The Pascal, Negative Binomial and Poisson Distributions

Generalizing the Geometric Distribution

Consider a sequence of independent, identical Bernoulli trials with probability $0 < p < 1$ of success.

Question: How many trials are necessary to obtain $r > 0$ successes, where r is a fixed parameter?

(The situation described by the geometric distribution corresponds to the case $r = 1$ here.)

We calculate the probability that $x \geq r$ trials are needed to obtain r successes.

Counting Trials for r Successes

Main idea: If the r^{th} success is obtained in the x^{th} trial, then there must have been ***exactly $r-1$ successes in the previous $x-1$ trials***.

$$P[\text{exactly } r-1 \text{ successes in } x-1 \text{ trials}] = \binom{x-1}{r-1} p^{r-1}(1-p)^{x-r}.$$

Now with probability p the xth trial will be a success, so

$$P[\text{obtain } r^{\text{th}} \text{ success in the } x^{\text{th}} \text{ trial}] = \binom{x-1}{r-1} p^{r}(1-p)^{x-r}.$$

The Pascal Distribution


Blaise Pascal (1623-1672). Anonym. ca. 1690. Painting. Palais de Versailles. Paris. File:Blaise Pascal Versailles.JPG. (2020, February 12). Wikimedia Commons, the free media repository.

4.1. Definition. Let $r \in \mathbb{N} \setminus \{0\}$. A random variable (X, f_X) with

$$\begin{aligned} X\colon S \to \Omega &= \mathbb{N} \setminus \{0, 1, \dots, r-1\} \\ &= \{r, r+1, r+2, \dots\} \end{aligned}$$

and distribution function $f_X\colon \Omega \to \mathbb{R}$ given by

$$f_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad 0 < p < 1,$$

is said to follow a ***Pascal distribution*** with parameters p and r.

The Negative Binomial Distribution

Instead of counting the number of trials needed to obtain r successes, we may count ***the number of failures obtained before r successes***:

4.2. Definition. Let $r \in \mathbb{N} \setminus \{0\}$. A random variable (X, f_X) with

$$X : S \to \Omega = \mathbb{N}$$

and distribution function $f_X : \Omega \to \mathbb{R}$ given by

$$f_X(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \qquad 0 < p < 1,$$

is said to follow a ***negative binomial distribution***with parameters p and r.

The Negative Binomial Distribution

The term "negative binomial" comes from the fact that

$$\begin{aligned}\binom{-r}{x} &= \frac{(-r)\cdot(-r-1)\cdots(-r-x+1)}{x!} \\ &= \frac{r\cdot(r+1)\cdots(r+x-1)}{x!}(-1)^x \\ &= (-1)^x\frac{(r+x-1)!}{x!(r-1)!} = (-1)^x\binom{r-1+x}{r-1}\end{aligned}$$

so that the density of the negative binomial distribution may be expressed as

$$f_X(x) = \binom{-r}{x}(-1)^x p^r(1-p)^x$$

We now return to the Pascal distribution.

The M.G.F. for the Pascal Distribution

4.3. Theorem. Let (X, f_X) be a Pascal random variable with parameters p and r.

(i) The moment generating function of X is given by

$$m_X\colon (-\infty, -\ln q) \to \mathbb{R}, \qquad m_X(t) = \frac{(pe^t)^r}{(1-qe^t)^r}, \qquad q = 1-p.$$

(ii) $\mathrm{E}[X] = r/p$.

(iii) $\operatorname{Var} X = rq/p^2$.

Using Mathematica:

```
MomentGeneratingFunction[PascalDistribution[r, p], t]
```

$$\left(\frac{e^t\, p}{1 - e^t\,(1-p)}\right)^r$$

The M.G.F. for the Pascal Distribution

Proof.

We derive the moment-generating function only. It is given by

$$\begin{aligned} m_X(t) = \mathrm{E}[e^{Xt}] &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= \sum_{x=0}^{\infty} e^{t(r+x)} \binom{r+x-1}{r-1} p^r (1-p)^x \\ &= p^r e^{tr} \sum_{x=0}^{\infty} \binom{-r}{x} [-e^t(1-p)]^x \end{aligned}$$

Recall the ***binomial series***

$$(1-y)^{-r} = \sum_{x=0}^{\infty} \binom{-r}{x} (-y)^x \qquad \text{for } |y| < 1.$$

The M.G.F. for the Pascal Distribution

Proof (continued).

It follows that, as long as $e^t(1-p) < 1$,

$$\begin{aligned} m_X(t) &= p^r e^{tr} \sum_{x=0}^{\infty} \binom{-r}{x} [-e^t(1-p)]^x \\ &= p^r e^{tr}(1-(1-p)e^t)^{-r} = \frac{(pe^t)^r}{(1-qe^t)^r} \end{aligned}$$

with $q = 1-p$. $\square$

4.4. Remark. A random variable following the Pascal distribution with parameters r and p is the sum of r independent geometric random variables with parameter p.

Counting Successes in a Continuous Environment

Binomial distribution: counts successes in n trials.

Now: count successes in a continuous interval $[a, b] \subset \mathbb{R}$.

Examples:

- number of earthquakes in a century;
- number of child births in a day;
- number of bacteria in a unit volume of water.

We will talk about ***arrivals in a time interval*** $[0, t]$ for some $t > 0$. The number of arrivals will be denoted by X_t.

Assumptions:

(i) ***Independence:*** If the intervals $T_1, T_2 \subset [0, t]$ do not overlap (except perhaps at one point), then the numbers of arrivals in these intervals are independent of each other.

(ii) ***Constant rate of arrivals.***

Rate of Arrivals (Heuristic Postulates)

Assumption: There exists a number $\lambda > 0$ (arrival rate) such that for any small time interval of size Δt the following postulates are satisfied:

(i) The probability that exactly one arrival will occur in an interval of width Δt is approximately $\lambda \cdot \Delta t$.

(ii) The probability that exactly zero arrivals will occur in the interval is approximately $1 - \lambda \cdot \Delta t$.

(iii) The probability that two or more arrivals ocur in the interval is approximately zero (very small).

Wanted: a more precise (mathematical) expression of these principles.

Rate of Arrivals (Heuristic Postulates)

4.5. Example. If a hospital ward experiences, on average, about 12 child births per day, spread completely randomly throughout 24 hours, then in any given 10-minute period

(i) The probability that exactly one child birth will occur is approximately

$$\lambda \cdot \Delta t = \frac{12}{24\,\text{hours}} \cdot \frac{1}{6}\,\text{hours} = \frac{1}{12}.$$

(ii) The probability that exactly zero births will occur is approximately

$$1 - \lambda \cdot \Delta t = \frac{11}{12}$$

(iii) The probability that two or more births occur is approximately zero (very small).

“Little-o” Notation

We denote by $o(t)$ any function f such that

$$\lim_{t \to 0} \frac{f(t)}{t} = 0.$$

Hence $o(t)$ does not denote a particular function, rather a class of functions. For example,

- $t^2 = o(t)$,
- $(1+t)^2 = 1 + 2t + o(t)$,
- $\sin t = t + o(t)$.

In particular,

- $o(t) + o(t) = o(t)$,
- $t^n \cdot o(t) = o(t)$ for all $n \in \mathbb{N}$,
- $o(t) \cdot o(t) = o(t)$.

Rate of Arrivals (Precise Postulates)

(i) The probability that ***exactly one arrival*** will occur in an interval of width Δt is

$$\lambda \cdot \Delta t + o(\Delta t).$$

(ii) The probability that ***exactly zero arrivals*** will occur in the interval is

$$1 - \lambda \cdot \Delta t + o(\Delta t).$$

(iii) The probability that ***two or more arrivals*** occur in the interval is

$$o(\Delta t).$$

We denote by X_t the number of arrivals in the interval $[0, t]$ and write

$$P[X_t = x] =: p_x(t) \qquad \text{with } x = 0, 1, 2, 3, \ldots$$

Probability of Zero Arrivals

Consider the time interval

$$[0, t + \Delta t] = [0, t] \cup [t, \Delta t]$$

Due to independence of non-overlapping time intervals,

$$\begin{aligned} p_0(t + \Delta t) &= P[0 \text{ arrivals in } [0, t + \Delta t]] \\ &= P[0 \text{ arrivals in } [0, t]] \cdot P[0 \text{ arrivals in } [t, t + \Delta t]] \\ &= p_0(t) \cdot \big(1 - \lambda \Delta t + o(\Delta t)\big) \end{aligned}$$

It follows that

$$-\lambda p_0(t) = \frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} + \frac{o(\Delta t)}{\Delta t}.$$

Probability of Zero Arrivals

We can take the limit as $\Delta t \to 0$ on both sides. Then we have

$$-\lambda p_0(t) = \lim_{\Delta t \to 0} \frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = p_0'(t).$$

This is a linear, homogeneous ordinary differential equation for p_0.

Question. What is a suitable initial value for this ODE?

(a) 0.

(b) 1.

(c) Some other value.

Probability of Several Arrivals

Now let $x > 0$. Then

$$
\begin{aligned}
p_x(t+\Delta t) &= P[x \text{ arrivals in } [0, t+\Delta t]] \\
&= \sum_{y=0}^{x} P[x-y \text{ arrivals in } [0,t]] \cdot P[y \text{ arrivals in } [t, t+\Delta t]] \\
&= p_x(t) \cdot \big(1 - \lambda \Delta t + o(\Delta t)\big) + p_{x-1}(t) \cdot \big(\lambda \Delta t + o(\Delta t)\big) \\
&\quad + p_{x-2}(t) \cdot o(\Delta t) + \dots + p_0(t) \cdot o(\Delta t) \\
&= \lambda \Delta t \, p_{x-1}(t) + (1 - \lambda \Delta t) p_x(t) + o(\Delta t)
\end{aligned}
$$

so that

$$
\lambda p_{x-1}(t) - \lambda p_x(t) = \frac{p_x(t+\Delta t) - p_x(t)}{\Delta t} + \frac{o(\Delta t)}{\Delta t}.
$$

Probability of Several Arrivals

Taking the limit as $\Delta t \to 0$, we obtain

$$p_x'(t) = \lambda p_{x-1}(t) - \lambda p_x(t).$$

Together with

$$p_0' = -\lambda p_0$$

and suitable initial conditions we have a system of differential equations that can be solved inductively to determine p_0, p_1, p_2,....

The solution to these equations is

$$p_x(t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}.$$

We often define $k := \lambda t$ (“rate times interval”).

The Poisson Distribution

Siméon Poisson (1781-1840). Delpech, François Séraphim before 1840. Lithograph. File:Simeon Poisson.jpg. (2019, August 13). Wikimedia Commons, the free media repository.

4.6. Definition. Let $k \in \mathbb{R}$. A random variable (X, f_X) with

$$X: S \to \mathbb{N}$$

and density function $f_X: \mathbb{N} \to \mathbb{R}$ given by

$$f_X(x) = \frac{k^x e^{-k}}{x!}$$

is said to follow a ***Poisson distribution*** with parameter k.

The Poisson distribution describes the occurrence of events that occur at a ***constant rate*** in a ***continuous environment***.

M.G.F. and C.D.F. of the Poisson Distribution

4.7. Theorem. Let (X, f_X) be a Poisson distributed random variable with parameter k.

(i) The moment generating function of X is given by

$$m_X \colon \mathbb{R} \to \mathbb{R}, \qquad m_X(t) = e^{k(e^t - 1)}.$$

(ii) $\mathrm{E}[X] = k$.

(iii) $\mathrm{Var}[X] = k$.

The cumulative distribution function

$$F(x) = P[X \le x] = \sum_{y=0}^{\lfloor x \rfloor} \frac{e^{-k} k^y}{y!}$$

is found in Table II of Appendix A of the text book.

The Poisson Distribution

4.8. Example. A healthy individual may have an average white blood cell count of as low as $4500/\,\text{mm}^3$ of blood. To detect a white-cell deficiency, a $0.001\,\text{mm}^3$ drop of blood is taken and the number X of white blood cells is found.

If at most one is found, is there evidence of a white-cell deficiency?

Here the volume of blood (in mm^3) takes the role of the continuous variable and each observed white cell counts as an "arrival."

The number of arrivals per unit volume is $\lambda = 4500$, the volume under consideration is $s = 0.001$. Hence we have a Poisson-distributed random variable with parameter

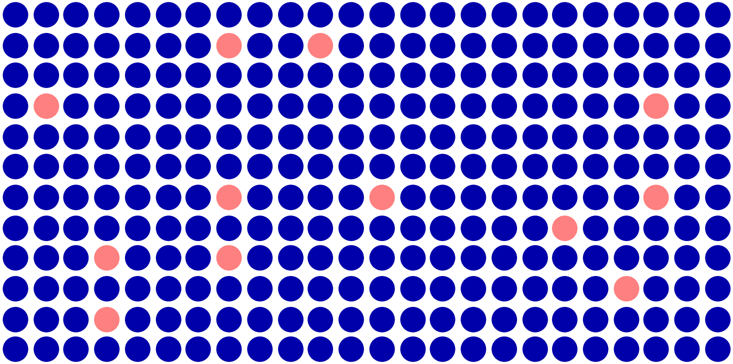
$$k = \lambda s = 4.5.$$

The expected value is $\mathrm{E}[X] = k = 4.5$. Furthermore,

$$P[X \le 1] = \sum_{x=0}^{1} \frac{e^{-4.5} 4.5^x}{x!} = 0.061.$$

Approximating the Binomial Distribution

Suppose a binomial random variable is given with ***large*** n. Then we can approximate the density function using that of a Poisson distribution:



Within many trials (represented by each disk) the successes (orange disks) occur as within a continuum of trials.

Approximating the Binomial Distribution

Mathematically, this is actually a limit statement: If $n \to \infty$ while $n \cdot p =: \lambda$ remains constant,

$$\binom{n}{m} p^m (1-p)^{n-m} \qquad \xrightarrow[n \cdot p = k]{n \to \infty} \qquad \frac{k^m}{m!} e^{-k}$$

Therefore, we can approximate the binomial distribution by a Poisson distribution with parameter

$$k = pn$$

if n is large.

In general, one does this if $p < 0.1$. The smaller p and the larger n are, the better the approximation.

Approximating the Binomial Distribution

4.9. Example. A typical aircraft wing has 40, 000 rivets. Suppose that the probability of a given rivet being defective is 0.001. What is the probability that not more than fifty rivets are defective?

The actual probability is

$$P[X \le 50] = \sum_{x=0}^{50} \binom{40\,000}{x} (0.001)^x (0.999)^{40\,000-x}$$
$$= 0.94746.$$

Using the Poisson approximation, $k = 40\,000 \cdot 0.001 = 40$ and

$$P[X \le 50] \approx \sum_{x=0}^{50} e^{-40} \frac{40^x}{x!} = 0.94737.$$