В изложении доцента БГТУ Сахарова В.Ю.

Пусть произведено N опытов со случайной величиной Бернулли с вероятностью p выпадения единицы в одном отдельно взятом испытании. Получены N значений (нули и единицы): X_1, X_2, … X_N. Пусть единиц из них K. Это число можно трактовать как сумму независимых случайных величин $K = \sum_{i=1}^{N} X_i$. Согласно центральной предельной теореме сумма большого количества одинаково распределённых независимых случайных величин (имеющих математические ожидания и дисперсии) с ростом числа слагаемых стремиться к нормальной случайной величине. Количество успехов $K = \sum_{i=1}^{N} X_i$ как раз и есть такая сумма. То есть при большом числе испытаний N случайная величина K приближённо может считаться нормальной. Для упрощения её можно «центрировать» (то есть вычесть из неё её же математическое ожидание, чтобы в итоге математическое ожидание оказалось равно нулю) и «нормировать» (то есть разделить на её же среднеквадратическое отклонение, чтобы в итоге дисперсия оказалась равна единице). Математическое ожидание $EK=Np$. Дисперсия $DK=Np(1\text{-}p)$. Средневкадратическое отклонение $\sigma(K) = \sqrt{DK} = \sqrt{Np(1-p)}$. Итак, стандартное нормальное распределение будет иметь величина $Z = \dfrac{K - EK}{\sigma(K)} = \dfrac{K - Np}{\sqrt{Np(1-p)}}$. Выберем достаточно большую вероятность β (по мнению исследователя равную вероятности практически достоверного события) и построим интервал ($-c$;c) практически гарантировано содержащий случайную величину Z. Очевидно, что чем больше будет эта вероятность, тем длиннее окажется этот интервал. Такую вероятность просто находить как разность значений функции распределения Ф(x) стандартного нормального распределения

$$P\left(-c < \frac{K - Np}{\sqrt{Np(1-p)}} < c\right) = P(-c < Z < c) = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1$$

По нашему условию эта вероятность равна β.

$$2\Phi(c) - 1 = \beta$$

$$2\Phi(c) = 1 + \beta$$

$$\Phi(c) = \frac{1+\beta}{2}$$

В программе Microsoft Office Excel встроена функция НОРМСТОБР(вероятность), с помощью которой можно находить значения функции, обратной к функции распределения Ф(x). Обозначим здесь эту функцию G(x). Тогда

$$c = G\left(\frac{1+\beta}{2}\right)$$

Итак, c это известное число и можно относительно неизвестного параметра p разрешить неравенство

$$-c < \frac{K - Np}{\sqrt{Np(1-p)}} < c$$

Возведём это неравенство в квадрат.

$$\frac{(K - Np)^2}{Np(1-p)} < c^2$$

И получим квадратное неравенство.

$$(K - Np)^2 < c^2 Np(1-p)$$

$$K^2 - 2KNp + N^2p^2 < c^2Np - c^2Np^2$$

$$\left(N^2 + c^2N\right)p^2 - \left(2KN + c^2N\right)p + K^2 < 0$$

Коэффициент при квадрате положительный, поэтому решением неравенства будет промежуток, ограниченный корнями соответствующего квадратного уравнения. Найдём эти корни с помощью дискриминанта.

$$D = \left(2KN + c^2N\right)^2 - 4\left(N^2 + c^2N\right)K^2 = 4K^2N^2 + 4Kc^2N^2 + c^4N^2 - 4N^2K^2 - 4c^2NK^2 =$$

$$= 4Kc^2N^2 + c^4N^2 - 4c^2NK^2 = 4Kc^2N(N-K) + c^4N^2 > 0$$

Корни квадратного уравнения найдутся по известной формуле

$$p_{1,2} = \frac{2KN + c^2N \pm \sqrt{4Kc^2N(N-K) + c^4N^2}}{2\left(N^2 + c^2N\right)} = \frac{2K + c^2 \pm \sqrt{4Kc^2\left(1 - \frac{K}{N}\right) + c^4}}{2\left(N + c^2\right)} =$$

$$= \frac{K + \frac{c^2}{2} \pm \sqrt{Kc^2\left(1 - \frac{K}{N}\right) + \frac{c^4}{4}}}{N + c^2} = \frac{K + \frac{c^2}{2} \pm \sqrt{Kc^2\left(1 - \frac{K}{N}\right) + \left(\frac{c^2}{2}\right)^2}}{N + c^2}$$

Решением квадратного неравенства будет

$$\frac{K + \frac{c^2}{2} - \sqrt{Kc^2\left(1 - \frac{K}{N}\right) + \left(\frac{c^2}{2}\right)^2}}{N + c^2} < p < \frac{K + \frac{c^2}{2} + \sqrt{Kc^2\left(1 - \frac{K}{N}\right) + \left(\frac{c^2}{2}\right)^2}}{N + c^2}$$

То есть доверительным интервалом для параметра p при доверительной вероятности β является промежуток

$$\left(\frac{K + \frac{c^2}{2} - \sqrt{Kc^2\left(1 - \frac{K}{N}\right) + \left(\frac{c^2}{2}\right)^2}}{N + c^2}; \frac{K + \frac{c^2}{2} + \sqrt{Kc^2\left(1 - \frac{K}{N}\right) + \left(\frac{c^2}{2}\right)^2}}{N + c^2}\right)$$

Где $c = G\left(\frac{1+\beta}{2}\right)$.

Заметим, что число K/N не является центром этого интервала.

Можно установить, что центр интервала несколько смещён относительно дроби K/N к числу ½.