Об одном видоизменении неравенства Чебышева и о погрешности формулы Лапласа.

С. Бернштейн.

§ 1. В настоящей статье я имею в виду, главным образом, показать, что, применяя надлежащим образом классическое рассуждение Чебышева, возможно получить неравенство значительно более точное, чем неравенство Чебышева, если только допустить некоторые ограничительные условия, обычно осуществляющиеся на практике. Затем мы воспользуемся полученным результатом для исследования погрешности формулы Лапласа.

Пусть x, y, z... представляют собой независимые величины, математические ожидания которых равны 0; пусть, далее, a_k, b_k, c_k ... будут математические ожидания x^k, y^k, z^k, причем можно указать такое число L, что математические ожидания степени выше третьей удовлетворяют неравенствам вида

$$\left|a_k\right| \leq \frac{k!}{4!}\left(\frac{L}{5}\right)^{k-4} a_4, \quad \left|b_k\right| \leq \frac{k!}{4!}\left(\frac{L}{5}\right)^{k-4} b_4 \text{ и т. д.}$$

В таком случае, полагая $A_2 = a_2 + b_2 + c_2 ...$, $A_3 = a_3 + b_3 + c_3 ...$, $A_4 = a_4 + b_4 + c_4 + ...$, неравенство

$$\left|(x+y+z...) - \frac{t^2}{3}\frac{A_3}{A_2}\right| < t\sqrt{2A_2}\left[1 + \frac{A_4 t^2}{6A_2^2}\right] \tag{1}$$

имеет вероятность большую, чем $1 - 2e^{-t^2}$ $\left(\text{если только } t \leq \frac{5}{4L}\sqrt{2A_2}\right)$.

В самом деле,

$$I = \text{Мат. ожид.}\left[e^{\varepsilon(x+y+z+...)}\right] = \text{Мат. ож.}\left[e^{\varepsilon x}\right] . \text{Мат. ож.}\left[e^{\varepsilon y}\right] . \text{Мат. ож.}\left[e^{\varepsilon z}\right] ...$$

Но

$$\text{мат. ож.}\left[e^{\varepsilon x}\right] = 1 + \frac{\varepsilon^2}{2}a_2 + \frac{\varepsilon^3}{3!}a_3 + \frac{\varepsilon^4}{4!}a_4 + \ldots =$$

$$= 1 + \frac{\varepsilon^2}{2}a_2 + \frac{\varepsilon^3}{3!}a_3 + \frac{\varepsilon^4}{4!}a_4\left[1 + \Theta\sum_1^\infty\left[\left(\frac{\varepsilon L}{5}\right)^n\right]\right],$$

где $0<\Theta<1$. Поэтому, полагая $\frac{\varepsilon L}{5} \leq \frac{1}{2}$, находим

$$\text{мат. ож.}\left[e^{\varepsilon x}\right] = 1 + \frac{\varepsilon^2}{2}a_2 + \frac{\varepsilon^3}{3!}a_3 + \frac{\Theta\varepsilon^4}{12}a_4$$

Откуда

$$\log\left[\text{Мат. ож. } e^{\varepsilon x}\right] < \frac{\varepsilon^2 a_2}{2} + \frac{\varepsilon^3 a_3}{6} + \frac{\varepsilon^4 a_4}{12}$$

Применяя аналогичные неравенства к остальным величинам и складывая их, получим

$$\log I < \frac{\varepsilon^2 A_2}{2} + \frac{\varepsilon^3 A_3}{6} + \frac{\varepsilon^4 A_4}{12} \tag{2}$$

Точно таким же образом, полагая $I_1 = \text{Мат. ож.}\left[e^{-\varepsilon(x+y+z+\ldots)}\right]$, найдем

$$\log I_1 < \frac{\varepsilon^2 A_2}{2} - \frac{\varepsilon^3 A_3}{6} + \frac{\varepsilon^4 A_4}{12}. \tag{3}$$

Применяя теперь известное рассуждение Чебышева, заключаем, что вероятность неравенства

$$e^{\varepsilon(x+y+z+\ldots)} > e^{t^2} I \tag{4}$$

менее, чем e^{-t^2}. Но неравенство (4) равнозначно неравенству

$$x+y+z+\ldots > \frac{t^2+\log I}{\varepsilon}.$$

Подставляя (2) находим, что вероятность неравенства

$$x+y+z+\ldots \geq \frac{t^2}{\varepsilon} + \frac{\varepsilon A_2}{2} + \frac{\varepsilon^2 A_3}{6} + \frac{\varepsilon^3 A_4}{12} \tag{5}$$

менее, чем e^{-t^2}, и полагая, наконец, $\varepsilon^2 A_2 = 2t^2$, заменяем неравенство (5) неравенством

$$(x+y+z+\ldots) - \frac{t^2}{3}\frac{A_3}{A_2} \geq t\sqrt{2A_2}\left(1+\frac{A_4 t^2}{6A_2^2}\right) \tag{6}$$

Подобным же образом убеждаемся, пользуясь неравенством (3), что и неравенство

$$(x+y+z+\dots)-\frac{t^2}{3}\frac{A_3}{A_2}\leq -t\sqrt{2A_2}\left(1+\frac{A_4 t^2}{6A_2^2}\right) \tag{6 bis}$$

имеет вероятность меньшую, чем e^{-t^2}.

Следовательно, вероятность неравенства

$$\left|(x+y+z+\cdots)-\frac{t^2}{3}\frac{A_3}{A_2}\right|<t\sqrt{2A_2}\left[1+\frac{A_4t^2}{6A_2^2}\right]. \tag{1}$$

более, чем $1-2e^{-t^2}$, при условии, что $t=\varepsilon\sqrt{\frac{A_2}{2}}\leq\frac{5}{4L}\sqrt{2A_2}$, что и требовалось доказать.

§ 2. Наше неравенство становится точнее неравенства Чебышева только при значениях t, для которых $e^{t^2}>4t^2$ и представляет особый интерес для значений t, превышающих несколько единиц, когда точность его приближается к точности предельной формулы Лапласа-Ляпунова, согласно которой вероятность неравенства (1) имеет пределом

$$\Phi(t)=\frac{2}{\sqrt{\pi}}\int_0^t e^{-t^2}dt,$$

при условии, что $\frac{A_4}{A_2^2}$ стремится к 0.

Остановимся для примера на случае ряда независимых опытов, при которых вероятность появления события E равна соответственно $p_1, p_2, p_3, \dots, p_n$. Очевидно, что в данном случае, как и всегда, когда рассматриваемые независимые величины ограничены, теорема применима. Поэтому вероятность неравенства

$$\left|m-\sum_{i=1}^{n}p_i-\frac{t^2}{3}\frac{A_3}{A_2}\right|<t\sqrt{2A_2}\left(1+\frac{A_4t^2}{6A_2^2}\right) \tag{7}$$

больше, чем $1-2e^{-t^2}$, где m число появлений E при n опытах, $A_2=\sum_1^n p_iq_i$, $A_3=\sum_1^n p_iq_i(q_i-p_i)$, $A_4=\sum_1^n p_iq_i\,(p_i^3+q_i^3)$, для всех значений $t\leq\frac{5}{4}\sqrt{2A_2}$.

В частности, если все вероятности $p_i=p$, неравенство (7) получит форму

$$\left|m-np-\frac{t^2}{3}(q-p)\right|<t\sqrt{2npq}\left[1+\frac{t^2(p^3+q^3)^2}{6npq}\right]. \tag{8}$$

Пусть, например, $p=\frac{1}{50}$, $n=20.000$, $t=5$; тогда неравенство (8) получит форму [1])

$$|m-408|<140\left(1+\frac{5}{784}\right),$$

вероятность которого, следовательно, более, чем $12-e^{-25}=0,99999.99999.7$, что позволяет нам утверждать, что предельное значение даваемое в данном случае предельною формулой Лапласа, которое, как не трудно вычислить, равно 0,99999.99999.8, правильно с точностью до $\frac{1}{10^{10}}$. Неравенство Чебышева даже для более широкого неравенства

$$|m-400|\leq 150$$

дает в качестве нижнего предела для вероятности только 0,984.

Если наше неравенство во многих случаях позволяет установить, что погрешность формулы Лапласа для больших значений t весьма незначительна, даже при сравнительно небольших значениях n, то с другой стороны,—этим неравенством можно также воспользоваться, чтобы показать, что, если $\sum_1^n p_i$ медленно возрастает, то даже для чрезвычайно больших значений n, предельная формула Лапласа в том виде, как она обычно применяется, дает погрешность бóльшую, чем наше неравенство. Действительно, положим n настолько большим, чтобы $\frac{t^3A_4}{3\sqrt{2A^3_2}}$ было правильной дробью; тогда выбирая t так, чтобы левая часть неравенства (7) была целым числом, можем заменить его эквивалентным неравенством

$$\left|m-\Sigma p_i-\frac{t^2}{3}\frac{A_3}{A_2}\right|<t\sqrt{2A_2} \tag{9}$$

По теореме Ляпунова, вероятность этого неравенства (9) при бесконечном возрастании A_2 имеет пределом

$$F(t)=\frac{1}{2}\left[\Phi\left(t+\frac{t^3A_3}{3\sqrt{2A^3_2}}\right)+\Phi\left(t-\frac{t^3A_3}{3\sqrt{2A^3_2}}\right)\right] \tag{10}$$

Замечая, что из асимптотического равенства, для больших значений t,

$$\Phi(t)\sim 1-\frac{e^{-t^2}}{t\sqrt{\pi}} \tag{11}$$

[1]) Заметим, что во второй части неравенства можно отбросить дробь $\frac{5}{784}$, так как m должно быть целым числом.

вытекает асимптотическое равенство

$$F(t) \sim 1 - \frac{e^{-(t-k)^2}}{(t-k)2\sqrt{\pi}}, \tag{12}$$

где $k = \frac{t^3 |A_3|}{3\sqrt{2A_2^3}}$, заключаем, что, если k не очень мало, то $F(t) < 1 - 2e^{-t^2}$, между тем, как по нашей теореме, точное значение рассматриваемой вероятности должно быть более, чем $1 - 2e^{-t^2}$. Например, если $p_i = \frac{1}{i}$, то даже для огромного значения $n = e^{2048}$, полагая $t = 5$, находим для $F(t) = 0{,}99999.99995.8$, между тем, как $1 - 2e^{-t^2} = 0{,}99999.999997$, откуда видно, что в то время, как вероятность неосуществления неравенства (9) в действительности, согласно нашей теореме, менее, чем $3 . 10^{-11}$, предельная формула дает $42 . 10^{-11}$.

§ 3. Из предыдущего видно, что предельная формула Лапласа, которой обычно пользуются, как приближенным значением для вероятности неравенства

$$\left| m - \sum p_i \right| < t\sqrt{2A_2},$$

более точно рассматривать, как приближенное значение вероятности неравенства

$$\left| m - \sum p_i - \frac{t^2}{3}\frac{A_3}{A_2} \right| < t\sqrt{2A_2}.$$

Для более полного обоснования нашего замечания, припомним, ограничиваясь случаем постоянной вероятности p, классический вывод формулы Лапласа, который состоит в том, что при помощи формулы Стирлинга для вероятности $P_{m,n} = C_n^m p^m q^{n-m}$ получается ассимтоптическое выражение

$$P_{m,n} \sim \sqrt{\frac{n}{2\pi m(n-m)}} \left[\frac{np}{m}\right]^m \left[\frac{nq}{n-m}\right]^{n-m}; \tag{13}$$

относительная погрешность этого выражения весьма мала даже для небольших значений n, m, $n-m$. После этого доказывается, что, для

$$m = np + z\sqrt{2npq}, \tag{14}$$

выражение $w=\left[\frac{m}{np}\right]^m\left[\frac{n-m}{nq}\right]^{n-m}$ имеет пределом e^{z^2} при безконечном возрастании n. На эту последнюю часть рассуждения я и хотел бы обратить внимание, чтобы внести в нее соответствующее изменение.

Не ограничиваясь целыми значениями m, поставим себе вопрос: определить m в функции z таким образом, чтобы

$$w=e^{z^2} \tag{15}$$

при всяком значении n. С этою целью заменяем равенство (14) равенством

$$m=n\left[p+y\left(\sqrt{2pq}+\varphi(y)\right)\right], \tag{16}$$

где $y=\frac{z}{\sqrt{n}}$, а $\varphi(y)$ неизвестная функция, которая определится из уравнения (15), получающего вид, после логарифмирования,

$$m\log\frac{m}{np}+\left(n-m\right)\log\frac{n-m}{nq}=ny^2,$$

или

$$\left[p+y\left(\sqrt{2pq}+\varphi(y)\right)\right]\log\left[1+\frac{y}{p}\left(\sqrt{2pq}+\varphi(y)\right)\right]+$$

$$\left[q-y\left(\sqrt{2pq}+\varphi(y)\right)\right]\log\left[1-\frac{y}{q}\left(\sqrt{2pq}+\varphi(y)\right)\right]=y^2.$$

Полагая, для краткости, $\sqrt{2pq}+\varphi(y)=u$, находим, наконец, пользуясь разложением логарифма, уравнение

$$\frac{u^2}{2}\left(\frac{1}{p}+\frac{1}{q}\right)+\frac{u^3y}{2.3}\left(\frac{1}{q^2}-\frac{1}{p^2}\right)+\ldots+\frac{u^ky^{k-2}}{k(k-1)}\left(\frac{1}{q^{k-1}}-\frac{1}{(-p)^{k-1}}\right)+\cdots=1, \tag{17}$$

которое, при $y=0$, $u=\sqrt{2pq}$, обращается в тождество, и т. к. для этих значений производная уравнения (17) отлична от 0, то, на основании известной теоремы о неявных функциях, u разлагается в строку Тейлора

$$u=\sqrt{2pq}+\frac{q-p}{3}y+\ldots, \tag{18}$$

сходящуюся для значений y не превышающих некоторого определенного числа R. Вычисление дальнейших коэффициентов ряда (18) имеет мало значения, т. к. подставляя (18) в формулу (16), получим

$$m=np+z\sqrt{2npq}+\frac{q-p}{3}z^2+\cdots \tag{19}$$

решение поставленного нами вопроса, откуда видно, что дальнейшие члены сходящегося ряда (19) с возрастанием n стремятся к 0, чего нельзя сказать о последнем написанном члене $\frac{q-p}{3}z^2$. Таким образом,

вероятность равную $\sqrt{\frac{n}{2\pi m\,(n-m)}e}$ точнее всего приписать целому числу ближайшему к

$$np+z\sqrt{2npq}+z^2\frac{q-p}{3}.$$

Более детальное развитие этих соображений дает возможность определить погрешность формулы Лапласа, применяемой с предлагаемой нами поправкой.

§ 4. Не останавливаясь на промежуточных [1]), довольно кропотливых (но не представляющих принципиальных трудностей) вычислениях, в которых я предполагаю, что число опытов n удовлетворяет неравенству $npq \geq 365$, мы получаем следующую точную формулу

$$P_{m,n}=\frac{\Delta z}{\sqrt{\pi}}e^{-\left(z+\frac{\psi\Delta z}{2}\right)^2} \tag{20}$$

где $|\psi|<1$, если z и Δz определяются условиями

$$m+\alpha=np+z\sqrt{2npq}+\frac{z^2(q-p)}{3},$$

$$m+1+\alpha=np+(z+\Delta z)\sqrt{2npq}+\frac{(z+\Delta z)^2}{3}(q-p),$$

причем $-1<\alpha<\frac{1}{2}$ и $\frac{2}{\sqrt{2npq}}<z\leq\sqrt[6]{2npq}$

Из (20) заключаем немедленно, что

$$P_{m,n}>\frac{\Delta z}{\sqrt{\pi}}e^{-\left(z+\frac{\Delta z}{2}\right)^2}>\frac{1}{\sqrt{\pi}}\int_a^b e^{-z^2}dz,$$

где a и b определяются соответственно из уравнений

$$m+\frac{1}{2}=np+a\sqrt{2npq}+\frac{q-p}{3}a^2,\quad m+\frac{3}{2}=np+b\sqrt{2npq}+\frac{q-p}{3}b^2;$$

---

[1]) Укажем, еще одно выражение для $P_{m;n}=\sqrt{\frac{n}{2\pi m\,(n-m)}}e^{-z^2-\frac{\varphi z^4}{4npq}-\frac{\theta}{10npq}}$ где $0<\theta<1$ и $-\frac{1}{4}<\varphi<1$ верное для всякого n, при условии, что $|z|<\frac{1}{4}\sqrt{2npq}$

и с другой стороны, замечая, что

$$\Delta z\, e^{-\left(z_0-\frac{\Delta z}{2}\right)^2} < \int_{z_0-\Delta z}^{z_0} e^{-z^2}dz,$$

видим, что

$$P_{m,\,n} < \frac{\Delta z}{\sqrt{\pi}} e^{-\left(z-\frac{\Delta z}{2}\right)^2} < \frac{1}{\sqrt{\pi}} \int_{a'}^{b'} e^{-z^2}dz,$$

где a' и b' определятся соответственно из уравнений

$$m-1=np+a'\sqrt{2npq}+\frac{q-p}{3}a'^2$$

$$m=np+b'\sqrt{2npq}+\frac{q-p}{3}b'^2.$$

Следовательно, формулу (20) можно заменить формулой

$$P_{m,\,n}=\frac{1}{\sqrt{\pi}}\int_{z_0}^{z_1} e^{-z^2}dz \tag{21}$$

где z_0 и z_1, соответственно, положительные корни уравнений

$$m+\alpha=np+z_0\sqrt{2npq}+\frac{q-p}{3}z_0^2,\quad m+1+\alpha=np+z_1\sqrt{2npq}+\frac{q-p}{3}z_1^2,$$

при условии, что $-1<\alpha<\frac{1}{2}$, $npq\leq 365$, и положительный корень z уравнения $m=np+z\sqrt{2npq}+\frac{q-p}{3}z^2$ удовлетворяет неравенству

$$\frac{2}{\sqrt{2npq}}\leq z\leq \sqrt[6]{2npq}.$$

Из (21) выводим, наконец, что вероятность неравенства $m_0\leq m<m_1$ равна

$$\sum_{m_0}^{m_1} P_{m,\,n}=\frac{1}{\sqrt{\pi}}\int_{z_0}^{z_1} e^{-z^2}dz, \tag{22}$$

где z_0 и z_1 определятся из уравнений

$$m_0+\alpha=np+z_0\sqrt{2npq}+\frac{q-p}{3}z_0^2,$$

$$m_0+\alpha=np+z_1\sqrt{2npq}+\frac{q-p}{3}z_1^2,$$

(сохраняя прежние предположения).

Полученному результату можно придать несколько иную форму: пусть z_1 и z_0 два положительные числа, удовлетворяющие условию, что разность

$$\left(z_1\sqrt{2npq}+\frac{q-p}{3}z_1^2\right)-\left(z_0\sqrt{npq}+\frac{q-p}{3}z_0^2\right)$$

равна целому числу; в таком случае

$$\frac{1}{\sqrt{\pi}}\int\limits_{z_0}^{z_1} e^{-z^2}dz=\frac{1}{2}\left[\Phi(z_1)-\Phi(z_0)\right]$$

более, чем вероятность неравенства

$$z_0\sqrt{2npq}+\frac{q-p}{3}z_0^2+1\leq m-np<z_1\sqrt{2npq}+\frac{q-p}{3}z_1^2+1$$

и менее, чем вероятность неравенства

$$z_0\sqrt{2npq}+\frac{q-p}{3}z_0^2-\frac{1}{2}\leq m-np<z_1\sqrt{2npq}+\frac{q-p}{3}z_1^2-\frac{1}{2}$$

Для того, чтобы получить аналогичные выводы для случая, когда z_0 и z_1 противоположных знаков, следует воспользоваться сначала неравенством (8), при помощи которого совместно с (22) легко вывести, что вероятность неравенства $m\geq m_0$ равна

$$\frac{1}{2}\left[\Phi\left(\sqrt[6]{2npq}\right)-\Phi(z_0)\right]+\Theta e^{-\sqrt[3]{2npq}},$$

где $o<\Theta<1$, а z_0 есть положительный корень уравнения

$$m_0+\alpha=np+z_0\sqrt{2npq}+z_0^2\frac{q-p}{3},\ \left(\text{где}-1<\alpha<\frac{1}{2}\right).$$

Отсюда можно заключить, что вероятность неравенства

$$\left|m-np-\frac{t_0^2}{3}(q-p)\right|<t_0\sqrt{2npq}, \tag{23}$$

рассматриваемого с точностью до одной единицы, имеет вероятностью $\Phi(t_0)$ с точностью до

$$2e^{-\sqrt[3]{2npq}}\ (\text{при}\ t_0\leq\sqrt[6]{2npq},\ npq\geq 365).$$

Полагая, например, $p=\frac{2}{5}$, $n=28750$ находим, что вероятность неравенства

$$11322 < m < 11678$$

$$\text{более, чем } \Phi(1{,}5)-\frac{1}{10^{10}} \doteq 0{,}966105;$$

вероятность же неравенства

$$11324 < m < 11676$$

$$\text{менее, чем } \Phi(1{,}5)+\frac{1}{10^{10}} \doteq 0{,}966105.$$

§ 5. В заключение сопоставим приближенное выражение $P'_{m,n}$ данное Пирсоном [1]) для $P_{m,n}$ с произведенными нами вычислениями.

По Пирсону,

$$P'_{m,n}=\left[1+\frac{p-q}{2pqn}(np-m)\right]^{\frac{4pqn}{(p-q)^2}} e^{-\frac{2}{p-q}(np-m)}\lambda_n\,,$$

где λ_n независимый от m коэффициент.

Полагая $np-m=t\sqrt{2npq}$, мы нашли точное значение

$$\lambda_n\, P_{m,n}=P'_{m,n}\; e^{-\left[\frac{\Theta n}{12m(n-m)}+\frac{t^4(1+\varepsilon)}{12pqn}\right]}\sqrt{\frac{n}{2\pi m(n-m)}},$$

где

$$\left|\varepsilon\right|<\frac{1}{2},\ 0<\Theta<1,\ \left|t\right|<\frac{1}{8}\sqrt{2pqn}\,,$$

Таким образом, функция Пирсона

$$\left(1+\frac{(p-q)t}{\sqrt{2pqn}}\right)^{\frac{4pqn}{(p-q)^2}} e^{-\frac{2t\sqrt{2pqn}}{p-q}}$$

выражает $P_{m,n}$ с тою же степенью точности [2]), как и e^{-z^2}, где z определяется из равенства

$$z\sqrt{2npq}+\frac{q-p}{3}z^2=m-np=-t\sqrt{2npq}.$$

---

[1]) Philosophical Transactions. Vol 186.

[2]) См. выноску на стр. 44.

Следовательно, пределы, в которых формула Пирсона дает точны[...] результаты, определяются еще требованием, чтобы $\frac{t^4}{12pqn}$ стремилось к 0. Поэтому формула Пирсона перестает быть удовлетворительной, когд[...] $\frac{(np-m)^{4/3}}{2npq}$ не очень мало. Например, полагая $p=\frac{3}{4}$, $q=\frac{1}{4}$, $n=20$, по Пирсону: $\frac{P_{5,20}}{P_{15,20}}=\left(1+\frac{2}{3}\right)^{60}e^{-40}=87.10^{-5}$, между тем, как в дей-ствительности $\frac{P_{5,20}}{P_{15,20}}=17.10^{-5}$; кроме того, с увеличением n, ошибк[...] быстро возрастает.

RESUMÉ

Le résultat principal contenu dans cette Note consiste dans la dé-monstration de la proposition suivante: Soient x, y, z,.. des quantités indé-pendantes dont les èspérances mathématiques sont nulles; soient a_k, b_k, c_k etc. les espérances mathématiques respectives de x^k, y^k, z^k etc. Dans ces conditions la probabilité de l'inégalité

$$\left|(x+y+z+\ldots)-\frac{t^2}{3}\frac{A_3}{A_2}\right|<t\sqrt{2A_2}\left[1+\frac{A_4t^2}{6A_2^2}\right], \qquad (1)$$

où $A_2=a_2+b_2+c_2+\ldots$; $A_3=a_3+b_3+c_3+\ldots$; $A_4=a_4+b_4+c_4+\ldots$

est supérieure à $1-2e^{-t^2}$, pour toute valeur positive de $t\leq\frac{1}{4L}\sqrt{2A_2}$, le nombre L étaut une borne supérieure des quantités

$$\sqrt[K-4]{\frac{4!}{K!}\left|\frac{a_k}{a_4}\right|},\quad \sqrt[K-4]{\frac{4!}{K!}\left|\frac{b_k}{b_4}\right|},\ldots.$$

Ensuite, je demontre que pour n assez grand la probabilité $P_{m,n}$ que l'évenement ayant la probabilité p se produit m fois dans n expérience[s] se rapproche le plus de $\sqrt{\frac{n}{2\pi m(n-m)}}$, e^{-z^2}, lorsque m est l'entier l[e] plus approché de $np+z\sqrt{2npq+\frac{z^2(q-p)}{3}}$. (2)

J'en déduis que la probabilité que

$$m_0\leq m<m_1$$

est égale â $\frac{1}{\sqrt{\pi}} \int_{z_0}^{z_1} e^{-z^2} dz$,

où $m_0 + \alpha = np + z_0 \sqrt{2npq} + \frac{q-p}{3} z_0^2$ (3)

$m_1 + \alpha = np + z_1 \sqrt{2npq} + \frac{q-p}{3} z_1^2$

$$\text{avec} -1 < \alpha < \frac{1}{2}, \text{ si } npq \geq 365, \; z_0 \geq \frac{2}{\sqrt{2npq}}, \; z_1 \leq \sqrt[6]{2npq}$$

En tenant compte de l'inégalité (3), on montre que l'inégalité $\left| m - np - \frac{t_0^2}{3}(q-p) \right| < t_0 \sqrt{2npq}$, considéreé à une unité pres, a la probabilité $\Phi(t_0)$, à $2e^{-\sqrt[3]{2npq}}$ prés, $\left(\text{si } t_0 \leq \sqrt[6]{2npq}, \; npq \geq 365 \right)$.

La Note se termine par la remarque que la formule approhée de Pearson

$$P_{m,n} = \left[1 - \frac{p-q}{2pqn}(np-m) \right]^{\frac{4pq}{(p-q)^2}} e^{-\frac{2}{p-q}(np-m)} \lambda_n$$

ne donne de valeurs approcheès de $P_{m,n}$ que tant que $\frac{(np-m)^{4/3}}{2pqn}$ tend vers 0 et dans ce cas elle fournit la même approximation que la formule (2).