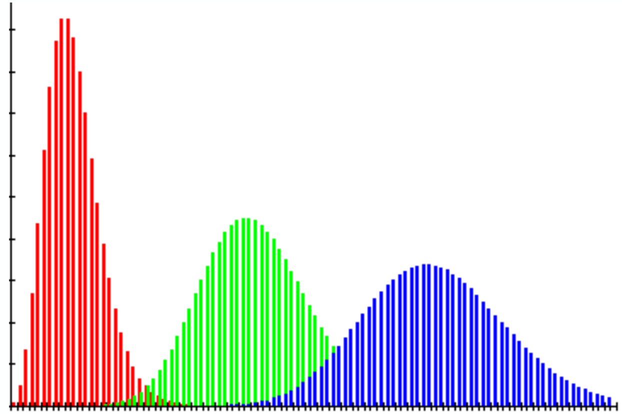
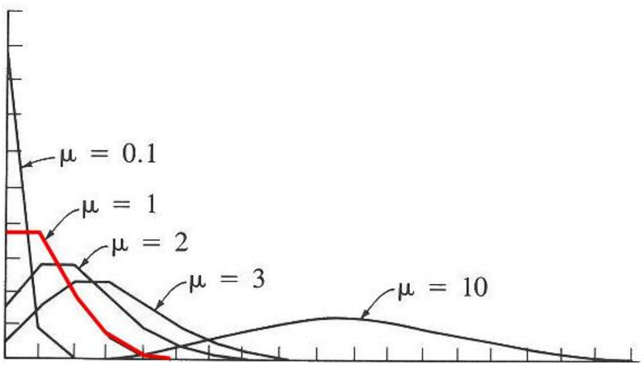
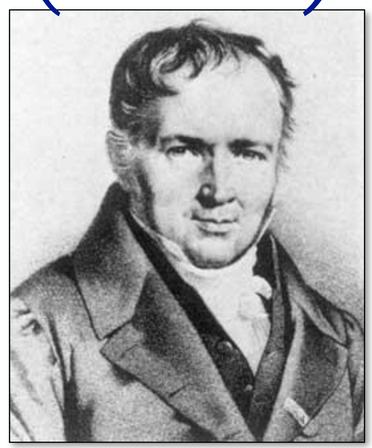


a zoo of (discrete) random variables (cont)





(Ω, P)

$X: \Omega \to \mathbb{R}$

(Ω_X, P_X)

Range pmf.

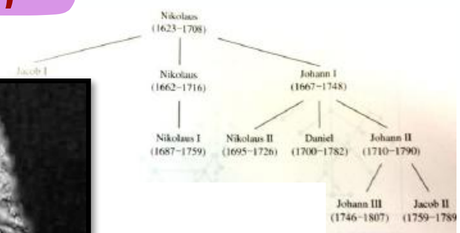
$\Omega_X \subseteq \mathbb{R}$

$$\sum_{x \in \Omega_X} P_X(x) = 1$$

$P_X(x) \geq 0$

An experiment results in “Success” or “Failure”

- X is an *indicator random variable* (1 = success, 0 = failure)
  - $P(X=1) = p$ and $P(X=0) = 1\text{-}p$
- X is called a *Bernoulli* random variable: $X \sim \text{Ber}(p)$
- $E[X] = E[X^2] = p$
- $\text{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1\text{-}p)$



Examples:

- coin flip
- random binary digit
- whether a disk drive crashed

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

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

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

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

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

$$\boxed{E[X] = np}$$

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

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

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

Poisson motivation

Model: # events that occur in an hour

→ rate 3 events per hour.

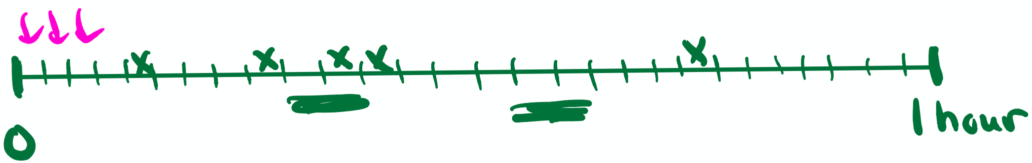
→ exp # events that occur in interval of t hours
$3t$

→ occurence of events in disjoint time intervals is indep.

$\boxed{X \text{ \# events that occur in one hour}}$

$E(X) = 3$

$\boxed{0.2n \cdot \frac{3}{n}} = 0.2 \cdot 3$



divide 1 hour into n intervals
each of length $\frac{1}{n}$ hour.

- in each interval, prob event is p

Bin(n,p) → count # events that occur.

take limit as $n \to \infty$

$p = ?$

a) $\boxed{\frac{3}{n}}$

b) $3n$

c) 3

d) $\frac{3}{60}$

$\Downarrow$ Exp value $= n \cdot p = 3$

$n \cdot \frac{3}{n} = \boxed{3}$

X # events in 1 hour

define as $\underbrace{\text{limit } Y}_{n \to \infty}$ $\quad Y \sim Bin(n, \frac{3}{n})$

what is this limit.

$Pr(X=0) = \lim_{n\to\infty} \underbrace{Pr(Y=0)}$

$$\underbrace{\left(1 - \frac{3}{n}\right)^n}$$

$$\left(e^{-\frac{3}{n}}\right)^n$$

$$e^{-3}$$

$1 - x \approx e^{-x}$

x very small

$e^{-x} = 1 - x + \underbrace{\frac{x^2}{2} - \frac{x^3}{3!} \cdots}$

$Bin(n, \frac{3}{n})$

$Pr(X=1) = \lim_{n\to\infty} \underbrace{Pr(Y=1)}$

$$\binom{n}{1}\left(\frac{3}{n}\right)^1 \underbrace{\left(1 - \frac{3}{n}\right)^{n-1}}$$

$\left(e^{-\frac{3}{n}}\right)^{n-1}$

$$= n \cdot \frac{3}{n} \left(e^{-n}\right)$$

$$= 3 \cdot e^{-3\left(\frac{n-1}{n}\right)}$$

$\uparrow$

1

$$= 3 \cdot e^{-3}$$

$$\Pr(X = i) = e^{-3} \frac{3^i}{i!} \Leftarrow \quad i = 0, 1, 2, \ldots$$

Poisson random variables

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):

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

$n \cdot p = \lambda$

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

$\boxed{\Omega_X = ?}$

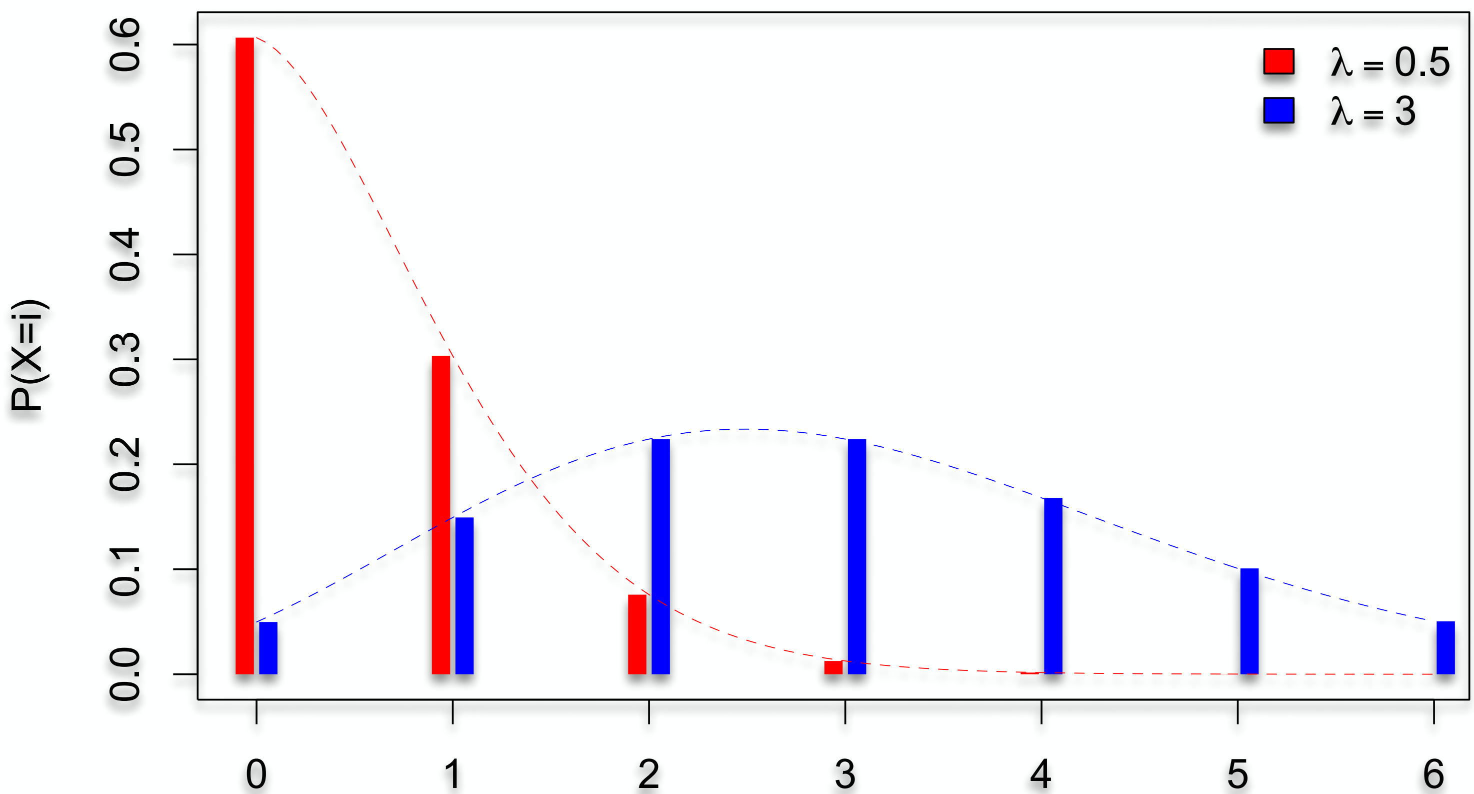
a) 0, 1, 2, ...

b) 1, 2, ...

c) 0, 1, 2, ..., n

d) I don't know

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



$\lambda \geq 0$

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

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

$i = 0, 1, 2, \ldots$

$p_X(i) \geq 0$ ✓

$$\sum_{i=0}^{\infty} p_X(i) = 1 \; ?$$

$$\sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\underbrace{\boxed{\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}}}_{e^{\lambda}} = e^{-\lambda}e^{\lambda} = 1$$

✓

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

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

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

E(X) = ?

exp(# of events in 1 unit of time)

a) $\boxed{\lambda}$
b) λt
c) 60λ
d) I don't know

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

$= \sum_{i=0}^{\infty} i \, \boxed{p_X(i)}$

$\sum_{i=1}^{\infty} i \, p_X(i)$

i = 0 term is zero

j = i-1

$\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$

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

(Var[X] = λ, too; proof similar)

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 $np = \lambda$

$\text{Bin}(n,p) \Longrightarrow$ approx with $\text{Poisson}(\lambda = np)$

X ~ Binomial(n,p)

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

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

#corrupted bits $\sim \text{Bin}(10^4, 10^{-6})$

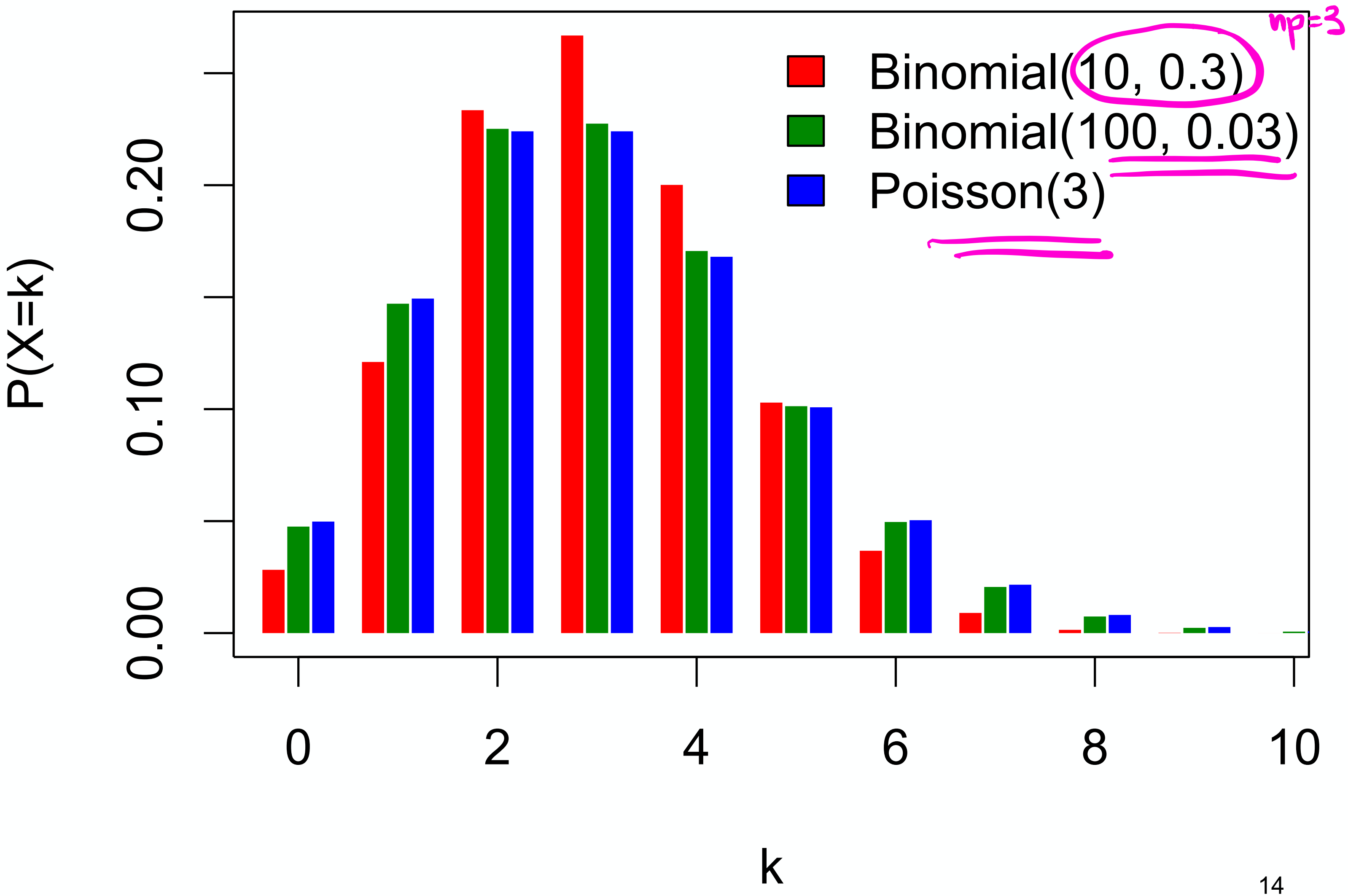
- Send bit string of length n = 10^4
- Probability of (independent) bit corruption is p = 10^{-6}
- $\boxed{X \sim \text{Poi}(\lambda = 10^4 \cdot 10^{-6} = 0.01)}$
- What is probability that message arrives uncorrupted?

$$P(X = 0) = \boxed{e^{-\lambda}\frac{\lambda^0}{0!}} = e^{-0.01}\frac{0.01^0}{0!} \approx 0.990049834$$

Using Y ~ Bin(10^4, 10^{-6}):

P(Y=0) ≈ 0.990049829

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


np=3
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 ~ Bin(n,p), then:

$\mathrm{Bin}(n,p) \Leftrightarrow \mathrm{Poi}(\lambda)$

$\lambda = np$

$E[Y] = pn$

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

$p = \frac{\lambda}{n}$

And if X ~ Poi(λ) where $\boxed{\lambda = np}$ ($n \to \infty$, $p \to 0$) then

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

$Var[X] = \lambda \approx \underbrace{\lambda(1-\lambda/n)}_{\downarrow\, n\to\infty\; 0} = \underbrace{np(1-p)} = Var[Y]$

Sum of independent Poissons is Poisson

$X \sim \mathrm{Poi}(\lambda_1)$ $\quad$ $Y \sim \mathrm{Poi}(\lambda_2)$

X, Y independent.

$X+Y \sim \mathrm{Poi}(\lambda_1+\lambda_2)$

$\Pr(X+Y=k)$

Use Law of total probability

partition: $X=0$, $X=1$, $\vdots$

Sum of independent Poissons is Poisson

Law of total prob.

Partition: $\underbrace{\{X=0\}}_{E_1}, \underbrace{\{X=1\}}_{E_2}, \underbrace{\{X=2\}}_{E_3}, \ldots$

$\underline{\Pr(X+Y=k) = ?}$

Which is correct?

(a) all of (1),(2)(3)(4)

$\boxed{\text{(b) only (1)(3)(4)}}$

(c) only (2)

(d) I don't know

$\Rightarrow$ (1) $\sum_{i=0}^{\infty} \Pr(X+Y=k \mid Y=i)\Pr(Y=i)$

(2) $\sum_{i=0}^{k} \Pr(X+Y=k \mid Y=i)$

$\Rightarrow$ (3) $\sum_{i=0}^{k} \Pr(X+Y=k \cap Y=i)$

$\Rightarrow$ (4) $\sum_{i=0}^{k} \Pr(X+Y=k \mid Y=i)\Pr(Y=i)$

$$\Pr(F) = \sum_{i=0}^{\infty} \underline{\Pr(F \cap E_i)} = \sum_{i=0}^{\infty} \underline{\Pr(F \mid E_i)\Pr(E_i)}$$

$X \sim Poi(\lambda_1)$ $Y \sim Poi(\lambda_2)$

Sum of independent Poissons is Poisson

$$\Pr(X+Y=k) = \sum_{i=0}^{k} \Pr(X+Y=k \mid X=i)\,\underbrace{\Pr(X=i)}_{e^{-\lambda_1}\frac{\lambda_1^{i}}{i!}}$$

$$\frac{\Pr(X+Y=k,\ X=i)}{\cancel{P(X=i)}}$$

$$\Pr(Y=k-i,\ X=i)$$

$$\overset{\text{indep}}{=} \Pr(Y=k-i)\,\cancel{\Pr(X=i)}$$

$$\sum_{i=0}^{k} e^{-\lambda_2}\frac{\lambda_2^{k-i}}{(k-i)!}\cdot e^{-\lambda_1}\frac{\lambda_1^{i}}{i!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\frac{k!}{i!\,(k-i)!}\lambda_1^{i}\lambda_2^{k-i}$$

$$\overset{\text{Binomial Thm}}{=} \frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1+\lambda_2)^{k}$$

$$(a+b)^k = \sum_{i=0}^{k}\binom{k}{i}a^i b^{k-i}, \quad \binom{k}{i} = \frac{k!}{i!\,(k-i)!}$$

p.m.f. $\sim Poi(\lambda_1+\lambda_2)$

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