

# Higher order approximation for the tail probability of discrete distributions

筑波大・数学 高橋 邦彦 (Kunihiko Takahashi)  
筑波大・数学 赤平 昌文 (Masafumi Akahira)

## 1. はじめに

統計的推測の高次漸近理論では Edgeworth(展開による) 近似がよく用いられ, 推定量, 検定などの高次の漸近的性質が論じられてきた. 実際, 推定量の真の母数の周りでの集中確率を漸近的に求めるときには漸近分布の中心部分を必要とし, そのとき Edgeworth 近似が重要な役割を果たした (Barndorff-Nielsen and Cox [BC89]). 一方, Bahadur 効率は漸近分布の裾の部分に基づいていて, 大偏差確率を必要とする. また, 裾確率の近似についてもいろいろ論じられている (Daniels [D87], Jensen [J95]). 最近, 我々は独立な離散型確率変数の和の分布に対する大偏差 (原理による) 近似を高次の次数まで求めて, 通常用いられている正規近似, Edgeworth 近似と比較検討し, その近似精度が良いことを確めた (Akahira, Takahashi and Takeuchi [ATT99], 高橋・赤平 [TA98]). 実際, 各点における確率関数の近似は分布の裾の部分はもちろん, 中心部分でも良い近似精度を与えた.

本論では, 裾確率の近似式を新たに導出し, Edgeworth 近似, [ATT99] の近似等との数値的比較を 2 項分布, 負の 2 項分布, 対数級数分布等の場合に行う. また, その近似式を仮説検定問題における検出力の数値計算に適用し, さらに, パーセント点の近似を通して予測問題にも適用する.

## 2. 大偏差近似

本節において [ATT99] に従って, 大偏差近似の概略を述べる. まず,  $X_1, \dots, X_n, \dots$  を互いに独立に, 各  $j = 1, 2, \dots$  について  $X_j$  が確率関数

$$p_j(x) = P\{X_j = x\} \quad (x = 0, \pm 1, \pm 2, \dots)$$

に従う離散型確率変数列とする. それらの和  $S_n := \sum_{j=1}^n X_j$  の確率関数を

$$p_n^*(y) := P\{S_n = y\} \quad (y = 0, \pm 1, \pm 2, \dots)$$

とする. また, 各  $j$  について,  $X_j$  の積率母関数を  $M_j(\theta) := E[e^{\theta X_j}]$  ( $\theta \in \Theta$ ) とする. ただし,  $\Theta$  は原点を含むある開区間とする. さて, 各  $j$  について, 確率関数

$$p_{j,\theta}(x) := P_\theta\{X_j = x\} = p_j(x)e^{\theta x} M_j(\theta)^{-1} \quad (x = 0, \pm 1, \pm 2, \dots)$$

の離散指数型分布族  $\mathcal{P}_j := \{p_{j,\theta}(x) : \theta \in \Theta\}$  を考える. ここで,  $p_{j,0}(x) = p_j(x)$  になる. また,  $S_n$  の確率関数を

$$p_{n,\theta}^*(y) := P_\theta \{S_n = y\} \quad (y = 0, \pm 1, \pm 2, \dots)$$

で表わす. ただし  $p_{n,0}^*(y) = p_n^*(y)$  とする. このとき

$$p_{n,\theta}^*(y) = p_n^*(y) e^{\theta y} \prod_{j=1}^n M_j(\theta)^{-1} \quad (2.1)$$

となる. また, 各  $j$  について,  $\mathcal{P}_j$  の下で,  $X_j$  の特性関数は

$$E_\theta [e^{itX_j}] = \sum_x e^{itx} p_{j,\theta}(x) = M_j(\theta)^{-1} M_j(\theta + it)$$

になる. ただし,  $i$  は虚数単位とする. 従って,  $\{\mathcal{P}_j\}$  の下で,  $S_n$  の特性関数は

$$E_\theta [e^{itS_n}] = \prod_{j=1}^n M_j(\theta + it) \prod_{j=1}^n M_j(\theta)^{-1}$$

となるから, フーリエ逆変換を用いて

$$p_{n,\theta}^*(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{j=1}^n M_j(\theta + it) \prod_{j=1}^n M_j(\theta)^{-1} e^{-ity} dt \quad (2.2)$$

になる. そこで, (2.1), (2.2) から

$$p_n^*(y) = e^{-\theta y} \prod_{j=1}^n M_j(\theta) \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{j=1}^n M_j(\theta + it) \prod_{j=1}^n M_j(\theta)^{-1} e^{-ity} dt$$

を得る. また,  $K_n(\theta) := \sum_{j=1}^n \log M_j(\theta)$  とおくと

$$p_n^*(y) = e^{K_n(\theta) - \theta y} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{K_n(\theta + it) - K_n(\theta) - ity} dt \quad (2.3)$$

になる. さらに, Taylor 展開によって, 十分に小さい  $|t|$  に対して

$$K_n(\theta + it) - K_n(\theta) = K_n^{(1)}(\theta)it + \frac{1}{2}K_n^{(2)}(\theta)(it)^2 + \frac{1}{6}K_n^{(3)}(\theta)(it)^3 + \frac{1}{24}K_n^{(4)}(\theta)(it)^4 + \dots \quad (2.4)$$

になる. ただし, 各  $\alpha = 1, 2, \dots$  について

$$K_n^{(\alpha)}(\theta) = (d^\alpha / d\theta^\alpha) K_n(\theta)$$

とおく. このとき,  $S_n = y$  ( $y = 0, \pm 1, \pm 2, \dots$ ) に対して,  $K_n^{(1)}(\hat{\theta}_0) = y$  となる  $\theta$  の推定量  $\hat{\theta}_0 := \hat{\theta}_0(S_n)$  を考える. ここで,  $K_n^{(\alpha)}(\hat{\theta}_0) = O(n)$  ( $\alpha = 2, 3, \dots$ ) を仮定する.

定理 1([ATT99]).  $S_n$  の確率関数  $p_n^*(y)$  は漸近的に次のように与えられる.

$$p_n^*(y) = \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_0)}} e^{K_n(\hat{\theta}_0) - \hat{\theta}_0 y} \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_0)}{8 \{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5 \{K_n^{(3)}(\hat{\theta}_0)\}^2}{24 \{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right) \right].$$

各  $j = 1, 2, \dots$  について, 確率関数  $p_{j,\theta}(\cdot)$  と  $p_j(\cdot)$  の間の Kullback-Leibler 情報量を

$$I_j(\theta, 0) := \sum_x p_{j,\theta}(x) \log \frac{p_{j,\theta}(x)}{p_j(x)}$$

で定義する. このとき, 確率関数  $p_{n,\theta}^*(\cdot)$  と  $p_n^*(\cdot)$  の間の Kullback-Leibler 情報量は

$$I_n^*(\theta, 0) = \sum_{j=1}^n I_j(\theta, 0) = \theta K_n^{(1)}(\theta) - K_n(\theta)$$

となる. 従って, 定理 1 から次の系を得る.

系 1.  $S_n$  の確率関数  $p_n^*(y)$  は漸近的に次のように与えられる.

$$p_n^*(y) = \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_0)}} e^{-I_n^*(\hat{\theta}_0, 0)} \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_0)}{8 \{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5 \{K_n^{(3)}(\hat{\theta}_0)\}^2}{24 \{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right) \right].$$

### 3. 裾確率の近似

まず, [ATT99] による裾確率の近似を概説する. まず, (2.3) より, 任意の  $z \geq 0$  に対して

$$p_n^*(y+z) = e^{K_n(\theta) - \theta(y+z)} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{K_n(\theta+it) - K_n(\theta) - it(y+z)} dt$$

が成り立つ. また,  $\hat{\theta}_0$  を用いれば

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{K_n(\hat{\theta}_0+it) - K_n(\hat{\theta}_0) - it(y+z)} dt \\ &= \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_0)}} \left\{ 1 - \frac{z^2}{2K_n^{(2)}(\hat{\theta}_0)} - \frac{K_n^{(3)}(\hat{\theta}_0)z}{2\{K_n^{(2)}(\hat{\theta}_0)\}^2} + \frac{K_n^{(4)}(\hat{\theta}_0)}{8\{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_0)\}^2}{24\{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right) \right\} \end{aligned}$$

となる. よって,  $y > E(S_n)$  なる任意の整数  $y$  について,  $S_n$  の分布の上側裾確率は

$$\begin{aligned} P\{S_n \geq y\} &= \sum_{z=0}^{\infty} p_n^*(y+z) \\ &= \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_0)}} e^{K_n(\hat{\theta}_0) - \hat{\theta}_0 y} \sum_{z=0}^{\infty} e^{-\hat{\theta}_0 z} \left\{ 1 - \frac{z^2}{2K_n^{(2)}(\hat{\theta}_0)} - \frac{K_n^{(3)}(\hat{\theta}_0)z}{2\{K_n^{(2)}(\hat{\theta}_0)\}^2} \right. \end{aligned}$$

$$+ \frac{K_n^{(4)}(\hat{\theta}_0)}{8\{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_0)\}^2}{24\{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right)\Bigg\}$$

として求められる。

定理 2([ATT99]).  $S_n$  の分布の上側裾確率は, 漸近的に次のように与えられる。

$$P\{S_n \geq y\} = \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_0)}} e^{K_n(\hat{\theta}_0) - \hat{\theta}_0 y} \sum_{z=0}^{\infty} e^{-\hat{\theta}_0 z} \left[ 1 - \frac{z^2}{2K_n^{(2)}(\hat{\theta}_0)} - \frac{K_n^{(3)}(\hat{\theta}_0)z}{2\{K_n^{(2)}(\hat{\theta}_0)\}^2} + \frac{K_n^{(4)}(\hat{\theta}_0)}{8\{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_0)\}^2}{24\{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right) \right], \quad y > E(S_n).$$

定理 2 の上側裾確率の近似式は数値計算の結果から見るとやや安定性に欠けることがある。そこで, この裾確率の近似式の改善を試みる。まず

$$K_n^{(1)}(\hat{\theta}_0) = y, \quad K_n^{(1)}(\hat{\theta}_k) = y + k, \quad (3.1)$$

となる  $\theta$  の推定量  $\hat{\theta}_0, \hat{\theta}_k$  を考える。ただし,  $k = 1, 2, \dots$  とする。また, 各  $k$  について

$$\Delta_k = \hat{\theta}_k - \hat{\theta}_0 \quad (3.2)$$

とおけば, 定理 1 より

$$p_n^*(y) = \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_0)}} e^{K_n(\hat{\theta}_0) - \hat{\theta}_0 y} \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_0)}{8\{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_0)\}^2}{24\{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right) \right],$$

$$p_n^*(y+k) = \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_k)}} e^{K_n(\hat{\theta}_k) - \hat{\theta}_k(y+k)} \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_k)}{8\{K_n^{(2)}(\hat{\theta}_k)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_k)\}^2}{24\{K_n^{(2)}(\hat{\theta}_k)\}^3} + O\left(\frac{1}{n^2}\right) \right]$$

となる。このとき

$$\begin{aligned} & \log \frac{p_n^*(y+k)}{p_n^*(y)} \\ &= K_n(\hat{\theta}_k) - K_n(\hat{\theta}_0) - \{\hat{\theta}_k(y+k) - \hat{\theta}_0 y\} - \frac{1}{2} \left\{ \log K_n^{(2)}(\hat{\theta}_k) - \log K_n^{(2)}(\hat{\theta}_0) \right\} \\ &+ \log \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_k)}{8\{K_n^{(2)}(\hat{\theta}_k)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_k)\}^2}{24\{K_n^{(2)}(\hat{\theta}_k)\}^3} + O\left(\frac{1}{n^2}\right) \right] \\ &- \log \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_0)}{8\{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_0)\}^2}{24\{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right) \right] \end{aligned} \quad (3.3)$$

となる。また, (3.1) より

$$k = K_n^{(1)}(\hat{\theta}_k) - K_n^{(1)}(\hat{\theta}_0)$$

$$= K_n^{(2)}(\hat{\theta}_0)\Delta_k + \frac{1}{2}K_n^{(3)}(\hat{\theta}_0)\Delta_k^2 + \frac{1}{6}K_n^{(4)}(\hat{\theta}_0)\Delta_k^3 + \dots$$

になるから

$$\Delta_k = \frac{k}{K_n^{(2)}(\hat{\theta}_0)} - \frac{K_n^{(3)}(\hat{\theta}_0)k^2}{2\{K_n^{(2)}(\hat{\theta}_0)\}^3} - \frac{K_n^{(4)}(\hat{\theta}_0)k^3}{6\{K_n^{(2)}(\hat{\theta}_0)\}^4} + \dots \quad (3.4)$$

となって

$$\begin{aligned} K_n(\hat{\theta}_k) - K_n(\hat{\theta}_0) &= K_n^{(1)}(\hat{\theta}_0)\Delta_k + \frac{K_n^{(2)}(\hat{\theta}_0)}{2}\Delta_k^2 + \dots \\ &= y\Delta_k + O\left(\frac{1}{n}\right) \end{aligned} \quad (3.5)$$

になる. さらに, (3.2) より

$$k\hat{\theta}_k = k\hat{\theta}_0 + k\Delta_k \quad (3.6)$$

となり,

$$\begin{aligned} &\frac{1}{2} \left\{ \log K_n^{(2)}(\hat{\theta}_k) - \log K_n^{(2)}(\hat{\theta}_0) \right\} \\ &= \frac{1}{2} \left[ \frac{K_n^{(3)}(\hat{\theta}_0)}{K_n^{(2)}(\hat{\theta}_0)} \Delta_k + \frac{1}{2} \left\{ \frac{K_n^{(4)}(\hat{\theta}_0)}{K_n^{(2)}(\hat{\theta}_0)} - \left( \frac{K_n^{(3)}(\hat{\theta}_0)}{K_n^{(2)}(\hat{\theta}_0)} \right)^2 \right\} \Delta_k^2 + \dots \right], \end{aligned} \quad (3.7)$$

$$\begin{aligned} &\log \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_k)}{8\{K_n^{(2)}(\hat{\theta}_k)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_k)\}^2}{24\{K_n^{(2)}(\hat{\theta}_k)\}^3} + O\left(\frac{1}{n^2}\right) \right] \\ &= \log \left[ 1 + \frac{K_n^{(4)}(\hat{\theta}_0)}{8\{K_n^{(2)}(\hat{\theta}_0)\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta}_0)\}^2}{24\{K_n^{(2)}(\hat{\theta}_0)\}^3} + O\left(\frac{1}{n^2}\right) \right] + \frac{g^{(1)}(\hat{\theta}_0)}{1+g(\hat{\theta}_0)}\Delta_k + O\left(\frac{1}{n^2}\right) \end{aligned} \quad (3.8)$$

となる. ただし

$$\begin{aligned} g(\theta) &:= \frac{K_n^{(4)}(\theta)}{8\{K_n^{(2)}(\theta)\}^2} - \frac{5\{K_n^{(3)}(\theta)\}^2}{24\{K_n^{(2)}(\theta)\}^3} = O\left(\frac{1}{n}\right) \\ g^{(1)}(\theta) &:= \frac{dg(\theta)}{d\theta} = \frac{K_n^{(5)}(\theta)}{8\{K_n^{(2)}(\theta)\}^2} - \frac{K_n^{(3)}(\theta)K_n^{(4)}(\theta)}{8\{K_n^{(2)}(\theta)\}^3} - \frac{5K_n^{(3)}(\theta)K_n^{(4)}(\theta)}{12\{K_n^{(2)}(\theta)\}^3} + \frac{5\{K_n^{(3)}(\theta)\}^3}{8\{K_n^{(2)}(\theta)\}^4} \\ &= O\left(\frac{1}{n}\right) \end{aligned}$$

とする. よって, (3.3)~(3.8) より

$$\log \frac{p_n^*(y+k)}{p_n^*(y)} = -k\hat{\theta}_0 - k\Delta_k + \frac{1}{2}K_n^{(2)}(\hat{\theta}_0)\Delta_k^2 - \frac{K_n^{(3)}(\hat{\theta}_0)}{2K_n^{(2)}(\hat{\theta}_0)}\Delta_k + O\left(\frac{1}{n^2}\right)$$

になり,

$$p_n^*(y+k) = p_n^*(y) \exp \left\{ -k\hat{\theta}_0 - k\Delta_k + \frac{1}{2}K_n^{(2)}(\hat{\theta}_0)\Delta_k^2 - \frac{K_n^{(3)}(\hat{\theta}_0)}{2K_n^{(2)}(\hat{\theta}_0)}\Delta_k + O\left(\frac{1}{n^2}\right) \right\} \quad (3.9)$$

になる. このとき, (3.4) より

$$\Delta_k = \frac{k}{K_n^{(2)}(\hat{\theta}_0)} + O\left(\frac{1}{n^2}\right)$$

となるから, (3.9) より

$$p_n^*(y+k) = p_n^*(y) \exp \left\{ -k\hat{\theta}_0 - \frac{k^2}{2K_n^{(2)}(\hat{\theta}_0)} - \frac{K_n^{(3)}(\hat{\theta}_0)k}{2\{K_n^{(2)}(\hat{\theta}_0)\}^2} + O\left(\frac{1}{n^2}\right) \right\}$$

となり,  $P\{S_n \geq y\} = \sum_{k=0}^{\infty} p_n^*(y+k)$  から,  $S_n$  の裾確率の近似式を得ることができる.

**定理 3.**  $S_n$  の分布の上側裾確率は, 漸近的に次のように与えられる.

$$P\{S_n \geq y\} = p_n^*(y) \sum_{k=0}^{\infty} \exp \left\{ -k\hat{\theta}_0 - \frac{k^2}{2K_n^{(2)}(\hat{\theta}_0)} - \frac{K_n^{(3)}(\hat{\theta}_0)k}{2\{K_n^{(2)}(\hat{\theta}_0)\}^2} + O\left(\frac{1}{n^2}\right) \right\}, \quad y > E(S_n)$$

ただし,  $p_n^*(y)$  は定理 1 で与えられたものとする.

次に

$$K_n^{(1)}(\hat{\theta}_0) = y, \quad K_n^{(1)}(\hat{\theta}_k) = y - k$$

となる推定量  $\hat{\theta}_0$  と  $\hat{\theta}_k$  を考えると, 上側裾確率の近似式の導出と同様にして

$$p_n^*(y-k) = p_n^*(y) \exp \left\{ k\hat{\theta}_0 - \frac{k^2}{2K_n^{(2)}(\hat{\theta}_0)} + \frac{K_n^{(3)}(\hat{\theta}_0)k}{2\{K_n^{(2)}(\hat{\theta}_0)\}^2} + O\left(\frac{1}{n^2}\right) \right\}$$

を得て, これより下側裾確率の近似式も導出できる.

**系 2.**  $S_n$  の分布の下側裾確率は, 漸近的に次のように与えられる.

$$P\{S_n \leq y\} = p_n^*(y) \sum_{k=0}^{\infty} \exp \left\{ k\hat{\theta}_0 - \frac{k^2}{2K_n^{(2)}(\hat{\theta}_0)} + \frac{K_n^{(3)}(\hat{\theta}_0)k}{2\{K_n^{(2)}(\hat{\theta}_0)\}^2} + O\left(\frac{1}{n^2}\right) \right\}, \quad y < E(S_n)$$

ただし,  $p_n^*(y)$  は定理 1 で与えられたものとする.

#### 4. 数値計算による精度の比較

実際に、2項分布、負の2項分布、対数級数分布などの場合について、上記の定理の近似式による近似値、Edgeworth近似による値を数値計算によりそれぞれ求め、真値と比較することによって、その近似精度を確める。なお、定理1において

$$p_n^*(y) = \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta}_0)}} e^{K_n(\hat{\theta}_0) - \hat{\theta}_0 y} (1 + o(1))$$

を1次近似  $LD_1$ 、定理1の近似式を2次近似  $LD_2$  ということにする。実は、定理1で求めた各点確率はかなり良い近似精度であることが分かっている ([ATT99])。また、裾確率の近似については定理3の近似式から良い精度の値を求めることができることがわかる。

##### 4.1. 2項分布の場合

$X_1, X_2, \dots, X_n, \dots$  が互いに独立に、各  $j$  について  $X_j$  が確率関数

$$f_B(x) := P\{X_j = x\} = p_j^x q_j^{1-x}, \quad (x = 0, 1)$$

をもつ2項分布  $B(1, p_j)$  に従う確率変数列とする。ただし、 $0 < p_j < 1$ ,  $q_j = 1 - p_j$  ( $j = 1, 2, \dots$ ) とする。このとき、各  $y = 0, 1, \dots, n$  について

$$K_n^{(1)}(\theta) = \sum_{j=1}^n \frac{p_j e^\theta}{p_j e^\theta + q_j} = y$$

となる  $\theta$  を  $\hat{\theta}_0 = \hat{\theta}_0(y)$  とする。このとき  $\hat{\tau} := e^{\hat{\theta}_0}$  とおいて、各  $j$  について  $\hat{p}_j := p_j \hat{\tau} / (p_j \hat{\tau} + q_j)$ ,  $\hat{q}_j := 1 - \hat{p}_j$  とすれば

$$K_n^{(2)}(\hat{\theta}_0) = \sum_{j=1}^n \hat{p}_j \hat{q}_j, \quad K_n^{(3)}(\hat{\theta}_0) = \sum_{j=1}^n \hat{p}_j \hat{q}_j (\hat{q}_j - \hat{p}_j), \quad K_n^{(4)}(\hat{\theta}_0) = \sum_{j=1}^n \hat{p}_j \hat{q}_j (1 - 6\hat{p}_j \hat{q}_j) \quad (4.1)$$

になる。従って、定理1から  $S_n := \sum_{j=1}^n X_j = y$  となる確率は

$$p_n^*(y) = \frac{1}{\sqrt{2\pi \sum_{j=1}^n \hat{p}_j \hat{q}_j}} \left( \prod_{j=1}^n \frac{q_j}{\hat{q}_j} \right) \hat{\tau}^{-y} \left[ 1 + \frac{K_n^{(4)}(\hat{\theta})}{8 \left( \sum_{j=1}^n \hat{p}_j \hat{q}_j \right)^2} - \frac{5\{K_n^{(3)}(\hat{\theta})\}^2}{24 \left( \sum_{j=1}^n \hat{p}_j \hat{q}_j \right)^3} + O\left(\frac{1}{n^2}\right) \right]$$

となる。

一方、 $S_n$  のキュムラントは

$$\begin{aligned} \mu_n &:= E(S_n) = \sum_{j=1}^n p_j, & v_n &:= V(S_n) = \sum_{j=1}^n p_j q_j, \\ \kappa_{3,n} &:= \kappa_3(S_n) = \sum_{j=1}^n p_j q_j (q_j - p_j), \end{aligned}$$

$$\kappa_{4,n} := \kappa_4(S_n) = \sum_{j=1}^n p_j q_j (1 - 6p_j q_j)$$

となるから,  $S_n$  の分布の Edgeworth 展開は

$$\begin{aligned} P\{S_n = t\} &= P\left\{\frac{S_n - \mu_n}{\sqrt{v_n}} = \frac{t - \mu_n}{\sqrt{v_n}}\right\} \\ &= \frac{1}{\sqrt{v_n}} \phi(y) \left\{1 + \frac{\kappa_{3,n}}{6v_n^{3/2}}(y^3 - 3y) + \frac{\kappa_{4,n}}{24v_n^2}(y^4 - 6y^2 + 3) \right. \\ &\quad \left. + \frac{\kappa_{3,n}^2}{72v_n^3}(y^6 - 15y^4 + 45y^2 - 15)\right\} + o\left(\frac{1}{n\sqrt{n}}\right) \end{aligned}$$

になる. ただし  $y := (t - \mu_n) / \sqrt{v_n}$ ,  $\phi(y) = (1/\sqrt{2\pi}) e^{-y^2/2}$  とする. 上式の右辺の  $\{\dots\}$  の第 1 項目までの近似を正規近似といい, 右辺のすべての項による近似を Edgeworth 近似という. このとき大偏差近似  $LD_1$ ,  $LD_2$  はそれぞれ正規近似, Edgeworth 近似より端において精確になることをいくつかの分布の場合について数値計算を行って確認した ([ATT99], [TA98] 参照).

特に, 各  $j$  について  $p_j = p$  のとき, すなわち  $X_j$  がいずれも独立でいずれも  $B(1, p)$  に従う確率変数の場合, 定理 1 より

$$p_n^*(y) = \frac{p^y q^{n-y}}{\sqrt{2\pi}} \cdot \frac{n^{n+(1/2)}}{y^{y+(1/2)}(n-y)^{n-y+(1/2)}} \left[1 - \frac{1}{12n} \left\{\frac{1}{\frac{y}{n}(1-\frac{y}{n})} - 1\right\} + O\left(\frac{1}{n^2}\right)\right]$$

になる. 一方,  $S_n$  は 2 項分布  $B(n, p)$  に従うので

$$p_n^*(y) = \binom{n}{y} p^y q^{n-y}$$

となる. よって, この場合の相対誤差は  $p$  に無関係になることが分かる ([J95]).

さらに, (4.1) の値を用いて, 定理 2, 定理 3 及び系 2 から裾確率もそれぞれ求められる (表 4.1.1 ~ 4.1.11 参照). 特に, 上側裾確率については Jensen [J95] において得られている値とも比較を行った. その結果, 定理 3, 系 2 による裾確率の近似がより精確なものになっていることが分かった.

表 4.1.1:  $B(1, p_j)$  ( $\{p_j\}_{j=1}^{19} = 0.05(0.05)0.95$ ) のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似  $LD$  (定理 2),  $LD$  (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 2	定理 3
10	50.0000	0.0000	—	-0.0029
11	29.0651	-0.0000	-0.8356	0.0001
12	13.4452	0.0000	-0.2010	0.0009
13	4.7942	0.0004	-0.0627	0.0009
14	1.2761	0.0014	-0.0232	0.0007
15	0.2442	-0.0006	-0.0097	0.0003
16	0.0320	-0.0293	-0.0048	-0.0003
17	0.0027	0.2154	-0.0038	-0.0018
18	0.0001	—	-0.0083	-0.0077
19	0.0000	—	—	—



表 4.1.2:  $B(1, p_j)$  ( $\{p_j\}_{j=1}^{10} = 0.05(0.05)0.50$ ) のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似  $LD$ (定理 2),  $LD$ (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 2	定理 3
3	55.6619	0.0014	—	-0.0516
4	27.7313	0.0013	-0.6407	-0.0154
5	9.7465	-0.0050	-0.1133	-0.0048
6	2.3314	-0.0243	-0.0301	-0.0016
7	0.3634	0.0259	-0.0103	-0.0010
8	0.0345	0.3918	-0.0054	-0.0021
9	0.0018	1.6628	-0.0089	-0.0080
10	0.0000	7.0938	—	—

表 4.1.3:  $B(1, p_j)$  ( $\{p_j\}_{j=1}^{19} = 0.03(0.03)0.60$ ) のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似  $LD$ (定理 2),  $LD$ (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 2	定理 3
7	45.0120	0.0002	—	-0.0213
8	26.2247	0.0003	-0.9933	-0.0095
9	12.7076	-0.0003	-0.2467	-0.0043
10	5.0524	-0.0032	-0.0840	-0.0020
11	1.6288	-0.0067	-0.0346	-0.0009
12	0.4207	0.0005	-0.0161	-0.0003
13	0.0859	0.0451	-0.0082	-0.0001
14	0.0136	0.1758	-0.0044	-0.0001
15	0.0016	0.4772	-0.0026	-0.0001
16	0.0001	1.1191	-0.0018	-0.0003
17	0.0000	2.5290	-0.0017	-0.0008
18	0.0000	6.0939	-0.0025	-0.0020
19	0.0000	18.1006	-0.0076	-0.0073
20	0.0000	91.0675	—	—

表 4.1.4:  $B(10, 0.5)$  のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似  $LD$ (定理 2),  $LD$ (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 2	定理 3
6	37.6953	-0.0004	-0.4706	-0.0026
7	17.1875	-0.0016	-0.0921	0.0014
8	5.4688	-0.0004	-0.0188	0.0005
9	1.0742	0.0195	-0.0082	-0.0065
10	0.0977	0.0679	—	—

表 4.1.5:  $B(20, 0.5)$  のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似  $LD$ (定理 2),  $LD$ (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 2	定理 3
11	41.1901	-0.0001	—	-0.0027
12	25.1722	-0.0003	-0.5658	0.0005
13	13.1588	-0.0005	-0.1912	0.0015
14	5.7659	-0.0001	-0.0739	0.0017
15	2.0695	0.0021	-0.0312	0.0015
16	0.5909	0.0070	-0.0137	0.0012
17	0.1288	0.0080	-0.0061	0.0007
18	0.0201	-0.0458	0.0031	-0.0004
19	0.0020	-0.4966	-0.0063	-0.0060
20	0.0001	—	—	—

表 4.1.7:  $B(n, p_j)$ ,  $p_j = p = 0.15 (j = 1, \dots, n)$  のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, Jensen ([J95], page 44) の Table 2.4.4 のその値の各近似値, および大偏差近似  $LD$ (定理 2),  $LD$ (定理 3) の相対誤差

$n = 10, p = 0.15$

$y$	2	4	5	8	9
真値 (%)	45.57	5.00	0.987	$8.67 \times 10^{-4}$	$3.33 \times 10^{-5}$
Normal	0.097	-0.234	-0.601	-0.994	-1.000
Saddlepoint	-0.137	-0.050	-0.035	-0.007	0.021
Sp.-expansion	0.004	0.004	0.005	0.010	0.024
Lugannani-Rice	0.004	0.008	0.013	0.038	0.081
定理 2	—	-0.070	-0.021	-0.003	-0.009
定理 3	-0.069	-0.008	-0.003	-0.002	-0.006

$n = 20, p = 0.15$

$y$	6	8	18	19
真値 (%)	6.73	0.592	$2.07 \times 10^{-11}$	$3.80 \times 10^{-13}$
Normal	-0.128	-0.591	-1.000	-1.000
Saddlepoint	-0.043	-0.027	0.010	0.039
Sp.-expansion	0.001	0.002	0.005	0.021
Lugannani-Rice	0.003	0.005	0.039	0.084
定理 2	-0.126	-0.022	-0.005	-0.006
定理 3	-0.010	-0.003	-0.004	-0.006

表 4.1.8:  $B(1, p_j)$  ( $\{p_j\}_{j=1}^{19} = 0.05(0.05)0.95$ ) のときの  $S_n$  の下側確率  $P\{S_n \leq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似 LD(系 2) の相対誤差

$y$	真値 (%)	Edgeworth	LD
0	0.0000	—	—
1	0.0001	—	-0.0077
2	0.0027	-0.2154	-0.0018
3	0.0320	-0.0293	-0.0003
4	0.2442	-0.0006	0.0003
5	1.2761	0.0014	0.0007
6	4.7942	0.0004	0.0009
7	13.4452	0.0000	0.0009
8	29.0651	-0.0000	0.0001
9	50.0000	0.0000	-0.0029

表 4.1.9:  $B(1, p_j)$  ( $\{p_j\}_{j=1}^{20} = 0.03(0.03)0.60$ ) のときの  $S_n$  の下側確率  $P\{S_n \leq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似 LD(系 2) の相対誤差

$y$	真値 (%)	Edgeworth	LD
0	0.0263	-0.2603	—
1	0.3232	0.0331	-0.0060
2	1.8695	0.0156	0.0015
3	6.7942	0.0019	0.0045
4	17.5403	-0.0011	0.0076
5	34.5942	-0.0006	0.0118
6	54.9880	-0.0002	0.0172

表 4.1.10:  $B(10, 0.5)$  のときの  $S_n$  の下側確率  $P\{S_n \leq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似 LD(系 2) の相対誤差

$y$	真値 (%)	Edgeworth	LD
0	0.0977	0.0679	—
1	1.0742	0.0195	-0.0065
2	5.4688	-0.0004	0.0005
3	17.1875	-0.0016	0.0014
4	37.6953	-0.0004	-0.0026
5	62.3047	0.0002	-0.0205

表 4.1.11:  $B(20, 0.5)$  のときの  $S_n$  の下側確率  $P\{S_n \leq y\}$  の値と, その値の Edgeworth 近似, 大偏差近似 LD(系 2) の相対誤差

$y$	真値 (%)	Edgeworth	LD
0	0.0001	—	—
1	0.0020	-0.4966	-0.0060
2	0.0201	-0.0458	-0.0004
3	0.1288	0.0080	0.0007
4	0.5909	0.0070	0.0012
5	2.0695	0.0021	0.0015
6	5.7659	-0.0001	0.0017
7	13.1588	-0.0005	0.0015
8	25.1722	-0.0003	0.0005
9	41.1901	-0.0001	-0.0027
10	58.8099	0.0000	-0.0106

#### 4.2. 負の 2 項分布の場合

確率変数  $X_1, \dots, X_n, \dots$  を互いに独立に, 各  $j = 1, 2, \dots$  について,  $X_j$  が

$$f_{X_j}(x) = P\{X_j = x\} = \binom{x+r_j-1}{x} p_j^{r_j} q_j^x$$

を確率関数とする負の二項分布  $NB(r_j, p_j)$  に従うとする. ただし  $0 < p_j < 1$  とし, また  $q_j = 1 - p_j$  ( $j = 1, 2, \dots$ ) とする. このとき  $S_n = \sum_{j=1}^n X_j$  の分布とその大偏差近似, 正規近似, Edgeworth 近似をそれぞれ求めて数値的に比較する.

各  $j = 1, 2, \dots$  について,  $|\theta|$  が小さいとき  $X_j$  の積率母関数は

$$M_j(\theta) = E(e^{\theta X_j}) = p_j^{r_j} (1 - q_j e^{\theta})^{-r_j}$$

で与えられる. このとき

$$K_n(\theta) = \sum_{j=1}^n \log M_j(\theta) = \sum_{j=1}^n \{r_j \log p_j - r_j \log(1 - q_j e^{\theta})\}$$

であるから, 各  $y = 0, 1, \dots$  について

$$K_n^{(1)}(\theta) = \sum_{j=1}^n \frac{r_j q_j e^{\theta}}{1 - q_j e^{\theta}} = y$$

となる  $\theta$  を  $\hat{\theta}$  とする. ここで,  $\hat{\tau} = e^{\hat{\theta}}$  とおくと

$$y = \sum_{j=1}^n \frac{r_j q_j \hat{\tau}}{1 - q_j \hat{\tau}}$$

になる。また

$$\begin{aligned} K_n^{(2)}(\hat{\theta}) &= \sum_{j=1}^n \frac{r_j q_j \hat{\tau}}{(1 - q_j \hat{\tau})^2}, \\ K_n^{(3)}(\hat{\theta}) &= \sum_{j=1}^n \frac{r_j q_j \hat{\tau} (1 + q_j \hat{\tau})}{(1 - q_j \hat{\tau})^3}, \\ K_n^{(4)}(\hat{\theta}) &= \sum_{j=1}^n \frac{r_j q_j (1 + 4q_j \hat{\tau} + q_j^2 \hat{\tau}^2)}{(1 - q_j \hat{\tau})^4} \end{aligned} \quad (4.2)$$

を得る。従って、定理 1 から大偏差近似として

$$p_n^*(y) = \frac{1}{\sqrt{2\pi K_n^{(2)}(\hat{\theta})}} \left\{ \prod_{j=1}^n \left( \frac{p_j}{1 - q_j \hat{\tau}} \right)^{r_j} \right\} \hat{\tau}^{-y} \left[ 1 + \frac{K_n^{(4)}(\hat{\theta})}{8\{K_n^{(2)}(\hat{\theta})\}^2} - \frac{5\{K_n^{(3)}(\hat{\theta})\}^2}{24\{K_n^{(2)}(\hat{\theta})\}^3} + O\left(\frac{1}{n^2}\right) \right]$$

を得る。

一方、 $S_n$  のキュムラントは

$$\begin{aligned} \mu_n &:= E(S_n) = \sum_{j=1}^n \frac{q_j r_j}{p_j}, & v_n &:= V(S_n) = \sum_{j=1}^n \frac{q_j r_j}{p_j^2}, \\ \kappa_{3,n} &:= \kappa_3(S_n) = \sum_{j=1}^n \frac{q_j (1 + q_j) r_j}{p_j^3}, \\ \kappa_{4,n} &:= \kappa_4(S_n) = \sum_{j=1}^n \frac{q_j r_j}{p_j^4} (1 + 4q_j + q_j^2) \end{aligned}$$

となるから、 $S_n$  の分布の Edgeworth 展開は

$$\begin{aligned} P\{S_n = t\} &= P\left\{ \frac{S_n - \mu_n}{\sqrt{v_n}} = \frac{t - \mu_n}{\sqrt{v_n}} \right\} \\ &= \frac{1}{\sqrt{v_n}} \phi(y) \left\{ 1 + \frac{\kappa_{3,n}}{6v_n^{3/2}} (y^3 - 3y) + \frac{\kappa_{4,n}}{24v_n^2} (y^4 - 6y^2 + 3) \right. \\ &\quad \left. + \frac{\kappa_{3,n}^2}{72v_n^3} (y^6 - 15y^4 + 45y^2 - 15) \right\} + o\left(\frac{1}{n\sqrt{n}}\right) \end{aligned}$$

になる。ただし  $y := (t - \mu_n) / \sqrt{v_n}$  とする。

いくつかの場合にこの近似精度を数値的に比較した ([TA98])。

さらに裾確率については、(4.2) の各式を定理 2, 定理 3 および系 2 にそれぞれ代入することによって求められる。ここでは、上側裾確率の真値とその Edgeworth 近似, 定理 3 による大偏差近似を数値的に比較を行った (表 4.2.1~4.2.4 参照)。ただしこの場合の無限和は十分な大きさの部分和 ( $\sum_0^{100}$ ) で数値的に求めた。その結果、大偏差近似は裾確率の近似として精確であることが分かった。

表 4.2.1:  $NB(1, p_j)$  ( $p_1 = \dots = p_5 = 0.9$ ,  $p_6 = \dots = p_{10} = 0.8$ ,  $p_{11} = \dots = p_{15} = 0.7$ ,  
 $p_{16} = \dots = p_{20} = 0.6$ ) のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似,  
 $LD$ (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 3
8	43.2981	0.0006	-0.0997
9	32.2869	0.0006	-0.0724
10	23.1125	-0.0003	-0.0537
11	15.9331	-0.0040	-0.0406
12	10.6106	-0.0105	-0.0313
13	6.8460	-0.0143	-0.0245
14	4.2912	-0.0057	-0.0195
15	2.6197	0.0228	-0.0158
16	1.5611	-0.8933	-0.0130
17	0.9099	0.1075	-0.0108
18	0.5197	0.1124	-0.0091
19	0.2913	0.0558	-0.0077
20	0.1605	-0.0695	-0.0066
21	0.0871	-0.2466	-0.0057
22	0.0465	-0.4431	-0.0050
23	0.0245	-0.6253	-0.0044
24	0.0128	-0.7709	-0.0039
25	0.0066	-0.8732	-0.0035
26	0.0033	-0.9358	-0.0032
27	0.0017	-0.9706	-0.0029
28	0.0008	-0.9878	-0.0027
29	0.0004	-0.9954	-0.0025
30	0.0002	-0.9984	-0.0023

表 4.2.2:  $NB(2, p_j)$  ( $p_1 = \dots = p_5 = 0.9$ ,  $p_6 = \dots = p_{10} = 0.8$ ,  $p_{11} = \dots = p_{15} = 0.7$ ,  
 $p_{16} = \dots = p_{20} = 0.6$ ) のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似,  
 $LD$ (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 3
15	47.6524	0.0002	-0.0869
16	39.3206	0.0002	-0.0688
17	31.6714	0.0002	-0.0550
18	24.9155	0.0000	-0.0443
19	19.1573	-0.0007	-0.0360
20	14.4080	-0.0020	-0.0295
21	10.6084	-0.0039	-0.0244
22	7.6534	-0.0053	-0.0204
23	5.4149	-0.0050	-0.0171
24	3.7604	-0.0011	-0.0145
25	2.5654	0.0074	-0.0124
26	1.7207	0.0207	-0.0106
27	1.1356	0.0365	-0.0092
28	0.7380	0.0507	-0.0080
29	0.4725	0.0567	-0.0070
30	0.2984	0.0477	-0.0062
31	0.1859	0.0173	-0.0055
32	0.1143	-0.0378	-0.0048
33	0.0694	-0.1178	-0.0043
34	0.0417	-0.2186	-0.0039
35	0.0248	-0.3333	-0.0035
36	0.0145	-0.4531	-0.0032
37	0.0085	-0.5693	-0.0029
38	0.0049	-0.6746	-0.0026
39	0.0028	-0.7645	-0.0024
40	0.0016	-0.8367	-0.0022

表 4.2.3:  $NB(1, p_j)$  ( $p_1 = \dots = p_5 = 0.8$ ,  $p_6 = \dots = p_{10} = 0.6$ ,  $p_{11} = \dots = p_{15} = 0.4$ ,  $p_{16} = \dots = p_{20} = 0.2$ ) のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その近似の相対誤差

$y$	真値 (%)	Edgeworth	定理 3
33	43.9044	-0.0008	-0.1352
34	40.4687	-0.0008	-0.1240
35	37.1713	-0.0009	-0.1139
36	34.0266	-0.0010	-0.1048
37	31.0456	-0.0012	-0.0966
38	28.2358	-0.0015	-0.0891
39	25.6014	-0.0020	-0.0824
40	23.1439	-0.0028	-0.0762
41	20.8624	-0.0038	-0.0707
42	18.7538	-0.0052	-0.0656
43	16.8134	-0.0068	-0.0610
44	15.0350	-0.0086	-0.0567
45	13.4112	-0.0106	-0.0529
46	11.9342	-0.0125	-0.0494
47	10.5953	-0.0141	-0.0462
48	9.3856	-0.0152	-0.0432
49	8.2962	-0.0154	-0.0405
50	7.3180	-0.0144	-0.0380
51	6.4421	-0.0119	-0.0357
52	5.6601	-0.0075	-0.0336
53	4.9638	-0.0010	-0.0316
54	4.3452	0.0078	-0.0298
55	3.7971	0.0188	-0.0281
56	3.3125	0.0320	-0.0266
57	2.8851	0.0472	-0.0252
58	2.5088	0.0638	-0.0238
59	2.1783	0.0814	-0.0226
60	1.8885	0.0990	-0.0214
61	1.6350	0.1160	-0.0204
62	1.4135	0.1312	-0.0194
63	1.2204	0.1438	-0.0184
64	1.0522	0.1527	-0.0175
65	0.9062	0.1569	-0.0167
66	0.7794	0.1556	-0.0160
67	0.6696	0.1482	-0.0152
68	0.5746	0.1340	-0.0146



表 4.2.4:  $NB(5, 0.2)$  のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似,  $LD$ (定理 3) の相対誤差

$y$	真値 (%)	Edgeworth	定理 3
21	42.0674	0.0003	-0.1678
22	38.3338	0.0003	-0.1521
23	34.8038	0.0005	-0.1382
24	31.4887	0.0005	-0.1257
25	28.3946	0.0003	-0.1146
26	25.5233	-0.0003	-0.1047
27	22.8729	-0.0016	-0.0959
28	20.4384	-0.0036	-0.0880
29	18.2126	-0.0066	-0.0808
30	16.1863	-0.0106	-0.0744
31	14.3492	-0.0153	-0.0687
32	12.6898	-0.0205	-0.0635
33	11.1964	-0.0257	-0.0588
34	9.8569	-0.0303	-0.0545
35	8.6595	-0.0334	-0.0507
36	7.5915	-0.0341	-0.0472
37	6.6424	-0.0316	-0.0440
38	5.8011	-0.0251	-0.0410
39	5.0572	-0.0140	-0.0384
40	4.4010	0.0019	-0.0359
41	3.8236	0.0228	-0.0337
42	3.3166	0.0480	-0.0317
43	2.8724	0.0769	-0.0298
44	2.4839	0.1080	-0.0280
45	2.1449	0.1398	-0.0264
46	1.8496	0.1703	-0.0250
47	1.5928	0.1976	-0.0236
48	1.3699	0.2196	-0.0223
49	1.1767	0.2343	-0.0211
50	1.0095	0.2401	-0.0200
51	0.8651	0.2356	-0.0190
52	0.7405	0.2200	-0.0181
53	0.6331	0.1929	-0.0172
54	0.5407	0.1546	-0.0164
55	0.4614	0.1057	-0.0156

### 4.3. ポアソン分布の場合

確率変数  $X_1, \dots, X_n, \dots$  を互いに独立に, 各  $j = 1, 2, \dots$  について,  $X_j$  が

$$f_{X_j}(x) = P\{X_j = x\} = \frac{e^{-1}}{x!}$$

を確率関数とするポアソン分布  $Po(1)$  に従うとする. 各  $j(j = 1, 2, \dots)$  について  $X_j$  の積率母関数は

$$M_j(\theta) = \exp\left\{\left(e^\theta - 1\right)\right\}$$

となり

$$K_n(\theta) = \sum_{j=1}^n \left(e^\theta - 1\right) = n \left(e^\theta - 1\right)$$

になる. すなわち  $S_n = \sum_{j=1}^n X_j$  の分布は  $Po(n)$  になる. よって  $K_n^{(i)}(i = 1, 2, 3, 4)$  を求め, 定理などに代入して各点確率, 裾確率の近似値を求めることができる. なお, このとき  $\mu_n = \nu_n = \kappa_{3,n} = \kappa_{4,n} = n$  を用いて, Edgeworth 近似も求めて比較した (表 4.3.1 参照). その結果, 大偏差近似の精確性を確認できた.

表 4.3.1:  $Po(10)$  のときの  $S_n$  の各点確率  $P\{S_n = y\}$  の値と. その値の Edgeworth 近似, 大偏差近似  $LD_1, LD_2$  の相対誤差

$y$	Exact(%)	Edgeworth	$LD_1$	$LD_2$
0	0.0045	38.6173	—	—
1	0.0454	7.7963	0.0844	-0.0059
2	0.2270	2.7388	0.0422	-0.0012
3	0.7567	1.2256	0.0281	-0.0005
4	0.0189	0.6152	0.0210	-0.0003
5	0.0378	0.3229	0.0168	-0.0002
6	6.3056	0.1678	0.0140	-0.0001
7	9.0079	0.0809	0.0120	-0.0001
8	11.2599	0.0321	0.0105	-0.0001
9	12.5110	0.0076	0.0093	-0.0000
10	12.5110	-0.0000	0.0084	-0.0000
11	11.3736	0.0029	0.0076	-0.0000
12	9.4780	0.0082	0.0070	-0.0000
13	7.2908	0.0044	0.0064	-0.0000
14	5.2077	-0.0214	0.0060	-0.0000
15	3.4718	-0.0789	0.0056	-0.0000
16	2.1699	-0.1685	0.0052	-0.0000
17	1.2764	-0.2777	0.0049	-0.0000
18	0.7091	-0.3833	0.0046	-0.0000
19	0.3732	-0.4600	0.0044	-0.0000
20	0.1866	-0.4917	0.0042	-0.0000
21	0.0889	-0.4800	0.0040	-0.0000
22	0.0404	-0.4446	0.0038	-0.0000
23	0.0176	-0.4146	0.0036	-0.0000
24	0.0073	-0.4154	0.0035	-0.0000
25	0.0029	-0.4588	0.0033	-0.0000
26	0.0011	-0.5396	0.0032	-0.0000
27	0.0004	-0.6414	0.0031	-0.0000
28	0.0001	-0.7443	0.0030	-0.0000
29	0.0001	-0.8330	0.0029	-0.0000
30	0.0000	-0.9000	0.0028	-0.0000

#### 4.4. 対数級数分布の場合

確率変数  $X_1, \dots, X_n, \dots$  を互いに独立に, 各  $j = 1, 2, \dots$  について,  $X_j$  が

$$f_{X_j}(x) = P\{X_j = x\} = \frac{a_j p_j^x}{x}$$

を確率関数とする対数級数分布に従うとする. ただし  $0 < p_j < 1$ ,  $q_j = 1 - p_j$ ,  $a_j = -1/\log q_j$  とする. 各  $j = 1, 2, \dots$  について,  $X_j$  の積率母関数は

$$M_j(\theta) = -a_j \log(1 - p_j e^\theta)$$

となり,  $K_n(\theta) = \sum_{j=1}^n \log M_j(\theta)$  として  $K_n^{(i)}(\theta)$  ( $i = 1, 2, 3, 4$ ) を求め, ここから大偏差近似を得ることができる.

なお, 対数級数分布には再生性がないため, 真値は合成積によって求めることになる. また,  $X_j$  の4次までのキュムラントは

$$\begin{aligned} \kappa_1(p_j) &= \mu(p_j) = \frac{a_j p_j}{q_j}, & \kappa_2(p_j) &= \sigma^2(p_j) = \frac{a_j p_j (1 - a_j p_j)}{q_j^2}, \\ \kappa_3(p_j) &= \frac{a_j p_j (1 + p_j - 3a_j p_j + 2a_j^2 p_j^2)}{q_j^3}, \\ \kappa_4(p_j) &= \frac{a_j p_j \{1 + 4p_j + p_j^2 - 4a_j p_j (1 + p_j) + 6a_j^2 p_j^2 - 3a_j^3 p_j^3\}}{q_j^4} - 3\sigma^4(p_j), \end{aligned}$$

となり, これを用いて Edgeworth 近似を求めることができる. ここでは  $p = 0.4$ ,  $n = 10, 25$  の場合の  $S_n$  の各点確率, 定理3による上側裾確率の近似値の数値的比較を行った (表 4.4.1~4.4.4 参照). その結果, 大偏差近似は精確であることが分かった.

表 4.4.1: 対数級数分布のときの  $S_n$  の各点確率  $P\{S_n = y\}$  の値と, その値の Edgeworth 近似, 大偏差近似  $LD_1, LD_2$  の相対誤差 ( $p = 0.4, n = 10$ )

$y$	真値 (%)	Edgeworth	$LD_1$	$LD_2$
10	8.6671	-0.0020	—	—
11	17.3342	-0.0349	0.0853	-0.0061
12	20.2232	0.0033	0.0437	-0.0014
13	18.0275	0.0001	0.0302	-0.0007
14	13.6177	-0.0019	0.0238	-0.0005
15	9.1894	0.0309	0.0201	-0.0005
16	5.7144	0.0084	0.0179	-0.0004
17	3.3404	-0.1210	0.0164	-0.0004
18	1.8606	-0.1625	0.0154	-0.0005
19	0.9970	0.0818	0.0147	-0.0005
20	0.5177	0.4245	0.0142	-0.0005
21	0.2619	0.4743	0.0138	-0.0005
22	0.1296	0.1329	0.0136	-0.0005
23	0.0629	-0.3496	0.0135	-0.0006
24	0.0301	-0.7168	0.0134	-0.0006
25	0.0142	-0.9053	0.0133	-0.0006
26	0.0066	-0.9754	0.0133	-0.0006
27	0.0030	-0.9950	0.0134	-0.0007
28	0.0014	-0.9992	0.0134	-0.0007
29	0.0006	-0.9999	0.0135	-0.0007
30	0.0002	-1.0000	0.0136	-0.0007

表 4.4.2: 対数級数分布のときの  $S_n$  の裾確率  $P\{S_n = y\}$  の値と, その値の Edgeworth 近似, 大偏差近似  $LD_1, LD_2$  の相対誤差 ( $p = 0.4, n = 25$ )

$y$	真値 (%)	Edgeworth	$LD_1$	$LD_2$
25	0.2211	0.3521	—	—
26	1.1057	0.1617	0.0846	-0.0060
27	2.9486	0.0402	0.0425	-0.0012
28	5.5729	-0.0036	0.0284	-0.0005
29	8.3730	-0.0105	0.0215	-0.0003
30	10.6375	-0.0048	0.0174	-0.0002
31	11.8741	0.0007	0.0147	-0.0001
32	11.9506	0.0013	0.0128	-0.0001
33	11.0468	-0.0009	0.0114	-0.0001
34	9.5088	-0.0009	0.0103	-0.0001
35	7.7030	0.0037	0.0095	-0.0001
36	5.9220	0.0089	0.0088	-0.0001
37	4.3497	0.0057	0.0083	-0.0001
38	3.0693	-0.0119	0.0079	-0.0001
39	2.0902	-0.0375	0.0075	-0.0001
40	1.3789	-0.0512	0.0072	-0.0001
41	0.8842	-0.0297	0.0069	-0.0001
42	0.5525	0.0356	0.0067	-0.0001
43	0.3373	0.1255	0.0065	-0.0001
44	0.2016	0.1971	0.0063	-0.0001
45	0.1182	0.2045	0.0062	-0.0001
46	0.0681	0.1220	0.0061	-0.0001
47	0.0386	-0.0429	0.0060	-0.0001
48	0.0215	-0.2562	0.0059	-0.0001
49	0.0118	-0.4740	0.0058	-0.0001
50	0.0064	-0.6617	0.0057	-0.0001

表 4.4.3: 対数級数分布のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似,  
LD(定理 3) の相対誤差 ( $p = 0.4, n = 10$ )

$y$	真値 (%)	Edgeworth	定理 3
14	35.7480	0.0027	-0.1114
15	22.1303	-0.0008	-0.0720
16	12.9408	-0.0261	-0.0490
17	7.2265	-0.0513	-0.0349
18	3.8861	0.0099	-0.0258
19	2.0255	0.1747	-0.0196
20	1.0285	0.2860	-0.0154
21	0.5108	0.1661	-0.0124
22	0.2489	-0.1682	-0.0101
23	0.1193	-0.5388	-0.0085
24	0.0564	-0.8009	-0.0072
25	0.0263	-0.9328	-0.0062
26	0.0122	-0.9822	-0.0054
27	0.0056	-0.9963	-0.0048
28	0.0025	-0.9994	-0.0043
29	0.0011	-0.9999	-0.0039
30	0.0005	-1.0000	-0.0035

表 4.4.4: 対数級数分布のときの  $S_n$  の裾確率  $P\{S_n \geq y\}$  の値と, その値の Edgeworth 近似,  
LD(定理 3) の相対誤差 ( $p = 0.4, n = 25$ )

$y$	真値 (%)	Edgeworth	定理 3
33	47.3163	0.0007	-0.1143
34	36.2695	0.0007	-0.0837
35	26.7607	0.0005	-0.0624
36	19.0577	-0.0017	-0.0474
37	13.1358	-0.0071	-0.0366
38	8.7850	-0.0133	-0.0286
39	5.7167	-0.0136	-0.0229
40	3.6265	0.0014	-0.0185
41	2.2476	0.0355	-0.0152
42	1.3634	0.0804	-0.0126
43	0.8109	0.1142	-0.0106
44	0.4735	0.1090	-0.0090
45	0.2719	0.0441	-0.0077
46	0.1537	-0.0832	-0.0066
47	0.0856	-0.2558	-0.0057
48	0.0471	-0.4443	-0.0050
49	0.0255	-0.6193	-0.0044
50	0.0137	-0.7610	-0.0039

## 4.5. 2項分布と負の2項分布の和の分布の場合

確率変数  $X_1, X_2$  が互いに独立に  $X_1$  が2項分布  $B(n, p_1)$ ,  $X_2$  が負の2項分布  $NB(m, p_2)$  に従うとする. このとき  $Y := X_1 + X_2$  の分布について,  $m = n = 10, p_1 = p_2 = 0.3$  の場合,  $m = n = 10, p_1 = 0.6, p_2 = 0.3$  の場合の各点確率, 裾確率について精度の数値比較を行った (表 4.5.1~4.5.4 参照).

表 4.5.1:  $B(10, 0.3)$  と  $NB(10, 0.3)$  の混合分布のときの  $Y$  の各点確率  $P\{Y = y\}$  の値と, その Edgeworth 近似, 大偏差近似  $LD_1, LD_2$  の相対誤差

$y$	真値 (%)	Edgeworth	$LD_1$	$LD_2$
0	0.0000	10137.4900	—	—
1	0.0002	1119.7422	0.0829	-0.0059
2	0.0011	241.9324	0.0396	-0.0013
3	0.0043	73.4905	0.0247	-0.0007
4	0.0132	30.6834	0.0173	-0.0005
5	0.0334	14.7369	0.0130	-0.0005
6	0.0731	8.0290	0.0104	-0.0004
7	0.1420	4.7957	0.0087	-0.0004
8	0.2504	3.0646	0.0076	-0.0004
9	0.4076	2.0583	0.0070	-0.0004
10	0.6196	1.4343	0.0066	-0.0003
11	0.8886	1.0269	0.0063	-0.0003
12	1.2117	0.7499	0.0062	-0.0003
13	1.5811	0.5553	0.0062	-0.0002
14	1.9847	0.4150	0.0062	-0.0002
15	1.2407	0.3116	0.0062	-0.0002
16	1.2831	0.2341	0.0063	-0.0002
17	1.3240	0.1752	0.0063	-0.0002
18	3.6171	0.1298	0.0064	-0.0001
19	3.9484	0.0946	0.0065	-0.0001
20	4.2228	0.0671	0.0066	-0.0001
21	4.4327	0.0457	0.0067	-0.0001
22	4.5740	0.0292	0.0068	-0.0001
23	4.6461	0.0168	0.0068	-0.0001
24	4.6512	0.0081	0.0069	-0.0001
25	4.5940	0.0026	0.0070	-0.0001
26	4.4811	0.0001	0.0070	-0.0001
27	4.3204	0.0004	0.0071	-0.0001
28	4.1205	0.0029	0.0072	-0.0001
29	3.8901	0.0072	0.0072	-0.0001
30	3.6378	0.0125	0.0073	-0.0001



表 4.5.1(続)

$y$	真値 (%)	Edgeworth	$LD_1$	$LD_2$
31	3.3717	0.0178	0.0073	-0.0001
32	3.0988	0.0219	0.0074	-0.0001
33	2.8256	0.0233	0.0074	-0.0001
34	2.5573	0.0205	0.0075	-0.0001
35	2.2981	0.0118	0.0075	-0.0001
36	2.0514	-0.0042	0.0075	-0.0000
37	1.8197	-0.0287	0.0076	-0.0000
38	1.6044	-0.0625	0.0076	-0.0000
39	1.4065	-0.1056	0.0076	-0.0000
40	1.2263	-0.1576	0.0077	-0.0000
41	1.0637	-0.2171	0.0077	-0.0000
42	0.9180	-0.2821	0.0077	-0.0000
43	0.7886	-0.3500	0.0077	-0.0000
44	0.6743	-0.4174	0.0078	-0.0000
45	0.5741	-0.4810	0.0078	-0.0000

表 4.5.2:  $B(10, 0.6)$  と  $NB(10, 0.3)$  の混合分布のときの  $Y$  の各点確率  $P\{Y = y\}$  の値と, その Edgeworth 近似, 大偏差近似  $LD_1, LD_2$  の相対誤差

$y$	真値 (%)	Edgeworth	$LD_1$	$LD_2$
5	0.0019	138.2235	0.0078	-0.0004
6	0.0061	54.0080	0.0038	-0.0005
7	0.0163	24.7027	0.0011	-0.0007
8	0.0383	12.7966	-0.0004	-0.0008
9	0.0797	7.3051	-0.0011	-0.0008
10	0.1501	4.4944	-0.0012	-0.0007
11	0.2595	2.9279	-0.0010	-0.0006
12	0.4169	1.9920	-0.0005	-0.0005
13	0.6282	1.4003	0.0001	-0.0004
14	0.8957	1.0087	0.0008	-0.0004
15	1.2166	0.7397	0.0014	-0.0003
16	1.5833	0.5493	0.0020	-0.0002
17	1.9839	0.4112	0.0026	-0.0002
18	2.4035	0.3090	0.0031	-0.0001
19	2.8253	0.2322	0.0035	-0.0001
20	3.2319	0.1737	0.0040	-0.0001

表 4.5.2(続)

$y$	真値 (%)	Edgeworth	$LD_1$	$LD_2$
21	3.6074	0.1285	0.0043	-0.0001
22	3.9377	0.0935	0.0047	-0.0001
23	4.2117	0.0661	0.0050	-0.0001
24	4.4218	0.0449	0.0052	-0.0000
25	4.5638	0.0285	0.0055	-0.0000
26	4.6368	0.0163	0.0057	-0.0000
27	4.6431	0.0078	0.0059	-0.0000
28	4.5872	0.0025	0.0061	-0.0000
29	4.4757	0.0001	0.0062	-0.0000
30	4.3164	0.0005	0.0064	-0.0000
31	4.1178	0.0031	0.0065	-0.0000
32	3.8887	0.0074	0.0066	-0.0000
33	3.6375	0.0126	0.0067	-0.0000
34	3.3722	0.0179	0.0068	-0.0000
35	3.1001	0.0218	0.0069	-0.0000
36	2.8274	0.0232	0.0070	-0.0000
37	2.5595	0.0203	0.0071	-0.0000
38	2.3006	0.0116	0.0072	-0.0000
39	2.0541	-0.0042	0.0072	-0.0000
40	1.8224	-0.0283	0.0073	-0.0000
41	1.6071	-0.0616	0.0073	-0.0000
42	1.4092	-0.1040	0.0074	-0.0000
43	1.2289	-0.1552	0.0074	-0.0000
44	1.0661	-0.2139	0.0075	-0.0000
45	0.9202	-0.2780	0.0075	-0.0000
46	0.7906	-0.3449	0.0076	-0.0000
47	0.6762	-0.4116	0.0076	-0.0000
48	0.5758	-0.4746	0.0076	-0.0000
49	0.4883	-0.5305	0.0077	-0.0000
50	0.4124	-0.5763	0.0077	-0.0000
51	0.3470	-0.6093	0.0077	-0.0000
52	0.2909	-0.6278	0.0078	-0.0000
53	0.2430	-0.6310	0.0078	-0.0000
54	0.2024	-0.6192	0.0078	-0.0000
55	0.1679	-0.5938	0.0078	-0.0000
56	0.1389	-0.5572	0.0078	-0.0000
57	0.1146	-0.5124	0.0079	-0.0000
58	0.0942	-0.4632	0.0079	-0.0000
59	0.0773	-0.4134	0.0079	-0.0000
60	0.0632	-0.3669	0.0079	-0.0000

表 4.5.3:  $B(10, 0.3)$  と  $NB(10, 0.3)$  の混合分布のときの  $Y$  の裾確率  $P\{Y \geq y\}$  の値と, その近似の相対誤差

$y$	真値 (%)	Edgeworth	定理 3
27	45.1432	0.0002	-0.1243
28	40.8228	0.0002	-0.1110
29	36.7023	0.0002	-0.0993
30	32.8122	0.0002	-0.0890
31	29.1744	0.0000	-0.0800
32	25.8027	-0.0003	-0.0721
33	22.7039	-0.0010	-0.0651
34	19.8783	-0.0020	-0.0589
35	17.3210	-0.0034	-0.0534
36	15.0229	-0.0052	-0.0485
37	12.9715	-0.0073	-0.0442
38	11.1518	-0.0094	-0.0404
39	9.5474	-0.0112	-0.0369
40	8.1409	-0.0122	-0.0339
41	6.9147	-0.0120	-0.0311
42	5.8510	-0.0100	-0.0286
43	4.9330	-0.0057	-0.0264
44	4.1444	0.0012	-0.0244
45	3.4701	0.0108	-0.0226
46	2.8960	0.0232	-0.0210
47	2.4092	0.0380	-0.0195
48	1.9980	0.0543	-0.0181
49	1.6521	0.0713	-0.0169
50	1.3622	0.0876	-0.0158
51	1.1200	0.1017	-0.0148
52	0.9184	0.1119	-0.0138
53	0.7511	0.1167	-0.0130
54	0.6127	0.1145	-0.0122
55	0.4986	0.1042	-0.0114
56	0.4048	0.0848	-0.0108
57	0.3279	0.0559	-0.0102
58	0.2650	0.0175	-0.0096
59	0.2167	-0.0299	-0.0091
60	0.1719	-0.0854	-0.0086
61	0.1381	-0.1477	-0.0081
62	0.1106	-0.2152	-0.0077
63	0.0885	-0.2862	-0.0073
64	0.0707	-0.3587	-0.0070
65	0.0563	-0.4311	-0.0066

表 4.5.4:  $B(10, 0.6)$  と  $NB(10, 0.3)$  の混合分布のときの  $Y$  の裾確率  $P\{Y \geq y\}$  の値と, その近似の相対誤差

$y$	真値 (%)	Edgeworth	定理 3
30	45.1771	0.0002	-0.1229
31	40.8607	0.0002	-0.1098
32	36.7428	0.0002	-0.0984
33	32.8542	0.0002	-0.0883
34	29.2167	-0.0000	-0.0794
35	25.8445	-0.0004	-0.0715
36	22.7445	-0.0011	-0.0646
37	19.9170	-0.0021	-0.0585
38	17.3576	-0.0035	-0.0531
39	15.0569	-0.0053	-0.0483
40	13.0028	-0.0073	-0.0440
41	11.1804	-0.0093	-0.0402
42	9.5732	-0.0110	-0.0368
43	8.1641	-0.0120	-0.0337
44	6.9352	-0.0117	-0.0310
45	5.8692	-0.0096	-0.0285
46	4.9489	-0.0054	-0.0263
47	4.1583	0.0014	-0.0243
48	3.4821	0.0110	-0.0225
49	2.9063	0.0232	-0.0209
50	2.4180	0.0377	-0.0194
51	2.0056	0.0538	-0.0181
52	1.6586	0.0705	-0.0169
53	1.3676	0.0865	-0.0157
54	1.1246	0.1004	-0.0147
55	0.9222	0.1105	-0.0138
56	0.7543	0.1153	-0.0129
57	0.6154	0.1133	-0.0122
58	0.5009	0.1032	-0.0114

## 5. 検出力の数値計算への応用

定理3の応用として仮説検定問題における検出力の数値計算を行う。まず,  $X_1, \dots, X_n$  が互いに独立にいずれも2項分布  $B(1, p)$  に従うとする。このとき,  $T := \sum_{i=1}^n X_i$  とおくと,  $T$  は2項分布  $B(n, p)$  に従う。従って, 帰無仮説  $H: p = p_0$  の対立仮説  $K: p = p_1 (> p_0)$  に対する検定問題を考えると, 水準  $\alpha$  ( $0 < \alpha < 1$ ) のランダム検定関数は

$$\phi(T) = \begin{cases} 1 & (T > t), \\ u & (T = t), \\ 0 & (T < t) \end{cases}$$

になる。ただし,  $u$  は  $0 < u < 1$  とし  $E_{p_0}[\phi(T)] = \alpha$  を満たすものとする。次に,  $T$  とは独立に区間  $[0, 1]$  上の一様分布に従う確率変数  $U$  を考え,  $Y := T + U$  とすれば, 上のランダム検定関数は  $Y$  に基づく非ランダム検定関数

$$\phi^*(Y) = \begin{cases} 1 & (Y > t - u + 1), \\ 0 & (Y \leq t - u + 1) \end{cases}$$

と同等になる。このとき,  $\bar{y}(p_0) := t - u + 1$  とおいて, 対立仮説の下で定理3の上側裾確率の近似式を用いることによって, 検出力  $P_{p_1}[T \geq \bar{y}(p_0)]$  を漸近的に求めることができる(表5.1参照)。

表5.1: 片側検定における検出力の近似 ( $\alpha = 0.05$ )

$p_0$	$p_1$	$n = 20$	$n = 50$	$p_0$	$p_1$	$n = 20$	$n = 50$
0.1	0.15	0.1773	0.2916	0.6	0.62	0.0710	0.0860
0.2	0.22	0.0725	0.0962	0.6	0.65	0.1160	0.1767
0.2	0.25	0.1309	0.2145	0.7	0.75	0.1133	0.1885
0.3	0.35	0.1240	0.1896	0.8	0.82	0.0556	0.0964
0.4	0.45	0.1141	0.1750	0.8	0.85	0.0987	0.2285
0.5	0.55	0.1111	0.1710	0.9	0.95	—	0.3245

## 6. 統計的予測への応用

観測されるデータを確率変数  $X$ , 未観測の確率変数  $Y$  とし,  $X$  と  $Y$  は互いに独立に,  $X, Y$  はそれぞれポアソン分布  $Po(m\lambda)$ ,  $Po(n\lambda)$  に従うとする。ただし,  $m, n$  は既知の自然数,  $\lambda$  は正值で未知母数とする。このとき,  $X$  に基づいて  $Y$  の区間予測を行う ([HA99])。まず, 統計量  $T := X + Y$  は  $\lambda$  に対する完備十分統計量であり,  $T$  の分布は  $Po((m+n)\lambda)$  になる。このとき,  $T = t$  を与えたときの  $Y$  の条件付分布は2項分布  $B(t, n/(m+n))$  になり, これは  $\lambda$  に無関係になる。このことは, 完備十分統計量  $T$  に基づく  $Y$  の予測区間が未知の母数  $\lambda$  に無関係に構成できることを意味している。そこで, 任意の  $\alpha$  ( $0 < \alpha < 1$ ) について2項分布  $B(t, n/(m+n))$  の上側  $100(\alpha/2)\%$  点を  $y_{\alpha/2}(t)$  とすれば,

$$P\{t - y_{\alpha/2}(t) \leq Y \leq y_{\alpha/2}(t) | T = t\} = 1 - \alpha \quad (6.1)$$

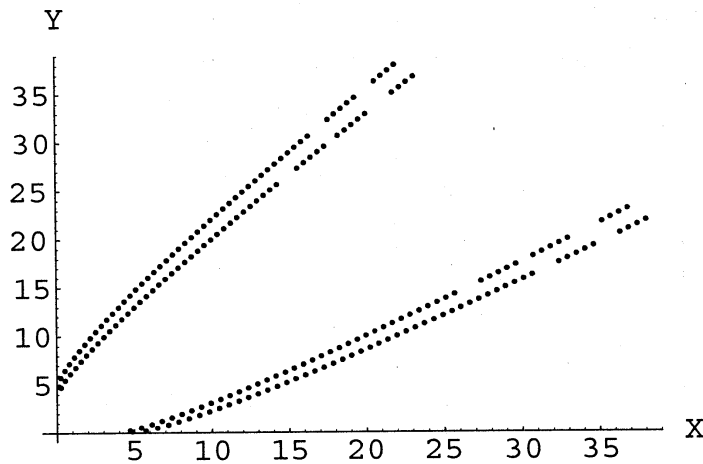


図 1:  $m = n = 25$  のときの  $Y$  の予測区間 ( $\alpha = 0.05, \alpha = 0.1$ )

より

$$P\{a(X) \leq Y \leq b(X)\} = 1 - \alpha$$

の形になり,  $[a(X), b(X)]$  は  $Y$  の信頼係数  $1 - \alpha$  の予測区間になる. よって, 自然数  $m, n$  と  $t = x + y$  について 2 項分布  $B(t, n/(m+n))$  の上側  $100(\alpha/2)\%$  点  $y_{\alpha/2}(t)$  の近似値を, 定理 3 の上側裾確率の近似式から求めて, (6.1) から  $Y$  の予測区間を漸近的に得ることができる (図 1 参照).

実際には分布のパーセント点の近似を求めることになるので, この場合について定理 3 から上側パーセント点の近似, 系 2 から下側パーセント点の近似をそれぞれ求め, Cornish-Fisher 近似と比較を行った (表 6.1~6.3 参照).

表 6.1:  $B(1, p_j)$  ( $\{p_j\}_{j=1}^9 = 0.05(0.05)0.95$ ) のときの  $S_n$  のパーセント点の値と, その値の Cornish-Fisher 近似, 大偏差近似 LD の誤差

	真値	Cornish-Fisher	LD
上側 0.5%	14.7521	-0.2497	-0.1594
1%	14.2675	-0.1645	-0.1055
2.5%	13.6521	-0.1579	-0.1304
5%	12.9762	-0.0233	-0.0114
10%	12.3982	-0.0847	-0.0815
下側 10%	6.6018	0.0603	0.0815
5%	6.0238	-0.0432	0.0115
2.5%	5.3479	0.0232	0.1304
1%	4.7325	-0.1066	0.1055
0.5%	4.2479	-0.1662	0.1594

表 6.2:  $B(1, p_j)$  ( $\{p_j\}_{j=1}^2 = 0.0(0.03)0.60$ ) のときの  $S_n$  のパーセント点の値と, その値の Cornish-Fisher 近似, 大偏差近似  $LD$  の誤差

$y$	真値	Chonish-Fisher	LD
上側 0.5%	11.9344	-0.2373	-0.0534
1%	11.5205	-0.2605	-0.1408
2.5%	10.7455	-0.1573	-0.1031
5%	10.0153	-0.0279	-0.0070
10%	9.3537	-0.0764	-0.0756
下側 10%	3.2983	0.0462	0.0658
5%	2.6357	0.0328	0.0942
2.5%	2.1280	-0.0521	0.0701
1%	1.4377	-0.0679	0.1679
0.5%	1.1143	-0.2437	0.1064

表 6.3:  $B(20, 0.5)$  のときの  $S_n$  のパーセント点の値と, その値の Cornish-Fisher 近似, 大偏差近似  $LD$  の誤差

$y$	真値	Chonish-Fisher	LD
上側 0.5%	16.1967	-0.2438	-0.0762
1%	15.7233	-0.2319	-0.1183
2.5%	14.8835	-0.1076	-0.0517
5%	14.2072	-0.0787	-0.0531
10%	13.4273	-0.0745	-0.0666
下側 10%	6.5727	0.0230	0.0666
5%	5.7928	-0.0617	0.0531
2.5%	5.1165	0.1797	0.0517
1%	4.2767	-0.3654	0.1184
0.5%	3.8033	-0.7226	0.0763

## 参考文献

- [ATT99] Akahira, M., Takahashi, K. and Takeuchi, K. (1999). The higher order large-deviation approximation for the distribution of the sum of independent discrete random variables. *Commun. Statist. - Theory Meth.*, **28**(3&4), 705–726.
- [BC89] Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.
- [D87] Daniels, H. E. (1987). Tail probability approximations. *Int. Statist. Review* **55**, 37–48.
- [HA99] 飛田英祐, 赤平昌文 (1999). 離散指数型分布族における区間予測とその応用. 京都大学 数理解析研究所講究録 **1101**, 43–68.
- [J95] Jensen, J. L. (1995). *Saddlepoint Approximations*. Clarendon Press, Oxford.
- [JKK92] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions* (2nd ed.). Wiley, New York.
- [T75] 竹内啓 (1975). 確率分布の近似. 教育出版.
- [TA98] 高橋邦彦, 赤平昌文 (1998). Large-deviation approximations for the discrete distribution. 京都大学 数理解析研究所講究録 **1055**, 45–67.
- [TF81] 竹内啓, 藤野和建 (1981). 2項分布とポアソン分布. 東京大学出版会.