Example 1.22 (p. 53)

![Diagram showing batches and sample sizes](image)

\[ X = 3 \text{ (e.g.)} \quad 5 \text{ (e.g.)} \quad 0 \text{ (e.g.)} \quad \cdots = \begin{cases} \# \text{ of Bads.} \\ \text{in sample of} \\ \text{size 100} \end{cases} \]

Assume the lots are identical, i.e.

The company manufacturing the 5,000 things is extremely consistent.

Then, the picture looks like this:

\[ \text{Sample 1} \quad 2 \quad \cdots \quad \text{Sample} \quad 10^8 \]

What proportion of these \( 10^8 \) lots will have \( X = 0, 1, \ldots, 100 \)?

\[ G = \text{Good} \quad \text{Sample} = \{G, G, \ldots, G\} \]

\[ B = \text{Bad} \quad \text{Sample} = \{B, B, \ldots, B\} \]

Suppose we know the prop. of Bads, period, in the pop. = 0.5%

Then \[ P(X = k) = \binom{100}{k} \pi^k (1-\pi)^{100-k} \]

\[ \pi = 0.005 \]

prop. of lots with \( X = 0 \):

\[ (100)^0 \pi^0 (1-\pi)^{100} = 0.6058 \]

\[ = 1 : \quad (100)^1 \pi^1 (1-\pi)^{99} = 0.3044 \]

\[ = 2 : \quad (100)^2 \pi^2 (1-\pi)^{98} = 0.0757 \]

\[ = 3 : \quad \ldots \quad \text{Etc.} \quad = 0.0124 \]

Important Interpretation

In the long run, we expect \( \sim 60\% \) of the lots to be all good.

\( \sim 30\% \) \( \ldots \) to have 1 bad out of 100.

\( \sim 7\% \) \( \ldots \) \( \ldots \) 2 bads \( \ldots \)

(i.e. 7% of the lots to be 2% defective)
Q1: A sample is taken from a population of boys & girls. The binomial mass function provides the proportion of ______ with certain characteristics.

A) people in the sample  B) people in the pop.  C) Samples  D) None of the above

For large $n$, all the factorials get nasty.

How did the French handle this problem?

Poisson noted that if

$n \rightarrow \text{large}$, \hspace{1cm} n\pi = \text{const} = \lambda$

$\pi \rightarrow \text{small}$ [rare event]

\[
\binom{n}{x} \pi^x (1-\pi)^{n-x} \approx \frac{e^{-\lambda} \lambda^x}{x!} = \text{Poisson mass function with parameter } \lambda
\]

Recall $x = \# \text{ of 1's} = \text{discrete}$

No $n!$. No $\pi$! Just $\lambda (= \text{avg. rate of occurrence})$

In the example, since we know $n$ & $\pi$, we can compute $\lambda$:

$\lambda = n - \pi = 100 - 0.005 = 0.5$

Then

\[
\text{prop. } (x=0) \approx \frac{e^{-\lambda} \lambda^0}{0!} = e^{-0.5} \approx 0.6065
\]

Similarly, for prop $(x=1, 2, 3)$, ..., $\approx$ exact answers from binomial.

As mentioned previously, and proved later, $\lambda$ of the Poisson dist. can be interpreted as the "average" of $x$. 
Although I derived Poisson as a large-$n$ limit of binomial, it turns out that some problems can be solved with Poisson, quite independently of Binomial, e.g. when you have a (average rate) but not $n$ or $\pi$.

Examples of data that follow the Poisson distr:
- # of bombs dropped over London per block.
- # of knots per unit length of wood.
- # of crashes (cars, planes, buildings) per year.
- # of people arriving at a teller per unit time, $\lambda$.

Eg: An avg of 4 people arrive at a teller per hour. What's the proportion (prob) of 3 people arriving in any hour?

Assume $X = \text{poisson with } \lambda = 4 \text{ people/hr}$.

$p(X=3) = e^{-4} \frac{4^3}{3!} = 0.19$

$\theta = 1 = \text{average arrival rate}$
In the prev. chs we played with histograms of sample data and distributions of (random) variables (cont. and discrete).

Histograms and distributions are the pillars of statistics.

In statistics, we describe the population in terms of distributions, and then ask: “Based on the histogram of my sample (data), could the sample have come from, say, a normal distribution with parameters μ = 13, σ = 3?”

If “No,” then we know something about the population.

One way to compare the hist. with the distr. is in terms of their summary measures. For example, we compare the “location” of the distribution (e.g., μ) with the location of the histogram (e.g., sample mean).

The location of a distr. is (usually) one of its parameters.

A histogram is called a statistic.

In short, one compares parameters with statistics. Later Ch. 7, ...
Examples of statistics for location are: The $x_i$ for the $i^{th}$ case.

- **Sample mean**: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i

- **Sample median**: $\tilde{x}$ = middle of the ordered data.

Examples of statistics for spread are:

- **Sample Range** (same units as $\bar{x}$)

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

  - Average of (deviations) $^2$

  $S \sim \text{"typical" spread/deviation.}$

$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = n \bar{x} - n \bar{x} = 0$.

In summary, we will use $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and

$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$ as our measures of location and spread of data.

**Example:**

$x = \{1, 3, 8\}$

$\bar{x} = \frac{1}{3} (1+3+8) = 4$

$S^2 = \frac{1}{2-1} \left[ (1-4)^2 + (3-4)^2 + (8-4)^2 \right] = \frac{1}{2} (9+1+16) = 13 \Rightarrow S = \sqrt{13}$
The Poisson mass function in the Teller example is “flat” at the top, i.e., \( p(x) \) has the same value at \( x=3 \) and \( x=4 \). Show that, quite generally,

The Poisson mass function has the same value at \( \lambda \) (i.e., at the average) and at \( (\lambda - 1) \).

Consider the examples of Poisson in lecture.

a) Find another example (google, books,...) that qualifies as a Poisson variable. Call it \( X \), and define it clearly.

b) Assume, or even guess, what the value of the \( \lambda \) parameter may be for your example. Remember, \( \lambda \) is the average \( x \). State that value, with the correct units.

c) Plot the Poisson dist. with that value of \( \lambda \) (by \( R \) or)

d) Compute \( p(x=0) \), and interpret it.