Свойства оценок параметров распределений.

Часто о случайной величине судят не по ее распределению, а по различного рода числовым характеристикам. Обычно это бывают математическое ожидание и дисперсия (или же среднеквадратическое отклонение - корень из дисперсии).

Приведу формулы подсчета матожидания и дисперсии для дискретной случайной величины (для непрерывной нужно заменить суммы на интегралы):

Матожидание: $m=\sum_i x_i p_i$, где x_i - значение случайной величины, а p_i - вероятность, с которой случайная величина принимает значение x_i .

Дисперсия: $d=\sum_i (x_i-m)^2 p_i$.

Понятно, что если закон распределения случайной величины неизвестен, то мы не можем вычислить ни матожидания, ни дисперсии. В этом случае их можно попытаться оценить по выборке из генеральной совокупности. Но оценки могут быть разными — какая-то лучше, какая-то хуже. Для того, чтобы их классифицировать по полезным свойствам выделяют следующие типы оценок:

- Несмещенная оценка
- Эффективная оценка
- Состоятельная оценка

Итак, что же это такое? Пойдем по порядку.

Во-первых, оценка какой-либо числовой характеристики выборки (например, оценка матожидания) является функцией от выборки[1], то есть оценка тоже является случайной величиной. Следовательно, для самой оценки можно подсчитать матожидание и дисперсию.

Во-вторых, подумаем, чего бы мы хотели от оценок? Вполне разумных требований:

- Оценка не должна содержать систематической ошибки. Это означает, что ее математическое ожидание должно совпадать с оцениваемым параметром. **Такая оценка называется несмещенной.**
- Оценка должна приближаться к оцениваемому параметру с увеличением объема выборки (то есть подчиняться закону больших чисел). **Если оценка стремится по вероятности к оцениваемому параметру, то она называется состоятельной.**
- Наконец, разные оценки имеют разные дисперсии. Чем меньше дисперсия (у несмещенной) оценки, тем точнее по ней можно судить об интересующем параметре. **Оценка, имеющая минимальную дисперсию из всех возможных несмещенных оценок называется эффективной (величина выборки фиксирована).**

Фишка состоит в том, что бывает так, что нельзя указать оценку, которая является одновременно несмещенной, состоятельной и эффективной.

Примеры оценок.

Оценивать матожидание m_x случайной величины x по выборке можно разными способами, например:

1 Понятно, что для разных выборок оценки в общем случае будут разными по величине.

1) $m_x = \frac{1}{n} \sum_i x_i$

2) $m_x = \frac{x_{min} + x_{max}}{2}$

3) Можно придумать еще множество вариантов оценок.

Обычно пользуются оценкой под номером 1. Почему? Да потому, что доказано, что эта оценка является несмещенной и состоятельной. А вот эффективной она (в общем случае) не является. Тем не менее, если распределение случайной величины x нормальное, то $m_x = \frac{1}{n} \sum_i x_i$ будет несмещенной, состоятельной и эффективной.

Перейдем к дисперсии.

Часто при оценке дисперсии используют:

1) $S^2 = \frac{1}{n} \sum_i (x - m_x)^2$

2) $\hat{S}^2 = \frac{1}{n-1} \sum_i (x - m_x)^2$

Вторая оценка, в отличие от первой, является несмещенной. Зато обе эти оценки являются состоятельными, но не являются эффективными. Кстати, разница между $\hat{S}^2$ и S^2 при больших n почти не заметна (например, при $n=50$ разница будет составлять 2%).

А вот состоятельной, несмещенной и эффективной оценкой будет оценка:

$S_@^2 = \frac{1}{n} \sum_i (x - m)^2$, где m — известное математическое ожидание.