東 海 大 學 統 計 學 研 究 所

碩 士 論 文

指導教授: 王堯弘 博士

ON THE POISSON APPROXIMATION

研究生: 陳怡君

中 華 民 國 九 十 五 年 六 月

ON THE POISSON APPROXIMATION

By

Yi-Chun, Chen

Dept. of Statistics

Tunghai University

Taichung, 40704, Taiwan

R. O. C.

June, 2006

ABSTRACT

Statistical applications and repercussions of the Poisson distribution. In this thesis, we introduce the relationship between the Poisson distribution and the Binomial , Negative Binomial, and hypergeometric distributions. Moreover, we consider the Poisson approximation to the distribution of the sums of iid Bernoulli / geometric and non-identically distributed random variables.

CONTENTS

1. Introduction 3

2. Some Discrete Distribution 5

2.1 Bernoulli Trials 6

2.2 Binomial Distribution 7

2.3 Geometric Distribution 9

2.4 Negative Binomial Distribution 10

2.5 Poisson Distribution 12

3. The Poisson Approximation for Sums of Independent and Identically Distributed Random Variables 16

3.1 Limits of the Binomial Distributions 17

3.2 Limits of the Negative Binomial Distribution 26

3.3 Limits of the Hypergeometric Distributions 27

4. The Poisson Approximation for Sums of Independent but Non-Identically Distributed Random Variables 28

CHAPTER 1 INTRODUCTION

Simeon Denis Poisson (1781-1840) was a famous French mathematician and physicist. The distribution which attributed to him now called the Poisson distribution is of great importance in theory and in practice. The Poisson approximation is a subject with a long history and many ramifications. It can be used as a tool for estimating probabilities connected with rare or exceptional occurrences. Poisson approximations are essential in extreme value theory, in reliability, in actuarial mathematics, and in many other applied fields. Following the work of Poisson(1837), there has been considerable practical and theoretical interest in how well the Poisson distribution approximates the binomial and many other distributions. This thesis contained a good deal of mathematics, including a limit theorem which derive the Poisson distributions as the limiting distributions of the binomial distribution. He approached the distribution by considering limiting forms of the binomial distribution. Although initially viewed as a little more than a welcome approximation for hard-to-compute binomial probabilities, this particular result was destined for bigger things: it was the analytical seed out of which grew what is now one of the most important of all probability models, the *Poisson distribution.*

Actually, as will be seen below, Poisson derived the distribution directly as an approximation to the negative binomial cumulative distribution. There is no indication that he sensed the wide applicability of the distribution; rather, it was one of several approximations and received no special comment.

Bortkiewicz(1898) considered circumstances in which Poisson's distribution might arise. He wrote a monograph included the first consideration of the Poisson limit as a probability distribution, and, most important, the use of the "Poisson" to model real-world phenomena. From the point of view of Poisson's own approach, these are situations where in addition to the requirements of independence of trials, and constancy of probability from trial to trial , the number of trials must be very large while the probability of occurrence of the outcome under observation must be small.

In this thesis, we shall present the proofs various distributions converging to the Poisson distributions. It is then expanded into a section of its own with detailed descriptions of the applications. In early 20^{th} century, the development of applied probability, especially the queueing theory, owns a great deal to the Poisson distributions. Poisson process is a natural model in input process and output process, such as arrival and the departure of customers at certain business institutions.

CHAPTER 2 SOME DISCRETE DISTRIBUTIONS

In this chapter, we briefly summarize some related discrete distributions used in this thesis.

2.1 Bernoulli Trials

Repeated independent trials are called Bernoulli trials[1] if there are only two possible outcomes at each trial. Bernoulli trials are the simplest type of random variable with two possible outcomes. One outcome is usually called a *"success"*, denoted by S, the other outcome is called a *"failure"*, denoted by F.

The sample space of a Bernoulli trial contains two points, S and F, the random variable defined by $X(S) = 1$ and $X(F) = 0$ is called a Bernoulli

[1] Bernoulli trials, named after the Swiss mathematician James Bernoulli(1654-1705), are perhaps the simplest type of random variable.

random variable. The probability function of X is

$$P(X=x)=\begin{cases} p & \text{if x = 1} \\ q=1-p & \text{if x = 0} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, p and q must be non-negative, and $p+q=1$.

Lemma 1. *X has a Bernoulli(p) distributions, then the probability generating function (pgf) $g(s)=E(s^X)$ is $g(s)=(1-p+ps), |s|\leq 1$.*

Proof.

$$\begin{aligned} g(s) &= E\left(s^x\right) \\ &= P\left(X=0\right)s^0+P\left(X=1\right)s^1 \\ &= (1-p)+ps. \end{aligned}$$

□

2.2 Binomial Distribution

Now we are interested in the total number of successes in a succession of n independent Bernoulli trials. The number of successes can be $0,1,\ldots n$. The event n trials results in k successes and $n-k$ failures can happen in as many different ways as k letters S can be distributed among n places. The event contains $\binom{n}{k}$ points, and each point has the probability $p^k\left(1-p\right)^{n-k}$. If n

Bernoulli trials all with probability of success p are performed independently, then X, the number of successes is one of the most important random variables called a binomial with parameters n and p. The probability function is given by the following theorem.

Theorem 1. *Let S_n be a Binomial random variable with parameters n and p, then the probability function of X, is*

$$P(S_n = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \textit{if } x = 0,1,2,\ldots,n \\ 0 & \textit{otherwise.} \end{cases}$$

If we let the random variable X equal the number of observed successes in n independent Bernoulli trials, the possible values of X are $0, 1, \ldots, n$. If x successes occur, where $x = 0, 1, \ldots, n$, then $n-x$ failures occur. The number of ways of selecting the x positions for the x successes in the n trials is

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Summarizing a binomial experiment satisfies these properties as follows:

1. A Bernoulli experiment is performed n times.
2. The random variable X equals the number of successes in the n trials.
3. The trials are independent.
4. The probability of success on each trial is a constant p; the probability of failure is $q = 1 - p$.

Lemma 2. *X has a Binomial (n, p) distributions, then the probability generating function (pgf) $g_n(s) = E(s^X)$ is $g_n(s) = (1 - p + ps)^n$, for $|s| \leq 1$.*

Proof.

$$\begin{aligned} g_n(s) &= E(s^x) \\ &= \sum_{x=0}^{n} P(S_n = x) s^x \\ &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} s^x \\ &= \sum_{x=0}^{n} \binom{n}{x} (ps)^x (1-p)^{n-x} \\ &= [ps + (1-p)]^n . \end{aligned}$$

□

2.3 Geometric Distribution

Suppose that a sequence of independent Bernoulli trials, each with probability of success p, $0 < p < 1$. Let X be the number of failures until the first success occurs, then X is a discrete random variable called *geometric*.

Theorem 2. *Let X be a geometric random variable with parameter p, $0 < p < 1$, and its set of possible values is $\{0, 1, 2, \dots\}$, then the probability function of X is*

$$P(X = x) = \begin{cases} (1-p)^x p & \text{if } x = 0,\ 1,\ \dots, \\ 0 & \text{otherwise.} \end{cases}$$

where there are x trials are all failures, $(x+1)^{th}$ trial is a success, and the successive Bernoulli trials are all independent.

Lemma 3. *X has a Geometric(p) distributions, then the probability generating function (pgf) $g(s) = E(s^X)$ is $g(s) = \frac{p}{1-(1-p)s}$, for $|s| \leq 1$.*

Proof.

$$\begin{aligned} g_(s) &= E(s^x) \\ &= \sum_{x=0}^{\infty} s^x (1-p)^x p \\ &= p \sum_{x=0}^{\infty} ((1-p)s)^x \\ &= \frac{p}{1-(1-p)s}, \quad \text{for} \quad |(1-p)s| \leq 1. \end{aligned}$$

□

2.4 Negative Binomial Distribution

Negative binomial random variables are generalizations of geometric random variables. Suppose that a sequence of independent Bernoulli trials, each with probability of success p, $0 < p < 1$. Let X be the number of failures until the r^{th} success occurs, where r is a fixed integer, then X is a discrete random variable called a *negative binomial*.

Theorem 3. *Let X be a negative binomial random variable with parameters r and p, $0 < p < 1$, and its set of possible values is $\{r, r+1, r+2, \dots\}$, then* *the probability function of X is*

$$P(X=x)=\begin{cases}\binom{x-1}{r-1}p^r\,(1-p)^{x-r} & \text{if } x=r,\ r+1,\ \dots,\\ 0 & \text{otherwise,}\end{cases}$$

where the outcome of the x^{th} trial is the r^{th} success, then in the first $(x-1)$ trials exactly $(r-1)$ successes have occurred and the x^{th} trial is a success.

The negative binomial distribution is sometimes defined in terms of the random variable Y = number of failures before the r^{th} success. This formulation is statistically equivalent to X = trial at which the r^{th} success occurs, since $Y=X-r$.

Theorem 4. *Let Y be a negative binomial random variable with parameter p, $0<p<1$, $Y = X$ - r, and its set of possible values is $\{0,1,2,\dots\}$, then the probability function of Y, $p(y)$ is*

$$p(Y=y)=\begin{cases}\binom{r+y-1}{y}p^r\,(1-p)^{y} & \text{if } y=0,\ 1,\ \dots,\\ 0 & \text{otherwise.}\end{cases}$$

Lemma 4. *Y has a Negative Binomial(r,p) distributions, then the probability generating function (pgf) $g_r(s)=E(s^Y)$ is $g_r(s)=\left[\frac{p}{1-(1-p)s}\right]^r$, for $|s|\leq 1$.*

Proof.

$$\begin{aligned}
g_n(s) &= E(s^y) \\
&= \sum_{y=0}^{\infty} P(S_n = y) \\
&= \sum_{y=0}^{\infty} s^y \binom{r+y-1}{y} p^r (1-p)^y \\
&= p^r \sum_{y=0}^{r} (-1)^y \binom{-r}{y} (s(1-p))^y \\
&= \left[\frac{p}{1-(1-p)s}\right]^r .
\end{aligned}$$

□

2.5 Poisson Distribution

In many situations the number of events of a specified kind has approximately the Poisson distribution such as

- Accidents happened in a given stretch of road during a fixed period.
- Sharks attacked the sightseer at a beach.
- The number of misprints on a document page typed by a secretary.
- The soldiers killed by the kick of horses.
- Supreme Court vacancies.

Many textbooks in statistics content themselves with an explanation of this phenomenon that runs something like this: There is a large number, say n, of events that might occur, for example, there are many telephone subscribers who might place a call in a minute. The chance, say p, that any specified one of these events will occur. The Poisson distribution can be derived from a set of basic assumptions, sometimes called the Poisson postulates, and these processes are called Poisson processes. Now we state as a theorem as follows:

Theorem 5. *For each $t \geq 0$, let $N(t)$ denote the number of events that occur in a given interval $[0, t]$, and suppose that the following assumptions hold. a) The probability that an event will occur in a given short interval $[t, t + \Delta t]$ is approximately proportional to the length of the interval, Δt; b) Does not depend on the position of the interval; c) That the occurrences of events in nonoverlapping intervals are independent; and d) The probability of two or more events in a short interval $[t, t + \Delta t]$ is negligible, if these assumptions are valid as $\Delta t \to 0$, then the distribution of N_t is Poisson. Note that $\boldsymbol{o}(\Delta t)$ denotes a function of $\Delta(t)$ such that $\lim_{\Delta t \to 0} \frac{\boldsymbol{o}(\Delta t)}{\Delta(t)} = 0$, $\boldsymbol{o}(\Delta t)$ is negligible relative to $\Delta(t)$.*

Now let $N(t)$ be the number of occurrences in the interval $[0, t]$, and $P_n(t) = P(N(t) = n)$. Consider the properties as follows:

(1) Start with no arrivals, denoted $N(0) = 0$*.*

(2) $P[N(t+h) - N(t) = n | N(s) = m] = P[N(t+h) - N(t) = n]$ *for all* $0 \leq s \leq t$ *and* $h > 0$*.*

(3) $P[N(t+\Delta t) - N(t) = 1] = \lambda \Delta t + \boldsymbol{o}(\Delta t)$ *for some constant* $\lambda > 0$*.*

(4) $P[N(t+\Delta t) - N(t) \geq 2] = \boldsymbol{o}(\Delta t)$*.*

If it satisfies conditions (1)-(4), then for any integer n *and for all* $t > 0$*,*

$$P_n(t) = P(N(t) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}.$$

That is, $N(t) \sim Poisson(\lambda t)$*.*

Proof. Now n events may occur in the interval $[0, t+\Delta t]$ by having 0 events in $[t, t+\Delta t]$ and n events in $[0, t]$, or one event in $[t, t+\Delta t]$, and $n-1$ events in $[0, t]$, or two or more events in $[t, t+\Delta t]$; thus for $n > 0$,

$$\begin{aligned} P_n(t+\Delta t) &= P_{n-1}(t)P_1(\Delta t) + P_n(t)P_0(\Delta t) + \mathbf{o}(\Delta t) \\ &= P_{n-1}(t)\left[\lambda \Delta t + \mathbf{o}(\Delta t)\right] + P_n(t)\left[1 - \lambda \Delta t - \mathbf{o}(\Delta t)\right] + \mathbf{o}(\Delta t) \end{aligned}$$

but

$$\begin{aligned} \frac{dP_n(t)}{dt} &= \lim_{\Delta t \to 0} \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{P_{n-1}(t)\lambda \Delta t + P_n(t) - P_n(t)\lambda \Delta t - P_n(t)}{\Delta t} \\ &= \lambda\left[P_{n-1}(t) - P_n(t)\right]. \end{aligned}$$

For $n = 0$,

$$P_0(t+\Delta t) = P_0(t)\, P_0(\Delta t)$$

$$= P_0(t)\left[1 - \lambda\Delta t - \mathbf{o}(\mathit{\Delta} t)\right]$$

$$\frac{dP_0(t)}{dt} = \lim_{\Delta t \to 0} \frac{P_0(t+\Delta t) - P_0(t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{-P_0(t)\,\lambda\Delta t - P_0(t)\,\mathbf{o}(\mathit{\Delta} t)}{\Delta t}$$

$$= -\lambda\left[P_0(t)\right].$$

Assuming that the initial condition $P_0(0) = 1$, the solution to the above differential equation is verified to be $P_0(t) = e^{-\lambda t}$, by mathematical induction, let $n = 1$, then

$$\frac{dP_1(t)}{dt} = \lambda\left[P_0(t) - P_1(t)\right]$$

$$= \lambda\left[e^{-\lambda t} - P_1(t)\right],$$

which gives $P_1(t) = \lambda t e^{-\lambda t}$. Assume $n = k, P_k(t) = e^{-\lambda t}(\lambda t)^k / k!$, now for $n = k + 1$,

$$\frac{dP_{k+1}(t)}{dt} = \lambda\left[P_k(t) - P_{k+1}(t)\right]$$

$$= \lambda\left[e^{-\lambda t}(\lambda t)^k / k! - P_{k+1}(t)\right]$$

which gives $P_{k+1}(t) = \frac{e^{-\lambda t}(\lambda t)^{k+1}}{(k+1)!}$. □

Thus, $N(t) \sim Poisson(\lambda t)$, where $\mu = E[N(t)] = \lambda t$. The proportionality constant λ reflects are *rate of occurrence* or *intensity* of the Poisson process. Since λ is assumed constant over t, the process is referred to as a *homogeneous Poisson process*(HPP). Now we define Poisson distribution as follows,

Theorem 6. *A discrete random variable X with possible values* $0, 1, 2, \ldots$ *is called Poisson with parameter* $\lambda, \lambda > 0$*, if*

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

Lemma 5. *Let X be a Poisson*(λ) *distribution, the probability generating function (pgf)* $g(s)$ *of* X *is* $g(s) = e^{-\lambda(1-s)}$*, for all* $s \in R$.

Proof.

$$\begin{aligned} g(s) = E\left(s^x\right) &= \sum_{x=0}^{\infty} P\left(X = x\right) s^x \\ &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x s^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda s)^x}{x!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}. \end{aligned}$$

□

CHAPTER 3 THE POISSON APPROXIMATION FOR SUMS OF INDEPENDENT AND IDENTICALLY DISTRIBUTED RANDOM VARIABLES

We first consider the Bernoulli random variables.

3.1 Limits of the Binomial Distributions

Assuming that the events are independent, has exactly the binomial distribution $B(n, p)$. Now let $n \to \infty$, and $p \to 0$, so that $np \to \lambda$, where $0 < \lambda < \infty$, is fixed.

Now we prove that $B(n, p)$ tends to the Poisson distribution $\mathrm{P}(\lambda)$ with expectation λ.

Lemma 6. *Let a be a real number, then* $\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a$.

Proof. • Method I

Define $f(x) = a \ln x$, then $f'(x) = \frac{a}{x}$, $f'(1) = a$

$$\begin{aligned} f'(1) &= \lim_{h\to 0} \frac{f(1+h) - f(1)}{h} \\ &= \lim_{h\to 0} \frac{a}{h} \ln(1+h) \\ &= \lim_{h\to 0} [\ln(1+h)]^{\frac{a}{h}} \\ &= \ln\left[\lim_{h\to 0} (1+h)^{\frac{a}{h}}\right] = a. \end{aligned}$$

so that

$$\lim_{h\to 0} (1+h)^{\frac{a}{h}} = e^a.$$

By replacing h with x , then

$$\lim_{x\to 0} (1+x)^{\frac{a}{x}} = e^a.$$

Let $n = \frac{a}{x}$, $x = \frac{a}{n}$, as $n \to \infty$, $x \to 0$, so that

$$\begin{aligned} \lim_{x\to 0} (1+x)^{\frac{a}{x}} &= \lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n \\ &= e^a. \end{aligned}$$

- Method II

Let $y_n = \left(1 + \frac{a}{n}\right)^n$, then $\log(y_n) = n \log\left(1 + \frac{a}{n}\right)$.

Using the Maclaurin series for $\log(y_n)$ as follows:

$$\begin{aligned} \log(y_n) &= n \log\left(1 + \frac{a}{n}\right) \\ &= n\left(\frac{a}{n} - \frac{a^2}{2n^2} + \frac{a^3}{3n^3} - \frac{a^4}{4n^4} + \cdots\right) \\ &= n\left(\frac{a}{n} + \boldsymbol{o}(\frac{1}{n})\right), \end{aligned}$$

$\boldsymbol{o}\left(\frac{1}{n}\right)$ must be expressly stated, where $\boldsymbol{o}\,(h)$ denotes $\frac{o(h)}{h} \to 0$ as $h \to 0$.

The limit of $\log(y_n)$ as n approaches infinite is the number a, written

$$\lim_{n\to\infty} \log(y_n) = \lim_{n\to\infty} n\left(\frac{a}{n} + \boldsymbol{o}(\frac{1}{n})\right)$$

$$= a.$$

Now we can easily find that

$$\lim_{n\to\infty} (y_n) = e^a.$$

Then

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

□

Theorem 7. *Next we prove, if $S_n \sim$ Binomial(n,p), for each value $x =$ $0, 1, 2, \ldots,$ and as $n \to \infty$, $p \to 0$ with $np = \mu$ constant, then*

$$\lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{e^{-\mu}\mu^x}{x!} \quad \textit{for fixed } x = 0, 1, 2, \ldots.$$

Proof.

- Method I

Let $\mu = np$,

$$\begin{aligned}\binom{n}{x} p^x (1-p)^{n-x} &= \frac{n!}{x!(n-x)!}\left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x} \\ &= \frac{\mu^x}{x!}\left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x} \\ &= \frac{\mu^x}{x!}\underbrace{\left(\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n}\right)}_{\text{x times}}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty}\binom{n}{x}p^x(1-p)^{n-x}\frac{n!}{x!(n-x)!} &= \frac{\mu^x}{x!}\lim_{n\to\infty}\left[\left(\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n}\right)\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}\right] \\
&= \frac{\mu^x}{x!}\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{-x} \\
&= \frac{\mu^x}{x!}\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n \\
&= \frac{e^{-\mu}\mu^x}{x!}
\end{aligned}$$

Since

$$\lim_{n\to\infty}\left(\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n}\right) = \lim_{n\to\infty}\frac{n}{n}\lim_{n\to\infty}\frac{n-1}{n}\cdots\lim_{n\to\infty}\frac{n-x+1}{n}$$

$$= 1\cdot 1\ldots\cdot 1 = 1$$

and

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

Hence

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

□

- Method II

X has a Binomial(p) distributions, then the probability generating function (pgf) $g_n(s) = E(s^X)$ is $g_n(s) = (1-p+ps)^n$, for $|s| \le 1$, and Y has a Poisson(λ) distribution, the probability generating function (pgf) $g(s)$ of X is $g(s) = e^{-\lambda(1-s)}$, for all $s \in R$.If $n \to \infty$, $p \to 0$ such that $\lambda = np$, then $g_n(s) \to g(s)$.

Proof.

$$\text{If } \lambda = np, \text{then } p = \frac{\lambda}{n}$$

$$g_n(s) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}s\right)^n = \{1 - \lambda(1-s)/n\}^n$$

$$\lim_{n\to\infty} \{1 - \lambda(1-s)/n\}^n = e^{-\lambda(1-s)}.$$

□

Now we rewrite the prove of Shey(1984), it can be easily used to show the Poisson approximation to the Binomial distribution uniformly in x.

Theorem 8. *Let S_n have the Binomial(n,p) distribution and Y_n have the Poisson(np) distribution, then*

$$\lim_{n\to\infty} \sum_{x=0}^{\infty} |P(S_n = x) - P(Y_n = x)| = 0, \quad \textit{for all } x = 0, 1, 2, \ldots.$$

Proof. First, we show that $\sum_{x=0}^{\infty} |P(S_n = x|n,p) - P(Y_n = x|\lambda = np)| \leq 2np^2$

For n = 1,

$$\sum_{x=0}^{\infty} |P(S_n = x|n,p) - P(Y_n = x|\lambda = np)|$$

$$= \left|\binom{1}{0}p^0(1-p)^{1-0} - \frac{e^{-p}p^0}{0!}\right| + \left|\binom{1}{1}p^1(1-p)^{1-1} - \frac{e^{-p}p^1}{1!}\right| + \sum_{x=2}^{\infty} \frac{e^{-p}p^x}{x!}$$

$$= \left|(1-p) - e^{-p}\right| + \left|p - pe^{-p}\right| + [1 - P(Y = 0|\lambda = p) - P(Y = 1|\lambda = p)]$$

$$= \left[e^{-p} - (1-p)\right] + \left[p - pe^{-p}\right] + \left[1 - e^{-p} - pe^{-p}\right]$$

$$= 2p - 2pe^{-p}.$$

$$= 2p(1 - e^{-p}) \leq 2p^2, \quad \text{because } 1 - p < e^{-p}.$$

For general n, we use the identities as follows:

$$P(S_n = x) = (1-p)P(S_{n-1} = x) + (p)P(S_{n-1} = x-1)$$

and

$$P(Y_n = x) = \frac{e^{-np}(np)^x}{x!} = \sum_{j=0}^{x} \frac{e^{-p}(p)^j}{j!} P(Y_{n-1} = x-j)$$

then

$$\begin{aligned}
&|P(S_n = x) - P(Y_n = x)| \\
&\leq \left|(1-p)P(S_{n-1} = x) - e^{-p}P(Y_{n-1} = x)\right| \\
&+ \left|(p)P(S_{n-1} = x-1) - pe^{-p}P(Y_{n-1} = x-1)\right| \\
&+ \sum_{j=2}^{x} \frac{e^{-p}p^j}{j!} P(Y_{n-1} = x-j) \\
&\leq \left|1 - p - e^{-p}\right| P(Y_{n-1} = x) \\
&+ \left|p - pe^{-p}\right| P(Y_{n-1} = x-1) \\
&+ (1-p)\left|P(S_n = x) - P(Y_{n-1} = x)\right| \\
&+ (p)\left|P(S_{n-1} = x-1) - P(Y_{n-1} = x-1)\right| \\
&+ \sum_{j=2}^{x} \frac{e^{-p}p^j}{j!} P(Y_{n-1} = x-j)
\end{aligned}$$

For $x = 1, 2, \ldots n$, then by induction,

$$\sum_{x=0}^{\infty} |P(S_n = x) - P(Y_n = x)|$$

$$\leq \left[e^{-p} - (1-p)\right] + \left[p - pe^{-p}\right] + \left[1 - e^{-p} - pe^{-p}\right]$$

$$+ \sum_{x=0}^{\infty} |P(S_n = x) - P(Y_{n-1} = x)|$$

$$\leq 2p^2 + 2(n-1)p^2$$

$$= 2np^2.$$

Next, we show that

$$\sum_{x=0}^{\infty} |P(S_n = x) - P(Y_n = x)| \leq 3p$$

Now, we assume $2np \geq 3$ and $3p < 2$, following Le Cam(1965).

Let

$$\begin{aligned}\psi(x) &= \frac{P(Y_n = n-x)}{P(S_n = n-x)} \\ &= \frac{e^{-(np)}(np)^{n-(n-x)}/(n-x)!}{\frac{n!}{(n-x)!x!}p^{n-x}(1-p)^x} \\ &= \frac{x!e^{-(np)n^n}}{n!(nq)^x}, \quad \text{where } x = 0, 1, 2, \ldots n, q = 1-p\end{aligned}$$

$\psi(x)$ is minimum if $x = [nq] = [n(1-p)]$, and $[nq]$ is a integer such that $nq - 1 < [nq] \leq nq$.

Let $g = [nq] = nq - \epsilon, 0 \leq \epsilon < 1$.

Using the inequalities in Feller(1968) as follows:

$$\sqrt{2\pi n}\ n^n e^{-n} e^{(12n+1)^{-1}} < n! < \sqrt{2\pi n}\ n^n e^{-n} e^{(12n)^{-1}}$$

$$e^{-t/1-t} \leq 1-t, 0 \leq t \leq 1$$

then

$$\psi(x) \geq \psi(g) \geq \sqrt{g/n}\ e^{n-np-g}(g/nq)^g = \sqrt{g/n}\ e^{\epsilon}e^{-\epsilon} = \sqrt{g/n}$$

where

$$\psi(x) = \frac{x!e^{-np}n^n}{n!(nq)^x}$$
$$\psi(g) = \frac{g!e^{-np}n^n}{n!(nq)^g}$$

Hence

$$\sum_{x=0}^{\infty} |P(S_n = x) - P(Y_n = x)|$$
$$= 2 \sum_{\{X:\ x\in A\}} |P(S_n = x) - P(Y_n = x)|$$
$$\leq 2(1-\sqrt{g/n})$$

where $A = \{x : P(S_n = x) - P(S_n = x|\lambda = np) > 0\}$

Also, we prove

$$2(1-\sqrt{g/n}) \leq 3p$$

Consider $1 - \frac{g}{n} = 1 - \left(\frac{nq-\epsilon}{n}\right) = 1 - \left(q - \frac{\epsilon}{n}\right) = 1 - (1 - p - \frac{\epsilon}{n}) = p + \frac{\epsilon}{n}$,

then we have

$$1 - \frac{g}{n} \leq p + \frac{\epsilon}{n} \leq p + \frac{1}{n} \leq p + \frac{2}{3}p = \frac{5}{3}p.$$

Since

$$\frac{1}{n} \leq \frac{2}{3}p, \text{ and } p \leq \frac{2}{3}$$

then

$$\frac{1}{n} \leq \frac{2}{3}p \leq \frac{4}{9}.$$

We only consider the situation for $n \geq 3$.

If $n = 3$, we have

$n(1-p) = nq \geq 1$, and $p \leq \frac{2}{3}, 1-p = q \geq \frac{1}{3}$, then

$$\frac{(10/3)p}{1+\sqrt{[nq]/n}} \leq \frac{(10/3)p}{1+\sqrt{1/3}} \leq 3p$$

If $n \geq 4$, we have

$$\frac{[nq]}{n} \geq q - \frac{1}{n} \geq \frac{1}{12},$$

then

$$\frac{(10/3)p}{1+\sqrt{[nq]/n}} \leq \frac{(10/3)p}{1+\sqrt{1/12}} \leq 3p.$$

□

Corollary 1. *If* $np_n \to \lambda, 0 < \lambda < \infty$, *then* $P(X_n = x|n, p_n) \to P(Y_n = x|\lambda = np_n)$ *uniformly in* x *as* $n \to \infty$.

Proof. Let $\lambda_n = np_n$, by the above theorem,

$$P(X_n = x|n, p_n) - P(Y_n = x|\lambda = np_n) \to 0$$

uniformly in x. □

Next we consider the non-negative integer-valued random variables.

3.2 Limits of the Negative Binomial Distribution

Theorem 9. *Let X_r have the negative binomial (n,p) distributions with density $P(X = x|r,p) = \binom{r+x-1}{x} p^r (1-p)^x$ then $P(X = x|r,p) \to \frac{e^{-\lambda}\lambda^x}{x!}$, as $r(1-p) = \lambda$, and $r \to \infty$.*

Proof. • Method I

Let $\lambda = r(1-p)$, then $\left(1 - \frac{\lambda}{r}\right) = p$.

We write

$$\begin{aligned}\binom{r+x-1}{x} p^r (1-p)^x &= \frac{(r+x-1)!}{x!\,(r-1)!}\left(1-\frac{\lambda}{r}\right)^r \left(\frac{\lambda}{r}\right)^x \\ &= \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{r}\right)^r \frac{(r+x-1)!}{(r-1)!}\left(\frac{1}{r}\right)^x.\end{aligned}$$

Then

$$\begin{aligned}\lim_{n\to\infty}\binom{r+x-1}{x} p^r (1-p)^x &= \frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{r}\right)^r \frac{(r+x-1)!}{(r-1)!}\left(\frac{1}{r}\right)^x \\ &= \frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{r}\right)^r \frac{(r+x-1)(r+x-2)\cdots n}{n^x} \\ &= \frac{\lambda^x}{x!}e^{-\lambda}.\end{aligned}$$

• Method II

Now we write the term $p/(1-(1-p)s)$ as $p/[p+(1-p)(1-s)]^{-1} = [1+(1-p)(1-s)/p]^{-1}$, so that the probability generating function $g_r(s)$ as

$$[1+(1-p)(1-s)/p]^{-r} = \left[1 + \frac{\lambda}{r}\frac{(1-s)}{p}\right]^{-r}.$$

which converges, under the conditions of limit to $e^{-\lambda(1-s)}$ which is the probability generating function of the Poisson distribution with parameter λ (note $p \to 1$, as $r \to \infty$). □

3.3 Limits of the Hypergeometric Distributions

In the next subsections, we consider the hypergeometric random variables, which is the distribution of sums of dependent random variables.

Theorem 10. *The hypergeometric*(z, n_1, n_2) *distribution is defined as*

$$P(X = x|z, n_1, n_2) = \frac{\binom{n_1}{x}\binom{n_2}{z-x}}{\binom{n_1+n_2}{z}},$$

for $x = 0, 1, 2, \dots z$, *and* $z \leq \min(n_1, n_2)$. *We shall show that under the limiting conditions, (a)* $\min(n_1, n_2) \to \infty$, *(b)* $z\left(\frac{n_1}{n_1+n_2}\right)$, *as* $z \to \infty$, *it converges to the Poisson*(λ) *distribution.*

Proof. Denote $n = n_1 + n_2$, we write

$$\begin{aligned}
\frac{\binom{n_1}{x}\binom{n_2}{z-x}}{\binom{n}{z}} &= \frac{n_1!}{x!(n_1-x)!}\frac{n_2!}{(z-x)!(n_2-z+x)!}\frac{z!(n-z)!}{n!} \\
&= \binom{z}{x}\left[\left(\frac{n_1}{n}\right)\cdots\left(\frac{n_1-x+1}{n}\right)\right]\left[\left(\frac{n_2}{n}\right)\cdots\left(\frac{n_2-z+x+1}{n}\right)\right] \\
&\times \left[\left(\frac{n}{n}\right)\cdots\left(\frac{n}{n-z+1}\right)\right].
\end{aligned}$$

So that, for fixed x and z, if

$$\frac{n_1}{n_1+n_2} \to p \text{ as } \min(n_1, n_2) \to \infty,$$

then

$$\frac{\binom{n_1}{x}\binom{n_2}{z-x}}{\binom{n}{z}} \to \binom{z}{x} p^x (1-p)^{z-x},$$

the binomial(z,p) distribution, because under the condition $\frac{n_1}{n} \to p$ as $\min(n_1,n_2) \to \infty$,

$$\frac{n_1 - k}{n} \to p$$

and

$$\frac{n_2 - k}{n} \to 1 - p,$$

for all fixed k.

Now if $zp \to \lambda$, as $z \to \infty$, then $\binom{z}{x} p^x (1-p)^{z-x}$ converges to the Poisson(λ) distribution. Since $z \leq \min(n_1,n_2)$, the limit "$z \to \infty$" must be preceded by the limit "$\min(n_1,n_2) \to \infty$". From this we conclude that the hypergeometric(z,n_1,n_2) distribution converges to the Poisson(λ) distribution under the conditions $\min(n_1,n_2) \to \infty$, and $z\left(\frac{n_1}{n_1+n_2}\right) \to \lambda$ as $z \to \infty$ □

CHAPTER 4 THE POISSON APPROXIMATION FOR SUMS OF INDEPENDENT BUT NON-IDENTICALLY DISTRIBUTED RANDOM VARIABLES

Let $X_1, X_2, \ldots X_n$ be independent Bernoulli random variables with $p_i = P(X_i = 1) = 1 - P(X_i = 0), 0 < p_i < 1, i = 1, 2, \ldots, n,$ and $Y_1, \ldots, Y_n$ be independent Poisson random variables with expectations $\lambda_i, i = 1, 2, \ldots$. Let $U_n = \sum_{i=1}^{n} X_i$ and $V_n = \sum_{i=1}^{n} Y_i$. Now we consider the total variation distance which is the approximation of the distribution of U_n by the distribution of V_n as follows:

$$\begin{aligned} d\left(U_n, V_n\right) &= \sup_A |P(U_n \in A) - P(V_n \in A)| \\ &= \frac{1}{2} \sum_{k=0}^{\infty} |P(U_n = k) - P(V_n = k)|. \end{aligned}$$

Estimations of this distance have been given by different authors, such as Le Cam(1960), and recently by Yannaros(1991) and Wang(1993).

Lemma 7. *Let* $|\theta| < 1$*, then* $-\ln(1-\theta) = \theta + \theta^2 K(\theta)$*, where* $K(\theta) < 1$ *if* $|\theta| < \frac{1}{2}$*.*

Proof.

$$-\ln(1-\theta) = \theta + \frac{1}{2}\theta^2 + \frac{1}{3}\theta^3 + \ldots, \quad |\theta| < 1$$

$$K(\theta) = \sum_{k=0}^{\infty} \frac{\theta^k}{k+2} \leq \frac{1}{2(1-\theta)}.$$

$$\text{For } \; |\theta| < \frac{1}{2}, \;\; |K(\theta)| < 1.$$

□

Let X_i be independent $Bernoulli(\theta_i)$ distribution, and let Y be $Poisson(\lambda)$, $\lambda = \sum_{i=1}^{n} \theta_i, m_n = \max(\theta_1, \theta_2 \ldots, \theta_n)$

Theorem 11. *If* $m_n \to 0$*, then* $S_n = \sum_{i=1}^{n} X_i \to Y$ *in distribution.*

Proof.

$$g_n(s) = E(s^{S_n}) = \prod_{i=1}^{n} (1 - \theta_i(1-s)), |s| < 1$$

$$h(s) = E(s^Y) = e^{-\lambda(1-s)}, |s| < 1.$$

$$\ln g_n(s) = \sum_{i=1}^{n} (1 - \theta_i(1-s))$$
$$= -\sum_{i=1}^{n} \theta_i (1-s) + \sum_{i=1}^{n} \theta_i^2 K(\theta_i).$$

Take $m_n < \frac{1}{2}$, then $K(\theta_i) < 1$ and $\sum_{i=1}^{n} \theta_i^2 \leq \lambda m_n$, so that

$\ln g_n(s) = -\lambda(1-s) + \lambda m_n \to h(s)$ as $n \to \infty$, for all $|s| < 1$. □

Theorem 12. *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables, where $P(X_i = 1) = \theta_i$ and $P(X_i = 0) = 1 - \theta_i$, Y be Poisson(λ). Let $S_n = \sum_{i=1}^n X_i, \lambda = \sum_{i=1}^n \theta_i$, and $m_n = \max \theta_i$. If $m_n \to 0$, then $S_n = \sum_{i=1}^n X_i \to Y$ uniformly in x.*

Proof.

$$\lambda = \sum_{i=1}^{n} \theta_i, \text{ with } \theta_1 \ldots \theta_n \text{ assumed to be independent.}$$

Now comparing $P(S_n = k)$ with $P(Y = k)$, as a first step, we observe that if S_n and Y are unequal, then at least one of the pairs $P(X_i = i) \neq P(Y_i = i)$. In the following table, we define a joint distribution of X and Y.

		$Y \sim Poisson(\lambda)$						
		0	**1**	**2**	**3**	$\ldots$	$\ldots$	marginal of X
$X \sim Bernoulli(\theta_i)$	**0**	$1 - p$	0	0	0	$\ldots$	0	$(1 - p)$
	1	$e^{-p} - (1 - p)$	pe^{-p}	$\frac{p^2 e^{-2}}{2!}$	$\frac{p^3 e^{-3}}{3!}$	$\ldots$	$\ldots$	p
marginal of Y		e^{-p}	pe^{-p}	$\frac{p^2 e^{-2}}{2!}$	$\frac{p^3 e^{-3}}{3!}$	$\ldots$	$\ldots$	

By the table, we can calculate $|P(X_i) \neq P(Y_i)| \leq p_2^2$, where

$$|P(X_i) \neq P(Y_i)| = 1 - P(X_i = 0, Y_i = 0) - P(X_i = 1, Y_i = 1)$$

$$= 1 - (1 - p) - pe^{-p}$$

$$= p(1 - e^{-p}).$$

Obtain $|P(X_i) \neq P(Y_i)| = p_2(1 - e^{-p_2}) \leq p_2^2$.

According to the above result we can establish the stronger result

$$\sum_{k=0}^{\infty} |P(S_n = k) - P(Y = k)| \leq \sum_{k=1}^{n} p_i^2$$
$$\leq (\max(p_i))\lambda \to 0 \text{ as } p_i \to 0.$$

So that $S_n \to Y$ uniformly in k. □

BIBLIOGRAPHY

Bernard, W. L. (1993). *Statistical Theory.* Chapman-Hall, New York.

Bortkiewicz, L. von. (1898). *Dzs Gesetz der Kleinen Zahlen.* Leipzig: Teubner.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications.* Vol. 1. Wiley, New York.

Ghahramani, S. (2000). *Fundamentals of Probability.* Prentice-Hall, New Jersey.

Good, I. J. (1986). Some Statistical Applications of Poisson's work. *Statist. Sci.* **1** 157-180.

Le Cam, L. (1965). *Bernoulli, Bayes, Laplace Anniversary Volume, eds.*. Berkeley: University of California Press.

Morris L. M. and Richard J. L. (1986). *An Introduction to Mathematical Statistics and Its Applications.* Prentice-Hall, New Jersey.

Norman L. J. and Samuel K. (1969). *Discrete Distributions.* Wiley, New York.

Poisson S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en mtière civile, précedées des règles générales du calcul des probabilités.*

Shey, S. S. (1984). The Poisson Approximation to the Binomial Distribution. *Amer. Statist.* **38** 206-207.

Wang, Y. H. (1986). Coupling Methods in Approximations. *Cana. J. Statist.* **14** 69-74.

Wang, Y. H. (1993). On the Number of Successes in Independent Trials. *Statist. Sinica* **3** 295-312.

Yannaros, N. (1991). Poisson Approximation for Random Sums of Bernoulli Random Variables. *Statist. Probab. Lett.* **11** 161-165.

Ehm, W. (1991). Binomial Approximation to the Poisson Binomial Distribution. *Statist. Probab. Lett.* **11** 7-16.