SUITE OF OUGHT-TO-KNOW PROBABILITY DISTRIBUTIONS FOR DISCRETE RANDOM VARIABLES

T.O.C.

Note the convention: Distribution Name (parameters)
↑values relevant to the distribution

(i) Bernoulli(p):

If X encodes the outcome of a Bernoulli trial with success probability p such that

$$X = \begin{cases} 1 & \text{if success} \\ 0 & \text{if failure,} \end{cases}$$

then $X \sim \text{Bernoulli}(p)$.

We often write parameters after a semi colon in the pmf/cdf

$$p_X(x;p) = \begin{cases} p^x(1-p)^{1-x} & \text{if } x \in \{0,1\} \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x;p) = \begin{cases} \sum_{t=0}^{\lfloor x \rfloor} p^t(1-p)^{1-t} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases},$$

where $\lfloor x \rfloor$ is the greatest integer not exceeding x.

$$\mathbb{E}X = p$$

$$\text{Var } X = p(1-p)$$

(ii) <u>Binomial(n,p)</u>: If $X = \#$ successes in n independent Bernoulli trials, each with success probability p, then $X \sim \text{Binomial}(n,p)$

$$p_X(x;n,p) = \begin{cases} \overbrace{\binom{n}{x}}^{\#\text{ of sequences of length } n \text{ with } x \text{ successes}} \underbrace{p^x(1-p)^{n-x}}_{\text{probability we assign to each sequence with } x \text{ successes and } n-x \text{ failures}} & \text{if } x \in \{0,1,\dots,n\} \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x;n,p) = \begin{cases} \sum_{t=0}^{\lfloor x \rfloor} \binom{n}{t} p^t (1-p)^{n-t} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

$$\mathbb{E}X = np$$

$$\operatorname{Var} X = np(1-p)$$

<u>Derivations</u>:

Show $\mathbb{E}X = np$

$$\mathbb{E}X = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{((n-1)-(x-1))!\,(x-1)!} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

set $m = n-1$
$y = x-1$

then $1 \leq x \leq n \Rightarrow 0 \leq y \leq m$

$$= np \underbrace{\sum_{y=0}^{m} \binom{m}{y} p^y (1-p)^{m-y}}_{=1}$$

$$= np$$

$=1$ (summation over Binomial pmf)

Show $\operatorname{Var} X = np(1-p)$

Use $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X^2 - (np)^2$, where

$$\mathbb{E}\, X^2 = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{x \quad (n-1)}{\big((n-1)-(x-1)\big)!\;(x-1)!}\, p^{x-1}\,(1-p)^{(n-1)-(x-1)}$$

set $m = n-1$
$y = x-1$

so $1 \le x \le n \Rightarrow 0 \le y \le m$

$$= np \sum_{y=0}^{m} (y+1) \binom{m}{y} p^y (1-p)^{m-y}$$

$$= np \left[\underbrace{\sum_{y=0}^{m} y \binom{m}{y} p^y (1-p)^{m-y}}_{=mp} + \underbrace{\sum_{y=0}^{m} 1 \binom{m}{y} p^y (1-p)^{m-y}}_{=1} \right]$$

$$= np\big((n-1)p + 1\big)$$

$$= np\,(np - p + 1)$$

$$= (np)^2 - np^2 + np$$

Now

$$\mathrm{Var}\, X = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2$$

$$= (np)^2 - np^2 + np - (np)^2$$

$$= np - np^2$$

$$= np(1-p)$$

(iii) __Geometric(p)__: If independent Bernoulli trials are conducted until the first success occurs, and X is the number of the trial on which the first success occurs, then $X \sim \text{Geometric}(p)$.

Sample space: $S = \{ \underbrace{S}_{p},\ \underbrace{FS}_{(1-p)p},\ \underbrace{FF}_{(1-p)^2}\underbrace{S}_{p},\ \underbrace{FFF}_{(1-p)^3}\underbrace{S}_{p},\ \underbrace{FFFF}_{(1-p)^4}\underbrace{S}_{p},\ \dots \}$

$$p_X(x;p) = \begin{cases} (1-p)^{x-1} p & \text{for } x = 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x;p) = \begin{cases} 1-(1-p)^{\lfloor x \rfloor} & \text{for } x \geq 1 \\ 0 & \text{for } x < 1 \end{cases}$$

$$\mathbb{E}X = 1/p$$

$$\operatorname{Var} X = (1-p)/p^2$$

Derivations

Show $F_X(x;p) = \begin{cases} 1-(1-p)^{\lfloor x \rfloor} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases}$

For $x \geq 0$,
$$\begin{aligned} F_X(x;p) &= P_X(X \leq x) \\ &= P_X(X \leq \lfloor x \rfloor) \\ &= \sum_{t=1}^{\lfloor x \rfloor} (1-p)^{t-1} p \\ &= p \sum_{t=1}^{\lfloor x \rfloor} (1-p)^{t-1} \\ &= p \sum_{t=0}^{\lfloor x \rfloor - 1} (1-p)^{t} \\ &= p \, \frac{1-(1-p)^{\lfloor x \rfloor}}{1-(1-p)} \\ &= 1-(1-p)^{\lfloor x \rfloor} \end{aligned}$$

$$p\left[(1-p)^0 + (1-p)^1 + \dots + (1-p)^{\lfloor x \rfloor - 1}\right]$$

partial geometric series

Result:

For $t \neq 1$,
$$\sum_{i=0}^{n} t^i = \frac{1-t^{n+1}}{1-t} \quad \text{for } n = 1, 2, \dots$$

Proof:

Let $S_n = \sum_{i=0}^{n} t^i = 1 + t + \dots + t^n$

Then
$$\begin{aligned} S_n - t S_n &= 1 + t + \dots + t^n \\ &\quad - (t + t^2 + \dots + t^{n+1}) \\ &= 1 - t^{n+1} \end{aligned}$$

So $S_n = \dfrac{1-t^{n+1}}{1-t}$

Show $\mathbb{E}X = \sum_{x=1}^{\infty} x(1-p)^{x-1}p$

$$= p\sum_{x=1}^{\infty} x(1-p)^{x-1}$$

$$= p\sum_{x=1}^{\infty} \frac{d}{dp}\left[-(1-p)^{x}\right]$$

We cannot always exchange differentiation and infinite summation, but here we can because the sum converges for all $p \in (0,1)$.

$$= p\frac{d}{dp}\left[-\sum_{x=1}^{\infty}(1-p)^{x}\right]$$

$$= p\frac{d}{dp}\left[-\frac{1-p}{p}\right]$$

$$= p\frac{d}{dp}\left[-\frac{1}{p}+1\right]$$

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

Show $\mathrm{Var}\,X = (1-p)/p^2$

Use $\mathrm{Var}\,X = \mathbb{E}X^2 - (\mathbb{E}X)^2$, where

$$\mathbb{E}X^2 = \sum_{x=1}^{\infty} x^2(1-p)^{x-1}p$$

$$= p\sum_{x=1}^{\infty}\frac{d}{dp}\left[-x(1-p)^{x}\right]$$

$$= p\frac{d}{dp}\left[-\sum_{x=1}^{\infty}x(1-p)^{x}\right]$$

$$= p\frac{d}{dp}\left[-\frac{1-p}{p}\underbrace{\sum_{x=1}^{\infty}x(1-p)^{x-1}p}_{=1/p \text{ from above}}\right]$$

$$= p\frac{d}{dp}\left[-\frac{1-p}{p^2}\right]$$

$$= p\left[\frac{2}{p^3}-\frac{1}{p^2}\right]$$

$$= p\left[\frac{2-p}{p^3}\right]$$

$$= \frac{2-p}{p^2}$$

So $\quad \text{Var } X = \dfrac{2-p}{p^2} - \dfrac{1}{p^2} = \dfrac{1-p}{p^2}$

(iv) $\underline{\underline{\text{Negative Binomial }(p, r)}}$: Let X be the trial on which the r^{th} success occurs in repeated independent Bernoulli trials with success probability p. Then $X \sim \text{Negative Binomial }(p, r)$.

Sample space: $\quad S = \left\{ \begin{array}{l} \text{Sequences of any length} \\ \text{where last outcome a success} \\ \text{and with } r-1 \text{ previous successes} \end{array} \right\}$

E.g.

$$\underbrace{\overbrace{\underset{1}{S}\quad \underset{2}{F}\quad \underset{3}{S}\cdots \underset{x-3}{F}\quad \underset{x-2}{F}\quad \underset{x-1}{F}}^{r-1 \text{ successes in these } x-1 \text{ trials}}}_{(x-1)-(r-1) = x-r \text{ failures in these } x-1 \text{ trials}} \quad \underset{x^{th}\text{ trial}}{\overset{r^{th}\text{ success}\ \downarrow}{S}}$$

$$p_X(x; p, r) = \begin{cases} \dbinom{x-1}{r-1}(1-p)^{x-r}\, p^r & \text{for } x = r, r+1, r+2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x; p, r) = \begin{cases} \displaystyle\sum_{t=r}^{\lfloor x \rfloor} \binom{t-1}{r-1}(1-p)^{t-r}\, p^r & \text{for } x \geq r \\ 0 & \text{for } x < r \end{cases}$$

$$\mathbb{E} X = \frac{r}{p}$$

$$\operatorname{Var} X = \frac{r(1-p)}{p^2}$$

Derivations:

Show $\mathbb{E} X = r/p$

$$\mathbb{E} X = \sum_{x=r}^{\infty} x \binom{x-1}{r-1} (1-p)^{x-r} p^r$$

$$= r \sum_{x=r}^{\infty} \frac{x(x-1)!}{((x-1)-(r-1))!\, r(r-1)!} (1-p)^{x-r} p^r$$

let $y = x+1$, $m = r+1$

$$= \frac{r}{p} \sum_{x=r}^{\infty} \frac{((x+1)-1)!}{(x-r)!\,((r+1)-1)!} (1-p)^{(x+1)-(r+1)} p^{r+1}$$

$r \leq x < \infty$ $\Rightarrow$ $m \leq y < \infty$

$$= \frac{r}{p} \sum_{y=m}^{\infty} \frac{(y-1)!}{((y-1)-(m-1))!\,(m-1)!} (1-p)^{y-m} p^m$$

$$= \frac{r}{p} \underbrace{\sum_{y=m}^{\infty} \binom{y-1}{m-1} (1-p)^{y-m} p^m}_{=1}$$

$$= \frac{r}{p}$$

Show $\operatorname{Var} X = r(1-p)/p^2$

$$\mathbb{E} X^2 = \sum_{x=r}^{\infty} x^2 \binom{x-1}{r-1} (1-p)^{x-r} p^r$$

$$= \frac{r}{p} \sum_{x=r}^{\infty} x \frac{((x+1)-1)!}{(x-r)!\,((r+1)-1)!} (1-p)^{(x+1)-(r+1)} p^{r+1}$$

let $\begin{matrix} y = x+1 \\ m = r+1 \end{matrix}$

$r \le x < \infty \Rightarrow m \le y < \infty$

$$= \frac{r}{p} \underbrace{\sum_{y=m}^{\infty} (y-1) \binom{y-1}{m-1} (1-p)^{y-m} p^{m}}_{= \mathbb{E}[Y-1]\ ,\ Y \sim \text{Negative Binomial}(p, m)}$$

$$= \frac{r}{p}\left[\frac{m}{p} - 1\right]$$

$$= \frac{r}{p}\left[\frac{r+1}{p} - 1\right]$$

So $\quad \operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2$

$$= \frac{r}{p}\left[\frac{r+1}{p} - 1\right] - \left(\frac{r}{p}\right)^2$$

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

(v) **Poisson(λ)**: If $X = \#$ occurrences of an event per unit time/space, where on average $\lambda \geq 0$ events occur per unit time/space, then we often assume $X \sim \text{Poisson}(\lambda)$.

$$p_X(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & x = 0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x;\lambda) = \begin{cases} \displaystyle\sum_{t=0}^{\lfloor x \rfloor} \frac{e^{-\lambda}\lambda^{t}}{t!} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

$\mathbb{E}X = \lambda$

$\operatorname{Var} X = \lambda$

Quite special that $\mathbb{E}X = \operatorname{Var} X$

Derivations:

Show $\mathbb{E}X = \lambda$

$$\mathbb{E}X = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}$$

set $y = x-1$

$1 \le x < \infty \Rightarrow 0 \le y < \infty$

$$= \lambda \underbrace{\sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!}}_{=1,\ \text{summation over pmf}}$$

$$= \lambda$$

Show $\text{Var}\, X = \lambda$

$$\mathbb{E}X^2 = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \lambda \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}$$

let $y = x-1$

so $1 \le x < \infty \Rightarrow 0 \le y < \infty$

$$= \lambda \underbrace{\sum_{y=0}^{\infty} (y+1) \frac{e^{-\lambda}\lambda^{y}}{y!}}_{\mathbb{E}[Y+1],\ Y \sim \text{Poisson}(\lambda)}$$

$$= \lambda(\lambda+1)$$

So $\text{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$

(vi) Hypergeometric(N, M, K):

Draw $K \geq 0$ marbles from a bag of $N \geq 0$ marbles, where $M \geq 0$ of the N marbles are red. If X is the number of red marbles drawn, $X \sim \text{Hypergeometric}(N, M, K)$.

$$p_X(x; N, M, K) = \begin{cases} \dfrac{\overbrace{\binom{M}{x}}^{\text{\# ways to draw } x \text{ of the } M \text{ red marbles}} \overbrace{\binom{N-M}{K-x}}^{\text{\# ways to draw } K-x \text{ of the } N-M \text{ non-red marbles}}}{\underbrace{\binom{N}{K}}_{\text{\# ways to draw } K \text{ marbles from bag of } N}} & \text{for } x = \underbrace{\max\big(0, K-(N-M)\big)}_{\substack{= 0 \text{ unless there} \\ \text{are fewer than } K \\ \text{non-red marbles}}}, \ldots, \underbrace{\min(K, M)}_{\substack{= K \text{ unless} \\ \text{there are} \\ \text{fewer than } K \\ \text{red marbles}}} \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x; N, M, K) = \begin{cases} \displaystyle\sum_{t=0}^{\lfloor x \rfloor} p_X(t; N, M, K) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

$$\mathbb{E} X = \frac{KM}{N}$$

$$\operatorname{Var} X = \frac{KM}{N}\left[\frac{(N-K)(N-M)}{N(N-1)}\right]$$

Derivations:

Show $\mathbb{E}X = KM/N$

$$\mathbb{E}X = \sum_{x=0}^{K} x \, \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$$

$$= \sum_{x=1}^{K} \frac{M}{N/K} \, \frac{\dfrac{(M-1)!}{\big((M-1)-(x-1)\big)!\,(x-1)!} \; \dfrac{\big((N-1)-(M-1)\big)!}{\big((N-1)-(M-1)-[(K-1)-(x-1)]\big)!\,\big((K-1)-(x-1)\big)!}}{\dfrac{(N-1)!}{\big((N-1)-(K-1)\big)!\,(K-1)!}}$$

set $y = x-1$

$L = K-1$

$Q = M-1$

$R = N-1$

so $0 \le y \le L$

$$= \frac{KM}{N} \underbrace{\sum_{y=1}^{L} \frac{\binom{Q}{y}\binom{R-Q}{L-y}}{\binom{R}{L}}}_{=1\text{, sum over pmf of Hypergeometric}(R,Q,L)}$$

$$= \frac{KM}{N}$$

Show $\operatorname{Var} X = \frac{KM}{N}\left[\frac{(N-K)(N-M)}{N(N-1)}\right]$

$$\mathbb{E}X^2 = \sum_{x=0}^{K} x^2 \binom{M}{x}\binom{N-M}{K-x} \Big/ \binom{N}{K}$$

As before, set

$y = x-1$

$L = K-1$

$Q = M-1$

$R = N-1$

so $0 \le y \le L$

$$= \frac{KM}{N} \underbrace{\sum_{y=1}^{L} (y+1) \frac{\binom{Q}{y}\binom{R-Q}{L-y}}{\binom{R}{L}}}_{\mathbb{E}[Y+1],\ Y \sim \text{Hypergeometric}(R,Q,L)}$$

$$= \frac{KM}{N}\left[\frac{LQ}{R} + 1\right]$$

$$= \frac{KM}{N}\left[\frac{(K-1)(M-1)}{N-1} + 1\right]$$

So $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2$

$$= \frac{KM}{N}\left[\frac{(K-1)(M-1)}{N-1} + 1\right] - \left(\frac{KM}{N}\right)^2$$

$$= \frac{KM}{N}\left[\frac{N(K-1)(M-1) + N(N-1) - (N-1)KM}{N(N-1)}\right]$$

$$= \frac{KM}{N}\left[\frac{NKM - NK - NM + N + N^2 - N - NKM + KM}{N(N-1)}\right]$$

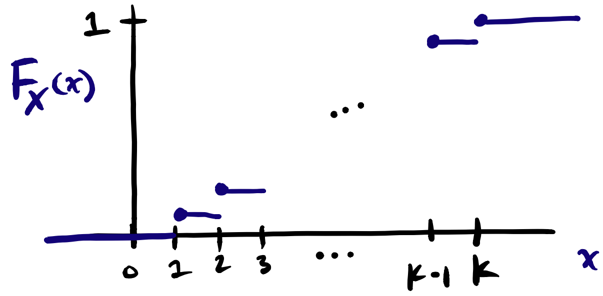
$$= \frac{KM}{N}\left[\frac{(N-K)(N-M)}{N(N-1)}\right]$$

(vii) Discrete Uniform (K)

If X takes the values $1, \dots, K$ each with probability $1/K$, then $X \sim$ Discrete Uniform (K).

$$p_X(x;K) = \begin{cases} 1/K & \text{for} \quad x = 1, 2, \dots, K \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x;K) = \begin{cases} \frac{\lfloor x \rfloor}{K} & \text{for} \quad x \geq 0 \\ 0 & \text{for} \quad x < 0 \end{cases}$$



$$\mathbb{E}X = \frac{K+1}{2}$$

$$\text{Var } X = \frac{(K+1)(K-1)}{12}$$

Derivations:

Show $\mathbb{E}X = (K+1)/2$

$$\mathbb{E}X = \sum_{x=1}^{K} x \cdot \frac{1}{K}$$

$$= \frac{1}{K} \sum_{x=1}^{K} x$$

Sum of first n integers
$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

$$= \frac{1}{K} \frac{K(K+1)}{2}$$

$$= \frac{K+1}{2}$$

Show $\text{Var } X = (K+1)(K-1)/12$

$$\mathbb{E} X^2 = \sum_{x=1}^{K} x^2 \cdot \frac{1}{K}$$

$$= \frac{1}{K} \sum_{x=1}^{K} x^2$$

$$= \frac{1}{k} \frac{k(k+1)(2k+1)}{6}$$

Sum of first n squares

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

$$= \frac{(k+1)(2k+1)}{6}$$

So $\quad \operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2$

$$= \frac{(k+1)(2k+1)}{6} - \left(\frac{k+1}{2}\right)^2$$

$$= (k+1)\left[\frac{2k+1}{6} - \frac{k+1}{4}\right]$$

$$= (k+1)\left[\frac{4k+2-3k-3}{12}\right]$$

$$= \frac{(k+1)(k-1)}{12}$$

(viii) Empirical Distribution $(x_1, \dots, x_n)$

If X takes each of the values $x_1, \dots, x_n$ with probability $1/n$, then $X \sim$ Empirical Distribution $(x_1, \dots, x_n)$

$$p_X(x; x_1, \dots, x_n) = \begin{cases} 1/n & \text{for } x \in \{x_1, \dots, x_n\} \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x; x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \leq x),$$

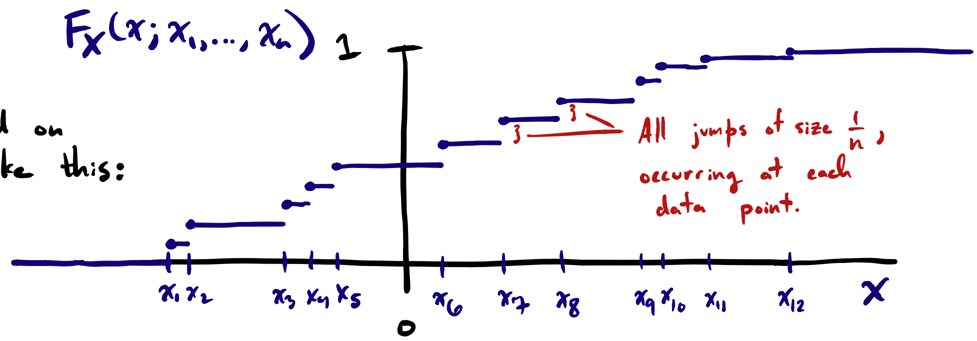
where $\underbrace{\mathbb{1}(\text{statement})}_{\text{"indicator function"}} = \begin{cases} 1 & \text{if statement true} \\ 0 & \text{if statement false} \end{cases}$

$$\mathbb{E}X = \frac{1}{n} \sum_{i=1}^{n} x_i =: \bar{x}$$

Average of $x_1, \dots, x_n$

$$\operatorname{Var} X = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Empirical distribution based on "data" $x_1, \dots, x_{12}$ looks like this:



Derivations:

Show $F_X(x; x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \le x)$

$$F_X(x; x_1, \dots, x_n) = \sum_{\{t \in \mathcal{I} : t \le x\}} p_X(t)$$

$$= \sum_{\{t \in \{x_1, \dots, x_n\} : t \le x\}} (1/n)$$

$$= \frac{1}{n} \left| \{ t \in \{x_1, \dots, x_n\} : t \le x \} \right|$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \le x)$$

Show $\mathbb{E} X = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$\mathbb{E} X = \sum_{x \in \mathcal{I}} x \, p_X(x)$$

$$= \sum_{x \in \{x_1, \dots, x_n\}} x \, p_X(x)$$

$$= \sum_{i=1}^{n} x_i \left(\frac{1}{n}\right)$$

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

Show $\operatorname{Var} X = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$$\operatorname{Var} X = \sum_{x \in \mathcal{I}} (x - \bar{x})^2 \, p_X(x)$$

$$= \sum_{x \in \{x_1, \dots, x_n\}} (x - \bar{x})^2 \, p_X(x)$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 \, (1/n)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$\boxed{\text{EXAMPLES}}$ (R code in green.)

You are selling goat's milk soap.

1. For an attempt to sell you have probability $1/10$ of success and attempts are independent.

(a) You make 1 attempt to sell. Let $X = \begin{cases} 1 & \text{if sale} \\ 0 & \text{if not} \end{cases}$.

(i) $\mathbb{E} X = .1$

(ii) $P_X(X=0) = .9$

(b) You make 30 attempts to sell. Let $X = \#$ sales.

(i) $\mathbb{E} X = 3$

(ii) $P_X(X=5) = \binom{30}{5} \left(\frac{1}{10}\right)^5 \left(1 - \frac{1}{10}\right)^{30-5}$

$= $ dbinom(x = 5, size = 30,, prob = .1) = .102

(iii) $P_X(X \geq 1) = 1 - \binom{30}{0} \left(\frac{1}{10}\right)^0 \left(1 - \frac{1}{10}\right)^{30}$

$= 1 - $ dbinom(x = 0, size = 30, prob = .1) = .958

(iv) $P_X(X \leq 5) = \sum_{x=0}^{5} \binom{30}{x} \left(\frac{1}{10}\right)^x \left(1 - \frac{1}{10}\right)^{30-x}$

$= $ pbinom(q = 5, size = 30, prob = .1) = .927

(v) $P_X(X \geq 7) = 1 - \sum_{x=0}^{6} \binom{30}{x} \left(\frac{1}{10}\right)^x \left(1 - \frac{1}{10}\right)^{30-x}$

$= 1 - $ pbinom(q = 6, size = 30, prob = .1) = .026

(c) Today you will make attempts until succeeding. Let $X = \#$ attempts.

(i) $\mathbb{E}X = 1/(1/10) = 10$

(ii) $P_X(X \geq 5) = 1 - P_X(X \leq 4) = 1 - [1 - (1-.1)^4] = .6561$

(iii) $P_X(X > 1) = 1 - P_X(X=1) = 1 - .1 = .9$

(d) Today you will make attempts until you make 5 sales. Let $X = \#$ attempts

(i) $\mathbb{E}X = 5/(1/10) = 50$

(ii) $P_X(X = 40) = \binom{40-1}{5-1}\left(1-\frac{1}{10}\right)^{40-5}\left(\frac{1}{10}\right)^5$

$= \text{choose}(40-1, 5-1) * (1-.1)\text{\^{}}(40-5) * .1\text{\^{}}5$

$= \underbrace{\text{dnbinom}}_{\text{R uses a version of negative binomial in which } X = \#\text{ failures before } r^{th}\text{ success.}}(x = 40-5,\ \text{size}=5,\ \text{prob}=.1) = .0206$

(iii) $P_X(X \leq 30) = \sum_{x=5}^{30}\binom{x-1}{5-1}\left(1-\frac{1}{10}\right)^{x-5}\left(\frac{1}{10}\right)^5$

$= \text{pnbinom}(q = 30-5,\ \text{size}=5,\ \text{prob}=.1) = .1755$

(iv) $P_X(X \geq 50) = 1 - \sum_{x=5}^{49}\binom{x-1}{5-1}\left(1-\frac{1}{10}\right)^{x-5}\left(\frac{1}{10}\right)^5$

$= 1 - \text{pnbinom}(q = 49-5,\ \text{size}=5,\ \text{prob}=.1) = .4497$

2. Now you sell goat's milk soap online. Let $X = \#$ sales tomorrow and assume $X \sim \text{Poisson}(5)$.

(i) $\mathbb{E}X = 5$

(ii) $P_X(X > 0) = 1 - P_X(X=0) = 1 - e^{-5} = .993$

(iii) $P_X(X = 4) = e^{-5}5^4/4!$

$= \text{dpois}(x=4,\ \text{lambda}=5) = .176$

(iv) $P_X(X \leq 6) = \sum_{x=0}^{6} e^{-5}5^x/x!$

$= \text{ppois}(q=6,\ \text{lambda}=5) = .762$

(v) $P_X(X > 10) = 1 - \sum_{x=0}^{10} e^{-5}5^x/x!$

$= 1 - \text{ppois}(q=10,\ \text{lambda}=5) = .014$

3. Suppose 10 of 30 residents of a neighborhood will buy if solicited. You will solicit 5 at random. Let X = # sales.

(i) $EX = 5\left(\frac{10}{30}\right) = 5/3$

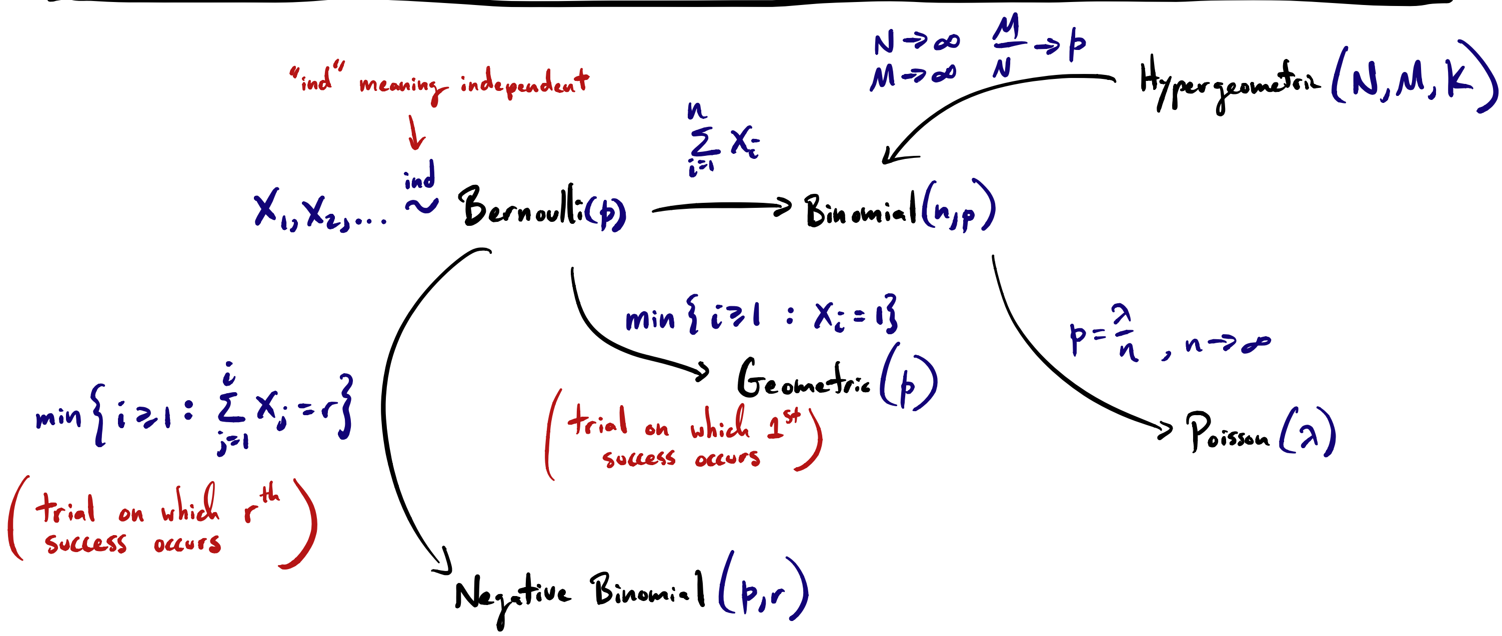
(ii) $P_X(X=2) = \binom{10}{2}\binom{30-10}{5-2} \Big/ \binom{30}{5}$

$= \text{dhyper}(x=2,\ m=10,\ \underbrace{n=30-10}_{\text{In R, } n = \text{ \# non-red marbles}},\ k=5) = .360$

(iii) $P_X(X \le 3) = \sum_{x=0}^{3} \binom{10}{x}\binom{30-10}{5-x} \Big/ \binom{30}{5}$

$= \text{phyper}(q=3,\ m=10,\ n=30-10,\ k=5) = .969$

SOME RELATIONSHIPS BETWEEN THESE DISTRIBUTIONS



I. Hypergeometric(N, M, K) → Binomial(n, p):

Draw $K \ge 0$ marbles from a bag of $N \ge 0$ marbles, where $M \ge 0$ of the N marbles are red. If X is the number of red marbles drawn, $X \sim$ Hypergeometric(N, M, K).

* Increase # marbles in bag s.t. $N \to \infty$, $M \to \infty$ and $\frac{M}{N} \to p$.

* set $n = K$

$$p_X(x;N,M,K) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$= \frac{\dfrac{M!}{(M-x)!\,x!}\;\dfrac{(N-M)!}{((N-M)-(n-x))!\,(n-x)!}}{\dfrac{N!}{(N-n)!\,n!}}$$

$$= \underbrace{\frac{n!}{(n-x)!\,x!}}_{\binom{n}{x}}\;\frac{\overbrace{M!/(M-x)!}^{M\cdot\ldots\cdot(M-x+1) \;\text{ for }\; x=1,\ldots,n}}{N!/(N-x)!}\;\frac{\overbrace{(N-M)!/((N-M)-(n-x))!}^{(N-M)\cdot\ldots\cdot((N-M)-(n-x)+1)}}{\underbrace{(N-x)!\,/\,(N-n)!}_{(N-x)\cdot\ldots\cdot(N-n+1)=(N-x)\cdot\ldots\cdot((N-x)-(n-x)+1)}}$$

$$= \begin{cases} \displaystyle\prod_{t=1}^{n}\left[\frac{N-M-n+t}{N-n+t}\right] & \text{if } x=0 \\ \displaystyle\binom{n}{x}\prod_{t=1}^{x}\left[\frac{M-x+t}{N-x+t}\right]\prod_{t=1}^{n-x}\left[\frac{N-M-(n-x)+t}{N-n+t}\right] & \text{if } x=1,\ldots,n-1 \\ \displaystyle\prod_{t=1}^{x}\left[\frac{M-x+t}{N-x+t}\right] & \text{if } x=n \end{cases}$$

Now for all $x=0,\ldots,n$ and $t=1,\ldots,x$

$$\lim_{N\to\infty}\left[\frac{M-x-t}{N-x+t}\right] = \lim_{N\to\infty}\left[\frac{M}{N}\right]\lim_{N\to\infty}\left[\frac{1-x/M-t/M}{1-x/N+t/N}\right] = p$$

and for all $t=1,\ldots,n-x$

$$\lim_{N\to\infty}\left[\frac{N-M-(n-x)+t}{N-n+t}\right] = \lim_{N\to\infty}\left[1-\frac{M}{N}\right]\lim_{N\to\infty}\left[\frac{1-\frac{n-x}{N-M}+\frac{t}{N-M}}{1-n/N+t/N}\right] = 1-p,$$

so $\prod_{t=1}^{x}\left[\dfrac{M-x+t}{N-x+t}\right] \to p^x$ and $\prod_{t=1}^{n-x}\left[\dfrac{N-M-(n-x)+t}{N-n+t}\right] \to (1-p)^{n-x}$

as $N\to\infty$. Plug in to see result.

II. Binomial(n,p) $\to$ Poisson(λ)

Let X be # successes in n independent Bernoulli trials with success probability λ/n. Then

$$p_X(x;n,\tfrac{\lambda}{n}) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n!}{(n-x)!\,x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n}$$

$$= \begin{cases}\left(1-\frac{\lambda}{n}\right)^n & \text{if } x=0\\ \left[\prod_{t=1}^{x}\frac{(n-x+t)}{n}\right]\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^n & \text{if } x\geq 1\end{cases}$$

We have for any finite x,

$$\lim_{n\to\infty}\frac{n-x+t}{n} = 1$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^x = 1$$

and we have

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

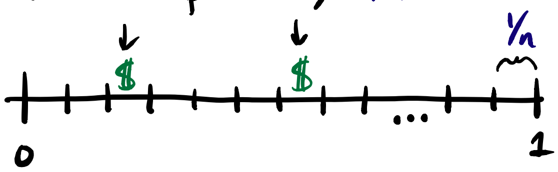
So $\displaystyle\lim_{n\to\infty}\binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}$ for all $x\geq 0$.

Illustration: Poisson(λ) for # occurrences per unit time/space.

Say I expect λ online purchases of goats' milk soap per day.
Break day into n time-bins of width $1/n$.
Assume purchase equally likely in any bin, purchases independent:

Let purchase occur in bin with probability λ/n

1 day:



Then $X = \#$ purchases $\sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$, and $\mathbb{E}X = n\left(\frac{\lambda}{n}\right) = \lambda$

Make grid finer: $n \to \infty$. Results in $X \sim \text{Poisson}(\lambda)$.