Лекция 4

Точечные оценки для математического ожидания и дисперсии

Пусть $(X_1, \ldots, X_n)$ – выборка, порожденная СВ X с неизвестным распределением. В качестве оценки неизвестного математического ожидания $M\{X\}$ рассмотрим статистику

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n},$$

называемую выборочным средним.

Утверждение 1 $\overline{X}_n$ есть несмещённая оценка для $M\{X\}$.

Док-во:

$$M\{\overline{X}_n\} = M\left\{\frac{X_1 + \ldots + X_n}{n}\right\} = \frac{1}{n}(MX_1 + \ldots + MX_n) =$$

$$= \frac{1}{n}(\underbrace{MX + \ldots + MX}_{n \text{ слагаемых}}) = MX, \text{ что и т.д.}$$

Утверждение 2. $\overline{X}_n$ есть состоятельная оценка для MX.

Док-во:

Из утверждения 1 следует $\lim\limits_{n \to \infty} M\{\overline{X}_n\} = MX$

Далее, дисперсия

$$D\{\overline{X}_n\} = D\left\{\frac{X_1 + \ldots + X_n}{n}\right\} = \frac{1}{n^2} D\{X_1 + \ldots + X_n\} =$$

$$= \left\{\begin{matrix}\text{в силу} \\ \text{независи-} \\ \text{мости} \\ X_1, \ldots, X_n\end{matrix}\right\} = \frac{1}{n^2}(DX_1 + \ldots + DX_n) = \frac{1}{n^2}(\underbrace{DX + \ldots + DX}_{n \text{ слагаемых}}) =$$

$= \frac{1}{n} DX \xrightarrow[n\to\infty]{} 0$ (предполагаем, что DX существует).

Таким образом, выполнены достаточные условия состоятельности (см. лекцию 3) $\Longrightarrow$ $\overline{X}_n$ есть состоятельная оценка для MX.

<u>Утверждение</u> 3 $\overline{X}_n$ является эффективной оценкой для $M\{X\}$ в классе всех линейных несмещенных оценок.

Док-во:

Рассмотрим произвольную линейную оценку

$$\hat{X}_n = c_1X_1 + \ldots + c_nX_n,$$

где $c_1,\ldots,c_n$ — некоторые константы. Тогда

$$M\{\hat{X}_n\} = (c_1 + \ldots + c_n)\cdot M\{X\}.$$

Поскольку рассматриваются только несмещенные оценки, то $c_1 + c_2 + \ldots + c_n = 1$. Для дисперсии $D\{\hat{X}_n\}$ получим

$$D\{\hat{X}_n\} = (c_1^2 + \ldots + c_n^2)\cdot DX$$ (предполагаем, что DX существует)

Минимальное значение дисперсии $D\{\hat{X}_n\}$ получится при таких $c_1,\ldots,c_n$, которые являются решением задачи на условный экстремум:

$$\begin{cases} c_1^2 + \ldots + c_n^2 \longrightarrow \min \\ \text{при условии } c_1 + \ldots + c_n = 1, \\ c_i \in (-\infty, +\infty)\ \forall i = 1,\ldots,n. \end{cases}$$

Эта задача легко решается: минимум достигается при

$$c_1 = c_2 = \ldots = c_n = \frac{1}{n}.$$

Таким образом, минимальное значение дисперсии $D\{\hat{X}_n\}$ получается, когда

$$\hat{X}_n = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n,$$

т.е. когда $\hat{X}_n$ совпадает с $\overline{X}_n$. Утверждение 3 доказано.

Рассмотрим теперь оценки для дисперсии $D\{X\}$. Будем сначала считать, что математическое ожидание MX известно: $MX = m$, m - известное число. В качестве оценки для $D\{X\}$ рассмотрим статистику

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2.$$

<u>Утверждение 4</u> S_n есть несмещённая оценка для $D\{X\}$.

Док-во:

$$M\{S_n\} = \frac{1}{n}\sum_{i=1}^{n} M\{(X_i-m)^2\} =$$

$$= \frac{1}{n}\Big(\underbrace{M(X-m)^2 + \ldots + M(X-m)^2}_{n \text{ слагаемых}}\Big) = M(X-m)^2 = DX,$$

что и т.д.

Утверждение 5. S_n есть состоятельная оценка для $D\{X\}$.

Док-во:

Достаточно показать, что $D\{S_n\} \xrightarrow[n\to\infty]{} 0$ (см. лекцию 3). Имеем:

$$D\{S_n\} = D\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2\right\} =$$

$$= \frac{1}{n^2} D\left\{\sum_{i=1}^{n}(X_i-m)^2\right\} = \left\{\begin{matrix}\text{в силу неза-}\\ \text{висимости}\\ X_1,\dots,X_n\end{matrix}\right\} =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} D\{(X_i-m)^2\} = \frac{1}{n^2}\Big(\underbrace{D\{(X-m)^2\}+\ldots+D\{(X-m)^2\}}_{n \text{ слагаемых}}\Big) =$$

$$= \frac{1}{n} D\{(X-m)^2\} \xrightarrow[n\to\infty]{} 0,$$

если $D\{(X-m)^2\}$ существует (для этого достаточно потребовать существование начальных моментов $M\{X^m\}$ при $m \le 4$).

Пусть теперь математическое ожидание $M\{X\}$ неизвестно. В этом случае в качестве оценки для DX рассмотрим статистику

$$S_n^{new} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2,$$

где $\overline{X}_n = \frac{1}{n}(X_1+\ldots+X_n)$ – выборочное среднее. Исследуем эту оценку на несмещённость:

$$M\{S_n^{new}\} = M\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right\} =$$

$$= \frac{1}{n}\sum_{i=1}^{n} M\{(X_i-\overline{X}_n)^2\} =$$

поскольку математическое ожидание $M\{(X_i-\overline{X}_n)^2\}$ одинаково для всех $i=1,\dots,n$

$$= M\{(X_1-\overline{X}_n)^2\} = M\{X_1^2 - 2X_1\overline{X}_n + \overline{X}_n^2\} =$$

$$= MX_1^2 - 2M\left\{X_1\cdot\frac{X_1+\dots+X_n}{n}\right\} + M\{\overline{X}_n^2\} =$$

$$= DX_1 + (MX_1)^2 - \frac{2}{n}M\{X_1^2 + X_1X_2 + \dots + X_1X_n\} + D\{\overline{X}_n\} + (M\overline{X}_n)^2$$

$$= DX + (MX)^2 - \frac{2}{n}\left(MX_1^2 + MX_1\cdot MX_2 + \dots + MX_1\cdot MX_n\right) + \frac{DX}{n} + (MX)^2 =$$

$$= DX + (MX)^2 - \frac{2}{n}\left(MX^2 + (n-1)(MX)^2\right) + \frac{DX}{n} + (MX)^2 =$$

$$= DX + (MX)^2 - \frac{2}{n}\left(DX + (MX)^2 + (n-1)(MX)^2\right) + \frac{DX}{n} + (MX)^2 =$$

$$= DX + (MX)^2 - \frac{2}{n}\left(DX + n(MX)^2\right) + \frac{DX}{n} + (MX)^2 =$$

$$= DX + (MX)^2 - \frac{2}{n}DX - 2(MX)^2 + \frac{DX}{n} + (MX)^2 =$$

$$= DX - \frac{1}{n}DX = \frac{n-1}{n}DX \neq DX, \text{ т.е.}$$

Оценка S_n^{new} для DX является смещенной, но при этом она авляется асимптотически несмещенной, ибо

$$\lim_{n\to\infty} M\{S_n^{new}\} = \lim_{n\to\infty} \frac{n-1}{n} DX = DX.$$

Если ввести ещё одну оценку

$$S_n^{ucn} = \frac{n}{n-1} S_n^{new} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

для дисперсии $D\{X\}$, то эта оценка будет несмещённой:

$$M\{S_n^{ucn.}\} = \frac{n}{n-1} M\{S_n^{new}\} = DX.$$

Оценку S_n^{new} называют выборочной дисперсией, а оценку S_n^{ucn} — исправленной выборочной дисперсией.

Утверждение 6 Оценка $S_n^{ucn.}$ является состоятельной оценкой для DX.

Док-во:

Можно проверить, что

$$D\{S_n^{ucn.}\} = \frac{1}{n}\left(M\{(X-MX)^4\} - \frac{n-3}{n-1}(DX)^2 \right),$$

так что выполнены достаточные условия состоятельности оценки (см. лекцию 3).

Заметим, что S_n^{new} тоже будет состоятельной оценкой для $D\{X\}$.