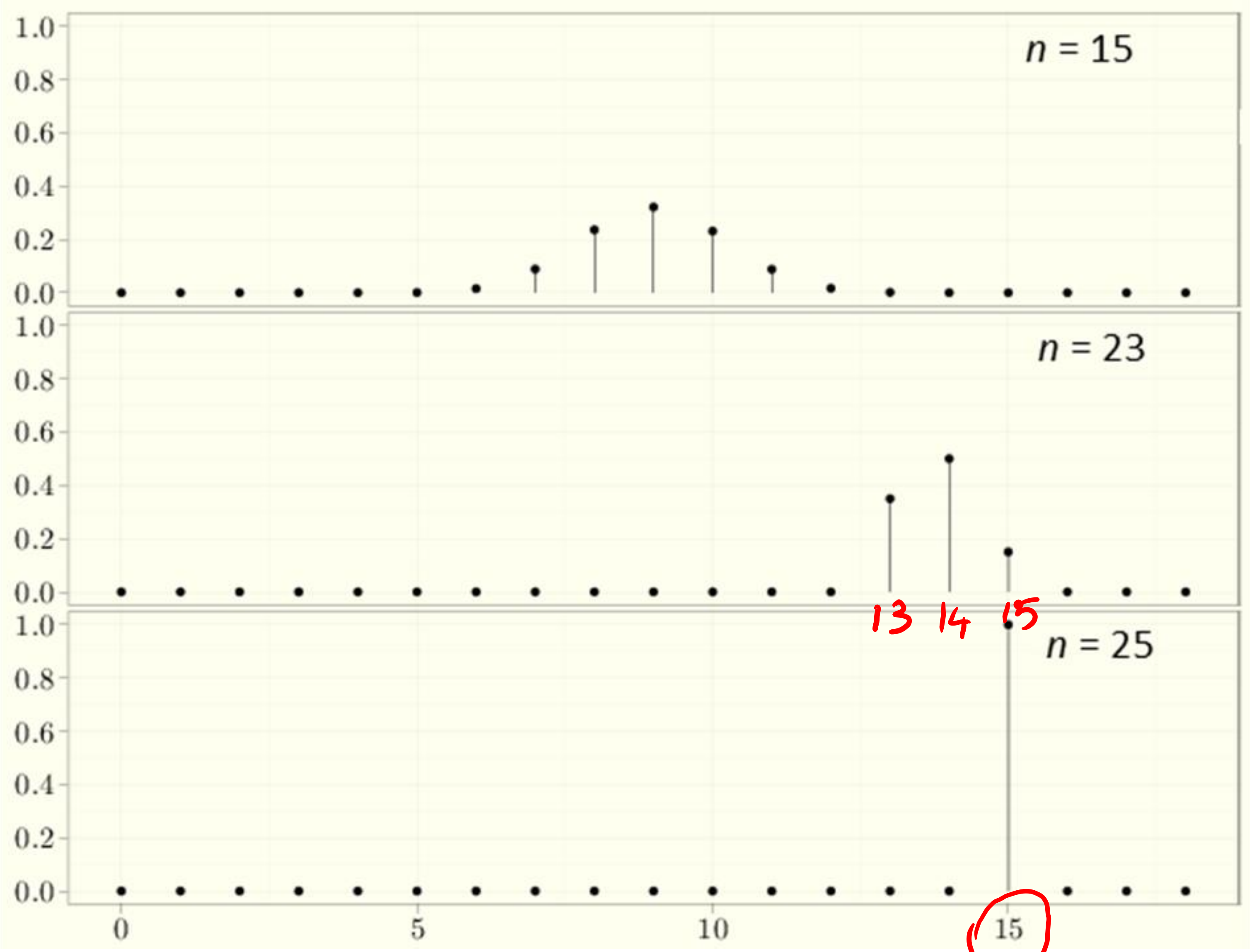
K: خراب ها

x: تعداد کارت های خراب انتخاب شده

n: تعداد کارت های انتخاب شده

توزیع فوق هندسی

$P_X(x)$



سالم 10
خراب 15

$N = 25$
$K = 15$

میانگین و واریانس توزیع فوق هندسی

$$P_X(x) = P\{X = x\} = \frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}} \qquad p = \frac{K}{N}$$

می‌توان نشان داد که:

$$E(X) = n\frac{K}{N} = \underline{np}$$

$$\mathrm{var}(X) = \underbrace{\frac{nK(N-K)}{N^2}}_{n\frac{K}{N}\left(1-\frac{K}{N}\right)} \times \underbrace{\frac{N-n}{N-1}}_{1}$$

$N \gg n$

npq

توزیع هندسی (Geometric Distribution)

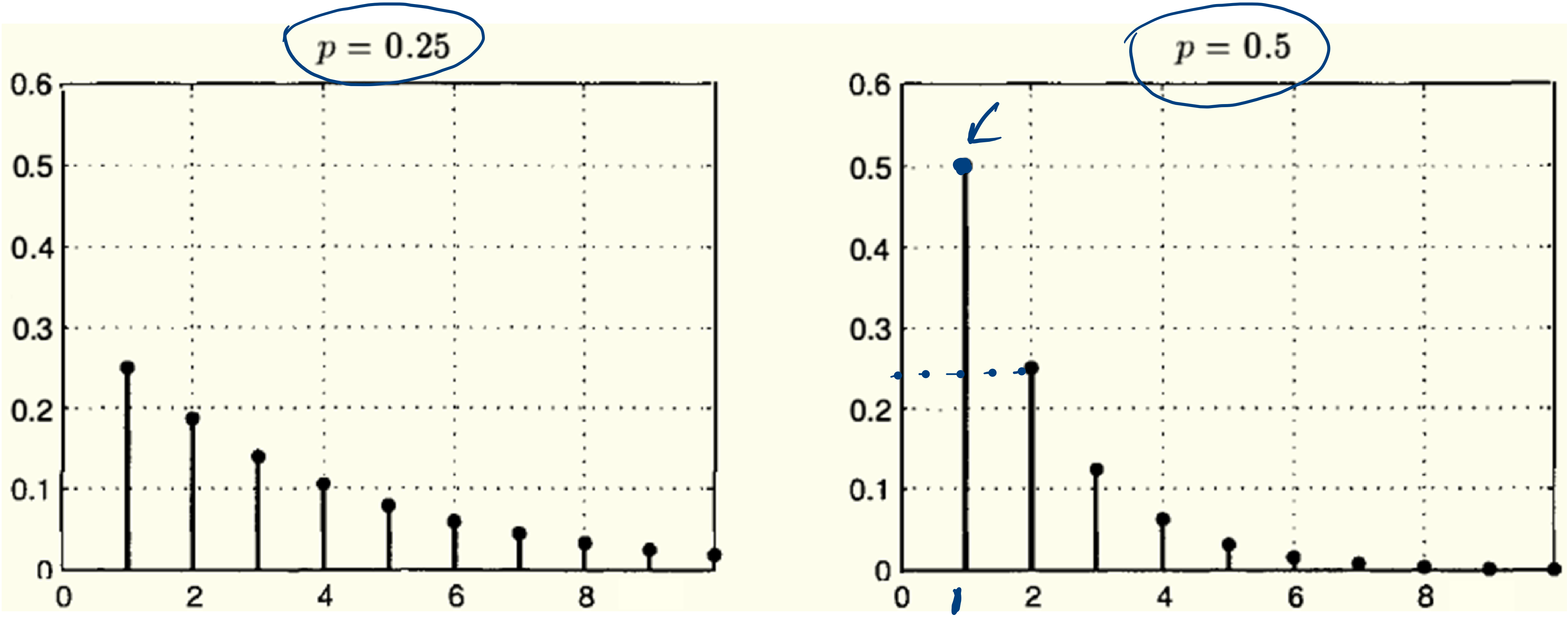
- توزیع متغیر تصادفی تعداد دفعات لازم برای تکرار یک آزمایش برنولی تا رسیدن به اولین موفقیت را توزیع هندسی می‌گوییم:

$$X \sim \text{Geo}(p) : \quad \underline{P\{X = k\}} = p\underbrace{(1-p)^{k-1}} \quad : \quad k = 1,2,\ldots$$

$$\sum_{k=1}^{+\infty} p(1-p)^{k-1} = p\sum_{k=1}^{+\infty}(1-p)^{k-1} = p\sum_{m=0}^{+\infty}(1-p)^{m}$$

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

توزیع هندسی



ویژگی‌های توزیع هندسی

- ویژگی‌های یک متغیر تصادفی با توزیع هندسی:
  - آزمایش‌های تصادفی مستقل از هم و در شرایط یکسان انجام می‌گیرند.
  - هر آزمایش دو خروجی ممکن دارد: شکست و موفقیت
  - احتمال موفقیت در همه آزمایش‌ها یکسان است: p
  - تعداد آزمایش‌های تصادفی ثابت نیست و تا زمانی که یک موفقیت مشاهده نشود ادامه می‌یابند.

- کاربردهای توزیع هندسی:
  - تعداد آزمایش‌های برنولی لازم برای این که یک فرایند گسسته تغییر حالت دهد.
  - تعداد تماس‌های لازم برای برقراری تماس موفق با یک خط تلفن که ۸۰٪ اوقات اشغال است
  - تعداد چاه‌هایی که در یک منطقه باید حفر شوند تا به یک چاه آب برسیم.
  - تعداد دفعاتی که یک تاس باید پرتاب شود تا عدد ۶ ظاهر شود.

میانگین توزیع هندسی

$$E[X] = \frac{1}{p}$$

$$f(x) = g(x)$$

$$\boxed{\sum_{k=0}^{+\infty} x^k = \frac{1}{1-x} \qquad |x|<1}$$

$$\boxed{\sum_{k=1}^{+\infty} k x^{k-1} = \frac{1}{(1-x)^2}} \longrightarrow$$

$$E[X] = \sum_{k=1}^{+\infty} k p (1-p)^{k-1} = p \sum_{k=1}^{+\infty} k (1-p)^{k-1} =$$

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

$$\sum_{k=1}^{+\infty} k x^{k-1} = \frac{1}{(1-x)^2} \longrightarrow \sum_{k=1}^{+\infty} k x^{k} = \frac{x}{(1-x)^2}$$

$$\Longrightarrow \sum_{k=1}^{+\infty} k^2 x^{k-1} = \frac{(1-x)^2 + 2(1-x)x}{(1-x)^4} = \frac{1+x}{(1-x)^3}$$

$$\operatorname{var}(X) = E[X^2] - E^2[X]$$

$$E[X^2] = \sum_{k=1}^{+\infty} k^2 p (1-p)^{k-1} = p \frac{1+1-p}{(1-(1-p))^3} = \frac{2-p}{p^2}$$

$$\operatorname{var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}$$

واریانس توزیع هندسی

بی‌حافظگی توزیع هندسی

- توزیع متغیر تصادفی X را بی‌حافظه (memoryless) گویند، اگر برای هر t و s، داشته باشیم:

$$P\{X \overset{>}{\geq} s+t \mid X \overset{>}{\geq} t\} = P\{X \overset{>}{\geq} s\}$$

- توزیع هندسی تنها توزیع احتمال گسسته دارای خاصیت بی‌حافظگی است.

$$P(X>t) = \sum_{k=t+1}^{+\infty} p(1-p)^{k-1} = (1-p)^{t} = q^{t}$$

$$P(X>t+s \mid X>t) = \frac{P(\overbrace{\{X>t+s\}} \cap \overbrace{\{X>t\}})}{P(X>t)} = \frac{q^{t+s}}{q^{t}}$$

$$= q^{s} = P(X>s)$$

اثبات بی‌حافظگی توزیع هندسی

$$P(\{X \geq t\}) = \sum_{k=t}^{+\infty} p(1-p)^{k-1} = p \sum_{m=0}^{+\infty} (1-p)^{m+t-1}$$

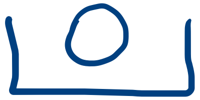
$$= p(1-p)^{t-1} \underbrace{\sum_{m=0}^{+\infty} (1-p)^{m}} = p(1-p)^{t-1} \frac{1}{p}$$

$$P(\{X \geq t+s \mid X \geq t\}) = \frac{P(\{X \geq t+s\} \cap \{X \geq t\})}{P(\{X \geq t\})}$$

$$= \frac{P(\{X \geq t+s\})}{P(\{X \geq t\})} = \frac{q^{t+s-1}}{q^{t-1}} = q^{s} = P\{X \geq s+1\}$$

مثال: مسأله جمع‌آوری کوپن (Coupon Collector)

یک جعبه که داخل آن n کوپن مختلف قرار دارد، در اختیار داریم. در هر بار آزمایش یک کوپن از جعبه خارج کرده و بعد از مشاهده به درون جعبه باز می‌گردانیم. به طور متوسط چند کوپن باید از جعبه خارج کنیم تا هر کوپن حداقل یک بار مشاهده شده باشد؟



$n=1$ $\qquad E[X]=1$

$n=2$ $\qquad E[X] = 1+2 = 1 + \frac{1}{\frac{1}{2}}$

$n=3$ $\qquad E[X] = 1 + \frac{1}{\frac{2}{3}} + \frac{1}{\frac{1}{3}} = 1 + \frac{3}{2} + 3$

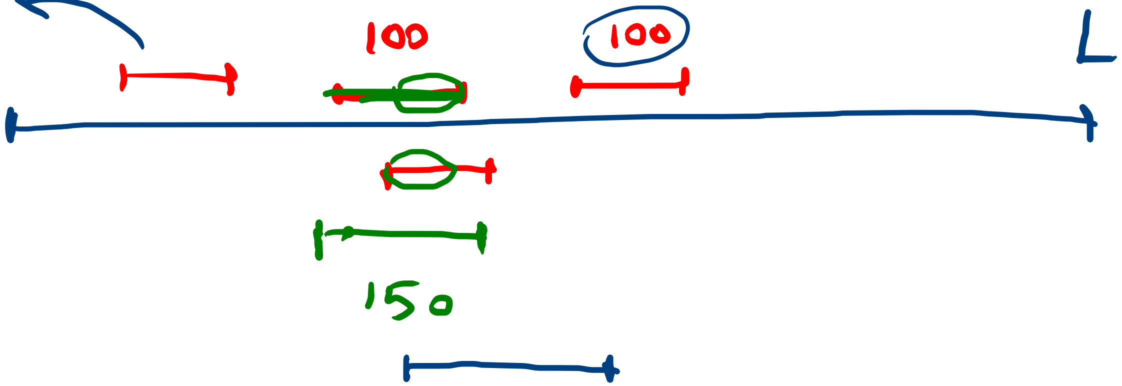
$$E[X] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} = n\sum_{i=1}^{n} \frac{1}{i} \simeq \underline{n \ln n}$$

$\{A, C, G, T\}$

AATTCAT...

3×10^9

Read



Illumina

$\frac{L}{K} \log L$

$\left(\frac{1}{4}\right)^{20}$

توزیع دوجمله‌ای منفی (Negative Binomial)
توزیع پاسکال (Pascal Distribution)

سکه‌ای را آنقدر پرتاب می‌کنیم که r بار شیر بیاید. اگر تعداد کل پرتاب‌ها را X بنامیم، احتمال $X = n$ چقدر است؟

$$P_X(x) = p\binom{x-1}{r-1} p^{r-1} \quad q^{x-1-(r-1)} \qquad x \geq r$$

$$= \binom{x-1}{r-1} p^{r} \quad q^{x-r} \qquad x \geq r$$

توزیع دوجمله‌ای منفی

این توزیع را دوجمله‌ای منفی می‌نامند زیرا بر خلاف توزیع دوجمله‌ای در اینجا تعداد موفقیت‌ها ثابت است و تعداد کل آزمایش‌های انجام شده متغیر.

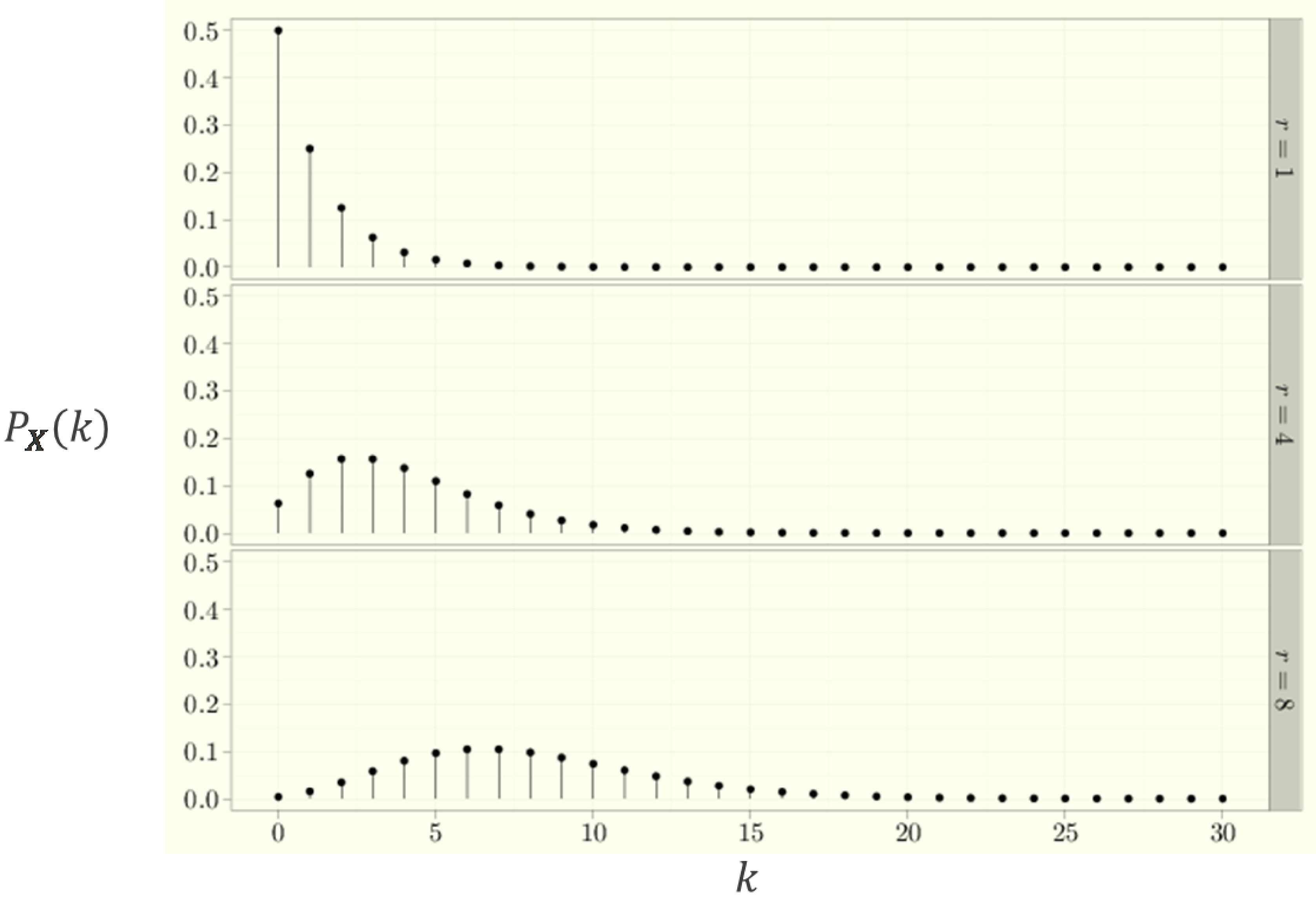
متغیر با توزیع دوجمله‌ای منفی را می‌توان به صورت مجموع r متغیر تصادفی هندسی نوشت:

NB(r, p)

$$X = X_1 + X_2 + \cdots + X_r \quad : \quad X_i \sim \mathrm{Geo}(p)$$

3 1

توزیع دوجمله‌ای منفی



$r = 1$

$\boxed{p = 0.5}$

$r = 4$

$r = 8$

ویژگی‌های توزیع دوجمله‌ای منفی

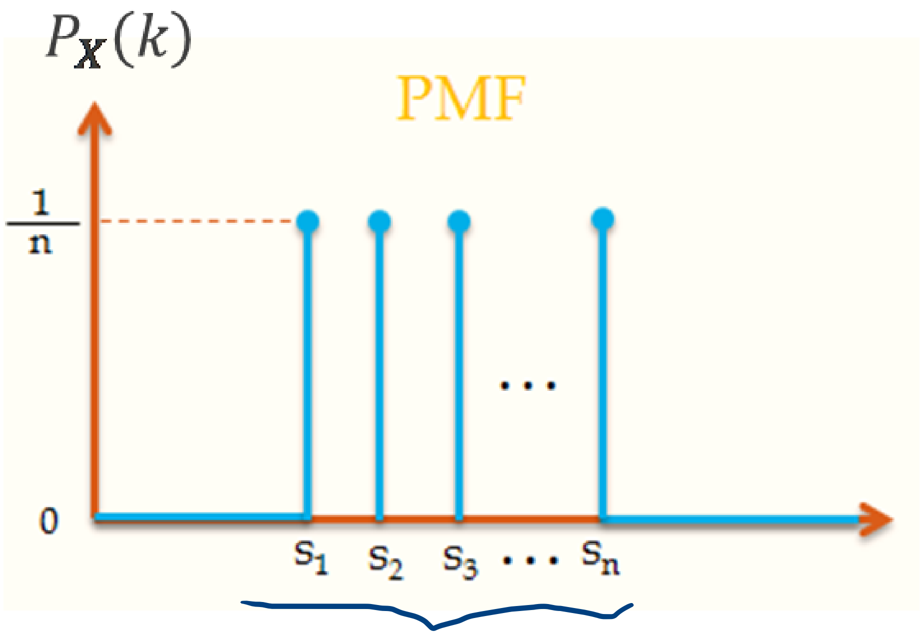
- ویژگی‌های یک متغیر تصادفی با توزیع دوجمله‌ای منفی:
  - آزمایش‌های تصادفی مستقل از هم و در شرایط یکسان انجام می‌گیرند.
  - هر آزمایش دو خروجی ممکن دارد: شکست و موفقیت
  - احتمال موفقیت در همه آزمایش‌ها یکسان است: p
  - تعداد آزمایش‌های تصادفی ثابت نیست و تا زمانی که r موفقیت مشاهده نشود ادامه می‌یابند.

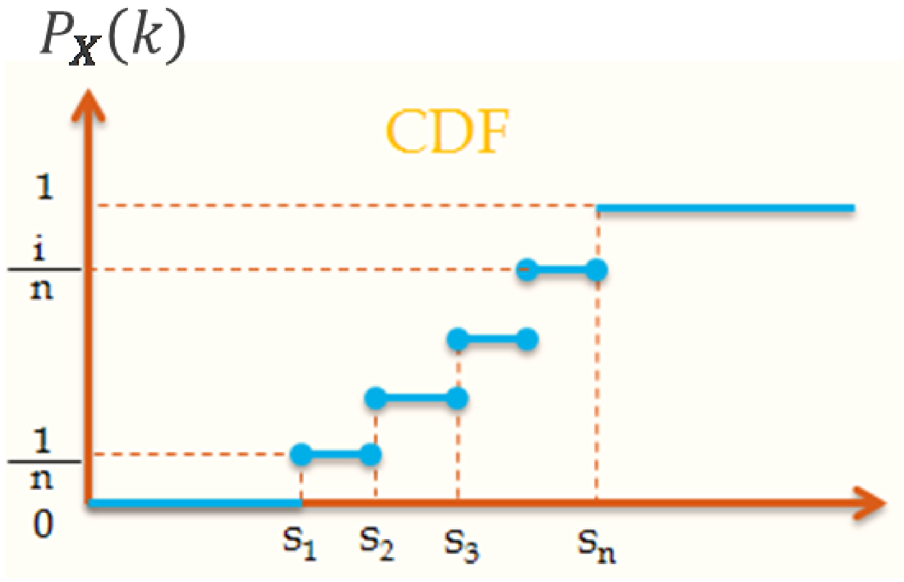
میانگین و واریانس توزیع دوجمله‌ای منفی

$$E[X] = E[\underbrace{X_1 + X_2 + \cdots + X_r}] = r\frac{1}{p}$$

$$var(x+y) \neq var(x) + var(y)$$

توزیع یکنواخت گسسته (Uniform Distribution)





$$P\{X = s_i\} = \frac{1}{n} \qquad i = 1, 2, \ldots, n$$