Poisson Distributions

Euler's number and natural exponential.

$$\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^{a}$$ ← It can be positive or negative

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = e^{-1} = \frac{1}{e}$$

The number

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = 2.7182\cdots = e^{1}$$

is known as the base of the natural logarithm (or, called Euler's number). The exponential function $f(x) = e^x$ is associated with the Taylor series

$$\boxed{e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}.} \tag{4.1}$$

Poisson distribution.

The **Poisson distribution** with parameter λ ($\lambda > 0$) has the frequency function

constant

$$P(X=k) = p(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

By applying (4.1) we can immediately see the following:

$$\boxed{1 \overset{?}{=} \sum_{k=0}^{\infty} p(k)} = e^{-\lambda} \underbrace{\boxed{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}}_{=e^{\lambda}} = e^{-\lambda}e^{\lambda} = 1.$$

A Poisson random variable X represents the number of successes when there are a large (and usually unknown) number of independent trials.

and usually small probability of success

Example

Random variable X

Suppose that the number of typographical errors per page has a Poisson distribution with parameter $\lambda = 0.5$.

1. Find the probability that there are no errors on a single page.
2. Find the probability that there is at least one error on a single page.

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$P(X=0) = p(0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda} = e^{-0.5}$$

$$P(X \geq 1) = 1 - P(X=0) = 1 - p(0) = 1 - e^{-0.5}$$

Example

Suppose that the number of typographical errors per page has a Poisson distribution with parameter $\lambda = 0.5$.

1. Find the probability that there are no errors on a single page.
2. Find the probability that there is at least one error on a single page.

1. $P(X = 0) = p(0) = e^{-0.5} \approx 0.607$
2. $P(X \geq 1) = 1 - p(0) = 1 - e^{-0.5} \approx 0.393$

Poisson approximation to binomial distribution.

$p(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

$0 < p = \frac{\lambda}{n} < 1$

The Poisson distribution can be used as an approximation for a binomial distribution with parameter (n, p) when n is large and p is small enough so that np is a moderate number λ. Let $\boxed{\lambda = np}$. Then the binomial frequency function becomes

It is fixed

$$
\begin{aligned}
p(k) &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \cdot \cancel{\frac{n!}{(n-k)!n^k}} \cdot \boxed{\left(1-\frac{\lambda}{n}\right)^n} \cdot \left(1-\frac{\lambda}{n}\right)^{-k} \\
&\to \frac{\lambda^k}{k!} \cdot 1 \cdot \boxed{e^{-\lambda}} \cdot 1 = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{as } n \to \infty.
\end{aligned}
$$

$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1$

$\underbrace{n(n-1)\times\cdots\times(n-k+1)}_{k \text{ terms}}$

$\approx n^k$ if n is very large

$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{\boxed{-\lambda}}$

Big picture formula

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}$$

if n is large and p is small and $\boxed{\lambda = np}$.

Example

Suppose that the a microchip is defective with probability 0.02. Find the probability that a sample of 100 microchips will contain at most 1 defective microchip.

X = # of defective items out of 100

$$P(X \le 1) = P(X=0) + P(X=1) = p(0) + p(1)$$

$$= \binom{100}{0} p^0 (1-p)^{100} + \binom{100}{1} p^1 (1-p)^{99}$$

What is p? $p = 0.02$

$$\approx e^{-\lambda} \frac{\lambda^0}{0!} + e^{-\lambda} \frac{\lambda^1}{1!}$$

What is λ? $\lambda = np = (100)(0.02) = 2$

Example

Suppose that the a microchip is defective with probability 0.02. Find the probability that a sample of 100 microchips will contain at most 1 defective microchip.

The number X of defective microchips has a binomial distribution with $n = 100$ and $p = 0.02$. Thus, we obtain

$$\begin{aligned} P(X \leq 1) &= p(0) + p(1) \\ &= \binom{100}{0}(0.98)^{100} + \binom{100}{1}(0.02)(0.98)^{99} \approx 0.403 \end{aligned}$$

Approximately X has a Poisson distribution with $\lambda = (100)(0.02) = 2$. We can calculate

$$P(X \leq 1) = p(0) + p(1) = e^{-2} + (2)(e^{-2}) \approx 0.406$$

Expectation and variance of Poisson distribution.

X = Binomial r.v. with (n, p) ; Y = Poisson r.v. with $\lambda = np \Rightarrow p = \frac{\lambda}{n}$

$E[X] = np \xleftrightarrow[\text{equal}]{\text{approximately}} E[Y] = \lambda$

$Var(X) = np(1-p) \longleftrightarrow Var(Y) \approx \lim_{n\to\infty} Var(X_n) = n\left(\frac{\lambda}{n}\right)\left[\underbrace{1-\left(\frac{\lambda}{n}\right)}_{1 \text{ as } n\to\infty}\right] \xrightarrow{n\to\infty} n\left(\frac{\lambda}{n}\right) = \lambda$

Let X_n, $n = 1, 2, \ldots$, be a sequence of binomial random variables with parameter $(n, \lambda/n)$. Then we regard the limit of the sequence as a Poisson random variable Y, and use the limit to find $E[Y]$ and $\mathrm{Var}(Y)$. Thus, we obtain

$\lambda = np$

$$E[X_n] = \lambda \quad \to \quad E[Y] = \lambda \quad \text{as } n \to \infty;$$

$$\mathrm{Var}(X_n) = \lambda \underbrace{\left(1 - \frac{\lambda}{n}\right)}_{1} \quad \to \quad \mathrm{Var}(Y) = \lambda \quad \text{as } n \to \infty.$$

Sum of independent Poisson random variables.

Theorem

The sum of independent Poisson random variables is a Poisson random variable: If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, then $Z = X + Y$ is distributed as the Poisson distribution with parameter $\lambda_1 + \lambda_2$. $= \lambda$

Binomial distribution has the same property:

X, Y are independent and binomial r.v's then so is X+Y.

$E[Z] = \lambda$ is the parameter for Z.

$$\begin{aligned} E[Z] &= E[X+Y] \\ &= E[X] + E[Y] \\ &= \lambda_1 + \lambda_2 \end{aligned}$$

Theorem

The sum of independent Poisson random variables is a Poisson random variable: If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, then $Z = X + Y$ is distributed as the Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Proof

Let X_n and Y_n be independent binomial random variables with respective parameter (k_n, p_n) and (l_n, p_n). Then the sum $X_n + Y_n$ of the random variables has the binomial distribution with parameter $(k_n + l_n, p_n)$. By letting $k_n p_n \to \lambda_1$ and $l_n p_n \to \lambda_2$ with $k_n, l_n \to \infty$ and $p_n \to 0$, the respective limits of X_n and Y_n are Poisson random variables X and Y with respective parameters λ_1 and λ_2. Meanwhile, the limit of $X_n + Y_n$ is the sum of the random variables X and Y, and has a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

$$\lambda = \lim_{n\to\infty} (k_n + l_n) p_n = \lim_{n\to\infty} k_n p_n + \lim_{n\to\infty} l_n p_n = \lambda_1 + \lambda_2$$

Assignment No.4

Supplementary Readings.

SS: Murray R. Spiegel, John Schiller, and R. Alu Srinivasan, *Probability and Statistics 4th ed.* McGraw-Hill.
Chapter 4: Poisson Distribution, Some Properties of Poisson Distribution, Relation between Binomial and Poisson Distributions.

TH: Elliot A. Tanis and Robert V. Hogg, *A Brief Course in Mathematical Statistics*. Prentice Hall, NJ.
Section 2.3.

WM: Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye, *Probability & Statistics for Engineers & Scientists*, 9th ed. Prentice Hall, NJ.
Section 5.5.

Problem

The probability that a man in a certain age group dies in the next four years is $p = 0.05$. Suppose that we observe 20 men.

1. *Find the probability that two of them or fewer die in the next four years.*
2. *Find an approximation by using a Poisson distribution.*

Problem

The probability of being dealt a royal straight flush (ace, king, queen, jack, and ten of the same suit) in poker is about 1.5×10^{-6}*. Suppose that an avid poker player sees* 100 *hands a week,* 52 *weeks a year, for* 20 *years.*

1. *What is the probability that she never sees a royal straight flush dealt?*
2. *What is the probability that she sees at least two royal straight flushes dealt?*

Problem

Professor Rice was told that he has only 1 chance in 10,000 of being trapped in a much-maligned elevator in the mathematics building. Assume that the outcomes on all the days are mutually independent. If he goes to work 5 days a week, 52 weeks a year, for 10 years and always rides the elevator up to his office when he first arrives.

1. *What is the probability that he will never be trapped?*
2. *What is the probability that he will be trapped once?*
3. *What is the probability that he will be trapped twice?*

Problem

Suppose that a rare disease has an incidence of 1 *in* 1000. *and that members of the population are affected independently. Let* X *be the number of cases in a population of* $100,000$.

1. *What is the average number of cases?*
2. *Find the frequency function.*

Problem

Suppose that in a city the number of suicides can be approximated by a Poisson random variable with $\lambda = 0.33$ per month.

1. *What is the probability of two suicides in January?*
2. *What is the distribution for the number of suicides per year?*
3. *What is the average number of suicides in one year?*
4. *What is the probability of two suicides in one year?*

Computer Project.

Improvised explosive devices (IED's) are planted in a city block of 25 square miles (exactly a 5×5-mile square). For every week the number of newly found IED's in this block is approximated by a Poisson distribution with $\lambda = 10$, and the locations of these IED's are scattered uniformly over the city block. Suppose that you are responsible for sweeping and clearing the south-west corner of 2×2-mile square block every week.

1. Develop the simulation to count the number of IED's in your area, and run it for 50 weeks to generate a random sample. Draw the relative frequency histogram from the simulation.
2. Calculate `mean()` and `var()` of the simulation. How many IED's in average do you expect to discover in your area for a particular week?

Computer Project, continued.

Let X be the number of IED's planted in the area of 2×2-mile square block in a week.

1. Discuss what distribution best approximates the random variable X, and how you can identify the parameter value associated with this distribution.
2. Continue from the previous question, what is the chance that you find more than one IED in a particular week? $P(X > 1) = 1 - \overbrace{P(X \leq 1)}^{P(0)+P(1)}$
3. If the number of IED's per week in the 5×5-mile square has a Poisson distribution with $\lambda = 20$, how many IED's do you expect to discover in the area of 2×2-mile square block every week? What is the chance that you find more than one IED in a particular week?

Answers

Problem 1.

1. The number X of men dying in the next four years has a binomial distribution with $n = 20$ and $p = 0.05$.

$$P(X \leq 2) = p(0) + p(1) + p(2)$$
$$= \binom{20}{0}(0.95)^{20} + \binom{20}{1}(0.05)(0.95)^{19} + \binom{20}{2}(0.05)^2(0.95)^{18}$$
$$\approx 0.925$$

2. The random variable X has approximately a Poisson distribution with $\lambda = (0.05)(20) = 1$.

$$P(X \leq 2) = e^{-1} + e^{-1} + \frac{1}{2}e^{-1} = \frac{5}{2}e^{-1} \approx 0.920$$

Problem 2.

1. The number X of royal straight flushes has a Poisson distribution with

$$\lambda = 1.5 \times 10^{-6} \times (100 \times 52 \times 20) = 0.156$$

Thus, we obtain

$$P(X = 0) = p(0) = e^{-0.156} \approx 0.856$$

2. $P(X \geq 2) = 1 - p(0) - p(1) = 1 - e^{-0.156} - (0.156)(e^{-0.156}) = 1 - (1.156)(e^{-0.156}) \approx 0.011$

Problem 3.

1. The number X of misfortunes for Professor Rice being trapped has a Poisson distribution with

$$\lambda = 10^{-4} \times (5 \times 52 \times 10) = 0.26$$

Thus, we obtain

$$P(X = 0) = p(0) = e^{-0.26} \approx 0.771$$

2. $P(X = 1) = p(1) = (0.26)(e^{-0.26}) \approx 0.20$
3. $P(X = 2) = p(2) = \frac{(0.26)^2}{2}(e^{-0.26}) \approx 0.026$

Problem 4.

1. The number X of cases has a Poisson distribution with

$$\lambda = 10^{-3} \times 100,000 = 100$$

   Thus, we obtain $E[X] = \lambda = 100$.

2. $p(k) = (e^{-100})\dfrac{100^k}{k!}$ for $k = 0, 1, \ldots$.

Problem 5.

1. Let X_1 be the number of suicides in January. Then we have

$$P(X_1 = 2) = \frac{(0.33)^2}{2}(e^{-0.33}) \approx 0.039$$

2. Let $X_1, \ldots, X_{12}$ be the number of suicides in i-th month. Then

$$Y = X_1 + X_2 + \cdots + X_{12}$$

   represents the number of suicides in one year, and it has a poisson distribution.

3. $E[Y] = E[X_1] + E[X_2] + \cdots + E[X_{12}] = (12)(0.33) = 3.96.$
4. Since $\lambda = E[Y] = 3.96$ is the parameter value for the Poisson random variable Y, we obtain

$$P(Y = 2) = \frac{(3.96)^2}{2}(e^{-3.96}) \approx 0.149$$