Hypergeometric distribution

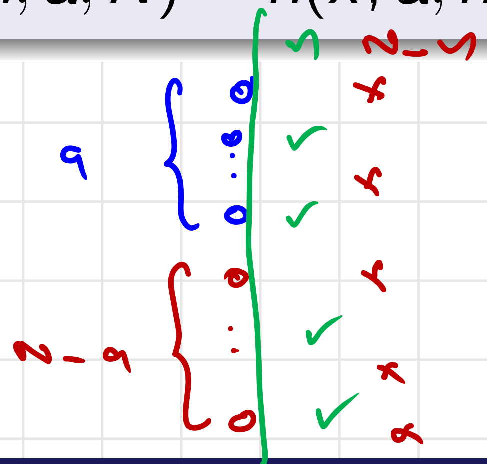
Definition 1.15

For $a, n, N \in \mathbb{N}$ such that $a \leq N, n \leq N$,

$$h(x; n, a, N) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}} \quad \text{for } x = 0, 1, \ldots, n.$$

Remark

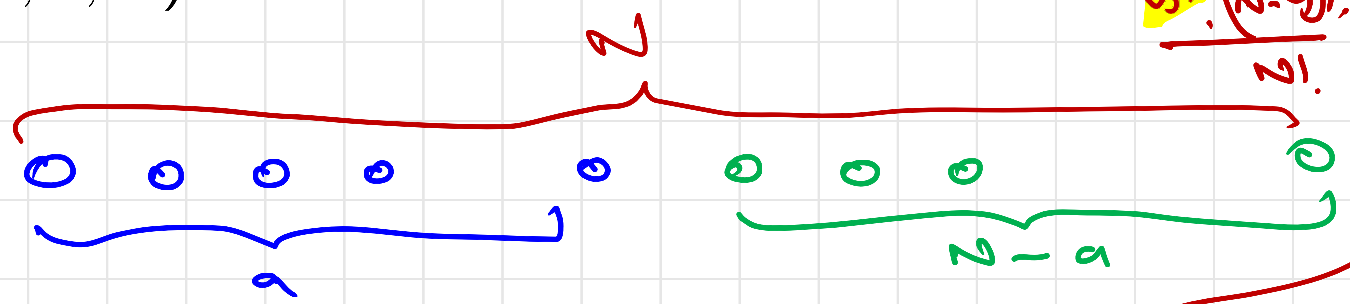
$h(x; n, a, N) = h(x; a, n, N)$



Hypergeometric distribution

$h(x; n, a, N)$ represents the probability of obtaining x items of Type I while choosing n items, without replacement, from a collection of N items of which a are of Type I (Success) and the remaining $N - a$ are of type II (Failure). Since items are picked without replacement, this distribution is different from Binomial distribution. But if n is small compared to N, the difference (between picking with replacement and without replacement) is not significant and so the binomial distribution with the parameters n and $p = \frac{a}{N}$ will be a good approximation to $h(x; n, a, N)$.

$$\frac{n!\,(N-n)!}{N!} \cdot \frac{(Np)!}{x!\,(Np-x)!} \cdot \frac{(N-Np)!}{(n-x)!\,(N-Np-n+x)!}$$



with replacement

$p = \frac{a}{N}$

$$\frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}} \xrightarrow{N \to \infty} \binom{n}{x} p^x (1-p)^{n-x}$$

$N \to \infty$

Hypergeometric distribution as a conditional p.d.f.

Let X and Y be independent binomial random variables having respective parameters (n, p) and (m, p). The conditional probability mass function of X given that $X + Y = k$ is as follows.

$$
\begin{aligned}
P\{X = i \mid X + Y = k\} &= \frac{P\{X = i, X + Y = k\}}{P\{X + Y = k\}} \\
&= \frac{P\{X = i, Y = k - i\}}{P\{X + Y = k\}} \\
&= \frac{P\{X = i\}P\{Y = k - i\}}{P\{X + Y = k\}} \\
&= \frac{\binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-(k-i)}}{\binom{n+m}{k} p^k (1-p)^{n+m-k}} \\
&= \frac{\binom{n}{i}\binom{m}{k-i}}{\binom{n+m}{k}} = h(i; k, n, n+m)
\end{aligned}
$$

$X = i$ $\quad$ $X+Y=k$ $\quad$ X $\quad$ $X+Y$

Poisson random variable

Definition 1.16

For a parameter $\lambda > 0$,

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0, 1, \ldots .$$

where e is Euler's number $\left(e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.71828\ldots\right)$.

$\underline{Ex'}$ $\quad \sum_{k=0}^{\infty} f(k,\lambda) = 1$

$\lambda,$ $\quad \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \qquad = 1.$

Mean of the Poisson distribution

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!}$$

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

$$= \lambda e^{-\lambda} \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots \right)$$

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

$$= \lambda$$

Variance of the Poisson distribution

$$
\begin{aligned}
E(X(X-1)) &= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=2}^{\infty}(x)(x-1)\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=2}^{\infty}\frac{e^{-\lambda}\lambda^x}{(x-2)!} = \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} \\
&= \lambda^2 e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots\right) \\
&= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathrm{Var}(X) &= E\left(X^2\right) - E(X)^2 = E(X(X-1)) + E(X) - E(X)^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}
$$

So both the expected value and the variance of X are equal to λ.

Exercise 1.16

Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week.

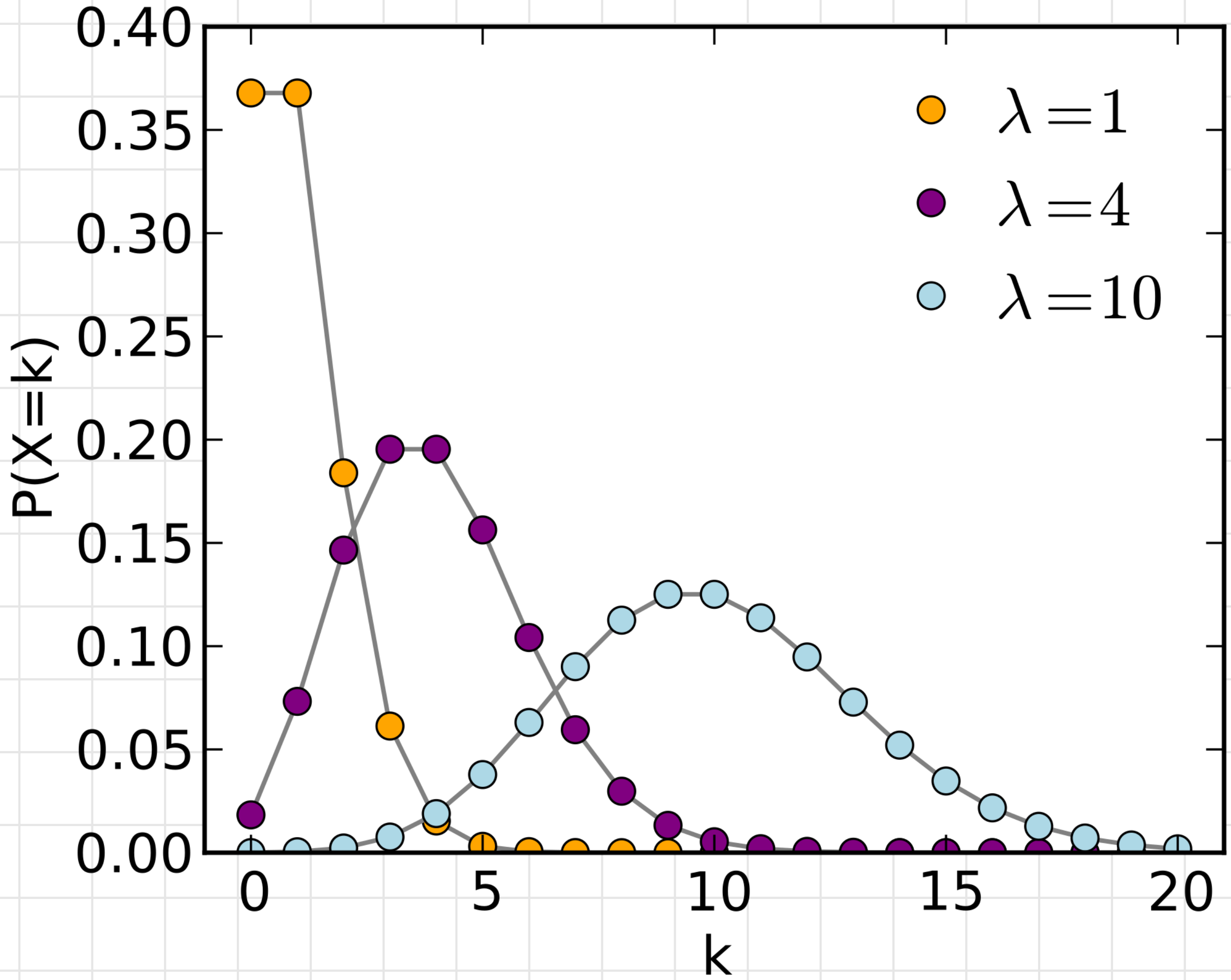
Assume that the number of accidents X follows $f(k, \lambda=3)$

ie, $P(X=k) = f(k,3) = \frac{e^{-3} 3^k}{k!}$

$P(X \geq 1) = 1 - P(X=0)$

$= 1 - \frac{e^{-3} 3^0}{0!} = 1 - e^{-3} \approx$ ______

The shape of the Poisson distribution



Poisson approximation of Binomial distribution

When the value of n in a binomial distribution is large and the value of p is very small, the binomial distribution can be approximated by a Poisson distribution with parameter $\lambda = np$.

Exercise 1.17

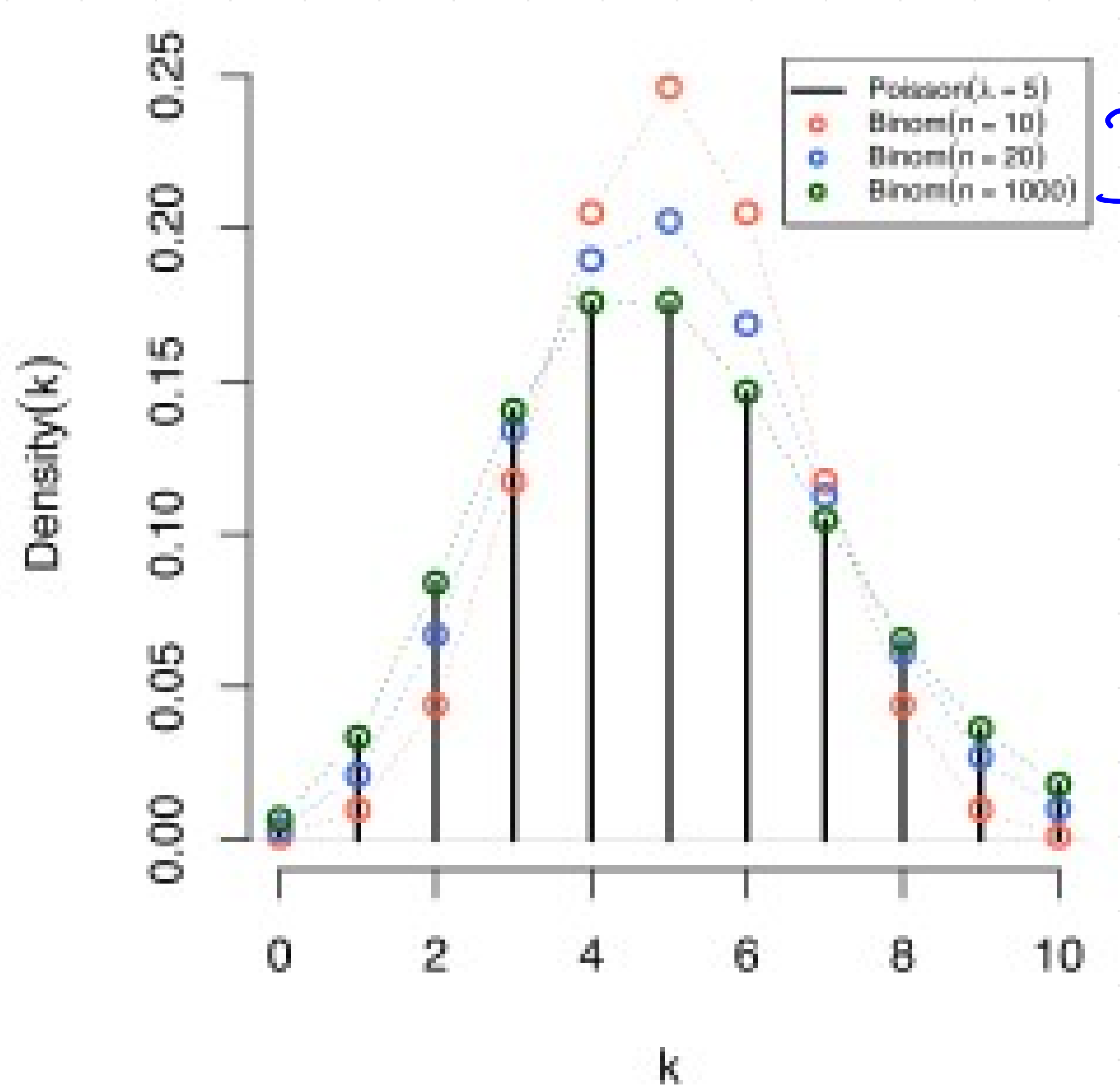
Suppose the probability that an item produced by a certain machine will be defective is .1. Let $\tilde{p}$ be the probability that a sample of 10 items will contain at most one defective item. Assume that the quality of successive items is independent and use the Poisson approximation to find $\tilde{p} \cdot e$.

$\lambda = np = 10 \times 0.1 = 1.$ — mean # defective items in a sample of 10

$\tilde{p} = P(X \leq 1) = P(X=0) + P(X=1)$

$= \frac{e^{-1} 1^0}{0!} + \frac{e^{-1} 1^1}{1!} = e^{-1}(1+1) = 2/e. \Rightarrow \tilde{p}e = 2$

Poisson approximation to Binomial distribution



$np = 5$

Exercise

Exercise 1.18

It is known that 5% of the books bound at a certain bindery have defective bindings. Find the probability that 2 of 100 books bound by this bindery will have defective bindings using (a) the formula for the binomial distribution; (b) the Poisson approximation to the binomial distribution.

$p = 5/100$

$\binom{100}{2} p^2 (1-p)^{98} =$

$\lambda = np = \frac{100 \times 5}{100} = 5$

$\frac{e^{-5} 5^2}{2!} =$