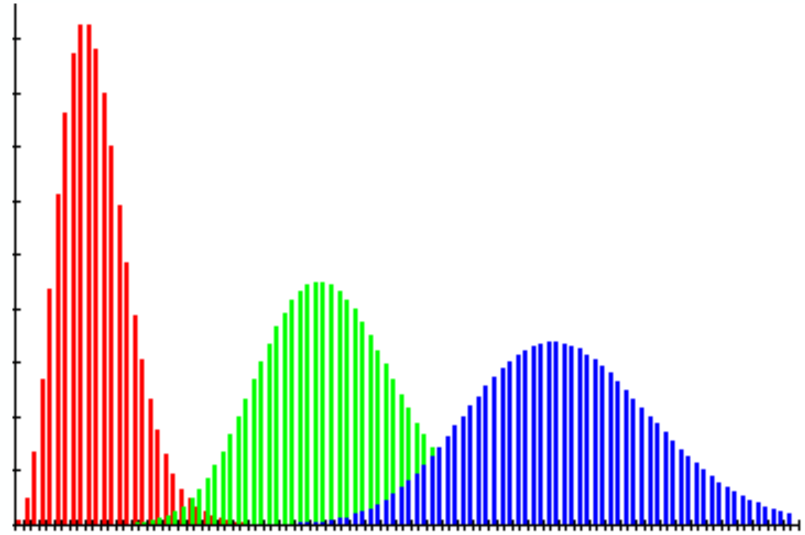
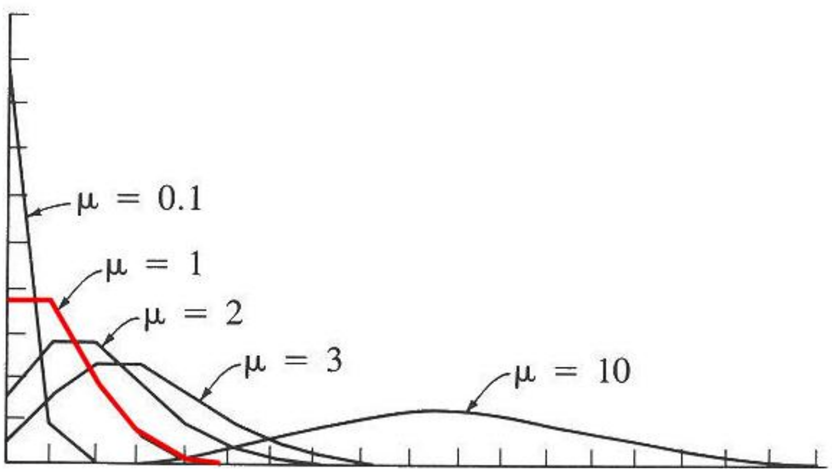
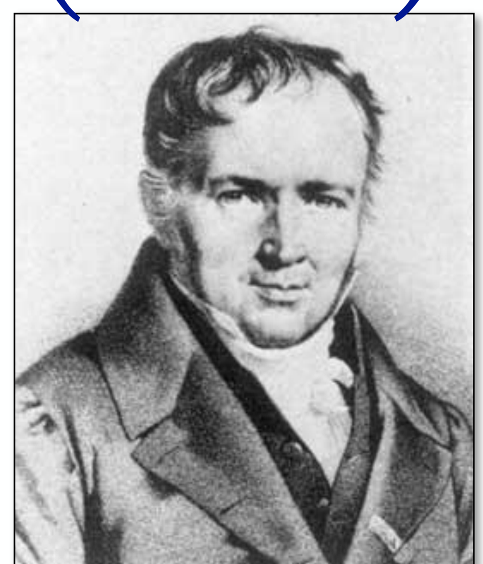


a zoo of (discrete) random variables (cont)





An experiment results in "Success" or "Failure"

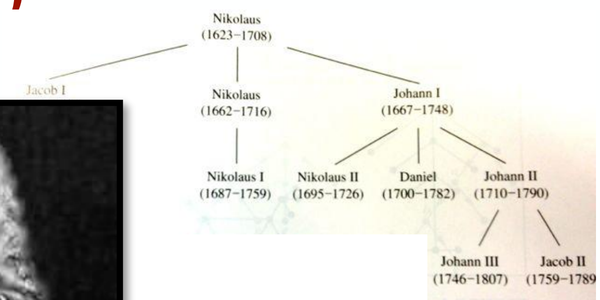
X is an *indicator random variable* (1 = success, 0 = failure)

$P(X=1) = p$ and $P(X=0) = 1-p$

X is called a *Bernoulli* random variable: $X \sim \mathrm{Ber}(p)$

$E[X] = E[X^2] = p$

$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$



Examples:

coin flip

random binary digit

whether a disk drive crashed

Jacob (aka James, Jacques) Bernoulli, 1654 – 1705

Consider n independent random variables $Y_i \sim Ber(p)$

$X = \Sigma_i Y_i$ is the number of successes in n trials

X is a *Binomial* random variable: $X \sim Bin(n,p)$

Examples

# of heads in n coin flips

# of 1's in a randomly generated length n bit string

# of disk drive crashes in a 1000 computer cluster

bit errors in file written to disk

of typos in a book

# of elements in particular bucket of large hash table

# of server crashes per day in giant data center

If $Y_1, Y_2, \ldots, Y_n \sim \mathrm{Ber}(p)$ and independent,

then $X = \sum_{i=1}^{n} Y_i \sim \mathrm{Bin}(n, p)$.

$$E[X] = np$$

$$E[X] = E\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} E\left[Y_i\right] = nE[Y_1] = np$$

$$\mathrm{Var}[X] = np(1-p)$$

$$\mathrm{Var}[X] = \mathrm{Var}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathrm{Var}\left[Y_i\right] = n\mathrm{Var}[Y_1] = np(1-p)$$

Siméon Poisson, 1781-1840

Suppose "events" happen, independently, at an *average* rate of λ per unit time. Let X be the *actual* number of events happening in a given time unit. Then X is a *Poisson* r.v. *with parameter* λ (denoted X ~ Poi(λ)) and has distribution (PMF):

$$P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$$

Examples:

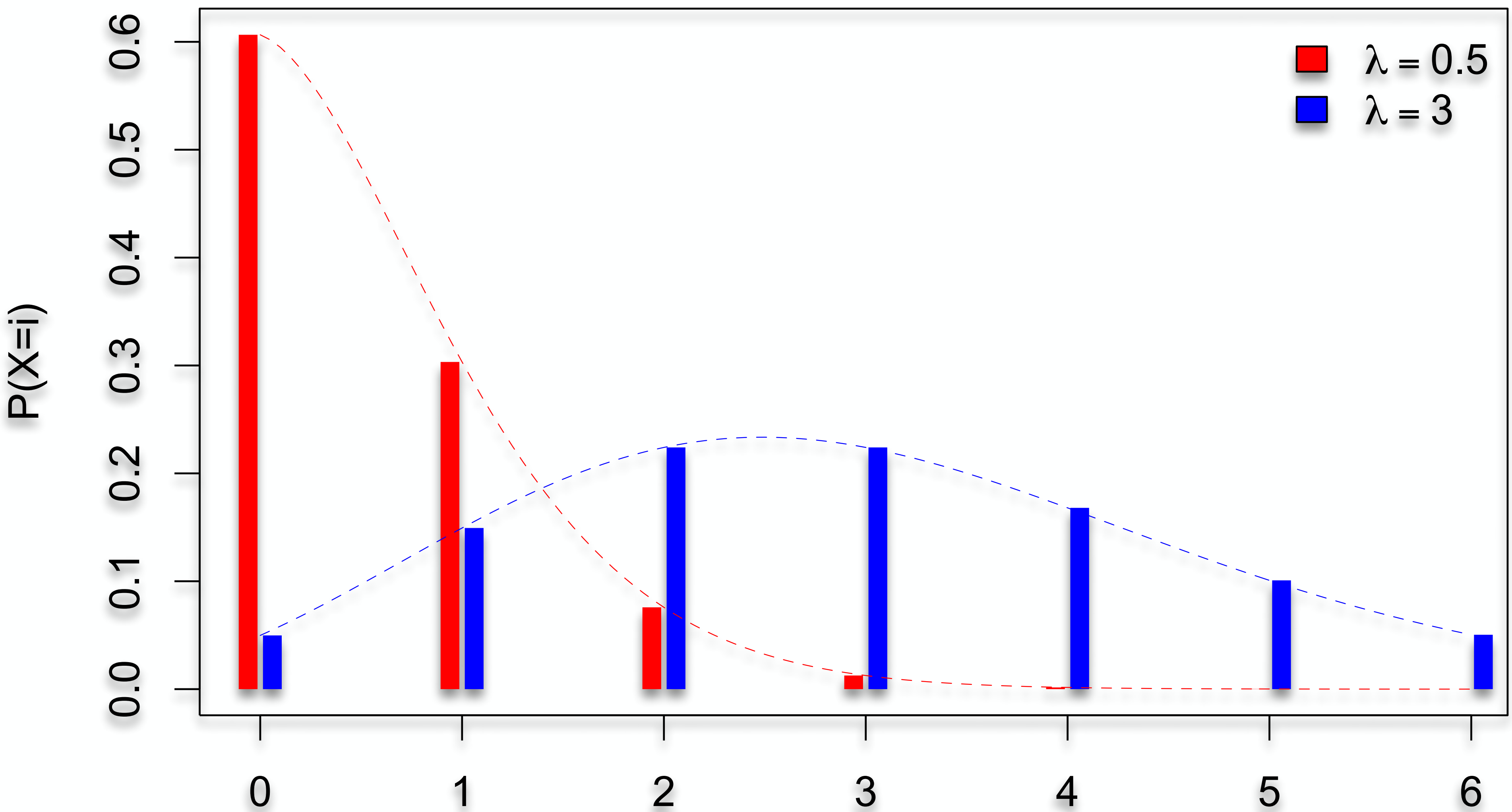
\# of alpha particles emitted by a lump of radium in 1 sec.

\# of traffic accidents in Seattle in one year

\# of babies born in a day at UW Med center

\# of visitors to my web page today

$$P(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}$$



X is a Poisson r.v. with parameter λ if it has PMF:

$$P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$$

Is it a valid distribution? Recall Taylor series:

$$e^{\lambda} = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \cdots = \sum_{0 \le i} \frac{\lambda^i}{i!}$$

So

$$\sum_{0 \le i} P(X = i) = \sum_{0 \le i} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{0 \le i} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

expected value of poisson r.v.s

$$
\begin{aligned}
E[X] &= \sum_{0\le i} i \cdot e^{-\lambda}\frac{\lambda^i}{i!} \\
&= \sum_{1\le i} i \cdot e^{-\lambda}\frac{\lambda^i}{i!} \\
&= \lambda e^{-\lambda}\sum_{1\le i}\frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda}\sum_{0\le j}\frac{\lambda^{j}}{j!} \\
&= \lambda e^{-\lambda}e^{\lambda} \\
&= \lambda
\end{aligned}
$$

i = 0 term is zero

j = i-1

As expected, given definition in terms of "average rate λ"

(Var[X] = λ, too; proof similar)

binomial random variable is poisson in the limit

Poisson approximates binomial when n is large, p is small, and $\lambda = np$ is "moderate"

Different interpretations of "moderate," e.g.

$n > 20$ and $p < 0.05$

$n > 100$ and $p < 0.1$

Formally, Binomial is Poisson in the limit as $n \to \infty$ (equivalently, $p \to 0$) while holding $np = \lambda$

X ~ Binomial(n,p)

$$
\begin{aligned}
P(X=i) &= \binom{n}{i} p^i (1-p)^{n-i} \\
&= \frac{n!}{i!(n-i)!} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}, \text{ where } \lambda = pn \\
&= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \\
&= \underbrace{\frac{n(n-1)\cdots(n-i+1)}{(n-\lambda)^i}} \frac{\lambda^i}{i!} \underbrace{(1-\lambda/n)^n} \\
&\approx \quad 1 \quad \cdot \frac{\lambda^i}{i!} \cdot \quad e^{-\lambda}
\end{aligned}
$$

I.e., Binomial ≈ Poisson for large n, small p, moderate i, λ.

Handy: Poisson has only 1 parameter–the expected # of successes

Consider sending bit string over a network

Send bit string of length $n = 10^4$

Probability of (independent) bit corruption is $p = 10^{-6}$

$X \sim \text{Poi}(\lambda = 10^4 \cdot 10^{-6} = 0.01)$

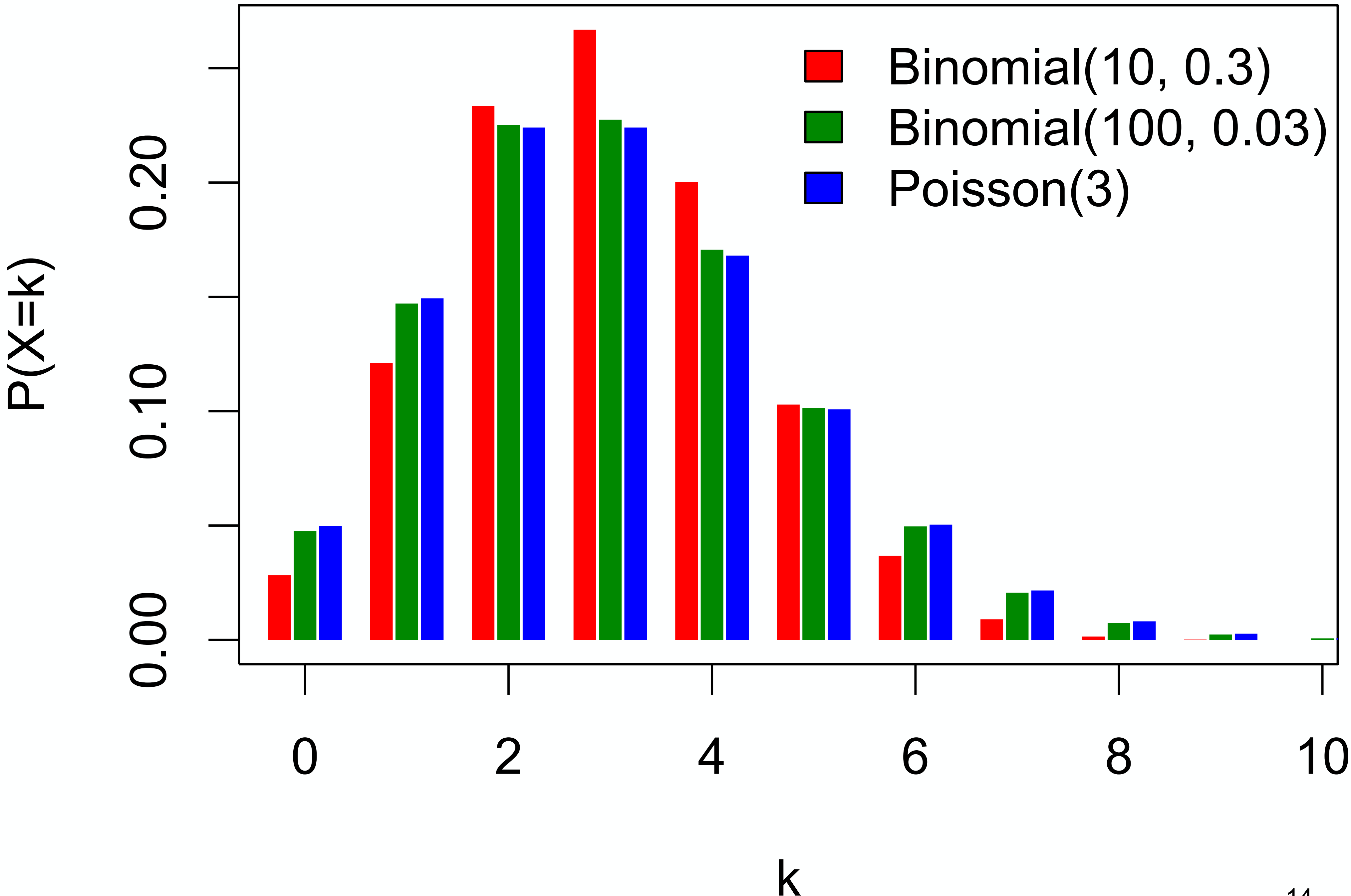
What is probability that message arrives uncorrupted?

$$P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-0.01} \frac{0.01^0}{0!} \approx 0.9900498\underline{34}$$

Using $Y \sim \text{Bin}(10^4, 10^{-6})$:

$P(Y=0) \approx 0.9900498\underline{29}$

I.e., Poisson approximation (here) is accurate to ~5 parts per billion


Binomial(10, 0.3)
Binomial(100, 0.03)
Poisson(3)
P(X=k)
0.00
0.10
0.20
0
2
4
6
8
10
k

Recall: if $Y \sim Bin(n,p)$, then:

$E[Y] = pn$

$Var[Y] = np(1-p)$

And if $X \sim Poi(\lambda)$ where $\lambda = np$ ($n \to \infty$, $p \to 0$) then

$E[X] = \lambda = np = E[Y]$

$Var[X] = \lambda \approx \lambda(1-\lambda/n) = np(1-p) = Var[Y]$

Sum of independent Poissons is Poisson

Important Examples:

Uniform(a,b): $P(X = i) = \frac{1}{b - a + 1}$ $\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)(b-a+2)}{12}$

Bernoulli(p): $P(X = 1) = p$, $P(X = 0) = 1\text{-}p$ $\mu = p$, $\sigma^2 = p(1\text{-}p)$

Binomial(n,p) $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$ $\mu = np$, $\sigma^2 = np(1\text{-}p)$

Poisson(λ): $P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$ $\mu = \lambda$, $\sigma^2 = \lambda$

$Bin(n,p) \approx Poi(\lambda)$ where $\lambda = np$ fixed, $n \to \infty$ (and so $p = \lambda/n \to 0$)

Geometric(p) $P(X = k) = (1\text{-}p)^{k\text{-}1}p$ $\mu = 1/p$, $\sigma^2 = (1\text{-}p)/p^2$