В изложении доцента БГТУ Сахарова В.Ю.

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.

Нужно найти вероятность p.

Сразу отметим, что задание невыполнимо.

Можно предложить несколько вариантов ответа. Правильным логично признать тот ответ, который будет лучше остальных во всех смыслах.

Напрашивается формула $\widehat{p} = \dfrac{K}{N}$. Кажется удивительным, но можно предложить альтернативную идею

$$\widehat{p} = \frac{K + \frac{\sqrt{N}}{2}}{N + \sqrt{N}}$$

Убедимся, что эта идея корректна в том смысле, что если вычислить по аналогичной формуле вероятность противоположного события

$$1-\widehat{p}=\frac{N-K+\frac{\sqrt{N}}{2}}{N+\sqrt{N}}$$

, то в сумме действительно будет единица:

$$\frac{K+\frac{\sqrt{N}}{2}}{N+\sqrt{N}}+\frac{N-K+\frac{\sqrt{N}}{2}}{N+\sqrt{N}}=$$

$$=\frac{K+\frac{\sqrt{N}}{2}+N-K+\frac{\sqrt{N}}{2}}{N+\sqrt{N}}=$$

$$=\frac{N+2\cdot\frac{\sqrt{N}}{2}}{N+\sqrt{N}}=\frac{N+\sqrt{N}}{N+\sqrt{N}}=1$$

Верно.

Теперь одновременно согласимся с двумя мнениями.

Первое: очень бы хотелось, чтобы математическое ожидание оценки было бы равно оцениваемому параметру (это свойство называется несмещённость).

Второе: из двух оценок лучшей надо признать ту, для которой мера её отклонения от оцениваемого параметра меньше (в этом случае говорят, что эта оценка эффективнее).

Теперь посмотрим, что получится для каждой из оценок.

Проверим на несмещённость первую формулу.

$$\widehat{p} = \frac{K}{N}$$

$$E\widehat{p} = E\left(\frac{K}{N}\right) = E\left(\frac{1}{N}\cdot K\right) =$$

Константа-множитель может быть вынесена за знак математического ожидания.

$$= \frac{1}{N}\cdot E(K) =$$

Поскольку в данном случае K это число успехов в серии из N независимых испытаний с постоянной вероятностью успеха p в каждом из испытаний, то его математическое ожидание равно $EK=Np$.

$$= \frac{1}{N} \cdot Np = p$$

. Свойство несмещённости выполняется. Теперь проверим на несмещённость альтернативную оценку.

$$\widehat{p} = \frac{K + \frac{\sqrt{N}}{2}}{N + \sqrt{N}}$$

$$E\widehat{p} = E\left(\frac{K + \frac{\sqrt{N}}{2}}{N + \sqrt{N}} \right) =$$

$$= E\left(\frac{1}{N + \sqrt{N}} \cdot \left(K + \frac{\sqrt{N}}{2} \right) \right) =$$

Выносим константу-множитель за знак математического ожидания.

$$= \frac{1}{N+\sqrt{N}} \cdot E\left(K + \frac{\sqrt{N}}{2}\right) =$$

Известно, что математическое ожидание суммы равно сумме математических ожиданий.

$$= \frac{1}{N+\sqrt{N}} \cdot \left(E(K) + E\left(\frac{\sqrt{N}}{2}\right)\right) =$$

То, что *EK*=*Np* уже отмечали, а математическое ожидание константы равно этой константе.

$$= \frac{1}{N+\sqrt{N}} \cdot \left(Np + \frac{\sqrt{N}}{2}\right) =$$

Уже очевидно, что полученное выражение отлично от *p*, и свойство несмещённости не выполнено, но сделаем ещё несколько преобразований, чтобы было видно, насколько результат отличается от желаемого.

$$= \frac{Np + \frac{\sqrt{N}}{2}}{N + \sqrt{N}} =$$

$$= \frac{Np + p\sqrt{N} + \frac{\sqrt{N}}{2} - p\sqrt{N}}{N + \sqrt{N}} =$$

$$= \frac{Np + p\sqrt{N}}{N + \sqrt{N}} + \frac{\frac{\sqrt{N}}{2} - p\sqrt{N}}{N + \sqrt{N}} =$$

$$= \frac{p\left(N + \sqrt{N}\right)}{N + \sqrt{N}} + \frac{\sqrt{N}\left(\frac{1}{2} - p\right)}{\sqrt{N}\left(\sqrt{N} + 1\right)} =$$

$$= p + \frac{\frac{1}{2} - p}{\sqrt{N} + 1}$$

Обратим внимание, что хотя это выражение и не совпадает с *p*, но его предел при *N* стремящимся к бесконечности всё же равен *p*.

$$\lim_{N\to\infty}\widehat{p} = \lim_{N\to\infty}\left(p + \frac{\frac{1}{2} - p}{\sqrt{N}+1}\right) = p + 0 = p$$

Это «утешительное» свойство называется асимптотическая несмещённость.

Теперь посмотрим, какая из двух оценок эффективнее. В качестве меры отклонения от *p* разумно принять выражение, аналогичное выражению для дисперсии $E\left(\widehat{p}-p\right)^2$, в котором вычитаемое равно той величины меру отклонения от которой находят.

Итак, в первом случае

$$E\left(\widehat{p}-p\right)^2 = E\left(\frac{K}{N} - p\right)^2 =$$

$$= E\left(\left(\frac{K}{N} - p\right)^2\right) =$$

$$= E\left(\left(\frac{K}{N}\right)^2 - 2\frac{K}{N}p + p^2\right) =$$

Пользуемся тем, что математическое ожидание суммы (разности) равно сумме (разности) математических ожиданий.

$$= E\left(\left(\frac{K}{N}\right)^2\right) - 2 \cdot E\left(\frac{K}{N}p\right) + E\left(p^2\right) =$$

Понимаем, что *p* это константа. И знаем, что математическое ожидание константы равно этой константе.

$$= E\left(\frac{1}{N^2} \cdot \left(K^2\right)\right) - 2 \cdot E\left(\frac{p}{N} \cdot K\right) + p^2 =$$

Постоянный множитель можно выносить за знак математического ожидания.

$$=\frac{1}{N^2}\cdot E\left(K^2\right)-2\cdot\frac{p}{N}\cdot E\left(K\right)+p^2=$$

То, что $EK=Np$ уже вспоминали, а как найти $E(K^2)$?

Для этого нужно знать дисперсию $DK=Np(1-p)$ и формулу $E(K^2)=DK+(EK)^2$.

Тогда $E(K^2)=Np(1-p)+(Np)^2$.

$$=\frac{1}{N^2}\cdot\left(Np(1-p)+(Np)^2\right)-\frac{p}{N}\cdot Np+p^2=$$

$$=\frac{Np(1-p)}{N^2}+\frac{N^2p^2}{N^2}-2p^2+p^2=$$

$$=\frac{p(1-p)}{N}+p^2-p^2=\frac{p(1-p)}{N}$$

Во втором случае

$$E\left(\widehat{p}-p\right)^2=E\left(\frac{K+\frac{\sqrt{N}}{2}}{N+\sqrt{N}}-p\right)^2=$$

$$= E\left(\left(\frac{K + \frac{\sqrt{N}}{2}}{N + \sqrt{N}} - p\right)^2\right) =$$

$$= E\left(\left(\frac{K + \frac{\sqrt{N}}{2} - p\left(N + \sqrt{N}\right)}{N + \sqrt{N}}\right)^2\right) =$$

$$= E\left(\left(\frac{K}{N + \sqrt{N}} + \frac{\frac{\sqrt{N}}{2} - p\left(N + \sqrt{N}\right)}{N + \sqrt{N}}\right)^2\right) =$$

Раскроем квадрат и одновременно воспользуемся тем, что математическое ожидание суммы (разности) равно сумме (разности) математических ожиданий.

$$= E\left(\left(\frac{K}{N+\sqrt{N}}\right)^2\right)+$$

$$+ E\left(2\frac{K}{N+\sqrt{N}}\cdot\frac{\frac{\sqrt{N}}{2}-p\left(N+\sqrt{N}\right)}{N+\sqrt{N}}\right)+$$

$$+ E\left(\left(\frac{\frac{\sqrt{N}}{2}-p\left(N+\sqrt{N}\right)}{N+\sqrt{N}}\right)^2\right)=$$

Вынесем множители-константы за знаки математических ожиданий и воспользуемся тем, что математическое ожидание константы есть сама эта константа (всё кроме *K* это постоянные).

$$=\frac{1}{\left(N+\sqrt{N}\right)^2}\cdot E\left(\left(K\right)^2\right)+$$

$$+\frac{2}{N+\sqrt{N}}\cdot\frac{\frac{\sqrt{N}}{2}-p\left(N+\sqrt{N}\right)}{N+\sqrt{N}}E\left(K\right)+$$

$$+\left(\frac{\frac{\sqrt{N}}{2}-p\left(N+\sqrt{N}\right)}{N+\sqrt{N}}\right)^2=$$

Помним, что $EK=Np$ и $E(K^2)=Np(1-p)+(Np)^2$.

$$=\frac{1}{\left(N+\sqrt{N}\right)^2}\cdot\left(Np(1-p)+(Np)^2\right)+$$

$$+\frac{2}{N+\sqrt{N}}\cdot\frac{\frac{\sqrt{N}}{2}-p\left(N+\sqrt{N}\right)}{N+\sqrt{N}}Np+$$

$$+\left(\frac{\frac{\sqrt{N}}{2}-p\left(N+\sqrt{N}\right)}{N+\sqrt{N}}\right)^2=$$

$$=\frac{Np(1-p)+(Np)^2}{\left(N+\sqrt{N}\right)^2}+$$

$$+\frac{\sqrt{N}-2p\left(N+\sqrt{N}\right)}{\left(N+\sqrt{N}\right)^2}Np+$$

$$+\frac{\left(\frac{\sqrt{N}}{2}-p\left(N+\sqrt{N}\right)\right)^2}{\left(N+\sqrt{N}\right)^2}=$$

Каждая из дробей может быть сокращена на *N*. Разумеется, при этом внутри квадратов в знаменателях и последнем из числителей сокращение произойдёт на корень из *N*.

$$= \frac{p(1-p)+Np^2}{\left(\sqrt{N}+1\right)^2} +$$

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

$$+ \frac{\left(\frac{1}{2}-p\left(\sqrt{N}+1\right)\right)^2}{\left(\sqrt{N}+1\right)^2} =$$

$$= \frac{p-p^2+Np^2}{\left(\sqrt{N}+1\right)^2} +$$

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

$$+ \frac{\frac{1}{4}-2\cdot\frac{1}{2}\cdot p\left(\sqrt{N}+1\right)+p^2\left(\sqrt{N}+1\right)^2}{\left(\sqrt{N}+1\right)^2} =$$

$$=\frac{p-p^2+Np^2+p\sqrt{N}-2p^2N-2p^2\sqrt{N}}{\left(\sqrt{N}+1\right)^2}+$$

$$+\frac{\frac{1}{4}-p\sqrt{N}-p+p^2\left(N+2\sqrt{N}+1\right)}{\left(\sqrt{N}+1\right)^2}=$$

Приведём подобные в числителях (знаменатели одинаковые).

$$=\frac{-p^2-p^2N-2p^2\sqrt{N}}{\left(\sqrt{N}+1\right)^2}+$$

$$+\frac{\frac{1}{4}+p^2N+2p^2\sqrt{N}+p^2}{\left(\sqrt{N}+1\right)^2}=$$

$$=\frac{\frac{1}{4}}{\left(\sqrt{N}+1\right)^2}=\frac{1}{4\left(\sqrt{N}+1\right)^2}$$

Теперь подумаем над вопросом, что меньше

$$\frac{p(1-p)}{N} \quad \text{или} \quad \frac{1}{4\left(\sqrt{N}+1\right)^2}.$$

Вспомним, что p это вероятность и поэтому $p \in [0;1]$.

Функция $f(p)=p(1-p)=p-p^2$ квадратичная, с отрицательным старшим коэффициентом. Поэтому она достигает своего наибольшего значения в точке, находящийся ровно между корнями. Корни очевидно равны нулю и единице и поэтому наибольшее значение равно значению функции в одной второй $\max f(p)=f(½)=½(1-½)=½\cdot½=¼$.

Итак, наибольшее значение меры отклонения от оцениваемого параметра в первом случае равно

$\frac{1}{4N}$, а во втором случае оно не зависит от настоящего значения p и равно

$$\frac{1}{4\left(\sqrt{N}+1\right)^2}=\frac{1}{4\left(N+2\sqrt{N}+1\right)}$$

Очевидно, что знаменатель второй дроби больше, поэтому сама дробь меньше.

В итоге получилось, что первая формула в отличие от второй обладает свойством несмещённости, но зато вторая формула эффективнее.