Section 3.2. The Poisson Distribution

Note. In this section, we define the Poisson distribution and see how it is related to the number of occurrences of an event in various time intervals. We give some properties (such as the moment generating function) and examples.

Definition. A random variable X with probability mass function

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{for } x \in \{0, 1, 2, \ldots\} \\ 0 & \text{elsewhere} \end{cases}$$

where $\lambda > 0$, is said to have a *Poisson distribution* with parameter λ.

Note. Since $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ for all $x \in \mathbb{R}$, then summing over the support of p gives

$$\sum_{n=0}^{\infty} p(n) = \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = (e^{-\lambda})(e^{\lambda}) = 1,$$

as needed.

Note. Siméon Poisson introduced this distribution in 1837 in *Recherches sur la probabilité des jugements en matière criminelle et matière civile* [Research on the likelihood of judgments in criminal and civil matters] (Paris: Bachelier, Imprimeur-Libraire pour lex mathématiques, La Physique, #55, 1837); a copy appears online from the Gallica online digital library.

Siméon Denis Poisson, June 21, 1781–April 25, 1840

Image from the MacTutor History of Mathematics Archive page on Poisson

Note 3.2.A. We consider a process that occurs repeatedly through time. Let X_t denote the number of occurrences of the process over the interval $(0, t]$. The range of X_t is $\{0, 1, 2, \ldots\}$. For $k \in \{0, 1, 2, \ldots\}$ and real number $t > 0$, define the probability mass function of X_t as $P(X_t = k) = g(k, t)$ (and we take $g(0, 0) = 1$). We assume

Axiom 1. $g(1, h) = \lambda h + o(h)$ for a constant $\lambda > 0$,

Axiom 1(a). $P(X_{t+h} = n + 1) - P(X_t = n) = g(n + 1, t + h) - g(n, t) = \lambda h + o(h)$ for constant λ and for all $n \in \mathbb{N}$,

Axiom 2. $\sum_{k=2}^{\infty} g(k, h) = o(h)$, and

Axiom 2(a). $\sum_{k=2}^{\infty} P(X_{t+h} = n + k) - P(X_t = n) = \sum_{k=2}^{\infty} g(n + k, t + h) - g(n, t) = o(h)$,

Axiom 3. the number of occurrences in nonoverlapping intervals are independent of one another.

Hogg, McKean, and Craig do not state Axioms 1(a) and 2(a) explicitly, but we need these ideas in our derivation below. They hint that these are necessary by stating on in Section 3.3 that "the Axioms (1) and (2) of the Poisson process only depend on λ and the length of the interval; in particular, they do not depend on the endpoints of the interval." See page 178. So we conclude from Axioms 1 and 1(a) that the probability of one occurrence of the process in any interval $(t, t+h]$ of length h is $\lambda h + o(h)$. We conclude from Axioms 2 and 2(a) that the probability of more than one occurrence of the process in any interval $(t, t+h]$ of length h is $o(h)$. The "little oh" notation $o(h)$ means that $o(h)$ is some function of h satisfying $\lim_{h\to 0} o(h)/h = 0$ (see my online Calculus 2 [MATH 1920] notes on 7.4. Relative Rates of Growth). From (1) and (2) we have

$$P(X_t \geq 1) = \sum_{k=1}^{\infty} g(k,t) = g(1,t) + \sum_{k=1}^{\infty} g(k,t) = \lambda t + o(t) + o(t).$$

So in a small interval of time, say $(0, h]$, $P(X_h \geq 1) = \lambda h + 2o(h)$. So for h "small," $P(X_h = 1)$ is approximately proportional to λ, $P(X_n > 1)$ is approximately 0, and hence $P(X_h = 0)$ is approximately $1 - \lambda h$. We now show that X_t has a Poisson distribution.

Note 3.2.B. Now no events occur in $(0, t+h]$ if and only if no event occur in $(0, t]$ and no events occur in $(t, t+h]$. As just discussed, the probability that no events occur in $(0, h]$ is $1 - \lambda h - 2o(h)$. Since "$o(h)$" simply represents *some* function of h such that $\lim_{h\to 0} o(h)/h = 0$, then any scalar multiple of $o(h)$ can be "absorbed" into the "$o(h)$" term. Hence the probability that no events occur in $(0, h]$ is $1 - \lambda h + o(h)$, as the text states (see page 168). Since intervals $(0, t]$ and $(t, t + h]$ do not overlap, by (3) (and the Multiplicative Rule) we have

$$g(0, t + h) = g(0, t)(1 - \lambda h + o(h)).$$

Therefore,

$$\lim_{h\to 0} \frac{g(0, t + h) - g(0, t)}{h} = \lim_{h\to 0} \frac{g(0, t)(1 - \lambda h + o(h)) - g(0, t)}{h}$$

$$= \lim_{h\to 0} \left(-\lambda g(0, t) + g(0, t)\frac{o(h)}{h}\right) = -\lambda g(0, t) + g(0, t) \lim_{h\to 0} \frac{o(h)}{h} = -\lambda g(0, t).$$

So with $g(0, t)$ as a function of t, it is differentiable and

$$\frac{d}{dt}[g(0, t)] = -\lambda g(0, t), \text{ or } \frac{g'(0, t)}{g(0, t)} = -\lambda$$

where the prime indicates differentiation with respect to t. So we have by integrating and using u-substitution with $u = g(0, t)$ we have

$$\int \frac{g'(0, t)}{g(0, t)}\, dt = \int -\lambda\, dt \text{ or } \lim |g(0, t)| = -\lambda t + c$$

for some $c \in \mathbb{R}$. Since $g(0, t) = P(X_t = 0) > 0$ then we have $|g(0, t)| = g(0, t)$ or $\ln g(0, t) = -\lambda + c$ or, by exponentiating, $g(0, t) = e^{-\lambda t + c}$. Since $g(0, 0) = 1$, then we see that $c = 0$, so that $g(0, t) = e^{-\lambda t}$.

Note 3.2.C. Next, consider $g(k, t)$ for $k \in \mathbb{Z}$, $k \geq 0$. We now show using mathematical induction that, under Axioms 1–3, $g(k, t) = e^{-\lambda t}(\lambda t)^k/k!$. The base case of $k = 0$ is established in Note 3.2.B. So suppose for given $k \geq 1$ we have $g(j, t) = \dfrac{e^{-\lambda t}(\lambda t)^j}{j!}$ for $j = k$ (the induction hypothesis). We address the case

$j = k+1$ and the value of $g(k+1,t)$ by first considering $g(k+1,t+h)$. In order to have $k+1$ occurrences in $(0,t+h]$, we must have $k+1-i$ occurrences in $(0,t]$ and i occurrences in $(t,t+h]$ for some i with $0 \le i \le k+1$. Now the probability that there are $k+1-i$ occurrences in $(0,t]$ is (by definition of probability mass function g of random variable X_t) $P(X_t = k+1-i) = g(k+1-i,t)$. The probability that there are i occurrences in $(t,t+h]$ is: (1) $1-(\lambda h + 2o(h))$ if $i=0$ by Axioms 1(a) and 2(a), (2) $\lambda h + o(h)$ if $i=1$ by Axiom 1(a), and (3) $o(h)$ if $i \ge 2$ by Axiom 2(a). These are independent events by Axiom 3, so summing over $i=0$ to $i=k$ give that the probability of $k+1$ occurrences of the process in the interval $(0,t+h]$ is

$$\begin{aligned} g(k+1,t+h) &= g(k+1,t)(1-(\lambda h+2o(h))+g(k,t)(\lambda h+o(h)) \\ &\quad +g(k-1,t)(o(h))+\cdots+g(1,t)(o(h))+g(0,t)(o(h)) \\ &= g(k+1,)(1-\lambda h-2o(h))+g(k,t)(\lambda h+o(h)) \\ &\quad +o(h)\sum_{i=2}^{k+1} g(k+1-i,t). \end{aligned}$$

Therefore

$$\frac{g(k+1,t+h)=g(k+1,t)}{h}$$

$$\begin{aligned} &= \left\{ g(k+1,t)(1-\lambda h-2o(h))+g(k,t)(\lambda h+o(h)) \right. \\ &\quad \left. +o(h)\sum_{i=2}^{k+1} g(k+1-i,t) - g(k+1,t) \right\} \Big/ h \\ &= \left\{ \lambda h g(k+1,t) - 2g(k+1,t)(o(h)) + \lambda h g(k,t) + g(k,t)(o(h)) \right. \\ &\quad \left. +o(h)\sum_{i=2}^{k+1} g(k+1-i,t) \right\} \Big/ h \end{aligned}$$

$$= \ -\lambda g(k+1,t) + \lambda g(k,t) + \frac{o(h)}{h}\left(-2g(k+1,t) + g(k,t) + \sum_{i=2}^{k+1} g(k+1-i,t)\right).$$

Letting $h \to 0$, since $o(h)/h \to 0$, we have

$$\frac{d}{dt}[g(k+1,t)] = -\lambda g(k+1,t) + \lambda g(k,t) = -\lambda g(k+1,t) + e^{-\lambda t}(\lambda t)^k/k!,$$

by the induction hypotheses that $g(j,t) = e^{-\lambda t}(\lambda t)^j/j!$ for $j = k$. So with $y = g(k+1,t)$, we need to solve the linear differential equation

$$y' + \lambda y = \lambda e^{-\lambda t}(\lambda t)^k/k!.$$

Now by "a theorem in differential equations," the linear differential equation $y' + P(t)y = Q(t)$ has the one parameter family of solutions of

$$y = e^{-\int P(t)\,dt}\left\{\int Q(t)e^{\int P(t)\,dt}\,dt + C\right\}.$$

Here, constant C is the "one parameter" and indefinite integral notation is meant to indicate *any antiderivative* of the integrand. For details, see my online notes for Differential Equations (MATH 2120) on Section 2.3. Linear Equations and Bernoulli Equations; notice Theorem 2.4. With $P(t) = \lambda$ and $Q(t) = \lambda e^{-\lambda t}(\lambda t)^k/k!$ we have

$$\begin{aligned} y &= g(k+1,t) = e^{-\int \lambda\,dt}\left\{\int \lambda e^{-\lambda t}(\lambda t)^k/k!(e^{\lambda t}\,dt + C\right\} \\ &= e^{-\lambda t}\left\{\int \lambda^{k+1}t^k/k!\,dt + C\right\} = e^{-\lambda t}\left\{\lambda^{k+1}t^{k+1}/(k+1)! + C\right\}. \end{aligned}$$

For the boundary condition $g(k+1,0) = 0$, we see that the parameter C must be 0 so that

$$g(k+1,t) = e^{-\lambda t}(\lambda t)^{k+1}/(k+1)!.$$

So the claim holds for $j = k+1$ and hence $g(k,t) = e^{-\lambda t}(\lambda t)^k/k!$ for all $k \in \mathbb{N}$ (as claimed).

Note. We have by Notes 3.2.A, 3.2.B, and 3.2.C that any process that occurs repeatedly such that Axioms 1, 1(a), 2, 2(a), and 3 are satisfied, must have a Poisson distribution.

Note. If X has a Poisson distribution with parameter λ then the moment generating function of X is

$$M(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} p(x) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)} \text{ for } t \in \mathbb{R}.$$

Now $M'(t) = e^{\lambda(e^t-1)}[\lambda e^t]$ and

$$M''(t) = [e^{\lambda * e^t - 1)}[\lambda e^t]](\lambda e^t) + (e^{\lambda(e^t-1)})[\lambda e^t] = \lambda e^{\lambda(e^t-1)} e^t(\lambda e^t+1) = \lambda e^{\lambda(e^t-1)+t}(\lambda e^t+1).$$

So by Note 1.9.B, $\mu = M'(0) = \lambda$ and $\sigma^2 = M''(0) - (M'(0))^2 = (\lambda^2 + \lambda) - (\lambda) = \lambda^2$. So a Poisson distribution has $\mu = \sigma^2 = \lambda > 0$.

Note. The next result shows that a sum of Poisson random variables is itself Poisson.

Theorem 3.2.1. Suppose $X_1, X_2, \ldots, X_n$ are independent random variables and suppose X_i has a Poisson distribution with parameters λ_i. Then $Y = \sum_{i=1}^n X_i$ has a Poisson distribution with parameter $\sum_{i=1}^n \lambda_i$.

Exercise 3.2.13. On the average, a grocer sells three of a certain article per week. How many of these should he have in stock so that the chance of his running out within a week is less than 0.01? Assume a Poisson distribution.

Revised: 7/11/2021