International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 20(2), 2020, pp 37-56

The Poisson Distribution and Its Convergence to the Normal Distribution

Subhash Bagui[1*] and K. L. Mehra[2]

[1]Department of Mathematics and Statistics, The University of West Florida, Pensacola, FL USA

[2]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada

[*]Correspondence should addressed to S. Bagui

(Email: sbagui@uwf.edu)

[Received July 10, 2020; Accepted November 01, 2020]

Abstract

In this paper, we have given a historical background for the Poisson distribution and have described some of its applications in the early days. We have also shown how Binomial and Negative-binomial distributions can be approximated by the Poisson distribution. Finally, we present four different methods of proof of the convergence of Poisson to the normal distribution. These proofs can be gainfully discussed in a senior classroom setting. The note may also serve as a useful pedagogical reference article in senior-level probability and mathematical statistics courses.

Keywords: Poisson distribution, Binomial distribution, Negative-binomial distribution, Exponential distribution, BMP Ratio method, Stirling's approximations, Moment generating functions, Central limit theorem.

AMS Classification: 62E15, 62E99.

1. Introduction

The Poisson distribution is a widely used discrete distribution in statistics. It has been shown to be highly effective in modeling the occurrences of rare events in a variety of real-world random phenomena. For example, it is applicable in modeling the number of severe traffic accidents that occur monthly in a town, the number of severe earthquakes occurring in a year, the number of life indemnity claims filed yearly with an insurance company, the number of alpha particles emitted from a piece of radioactive material in a fixed time interval, the number of mail carriers bitten by dogs per month and so on and on. All these scenarios deal with the total number of occurrences in a series of independent Bernoulli type experiments, each having a very small probability of occurrence. This total number, say X_λ, a random variable (r.v.) with possible values $0, 1, 2, \cdots$ and depending on a parameter λ ($\lambda > 0$), is said to have a Poisson distribution if its probability mass function (pmf) f_λ is given by

$$f_\lambda(x) = P(X_\lambda = x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \text{for } x = 0,1,2,\cdots, \tag{1.1}$$

where $e = \sum_0^\infty \frac{1}{n!} = 2.71828\ldots$ is the Euler number. The expected value (mean) and the variance of this discrete r.v. X_λ are $\mu = E(X_\lambda) = \lambda$ and $\sigma^2 = Var(X) = \lambda$, respectively. If $\mu_r(\lambda) = E(X_\lambda - \lambda)^r$ denotes its rth ($r \geq 1$) central moment, its third-order central moment is also λ and the skewness and kurtosis coefficients are $\gamma_1 = [\mu_3(\lambda)/\mu_2^{3/2}(\lambda)] = 1/\sqrt{\lambda}$ and $\gamma_2 = [\mu_4(\lambda)/\mu_2^2(\lambda) - 3] = 1/\lambda$, respectively. So, it is a positively skewed leptokurtic distribution. The recursion relation $f_\lambda(x) = (\lambda/x) f_\lambda(x-1)$ $x = 0,1,2,\cdots$ calculates all probability values of this pmf f_λ readily. Similarly, a straight forward recursion relation for the central moments $\mu_{r+1} = \lambda[r\mu_{r-1} + \frac{d\mu_r}{d\lambda}]$ (see [10], p. 158) helps to determine all central moments recursively. Likewise, if we let $\alpha_r(\lambda) = E(X_\lambda^r)$, $r \geq 1$, denote its rth integral moment about the origin, the recursive relation $\alpha_{r+1} = \lambda[\alpha_r + \frac{d\alpha_r}{d\lambda}]$ aids the easy calculation of all moments $\alpha_r(\lambda)$ (see [15], p.179). Further, the cumulative distribution function (cdf) of the Poisson r.v. X_λ can be expressed and easily evaluated in terms of the incomplete gamma function, viz.,

$$F_{X_\lambda}(x) = P(X_\lambda \leq x) = \sum_{k=0}^{x}\frac{e^{-\lambda}\lambda^k}{k!} = \frac{1}{\Gamma(x+1)}\int_\lambda^\infty e^{-y}y^x dy, \tag{1.2}$$

where $\Gamma(x+1) = x!$ for $x = 0,1,\cdots$ (see [15], p. 180). The Poisson probability distribution was named after an eminent French mathematician, physicist, and academic administrator Simeon Denis Poisson (1781-1840). It was so named due to his introduction of this distribution as a model for "laws of small numbers". According to a letter by the mathematician Abel in 1826 to his friend, Poisson was a man who knew "how to behave with a great deal of dignity" (see [12], p.129).

One of the early applications of this model was to the number of Prussian cavalry deaths by kicking horses in the Prussian army. One of Poisson's many interests was the application of probability theory to the administration of law in criminal trials and the like. This is evident from his book, titled "*Recherches sur la probabilitte des jugements en matiere criminelle et en matiere civile*", which was published in 1837 (see [13]). The French mathematician S. D. Poisson did not recognize the potential of the vast real-world applications of this distribution. It was L. von Bortkiewicz in 1898, a German professor, who first understood and explained the importance of Poisson distribution in his book titled, "*Das Gesetz der Lleinen Zahlen*" (see [8], [14]) that transformed Poisson's "limit" to Poisson "distribution". One of the striking examples in this book was the modeling, as mentioned above, of the number of Prussian cavalry deaths by kicking horses using Poisson distribution (see [18]). Another great application of this distribution was the study of hits of flying bombs in London during the Second World War. During this war, the British authorities were anxious to know if these bombs were aimed at particular targets or were simply dropped at random. The modeling of this data using Poisson distribution convinced the British military that the bombs struck at random and not with any advanced aiming ability.

There could be many more areas of applications: for example, in assessing a variety of rare occurrences in space or in detecting clusters of diseases whether they have high or low epidemicity and so on and on. For a thorough review of this distribution, see [11] and [17].

The paper is organized as follows. We have listed some elementary results in Section 2 that will be used in subsequent sections for proofs. In Section 3, we discuss Poisson approximations to the Binomial and Negative-binomial and Poisson's relationships with other distributions. In Section 4, four different methods of proof of the convergence of Poisson to the normal distribution are discussed. Some concluding remarks are included in Section 5. In the Appendix, we have provided brief derivations for the two recursion formulae above for easy calculation of all moments, and also for the formula (1.2) for Poisson P_λ cdf F_λ in terms of the 'Incomplete gamma function'.

2. Preliminaries

In this section, we list a few definitions, formulas, Lemmas and Theorems which will be used for proofs in Section 3.

Formula 2.1. The Stirling formula for approximating $n! = n(n-1)(n-2)\cdots(2)(1)$ is given by

$$n! \approx \sqrt{2\pi n}(n)^n e^{-n}\text{, for large } n. \tag{2.1}$$

This approximation ($\approx$) is in the sense that $[n!/\sqrt{2\pi n}n^n e^{-n}] \to 1$, as $n \to \infty$ (see [1]).

Formulas 2.2. The following are two well-known Tylor series expansions for natural logarithm:

$$\text{(i) } \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots = \sum_{i=1}^{\infty}(-1)^{i-1}\frac{x^i}{i} \quad \text{for } -1 \le x \le 1, \tag{2.2}$$

and

$$\text{(ii) } \ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots = -\sum_{i=1}^{\infty}\frac{x^i}{i} \quad \text{for } -1 \le x \le 1. \tag{2.3}$$

Lemma 2.1. Let $\{\delta_n, n \ge 1\}$ be a sequence of real numbers $\ni$ $\delta_n \to 0$, as $n \to \infty$. Then,

$\lim_{n\to\infty}(1 + \frac{\alpha}{n} + \frac{\delta_n}{n})^{\theta n} = e^{\theta\alpha}$, where θ and α do not depend on n.

Definition 2.1. The moment generating function (mgf) of a r.v. X is defined by $M_X(t) = E(e^{tX})$ for all $|t| < h$ with $h > 0$, provided $|E(e^{tX})| < \infty$.

If the mgf $M_X(t)$ exists, it is associated with a unique probability distribution. That is, there is a one-to-one correspondence between the classes of all probability distributions and their corresponding mgf's.

Lemma 2.2. Suppose Z is a (standard normal) $N(0,1)$ r.v. i.e., with density $f_Z(z) = \left(1/\sqrt{2\pi}\right)e^{-z^2/2}$, $-\infty < z < \infty$. Then the mgf of Z is $M_Z(t) = e^{t^2/2}$.

Theorem 2.1. (see [2]). Let X_n, $n = 1, 2, \cdots$ be a sequence of r.v.'s with well-defined mgf's $M_{X_n}(t)$, for $|t| < h$, $h > 0$. Let $M_X(t)$ denote the mgf of a r.v. X such that $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$ for all $|t| < h$; then $X_n \xrightarrow{d} X$, as $n \to \infty$.

(The notation $X_n \xrightarrow{d} X$ stands for the convergence of the distribution of r.v. X_n to the distribution of the r.v. X, as $n \to \infty$.)

Theorem 2.2. (see [1]). Let X_1, $X_2, \cdots$, X_n be a sequence of independent and identically distributed (i.i.d.) r.v.'s with a finite mean μ and a finite variance $\sigma^2 > 0$. Set $S_n = \sum_{i=1}^{n} X_i$ and $\bar{X}_n = (S_n/n)$. Then the sequence $Z_n = [(S_n - n\mu)/(\sqrt{n}\,\sigma)]$ = $[\sqrt{n}(\bar{X}_n - \mu)/\sigma] \xrightarrow{d} Z \sim N(0,1)$ the standard normal distribution, as $n \to \infty$.

The above is the simplest of Central Limit Theorems (CLT's), whose proof can be accomplished using the well-known characteristic function methodology (see [16], p.189).

Theorem 2.3. (see [1]). Let Y_1, $Y_2, \cdots, Y_n$ be a sequence of independent and identically distributed Poisson r.v.'s with parameter λ=1. If we set $X_n = \sum_{i=1}^{n} Y_i$, then X_n is a Poisson r.v. with parameter $\lambda = n$. That is, the r.v. X_n has the pmf given by

$$f_n(x) \; = P(X_n = x) = \frac{e^{-n} n^x}{x!}, \quad x = 0,1,2,\cdots \tag{2.4}$$

with mean $\mu_n = n$ and variance $\sigma_n^2 = n$.

Big O and Small o Notations

The Big O represented by the notation $f(n) = O(g(n))$ implies that the ratio $|f(n)/g(n)| < K$ for all n, as $n \to \infty$, where $K < \infty$ is some positive constant. Thus, if $|g(n)| \to 0$, as $n \to \infty$, $|f(n)| \to 0$ at the same or higher rate than that of $|g(n)|$.

The Small o represented by the notation $f(n) = o(g(n))$ implies that the ratio $|f(n)/g(n)| \to 0$, as $n \to \infty$. In this case, if $|g(n)| \to 0$, as $n \to \infty$, $|f(n)| \to 0$ necessarily at a higher rate than that of $|g(n)|$; (see [3], p. 402 or [5], p. 46).

3. Poisson Approximations and its Relationship to Other Distributions

Poisson as a limit of the binomial

In case of a large number of independent Bernoulli trials n, each having a small probability p, $0 < p < 1$, of success, it is well-known that the Binomial pmf of $X_n =$ the total number of successes, viz., $\eta_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0,1,\cdots,n$, can be approximated by the Poisson pmf $f_\lambda(k)$, $k = 0,1,\cdots$ provided as $n \to \infty$ and $p = p_n \to 0$, $\lambda_n = np_n$, $\lambda = np_n$ $\left(p_n = {}^{\lambda_n}\!/_{n}\right)$ remains constant or converges to λ;

$$\lim_{n\to\infty} \eta_n(k) = \lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!} = f_\lambda(k)\ ,\ k = 0,1,2,\cdots. \tag{3.1}$$

The proof of (3.1) is as follows: We can write

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \lim_{n\to\infty} \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n \frac{n!}{n^k (n-k)!}\left(1-\frac{\lambda}{n}\right)^{-k}. \tag{3.2}$$

The third term in (3.2) above can be simplified using

$$\frac{n!}{n^k (n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}$$

$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\cdots\left(1-\frac{k-1}{n}\right), \tag{3.2a}$$

so that for fixed k, as $n \to \infty$, in view of Lemma 2.1 and (3.2a) above, we clearly have

$$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}\ ,\ \frac{n!}{n^k (n-k)!} \to 1 \text{ and } \left(1-\frac{\lambda}{n}\right)^{-k} \to 1. \tag{3.2b}$$

The proof of (3.1) now follows from (3.2) in view of (3.2*b*). This completes the proof.

Poisson as a limit of the negative binomial

The negative binomial distribution is defined, in terms of an infinite series of independent Bernoulli trials, as the distribution of the random variable X_r that denotes the number of successes before the *r*th failure. This form of the negative binomial distribution has the probability mass function (pmf)

$$f_r^*(k) = P(X_r = k) = \binom{r+k-1}{k} p^k (1-p)^r \text{ for } k = 0,1,2,\cdots. \tag{3.3}$$

The mean and variance of this distribution are $E(X_r) = [rp/(1-p)]$ and Var(X_r) $= [rp/(1-p)^2]$. If $r \to \infty$ and $p = p_r \to 0$ with $\lambda_r = [rp/(1-p)]$ remaining constant or converging to λ; $(p = \lambda_r/(r+\lambda_r))^n$, then the negative binomial pmf (3.3) above can be approximated by the Poisson pmf f_λ. This is shown as follows:

$$\lim_{r\to\infty} f_r^*(k) = \lim_{r\to\infty}\binom{r+k-1}{k} p^k (1-p)^r \;= \lim_{r\to\infty}\frac{(r+k-1)!}{k!(r-1)!}\left(\frac{\lambda}{r+\lambda}\right)^k\left(\frac{r}{r+\lambda}\right)^r$$

$$= \lim_{r\to\infty}\frac{\lambda^k}{k!}\frac{1}{(1+\lambda/r)^r}\frac{1}{(1+\lambda/r)^k}\frac{(r+k-1)!}{r^k(r-1)!}\;; \tag{3.4}$$

the fourth term in (3.4) can be simplified as

$$\frac{(r+k-1)!}{r^k(r-1)!} = \frac{(r+k-1)(r+k-2)\cdots(r)(r-1)!}{r^k(r-1)!}$$

$$= \left(1+\frac{k-1}{r}\right)\left(1+\frac{k-2}{r}\right)\cdots\left(1+\frac{1}{r}\right), \tag{3.5}$$

so that that from Lemma 2.1 and equation (3.5), as $r\to\infty$ and k remains fixed, we clearly have

$$\left(1+\frac{\lambda}{r}\right)^r \to e^{-\lambda}\;,\; \left(1+\frac{\lambda}{r}\right)^k \to 1 \;\text{ and }\; \frac{(r+k-1)!}{r^k(r-1)!} \to 1. \tag{3.6}$$

Thus from (3.4), in view of (3.6), we may conclude that

$$\lim_{r\to\infty} f_r^*(k) = \lim_{r\to\infty}\binom{r+k-1}{k} p^k (1-p)^r \;= e^{-\lambda}\frac{\lambda^k}{k!} = f_\lambda(k)\,, \tag{3.7}$$

The proof is complete.

Relationship with Exponential Distribution

The Poisson distribution can be employed to model the arrival of customers at a counter in a given time interval. One of the underlying assumptions on which the Poisson distribution is built on is that, for a small interval, the probability of arrival is proportional to its length. It makes sense that the longer we wait, the more likely that a customer will arrive at the counter.

Consider a series of events that may occur at random time points, say, at $T_1, T_2, \cdots$ in a time period. For example, cars may arrive at a toll booth or light bulbs may fail at times $\{T_n; n\geq 1\}$. The underlying assumption to be made is that the times between consecutive events denoted by $\{X_n; n\geq 1\}$, are all independent and identically distributed random variables and follow an exponential distribution with parameter λ. Let $T_0 = 0$, $T_n = \sum_{i=1}^{n} X_i$, and set $X(t) = \max\{n: \mathrm{T}_n \le t,\ \mathrm{t} \ge 0\}$. Then $X(t)$ is a Poisson process with the parameter λt, i.e., the process $X(t)$, at a fixed point $t > 0$, has the pmf

$$f_{X(t)}(k) = e^{-\lambda t} \tfrac{(\lambda t)^k}{k!}, \quad k = 0,1,2,\cdots \tag{3.8}$$

The proof of this is shown below: From the definitions above, we note that the event $X(t) \ge k$ occurs if and only if $T_k \le t$. It follows that

$$P[X(t) \ge k] = P[T_k \le t]. \tag{3.9}$$

Since X_i's have an exponential density with shape parameter equal to 1, T_k has the gamma density with shape parameter k. Therefore, the pmf of the r.v. $X(t)$ can be obtained utilizing (3.9) as

$$f_{X(t)}(k) = P[X(t) \ge k] - P[X(t) \ge k+1] \ = P[T_k \le t] - P[T_{k+1} \le t]$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^t y^{k-1} e^{-\lambda y} dy - \frac{\lambda^{k+1}}{\Gamma(k+1)} \int_0^t y^k e^{-\lambda y} dy$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^t y^{k-1} e^{-\lambda y} dy - \frac{\lambda^{k+1}}{\Gamma(k+1)} \Big[(-y^k \frac{e^{-\lambda y}}{\lambda}) \big|_0^t + \tfrac{k}{\lambda} \int_0^t y^{k-1} e^{-\lambda y} dy \Big],$$

where we have used integration by parts for the second integral above, so that, in view of $\Gamma(k) = (k-1)!$,

$$f_{X(t)}(k) \ = \frac{\lambda^k}{\Gamma(k)} \int_0^t y^{k-1} e^{-\lambda y} dy + \frac{e^{-\lambda t} (\lambda t)^k}{k!} - \frac{\lambda^k}{\Gamma(k)} \int_0^t y^{k-1} e^{-\lambda y} dy$$

$$= \frac{e^{-\lambda t} (\lambda t)^k}{k!}. \tag{3.10}$$

The proof is complete.

Relationship with Negative Binomial Distribution

Let the random vector (X, Λ) be such that the r.v. $X \mid \Lambda = \lambda$ (that is, the conditional r.v. X given $\Lambda = \lambda$) follows a Poisson distribution with parameter λ and Λ has a Gamma(n , β) distribution; then

$$P(X=x)=\int_0^\infty f_\lambda(x)g_{n,\beta}(\lambda)d\lambda = \frac{1}{\beta^n\Gamma(n)}\int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\lambda^{n-1}e^{-\lambda/\beta}d\lambda$$

$$=\frac{(n+x-1)!}{x!\,(n-1)!}\frac{1}{\beta^n}\left(\frac{\beta}{1+\beta}\right)^{n+x}\int_0^\infty \frac{e^{-[(1+\beta)/\beta]\lambda}\lambda^{n+x-1}}{[\beta/(1+\beta)]^{n+x}\Gamma(n+x)}d\lambda$$

$$=\binom{n+x-1}{x}\left(\frac{1}{1+\beta}\right)^n\left(\frac{\beta}{1+\beta}\right)^x, \qquad (3.11)$$

which is a negative binomial pmf with parameters n and $p=\beta/(1+\beta)$ (see [5]). Green and Yule obtained this result in the year 1920. This model is useful when the parameter $\Lambda=\lambda$ is the expected number of accidents for an individual, which is assumed to vary from person to person following a Gamma distribution, as stated above.

4. Multiple Proofs for Asymptotic Normality

It is well-known that Poisson distribution converges to the normal, as $\lambda=n\to\infty$. In this section, we will present four different methods of proof for showing this convergence as $n\to\infty$. The four different methods of proof are the Bagui-Mehra-Proschan (BMP) Ratio method, the Stirling Approximation method, the MGF method, and the general CLT method. The last three are the historically well-known advanced methods but the first one, a recently devised one (see [4], [6]), is the easiest of all.

4.1. The BMP Ratio Method [6]

This method uses the ratio of two successive probability terms of the pmf $P(X_n=x)=f_n(x)$. Converting this ratio into one in terms of the pmf $f_n^*(z)=P(Z_n=z)$ of the standardized variable $Z_n=[(X_n-n)/\sqrt{n}]$, one can find the limit of the difference quotient of the function $\ln f_n^*(z)$, leading to a simple limiting differential equation. The solution of this equation yields the desired density of the limiting distribution, as $n\to\infty$.

Let X_λ denote a Poisson r.v. with parameter λ . For convenience, set $\lambda=n$. From equation (1.1), the pmf of r.v. X_n is given by

$$f_n(x) = P(X_n = x) = \frac{e^{-n}n^x}{x!}, \ x = 0, 1, 2, \cdots, \tag{4.1}$$

with mean $E(X_n) = \mu_n = n$ and the variance $\text{Var}(X_n) = \sigma_n^2 = n$. The ratio of two consecutive probability terms of the pmf in (4.1) simplifies to

$$\frac{P(X_n = x+1)}{P(X_n = x)} = \frac{n^{x+1}(x!)}{n^x(x+1)!} = \frac{n}{(x+1)}. \tag{4.2}$$

Setting $z = (x-n)/\sqrt{n}$, so that $x = n + z\sqrt{n}$, and substituting it into (4.2) leads to

$$\frac{P\left[(X_n - n)/\sqrt{n} = z + 1/\sqrt{n}\right]}{P\left[(X_n - n)/\sqrt{n} = z\right]} = \frac{1}{1 + z/\sqrt{n} + 1/n}, \tag{4.3}$$

so that by utilizing the notation $Z_n = (X_n - n)/\sqrt{n}$ and $\Delta_n = 1/\sqrt{n}$, we can write equation (4.3) as

$$\frac{P(Z_n = z + \Delta_n)}{P(Z_n = z)} = \frac{1}{1 + z\Delta_n + \Delta_n^2}. \tag{4.4}$$

Since the conditions of Theorem 2.1 of Bagui and Mehra (2020) are satisfied, there is a continuously differentiable pdf $f(z)$, $-\infty < z < \infty$, such that $P(Z_n = z + \Delta_n) \approx f(z + \Delta_n)dz$ and $P(Z_n = z) \approx f(z)dz$ for large n. Upon employing this approximation, (4.4) would reduce to

$$\frac{f(z + \Delta_n)}{f(z)} \approx \frac{1}{1 + z\Delta_n + \Delta_n^2}. \tag{4.5}$$

Now taking logarithms on both sides of (4.5), diving by Δ_n and taking limits as $n \to \infty$, or equivalently as $\Delta_n \to 0$, we get

$$\lim_{\Delta_n \to 0}\left[\frac{\ln f(z + \Delta_n) - \ln f(z)}{\Delta_n}\right] = -\lim_{\Delta_n \to 0}\left[\frac{\ln(1 + z\Delta_n + \Delta_n^2)}{\Delta_n}\right]. \tag{4.6}$$

The left-hand side of (4.6) is simply the derivative of $\ln f(z)$. By applying the L'Hospital's rule to the right-hand side of (4.6), we get the differential equation

$$\frac{d \ln f(z)}{dz} = -\lim_{\Delta_n \to 0}\left[\frac{z + 2\Delta_n}{1 + z\Delta_n + \Delta_n^2}\right] = -z. \tag{4.7}$$

Integrating the two sides of (4.7) with respect to z yields the equation $\ln f(z) = -z^2/2 + c$, where c is the constant of integration. By exponentiating, we obtain $f(z) = ke^{-z^2/2}$ with the constant k to be determined so that $f(z)$ is a valid density, which gives $k = 1/\sqrt{2\pi}$. Thus, we can conclude that the r.v. $Z_n = (X_n - n)/\sqrt{n}$ has the limiting standard normal $N(0,1)$ distribution, as $n \to \infty$; or equivalently, that the Poisson r.v. X_n follows approximately the normal distribution with mean $\mu_n = n$ and the variance $\sigma_n^2 = n$ for sufficiently large n.

4.2. Stirling Approximation Formula Method

First, we substitute Stirling's approximation formula given by (2.1) in the Poisson pmf (4.1). Then, after some algebraic simplification, we have

$$f_n(x) = P(X_n = x) \approx \frac{e^{-\lambda} n^x}{\sqrt{2\pi x}(x)^x e^{-x}} = \frac{n^{x+1/2} e^{-(n-x)}}{\sqrt{2\pi n}(x)^{x+1/2}}$$

$$= \frac{1}{\sqrt{2\pi n}} \left(\frac{x}{n} \right)^{-(x+1/2)} e^{-(n-x)} = C \left(\frac{x}{n} \right)^{-(x+1/2)} e^{-(n-x)}, \tag{4.8}$$

where $C = 1/\sqrt{2\pi n}$. Now taking natural logarithms on both sides of (4.8), we get

$$\ln P(X_n = x) \approx \ln C - (x + 1/2) \ln \left(\frac{x}{n} \right) - (n - x) . \tag{4.9}$$

As in Section 4.1, set $Z_n = (X_n - n)/\sqrt{n}$ and $z = (x-n)/\sqrt{n}$. The second setting implies that $x = n + z\sqrt{n}$, $x/n = (1 + z/\sqrt{n})$ and $(n - x) = n - n - z\sqrt{n} = -z\sqrt{n}$. We re-express (4.9) as

$$\ln P(Z_n = z) = \ln P(X_n = x) \approx \ln C - (n + z\sqrt{n} + 1/2) \ln \left(1 + z/\sqrt{n}\right) + z\sqrt{n}$$

$$= \ln C - \left(n + z\sqrt{n} + 1/2\right) \left[\frac{z}{\sqrt{n}} - \frac{z^2}{2n} + \frac{z^3}{3n^{3/2}} - \cdots \right] + z\sqrt{n}$$

$$= \ln C - \left[z\sqrt{n} - \frac{z^2}{2} + z^2 + O\left(1/\sqrt{n}\right) \right] + z\sqrt{n}$$

$$= \ln C - \frac{z^2}{2} + O\left(1/\sqrt{n}\right), \tag{4.10}$$

so that from (4.10), we obtain for large n that

$$f_n^*(z) = P(Z_n = z) \approx Ce^{-z^2/2} = \frac{1}{\sqrt{2\pi n}} e^{-z^2/2};$$

the above equation may also be written as

$$f_n(x) = P(X_n = x) \approx \frac{1}{\sqrt{2\pi}\sqrt{n}} e^{-\frac{(x-n)^2}{2n}}$$

The proof is complete.

4.3. The MGF Method [2]

Let X_n be a Poisson r.v. with pmf as given in (4.1). The mgf of X_n can be derived as

$$M_{X_n}(t) = E(e^{tX_n}) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-n}(n)^x}{x!} = e^{-n} \sum_{x=0}^{\infty} \frac{(ne^t)^x}{x!} = e^{-n} e^{ne^t} = e^{n(e^t-1)} . \tag{4.11}$$

Let $Z_n = (X_n - n)/\sqrt{n}$, the normed version of X_n. Below we derive the limiting mgf of Z_n to get the limiting distribution of Z_n. In view of (4.11), we obtain the mgf of Z_n as

$$M_{Z_n}(t) = E\left(e^{tZ_n}\right) = E\left[e^{t(X_n/\sqrt{n} - \sqrt{n})}\right] = e^{-t\sqrt{n}} E\left[e^{(t/\sqrt{n})X_n}\right]$$

$$= e^{-t\sqrt{n}} M_{X_n}(t/\sqrt{n}) = e^{-t\sqrt{n}} e^{n\left(e^{t/\sqrt{n}}-1\right)} = \left[e^{-t} e^{\sqrt{n}\left(e^{t/\sqrt{n}}-1\right)}\right]^{\sqrt{n}} . \tag{4.12}$$

Now consider the simplification of the exponent term $\sqrt{n}\left(e^{t/\sqrt{n}} - 1\right)$ in (4.12) as

$\sqrt{n}\left(e^{t/\sqrt{n}} - 1\right) = \sqrt{n}\left(1 + \frac{t}{\sqrt{n}} + \frac{t^2}{(2!)n} + \frac{t^3}{(3!)n^{3/2}} + \frac{t^4}{(4!)n^2}\exp[\varsigma(n)] - 1\right)$, where $\varsigma(n)$ is a number between 0 and $t/\sqrt{n}$ and $\varsigma(n) \to 0$, as $n \to \infty$. The exponent term $\sqrt{n}\left(e^{t/\sqrt{n}} - 1\right)$ further simplifies to $\sqrt{n}\left(e^{t/\sqrt{n}} - 1\right) = t + \frac{t^2}{(2!)\sqrt{n}} + \frac{t^3}{(3!)n} + \frac{t^4}{(4!)n^{3/2}}\exp[\varsigma(n)]$.

Now substitute this last expression on the RHS of the equation (4.12) for $M_{Z_n}(t)$ to obtain

$$M_{Z_n}(t) = \left[e^{-t} e^{t + t^2/[(2!)\sqrt{n}] + t^3/[(3!)n] + t^4 \exp[\varsigma(n)]/[(4!)n^{3/2}]} \right]^{\sqrt{n}} = e^{t^2/2} b(n) , \tag{4.13}$$

with $b(n) = e^{t^3/[(3!)\sqrt{n}]} e^{[t^4 \exp[\varsigma(n)]]/[(4!)n^{3/2}]}$ which tends to 1, as $n \to \infty$. The equation (4.13) yields

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2} \tag{4.14}$$

for all real values of t. By Theorem 2.1, thus, we can conclude from (4.14) that $Z_n = (X_n - n)/\sqrt{n}$ has the limiting standard normal distribution. Equivalently, we may state that the Poisson r.v. X_n follows approximately, for large n, a normal distribution with both mean and variance equal to n.

4.4. The CLT Method

Let $Y_1, Y_2, \cdots$ be a sequence of independent and identically distributed Poisson r.v.'s with parameter λ=1. Suppose $X_n = \sum_{i=1}^{n} Y_i$, then by Theorem 2.3, X_n is a Poisson r.v. with parameter $\lambda = n$. We also note that X_n is the sum of n i.i.d. r.v.'s with mean equal to 1 and the variance also equal to 1. By the CLT stated in Theorem 2.2, therefore, we may conclude that $Z_n = [(X_n - n)/\sqrt{n}] \xrightarrow{d} N(0,1)$, as $n \to \infty$; or equivalently, that the r.v. X_n- which is Poisson $P(n)$- follows approximately the normal distribution with mean $\mu_n = n$ and variance $\sigma_n^2 = n$ for large n.

Example 4.1 Normal approximation to the Poisson. Each year in Mythica (see [9], p. 217), an average of 25 postal delivery persons are bitten by dogs. Suppose one wants to know the probability that at least 33 such incidents occur in a particular year. The dog-bites being "rare" events, the distribution of the total number X - of dog-bites in a year - can be modeled by Poisson $P(\lambda)$ with the mean λ as the observed value $\lambda = 25$ above. The exact calculation of the probability $P(X \geq 33)$ will take some doing. It can be done using incomplete Γ-function tables or any statistical computing software such as R. But since λ is reasonably large, we can use the normal approximation to the Poisson to deduce its approximate value much more easily using the standard normal tables. Figure 4.1 below shows the closeness of the two distributions.

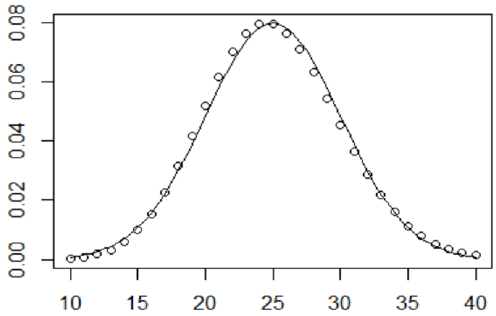

Figure 4.1: Plot of Poisson and Normal with mean $\lambda = 25$

The mean of the assumed model is $\lambda = 25$, while the standard deviation $\sqrt{\lambda} = \sqrt{25} = 5$. Thus, the normal approximation for $P(X \geq 33) \approx P(Z \geq (32.5 - 25)/5) = P(Z \geq 1.5) = 0.0668$, whereas the exact probability under Poisson $P(\lambda)$ distribution with $\lambda = 25$, namely, $P(X \geq 33) = 1 - P(X \leq 32)$ is 0.0714 approximately. The approximation error up to four decimal places is only 0.0046. The normal tables, thus, can provide a quick and fairly accurate answer without the use of cumbersome Γ-function tables or any statistical software.

5. Concluding Remarks

In this article, we have given in the introduction a historical background for the Poisson distribution and also mentioned one of its early applications (see Bortkewitsch (1898)) in modeling the numbers of Prussian cavalry killed by horse-kicks. Later during World war II, Poisson distribution was again employed effectively to conclude that the hits of flying bombs (V-1 and V-2 missiles) in London were landing at random and not at predetermined targeted spots. We discussed some basic properties of the Poisson distribution. We also derived Poisson approximations to the binomial and negative-binomial distributions and indicated its relationship to other distributions. We showed that a gamma mixture of Poisson distributions yields a negative-binomial distribution. Accordingly, negative-binomial can be an alternative to the Poisson in modeling similar but more complex type rare events data.

In Section 4, we presented four different methods of proof of the convergence of Poisson to a normal distribution, as the parameter $\lambda \to \infty$. These methods are the BMP Ratio method and the other three well-known ones based, respectively, on the Sterling's formula, the Moment Generating Functions (mgf's), and that of the general Central Limit Theorems (CLT). The new BMP Ratio method uses only a basic knowledge of Calculus. The other three, relatively more advanced, are respectively: the De-Moivre's approach based on Sterling's formula, the MGF method based on Laplace transforms, and that of the general Central Limit Theorems based on characteristic functions. These four methods of proof can be profitably discussed in a senior classroom setting. Bagui and Mehra (2017, 2019) had earlier employed these methods for showing the convergence of binomial and negative-binomial to the limiting normal, as the number of trials $n \to \infty$.

The Poisson Paradigm

As discussed in the preceding sections, Poisson distribution can be employed befittingly to model the number of occurrences that can take place over a period of time under "rare" events phenomena, such as those described in the introductory section, viz., the number of severe earthquakes, traffic accidents, hits of bombs, deaths by kicking horses and so on. We shall give a formal shape to this assertion in inequality (5.1) below and follow it up with a paradigmic example of its application:

Let $A_1, A_2, \cdots, A_n$ be independent (or weakly dependent) "rare" events with $p_k = P(A_k),\ 1 \le k \le n,\ p_k$ being small, $\lambda_n = \sum_{k=1}^{n} p_k$ moderate and n large, and set $X_n = \sum_{k=1}^{n} I(A_k)$, where $I(A_k)$ is equal to 1 if the event A_k occurs, otherwise it is 0. Thus, X_n counts the number among events $A_k,\ k = 1, 2, \cdots, n$ that occur. Note that $E(X_n) = \sum_{k=1}^{n} E[I(A_k)] = \sum_{k=1}^{n} p_k = \lambda_n$. Then for large n, X_n follows approximately a Pois(λ_n) distribution with parameter λ_n. More precisely, if N denotes a Pois(λ_n) r.v. and C any set of non-negative integers, then

$$|P(X_n \in C) - P(N \in C)| \le min\left(1, \frac{1}{\lambda_n}\right) \sum_{k=1}^{n} p_k^2 \ ; \tag{5.1}$$

(see [7], p.164 or [15], p.410). The upper bound on the right provides the maximum error that may be incurred from using Poisson approximation, not only for approximating the probability of a single point but also in approximating the probability of any set. In other words, how small the p_k should be so that $\sum_{k=1}^{n} p_k^2$ is very small, or at least very small compared to λ_n. For example, given $\varepsilon > 0$ we may choose n sufficiently large so that $\max_{1\le k\le n} p_k \le \sqrt{\varepsilon(\lambda_n/n)}$; then the maximum error would be less than or equal to ε .

As an application, we consider the well-known "birthday" problem: In a group of m persons, what is the probability that there is at least one pair in the group who have the same birthday? There are $n = \binom{m}{2}$ pairs in this group. The probability of each pair having the same birthday is equal to P(1st -person and 2nd person both have the same birthday "D") = P(1st -Person has a birthday "D")P(2nd person has birthday "D"|1st -Person has birthday "D") = $\frac{1}{365} \cdot 1 = \frac{1}{365}$. By the Poisson paradigm, the distribution of X_n - the number of birthday matches- is approximately Pois(λ_n) with $\lambda_n = \frac{n}{365} = \binom{m}{2}\frac{1}{365}$ for large n. Thus, the probability of at least one match is approximately

$$P(X_n \ge 1) = 1 - P(X_n = 0) \approx 1 - e^{-\lambda_n}.$$

For m = 23, $\lambda_n = 253/365$ and $1 - e^{-\lambda_n} \approx 0.500$, which is very close to the exact probability 0.507 of having at least one match in a group of 23 people.

This note may serve as a useful pedagogical reference article in senior-level probability courses. The material should also be of reading interest for senior students in probability and mathematical statistics. The teachers may assign these different methods of proof to students as class projects.

Acknowledgements: The authors would like to thank the Editors and the referee for their constructive comments and suggestions on the earlier version of the article.

References

[1] Bagui, S. C., Bagui, S. S., and Hemasinha, R. (1913). Nonrigorous proofs of Striling's formula, Mathematics and Computer Education, 47, 115-125.

[2] Bagui, S. C. and Mehra, K. L. (1916). Convergence of Binomial, Poisson, Negative Binomial, and Gamma to normal distribution: Moment

generating functions technique, American Journal of Mathematics and Statistics, **6**, 115-121.

[3] Bagui, S. C. and Mehra, K. L. (1917). Convergence of Binomial to Normal: Multiple proofs, International Mathematical Forum, 12, 399-422.

[4] Bagui, S. and Mehra, K. L. (2017a). Convergence of Known Distribution to Limiting Normal or Non-normal Distributions: An Elementary Ratio Technique, The American Statistician, 71:3, 265-271.

[5] Bagui, S. C. and Mehra, K. L. (2019). On the Convergence of Negative Binomial Distribution, American Journal of Mathematics and Statistics, 9, 44-50, DOI: 10.5923/j.ajms.20190901.06.

[6] Bagui, S. C. and Mehra, K. L. (2020). On an Elementary Ratio Technique for Proving Convergence of Known Distributions, Communications in Statistics-Theory and Methods, 49(22), 5535-5552, DOI: 10.1080/ 03610926.2019. 1619 771.

[7] Blitzstein, J. K. and Hwang, J. (2014). Introduction to Probability, CRC Press, Taylor & Francis Group, Boca Raton, FL.

[8] Bortkewitsch, L. von (1898). Das Gesetz der kleinen Zahlen, Leipzig: B. G. Teubner.

[9] Durrett, R. (2009). Elementary Probability for Applications, Cambridge University Press.

[10] Goon, A. M., Gupta, M. K., and Dasgupta, B. (1977). An Outline of Statistical Theory, The World Press Private Limited, Calcutta.

[11] Haight, F. A. (1967). Handbook of the Poisson distribution, John Wiley & Sons, Inc., New York.

[12] Larsen, R. J. and Marx, M. L. (2012). An Introduction to Mathematical Statistics and Its Applications, Prentice-Hall, Boston.

[13] Poisson, S. D. (1837). Recherches sur la probabilitte des jugements en matiere criminelle et en matiere civile, Paris, Bachelier, Imprimeur-Libraire.

[14] Preece, D. A., Ross, J. S., and Kirby, P. J. (1988). Bortkewitsch's horse-kicks and the generalized linear model, The Statistician, 37, 313-318.

[15] Ross, S. (1988). A First Course in Probability, 4th edition, Prentice-Hall, Englewood Cliffs, NJ.

[16] Roussas, G. (2009). A Course in Mathematical Statistics, 2nd edition, Academic Press, New York.

[17] Thedeen, Torbjörn (1964). A note on the Poisson tendency in traffic distribution. Annals of Mathematical Statistics, 35, 1823-1824.

[18] Tijms, Henk (2013). Understanding Probability, Cambridge University Press.

APPENDIX

1. Recursive Relations for Moments of P_λ Distribution

Let $X = X_\lambda$ be a Poisson P_λ r.v., that is, with parameter λ ($\lambda > 0$) and let $\alpha_r(\lambda) = \alpha_{r,X}(\lambda) = E(X^r)$ denote its rth integral ($r \geq 1$) moment about the origin. Further, let $\mu_r(\lambda) = \mu_{r,X}(\lambda) = E[X - \alpha_1(\lambda)]^r$ denote its rth (integral) central moment, that is, about the mean $\alpha_1(\lambda) = E(X)$. It is well-known that for the Poisson X_r r.v. its mean $\alpha_1(\lambda) = E(X)$ equals λ and so does its variance $\mu_2(\lambda)$. The following two recursion relations (see equations (A.1) and (A.4) below), respectively for the two preceding series of moments, assist us in calculating all their values fairly easily.

I. Recursion Relation for $\alpha_{r,X}(\lambda)$, $r \geq 1$.

For all integral values $r \geq 1$, the following recursion relation holds:

$$\alpha_{r+1,X}(\lambda) = \lambda[\alpha_{r,X}(\lambda) + \tfrac{d\alpha_{r,X}(\lambda)}{d\lambda}]; \tag{A.1}$$

(In fact, since $\alpha_{0,X}(\lambda) = E[X^0] = 1$, the formula (A.1) holds for $r = 0$ also.)

To see that the formula (A.1) holds for $r \geq 1$, first note that

$$\tfrac{d}{d\lambda}\alpha_{r,X}(\lambda) = \tfrac{d}{d\lambda}\sum_{x=0}^{\infty} x^r \tfrac{e^{-\lambda}\lambda^x}{x!} = -\alpha_{r,X}(\lambda) + \sum_{x=1}^{\infty} x^r \tfrac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}$$

$$= -\alpha_{r,X}(\lambda) + \sum_{y=0}^{\infty}(y+1)^r \tfrac{e^{-\lambda}\lambda^y}{y!} = -\alpha_{r,X}(\lambda) + \alpha_{r,(X+1)}(\lambda), \tag{A.2}$$

and then also that

$$\alpha_{r+1,X}(\lambda) = \sum_{x=0}^{\infty} x^{r+1} \tfrac{e^{-\lambda}\lambda^x}{x!} = \lambda\sum_{x=1}^{\infty} x^r \tfrac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}$$

$$= \lambda\sum_{y=0}^{\infty}(y+1)^r \tfrac{e^{-\lambda}\lambda^y}{y!} = \lambda\alpha_{r,(X+1)}(\lambda)\,. \tag{A.3}$$

Now substituting for $\alpha_{r,(X+1)}(\lambda)$ from (A.2) into (A.3), we obtain the recursion relation (A.1) readily.

Note that the recursion relation (A.1) holds for $r=0$ also. This is so because, on account of $\alpha_{0,X}(\lambda)=1$, the relation (A.1) if true should yield $\alpha_{1,X}(\lambda)=\lambda[1+0]=\lambda$, which being factually true provides the verification.

II. Recursion Relation for $\mu_{r,X}(\lambda),\ r\geq 1$.

The following recursion relation also holds for all integral values of $r\geq 1$:

$$\mu_{r+1,X}(\lambda)=\lambda[r\mu_{r-1,X}(\lambda)+\tfrac{d\mu_{r,X}(\lambda)}{d\lambda}]; \qquad \text{(A.4)}$$

(Again, since $\mu_{0,X}(\lambda)=1$ and $\mu_{1,X}(\lambda)=0$, the formula (A.4) holds for $r=0$ also.)

To see that it holds for all integral $r\geq 1$, first note that

$$\tfrac{d\mu_{r,X}(\lambda)}{d\lambda}=\tfrac{d}{d\lambda}[\textstyle\sum_{x=0}^{\infty}(x-\lambda)^r\tfrac{e^{-\lambda}\lambda^x}{x!}]=-r\mu_{r-1,X}(\lambda)-\mu_{r,X}(\lambda)+\textstyle\sum_{x=1}^{\infty}(x-\lambda)^r\tfrac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}$$

$$=-r\mu_{r-1,X}(\lambda)-\mu_{r,X}(\lambda)+\mu_{r,(X+1)}(\lambda), \qquad \text{(A.5)}$$

and then also that

$$\mu_{r+1,X}(\lambda)=\textstyle\sum_{x=0}^{\infty}(x-\lambda)^{r+1}\tfrac{e^{-\lambda}\lambda^x}{x!}=\textstyle\sum_{x=1}^{\infty}x(x-\lambda)^r\tfrac{e^{-\lambda}\lambda^x}{x!}-\lambda\mu_{r,X}(\lambda)$$

$$=\lambda\textstyle\sum_{x=0}^{\infty}(x+1-\lambda)^r\tfrac{e^{-\lambda}\lambda^x}{x!}-\lambda\mu_{r,X}(\lambda)=\lambda[\mu_{r,X+1}(\lambda)-\mu_{r,X}(\lambda)]. \qquad \text{(A.6)}$$

Now substituting for $\mu_{r,X+1}(\lambda)$ from (A.5) into (A.6), we obtain the relationship (A.4) as follows:

$$\mu_{r+1,X}(\lambda)=\lambda[r\mu_{r-1,X}(\lambda)+\mu_{r,X}(\lambda)+\tfrac{d\mu_{r,X}(\lambda)}{d\lambda}-\mu_{r,X}(\lambda)]=\lambda[r\mu_{r-1}(\lambda)+\tfrac{d\mu_{r,X}(\lambda)}{d\lambda}].$$

The derivation is complete. One can easily see that the relationship (A.4) holds for $r=0$ also. This is so because, on account of $\mu_{0,X}(\lambda)=1$, (A.4) if true should yield $\mu_{1,X}(\lambda)=\lambda[0\cdot\mu_{-1,X}(\lambda)+\tfrac{d(0)}{d\lambda}]=0$, which being true by definition provides the verification.

2. Derivation of the Formula (1.2) for the CDF F_{X_λ} of Poisson X_λ

The derivation of the above formula is as follows: Note that by virtue of 'integration by parts' formula, we have for each integral k, $1\leq k\leq x$, that

$$A(k)=\frac{1}{\Gamma(k+1)}\int_{\lambda}^{\infty}e^{-y}y^{k}dy \ = -e^{-y}\left(\frac{y^k}{k!}\right)\Big|_{\lambda}^{\infty}+\frac{k}{\Gamma(k+1)}\int_{\lambda}^{\infty}e^{-y}\,y^{k-1}dy=\frac{e^{-\lambda}\lambda^k}{k!}+A(k-1). \quad \text{(A.7)}$$

Summing both sides of (A.7) over $1\le k\le x$, we obtain

$$\sum_{k=1}^{x}A(k)=\sum_{k=1}^{x}\frac{e^{-\lambda}\lambda^k}{k!}+\sum_{k=0}^{x-1}A(k)\,, \quad \text{(A.8)}$$

or equivalently, upon canceling the sum $\sum_{k=0}^{x-1}A(k)$ on opposite sides of (A.8), that

$$A(x)=\sum_{k=1}^{x}\frac{e^{-\lambda}\lambda^k}{x!}+A(0)\ =\ F_{X_\lambda}(x) \quad \text{for all} \quad x\ge 0\,.$$

This is exactly the formula (1.2), viz., for all integral $x\ge 0$

$$F_{X_\lambda}(x)=P(X_\lambda\le x)\ =\sum_{k=0}^{x}\frac{e^{-\lambda}\lambda^k}{k!}=\frac{1}{\Gamma(x+1)}\int_{\lambda}^{\infty}e^{-y}y^{x}dy\,. \quad \text{(A.9)}$$

The derivation is complete.

An alternative derivation of the formula (1.2) is as follows:

Since by the definition of Gamma function $\Gamma(x-k+1)=\int_0^\infty t^{x-k}e^{-t}dt$ (see [13], p. 67), for each integral $x\ge 0$,

$$F_{X_\lambda}(x)=P(X_\lambda\le x)=\sum_{k=0}^{x}\frac{e^{-\lambda}\lambda^k}{k!}\ =\frac{e^{-\lambda}}{x!}\sum_{k=0}^{x}\frac{x!}{k!(x-k)!}(x-k)!\lambda^k$$

$$=\ \frac{e^{-\lambda}}{\Gamma(x+1)}\sum_{k=0}^{x}\binom{x}{k}\Gamma(x-k+1)\lambda^k\ =\frac{1}{\Gamma(x+1)}\sum_{k=0}^{x}\binom{x}{k}\left(\int_0^\infty t^{x-k}e^{-(\lambda+t)}dt\right)\lambda^k\,, \quad \text{(A.10)}$$

so that by interchanging the summation and integration signs in (A.10), we obtain

$$F_{X_\lambda}(x)=P(X_\lambda\le x)$$

$$=\frac{1}{\Gamma(x+1)}\ \int_0^\infty e^{-(\lambda+t)}\Big[\sum_{k=0}^{x}\binom{x}{k}\lambda^k t^{x-k}\Big]dt\ \ =\frac{1}{\Gamma(x+1)}\int_0^\infty e^{-(\lambda+t)}(\lambda+t)^x\,dt$$

$$=\frac{1}{\Gamma(x+1)}\int_{\lambda}^{\infty}e^{-y}y^{x}dy \quad \text{for all integral} \quad x\ge 0\,, \quad \text{(A.11)}$$

where for the last two equalities in (A.11), we have used, respectively, the binomial summation and a change in the variable of integration from t to y = $\lambda+t$. The equation (A.11) is precisely the same as (A.9) or (1.2). The derivation is complete.