Brazilian Journal of Probability and Statistics
2013, Vol. 27, No. 2, 150–161
DOI: 10.1214/11-BJPS162

On a saddlepoint approximation to the Markov binomial distribution

Jens Ledet Jensen

University of Aarhus

Abstract. A nonstandard saddlepoint approximation to the distribution of a sum of Markov dependent trials is introduced. The relative error of the approximation is studied, not only for the number of summands tending to infinity, but also for the parameter approaching the boundary of its definition range. A comparison is made with another recent study of Markov dependent trials.

1 Introduction

Motivated by recent studies, we consider in this paper a saddlepoint approximation to the Markov binomial distribution, that is, the distibution of $S_n = \sum_{i=1}^n X_i$, where $X_1, X_2, \dots$ is a Markov chain on the state space $\{0, 1\}$. Let the transition probabilities be parameterized by $\alpha = P(X_{n+1} = 1|X_n = 0)$ and $\beta = P(X_{n+1} = 1|X_n = 1)$. Broadly speaking, approximations can be divided into the Gaussian type and the compound Poisson type. In the first type the approximation becomes exact in the limit of a Gaussian distribution only, whereas the second type handles cases with $\alpha \to 0$, typically with α of order $\frac{1}{n}$. The possible limiting distributions when α and β depend on n can be seen in Dobrushin (1961). The approximation of Xia and Zhang (2009) is of the Gaussian type. The Markov binomial distribution is approximated by either a binomial distribution or a negative binomial distribution obtained by fitting the first two moments. This approximation is exact when $\alpha = \beta$, but otherwise becomes exact for a limiting Gaussian distribution only. Most importantly, though, Xia and Zhang (2009) provide an explicit upper bound on the total variation distance of the approximation, which is of order $1/\sqrt{n}$ for fixed values of α and β. In Čekanavičius and Vellaisamy (2010) approximations of the compound Poisson type are considered. If α is of order $\frac{1}{n}$ the error is of order $\frac{1}{n}$ as well, and for fixed values of α and β the error is of order $\frac{1}{\sqrt{n}}$. However, one should keep in mind that the approximating measure in Čekanavičius and Vellaisamy (2010) is by itself complicated to evaluate. The saddlepoint approximation we suggest in this paper is of the Gaussian type and is thus mostly inspired by Xia and Zhang (2009). For fixed values of α and β the *relative* error is of order $o(\frac{1}{n})$. Contrary to

Key words and phrases. Markov dependent trials, relative error, saddlepoint approximation.
Received May 2011; accepted June 2011.

the above mentioned approximations, the saddlepoint approximation has bounded relative error throughout the parameter space.

The different approximations to the Markov binomial distribution have mostly theoretical interest. The exact distribution can be calculated recursively. Define $p(k,a,n) = P(S_n = k, X_n = a)$. Then by splitting an event according to the value of X_n we find

$$\begin{aligned} p(k,0,n+1) &= p(k,0,n)(1-\alpha) + p(k,1,n)(1-\beta), \\ p(k,1,n+1) &= p(k-1,0,n)\alpha + p(k-1,1,n)\beta, \end{aligned} \tag{1.1}$$

with initial values $p(0,0,1) = P(X_1 = 0)$, $p(1,1,1) = P(X_1 = 1)$, $p(0,1,1) = p(1,0,1) = 0$, and with $p(k,a,n) = 0$ for $k > n$. This simple recursion was stated in Ladd (1975), and can be considered analogous to standard recursions within the field of hidden Markov chains. The recursion is useful when all of the distribution is wanted. For n of the order 10^4 the calculation is feasible. If only one point probability is needed, the sum formula of Gabriel (1959) can be used,

$$\begin{aligned} P(S_n = k) &= \{P(X_0 = 0)G_0(k,n) + P(X_0 = 1)G_1(k,n)\}\beta^k(1-\alpha)^{n-k}, \\ G_0(k,n) &= \sum_{m=1}^{c_0} \binom{n-k}{\lfloor m/2\rfloor}\binom{k-1}{\lceil m/2\rceil - 1}\left(\frac{1-\beta}{1-\alpha}\right)^{\lfloor m/2\rfloor}\left(\frac{\alpha}{\beta}\right)^{\lceil m/2\rceil}, \\ G_1(k,n) &= \sum_{m=1}^{c_1} \binom{k}{\lfloor m/2\rfloor}\binom{n-k-1}{\lceil m/2\rceil - 1}\left(\frac{1-\beta}{1-\alpha}\right)^{\lceil m/2\rceil}\left(\frac{\alpha}{\beta}\right)^{\lfloor m/2\rfloor}, \end{aligned}$$

where $c_0 = \min\{2k, 2(n-k)+1\}$ and $c_1 = \min\{2k+1, 2(n-k)\}$. For n of the order 10^6 the calculation using this formula is feasible. Thus, the different approximations become of practical interest only for very large n.

In Section 2 we introduce the saddlepoint approximation and study the relative error of the approximation. In Section 3 we compare the approximation with that of Xia and Zhang (2009) and study the upper bound given in that paper. The Appendix provides details of the saddlepoint approximation.

2 Saddlepoint approximation

When β is of order $1-\alpha$ and α is small the distribution of S_n has high point probabilities at zero and n and is almost uniform in between. Most approximations will fail for this case. To handle this, we calculate the probabilities $P(S_n = 0)$ and $P(S_n = n)$ exactly and use the saddlepoint approximation for the conditional distribution given that $0 < S_n < n$. This is motivated by numerical investigations which indicate that the density $P(S_n = k)$, $k = 1, \dots, n-1$, in between the two extremes is log concave. In relation to the saddlepoint approximation there are two extreme cases of a log concave density. One is the uniform distribution mentioned already above, and the other is a very peaked distribution. The latter case is encountered as $\alpha \to 1$ and a slight modification of the traditional saddlepoint approximation is needed to handle this case. For comparison with Xia and Zhang (2009) we consider the stationary case with $p_0 = P(X_1 = 0) = (1-\beta)/(1-\beta+\alpha)$ and $p_1 = P(X_1 = 1) = \alpha/(1-\beta+\alpha)$, but the investigations of this section can easily be redone with other choices of p_0 and p_1.

Let $q_k = P(S_n = k|0 < S_n < n)$ be the conditional probabilities and let $\phi(z) = \sum_{k=1}^{n-1} z^k q_k$ be the moment generating function. In terms of the moment generating function $\psi(z)$ of S_n, we have

$$\begin{aligned}\phi(z) &= [\psi(z) - P(S_n = 0) - z^n P(S_n = n)]/P(0 < S_n < n)\\ &= \frac{[\psi(z) - p_0(1-\alpha)^{n-1} - z^n p_1 \beta^{n-1}]}{[1 - p_0(1-\alpha)^{n-1} - p_1\beta^{n-1}]},\end{aligned} \tag{2.1}$$

and using the recursion (1.1) for $\psi(z) = \sum_{k=0}^{n} z^k [p(k,0,n) + p(k,1,n)]$, we find

$$\psi(z) = (p_0, zp_1) P_0(z)^{n-1}(1,1)^{\mathrm{T}}, \qquad P_0(z) = \begin{pmatrix} 1-\alpha & \alpha z \\ 1-\beta & \beta z\end{pmatrix}. \tag{2.2}$$

Define the *exponentially tilted* distribution as $q(k,z) = q_k z^k/\phi(z)$, $k = 1, \dots, n-1$. For a fixed $k = 2, \dots, n-2$ we choose z_k such that the tilted distribution has mean k and consider an approximation of the form

$$q_k = \phi(z_k) z_k^{-k} q(k, z_k) \approx \phi(z_k) z_k^{-k} A(k), \tag{2.3}$$

where $A(k)$ is an approximation to $q(k, z_k)$. In this way the approximation problem has been centered in that we seek an approximation to the point probability at the mean of the distribution. The traditional saddlepoint approximation, including $O(\frac{1}{n})$ terms, takes the form

$$A(k) = \frac{1}{\sqrt{2\pi\sigma(k)^2}} \left\{1 + \frac{1}{8}\gamma_4(k) - \frac{5}{24}\gamma_3(k)^2\right\}, \tag{2.4}$$

where $\sigma(k)^2$ is the variance and $\gamma_3(k)$ and $\gamma_4(k)$ are the third and fourth standardized cumulants of the exponentially tilted distribution. The saddlepoint approximation was originally developed for situations with a limiting normal distribution, but including the $O(\frac{1}{n})$ term makes the approximation widely applicable. For discrete distributions the approximation will, however, fail in cases where the variance $\sigma(k)^2$ is small, corresponding to a distribution almost concentrated in one point. We therefore make the following alternative approximation:

$$A(k) = 1 - \sigma(k)^2 \qquad \text{if } \sigma(k)^2 < 0.4. \tag{2.5}$$

The value $1 - \sigma^2$ comes from the probability at the center of a symmetric three point distribution with variance σ^2. Details of the approximation are given in the Appendix.

Before embarking on a detailed discussion of the Markov binomial distribution we illustrate how log concavity bounds the error of the saddlepoint approximation. Consider a variable X with a log concave density symmetric around $x = 0$, and consider the main term of the saddlepoint approximation (2.3), that is, $1/\sqrt{2\pi\sigma^2}$ if $\sigma^2 \geq 0.4$ from (2.4) and $1-\sigma^2$ if $\sigma^2 < 0.4$ from (2.5), where σ^2 is the variance. For a fixed value p_0 of $P(X = 0)$ the smallest variance is obtained with an almost uniform density: $P(X = j) = p_0$ for $|j| \leq j_*$, where j_* is the largest integer less than or equal to $(1 - p_0)/(2p_0)$, and $P(X = \pm(j_* + 1)) = (1 - p_0 - 2p_0 j_*)/2$. The largest variance is obtained with a discrete Laplace distribution: $P(X = j) = \theta^{|j|}(1-\theta)/(1+\theta)$ with $\theta = (1 - p_0)/(1 + p_0)$. We thus have bounds on the ratio of the approximation to the true probability. These bounds are shown in Figure 1 as the full drawn line and the dashed line. Also included in the figure is the improved approximation (2.4).

We start the investigation of the Markov binomial distribution with the case of fixed parameter values as n tends to infinity.

Proposition 1. *For fixed values of α and β the saddlepoint approximation given through (2.3) and (2.4) has relative error of order $o(\frac{1}{n})$ in a large deviation region for k. In particular, it follows that the total variation distance is of order $o(\frac{1}{n})$. For*

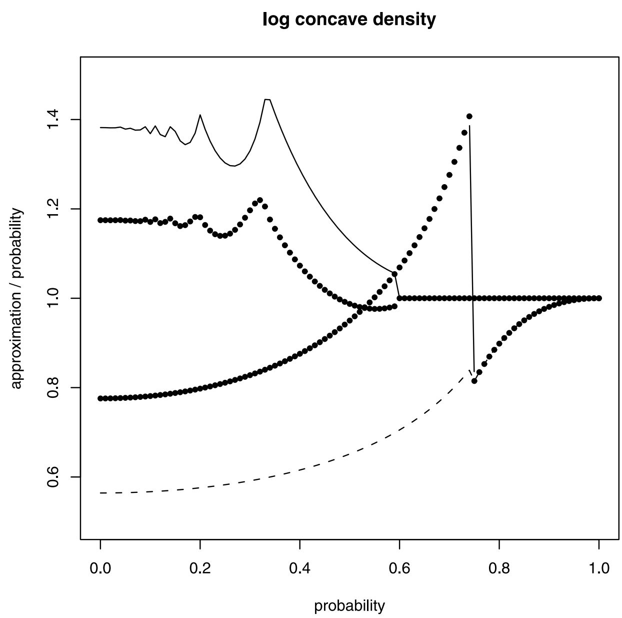

Figure 1 *Ratio of saddlepoint approximation to the true probability at the center point for a symmetric log concave density. The full drawn line is for the main term of the saddlepoint approximation and an almost uniform density, and the dashed line is for a discrete Laplace distribution. The dotted lines are for the same cases when using the saddlepoint approximation in* (2.4) *and* (2.5).

$n \to \infty$ the relative error of the saddlepoint approximation at the extremes of the distribution ($k = 2$ *and* $k = n - 2$) *is approximately* 0.0089.

Proof. The relative error in a large deviation region comes from standard results (see, e.g., Jensen (1995), Chapter 9).

In order to consider the extreme case with $k = 2$ (and similarly with $k = n - 2$), we consider the exponentially tilted distribution with moment generation function $\phi(\frac{\gamma}{n}z)/\phi(\frac{\gamma}{n})$. Using an eigenvalue decomposition of $P_0(\frac{\gamma}{n}z)$, we find that the limit as $n \to \infty$ is a Poisson distribution with mean $2\gamma\alpha(1-\beta)/(1-\alpha)^2$, conditioned on being greater than zero. The saddlepoint approximation to the Poisson distribution can be calculated numerically. This gives the stated relative error for the extreme cases. □

The result of Proposition 1 that the total variation distance of the approximation is $o(\frac{1}{n})$ may be compared to the order $O(\frac{1}{\sqrt{n}})$ of the approximation of Xia and Zhang (2009). A small relative error in a large deviation region is, however, a stronger statement, allowing us to approximate tail probabilities that are much smaller than $o(\frac{1}{n})$. The proposition also shows that in the very extreme tail of the distribution the approximation gives the correct order of the probability.

We next turn to a discussing of the relative error of the approximation when (α, β) approaches the boundary of the parameter space. We state the results in terms of $z^k P(S_n = k)$ for suitable z, in which case the exponentially tilted distribution $z^k P(S_n = k)/[P(0 < S_n < n)\phi(z)]$ is obtained by a normalization. We first consider the case with $\alpha \to 0$ where $P(S_n = 0) \to 1$, and the conditioning on $0 < S_n < n$ becomes important.

Proposition 2. *For* $\alpha \to 0$ *with* $\alpha/\beta \to 0$ *we have that*

$$\begin{aligned}\frac{\beta(1-\beta+\alpha)}{\alpha(1-\beta)\beta^k} P(S_n = k) \\ \sim 2 + (1-\beta)(n-1-k), \qquad k = 1, \ldots, n-1.\end{aligned} \tag{2.6}$$

The saddlepoint approximation (2.3) *to the limiting distribution in* (2.6) *has maximal relative error between* 0.17 *and* 0.18 *for* $n \geq 7$.

Proof. The event $\{S_n = k\}$ is the union of cases where $x_j = 1$ for k consecutive times and cases where these are not consecutive. In the former case the probability is of order $\alpha(1-\beta)\beta^{k-1}/(1-\beta+\alpha)$, and in the latter case the probability is of the same order multiplied by α/β. Thus, in the limit $\alpha \to 0$ with $\alpha/\beta \to 0$, we need only consider the consecutive cases. This gives (2.6).

The density in (2.6) is clearly log concave and the most extreme case, when using the saddlepoint approximation, is the uniform distribution obtained for $\beta \to 1$.

For a uniform discrete distribution on the numbers $\{1, 2, \ldots, m\}$ the cumulant transform is given by $\kappa(s) = \log(1 - e^{sm}) + \log(\omega/m)$ with $\omega = e^s/(1 - e^s)$. For $k = 2, \ldots, m - 1$ let s be the saddlepoint determined by $\kappa'(s) = k$. For symmetry reasons we need only consider $k \leq \frac{m}{2}$, which is covered by $-\frac{1}{2} \leq s \leq 0$. We first consider the tail of the distribution with $m^4 e^{sm} \to 0$ as $m \to \infty$. Then the relative error of the saddlepoint approximation tends to

$$\frac{\omega e^{-s[1+\omega]}}{\sqrt{2\pi\omega^2/e^s}}\left\{1 + \frac{\omega + 6\omega^2 + 6\omega^3}{8\omega^2/e^s} - \frac{5[\omega + 2\omega^2]^2}{\sqrt{e^s}\,\omega^{3/2}}\right\} - 1,$$

as $m \to \infty$. This is in the limiting situation of a geometric distribution, and the largest absolute relative error is less than 0.013 for $k = 2$ corresponding to $s = -\log(2)$. Next, for the center of the distribution let $s = \frac{a}{m+1}$ with $-5\log(m) \leq a \leq 0$ and let now $\omega = e^a/(1 - e^a)$. The limiting relative error as $m \to \infty$ is

$$\frac{e^a e^{1+a\omega}/[(-a)\omega]}{\sqrt{2\pi(1/a^2 - \omega - \omega^2)}}\{1 + f_1(a) - f_2(a)\} - 1,$$

$$f_1(a) = \frac{6/a^4 - \omega - 7\omega^2 - 12\omega^3 - 6\omega^4}{8(1/a^2 - \omega - \omega^2)^2},$$

$$f_2(a) = \frac{5[2/(-a)^3 - \omega - 3\omega^2 - 2\omega^3]^2}{24(1/a^2 - \omega - \omega^2)^{3/2}}.$$

The largest relative error is less than 0.18 and is obtained for $a = 0$ corresponding to the center value $k = (m + 1)/2$. Numerical investigations show that the maximal relative error is between 0.17 and 0.18 for m down to 6 (see in this connection also Figure 1). □

By an interchange of the two states Proposition 1 also covers the case $\beta \to 1$ with $(1 - \beta)/(1 - \alpha) \to 0$. Similarly, the case where α and β tend to zero at the same rate is covered by the next proposition, where we turn to limiting cases with $\alpha \to 1$.

Proposition 3. *For* $\alpha \to 1$ *with* $(1 - \alpha)\beta/(1 - \beta) \to 0$ *we get*

$$\beta(2-\beta)\left(\frac{\beta}{1-\beta}\right)^n\left(\frac{1-\beta}{\beta^2}\right)^k P(S_n = k)$$
$$\sim \begin{cases} \binom{k-1}{n-k} + 2\beta\binom{k-1}{n-k-1} + \beta^2\binom{k-1}{n-k-2}, & k \geq \dfrac{n-1}{2}, \\ o(1), & k < \dfrac{n-1}{2}, \end{cases} \tag{2.7}$$

and

$$\begin{aligned}
&\frac{2-\beta}{(1-\beta)(1-\alpha)^{n-1}}\left(\frac{(1-\alpha)^2}{1-\beta}\right)^k P(S_n=k) \\
&\sim \begin{cases} o(1), & k > \dfrac{n-1}{2}, \\ \dbinom{n-k-1}{k}, & k \le \dfrac{n-1}{2}. \end{cases}
\end{aligned} \tag{2.8}$$

In both of the above cases the nonvanishing part of the density is log concave.

Finally, for $\alpha \to 1$ *and* $(1-\beta)/(1-\alpha) \to \xi$, *where* $0 < \xi < \infty$, *we can express* $P(S_n = k)(1-\alpha)^k/[(1-\alpha)^n(1+\sqrt{\xi})^{n-1}]$ *through probabilities calculated from the Markov chain with transition probability* $\sqrt{\xi}/(1+\sqrt{\xi})$ *from state* 0 *to state* 1 *and transition probability* $1/(1+\sqrt{\xi})$ *from state* 1 *to state* 1.

Proof. The different cases are all based on the following rewriting of the moment generating function:

$$\begin{aligned}
\psi(\gamma z) &= \sum_{k=0}^{n}(\gamma z)^k p(S_n=k) = (p_0, \gamma z p_1)P_0(\gamma z)^{n-1}(1,1)^{\mathrm{T}} \\
&= \lambda(\gamma)^{n-1}(p_0, \gamma z v(\gamma) p_1)P_0(\gamma; z)^{n-1}\big(1, 1/v(\gamma)\big)^{\mathrm{T}},
\end{aligned}$$

where $\lambda(\gamma)$ is the maximal eigenvalue of $P_0(\gamma)$, $(1, v(\gamma))^{\mathrm{T}}$ is a right eigenvector, $v(\gamma) = [\lambda(\gamma) - (1-\alpha)]/(\alpha\gamma)$, and

$$P_0(\gamma; z) = \begin{pmatrix} (1-\alpha)/\lambda(\gamma) & z\alpha\gamma v(\gamma)/\lambda(\gamma) \\ (1-\beta)/(\lambda(\gamma)v(\gamma)) & z\beta\gamma/\lambda(\gamma) \end{pmatrix}.$$

When $(1-\alpha)\beta/(1-\beta) \to 0$ and $\gamma = (1-\beta)/\beta^2$ we find

$$\lambda(\gamma) \sim \frac{1-\beta}{\beta\varepsilon}, \qquad v(\lambda) \sim \frac{\beta}{\varepsilon}, \qquad P_0(\gamma; z) \to \begin{pmatrix} 0 & z \\ 1-\varepsilon & \varepsilon z \end{pmatrix},$$

where $\varepsilon = 2/(1+\sqrt{5})$. When instead $\gamma = (1-\alpha)^2/(1-\beta)$ we find

$$\lambda(\gamma) \sim \frac{1-\alpha}{\varepsilon}, \qquad v(\lambda) \sim \frac{(1-\beta)\varepsilon}{(1-\alpha)}, \qquad P_0(\gamma; z) \to \begin{pmatrix} \varepsilon & (1-\varepsilon)z \\ 1 & 0 \end{pmatrix}.$$

Finally, when $(1-\beta)/(1-\alpha) \to \xi$ and $\gamma = 1-\alpha$ we obtain

$$\lambda(\gamma) \sim (1-\alpha)\big(1+\sqrt{\xi}\big), \qquad v(\lambda) \to \sqrt{\xi},$$

$$P_0(\gamma; z) \to \begin{pmatrix} \dfrac{1}{1+\sqrt{\xi}} & \dfrac{z\sqrt{\xi}}{1+\sqrt{\xi}} \\ \dfrac{\sqrt{\xi}}{1+\sqrt{\xi}} & \dfrac{z}{1+\sqrt{\xi}} \end{pmatrix}.$$

Thus, to obtain the results of the proposition, we need to find powers of specialized matrices. By tedious calculations we obtain

$$\begin{aligned}
&\begin{pmatrix} 0 & z \\ 1-\varepsilon & \varepsilon z \end{pmatrix}^{n-1} \\
&\quad = \frac{\varepsilon^n}{1-\varepsilon} \sum_{k=n^*}^{n-1} z^k \left\{ \binom{k-1}{n-k-1} \begin{pmatrix} 0 & 1 \\ 0 & \varepsilon \end{pmatrix} \right. \\
&\qquad\qquad \left. + \binom{k-1}{n-k-2} \begin{pmatrix} \varepsilon & 0 \\ \varepsilon^2 & \varepsilon \end{pmatrix} + \binom{k-1}{n-k-3} \begin{pmatrix} 0 & 0 \\ \varepsilon^2 & 0 \end{pmatrix} \right\},
\end{aligned}$$

and

$$\begin{aligned}
&\begin{pmatrix} 1-\varepsilon & \varepsilon z \\ 1 & 0 \end{pmatrix}^{n-1} \\
&\quad = (1-\varepsilon)^{n-2} \sum_{k=0}^{n_*} z^k \left\{ \binom{n-k-2}{k} \begin{pmatrix} 1-\varepsilon & 0 \\ 1 & 0 \end{pmatrix} \right. \\
&\qquad\qquad + \binom{n-k-2}{k-1} \begin{pmatrix} 1-\varepsilon & (1-\varepsilon)^2 \\ 0 & 1-\varepsilon \end{pmatrix} \\
&\qquad\qquad \left. + \binom{n-k-2}{k-2} \begin{pmatrix} 0 & (1-\varepsilon)^2 \\ 0 & 0 \end{pmatrix} \right\},
\end{aligned}$$

where n^* is the smallest integer greater than or equal to $(n-1)/2$, and n_* is the largest integer smaller than or equal to $(n-1)/2$.

To show log concavity of a function $h(k)$, we must show that $h(k)h(k+2) \leq h(k+1)^2$. This is easy to show for the nonvanishing part of (2.8). For the nonvanishing part of (2.7) we must show that a fourth degree polynomial in $0 < \beta < 1$ is nonnegative, and a tedious calculation shows that the coefficients of the polynomial are indeed nonnegative. □

Since the two cases in (2.7) and (2.8) have different scalings, the exponentially tilted distribution with mean $(n-1)/2$ for n odd or $n/2$ or $(n-2)/2$ for n even become concentrated at one point for $\alpha \to 1$. Combined with the log concavity of the nonvanishing parts of (2.7) and (2.8), we see that we need to consider the relative error of the approximation at integers closest to $(n-1)/2$. We should therefore look at cases where we change from the approximation (2.4) to the approximation (2.5). This is when the variance of the exponentially tilted distribution is 0.4. Thus, we have not gone all the way to the limit $\alpha = 1$, and these cases have to be investigated by numerical calculations. It is found that the relative error increases with the value of β, and is less than 0.10 for (2.5) and less than 0.09 for (2.4) when the variance is 0.4. To understand these numbers, it may be of interest to compare with a distribution with probabilities proportional to $\exp(-\tau k^2 - s|k|)$, $k \in \mathbf{N}$, and where we want to approximate the probability at $k = 0$. For $\tau \approx 0.73$ and s chosen so that the variance is 0.4 the relative errors of the saddlepoint approximation (2.3) given through (2.4) and (2.5) are comparable to the numbers quoted above. The extreme case of a discrete Laplace distribution ($\tau = 0$) gives relative errors 0.43 and 0.24 for the approximations (2.4) and (2.5); see in this connection Figure 1.

In summary, the saddlepoint approximation of this paper has relative error of order $o(1/n)$ in a large deviation region for fixed values of (α, β), and has bounded relative error all over the parameter space with a maximum of 0.18 for the limiting case of a uniform distribution.

3 A comparison with Xia and Zhang (2009)

In this section we compare the saddlepoint approximation of this paper with the approximation of Xia and Zhang (2009). For the comparison we calculate exact probabilities using the recursion in (1.1). As mentioned in the introduction, the important aspect of the latter paper is that an explicit upper bound for the total variation distance of the approximation is given. Unfortunately, as we will demonstrate below, for very many cases the upper bound is actually above one and therefore gives no information on the quality of the approximation.

In Figure 2 it is shown for what parameter values the upper bound is less than one. The white region is where the bound is below one. The left subfigure is for $n = 1000$ and the right subfigure is for $n = 100{,}000$. Xia and Zhang (2009) say that the upper bound is "useful when both α and β are a reasonable distance from 0 and 1." However, as can be seen from Figure 2, even when $n = 100{,}000$ quite a large part of the parameter space is excluded.

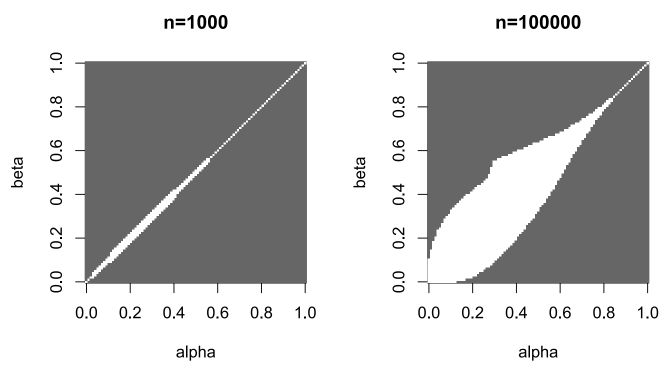

Figure 2 *The figure shows the upper bound on the total variation distance from Xia and Zhang (2009) truncated at one. Thus, the white region is where the upper bound is below one and the black region is where the bound is above one. The left figure is for $n = 1000$ and the right figure is for $n = 100{,}000$.*

Table 1 *Upper bound on the total variation distance from Xia and Zhang (2009). Four limiting situations for S_n from Dobrushin (1961) are considered*

		Upper bound		
α	β	$n = 100$	$n = 10{,}000$	$n = 1{,}000{,}000$
$4/\sqrt{n}$	$1/2$	43	55	10
$1/2$	$1 - 4/\sqrt{n}$	73	$3 \cdot 10^3$	$6 \cdot 10^4$
$1/2$	$1 - 40/n$	73	$3 \cdot 10^6$	$3 \cdot 10^{10}$
$1 - 1/n$	$2/n$	<0	<0	<0

Table 2 *Total variation distance (Total) and upper bound for the approximation in Xia and Zhang (2009). The parameter values are $\alpha = 0.4$ and $\beta = 0.5$*

	$n = 10$	$n = 1000$	$n = 10{,}000$
Total	0.015	0.001	0.0005
Upper bound	360	6.4	1.3

We next consider situations where $\alpha = \alpha_n$ and $\beta = \beta_n$ depend on n and let $n \to \infty$. We consider situations where S_n, properly normalized, has a limiting distribution. In Table 1 we have considered three of the cases in Dobrushin (1961). The table gives the upper bound of Xia and Zhang (2009) for the total variation distance for $n = 100, 10{,}000, 1{,}000{,}000$. In the two first rows S_n has a limiting normal distribution with $\mathrm{Var}(S_n)/E(S_n)$ larger than one and less than one, respectively. Because the limit is a normal distribution we expect the total variation distance to tend to zero. This is, however, not reflected well in the upper bound. The third row of Table 1 is for one of the remaining cases with a nonnormal limit, and the results shown are typical for these cases, that is, the upper bound tends to infinity. The fourth row has been included to show a flaw in the formulation of Theorem 1.1 of Xia and Zhang (2009). That theorem contains a success probability θ of a binomial distribution that becomes greater than one for certain values of (α, β, n).

As a final investigation of the upper bound of Xia and Zhang (2009) we compare the actual total variation distance with the upper bound as $n \to \infty$. We take $\alpha = 0.4$ and $\beta = 0.5$, staying well away from zero and one. The results can be seen in Table 2. Even for $n = 10{,}000$, where the total variation distance is very small, the upper bound is still above one.

Having considered the upper bound of Xia and Zhang (2009), we next turn to a direct comparison of the quality of the approximation of that paper and the saddlepoint approximation (2.3). Since we use the exact values of $P(S_n = k)$ for $k = 0, 1, n - 1, n$ [see (A.1)], we exclude these points when calculating the performance of the approximation of Xia and Zhang (2009). We consider both the

Table 3 *The total variation distance for the approximation of Xia and Zhang (2009) (T_{XZ}) and for the saddlepoint approximation (2.3) (T_{Sp}), together with the maximal relative error of the approximation to the point probabilities $P(S_n = k)$, $k = 2, \ldots, n-2$, (R_{XZ}: approximation of Xia and Zhang (2009); R_{Sp}: saddlepoint approximation)*

α	β	n	T_{XZ}	T_{Sp}	R_{XZ}	R_{Sp}
0.01	0.10	10	0.0003	0.0001	$9 \cdot 10^0$	0.0508
0.01	0.10	300	0.0010	0.0017	$5 \cdot 10^{81}$	0.0072
0.01	0.50	10	0.0036	0.0015	$6 \cdot 10^{-1}$	0.1294
0.01	0.50	300	0.0192	0.0020	$6 \cdot 10^{37}$	0.0075
0.01	0.99	10	0.2461	0.0031	$1 \cdot 10^1$	0.1764
0.01	0.99	300	0.1368	0.0104	$2 \cdot 10^2$	0.0441
0.50	0.60	10	0.0222	0.0023	$2 \cdot 10^{-1}$	0.0108
0.50	0.60	300	0.0033	0.0000	$2 \cdot 10^{17}$	0.0089
0.50	0.99	10	0.0170	0.0015	$2 \cdot 10^8$	0.1294
0.50	0.99	300	0.0770	0.0020	$>1 \cdot 10^{100}$	0.0075
0.90	0.99	10	0.0034	0.0001	$5 \cdot 10^5$	0.0508
0.90	0.99	300	0.0313	0.0017	$>1 \cdot 10^{100}$	0.0072

total variation distance and the maximal relative error of the approximation for the point probabilities. The relative error of an approximation a to a number x is in our comparisons computed as $\max\{a, x\}/\min\{a, x\} - 1$. Table 3 shows the typical behaviour using $n = 10$ and $n = 300$. The saddlepoint approximation of course has better properties in terms of relative error, but also, generally, performs better on the total deviation scale. Even for the case $\alpha = 0.5$ and $\beta = 0.6$, which is close to the binomial case $\alpha = \beta$, the saddlepoint approximation performs best.

Appendix: Technical details of the approximation

In this appendix we give details of the saddlepoint approximation (2.3). We let $p_0 = P(X_1 = 0)$ and $p_1 = P(X_1 = 1)$ and use the following exact values:

$$\begin{aligned} P(S_n = 0) &= p_0(1-\alpha)^{n-1}, \qquad P(S_n = n) = p_1\beta^{n-1}, \\ P(S_n = 1) &= (1-\alpha)^{n-3}\big[p_0\alpha\big(1-\alpha+(n-2)(1-\beta)\big) \\ &\qquad + p_1(1-\alpha)(1-\beta)\big], \\ P(S_n = n-1) &= \beta^{n-3}\big[p_1(1-\beta)\big(\beta+(n-2)\alpha\big) + p_0\beta\alpha\big]. \end{aligned} \tag{A.1}$$

The saddlepoint approximation is then used for $P(S_n = k)/P(0 < S_n < n)$ for $k = 2, \ldots, n-2$.

The moment generating function $\psi(z)$ from (2.2) can be evaluated through an eigenvalue decomposition of $P_0(z)$, and from this the first four derivatives can be found as well. Alternatively, we define

$$\psi(z,a,n;j)=\sum_{k=j}^{n}k_{(j)}z^{k-j}P(S_n=k,X_n=a),\qquad a=0,1,\ j=0,1,2,3,4,$$

where $k_{(j)}=k(k-1)\cdots(k-j+1)$ with $k_{(0)}=1$. Then the recursion (1.1) gives that the vector of these terms (with the index a running faster than j) is calculated as

$$(p_0,zp_1,0,p_1,0,0,0,0,0,0)M_5(z)^{n-1},$$

where $M_5(z)$ is the 10×10 matrix

$$M_5(z)=\begin{pmatrix} P_0(z) & B & 0 & 0 & 0\\ 0 & P_0(z) & 2B & 0 & 0\\ 0 & 0 & P_0(z) & 3B & 0\\ 0 & 0 & 0 & P_0(z) & 4B\\ 0 & 0 & 0 & 0 & P_0(z)\end{pmatrix},\qquad B=\begin{pmatrix}0 & \alpha\\ 0 & \beta\end{pmatrix}.$$

Let $\kappa(s)=\log(\phi(e^s))$ be the cumulant transform of the conditional distribution with $\phi(z)$ defined in (2.1). For a given value of k let $s(k)$ be the saddlepoint, that is, $\kappa'(s(k))=k$, and let $\sigma(k)^2=\kappa''(s(k))$ be the variance of the exponentially tilted distribution. The cumulants are $\kappa_3(k)=\kappa^{(3)}(s(k))$ and $\kappa_4(k)=\kappa^{(4)}(s(k))$, and the standardized cumulants are $\gamma_3(k)=\kappa_3(k)/\sigma(k)^3$ and $\gamma_4(k)=\kappa_4(k)/\sigma(k)^4$. We thus have all the quantities entering the approximation in (2.4) and (2.5).

For the numerical calculations of this paper we have found $M_5(z)^n$ by calculating $M_5^2, M_5^4,\dots,M_5^{2^m}$, where m is the largest integer with $2^m\le n$.

References

Čekanavičius, V. and P. Vellaisamy (2010). Compound Poisson and signed compound Poisson approximations to the Markov binomial law. *Bernoulli* **16**, 1114–1136. MR2759171

Dobrushin, R. L. (1961). Limit theorems for a Markov chain of two states. *Select. Transl. Math. Statist. Probab.* **1**, 97–134. MR0116371

Gabriel, K. R. (1959). The distribution of the number of successes in a sequence of dependent trials. *Biometrika* **46**, 454–460. MR0109363

Jensen, J. L. (1995). *Saddlepoint Approximations*. Oxford: Clarendon Press. MR1354837

Ladd, D. W. (1975). An algorithm for the binomial distribution with dependent trials. *J. Amer. Statist. Assoc.* **70**, 333–340.

Xia, A. and M. Zhang (2009). On approximation of Markov binomial distributions. *Bernoulli* **15**, 1335–1350. MR2597595

Department of Mathematical Sciences
University of Aarhus
Ny Munkegade
DK-8000 Aarhus C
Denmark
E-mail: jlj@imf.au.dk