Chapter 1

A generalized binomial distribution determined by a two-state Markov chain and a distribution by the Bayesian approach

1.1. Introduction

For a sequence of independent Bernoulli random variables, the distribution of a sum of the random variables is well known to be a binomial one. In relation to a bivariate Bernoulli distribution, bivariate binomial distributions are discussed by Kocherlakota and Kocherlakota [KK92]. In this chapter, following [AKT97], we consider a problem how the distribution of the sum Y_n of dependent Bernoulli random variables $X_1, \ldots, X_n$ is determined, and show that the sum Y_n is distributed as a generalized binomial distribution based on a two-state Markov chain. We also obtain the approximations of the distribution of Y_n using the central limit theorem and the Edgeworth expansion and numerically compare the generalized binomial distribution with the asymptotic distributions. Further we derive a distribution of the sum Y_n by the Bayesian approach and obtain asymptotic distributions similar to the Markov case. Numerical results are given.

1.2. An approach by a two-state Markov chain

We consider a generalized binomial distribution based on a two-state Markov chain. Suppose that $X_1, X_2, \ldots$ is a sequence of random variables each taking the values 0 and 1, with the joint distribution determined by the initial probability $P\{X_1 = 1\} = p$, and the transition probabilities

$$P\{X_{i+1} = 1 \mid X_i = 0\} = \pi_0, \quad P\{X_{i+1} = 1 \mid X_i = 1\} = \pi_1,$$

for $i=1,2,\ldots,$ of which we assume $0<\pi_i<1$ $(i=0,1)$. In such a situation the consistency of the sample mean is discussed in Lehmann ([Le83]). We also assume that $\{X_i\}$ is strongly stationary, i.e. $P\{X_i=1\}=p$ for $i=1,2,\ldots.$ Then we have $p=\pi_0/(1+\pi_0-\pi_1)$, since $P\{X_i=1\}=\pi_0(1-p)+\pi_1 p$.

Put $S_j:=\sum_{i=j}^{n} X_i$ $(j=1,\ldots,$ n-1). For $j=2,\ldots,n,$ we have

$$
\begin{aligned}
&P\{S_j=k \mid X_{j-1}=0\}\\
&=P\{X_j=0,S_j=k \mid X_{j-1}=0\}+P\{X_j=1,S_j=k \mid X_{j-1}=0\}\\
&=P\{X_j=0,S_{j+1}=k \mid X_{j-1}=0\}+P\{X_j=1,S_{j+1}=k-1 \mid X_{j-1}=0\}\\
&=P\{S_{j+1}=k \mid X_j=0\}P\{X_j=0 \mid X_{j-1}=0\}\\
&\quad+P\{S_{j+1}=k-1 \mid X_j=1\}P\{X_j=1 \mid X_{j-1}=0\}\\
&=P\{S_{j+1}=k \mid X_j=0\}(1-\pi_0)+P\{S_{j+1}=k-1 \mid X_j=1\}\pi_0. \qquad (1.1)
\end{aligned}
$$

Putting $P_j(k,0):=P\{S_j=k \mid X_{j-1}=0\}$, $P_j(k,1):=P\{S_j=k \mid X_{j-1}=1\}$ for $j=2,\ldots,n,$ we obtain from (1.1)

$$P_j(k,0)=P_{j+1}(k,0)(1-\pi_0)+P_{j+1}(k-1,1)\pi_0. \qquad (1.2)$$

In a similar way to the above we have

$$P_j(k,1)=P_{j+1}(k,0)(1-\pi_1)+P_{j+1}(k-1,1)\pi_1. \qquad (1.3)$$

We easily see that

$$
\begin{aligned}
&P_n(1,0)=\pi_0,\ P_n(0,0)=1-\pi_0,\ P_n(1,1)=\pi_1,\ P_n(0,1)=1-\pi_1,\\
&P_n(k,0)=0\ (k=2,\ldots,n),\ P_n(k,1)=0\ (k=2,\ldots,n).
\end{aligned} \qquad (1.4)
$$

We also define $P_j(\alpha,1)=0$ and $P_j(\alpha,0)=0$ for $\alpha<0$. Putting $Y_n=S_1$, we have for $k=0,1,\ldots,n$

$$
\begin{aligned}
P\{Y_n=k\}&=P_2(k,0)(1-p)+P_2(k-1,1)p\\
&=\Big\{P_3(k,0)(1-\pi_0)+P_3(k-1,1)\pi_0\Big\}(1-p)\\
&\quad+\Big\{P_3(k-1,0)(1-\pi_1)+P_3(k-2,1)\pi_1\Big\}p\\
&=\cdots. \qquad (1.5)
\end{aligned}
$$

From (1.2) to (1.5) it follows that the distribution of Y_n can be recursively obtained as a generalized binomial distribution.

Next we shall obtain the asymptotic distribution of Y_n as $n \to \infty$. First we have

$$E(Y_n) = np = \frac{n\pi_0}{1+\pi_0-\pi_1}. \tag{1.6}$$

Since, for $j > i+1$

$$P\{X_j = 1 \mid X_i = 1\} = P\{X_{j-1} = 1 \mid X_i = 0\}(1-\pi_1) + P\{X_{j-1} = 1 \mid X_i = 1\}\pi_1,$$
$$P\{X_j = 1 \mid X_i = 0\} = P\{X_{j-1} = 1 \mid X_i = 0\}(1-\pi_0) + P\{X_{j-1} = 1 \mid X_i = 1\}\pi_0,$$

and for $j = i+1$

$$P\{X_j = 1 \mid X_i = 1\} = \pi_1, \; P\{X_j = 1 \mid X_i = 0\} = \pi_0,$$

it follows that, for $i < j$

$$\begin{aligned} &P\{X_j = 1 \mid X_i = 1\} - P\{X_j = 1 \mid X_i = 0\} \\ &\quad = (\pi_1 - \pi_0)^{j-i-1}\Big[P\{X_{i+1} = 1 \mid X_i = 1\} - P\{X_{i+1} = 1 \mid X_i = 0\}\Big] \\ &\quad = (\pi_1 - \pi_0)^{j-i}. \end{aligned} \tag{1.7}$$

On the other hand, since, for $i < j$

$$p = E(X_j) = E[E(X_j \mid X_i)] = P\{X_j = 1 \mid X_i = 1\}p + P\{X_j = 1 \mid X_i = 0\}(1-p),$$

it follows from (1.7) that

$$P\{X_j = 1 \mid X_i = 0\} = p - p(\pi_1 - \pi_0)^{j-i},$$
$$P\{X_j = 1 \mid X_i = 1\} = p + (1-p)(\pi_1 - \pi_0)^{j-i},$$

where $p = \pi_0/(1+\pi_0-\pi_1)$. Hence, for $i < j$

$$\begin{aligned} E(X_iX_j) &= P\{X_i = 1, X_j = 1\} = P\{X_j = 1 \mid X_i = 1\}P\{X_i = 1\} \\ &= p^2 + p(1-p)(\pi_1 - \pi_0)^{j-i}. \end{aligned} \tag{1.8}$$

From (1.6) and (1.8) we have

$$\begin{aligned} Var(Y_n) &= E\left[(Y_n - np)^2\right] \\ &= \sum_{i=1}^{n} E\left[(X_i - p)^2\right] + \sum\sum_{i \neq j} E\big[(X_i - p)(X_j - p)\big] \\ &= np(1-p) + \sum\sum_{i \neq j} \left\{E(X_iX_j) - p^2\right\} \end{aligned}$$

$$
\begin{aligned}
&= np(1-p) + p(1-p)\sum\sum_{i \neq j}(\pi_1 - \pi_0)^{|j-i|} \\
&= np(1-p) + p(1-p)\frac{2\delta}{1-\delta}\left(n - \frac{1-\delta^n}{1-\delta}\right) \\
&= np(1-p)\frac{1+\delta}{1-\delta} - 2p(1-p)\frac{\delta(1-\delta^n)}{1-\delta},
\end{aligned}
$$

where $\delta = \pi_1 - \pi_0$, hence for sufficiently large n

$$
Var(Y_n) = np(1-p)\frac{1+\delta}{1-\delta} + O(1) \tag{1.9}
$$

From the central limit theorem, (1.6) and (1.9) it follows that the distribution of Y_n is asymptotically normal with mean np and variance $np(1-p)(1+\delta)/(1-\delta)$ for sufficiently large n, where $p = \pi_0/(1+\pi_0-\pi_1)$ and $\delta = \pi_1 - \pi_0$.

Further we consider a higher order asymptotic distribution of Y_n using the Edgeworth expansion. In order to do so we shall obtain the third cumulant $\kappa_3(Y_n)$ of Y_n. First we have

$$
\begin{aligned}
\kappa_3(Y_n) :&= E\left[(Y_n - np)^3\right] \\
&= \sum_{i=1}^{n} E\left[(X_i - p)^3\right] + 3\sum\sum_{i \neq j} E\left[(X_i - p)^2(X_j - p)\right] \\
&\quad + \sum\sum\sum_{i \neq j \neq k} E\big[(X_i - p)(X_j - p)(X_k - p)\big].
\end{aligned} \tag{1.10}
$$

We also obtain

$$
E\left[(X_i - p)^3\right] = p(1-p)(1-2p) \qquad (i = 1, \ldots, n) \tag{1.11}
$$

and

$$
E\left[(X_i - p)^2(X_j - p)\right] = p(1-p)(1-2p)\delta^{|i-j|} \qquad (i \neq j), \tag{1.12}
$$

where $\delta = \pi_1 - \pi_0$. Since $i < j < k$

$$
E\left[(X_i - p)(X_j - p)(X_k - p)\right] = p(1-p)(1-2p)\delta^{k-i},
$$

it follows that

$$
\begin{aligned}
&\sum\sum\sum_{i \neq j \neq k} E\big[(X_i - p)(X_j - p)(X_k - p)\big] \\
&= 6\sum\sum\sum_{j<i<k} p(1-p)(1-2p)\delta^{k-j} \\
&= 6p(1-p)(1-2p)\frac{\delta}{1-\delta}\left(\sum\sum_{j<i}\delta^{i-j} - \sum\sum_{j<i}\delta^{n-j}\right).
\end{aligned} \tag{1.13}
$$

Since

$$\sum\sum_{j<i} \delta^{i-j} = \frac{\delta}{1-\delta}\left(n - \frac{1-\delta^n}{1-\delta}\right)$$

and

$$\sum\sum_{j<i} \delta^{n-j} = -\frac{(n-1)\delta^n}{1-\delta} - \frac{\delta^n - \delta}{(1-\delta)^2},$$

it follows from (1.13) that

$$\begin{aligned}
&\sum\sum\sum_{i\neq j\neq k} E\big[(X_i - p)(X_j - p)(X_k - p)\big] \qquad (1.14)\\
&= 6p(1-p)(1-2p)\frac{\delta}{1-\delta}\left\{\frac{\delta}{1-\delta}\left(n - \frac{1-\delta^n}{1-\delta}\right) + \frac{(n-1)\delta^n}{1-\delta} + \frac{\delta^n-\delta}{(1-\delta)^2}\right\}.
\end{aligned}$$

From (1.10), (1.11), (1.12) and (1.14) we have

$$\begin{aligned}
\kappa_3(Y_n) &= np(1-p)(1-2p) + 3p(1-p)(1-2p)\frac{2\delta}{1-\delta}\left(n - \frac{1-\delta^n}{1-\delta}\right)\\
&\quad + 6p(1-p)(1-2p)\frac{\delta}{1-\delta}\left\{\frac{\delta}{1-\delta}\left(n - \frac{1-\delta^n}{1-\delta}\right) + \frac{(n-1)\delta^n}{1-\delta} + \frac{\delta^n-\delta}{(1-\delta)^2}\right\}\\
&= np(1-p)(1-2p)\left\{1 + \frac{6\delta}{1-\delta} + \frac{6\delta^2}{(1-\delta)^2}\right\} + O(1). \qquad (1.15)
\end{aligned}$$

Then it follows from (1.9) and (1.15) that the Edgeworth expansion of the distribution $F_{Y_n}(y)$ of Y_n is given by

$$\begin{aligned}
F_{Y_n}(y) &= P\{Y_n \leq y\} = P\left\{\frac{Y_n - np}{\sqrt{Var(Y_n)}} \leq \frac{y - np}{\sqrt{Var(Y_n)}}\right\}\\
&= \Phi\left(\frac{y + 0.5 - np}{\sqrt{v_n}}\right) - \frac{1}{6{v_n}^{3/2}} np(1-p)(1-2p)\\
&\quad \cdot \left\{1 + \frac{6\delta}{1-\delta} + \frac{6\delta^2}{(1-\delta)^2}\right\}\left\{\frac{(y+0.5-np)^2}{v_n} - 1\right\}\phi\left(\frac{y+0.5-np}{\sqrt{v_n}}\right) + o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

where

$$\Phi(t) = \int_{-\infty}^{t} \phi(x)\, dx \quad \text{with} \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and

$$v_n = np(1-p)\frac{1+\delta}{1-\delta}.$$

1.3. A numerical comparison

In this section we numerically compare the exact distribution with the asymptotic normal distribution of Y_n and its Edgeworth expansion given in Section 1.2. We consider the cases when (i) $\pi_0 = 0.2$ and $\pi_1 = 0.7$ and (ii) $\pi_0 = 0.2$ and $\pi_1 = 0.1$. In each case we give the errors between exact and asymptotic normal distributions of Y_n after continuity correction for $n = 10$ and these among exact, asymptotic normal ones of Y_n and its Edgeworth expansion after continuity correction for $n = 25$ as Tables 1.1 and 1.2 and their cumulative distribution functions (c.d.f.'s) as Figures 1.1 to 1.4. The tables and figures show that the approximations by the central limit theorem and the Edgeworth expansion are useful for each case ([AKT97]).

Table 1.1: Errors of the asymptotic normal c.d.f. of Y_n in the case (i)π_0=0.2, π_1=0.7 and (ii)π_0=0.2, π_1=0.1 for n=10

$n = 10$, π_0=0.2, π_1=0.7				$n = 10$, π_0=0.2, π_1=0.1			
k	Exact	Normal	Error	k	Exact	Normal	Error
0	8.0531 %	9.6053 %	1.5523 %	0	10.9815 %	11.6076 %	0.6262 %
1	18.1194	17.5747	−0.5447	1	41.1804	38.6517	−2.5287
2	30.8989	28.2075	−2.0914	2	74.3071	73.1718	−1.1353
3	45.0989	42.6089	−2.4900	3	92.9871	93.6301	0.6430
4	59.2452	57.3911	−1.8542	4	98.8441	99.2469	0.4028
5	72.0165	71.1925	−0.8240	5	99.8881	99.9577	0.0696
6	82.4930	82.4253	−0.0677	6	99.9938	99.9989	0.0051
7	90.2681	90.3947	0.1266	7	99.9998	100.0000	0.0002
8	95.4211	95.3234	−0.0977	8	100.0000	100.0000	0.0000
9	98.3859	97.9805	−0.4054	9	100.0000	100.0000	0.0000
10	100.0000	99.2291	−0.7709	10	100.0000	100.0000	0.0000

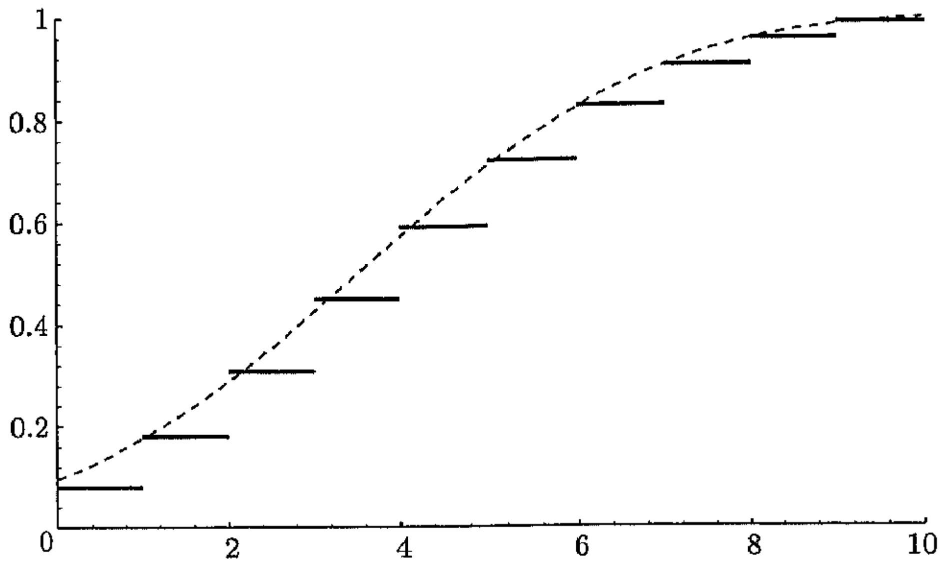

Figure 1.1: The exact and asymptotic normal c.d.f.'s of Y_n in the case (i) $\pi_0 = 0.2$, $\pi_1 = 0.7$ for $n = 10$

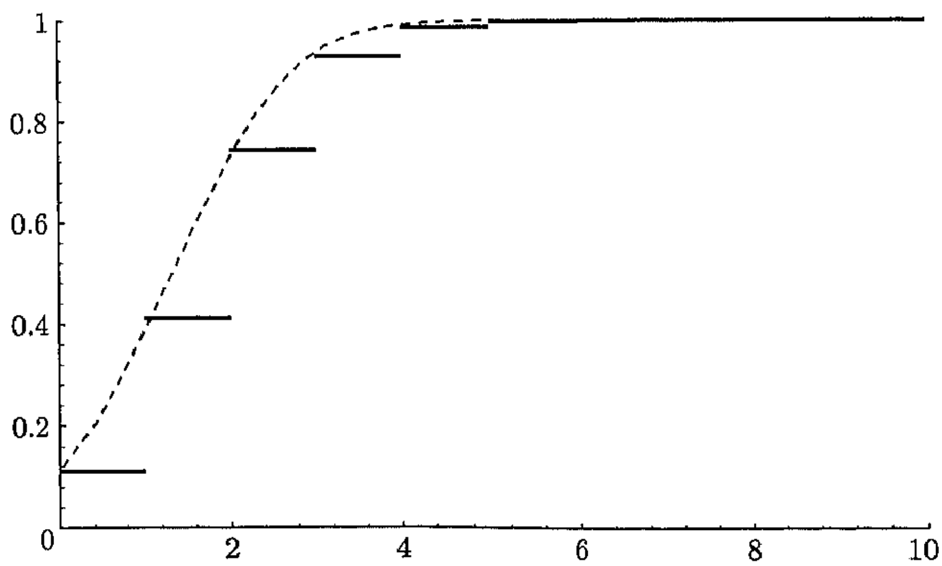

Figure 1.2: The exact and asymptotic normal c.d.f.'s of Y_n in the case (ii) $\pi_0 = 0.2$, $\pi_1 = 0.1$ for $n = 10$

Table 1.2: Errors of the asymptotic normal c.d.f. of Y_n and its Edgeworth expansion in the case (i)$\pi_0=0.2$, $\pi_1=0.7$ and (ii)$\pi_0=0.2$, $\pi_1=0.1$ for $n=25$

$n=25$, $\pi_0=0.2$, $\pi_1=0.7$				$n=25$, $\pi_0=0.2$, $\pi_1=0.1$			
k	Exact	Normal	Edgeworth	k	Exact	Normal	Edgeworth
0	0.2833 %	0.9739 %	0.5295 %	0	0.3864 %	0.6330 %	0.1884 %
1	1.0360	1.2204	0.6702	1	3.0789	0.9627	0.2959
2	2.5564	1.2986	0.6936	2	11.6814	0.3663	0.0845
3	5.1331	1.1422	0.5763	3	28.4001	−0.9525	−0.1522
4	8.9971	0.7455	0.3465	4	50.5192	−1.5586	−0.0655
5	14.2694	0.1728	0.0761	5	71.6499	−0.8614	0.0400
6	20.9258	−0.4560	−0.1473	6	86.7437	0.1310	−0.0728
7	28.7860	−1.0015	−0.2562	7	94.9840	0.5005	−0.1650
8	37.5312	−1.3475	−0.2311	8	98.4714	0.3592	−0.1144
9	46.7456	−1.4364	−0.1062	9	99.6260	0.1487	−0.0384
10	55.9727	−1.2820	0.0481	10	99.9267	0.0413	−0.0057
11	64.7749	−0.9586	0.1578	11	99.9885	0.0081	0.0003
12	72.7862	−0.5706	0.1747	12	99.9986	0.0012	0.0003
13	79.7471	−0.2168	0.0919	13	99.9999	0.0001	0.0001
14	85.5210	0.0368	−0.0599	14	100.0000	0.0000	0.0000
15	90.0895	0.1680	−0.2309	15	100.0000	0.0000	0.0000
16	93.5319	0.1927	−0.3731	16	100.0000	0.0000	0.0000
17	95.9961	0.1489	−0.4561	17	100.0000	0.0000	0.0000
18	97.6654	0.0782	−0.4719	18	100.0000	0.0000	0.0000
19	98.7300	0.0128	−0.4316	19	100.0000	0.0000	0.0000
20	99.3642	−0.0306	−0.3561	20	100.0000	0.0000	0.0000
21	99.7131	−0.0490	−0.2678	21	100.0000	0.0000	0.0000
22	99.8874	−0.0483	−0.1842	22	100.0000	0.0000	0.0000
23	99.9642	−0.0373	−0.1158	23	100.0000	0.0000	0.0000
24	99.9923	−0.0239	−0.0661	24	100.0000	0.0000	0.0000
25	100.0000	−0.0129	−0.0341	25	100.0000	0.0000	0.0000

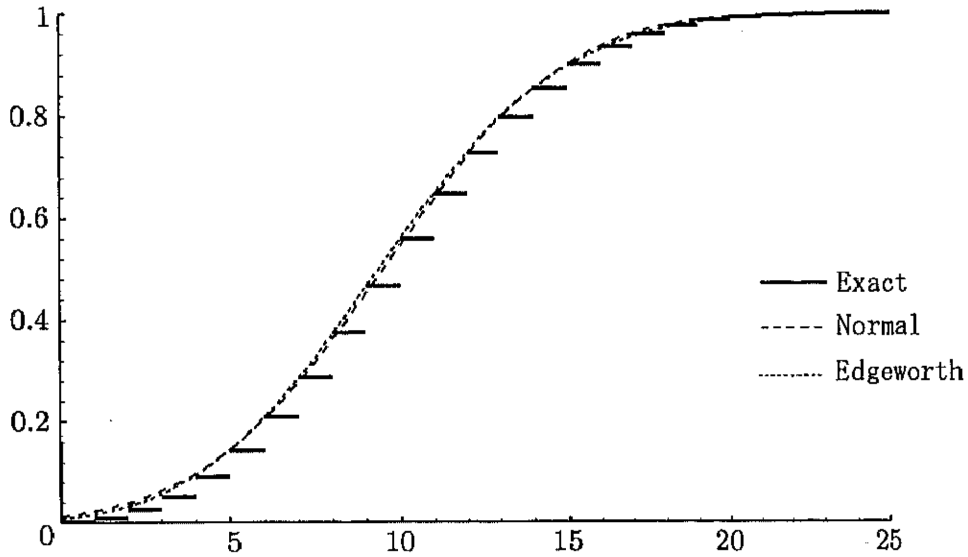

Figure 1.3: The exact and asymptotic normal c.d.f.'s of Y_n in the case (i) $\pi_0=0.2$, $\pi_1=0.7$ for $n=25$

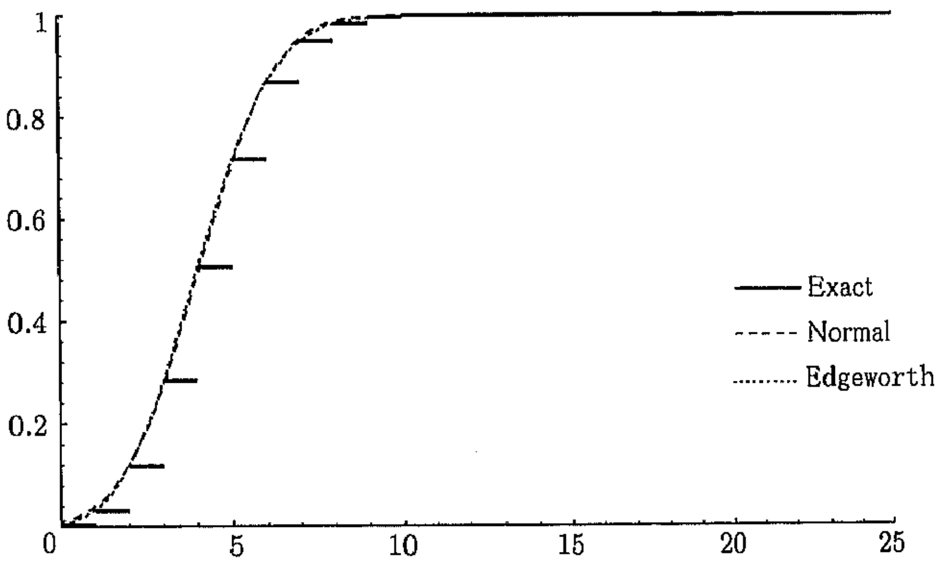

Figure 1.4: The exact and asymptotic normal c.d.f.'s of Y_n in the case (i) $\pi_0=0.2$, $\pi_1=0.1$ for $n=25$

1.4. A Bayesian approach

We assume that $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of identically distributed random variables each taking on the values 0 and 1 with the probability $P\{X_1=1\}=p$, and p is also a random variable. Suppose that $X_1, \ldots, X_n$ are conditionally independent given p, and a prior density $\pi(p)$ of p is given by

$$\pi(p) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1} & \text{for } 0<p<1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha, \beta > 0$ and $B(\alpha,\beta)$ is a Beta function. Then, for each $i=1,2,\ldots$, the mean of X_i is given by

$$\mu := E(X_i) = E\Big[E(X_i \mid p)\Big] = E(p) = \frac{\alpha}{\alpha+\beta},$$

and, for $i \neq j$, the covariance of X_i and X_j is given by

$$\begin{aligned} Cov(X_i, X_j) &= E(X_iX_j) - E(X_i)E(X_j) \\ &= E\Big[E(X_iX_j \mid p)\Big] - \mu^2 \\ &= E\Big[E(X_i \mid p)E(X_j \mid p)\Big] - \mu^2 \\ &= E(p^2) - \mu^2 \\ &= \frac{1}{\alpha+\beta+1}\,\mu(1-\mu). \end{aligned}$$

Let $Y_n := X_1 + \cdots + X_n$. Since the conditional distribution of Y_n given p is a binomial one $B(n,p)$, it follows that the distribution of Y_n is given by

$$\begin{aligned} P\{Y_n = y\} &= E\Big[P\{Y_n = y \mid p\}\Big] = \int_0^1 \binom{n}{y} p^y (1-p)^{n-y}\pi(p)dp \\ &= \left\{\binom{n}{y}\Big/B(\alpha,\beta)\right\} \int_0^1 p^{y+\alpha-1}(1-p)^{n-y+\beta-1}dp \\ &= \binom{\alpha+y-1}{y}\binom{\beta+n-y-1}{n-y}\Big/\binom{\alpha+\beta+n-1}{n} \end{aligned} \tag{1.16}$$

for $y=0,1,\ldots,n$. It is known that for any $a \in \mathbf{R}$ and any non-negative integer r

$$\binom{-a}{r} = (-1)^r \binom{a+r-1}{r}. \tag{1.17}$$

Then it follows from (1.16) and (1.17) that

$$P\{Y_n=y\} = \binom{-\alpha}{y}\binom{-\beta}{n-y}\Big/\binom{-\alpha-\beta}{n} \tag{1.18}$$

for $y=0,1,\ldots,n$, which may be called a negative hypergeometric distribution. Note that the distribution of Y_n is given in Stuart and Ord ([SO94]) as the Beta-binomial distribution which coincides with the negative hypergeometric distribution (1.18).

Next we obtain the asymptotic distribution of Y_n as $n \to \infty$. In order to do so we consider a hypergeometric distribution with a probability mass function

$$p_H(x) = p_H(x:n,M,N) = \binom{M}{x}\binom{N-M}{n-x} \Big/ \binom{N}{n} \tag{1.19}$$

for $x=0,1,\ldots,\min\{n,M\}$, where n, M and N are positive integers with $M \le N$. Without loss of generality we assume that $n \le M \le N/2 \le N-M \le N-n$. Then the mean μ and variance σ^2 of the hypergeometric distribution are given by

$$\mu = \frac{nM}{N}, \qquad \sigma^2 = \frac{n(N-n)M(N-M)}{N^2(N-1)} \fallingdotseq \frac{n}{N^3}(N-n)M(N-M), \tag{1.20}$$

respectively (see [SO94] and [TF81]). Put $p = M/N$, $q = (N-M)/N$, $\gamma = n/N$ and $\delta' = (N-n)/N$. If $\mu = Np\gamma \to 0$, then

$$\begin{aligned} F_H(x:n,M,N) &:= \sum_{t=0}^{x} p_H(t:n,M,N) \\ &= \Phi(u) + o(1), \end{aligned} \tag{1.21}$$

where $u=(x+0.5-\mu)/\sigma$. Since the third cumulant κ_3 of the hypergeometric distribution is given by

$$\kappa_3 = \frac{(N-n)(N-2n)}{(N-1)(N-2)} npq(q-p),$$

it is seen in a similar way to Section 1.2 that the Edgeworth expansion of the distribution is

$$F_H(x:n,M,N) = \Phi(u) - \frac{1}{6\sigma}(q-p)(\delta'-\gamma)(u^2-1)\phi(u) + o\left(\frac{1}{\sigma}\right) \tag{1.22}$$

(see also [TF81]). On the other hand, the probability mass function of the negative hypergeometric distribution corresponding to the hypergeometric one (1.19) can be defined by

$$p_{NH}(x) = p_{NH}(x:k,M,N) := \frac{M-k+1}{N-k-x+1}\binom{M}{k-1}\binom{N-M}{x} \Big/ \binom{N}{k+x-1} \tag{1.23}$$

for $x=0,1,\ldots,\min\{N-k,N-M\}$, where k, M and N are positive integers with $k \le N$ and $M \le N$. Letting $k=\alpha$, $M=\alpha+\beta-1$ and $N=\alpha+\beta+n-1$ in (1.23) we get (1.18) as a negative hypergeometric distribution. From (1.19) and (1.23) we have as a relationship between the hypergeometric and negative hypergeometric distributions

$$F_{NH}(x:k,M,N) := \sum_{t=0}^{x} p_{NH}(t:k,M,N)$$

$$
\begin{aligned}
&= \sum_{t=k}^{T} p_H(t : k+x, M, N) \\
&=: G_H(k : k+x, M, N), \qquad (1.24)
\end{aligned}
$$

where $T = \min\{k+x, M\}$ (see, e.g. [TF81]). Since, from (1.18) and (1.24),

$$
F_{NH}(y : \alpha, \alpha+\beta-1, \alpha+\beta+n-1) = G_H(\alpha : y+\alpha, \alpha+\beta-1, \alpha+\beta+n-1),
$$

it follows from (1.22) the Edgeworth expansion of the distribution of Y_n is given by

$$
\begin{aligned}
P\{Y_n \leq y\} &= G_H(\alpha : y+\alpha, \alpha+\beta-1, \alpha+\beta+n-1) \\
&= 1 - \left\{ \Phi(t) - \frac{1}{6\sigma}(q-p)(\delta' - \gamma)(t^2-1)\phi(t) \right\} + o\left(\frac{1}{\sigma}\right), \qquad (1.25)
\end{aligned}
$$

where

$$
t = \frac{1}{\sigma}(\alpha - 0.5 - \mu), \qquad \mu = \frac{(y+\alpha)(\alpha+\beta-1)}{\alpha+\beta+n-1},
$$

$$
\sigma = \sqrt{\frac{(y+\alpha)(\beta+n-y-1)(\alpha+\beta-1)n}{(\alpha+\beta+n-1)^3}}\,.
$$

We also numerically compare the exact distribution (1.18) with the asymptotic normal distribution $1 - \Phi(t)$ up to the order $o(1)$ and the Edgeworth expansion (1.25) of the c.d.f. of Y_n after continuity correction. We consider the cases when (i) $\alpha=10$, $\beta=11$, $n=80$ and (ii) $\alpha=40$, $\beta=41$, $n=20$. In each case we give errors among the above asymptotic c.d.f.'s as Table 1.3 and 1.4 and the c.d.f.'s as Figure 1.5 and 1.6. The table and figure show that the asymptotic c.d.f.'s are useful for each case ([AKT97]).

Table 1.3: Errors of the asymptotic normal c.d.f. of Y_n and its Edgeworth expansion in the case (i)$\alpha=10$, $\beta=11$, $n=80$

k	Exact	Normal	Edgeworth	k	Exact	Normal	Edgeworth
0	0.0000 %	-0.0000 %	-0.0000 %	41	63.6850 %	0.0001 %	0.0337 %
1	0.0000	-0.0000	-0.0000	42	67.3975	-0.0201	0.0374
2	0.0001	-0.0001	-0.0000	43	70.9519	-0.0338	0.0379
3	0.0002	-0.0002	-0.0001	44	74.3223	-0.0394	0.0350
4	0.0007	-0.0006	-0.0002	45	77.4875	-0.0366	0.0285
5	0.0019	-0.0015	-0.0004	46	80.4310	-0.0258	0.0189
6	0.0044	-0.0036	-0.0004	47	83.1411	-0.0078	0.0064
7	0.0096	-0.0072	-0.0000	48	85.6109	0.0159	-0.0082
8	0.0193	-0.0132	0.0012	49	87.8383	0.0434	-0.0242
9	0.0364	-0.0223	0.0039	50	89.8256	0.0727	-0.0406
10	0.0648	-0.0354	0.0086	51	91.5791	0.1016	-0.0565
11	0.1099	-0.0528	0.0157	52	93.1086	0.1281	-0.0709
12	0.1789	-0.0747	0.0255	53	94.4270	0.1506	-0.0829
13	0.2807	-0.1005	0.0377	54	95.5494	0.1676	-0.0920
14	0.4262	-0.1294	0.0519	55	96.4927	0.1783	-0.0975
15	0.6285	-0.1597	0.0669	56	97.2748	0.1823	-0.0993
16	0.9023	-0.1896	0.0816	57	97.9140	0.1798	-0.0974
17	1.2645	-0.2167	0.0947	58	98.4287	0.1715	-0.0921
18	1.7334	-0.2387	0.1050	59	98.8364	0.1582	-0.0839
19	2.3284	-0.2535	0.1113	60	99.1541	0.1414	-0.0736
20	3.0696	-0.2592	0.1130	61	99.3972	0.1222	-0.0621
21	3.9773	-0.2546	0.1097	62	99.5795	0.1022	-0.0502
22	5.0709	-0.2392	0.1016	63	99.7135	0.0825	-0.0387
23	6.3687	-0.2133	0.0893	64	99.8097	0.0642	-0.0283
24	7.8868	-0.1779	0.0736	65	99.8771	0.0481	-0.0195
25	9.6387	-0.1349	0.0557	66	99.9230	0.0346	-0.0125
26	11.6341	-0.0866	0.0369	67	99.9534	0.0237	-0.0073
27	13.8789	-0.0359	0.0184	68	99.9729	0.0155	-0.0038
28	16.3745	0.0144	0.0014	69	99.9849	0.0096	-0.0016
29	19.1171	0.0611	-0.0130	70	99.9920	0.0056	-0.0005
30	22.0980	0.1017	-0.0241	71	99.9960	0.0030	0.0000
31	25.3033	0.1340	-0.0313	72	99.9981	0.0015	0.0002
32	28.7140	0.1563	-0.0346	73	99.9992	0.0007	0.0002
33	32.3064	0.1679	-0.0339	74	99.9997	0.0003	0.0001
34	36.0527	0.1687	-0.0298	75	99.9999	0.0001	0.0000
35	39.9214	0.1593	-0.0227	76	100.0000	0.0000	0.0000
36	43.8779	0.1413	-0.0134	77	100.0000	0.0000	0.0000
37	47.8860	0.1165	-0.0029	78	100.0000	0.0000	0.0000
38	51.9080	0.0875	0.0080	79	100.0000	0.0000	0.0000
39	55.9061	0.0568	0.0182	80	100.0000	-0.0000	-0.0000
40	59.8436	0.0271	0.0271				

Table 1.4: Errors of the asymptotic normal c.d.f. of Y_n and its Edgeworth expansion in the case (ii) $\alpha=40$, $\beta=41$, $n=20$

k	Exact	Normal	Edgeworth	k	Exact	Normal	Edgeworth
0	0.0008 %	0.0057 %	0.0020 %	11	74.1615 %	0.0581 %	0.0395 %
1	0.0112	0.0218	0.0101	12	85.3382	−0.0048	0.0000
2	0.0801	0.0601	0.0306	13	92.7894	−0.1059	−0.0595
3	0.3792	0.1215	0.0639	14	96.9905	−0.1636	−0.0937
4	1.3381	0.1788	0.0935	15	98.9633	−0.1502	−0.0862
5	3.7492	0.1812	0.0918	16	99.7167	−0.0968	−0.0549
6	8.6811	0.1003	0.0469	17	99.9424	−0.0458	−0.0249
7	17.0834	−0.0213	−0.0142	18	99.9922	−0.0159	−0.0079
8	29.1916	−0.0907	−0.0429	19	99.9995	−0.0040	−0.0016
9	44.0939	−0.0568	−0.0182	20	100.0000	−0.0007	−0.0001
10	59.8436	0.0271	0.0271				

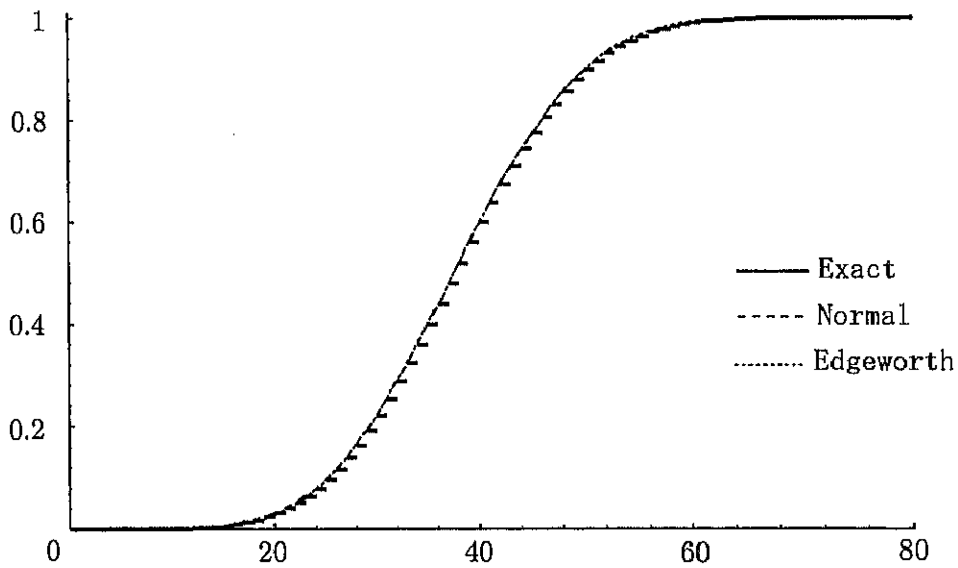

Figure 1.5: The exact, asymptotic normal c.d.f.'s of Y_n and its Edgeworth expansion in the case (i) $\alpha=10$, $\beta=11$, $n=80$

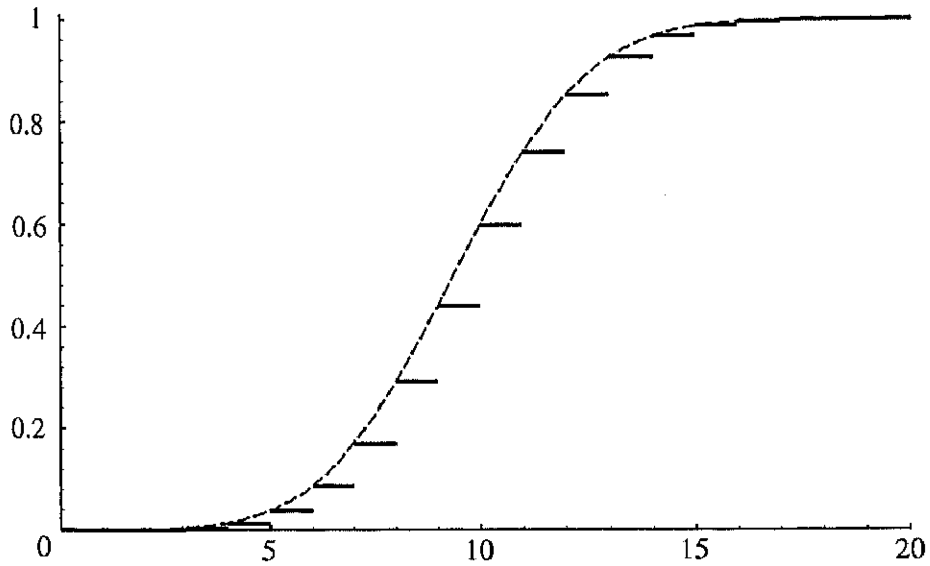

Figure 1.6: The exact, asymptotic normal c.d.f.'s of Y_n and its Edgeworth expansion in the case (ii) $\alpha=40$, $\beta=41$, $n=20$