# The Skew-Normal Approximation of the Binomial Distribution

Joyce Tipping

2010

The Skew-Normal Approximation of the Binomial Distribution Joyce Tipping 2010

## 1 Introduction

One of the most basic distributions in probability and statistics is the binomial, $X \sim Bin(n, p)$, $n = 1, 2, \ldots$, $p \in (0, 1)$ with probability density function (pdf)

$$f_X(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n,$$

where $q = 1 - p$. Calculating the binomial cumulative distribution function (cdf), $F_X(x) = P(X \leq x) = \sum_{k=0}^{x} f_X(k)$, by hand is manageable for small $n$ but quickly becomes cumbersome as $n$ grows even moderately large. A common strategy is to use the normal distribution[1] as an approximation:

$$F_X(x) \approx \Phi\left(\frac{x + 0.5 - \mu}{\sigma}\right), \tag{1}$$

where $\Phi$ is the standard normal cdf and $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

This approximation works best when the binomial is perfectly symmetrical, with $p = 0.5$. However as $p$ travels away from 0.5 in either direction, the binomial becomes increasingly skewed, and one must either provide larger values of $n$ to compensate or face growing inaccuracy.

[1] A random variable $X$ follows the normal distribution with mean $\mu$ and variance $\sigma^2$ (denoted $X \sim N(\mu, \sigma^2)$) if it has the pdf $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-[(x-\mu)/\sigma]^2/2}$ for $-\infty < x < \infty$, where $-\infty < \mu < \infty$ and $0 < \sigma < \infty$.

$N(0, 1)$ is an important special case known as the standard normal (denoted $Z$) and has pdf $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$ and cdf given by $\Phi(z) = \int_{-\infty}^{z} \phi(t)\, dt$. Equations 3.3.27, 3.3.29, and 3.3.30, Bain and Engelhardt (1992).

To protect against these cases, many textbooks impose a condition on the normal approximation, often stated as either (1) $np(1-p) > 9$ or (2) $np > 5$ for $0 < p \leq 0.5$, and $n(1-p) > 5$ for $0.5 < p < 1$. However, Schader and Schmid (1989) show that even when these conditions are met, the inaccuracy of the normal approximation can be surprisingly large. So, the natural question arises: Can we do better?

Perhaps so. Intuitively speaking, the main problem with the normal approximation is that the normal is always symmetric while the binomial is skewed for all values of $p$ not equal to 0.5. This fact suggests using the skew-normal distribution, which adds an extra parameter to capture an asymetrical lean.

In this paper, we will explore the aptness of the skew-normal as a method of approximating the binomial. First, we'll acquaint ourselves with its basic properties (section 2). We'll then derive an approximation via the method of moments (section 3) and compare its accuracy to that of the normal approximation (section 4). Finally, we will leave the reader with a few practical resources for applying our new method (section 5).

# 2 The Skew-Normal Distribution

The skew-normal distribution is similar to the normal but with an added parameter for skew that allows it to lean to the left or right. In this section, we'll get to know some of its basic properties.

## 2.1 Foundations

**Definition 1.** Let $Y$ be a skew-normal distribution, with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$, and shape parameter $\lambda \in \mathbb{R}$; we will denote it $SN(\mu, \sigma, \lambda)$.[2]Then $Y$ has pdf

$$f_Y(x|\mu, \sigma, \lambda) = \frac{2}{\sigma} \cdot \phi\left(\frac{x-\mu}{\sigma}\right) \cdot \Phi\left(\frac{\lambda(x-\mu)}{\sigma}\right), \quad x \in \mathbb{R}, \tag{2}$$

where $\phi$ is the standard normal pdf and $\Phi$ is the standard normal cdf. Chang et al. (2008)

The skew-normal was first introduced by O'Hagan and Leonard (1976) and was most notably developed by Azzalini (1985). It belongs to a class of skewed distributions conforming to the following guidelines:

**Lemma 1.** *If $f_0$ is a one-dimensional probability density function symmetric about 0, and $G$ is a one-dimensional distribution function such that $G'$ exists and is a density symmetric about 0, then*

[2]In this paper, we have followed the notation established by Chang et al. (2008) in naming our parameters $\mu$, $\sigma$, and $\lambda$. It does seem intuitive, after all, for the skew-normal to simply "extend" the normal with an extra parameter.

An unfortunate consequence of this choice, however, is the misconception that $\mu$ and $\sigma$ are related to the mean and variance the way they are in the normal distribution. In the case of the skew-normal, it is more productive to think of them as abstract location and scale parameters. In fact, a quick comparison of figures 2, 3, and 4 and the corresponding parameters in Table 1 shows that $\mu$ and $\sigma$ are not visually related to the curve in any intuitive way.

$$f(z) = 2 \cdot f_0(z) \cdot G\{w(z)\} \quad (-\infty < z < \infty) \tag{3}$$

*is a density function for any odd function $w(\cdot)$. (Lemma 1, Azzalini, 2005)*

In equation (3), $f_0$ acts as a weighting function referred to as a kernel. When $f_0 = \phi$, the distribution is said to be generated by the normal kernel. $G$ does not necessarily have to be taken from the same distribution as $f_0$; for example, the normal kernel has been paired with cdf's from various other distributions, such as Student's t, Cauchy, Laplace, and uniform. In the case of our skew-normal, however, both $f_0$ and $G$ come from the normal distribution.

The skew-normal's first three moments about the origin (raw moments), given by Pewsey (2000), are

$$\begin{aligned} E(Y) &= \mu + b\delta\sigma, \\ E(Y^2) &= \mu^2 + 2b\delta\mu\sigma + \sigma^2, \\ E(Y^3) &= \mu^3 + 3b\delta\mu^2\sigma + 3\mu\sigma^2 + 3b\delta\sigma^3 - b\delta^3\sigma^3, \end{aligned} \tag{4}$$

where $b = \sqrt{\frac{2}{\pi}}$ and $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. Using these raw moments, we can derive $Var(Y)$:

$$\begin{aligned} Var(Y) &= E(Y^2) - [E(Y)]^2 \\ &= \mu^2 + 2b\delta\mu\sigma + \sigma^2 - (\mu + b\delta\sigma)^2 \\ &= \cancel{\mu^2} + \cancel{2b\delta\mu\sigma} + \sigma^2 - \cancel{\mu^2} - \cancel{2b\delta\mu\sigma} - b^2\delta^2\sigma^2 \\ &= \sigma^2 - b^2\delta^2\sigma^2 \\ &= \sigma^2(1 - b^2\delta^2). \end{aligned} \tag{5}$$

The $SN(0, 1, \lambda)$ distribution is called the standard skew-normal; its pdf is

$$f_Z(x|\lambda) = 2 \cdot \phi(x) \cdot \Phi(\lambda x), \quad x \in \mathbb{R}. \tag{6}$$

Similar to the normal and standard normal, $Z = \frac{Y-\mu}{\sigma}$ and $Y = \sigma Z + \mu$.

A natural question to ask is how the skew-normal relates to the normal. Fortunately, the connection is very intuitive: When $\lambda = 0$, Equation (2) becomes

$$
\begin{aligned}
f(x|\mu,\sigma,\lambda=0) &= \frac{2}{\sigma}\cdot\phi\left(\frac{x-\mu}{\sigma}\right)\cdot\Phi(0) \\
&= \frac{2}{\sigma}\cdot\phi\left(\frac{x-\mu}{\sigma}\right)\cdot 0.5 \\
&= \frac{1}{\sigma}\cdot\phi\left(\frac{x-\mu}{\sigma}\right) \\
&= \frac{1}{\sigma}\cdot\frac{1}{\sqrt{2\pi}}\cdot\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),
\end{aligned}
$$

which is the pdf of the normal distribution ($\mu$, $\sigma$). Furthermore, when $\lambda > 0$, the curve skews to the right, and when $\lambda < 0$, it skews to the left (a property we will prove in section 2.2).

## 2.2 Properties of the Standard Skew-Normal Distribution

The following four properties of the standard skew-normal, given by Chang et al. (2008), help shed light on our enigmatic new distribution:

**Property 1.** *If $Z \sim SN(0,1,\lambda)$, then $(-Z) \sim SN(0,1,-\lambda)$.*

*Proof.* The standard normal pdf is an even function: $\phi(-x) = \frac{1}{\sqrt{2\pi}} e^{-(-x)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \phi(x)$. But the standard normal cdf, $\Phi(x) = \int_{-\infty}^{x}\phi(t)\,dt$, is not even, approaching 0 as $x \to -\infty$ and approaching 1 as $x \to \infty$. Thus,

$$
\begin{aligned}
f_{(-Z)}(x) &= f_Z(-x) \\
&= 2\cdot\phi(-x)\cdot\Phi(-\lambda x) \\
&= 2\cdot\phi(x)\cdot\Phi(-\lambda x),
\end{aligned}
$$

which is the pdf of $SN(0,1,-\lambda)$. $\mathcal{Q.E.D.}$

**Property 2.** *If $Z \sim SN(0,1,\lambda)$, then $Z^2 \sim \chi^2_1$ (chi-square with 1 degree of freedom).*

*Proof.* Lemma 1 from section 2.1 provides a very useful result for the above defined $f_0$ and $f$, which we restate here as a corollary:

**Corollary 1** (Perturbation Invariance)**.** *If $Y \sim f_0$ and $Z \sim f$, then $|Y| \stackrel{d}{=} |Z|$, where the notation $\stackrel{d}{=}$ denotes equality in distribution. (Page 161, Azzalini, 2005)*

Let $X \sim N(0,1)$. Take $f_0 = \phi$, and $G = \Phi$. Then, $f_Z(z) = 2 \cdot \phi(z) \cdot \Phi(\lambda z)$ conforms to Equation (3), and we can conclude that $X$ and $Z$ are equal in distribution.

We will now show that $X^2 \sim \chi^2_1$ by deriving its moment generating function (mgf):[3]

$$
\begin{aligned}
M_{X^2}(t) &= E[e^{tX^2}] \\
&= \int_{-\infty}^{\infty} e^{tx^2} \left[ \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx^2 - x^2/2} \, dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}(1-2t)} \, dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{1-2t}\, x)^2} \, dx \, ;
\end{aligned}
$$

let $u = (\sqrt{1-2t})\, x$; then $du = (\sqrt{1-2t})\, dx$, $dx = \frac{du}{\sqrt{1-2t}}$, and our limits become $x \to -\infty, x \to \infty \;\Rightarrow\; u \to -\infty, u \to \infty$:

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \left( \frac{1}{\sqrt{1-2t}} du \right) \\
&= \frac{1}{\sqrt{1-2t}} \underbrace{\left( \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du \right)}_{\phi(u) \text{ integrated over } (-\infty,\infty) = 1} \\
&= \frac{1}{\sqrt{1-2t}} \, ,
\end{aligned}
$$

which is the mgf of the $\chi^2_1$. Since $Z$ is equal in distribution to $X$, we can also conclude that $Z^2 \sim \chi^2_1$. $\mathcal{Q.E.D.}$

**Property 3.** *As $\lambda \to \pm\infty$, $SN(0,1,\lambda)$ tends to the half normal distribution, $\pm|N(0,1)|$.*

To prove our theorem, it is helpful to formally define the half normal distribution:

**Lemma 3.1.** *Let $X \sim N(0,\sigma^2)$. Then the distribution of $|X|$ is a half-normal random variable with parameter $\sigma$ and*

$$
f_{|X|}(x) = \begin{cases} 0 & \text{when } -\infty < x \leq 0 \\ 2 \cdot f_X(x) & \text{when } 0 < x < \infty \end{cases} .
$$

[3] If $X$ is a random variable, then the expected value $M_X(t) = E(e^{tX})$ is called the moment generating function (mgf) of X if this expected value exists for all values of $t$ in some interval of the form $-h < t < h$ for some $h > 0$. Definition 2.5.1, Bain and Engelhardt (1992).

*Proof.* Let $X \sim N(0, \sigma^2)$, defined over $A = (-\infty, \infty)$. Define

$$Y = |X| = \begin{cases} -x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ x & \text{if } x > 0 \end{cases}.$$

$Y$ is not one-to-one over $A$. However, we can partition $A$ into disjoint subsets $A_1 = (-\infty, 0)$, $A_2 = (0, \infty)$, and $A_3 = \{0\}$ such that $A = A_1 \cup A_2 \cup A_3$ and $Y$ is one-to-one over each $A_i$. We can then transform each piece separately using Theorem 6.3.2 from Bain and Engelhardt (1992):[4]

On $A_1$: $y = -x \;\Rightarrow\; x = -y$, and the absolute value of the Jacobian $|\mathbb{J}| = \left|\frac{dx}{dy}\right| = |-1| = 1$, yielding

$$\begin{aligned} f_Y(y) &= f_X(x) \cdot |\mathbb{J}| \\ &= f_X(-y) \cdot 1 \\ &= \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{(-y)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{y^2}{2\sigma^2}} \\ &= f_X(y) \end{aligned}$$

over the domain $A_1 : -\infty < x < 0 \;\Rightarrow\; -\infty < -y < 0 \;\Rightarrow\; 0 < y < \infty : B_1$.

Similarly, on $A_2$: $y = x \;\Rightarrow\; x = y$, and the absolute value of the Jacobian $|\mathbb{J}| = \left|\frac{dx}{dy}\right| = |1| = 1$, yielding

$$\begin{aligned} f_Y(y) &= f_X(x) \cdot |\mathbb{J}| \\ &= f_X(y) \cdot 1 \\ &= f_X(y) \end{aligned}$$

over the domain $A_2 : 0 < x < \infty \;\Rightarrow\; 0 < y < \infty : B_2$.

[4]**Continuous transformations that are one-to-one**: Suppose that $X$ is a continuous random variable with pdf $f_X(x)$, and assume that $Y = u(X)$ defines a one-to-one transformation from $A = \{x | f_X(x) > 0\}$ on to $B = \{y | f_Y(y) > 0\}$ with inverse transformation $x = w(y)$. If the derivative $(d/dy)w(y)$ is continuous and nonzero on $B$, then the pdf of $Y$ is

$$f_Y(y) = f_X(w(y)) \left| \frac{d}{dy} w(y) \right| \qquad y \in B. \tag{7}$$

Theorem 6.3.2, Bain and Engelhardt (1992).

On $A_3$, we have $x = 0 \Rightarrow y = 0$, and the absolute value of the Jacobian $|\mathbb{J}| = \left|\frac{dx}{dy}\right| = |0| = 0$, yielding $f_Y(y) = f_X(x) \cdot |\mathbb{J}| = f_X(x) \cdot 0 = 0$.

Then, by Equation 6.3.10 from Bain and Engelhardt (1992),[5]

$$
\begin{aligned}
f_Y(y) &= \{f_Y(y) \text{ over } A_1\} + \{f_Y(y) \text{ over } A_2\} \\
&= f_X(y) + f_X(y) \\
&= 2 \cdot f_X(y)
\end{aligned}
$$

over $B = B_1 \cup B_2 = (0, \infty)$, and 0 otherwise. $\mathcal{Q.E.D.}$

With Lemma 3.1, we can easily show our property:

*Proof of Property 3.* Let $Z \sim SN(0,1,\lambda)$. Recall that $f_Z(x) = 2 \cdot \phi(x) \cdot \Phi(\lambda x)$.

Consider $\lim_{\lambda\to\infty} f_X(x)$. When $x$ is negative, $\lambda x \to -\infty$ and thus $\Phi(\lambda x) \to 0$. When $x$ is positive, however, $\lambda x \to \infty$ and $\Phi(\lambda x) \to 1$. Thus,

$$
\lim_{\lambda\to\infty} 2 \cdot \phi(x) \cdot \Phi(\lambda x) \;=\; \begin{cases} 0 & \text{when } x \leq 0 \\ 2 \cdot \phi(x) & \text{when } x > 0 \end{cases} \;=\; |N(0,1)|. \tag{8}
$$

In $\lim_{\lambda\to-\infty} f_X(x)$, the signs are reversed. When $x$ is negative, $\lambda x \to \infty$ and $\Phi(\lambda x) \to 1$. When $x$ is positive, $\lambda x \to -\infty$ and $\Phi(\lambda x) \to 0$. Thus,

$$
\lim_{\lambda\to-\infty} 2 \cdot \phi(x) \cdot \Phi(\lambda x) \;=\; \begin{cases} 2 \cdot \phi(x) & \text{when } x < 0 \\ 0 & \text{when } x \geq 0 \end{cases} \;=\; -|N(0,1)|. \tag{9}
$$

$\mathcal{Q.E.D.}$

**Property 4.** *The moment generating function of* $SN(0,1,\lambda)$ *is*

$$
M(t|\lambda) = 2 \cdot \Phi(\delta t) \cdot e^{t^2/2}, \tag{10}
$$

*where* $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ *and* $t \in (-\infty, \infty)$.

[5]**Continuous transformations that are not one-to-one**: When $u(x)$ is not one-to-one over $A$, we can replace equation (7) in footnote 4 with

$$
f_Y(y) = \sum_j f_X(w_j(y)) \left| \frac{d}{dy} w_j(y) \right|.
$$

Equation 6.3.10, Bain and Engelhardt (1992).

*Proof.* According to Equation 5 in Azzalini (2005), the mgf of $SN(\mu, \sigma, \lambda)$ is

$$M(t) \;=\; E\{e^{tY}\} \;=\; 2 \cdot \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \cdot \Phi(\delta \sigma t),$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in (-1, 1)$. It follows that the mgf of the $SN(0, 1, \lambda)$ is

$$2 \cdot \exp\left(0 \cdot t + \frac{1 \cdot t^2}{2}\right) \cdot \Phi(\delta \cdot 1 \cdot t) \;=\; 2 \cdot e^{t^2/2} \cdot \Phi(\delta t).$$

$\mathcal{Q.E.D.}$

# 3 Developing an Approximation

Now that we have gotten to know our new distribution a little better, we can use it to develop an approximation for the binomial.

Let $B \sim Bin(n, p)$ and $Y \sim SN(\mu, \sigma, \lambda)$. We will find estimates for $\mu$, $\sigma$, and $\lambda$ using the method of moments, that is, by comparing the first, second, and third moments about the mean (central moments) of $B$ and $Y$.

## 3.1 The Central Moments of the Binomial Distribution

The first two central moments of the binomial are simply the mean and variance:

$$E(B) = np, \quad Var(B) = np(1-p).$$

From these facts, we can find $E(B^2)$: Recall that $Var(B) = E[(B - E(B))^2] = E(B^2) - [E(B)]^2$ and thus

$$E(B^2) = Var(B) + [E(B)]^2 = np(1-p) + n^2p^2 = np - np^2 + n^2p^2.$$

We will also need $E(B^3)$, which we will get via the third factorial moment:

$$E[B(B-1)(B-2)] = \sum_{x=0}^{n} x(x-1)(x-2) \cdot \left\{ \binom{n}{x} p^x q^{n-x} \right\}.$$

Notice that the first three terms of the sum on the right hand side are zero, so we can rewrite it beginning at $x = 3$:

$$\begin{aligned}
E[B(B-1)(B-2)] &= \sum_{x=3}^{n} x(x-1)(x-2) \cdot \left\{ \binom{n}{x} p^x q^{n-x} \right\}. \\
&= \sum_{x=3}^{n} x(x-1)(x-2) \cdot \frac{n!}{x!\,(n-x)!}\; p^x q^{n-x} \\
&= \sum_{x=3}^{n} \frac{n!}{(x-3)!\,(n-x)!}\; p^x q^{n-x} \\
&= \sum_{x=3}^{n} n(n-1)(n-2)p^3 \cdot \frac{(n-3)!}{(x-3)!\,(n-x)!}\; p^{x-3} q^{n-x}\,;
\end{aligned}$$

let $y = x - 3$; then $x = y + 3$, and $x = 3, x = n \;\Rightarrow\; y = 0, y = n - 3$:

$$\begin{aligned}
&= n(n-1)(n-2)p^3 \cdot \sum_{y=0}^{n-3} \frac{(n-3)!}{y!\,(n-(y+3))!}\; p^y q^{n-(y+3)} \\
&= n(n-1)(n-2)p^3 \cdot \underbrace{\sum_{y=0}^{n-3} \frac{(n-3)!}{y!\,((n-3)-y)!}\; p^y q^{(n-3)-y}}_{\text{[pdf of } Bin(n-3,p) \text{ summed over its domain] = 1}} \\
&= n(n-1)(n-2)p^3 \\
&= n^3p^3 - 3n^2p^3 + 2np^3. \qquad (11)
\end{aligned}$$

Here we take a moment to expand the left side of the previous equation:

$$\begin{aligned}
E[B(B-1)(B-2)] &= E\left[B^3 - 3B^2 + 2B\right] \\
&= E(B^3) - 3E(B^2) + 2E(B) \\
&= E(B^3) - 3(np - np^2 + n^2p^2) + 2np \\
&= E(B^3) - 3np + 3np^2 - 3n^2p^2 + 2np \\
&= E(B^3) + 3np^2 - 3n^2p^2 - np. \qquad (12)
\end{aligned}$$

Since (11) and (12) are equal to each other, we can solve for $E(B^3)$:

$$\begin{aligned}
E(B^3) + 3np^2 - 3n^2p^2 - np &= n^3p^3 - 3n^2p^3 + 2np^3 \\
\Rightarrow \qquad E(B^3) &= n^3p^3 - 3n^2p^3 + 2np^3 - 3np^2 + 3n^2p^2 + np. \qquad (13)
\end{aligned}$$

Now, finally, we have all the information necessary to obtain the third central moment:

$$
\begin{aligned}
E\left([B-E(B)]^3\right) &= E\left(B^3 - 3B^2E(B) + 3B[E(B)]^2 - [E(B)]^3\right) \\
&= E(B^3) - 3E(B^2)E(B) + 3E(B)[E(B)]^2 - [E(B)]^3 \\
&= E(B^3) - 3E(B^2)E(B) + 2[E(B)]^3 \\
&= (n^3p^3 - 3n^2p^3 + 2np^3 - 3np^2 + 3n^2p^2 + np) - 3(np - np^2 + n^2p^2)(np) + 2(np)^3 \\
&= \cancel{n^3p^3} - \cancel{3n^2p^3} + 2np^3 - 3np^2 + \cancel{3n^2p^2} + np - \cancel{3n^2p^2} + \cancel{3n^2p^3} - \cancel{3n^3p^3} + \cancel{2n^3p^3} \\
&= 2np^3 - 3np^2 + np \\
&= np(p-1)(2p-1).
\end{aligned}
$$

Our hard-earned results, restated for convenience:

$$
\begin{aligned}
E(B) &= np, \\
E([B-E(B)]^2) &= np(1-p), \\
E([B-E(B)]^3) &= np(p-1)(2p-1).
\end{aligned} \tag{14}
$$

## 3.2 The Central Moments of the Skew Normal Distribution

Now we'll take a look at the skew normal. Equations (4) and (5) takes care of the mean and variance; again the third central moment is a little more complicated:

$$
\begin{aligned}
E([Y-E(Y)]^3) &= E(Y^3) - 3E(Y^2)E(Y) + 2[E(Y)]^3 \\
&= (\mu^3 + 3b\delta\mu^2\sigma + 3\mu\sigma^2 + 3b\delta\sigma^3 - b\delta^3\sigma^3) - 3(\mu^2 + 2b\delta\mu\sigma + \sigma^2)(\mu + b\delta\sigma) \\
&\quad + 2\,(\mu + b\delta\sigma)^3 \\
&= \cancel{\mu^3} + \cancel{3b\delta\mu^2\sigma} + \cancel{3\mu\sigma^2} + \cancel{3b\delta\sigma^3} - b\delta^3\sigma^3 - \cancel{3\mu^3} - \cancel{3b\delta\mu^2\sigma} - \cancel{6b\delta\mu^2\sigma} - \cancel{6b^2\delta^2\mu\sigma^2} - \cancel{3\mu\sigma^2} \\
&\quad - \cancel{3b\delta\sigma^3} + \cancel{2\mu^3} + \cancel{6b\delta\mu^2\sigma} + \cancel{6b^2\delta^2\mu\sigma^2} + 2b^3\delta^3\sigma^3 \\
&= 2b^3\delta^3\sigma^3 - b\delta^3\sigma^3 \\
&= b\delta^3\sigma^3(2b^2 - 1).
\end{aligned}
$$

We restate our results:

$$
\begin{aligned}
E(Y) = \mu + b\delta\sigma \qquad &= \mu + \sigma \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{\lambda}{\sqrt{1+\lambda^2}}\,, \\
E([Y - E(Y)]^2) = \sigma^2(1 - b^2\delta^2) \quad &= \sigma^2 \left(1 - \frac{2}{\pi} \cdot \frac{\lambda^2}{1+\lambda^2}\right), \\
E([Y - E(Y)]^3) = b\delta^3\sigma^3(2b^2 - 1) &= \sigma^3 \sqrt{\frac{2}{\pi}} \left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)^3 \left(\frac{4}{\pi} - 1\right).
\end{aligned}
\tag{15}
$$

### 3.3 Solving for $\mu$, $\sigma$, $\lambda$

To derive our approximation, we set the above moments of our two distributions equal to each other and, taking $n$ and $p$ as constants, solve for $\mu$, $\sigma$ and $\lambda$:

$$
np = \mu + \sigma \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{\lambda}{\sqrt{1+\lambda^2}} \tag{16a}
$$

$$
np(1-p) = \sigma^2 \left(1 - \frac{2}{\pi} \cdot \frac{\lambda^2}{1+\lambda^2}\right) \tag{16b}
$$

$$
np(p-1)(2p-1) = \sigma^3 \sqrt{\frac{2}{\pi}} \left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)^3 \left(\frac{4}{\pi} - 1\right) \tag{16c}
$$

To get $\lambda$, we divide the cube of (16b) by the square of (16c):

$$
\begin{aligned}
&\frac{\sigma^6 \left(1 - \frac{2}{\pi} \cdot \frac{\lambda^2}{1+\lambda^2}\right)^3}{\sigma^6 \cdot \frac{2}{\pi} \left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)^6 \left(\frac{4}{\pi} - 1\right)^2} = \frac{n^3p^3(1-p)^3}{n^2p^2(p-1)^2(2p-1)^2} \\
\Rightarrow \quad &\frac{\left(1 - \frac{2}{\pi} \cdot \frac{\lambda^2}{1+\lambda^2}\right)^3}{\frac{2}{\pi} \left(\frac{\lambda^2}{1+\lambda^2}\right)^3 \left(\frac{4}{\pi} - 1\right)^2} = \frac{np(1-p)}{(1-2p)^2}\,.
\end{aligned}
\tag{17}
$$

The above equation (17) is a rational expression in $\lambda^2$ that can be solved with either a considerable amount of manual labor or, more efficiently, with a computer algebra system. Once we have $\lambda^2$, then $\lambda$ is simply the positive or negative square root, as determined by the sign of $(1 - 2p)$.

The sign can be explained with a little assistance from Property 3: When $p \to 0$, the binomial shifts its weight towards the smaller values on the left and converges toward the positive half normal, which by (8) corresponds to a positive $\lambda$. When $p \to 1$, the binomial shifts its weight towards the the larger values on the right and converges toward the negative half normal, which by (9) corresponds to a negative $\lambda$. When $p = 0.5$, the binomial is symmetric and $\lambda$ is 0, eliminating the need for a sign. Thus:

$$\lambda = \{\text{sign of } (1 - 2p)\}\sqrt{\lambda^2}. \tag{18}$$

Having secured $\lambda$, we can find $\sigma$ using (16b):

$$np(1-p) = \sigma^2 \left(1 - \frac{2}{\pi} \cdot \frac{\lambda^2}{1+\lambda^2}\right) \quad \Rightarrow \quad \sigma = \sqrt{\frac{np(1-p)}{1 - \frac{2}{\pi} \cdot \frac{\lambda^2}{1+\lambda^2}}}\,. \tag{19}$$

And with both $\lambda$ and $\sigma$, a simple rearrangement of (16a) yields $\mu$:

$$np = \mu + \sigma \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{\lambda}{\sqrt{1+\lambda^2}} \quad \Rightarrow \quad \mu = np - \sigma \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{\lambda}{\sqrt{1+\lambda^2}}\,. \tag{20}$$

One notable case where the above skew-normal approximation fails is when $p = 0.5$; the right side of (17) becomes undefined, and we are unable to obtain $\lambda$. Fortunately, here we can fall back on our intuition, which tells us that since the binomial is perfectly symmetrical, the skew should be 0. A few observations support this conclusion: When $p = 0.5$, equation (18) fails to yield a sign. Furthermore, when $\lambda = 0$, equations (19) and (20) return us to the normal approximation ($\mu = np$ and $\sigma = \sqrt{np(1-p)}$, respectively), which is after all a natural choice for a symmetric binomial distribution.

## 3.4 Restrictions

Unfortunately, the skew-normal is also not universally applicable. To obtain an estimate for $\lambda$, we must put a few restrictions on $n$ and $p$.

If we let $u = \frac{\lambda^2}{1+\lambda^2}$ and $v = 1/u$, we can rewrite the left hand side of (17) as

$$
\begin{aligned}
\frac{\left(1-\frac{2}{\pi}\cdot\frac{\lambda^2}{1+\lambda^2}\right)^3}{\frac{2}{\pi}\left(\frac{\lambda^2}{1+\lambda^2}\right)^3\left(\frac{4}{\pi}-1\right)^2} &= \left(1-\frac{2}{\pi}u\right)^3 \Big/ \frac{2}{\pi}u^3\left(\frac{4}{\pi}-1\right)^2 \\
&= \left(1-\frac{2}{\pi}u\right)^3 \cdot v^3 \cdot \frac{\pi}{2} \cdot \left(\frac{\pi}{4-\pi}\right)^2 \\
&= \left[v\left(1-\frac{2}{\pi}u\right)\right]^3 \left(\frac{\pi^3}{2(4-\pi)^2}\right) \\
&= \left(v-\frac{2}{\pi}\right)^3 \left(\frac{\pi^3}{2(4-\pi)^2}\right) \qquad = g(v). \qquad (21)
\end{aligned}
$$

We can see that $g(v)$ is increasing in $v = \frac{1+\lambda^2}{\lambda^2} \geq 1$. Therefore:

$$
\min_v g(v) = g(v)|_{v=1} = \left(1-\frac{2}{\pi}\right)^3 \left(\frac{\pi^3}{2(4-\pi)^2}\right) = 1.009524 \approx 1, \qquad (22)
$$

which means that the right hand side of (17), which is supposed to be equal to the left hand side of (17), can't ever be less than 1. Unfortunately, it sometimes is; in particular, $\frac{np(1-p)}{(1-2p)^2} \to 0$ when $p \to 0$ or $p \to 1$. So if we want a solution, we must restrict $n$ and $p$ such that

$$
\begin{aligned}
\{\text{right hand side of (17)}\} &\geq \{\text{min of left hand side of (17)}\} \\
\frac{np(1-p)}{(1-2p)^2} &\geq 1 \\
np(1-p) &\geq (1-2p)^2. \qquad (23)
\end{aligned}
$$

Here, two scenarios arise. The first is when we have a fixed $p$ and wish to find the minimum $n$ necessary to derive a skew-normal approximation. From (23), solving for $n$ is very simple:

$$
n \geq \frac{(1-2p)^2}{p(1-p)}\ . \qquad (24)
$$

Figure 1a in appendix A shows the least sample size required to estimate $\lambda$ as a function of $p$.[6] As we would expect, it is larger when $p$ is small and rapidly goes to 0 as $p$ approaches 0.5; for example, when $p = 0.01$, $n$ must be $\geq 98$, but at $p = 0.2$, $n$ need only be $\geq 3$, a trivial requirement to meet.

The second scenario, primarily of academic interest, is when $n$ is fixed and we wish to solve for $p$. In this case, we return to (23) for further factoring:

[6] Since $Bin(n, p)$ and $Bin(n, 1-p)$ are mirror image curves, it is often only necessary to examine either $p \in (0, 0.5)$ or $p \in (0.5, 1)$. We will usually consider the former range of $p$'s.

$$
\begin{aligned}
np(1-p) &\geq (1-2p)^2 \\
np - np^2 &\geq 1 - 4p + 4p^2 \\
1 - 4p + 4p^2 - np + np^2 &\leq 0 \\
(n+4)p^2 - (n+4)p + 1 &\leq 0. 
\end{aligned} \tag{25}
$$

We then apply the quadratic formula with $a = n+4$, $b = -(n+4)$, and $c = 1$:

$$
\begin{aligned}
p &= \frac{(n+4) \pm \sqrt{(n+4)^2 - 4 \cdot (n+4) \cdot 1}}{2(n+4)} \\
&= \frac{(n+4) \pm \sqrt{n^2 + 8n + 16 - 4n - 16}}{2(n+4)} \\
&= \frac{(n+4) \pm \sqrt{n^2 + 4n}}{2(n+4)} \\
&= \frac{n+4}{2(n+4)} \pm \frac{1}{2}\sqrt{\frac{n(n+4)}{(n+4)^2}} \\
&= \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{n}{n+4}}\,.
\end{aligned}
$$

Let $r_1 = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{n}{n+4}}$ and $r_2 = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{n}{n+4}}$. (Note that $r_1 < r_2$.) Now we can rewrite (25) as

$$
(p - r_1)(p - r_2) \leq 0.
$$

Examining the left hand side, when $p < r_1$, both terms are negative and so their product is positive; when $p > r_2$, both terms are positive, again leading the product to be positive. Therefore, our solution lies where $r_1 \leq p \leq r_2$, or more explicitly,

$$
\frac{1}{2} - \frac{1}{2}\sqrt{\frac{n}{n+4}} \;\leq\; p \;\leq\; \frac{1}{2} + \frac{1}{2}\sqrt{\frac{n}{n+4}}\,. \tag{26}
$$

Figure 1b in appendix A shows this range as a function of $n$, this interval grows quickly as $n$ increases, and for sufficiently large $n$, it becomes almost $(0, 1)$. For example, when $n = 100$, our interval is $(0.00971, 0.99029)$; when $n = 500$, it is (0.00199, 0.99801).

Although the presence of these restrictions are somewhat disappointing, we can console ourselves with the observation that at the same $n$ and $p$, the skew-normal yields substantially more accurate approximations than the normal (see section 4.2). Thus while imperfect, it is nevertheless an improvement.

# 4 Demonstrating Improved Accuracy

Now comes the time to justify our efforts by comparing the accuracy of our skew-normal approximation to that of the normal.

We are naturally most interested in cases where other solutions perform poorly. Recall that when $n$ is small, it is feasible to execute the binomial calculations directly, and when $n$ is large or $p$ is close to 0.5, the normal distribution provides an adequate approximation. Therefore, our interest is primarily in cases where $n$ is moderate and $p$ is extreme (close to 0 or 1).[7]

## 4.1 Visual Comparison

The first and most obvious way of judging accuracy is by visual inspection. Figures 2, 3, and 4 in appendix A compare the binomial, normal, and skew-normal at small values of $p = 0.05$, $p = 0.1$, and $p = 0.2$ for each of $n = 25$, $n = 50$, and $n = 100$, respectively.

The graphs in figures 2, 3, and 4 indicate that at moderate $n$ and small $p$, our skew-normal curve follows the shape of the binomial much more closely than the normal. As $n$ grows for each value of $p$, however, the Central Limit Theorem begins to exert its effect and the normal distribution "catches up" in accuracy, slowly for for more extreme values of $p$ and faster as $p$ approaches 0.5. Thus as we would expect, the skew-normal approximation is of greatest value when $n$ is moderate and $p$ is extreme.

## 4.2 Maximal Absolute Error

Another more quantitative method of judging accuracy is comparing the maximal absolute errors of our two approximations, defined by Schader and Schmid (1989) as

$$\mathrm{MABS}(n, p) \;=\; \max_{k \in \{0,1,\ldots,n\}} \left| F_{B(n,p)}(k) - F_{\mathrm{appr}(n,p)}(k + 0.5) \right| \tag{27}$$

where $F_{B(n,p)}$ is the cdf of the binomial and $F_{\mathrm{appr}(n,p)}$ is the cdf of either the normal or skew-normal approximation; the 0.5 is a continuity correction.

Figure 5 in appendix A shows the MABS of the skew-normal and the normal approximations as a function of $p$ for $n = 25$ and $n = 100$. When $p$ is very close to 0, the MABS of the normal approximation is more than four times that of the skew-normal, again demonstrating the benefit of using the skew-normal approximation for moderate $n$ and extreme $p$. The two error curves converge and eventually meet as $p$ approaches 0.5.

Figure 6, on the other hand, shows the MABS of our two approximations as a function of $n$ for $p = 0.05$ and $p = 0.1$. At small to moderate $n$, the normal MABS is roughly six times the skew-

[7] Again, we examine only $p \in (0, 0.5)$. (See footnote 6.)

normal MABS. Interestingly, the error curves converge much more slowly this time, leading to the pleasant (if surprising) conclusion that when $p$ is extreme, the skew-normal gains us a measure of accuracy even at large $n$.

# 5 Practical Resources

We conclude our dicussion by offering the reader a few practical resources:

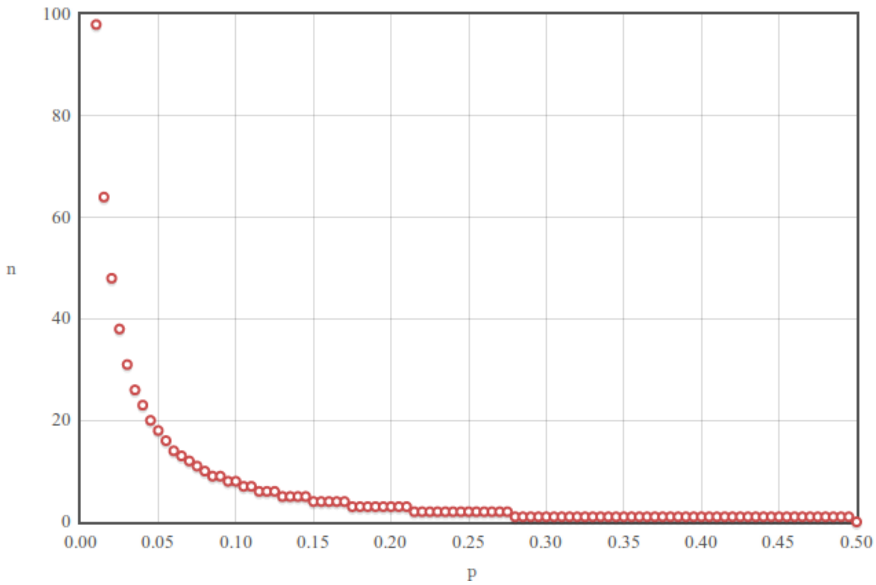
Table 1 in appendix A shows estimations of $SN(\mu, \sigma, \lambda)$ for common binomial distributions.

For those with an unusual combination of $n$ and $p$ not in the table, appendix B demonstrates how to calculate the skew-normal parameters by hand.

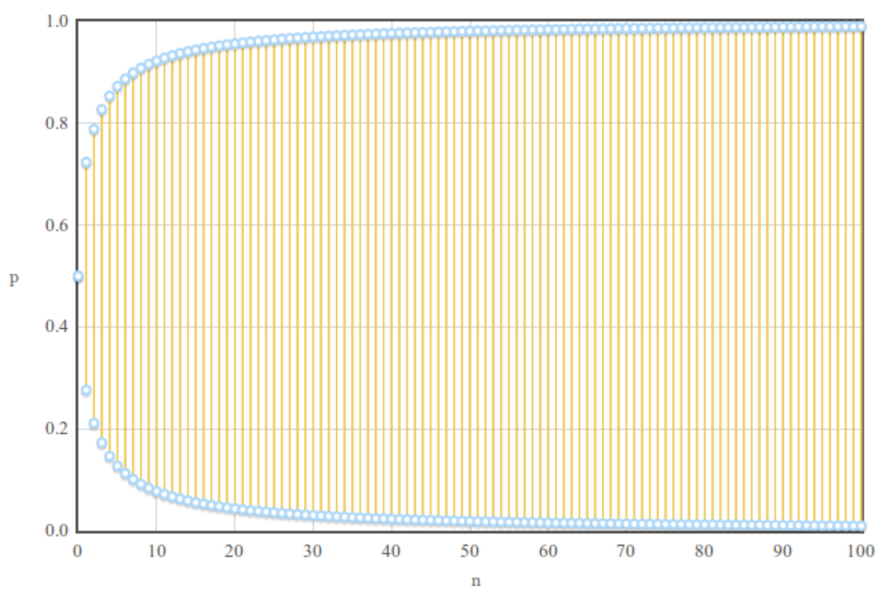
Finally, for rapidly approximating many binomial distributions, the author's Python library, which was used to compute all values presented in this paper, is freely available online:

http://github.com/joycetipping/skew-normal-capstone/

# A Figures



(a) Least possible $n$, given a fixed $p$



(b) Range of possible $p$, given a fixed $n$

Figure 1: Restrictions on $n$ and $p$ necessary to obtain $\lambda$

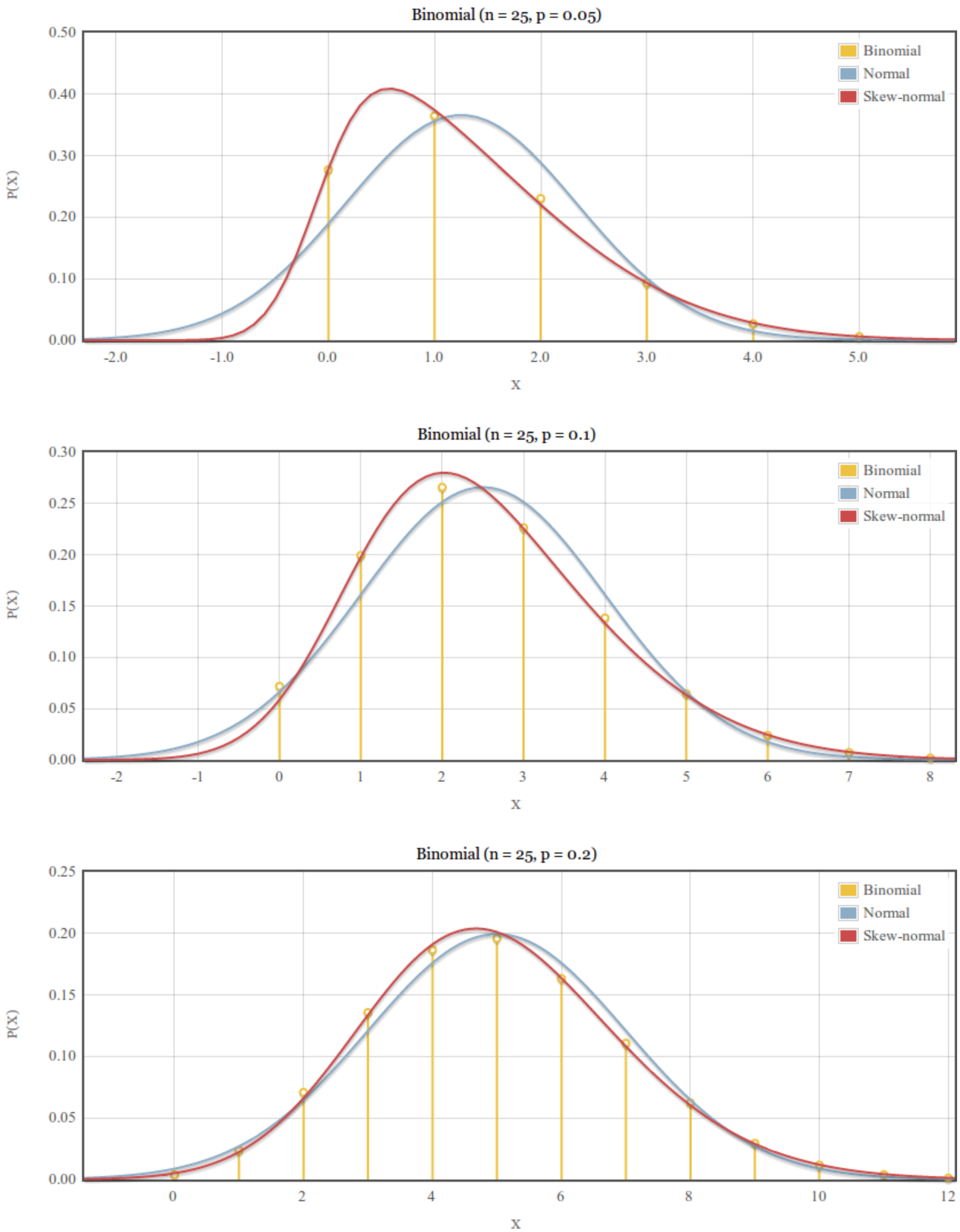



Figure 2: Binomial, normal, and skew-normal, $n = 25$

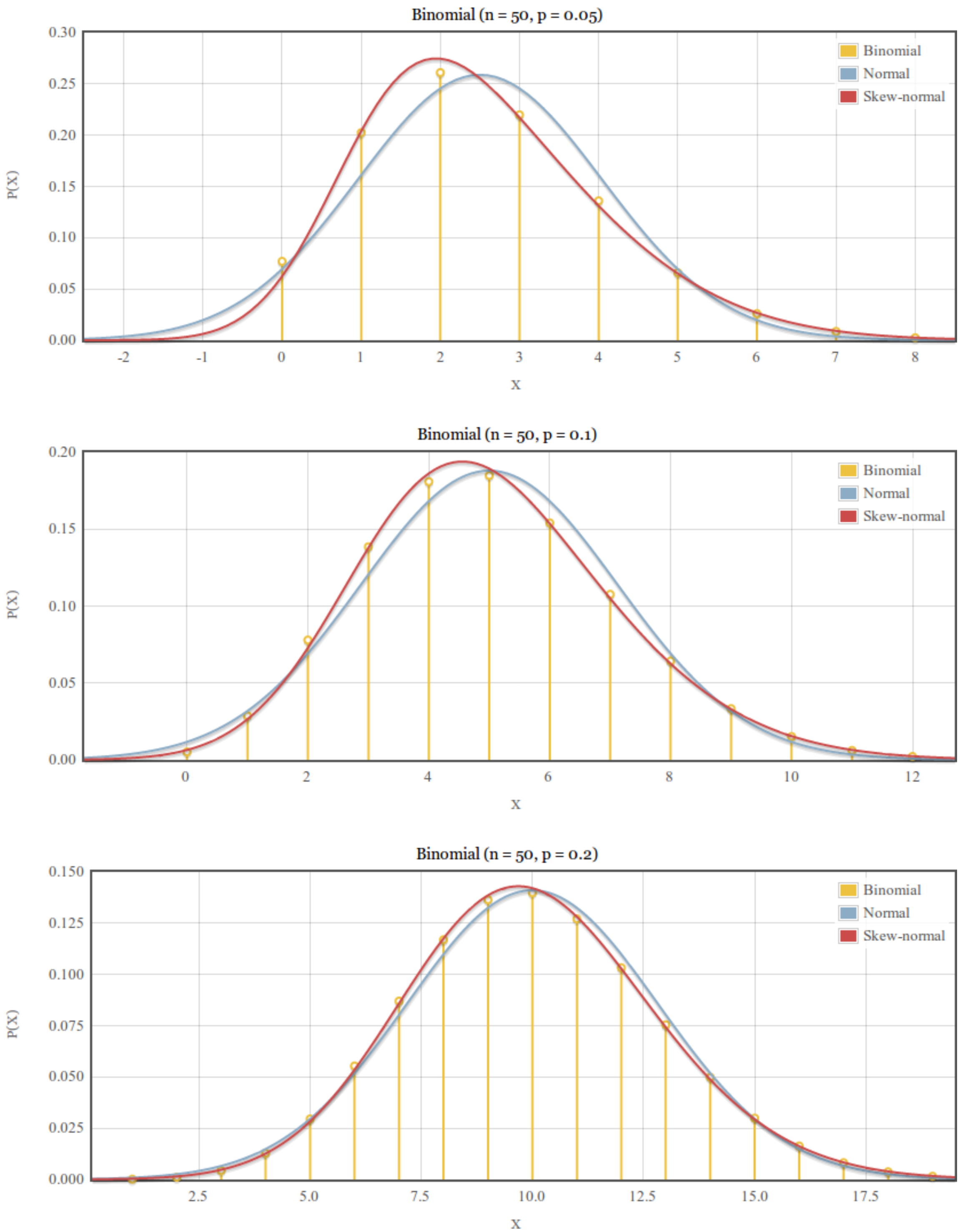


Figure 3: Binomial, normal, and skew-normal, $n = 50$

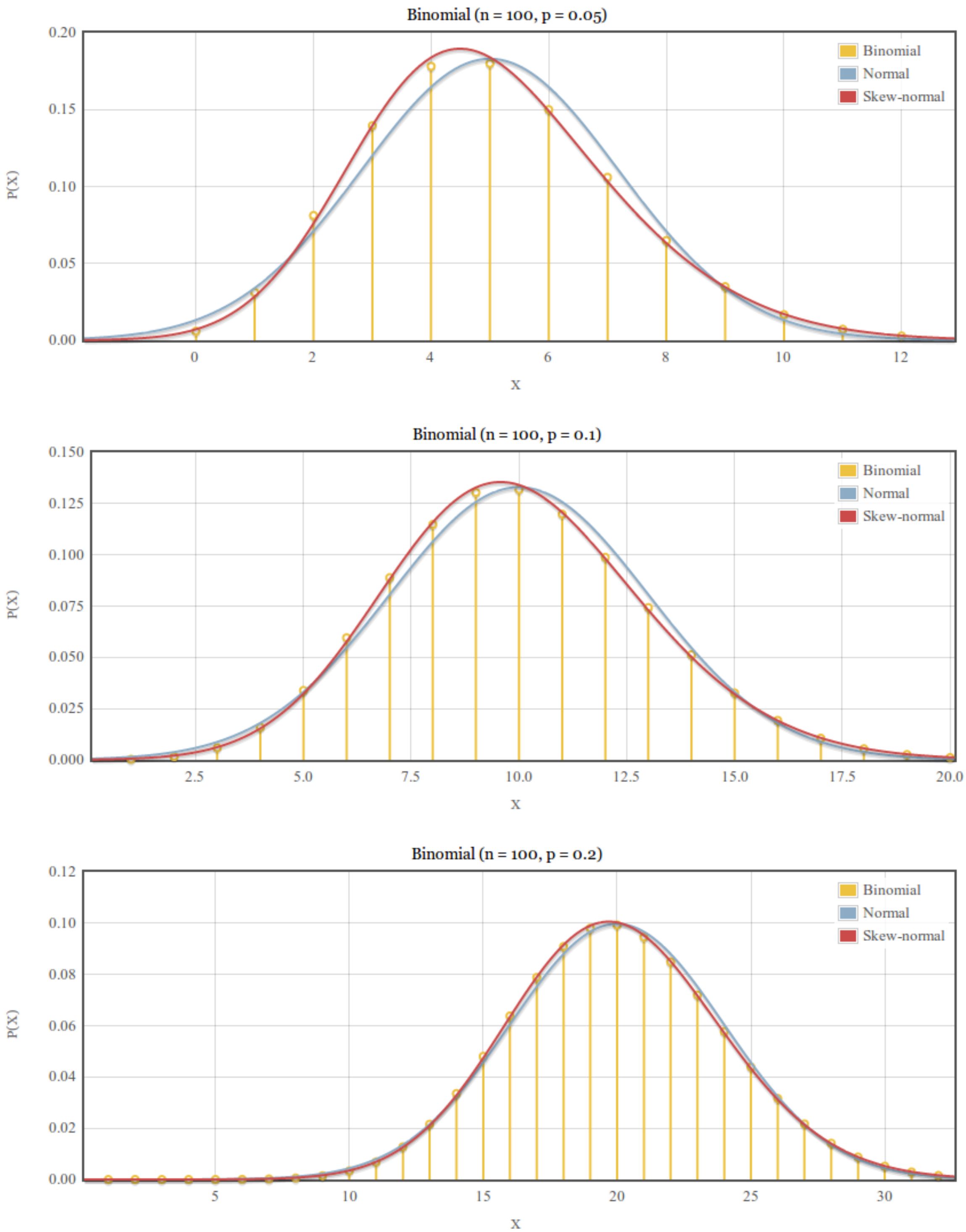



Figure 4: Binomial, normal, and skew-normal, $n = 100$

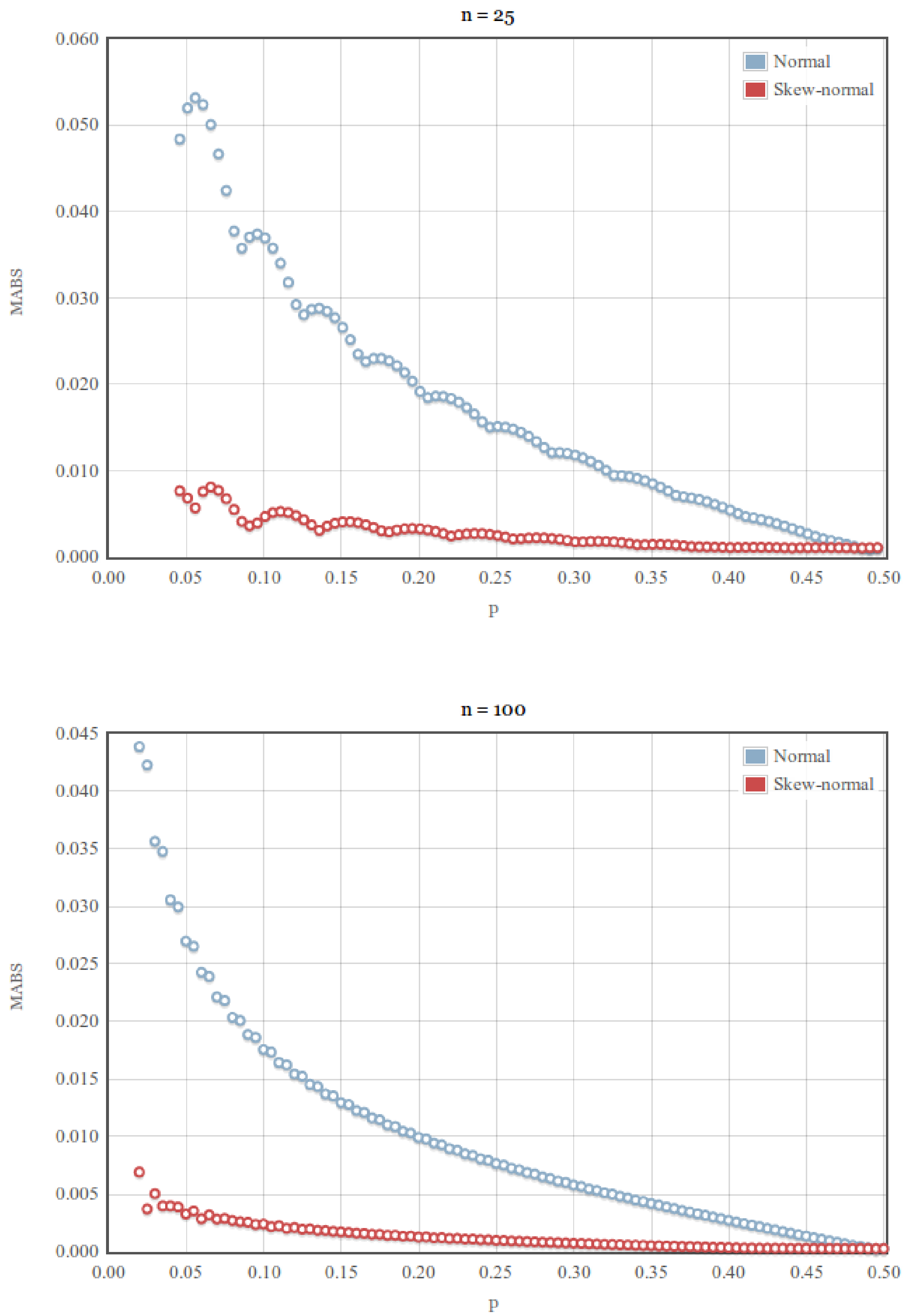



Figure 5: MABS as a function of p

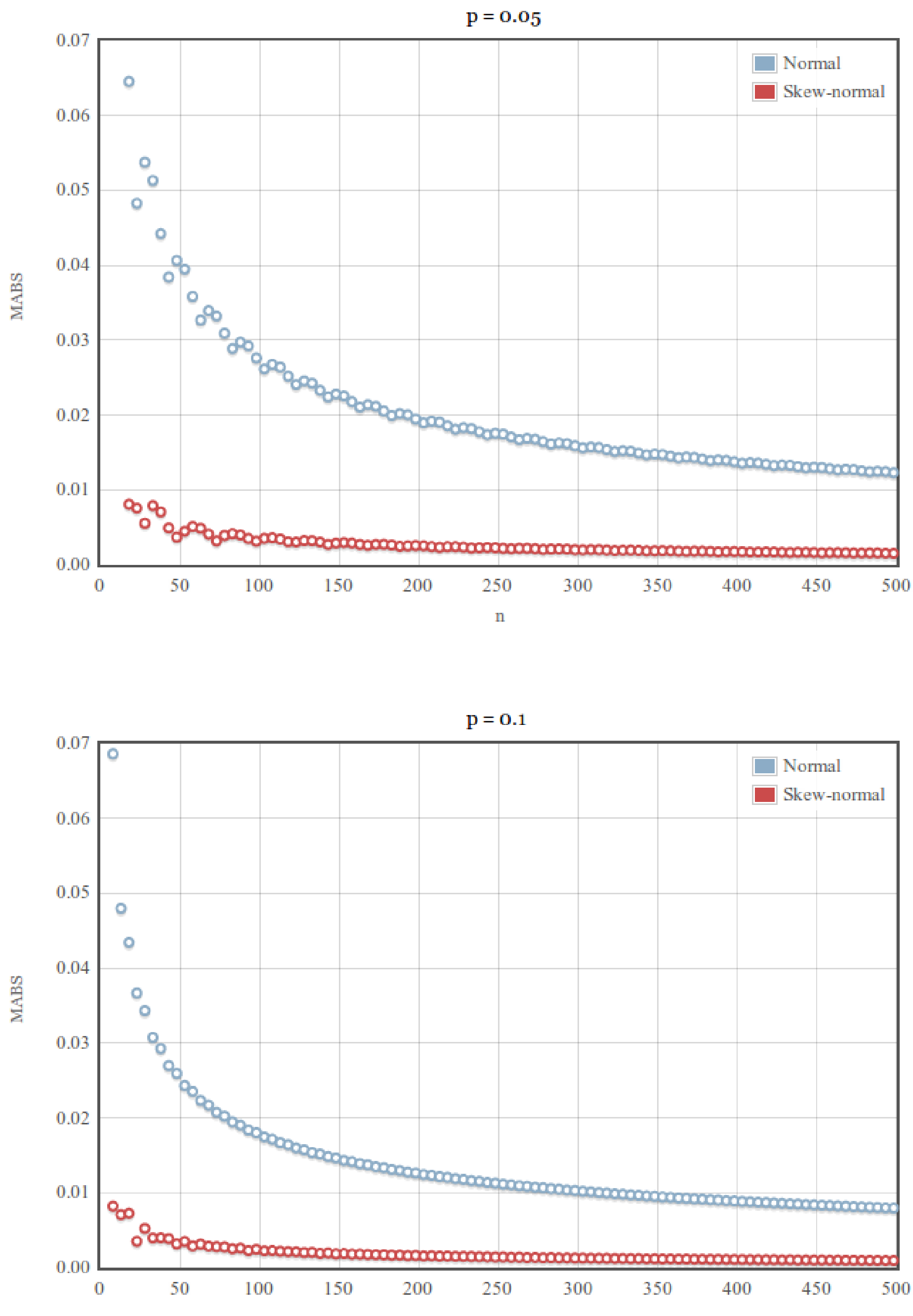



Figure 6: MABS as a function of n

Table 1: Estimations of $SN(\mu, \sigma, \lambda)$ for $Bin(n, p)$

| | | $n$ | | | | |
|---|---|---|---|---|---|---|
| | | 25 | 50 | 100 | 250 | 500 |
| $p$ | 0.05 | (-0.11, 1.74, 4.56) | (0.79, 2.30, 2.54) | (2.85, 3.06, 1.86) | (9.58, 4.52, 1.38) | (21.32, 6.11, 1.15) |
| | 0.10 | (0.89, 2.20, 2.31) | (2.97, 2.94, 1.74) | (7.44, 3.94, 1.40) | (21.53, 5.88, 1.10) | (45.62, 8.01, 0.94) |
| | 0.15 | (2.02, 2.49, 1.79) | (5.32, 3.34, 1.43) | (12.25, 4.51, 1.19) | (33.77, 6.77, 0.96) | (70.30, 9.27, 0.82) |
| | 0.20 | (3.23, 2.67, 1.50) | (7.76, 3.61, 1.24) | (17.18, 4.89, 1.04) | (46.18, 7.39, 0.85) | (95.18, 10.16, 0.74) |
| | 0.25 | (4.49, 2.79, 1.29) | (10.28, 3.78, 1.09) | (22.20, 5.15, 0.93) | (58.71, 7.83, 0.76) | (120.22, 10.80, 0.67) |
| | 0.30 | (5.80, 2.85, 1.12) | (12.86, 3.88, 0.95) | (27.31, 5.32, 0.82) | (71.34, 8.12, 0.68) | (145.39, 11.24, 0.60) |
| | 0.35 | (7.17, 2.86, 0.96) | (15.50, 3.92, 0.83) | (32.49, 5.39, 0.72) | (84.09, 8.28, 0.60) | (170.70, 11.50, 0.53) |
| | 0.40 | (8.59, 2.83, 0.80) | (18.23, 3.89, 0.70) | (37.76, 5.39, 0.61) | (96.96, 8.32, 0.51) | (196.18, 11.60, 0.45) |
| | 0.45 | (10.12, 2.73, 0.61) | (21.08, 3.79, 0.53) | (43.21, 5.29, 0.47) | (110.07, 8.23, 0.40) | (221.93, 11.54, 0.35) |
| | 0.50 | (12.50, 2.50, 0.00) | (25.00, 3.54, 0.00) | (50.00, 5.00, 0.00) | (125.00, 7.91, 0.00) | (250.00, 11.18, 0.00) |
| | 0.55 | (14.88, 2.73, -0.61) | (28.92, 3.79, -0.53) | (56.79, 5.29, -0.47) | (139.93, 8.23, -0.40) | (278.07, 11.54, -0.35) |
| | 0.60 | (16.41, 2.83, -0.80) | (31.77, 3.89, -0.70) | (62.24, 5.39, -0.61) | (153.04, 8.32, -0.51) | (303.82, 11.60, -0.45) |
| | 0.65 | (17.83, 2.86, -0.96) | (34.50, 3.92, -0.83) | (67.51, 5.39, -0.72) | (165.91, 8.28, -0.60) | (329.30, 11.50, -0.53) |
| | 0.70 | (19.20, 2.85, -1.12) | (37.14, 3.88, -0.95) | (72.69, 5.32, -0.82) | (178.66, 8.12, -0.68) | (354.61, 11.24, -0.60) |
| | 0.75 | (20.51, 2.79, -1.29) | (39.72, 3.78, -1.09) | (77.80, 5.15, -0.93) | (191.29, 7.83, -0.76) | (379.78, 10.80, -0.67) |
| | 0.80 | (21.77, 2.67, -1.50) | (42.24, 3.61, -1.24) | (82.82, 4.89, -1.04) | (203.82, 7.39, -0.85) | (404.82, 10.16, -0.74) |
| | 0.85 | (22.98, 2.49, -1.79) | (44.68, 3.34, -1.43) | (87.75, 4.51, -1.19) | (216.23, 6.77, -0.96) | (429.70, 9.27, -0.82) |
| | 0.90 | (24.11, 2.20, -2.31) | (47.03, 2.94, -1.74) | (92.56, 3.94, -1.40) | (228.47, 5.88, -1.10) | (454.38, 8.01, -0.94) |
| | 0.95 | (25.11, 1.74, -4.56) | (49.21, 2.30, -2.54) | (97.15, 3.06, -1.86) | (240.42, 4.52, -1.38) | (478.68, 6.11, -1.15) |

## B Calculating a Skew-Normal Approximation

Although easier with a computer program, calculating estimates for $\mu$, $\sigma$, and $\lambda$ by hand is perfectly possible. Here, we will demonstrate using $n = 25$, $p = 0.1$.

By far the largest battle is finding $\lambda$. We will use Equation (17) but with the simplified left hand side given by (21):

$$\left(\frac{1+\lambda^2}{\lambda^2} - \frac{2}{\pi}\right)^3 \left(\frac{\pi^3}{2(4-\pi)^2}\right) = \frac{np(1-p)}{(1-2p)^2}\,. \tag{28}$$

The closed-formed solution to this equation is long, hideous, and hard to work with, so for this demonstration, we will take a numerical approach.

The left hand side of (28) is a function of lambda; let us denote it $f(\lambda)$. The right hand side is a constant in $n$ and $p$; let us call it $k_{n,p}$. Our goal is to find a value of $\lambda$ such that $f(\lambda)$ is within a certain margin of error, $e$, of $k_{n,p}$. Since we are computing by hand, we will take $e$ to be a modest 0.01.

Recall that the sign of $\lambda$ is determined independently of the value. In fact, $f$ is never affected by the sign of $\lambda$, as all terms are squared. Thus, we can restrict our search for $\lambda$ to the interval $(0, \infty)$.

Next, by taking $f$'s derivative, we can show that it is monotonically decreasing for positive $\lambda$:

$$\begin{aligned}
\frac{d}{d\lambda}\left[\left(\frac{1+\lambda^2}{\lambda^2} - \frac{2}{\pi}\right)^3 \left(\frac{\pi^3}{2(4-\pi)^2}\right)\right] &= \left(\frac{\pi^3}{2(4-\pi)^2}\right)\cdot 3\left(\frac{1+\lambda^2}{\lambda^2} - \frac{2}{\pi}\right)^2 \cdot \left(\frac{2\lambda}{\lambda^2} - \frac{2(1+\lambda^2)}{\lambda^3}\right) \\
&= \left(\frac{\pi^3}{2(4-\pi)^2}\right)\cdot 3\left(\frac{1+\lambda^2}{\lambda^2} - \frac{2}{\pi}\right)^2 \cdot \left(\cancel{\frac{2}{\lambda}} - \frac{2}{\lambda^3} - \cancel{\frac{2}{\lambda}}\right) \\
&= \underbrace{\left(\frac{\pi^3}{2(4-\pi)^2}\right)\cdot 3\left(\frac{1+\lambda^2}{\lambda^2} - \frac{2}{\pi}\right)^2}_{\text{Always positive}} \cdot \underbrace{\left(-\frac{2}{\lambda^3}\right)}_{\text{Negative when } \lambda > 0}.
\end{aligned}$$

This convenient fact allows us to find lower and upper bounds for $\lambda$ and repeatedly bisect our interval until we are within $e$ of $k_{n,p}$.

(For the following calculations, it is helpful to keep in mind that because $f$ is decreasing in $\lambda$, smaller values of $\lambda$ will produce larger values of $f$, and vice versa.)

1. *Find $k_{n,p}$.*

   Our value: $k_{n,p} = \dfrac{25 \cdot 0.1 \cdot 0.9}{(1-2\cdot 0.1)^2} = 3.5156.$

2. *Find a and b such that $f(a) > k_{n,p} > f(b)$.*

   Our values: $a = 1$, $b = 3$.

3. *Repeatedly bisect $(a, b)$ until $f(c)$ is within e of $k_{n,p}$.*

   Calculate $c = \frac{a+b}{2}$.

   - If $f(c) \leq k_{n,p} - 0.01$, we need a small value of $c$, so we take our new interval to be $(a, c)$.
   - If $f(c) \geq k_{n,p} + 0.01$, we need a larger value of $c$, so we take our new interval to be $(c, b)$.

   Repeat this step until $f(c)$ is within $e$ of $k_{n,p}$, or more precisely $k_{n,p} - 0.01 < f(c) < k_{n,p} + 0.01$.

   The following table shows our iterations:

   | Iteration | $a$ | $b$ | $c$ | $f(c)$ | $f(c) \leq k_{n,p} - 0.01$ | $f(c) \geq k_{n,p} + 0.01$ |
   |---|---|---|---|---|---|---|
   | 1 | 2.00 | 3.000 | 2.5000 | 3.0164 | True | False |
   | 2 | 2.00 | 2.500 | 2.2500 | 3.7129 | False | True |
   | 3 | 2.25 | 2.500 | 2.3750 | 3.3252 | True | False |
   | 4 | 2.25 | 2.375 | 2.3125 | 3.5076 | False | False |

   We take the last value of $c$: 2.3125.

4. *Find the sign of $(1 - 2p)$.*

   Our $p = 0.1 \;\Rightarrow\; (1 - 2 \cdot 0.1) = 0.8 \;\Rightarrow\;$ positive.

5. *Final answer: {sign of $(1 - 2p)$}$\lambda$*

   Our final answer: $\lambda = 2.3125$.

Once we have $\lambda$, we can easily find $\sigma$:

$$\sigma = \sqrt{\frac{np(1-p)}{1 - \frac{2}{\pi} \cdot \frac{\lambda^2}{1+\lambda^2}}} = \sqrt{\frac{25 \cdot 0.1 \cdot 0.9}{1 - \frac{2}{\pi} \cdot \frac{2.3125^2}{1+2.3125^2}}} = 2.2029.$$

And with $\lambda$ and $\sigma$, we can also find $\mu$:

$$\mu = np - \sigma \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{\lambda}{\sqrt{1+\lambda^2}} = 25 \cdot 0.1 - 2.2029 \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{2.3125}{\sqrt{1+2.3125^2}} = 0.8867.$$

## References

Adelchi Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178, 1985.

Adelchi Azzalini. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32:159–188, 2005.

Lee J. Bain and Max Engelhardt. *Introduction to Probability and Mathematical Statistics*. Duxbury, 2nd edition, 1992.

Ching-Hui Chang, Jyh-Jiuan Lin, Nabendu Pal, and Miao-Chen Chiang. A note on improved approximation of the binomial distribution by the skew-normal distribution. *American Statistical Association*, 62(2):167–170, May 2008.

A. O'Hagan and Tom Leonard. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63:201–203, 1976.

Arthur Pewsey. Problems of inference for azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27(7):859–870, 2000.

Martin Schader and Friedrich Schmid. Two rules of thumb for the approximation of the binomial distribution by the normal distribution. *American Statistical Association*, 43(1):23–24, February 1989.