Many of the random variables we will encounter are counts or measures of something -- their possible values are positive numbers, such as:

- the number of deaths due to AIDS in Providence in 1995.
- the weight of a baby at birth
- the level of an antibody in a patient's blood
- my remaining lifetime
- the number of lottery tickets I will buy before winning a payoff of $500 or more

If we're observing how many boys there are in n births, or how many of the next n mothers who deliver babies are HIV+, we might use a binomial probability model:

(1) X = the number of boys ... or HIV+ mothers
(2) S = {0,1,2,...,n}
(3) $P(X=k) = \binom{n}{k}\theta^k (1-\theta)^{n-k} \qquad k=0,1,2,\ldots,n$

The binomial probability model is appropriate:

(1) when we're observing a count of how many times something happens, (**Bernoulli trials**)

(2) when there is a fixed number of independent trials on which it either happens or does not (**fixed number of independent trials**), and

(3) when the chance that it will happen is the same on each trial (**constant θ**).

The Poisson Probability Distribution

Here is another probability model for count data:

(1) X = the number of occurrences
(often out of a specified time period)
(2) S = {0,1,2, ...}
(3) P(X = k) = $\lambda^k e^{-\lambda}$ /k! k = 0,1,2,...

This is a **Poisson probability** distribution model. In this model the variable λ (the Greek letter, lambda) can be any positive number and is the rate of a single occurrence.

Note also that there is no upper limit on the possible values of X.

The Poisson probability model is used for random variables like

- the number of radioactive particles emitted from some source during a fixed time interval.
- the number of suicides in one week in the U.S.
- the number of crimes per day when there is a full moon in three small areas of India
- the number of deaths per day in a hospital in Montreal.

Example

It is estimated that in Providence, RI, the average number of births that occur in a taxi on the way to the hospital is 2 per day.

The poison model is appropriate here:

(1) X = number of taxi births
(2) S = {0,1,2,3,…}
(3) P(X = K) = $\lambda^k e^{-\lambda}$ /k!   k = 0,1,2,...

Where λ=2=average number of taxi births/day.
(The notation is X~Poiss(λ=2).)

On any given day:

- The probability of observing 3 births is

  P(X = 3) = $\lambda^3 e^{-\lambda}$ /3! = 0.1804 or 18%

- The probability of observing at least one birth is

$$P(X \geq 1) = \sum_{k=1}^{\infty} P(X = k) = 1 - P(X < 1) = 1 - P(X = 0)$$

1-P(X = 0) =1- $\lambda^0 e^{-\lambda}$ /0! = 1- 0.1353 = 86.47%

- The probability of observing 3 births, after sharing a taxi with a woman who gave birth?

$$
\begin{aligned}
P(X=3|X\geq 1) &= P(X=3 \text{ and } X\geq 1)/P(X\geq 1)\\
&= P(X=3)/P(X\geq 1)\\
&= 0.1804/0.8647\\
&= 0.2086\\
&= 20.9\%
\end{aligned}
$$

So after observing at least one birth, the probability of observing exactly 3 births rises from 18% to 21%.

The Poisson probability model is appropriate:

(1) when we're observing a count of how many times something happens (usually within a time period),

(2) when there is a possibility of observing an unlimited number of these events in that time period, and

(3) when the events that occur within, and between, varying time periods are all mutually independent.

Binomial and Poisson

The Poisson distribution is often used to approximate the Binomial distribution when either the number of trials (n) is large or the probability of success (θ) is near 0 or 1.

- This approximation can be especially useful because the $\binom{n}{k}$ term can be laborious to calculate when n is very large.

Pagano and Gauvreau (p.172) use the Poisson model to represent the number of people in a population of 10,000 who will be involved in a motor vehicle accident next year. They use the value $\lambda = 2.4$, and calculate various probabilities on pages 173 & 174.

To illustrate how similar the Poisson and Binomial models are in this situation, we can calculate the probabilities of the same values, using the binomial distribution model with n = 10,000 and θ = 2.4/10,000.

⇨ (As we will soon learn, these two models both give the same <u>expected value</u> for the number of persons who will be involved in an auto accident--both say the expected value is 2.4.)

X~Poisson(λ=2.4) **X~Bin(n, θ=0.00024) (n=10,000)**

$$P(X=0)=\frac{(2.4)^0 e^{-2.4}}{0!}=0.0907 \qquad P(X=0)=\binom{10{,}000}{0}\theta^0(1-\theta)^{10{,}000-0}=0.0907$$

$$P(X=1)=\frac{(2.4)^1 e^{-2.4}}{1!}=0.2177 \qquad P(X=1)=\binom{10{,}000}{1}\theta^1(1-\theta)^{10{,}000-1}=0.2177$$

$$P(X=2)=\frac{(2.4)^2 e^{-2.4}}{2!}=0.2613 \qquad P(X=2)=\binom{10{,}000}{2}\theta^2(1-\theta)^{10{,}000-2}=0.2613$$

$$P(X=3)=\frac{(2.4)^3 e^{-2.4}}{3!}=0.2090 \qquad P(X=3)=\binom{10{,}000}{3}\theta^3(1-\theta)^{10{,}000-3}=0.2090$$

etc.

These calculations show that the two models are essentially the same (under these conditions), both giving the same probabilities to the 4-decimal-place accuracy given here.

Does the event have to be 'rare'?

No, λ may be any positive real number. Events are not required to be rare in any sense. End of story.

The often thought 'Rareness requirement' comes from calculations like those on the previous slide. They show that the Poisson distribution is a good approximation to the Binomial distribution only when the events (or successes) are rare, i.e. n is large and θ is small.

⇨ The Poisson distribution is a probability model in is own right and does not have any restrictions on the magnitude of the frequency of events within a specified time interval.